

Statistical Evaluation of Metaproteomics and 16S rRNA Amplicon Sequencing Techniques for Study of Gut Microbiota Establishment in Infants with Cystic Fibrosis

Saralegui, Claudia; García-Durán, Carmen; Romeu, Eduardo; Hernández-Sánchez, María Luisa; Maruri, Ainhize; Bastón-Paz, Natalia; Lamas, Adelaida; Vicente, Saioa; Pérez-Ruiz, Estela; Delgado, Isabel; Luna-Paredes, Carmen; de Dios Caballero, Juan; Zamora, Javier; Monteoliva, Lucía; Gil, Concepción; del Campo, Rosa

DOI:

[10.1128/spectrum.01466-22](https://doi.org/10.1128/spectrum.01466-22)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Saralegui, C, García-Durán, C, Romeu, E, Hernández-Sánchez, ML, Maruri, A, Bastón-Paz, N, Lamas, A, Vicente, S, Pérez-Ruiz, E, Delgado, I, Luna-Paredes, C, de Dios Caballero, J, Zamora, J, Monteoliva, L, Gil, C & del Campo, R 2022, 'Statistical Evaluation of Metaproteomics and 16S rRNA Amplicon Sequencing Techniques for Study of Gut Microbiota Establishment in Infants with Cystic Fibrosis', *Microbiology spectrum*, vol. 10, no. 6, e01466-22. <https://doi.org/10.1128/spectrum.01466-22>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy



While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 29. May. 2023



Statistical Evaluation of Metaproteomics and 16S rRNA Amplicon Sequencing Techniques for Study of Gut Microbiota Establishment in Infants with Cystic Fibrosis

Claudia Saralegui,^{a,b,c} Carmen García-Durán,^d Eduardo Romeu,^e María Luisa Hernández-Sánchez,^e Ainhize Maruri,^{a,b,c} Natalia Bastón-Paz,^{a,b,c} Adelaida Lamas,^{c,f} Saioa Vicente,^{c,f} Estela Pérez-Ruiz,^g Isabel Delgado,^h Carmen Luna-Paredes,ⁱ Juan de Dios Caballero,^{a,b,c} Javier Zamora,^j  Lucía Monteoliva,^{d,e} Concepción Gil,^{d,e}  Rosa del Campo^{a,b,c,k}

^aServicio de Microbiología, Hospital Universitario Ramón y Cajal and Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain

^bCIBERINFEC, Madrid, Spain

^cUnidad de Fibrosis Quística, Hospital Universitario Ramón y Cajal, Madrid, Spain

^dDepartamento de Microbiología y Parasitología, Universidad Complutense de Madrid and Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain

^eUnidad de Proteómica, Universidad Complutense de Madrid, Madrid, Spain

^fServicio de Pediatría, Hospital Universitario Ramón y Cajal and Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain

^gUnidad de Fibrosis Quística, Hospital Regional Universitario de Málaga, Málaga, Spain

^hUnidad de Fibrosis Quística, Hospital Virgen del Rocío, Seville, Spain

ⁱSección de Neumología y Alergia Infantil, Unidad Multidisciplinar Fibrosis Quística, Hospital Doce de Octubre, Madrid, Spain

^jUnidad de Bioestadística, Hospital Universitario Ramón y Cajal and Instituto Ramón y Cajal de Investigación Sanitaria and CIBERESP, Madrid, Spain

^kUniversidad Alfonso X El Sabio, Madrid, Spain

Claudia Saralegui and Carmen García-Durán contributed equally to this work. The order of the authors was decided by their contributions.

ABSTRACT Newborn screening for cystic fibrosis (CF) can identify affected but asymptomatic infants. The selection of omic technique for gut microbiota study is crucial due to both the small amount of feces available and the low microorganism load. Our aims were to compare the agreement between 16S rRNA amplicon sequencing and metaproteomics by a robust statistical analysis, including both presence and abundance of taxa, to describe the sequential establishment of the gut microbiota during the first year of life in a small size sample (8 infants and 28 fecal samples). The taxonomic assignments by the two techniques were similar, whereas certain discrepancies were observed in the abundance detection, mostly the lower predicted relative abundance of *Bifidobacterium* and the higher predicted relative abundance of certain *Firmicutes* and *Proteobacteria* by amplicon sequencing. During the first months of life, the CF gut microbiota is characterized by a significant enrichment of *Ruminococcus gnavus*, the expression of certain virulent bacterial traits, and the detection of human inflammation-related proteins. Metaproteomics provides information on composition and functionality, as well as data on host-microbiome interactions. Its strength is the identification and quantification of *Actinobacteria* and certain classes of *Firmicutes*, but alpha diversity indices are not comparable to those of amplicon sequencing. Both techniques detected an aberrant microbiota in our small cohort of infants with CF during their first year of life, dominated by the enrichment of *R. gnavus* within a human inflammatory environment.

IMPORTANCE In recent years, some techniques have been incorporated for the study of microbial ecosystems, being 16S rRNA gene sequencing being the most widely used. Metaproteomics provides the advantage of identifying the interaction between microorganisms and human cells, but the available databases are less extensive as well as imprecise. Few studies compare the statistical differences between the two techniques to define the composition of an ecosystem. Our work shows that the two methods are comparable in terms of microorganism identification but provide different results in alpha diversity analysis. On the other hand, we have studied newborns with cystic fibrosis, for whom we have

Editor Joanna B. Goldberg, Emory University School of Medicine

Copyright © 2022 Saralegui et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Concepción Gil, conchagil@ucm.es, or Rosa del Campo, rosa.campo@salud.madrid.org.

The authors declare a conflict of interest. Rosa del Campo is recipient of a Vertex grant. The remaining authors have no conflict of interest.

[This article was published on 18 October 2022 with information missing from Acknowledgments. This information was added in the current version, posted on 15 November 2022.]

Received 21 April 2022

Accepted 23 September 2022

Published 18 October 2022

described the establishment of an intestinal ecosystem marked by the inflammatory response of the host and the enrichment of *Ruminococcus gnavus*.

KEYWORDS amplicon sequencing, metaproteomics, cystic fibrosis, gut microbiota establishment, *Ruminococcus gnavus*, Bland-Altman test

The establishment of the microbiota at birth affects human health during the rest of life (1). In cystic fibrosis (CF), a genetic disease diagnosed by neonatal screening, the colonization of the lung airways by pathogenic bacteria impacts survival (2). The life expectancy of individuals with CF has increased considerably in recent years, and comorbidities beyond those of the respiratory tract have been revealed, mostly related to the digestive tract (3). The metabolic activity of the gut microbiota has a direct impact on nutritional status, a relevant aspect in CF since most of the patients are malnourished, and proper nutrition has a major effect on their infectious lung disease (4).

In CF, the gut microbiota is conditioned by the following factors: transmembrane regulator dysfunction, which creates a thick mucus layer; the acidic environment caused by impaired bicarbonate secretion; maldigestion; malabsorption; steatorrhea due to pancreatic insufficiency; and the high intake of antibiotics for bacterial infections during pulmonary exacerbations (5, 6). Significant differences in the microbiota's bacterial composition and alpha diversity have been reported compared with that in healthy infants (7–9).

Currently, 16S rRNA amplicon sequencing (AS) is the most extensive, rapid, and cost-effective technique for characterizing complex bacterial communities, such as the gut microbiota (10). The technique's main weakness results from the variable affinity of the primers for flanking regions; consequently, a higher predicted relative abundance has been found for certain genera, whereas others, mainly *Bifidobacterium*, are consistently reported to have a lower predicted relative abundance than other methods (11). Even with optimized primers, amplicon sequencing often provides deficient taxonomic assignment at the genus and species levels due to the short length of the amplicons. On the other hand, metaproteomics applies high-resolution mass spectrometry (MS) for the detection and quantification of proteins, allowing both taxonomic identification and functional characterization of all present microorganisms, not just bacteria. The recent introduction of updated bioinformatics tools, such as MetaLab, has considerably improved data analysis (12); for lesser-represented species, however, the analysis remains inaccurate (13). Available data suggest that only 10% to 20% of the total protein content of a clinical sample is identified, corresponding to only one-third of the entire microbiota (14). Additional strategies have recently been suggested as prefractionation steps to circumvent instrument sensitivity problems and access low-abundance proteins of biological importance (15). The strengths of metaproteomics include a high resolution in bacterial species identification, the overview of functional and metabolic activities of microorganisms, and, most notably, the detection of host proteins, expanding the knowledge of host-microbiome interactions (16, 17), which is particularly interesting in CF (18). However, metaproteomics also has relevant limitations (19): standardization of the technique is needed to make it reproducible, and integration of meta-omics platforms is necessary for in-depth characterization. Moreover, although its cost has been reduced, it is still far from those for amplicon sequencing.

The combination of amplicon sequencing and metaproteomics techniques provides a wider characterization of microbial communities and their interaction with the host (20–22), and the integration of the results could support accurate predictions (23). However, few studies have included a statistical analysis comparing the results by the two methodologies, especially for challenging samples such as those from neonates, characterized by limited sample quantity and potentially low bacterial load. The aims of this study were to perform a robust statistical analysis on the concordance of amplicon sequencing and metaproteomics to study and describe the sequential establishment of the gut microbiota of a small representative cohort of infants with CF.

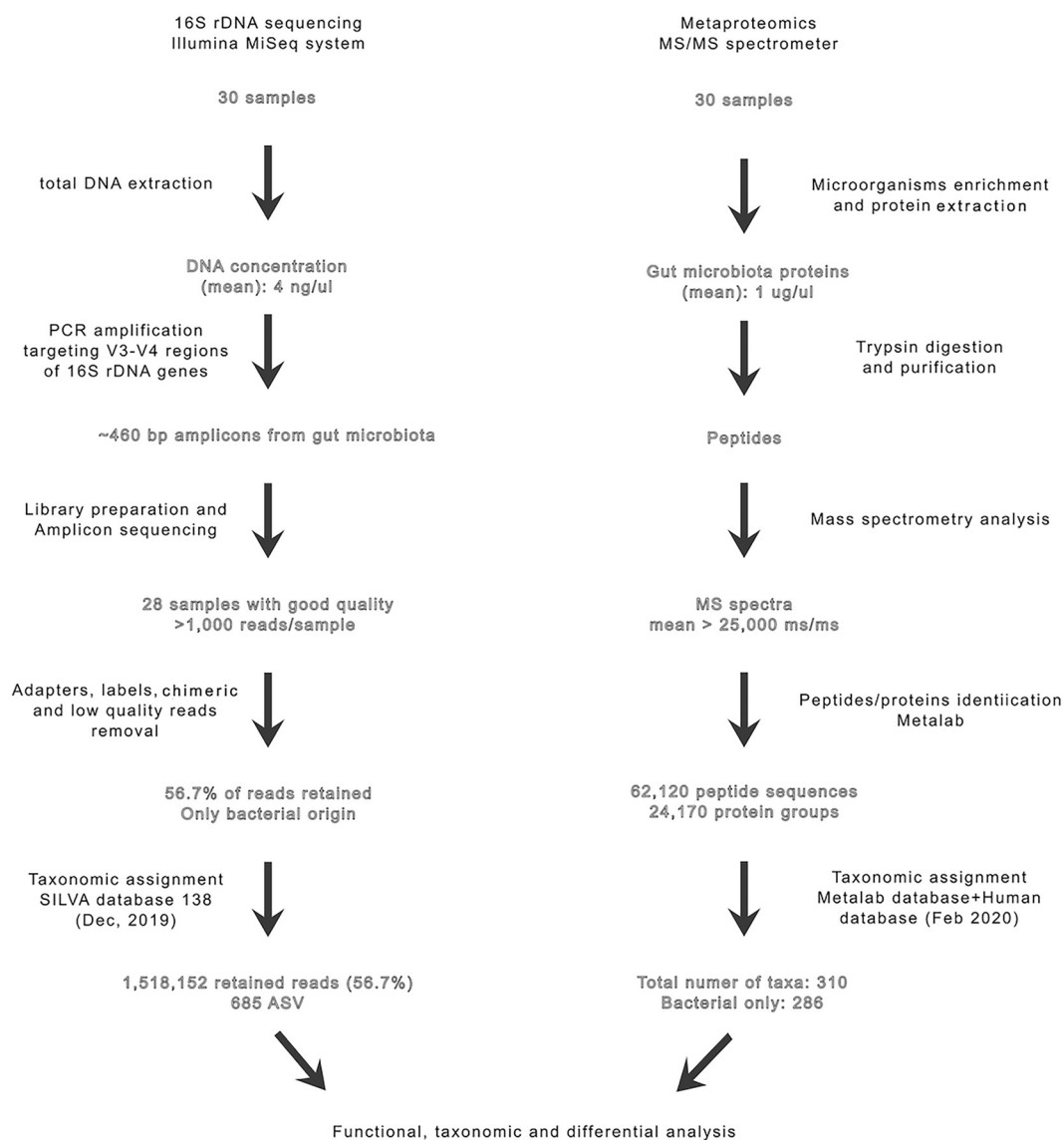


FIG 1 Workflow and main results obtained from the two methodologies used in this study.

RESULTS

Statistical concordance of the results of amplicon sequencing and metaproteomics.

We analyzed fecal samples ($n = 28$) using both techniques for bacterial identification (presence/absence) and quantification (abundance) and then respectively applied the McNemar and Wilcoxon tests to detect significant discrepancies (Fig. 1; see also “Amplicon sequencing data” and “Metaproteomics” in the supplemental material). Globally, 685 amplicon sequence variants (ASVs) (9 phyla, 13 classes, and 130 genera) were identified by amplicon sequencing (AS), and 310 taxa (8 phyla, 17 classes, and 65 genera) were identified by metaproteomics.

Taxa detected exclusively by a single method accounted for less than 10% of the total bacterial composition, being the genus level the one with most of the unique bacteria (Fig. 2). Minor statistical discrepancies were observed in taxon presence/absence by McNemar test (Table 1), most of which were related to the genus level (Fig. 2, Table 2, and “Common taxonomy” in the supplemental material). Furthermore, we detected certain nominal inconsistencies that should be accounted for related to the taxonomic classification by each database, such as the inclusion of the *Tenericutes* phylum from metaproteomics in the *Firmicutes* phylum

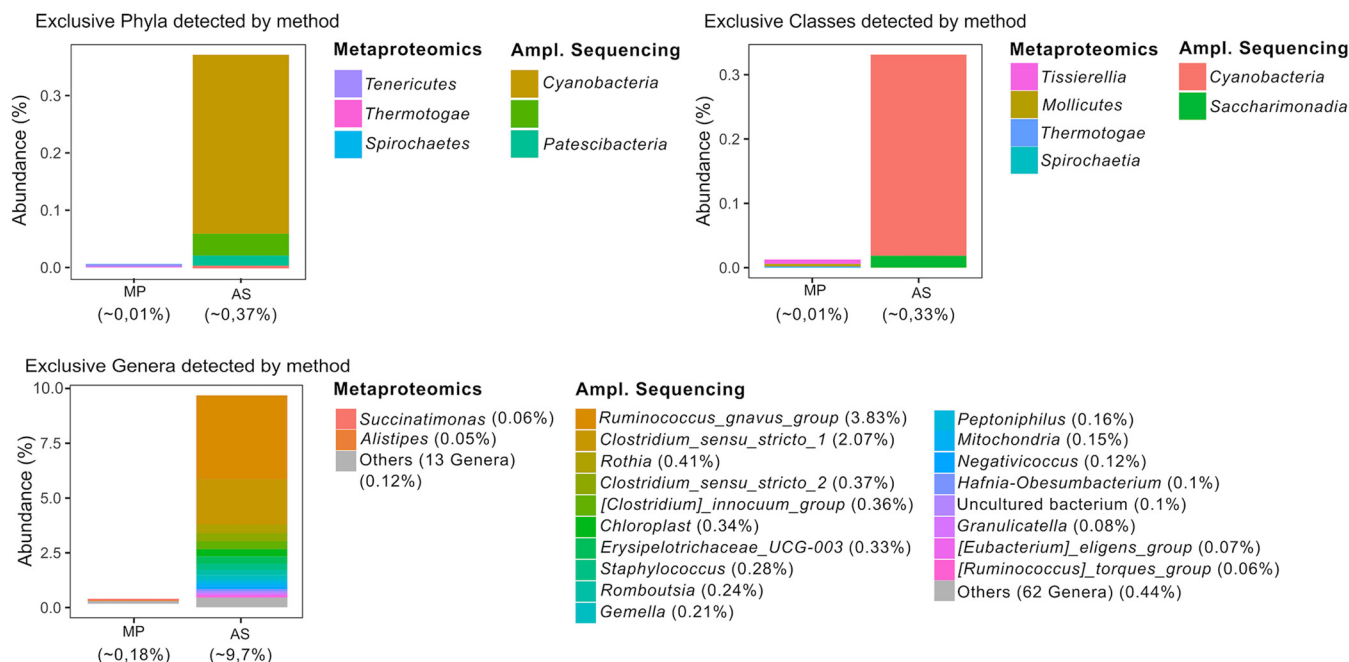


FIG 2 Bacterial phyla, classes, and genera uniquely identified by each of the methodologies. AS, amplicon sequencing; MP, metaproteomics. White spaces correspond to taxa correctly identified by both methods.

according to amplicon sequencing and the recent taxonomic reclassification of *Proteobacteria* classes (Table 2). Most of the reclassified taxa were at higher taxonomic levels.

Figure 3 summarizes the results of the compositional analysis from the top common taxa. Wilcoxon signed-rank paired tests showed significant differences in the relative abundances of the following taxa: the phylum *Actinobacteriota*, the classes *Bacilli* and *Actinobacteria*, and the genera *Blautia*, *Haemophilus*, *Escherichia*, *Enterococcus*, *Streptococcus*, and *Bifidobacterium* (Fig. 3). These top genera were separately explored to consider within-subject effect across the time series (Fig. 4A). The only genera that had a differential composition across time according to AS were *Ruminococcus* ($P = 0.022$) and *Eubacterium* ($P = 0.007$) (Fig. 4B, left). According to metaproteomics, the genera that have changed over time are *Fusobacterium*

TABLE 1 Contingency table for McNemar test accounting for the number of samples in which different bacteria (class and genus levels) were or were not identified by one, both, or neither of the technologies

Organism(s)	No. of samples detected by:				P value ^a
	Amplicon sequencing only	Metaproteomics only	Both	Neither	
Class					
<i>Alphaproteobacteria</i>	8	1	9	10	0.02
Genus					
<i>Eggerthella</i>	12	0	1	15	0.001
<i>Anaerococcus</i>	11	0	1	16	0.002
<i>Haemophilus</i>	15	2	5	6	0.003
<i>Coprococcus</i>	1	10	1	16	0.007
<i>Lachnospirillum</i>	0	9	2	17	0.007
<i>Roseburia</i>	0	9	2	17	0.007
<i>Eubacterium</i>	1	14	6	7	0.01
<i>Enterobacter</i>	10	1	3	14	0.01
<i>Enterococcus</i>	10	1	14	3	0.01
<i>Escherichia</i>	11	1	14	2	0.01
<i>Actinomyces</i>	14	4	6	4	0.03
<i>Collinsella</i>	8	1	5	14	0.04

^aSignificance level < 0.05.

TABLE 2 Discrepancies in nomenclature or taxonomic classification at different ranks (phylum/class/order/family/genus) by amplicon sequencing and metaproteomics^a

Taxa identified by amplicon sequencing	Taxa identified by metaproteomics
<i>Campilobacterota/Campylobacteria/Campylobacterales</i>	<i>Proteobacteria^a/Epsilonproteobacteria/Campylobacterales</i>
<i>Firmicutes/Bacilli/Mycoplasmatales</i>	<i>Tenericutes/Mollicutes/Mycoplasmatales</i>
<i>Actinobacteriota</i>	<i>Actinobacteria</i>
<i>Bacteroidota</i>	<i>Bacteroidetes</i>
.../Bacilli/Erysipelotrichales	.../Erysipelotrichia/Erysipelotrichales
.../Gammaproteobacteria/Burkholderiales	.../Betaproteobacteria/Burkholderiales
<i>Desulfobacterota/Desulfovibrionia/Desulfovibrionales</i>	<i>Proteobacteria/Deltaproteobacteria/Desulfovibrionales</i>
.../Clostridia/Peptostreptococcales-Tissierellales	.../Tissierellia/Tissierellales
.../.../.../Escherichia-Shigella	.../.../.../Escherichia
.../.../.../Lachnospirales/Lachnospiraceae/Ruminococcus_gnavus_group	.../.../.../Clostridiales/Ruminococcaceae/Ruminococcus
.../.../.../Lachnospirales/Lachnospiraceae/Ruminococcus_torques_group	

^aNames of taxonomic ranks that are identical are expressed with ".../..." The taxonomic assignment was made in 2019 before the latest nominal changes in the bacterial phyla. *Actinobacteriota* is now *Actinomycetota*, *Firmicutes* is *Bacillota*, and *Proteobacteria* is now *Pseudomonadota*.

($P = 0.02$), *Haemophilus* ($P = 0.029$) and *Veillonella* ($P = 0.03$) (Fig. 4B, right). The two methods showed similar trends for most genera, except for *Bifidobacterium*, *Escherichia*, *Enterococcus*, and *Ruminococcus*, which almost reach statistical significance ($P < 0.05$ but with a false-discovery rate [FDR] of 0.13). *Bifidobacterium* was the genus with the highest distance between methods (Fig. 4C to E). It is worth mentioning that significance could be affected by the small sample size and that sometimes changes are mostly detected for the lowest-abundance genera.

Alpha diversity metrics were significantly different between methodologies (Wilcoxon signed-rank paired test, $P < 0.05$). The methods also showed differences in temporal trends (Fig. 5), with only amplicon sequencing showing an increasing trend for the Shannon index in the latest samples. It must be mentioned that alpha diversity metrics calculations did not work well with metaproteomics data, as phyloseq software warned that there were no singletons in metaproteomics data frames. Calculating Shannon indices from rank-collapsed data and label free quantifications (LFQ) intensities as the ones that were available in metaproteomics workflow output is not common and should be considered for further works. Amplicon sequencing remained the best approach to calculate diversity metrics such as the aforementioned.

Functional assignment by metaproteomics. Functional information on the detected proteins was obtained by MetaLab software, using the sum of the peptide intensities for each protein (total intensity of the protein). Both the number and intensity of the proteins in the initial (5,718 and 4.86e12, respectively) and the early-CF (7,029 and 4.70e12, respectively) fecal samples were comparable. MetaLab assigned the functional profile by matching proteins with the Cluster of Orthologous Groups (COG) database with a 97% successful assignment rate for name/category. A previous process of normalization of the protein intensities (intensity of the function/sum of the intensities in that sample) was performed, and 2-way analysis of variance detected 14 COGs (6 from the initial and 8 from the early-CF samples) with statistical relevance (Table 3). However, and despite the small sample size of our work, there were no significant differences in the general functional profiles in the enrichment analysis, of which "energy production and conversion" (C), "carbohydrate transport and metabolism" (G), and "translation, ribosomal structure, and biogenesis" (J) were the COG categories with greater intensity (Fig. 6A).

The correlation of each bacterial taxon and its function was also performed with iMetaLab, using the iMetaShiny apps. *Bifidobacterium* spp. were the most abundant taxa and their proteins remained abundant in the early-CF samples, in which most of the functional pathways were promoted by *Ruminococcus gnavus* (Fig. 6B). A significant enrichment of *R. gnavus* proteins was observed from initial to early-CF times (2 versus 69 proteins), most of them related to translational processes (Fig. 6C). In contrast, the functions associated with *Clostridium perfringens* and *Blautia hansenii* showed a statistically significant decrease in the early-CF samples.

Metaproteomics also allowed the identification of 293 human proteins: 85 exclusively

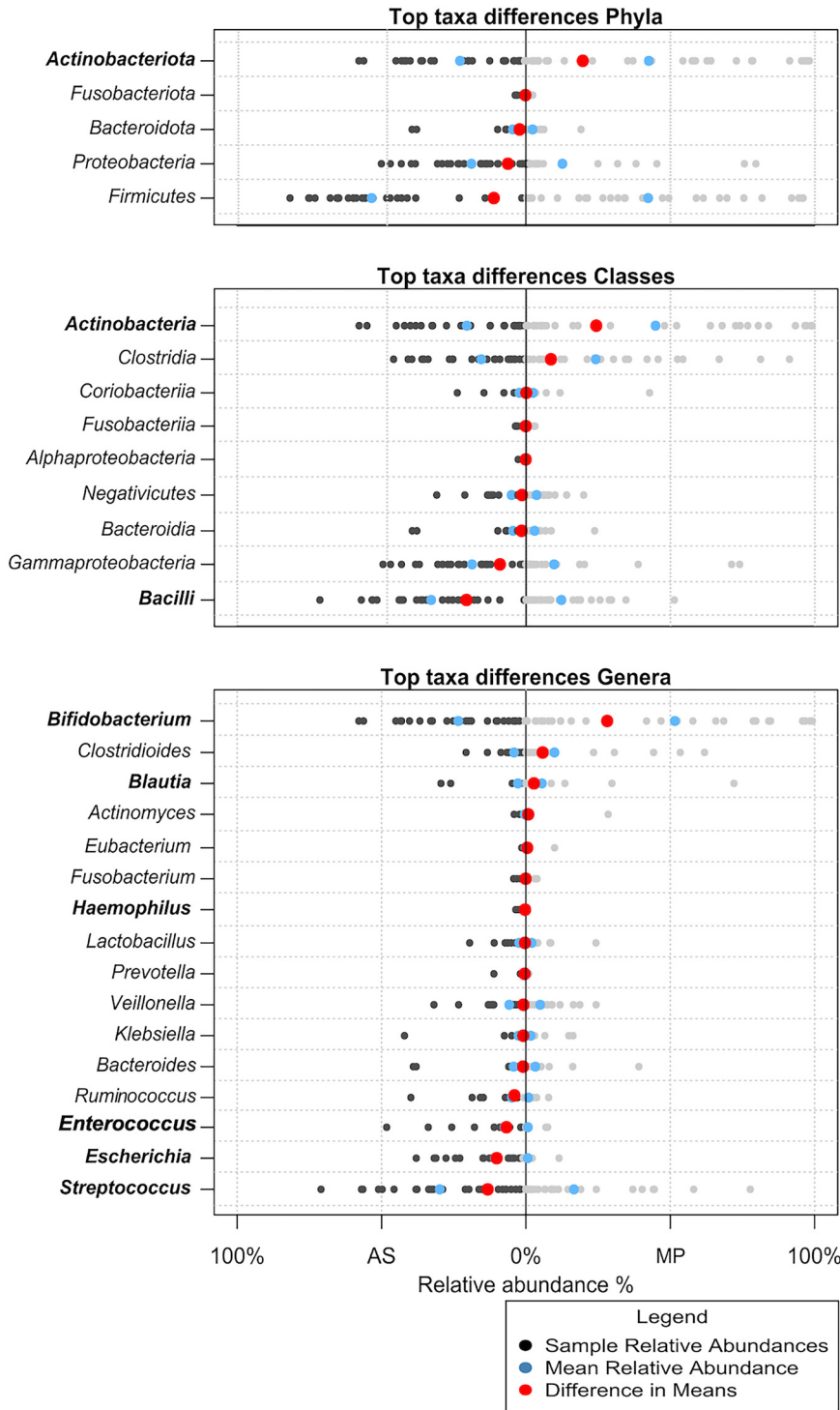


FIG 3 Relative abundances of the main taxa detected by both methodologies. Taxa in bold are those for which the paired Wilcoxon test result was significant ($P < 0.05$).

present in initial samples, 43 exclusively in early-CF samples, and 165 in both samples (“Human proteins” in the supplemental material). Six of the 165 shared proteins had significantly differential expression levels (Table 4). A STRING analysis (STRING version 11.5) was performed with the 3 groups of human proteins (initial, early CF, and shared proteins), including for the first 2 groups those proteins exclusively identified and with a significantly higher abundance in each of the samples (Fig. 7). The analysis of the proteins shared by the two groups

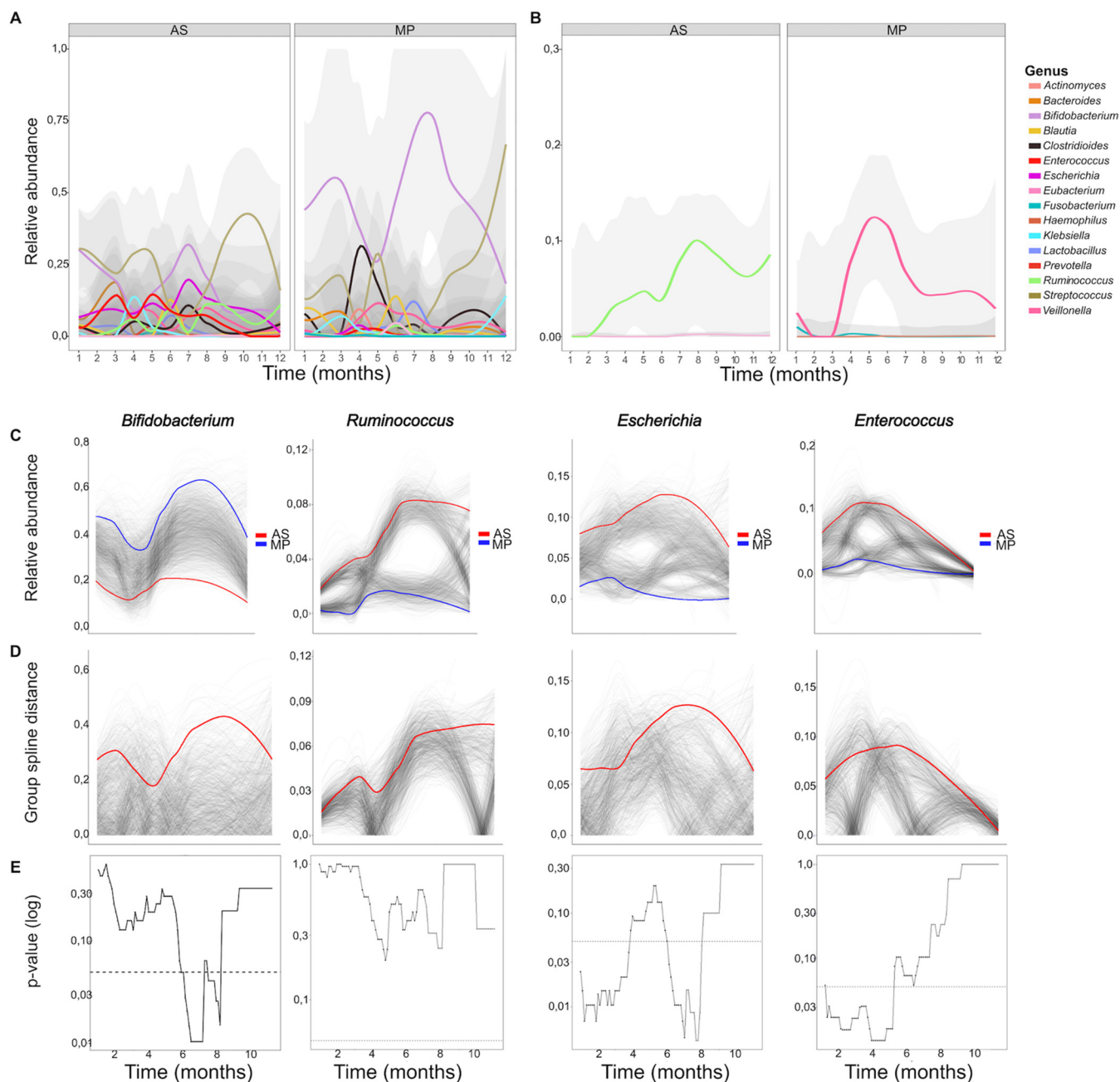


FIG 4 Compositional change of the major genera in each of the methodologies. (A) Evolution of the relative abundance of the main bacterial genera ($n = 16$). (B) Evolution of bacterial genera with a significant change over time detected by AS (*Ruminococcus* and *Eubacterium*) and by MP (*Fusobacterium*, *Haemophilus*, and *Veillonella*). (C) Graphs obtained by the permutest function (999 permutations) showing the temporal evolution of relative abundance in those genera with significant differences between the two methodologies. (D) Plots of distances between both methodologies. The difference (solid red line) is not significant if it is not above 95% of the permuted values (translucent gray). (E) Plots obtained by sliding spliner showing the P value at each specified interval (shown with 100 intervals by default). The dotted line indicates a P value of 0.05. At certain intervals, the differences became significant.

showed a significant protein-protein interaction (1.87×10^{-16}). Shared proteins were mainly related to neutrophil-mediated immunity (42 proteins; $P = 1.87 \times 10^{-25}$), such as calprotectin (S100-A9 and S100-A8), neutrophil elastase (ELANE), cathepsin G, and myeloperoxidase, or related to the intestinal barrier, such as mucins (MUC2, MUC4, and MUC13) or proteins present in cell junctions (40 proteins; $P = 1.24 \times 10^{-5}$) (Fig. 7A). The initial samples contained significantly more proteins associated with chylomicron assembly and the immune system than the early-CF samples (Fig. 7B), in which a cluster of muscle proteins stood out (Fig. 7C).

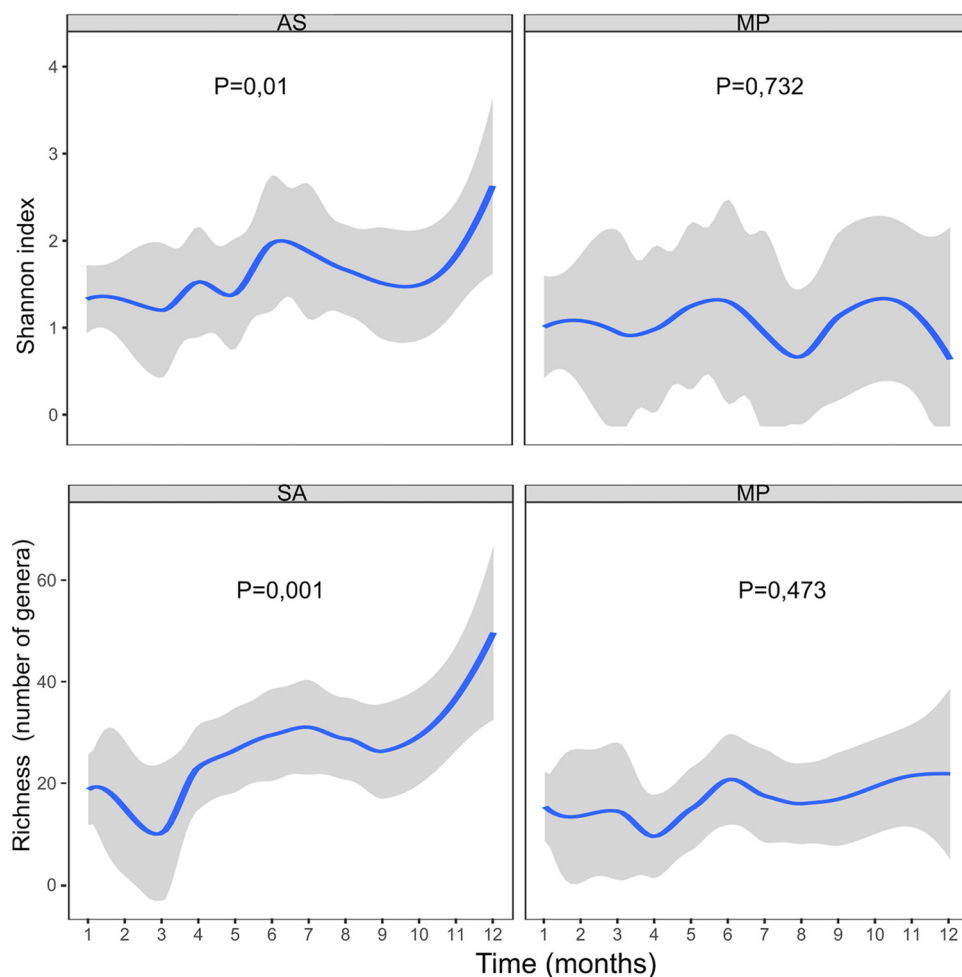


FIG 5 Evolution of alpha diversity (measured by Shannon index and richness) detected by one method and the other. The *P* value was obtained with the trendspliner spliner function (shown with 100 intervals by default), which evaluates the differences over time for each group of samples, separately.

The names of the proteins belonging to each function are listed in “Microbial data” in the supplemental material.

DISCUSSION

A major clinical feature of CF are digestive comorbidities, such as steatorrhea, mal-digestion/malabsorption, and pancreatic insufficiency, which also affect the quality of life and nutritional status of these individuals. A new classification of CF genetic mutations has been published (24), differentiating between high and low risk according to their clinical manifestations. Seven out of the 8 infants of this study carried high-risk mutations. The CF environment strongly conditions the gut microbiota composition (25), also contributing to its malfunction, even at the age of 6 weeks (9). Even though amplicon sequencing is the most widely used technique for determining the composition of gut microbiota, there is a growing demand for techniques that determine the global functionality of the ecosystem, such as metaproteomics.

The taxonomical incongruities detected were attributed to paraphyletic groups, which are continuously reclassified according to the latest advances in phylogeny. In this sense, SILVA 138 is a more accurate database with which to establish phylogenetic relationships, because protein databases are less frequently updated. By the time this paper is published, certain phylum names will have changed: *Firmicutes* is now *Bacillota*, *Proteobacteria* is now *Pseudomonadota*, and *Actinobacteriota* is currently named *Actinomycetota*.

TABLE 3 Functional categories of bacterial proteins with significantly different values between the initial group of samples and the early-CF group

COG name ^a	Mean difference	P value	COG category
Formyltetrahydrofolate synthetase (2,402)	2.64	<0.0001	Nucleotide transport and metabolism
ABC-type phosphate transport system, periplasmic component (2,605)	1.20	<0.0001	Inorganic ion transport and metabolism
Superfamily II DNA and RNA helicase (3,843)	1.25	<0.0001	Replication, recombination, and repair
Translation elongation factor EF-Tu, a GTPase (2,272)	0.74	0.0009	Translation, ribosomal structure, and biogenesis
F ₀ F ₁ -type ATP synthase, delta subunit (1,785)	0.65	0.01	Energy production and conversion
Pyruvate-formate lyase (4,465)	0.63	0.03	Energy production and conversion
Chaperonin GroEL (HSP60 family) (2,675)	-2.06	<0.0001	Posttranslational modification, protein turnover, and chaperones
Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase (2,333)	-1.03	<0.0001	Carbohydrate transport and metabolism
Carboxysome shell and ethanolamine utilization microcompartment protein CcmL/EutN (994)	-0.98	<0.0001	Secondary-metabolite biosynthesis, transport, and catabolism
Phosphoenolpyruvate-protein kinase (PTS system component in bacteria) (4,879)	-0.96	<0.0001	Carbohydrate transport and metabolism
Alcohol dehydrogenase, class IV (6,245)	-0.94	<0.0001	Energy production and conversion
Rubryerythrin (4,121)	-0.86	<0.0001	Energy production and conversion
Threonine dehydrogenase or related Zn-dependent dehydrogenase (5,594)	-0.79	<0.0001	General function prediction only
ABC-type sugar transport system, periplasmic component, contains N-terminal Xre family HTH domain (4,787)	-0.62	0.03	Carbohydrate transport and metabolism

^aPTS, phosphotransferase system; HTH, helix-turn-helix.

Curiously, metaproteomics detected proteins attributed to *Streptococcus thermophilus*, whereas when using amplicon sequencing, the proteins appeared to correspond to *Streptococcus salivarius*. Poor resolution of streptococcal species using the V3-V4 regions of the 16S rRNA gene has been previously reported (26), whereas the accuracy of metaproteomics in bacterial species identification has been demonstrated (12). Another amplicon sequencing limitation is the detection of what is known as sporobiota (27), most of the endospore-forming species belonging to *Firmicutes*. Spores are underrepresented because of their ability to resist DNA extraction techniques. Other bacterial phyla, such as *Actinobacteria* (particularly *Bifidobacterium*), have genomes with a high GC content, and this likewise contributes to their underestimation by sequencing techniques.

Metaproteomics has also failed to detect or properly estimate certain bacterial groups typically found in the human gut microbiota due to their resistance to cell lysis, especially Gram-positive bacteria (28). Moreover, certain microbial components and detergents used for cell lysis can affect postenzymatic digestion (14). Our workflow includes optimized steps, such as serial centrifugations for microbial pre-enrichment, sonication with bead beating, and lysis with sodium dodecyl sulfate (SDS)-based buffers. While our method improves the number of peptide and protein identifications and detection of the relative abundance of *Actinobacteria*, other protocols have shown a higher identification of proteins from other taxonomic groups, such as *Proteobacteria*, but a lower number of peptide and protein identifications (29). Metaproteomic analysis presents several bioinformatics challenges, the most common being adequate peptide identification, taxonomic assignment, and the quality of functional annotation (14).

The statistical analysis of our results demonstrated equivalent bacterial taxonomic assignments by both techniques, while for abundance detection, some discrepancies were found. The lower predicted relative abundance of *Bifidobacterium* by amplicon sequencing has been repeatedly reported, as well as the higher predicted relative abundance we detected for some *Firmicutes* and *Proteobacteria* taxa in comparison to other methodologies (30).

The small size of our cohort limits comparison with other studies. Alpha diversity indices of our samples did not reach those observed in other CF cohorts or in the control cohort of Antosca et al. (9), although other work showed lower values in CF patients than in healthy controls (31). We conclude that it is not possible to compare alpha diversity indices between metaproteomics and amplicon sequencing with the approach we followed in this study,

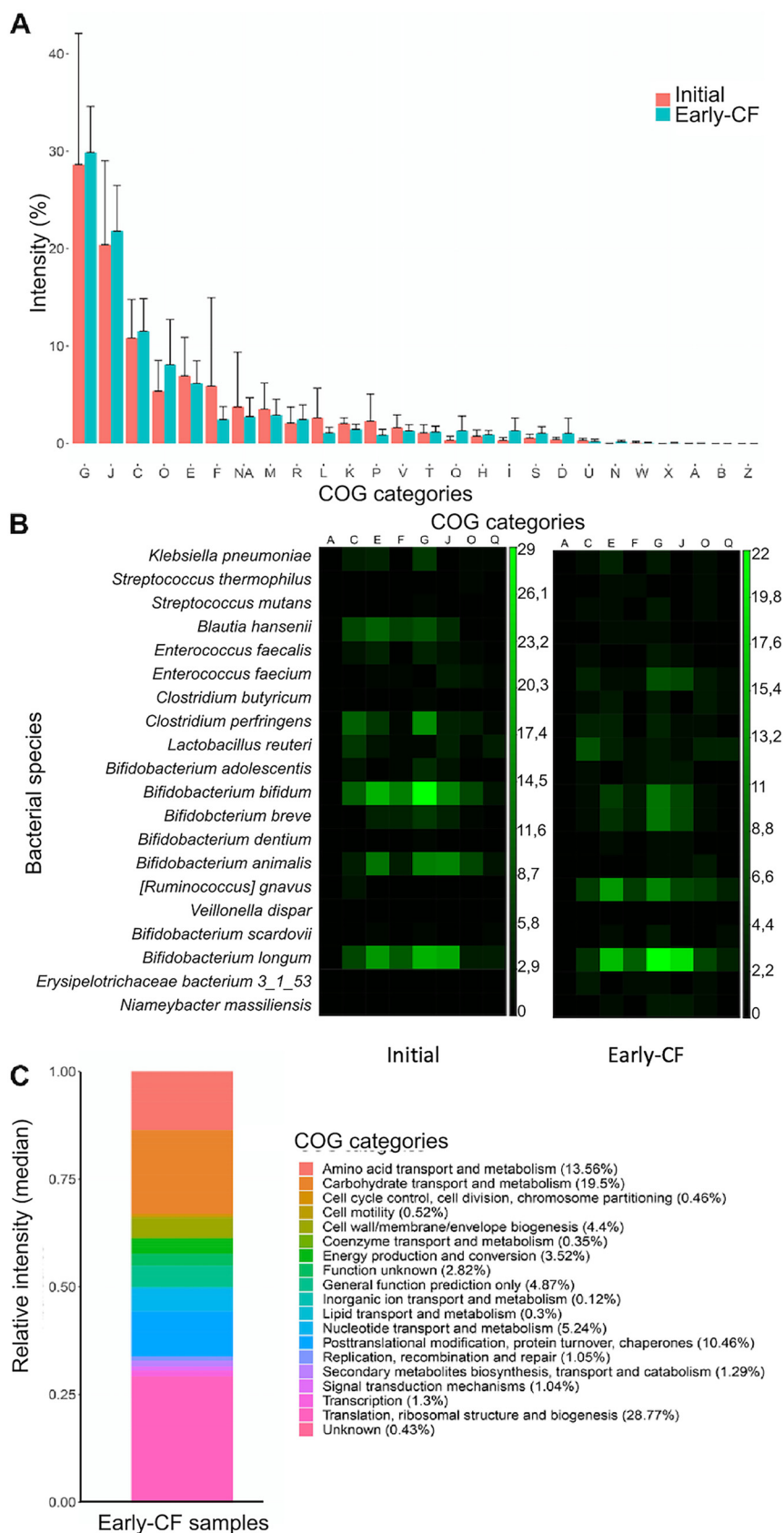


FIG 6 Functional analysis of bacterial proteins from this cohort, comparing the group of initial and final samples (early CF). (A) Relative intensity of each of the COG functional categories in each group of (Continued on next page)

TABLE 4 Human proteins enriched in each group of the cohort

Higher expression in initial sample	Higher expression in early-CF sample	P value
Lactotransferrin		<0.0001
Myeloperoxidase		<0.0001
	Actin cytoplasmic 2	<0.0001
	Chymotrypsin-C	<0.0001
	Carboxypeptidase A1	0.0409
	Myosin-1	<0.0001

although comparison between samples processed with identical proteomic techniques could be possible. In amplicon sequencing, several ASVs (each with a certain number of reads, typically on the order of 1 to 10⁴) may correspond to the same bacterial taxon. In metaproteomics, on the other hand, the sum of intensities (on the order of 10⁶ to 10¹¹) of all peptides corresponding to the same taxon is evaluated to calculate the taxon's relative abundance in every sample. This makes the calculation of metrics such as the Shannon index, which considers the proportion of each bacterium, and their comparison between techniques not entirely reliable; nonetheless, species richness (number of different species) could still be used.

Fecal proteomics for adults with CF has pointed to the dominance of *R. gnavus*, *Enterobacteriaceae*, and *Clostridia* species, combined with decreased butyrate producers, such as *Faecalibacterium prausnitzii* (18). We detected similar compositional changes in a very reduced cohort of 8 infants followed for a year right after being diagnosed with CF. *R. gnavus* has been associated with several human gastrointestinal diseases and is a mucolytic bacterium capable of degrading the human colonic mucin MUC2, which is a major form of secretory glycosylated mucin coating the epithelia of the intestine and therefore has a close relationship with the mucus layer (32, 33). It has been suggested that the bacterium can penetrate the mucus layer and lead to an activation of the immune system and inflammatory responses (34, 35). Nevertheless, the involvement of this species in allergic or autoimmune diseases has been also questioned (32, 36).

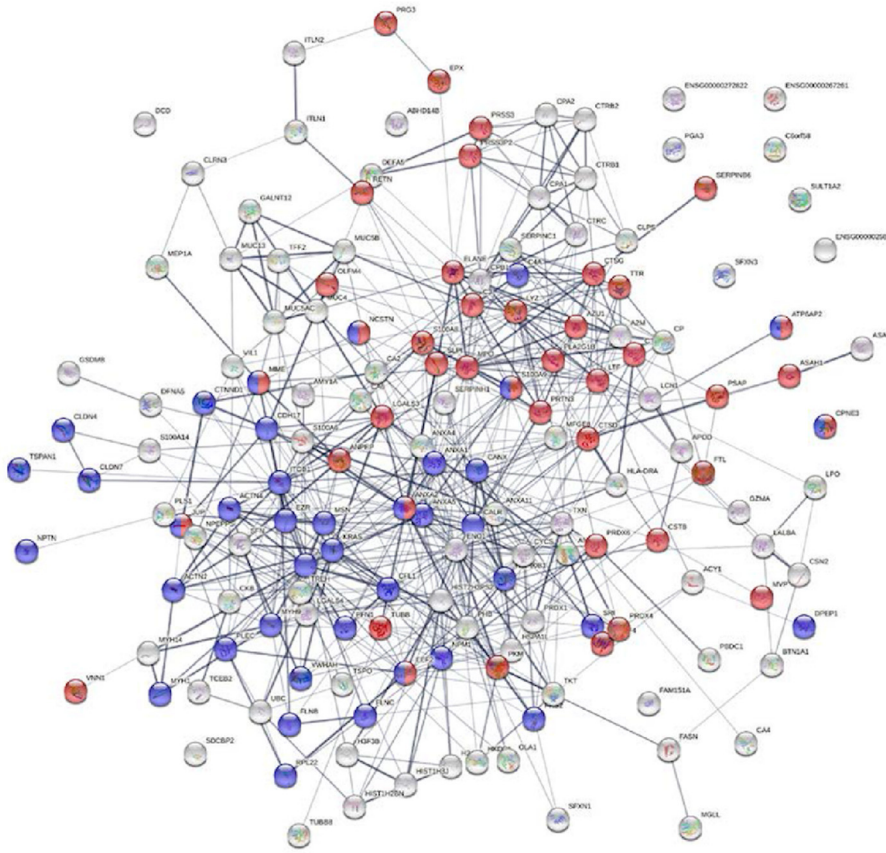
Drastic changes in the bacterial functional profiles of the initial and early-CF samples were not observed; the most abundant proteins in the initial samples were a formyltetrahydrofolate synthetase and a pyruvate-formate lyase previously related to biofilm upregulation (37). In the early-CF samples, the most relevant enrichment was that of the chaperonin GroEL, a stress response protein whose main role is protein folding, which has recently been observed to have roles in virulence and pathogenesis, such as bacterial adherence and immune invasion (38).

One of the most interesting advantages of metaproteomics is the detection of human proteins. In our samples, several proteins associated with the maintenance of the intestinal epithelium and the immune response were detected. Lactotransferrin, an iron sequestration protein abundant in human breast milk; and myeloperoxidase, regulator of inflammation in granulocytes of neutrophils, were significantly enriched in the initial samples (39, 40). All our results point to an altered establishment of the

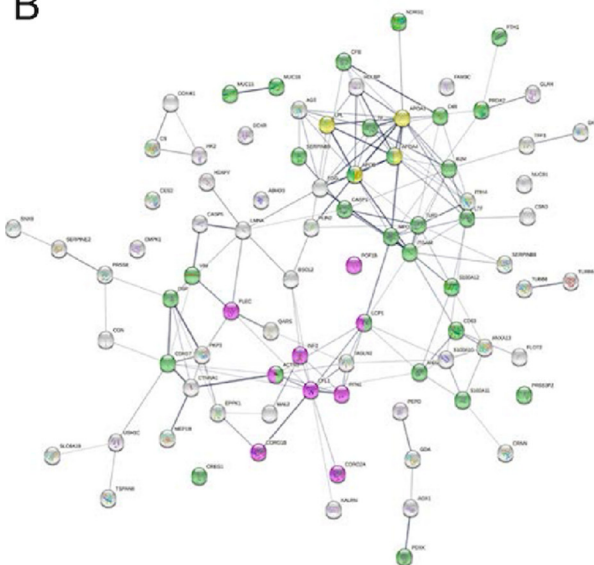
FIG 6 Legend (Continued)

samples. (B) Correlation between the most enriched functional categories and the bacterial species to which they are assigned. The color gradient indicates higher relative intensity (light green). (C) Proportions of COG functional categories of proteins assigned to *R. gnavus*. COG category nomenclature: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control, cell division, and chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure, and biogenesis; K, transcription; L, replication, recombination, and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, posttranslational modification, protein turnover, and chaperones; P, inorganic ion transport and metabolism; Q, secondary-metabolite biosynthesis, transport, and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defense mechanisms; W, extracellular structures; X, mobilome: prophages and transposons; Z, cytoskeleton; NA, not assigned to any category.

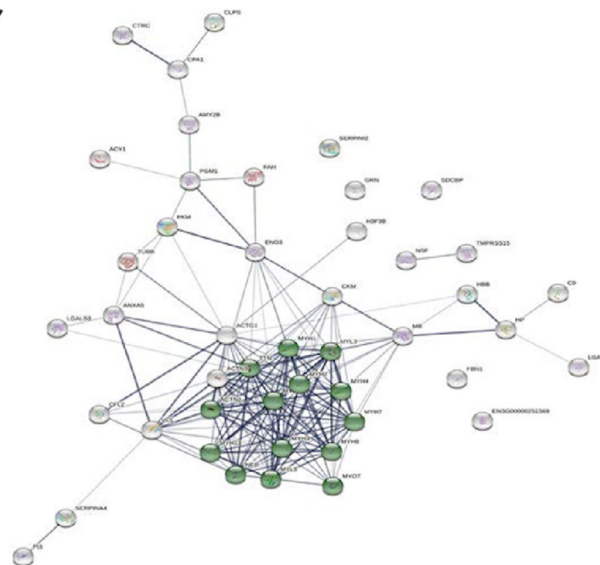
A



B



C



- Neutrophil mediated immunity
- Immune system process
- Actin binding
- Cell junction
- Chylomicron remodeling
- Muscle protein and sarcomere organization

FIG 7 Human proteins found interacting with the gut microbiome of the cohort. (A) Group of proteins unique to the initial samples; (B) group of proteins unique to the early-CF samples; (C) proteins shared by both groups.

TABLE 5 Relevant clinical characteristics of infants and sampling^a

Infant	Sex	Pancreatic insufficiency	Delivery mode	Birth feeding	CF mutation ^b	Sampling at mo (no. of sample)											
						1	2	3	4	5	6	7	8	9	10	11	12
1	Female	Yes	Vaginal	Formula	G542X/R1162X	1	2						3				
2	Male	Yes	Caesarean section	Formula	F508del/Y1092X	4				5		6					
3	Female	Yes	Vaginal	Exc. ^c breast milk	F508del/F508del	7			8						9		10
4	Male	No	Vaginal	Exc. breast milk	F508del/3272-26A->G	11		12			13		14				
5	Male	Yes	Vaginal	Exc. breast milk	F508del/F508del	15			16				17				
6	Male	Yes	Vaginal	Formula	F508del/N1303K	18					19			20			
7	Female	Yes	Vaginal	Formula	F508del/F508del	21			22	23		24					
8	Female	Yes	Vaginal	Formula	F508del/R1162X	25		26		27							28

^aAll samples ($n = 28$) were included in the methodological comparative analysis, whereas samples in gray ($n = 16$; 8 initial samples and 8 early-CF samples) were considered to analyze gut microbiota and human proteomic content in this cohort.

^bAll CFTR alleles except 3272-26A>G are classified as high risk, and therefore, all infants had a high-risk CF genotype except infant 4.

^cExc., exclusively breast milk. At endpoint, all infants had included solid food in their feeding status.

gut microbiota in infants with CF, with a dominance of *R. gnavus* within a human inflammatory state.

The main limitation of our study is the small sample size. Due to that, we have included unadjusted, slightly higher *P* values to inform about bacterial phyla, classes, and genera and about microbial and human proteins which appeared to change but did not reach statistical significance. These features could be considered in future studies with larger sample sizes or more exhaustive CF metabolic characterization.

Our study provides an extended comparative analysis with robust statistical support that could optimize the use of both approaches for gut microbiota research. Metaproteomics provides information on composition and functionality, as well as data on host-microbiome interactions (41). Its strength is the identification and quantification of *Actinobacteria* and certain classes of *Firmicutes*, whereas its main limitation is the lack of comparability of alpha diversity indices. Taking all the results into account, both techniques detected an aberrant microbiota in infants with CF during their first year of life, dominated by the enrichment of *R. gnavus* within a human inflammatory environment.

MATERIALS AND METHODS

Participants, sample collection, and medical records. Subjects included in this study had CF diagnosed by neonatal screening, sweat chloride test, and mutation sequencing and were recruited during their first month of life, when the first fecal sample was collected, between January 2018 and January 2019. The infants were attended at 3 Spanish CF reference hospitals, geographically distant. Ethical approval for the study was granted by the Ramón y Cajal Ethics Committee in 2017 with (file number 163/17), and all infants' parents gave informed consent. Subjects were excluded if they were born after that date or had fewer than 3 collected samples or if important clinical data were missing. Eight infants with CF were finally followed up, and the relevant data, including feeding habits, delivery mode, and CF mutations, are shown in Table 5. Data related to antibiotic and other treatments are summarized in "Antibiotic treatments" in the supplemental material. Fecal samples were sequentially collected from the diaper of each infant during the first year of life over routine medical checkups. Samples were stored at -80°C in 2 aliquots for subsequent amplicon sequencing and meta-proteomics analysis.

DNA extraction and 16S rRNA amplicon sequencing. Our study followed guidelines for the Strengthening The Organization and Reporting of Microbiome Studies (STORMS) reporting ("STORM checklist" in the supplemental material) (42). Total DNA from the fecal samples was obtained with the QiaAMP kit (Qiagen, Germany), and further MiSeq 2 × 300-bp paired-end (Illumina) 16S rRNA sequencing of the V3 and V4 regions (43) was performed at the Central Unit for Translational Genomics Support (Ramón y Cajal Health Research Institute). The sequencing data analysis was performed using the Qiime pipeline (44), which includes DADA2 for sequence quality filtering, and the SILVA 138 database (released December 2019) was used for taxonomical assignment, discharging those samples with fewer than 1,000 reads ($n = 2$). Nucleotide sequences were deposited in the National Center for Biotechnology Information's Sequence Read Archive (SRA) repository, BioProject identifier (ID) [PRJNA719717](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA719717).

Metaproteomics. Samples were prepared using differential centrifugation, and the microbial pellet was processed by sonication according to a previously published protocol. In brief, 0.1 to 0.3 g of feces was suspended in 10 mL of phosphate-buffered saline (PBS) and mixed in a tube rotator for 45 min at 4°C , as previously published (45, 46). The samples were centrifuged at $500 \times g$ for 5 min, and the supernatant was collected in a 50-mL tube. Ten milliliters of PBS was added again, repeating the process twice, and the 3 supernatants (~30 mL) were centrifuged at $11,000 \times g$. Microbial pellets were suspended in 500 μL of lysis buffer (4% sodium dodecyl sulfate [SDS], 50 mM Tris-HCl [pH 8.0]) and heated for 10 min at 95°C . Four sonication cycles (30 s with a 1-min interval on ice) were performed, with an amplitude of 40%. Silica beads were then added

(0.3 g) to each sample, and 5 rounds of bead beating were performed (30 s with a 5-min interval on ice) at a speed of 6.5 ms^{-1} . Centrifugation at $14,000 \times g$ was performed to remove the beads and cell debris. To remove the SDS, proteins were precipitated using methanol-chloroform and suspended in 8 M urea for in-solution trypsin digestion. Lastly, the peptides were quantified in a Qubit fluorimeter (Thermo Scientific), and 1 μg of peptides was loaded for reverse-phase nano-liquid chromatography electrospray ionization tandem mass spectrometric analysis on an EASY-nLC 1000 system (Proxeon) coupled to a Q-Exactive HF mass spectrometer (Thermo Scientific). Peptides were loaded on-line onto an Acclaim PepMap 100 trapping column (75 μm [inside diameter] by 20 mm, 3 μm , C_{18} resin with 100-Å pore size; Thermo Scientific) using buffer A (0.1% formic acid) and then separated on a C_{18} resin analytical column (75 μm [inside diameter] by 500 mm, 2 μm , 100-Å pore size; EASY-Spray column; Thermo Scientific). A 240-min gradient from 2% to 40% buffer B (0.1% formic acid in 100% acetonitrile) in buffer A was performed to separate the peptides.

Data were obtained by data-dependent acquisition in positive mode. From each MS scan (between 350 and 2,000 Da), the 15 most intense precursors (charges between 2+ and 5+) were selected for their high collision energy dissociation fragmentation, with a dynamic exclusion of 10 s and a normalized collision energy of 20, and the corresponding MS/MS spectra were acquired. Peptides were eluted using a 240-min gradient. The MS proteomics raw data have been submitted to the ProteomeXchange Consortium (<http://www.proteomexchange.org>) via the Proteomics Identifications Database partner repository with the database identifier PXD029284 (47). For data processing, we employed MetaLab software, which provides a human gut microbial database (12, 48). A human database downloaded from UniProt DB (<http://www.uniprot.org>) (49), restricted to human taxonomy (downloaded on 18 February 2020 with 74,451 sequences), was also used to identify human proteins. For peptide and protein identification, the FDR was set to 0.01. Only taxa identified with at least 2 peptides were considered and were manually filtered to eliminate human peptides. The sum of the intensities of all the distinctive peptides assigned to a taxon was used as the relative abundance of that taxon. A taxon-function analysis was also performed, using taxonomic information from the enrichment analysis from the iMetaLab platform (<http://shiny.imetalab.ca/>) (50).

Graphs and statistical analysis. The data were trimmed by selecting only the taxa identified by both methods (shared taxa), solving annotation incongruences. First, we evaluated the dichotomous dependent variable, "taxa presence," by a McNemar test with continuity correction at the phylum, class, and genus levels, considering in each sample any amplicon sequence variant (of at least 2 peptides) per taxon. Significance was set to a *P* value of <0.05 . Second, a differential abundance analysis was performed with taxa detected in more than 50% of the samples by at least one of the approaches. Compositional abundance of each phylum, class, and genus was normalized per sample, dividing by total bacterial read counts assigned to the specific taxonomical level and by total sum of LFQ intensity detected at that bacterial taxonomical level for AS and metaproteomics, respectively. Alpha diversity metrics (Shannon index and species richness) were calculated with `estimate_richness` function from the `phyloseq` package (version 1.40.0) of RStudio using genus abundances (LFQ or read counts). Package `ggpubr` (version 0.4.0) was used for plotting paired data. GraphPad Prism (version 9.1.1.225) was used for the functional statistical analysis of the microbial and human proteins.

Nonparametric Wilcoxon signed-rank paired tests were performed to compare the populations' mean ranks (individual taxa and alpha diversity metrics), pairing by sample ID. Significance values in Wilcoxon tests were adjusted via a Benjamini-Hochberg (BH) FDR of 0.1 for AS and metaproteomics abundances.

The CF infant microbiota through the first months of life was profiled using the `splinctomeR` package (version 0.1.0) in RStudio (version 4.2.1). Differences in the top taxon relative abundances over the time series detected by each method were evaluated with the `permuspliner` function (999 permutations). Likewise, each taxon was investigated individually to detect any divergence between methods at a particular time point along the time course with the `slidingspliner` function. Finally, we investigated the time effect for the top taxon relative abundance trends along the time series detected by each method separately using the `trendyspliner` function. The same approach was used to evaluate differences between methods for alpha diversity metrics. Significance values were adjusted by an FDR of 0.1.

Significant differences were verified with linear models adjusting for age, birth feeding (maternal only/formula), cystic fibrosis transmembrane conductance regulator (CFTR) mutation type (high risk/low risk), and respiratory exacerbations (yes/no), which were considered if one or more episodes were suffered during the sample collection.

Data availability. The sequence data from this study are deposited in the NCBI's SRA repository, BioProject ID [PRJNA719717](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA719717). The proteomics data set from this paper has been deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier [PXD029284](https://www.ebi.ac.uk/pride/archive/study/PXD029284).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, XLSX file, 0.01 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.

SUPPLEMENTAL FILE 3, XLSX file, 0.1 MB.

SUPPLEMENTAL FILE 4, XLSX file, 0.2 MB.

SUPPLEMENTAL FILE 5, XLSX file, 0.02 MB.

SUPPLEMENTAL FILE 6, XLSX file, 0.03 MB.

SUPPLEMENTAL FILE 7, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

We thank the families of the children for their participation and all the health personnel of the CF units, and we thank Marta Cobo for technical assistance.

R.D.C. is the recipient of a Vertex Pharmaceuticals grant. The other authors declare no conflicts of interest related to this work.

This work was supported by the Instituto de Salud Carlos III, cofunded by the European Union (PI17/00115 and PI20/00164 to R.D.C.), REIPI 307 (RD16/0016/0011) actions, cofinanced by the European Development Regional Fund “A way to achieve Europe” (ERDF), Vertex Pharmaceuticals. C.S. has grant support from “Fundación Mutua Madrileña” in Atenas Internacional (ATH) 017 call to R.D.C. (AP165902017). Proteomic experiments were supported by projects RTI2018-094004-B-100 and InGEMICS-CM B2017/BMD3691 via funds to C.G.

REFERENCES

- Selma-Royo M, Calatayud Arroyo M, García-Mantrana I, Parra-Llorca A, Escuriel R, Martínez-Costa C, Collado MC. 2020. Perinatal environment shapes microbiota colonization and infant growth: impact on host response and intestinal function. *Microbiome* 8:167. <https://doi.org/10.1186/s40168-020-00940-8>.
- Cuthbertson L, Walker AW, Oliver AE, Rogers GB, Rivett DW, Hampton TH, Ashare A, Elborn JS, De Soyza A, Carroll MP, Hoffman LR, Lanyon C, Moskowicz SM, O'Toole GA, Parkhill J, Planet PJ, Teneback CC, Tunney MM, Zuckerman JB, Bruce KD, van der Gast CJ. 2020. Lung function and microbiota diversity in cystic fibrosis. *Microbiome* 8:45. <https://doi.org/10.1186/s40168-020-00810-3>.
- Dayama G, Priya S, Niccum DE, Khoruts A, Blekhnman R. 2020. Interactions between the gut microbiome and host gene regulation in cystic fibrosis. *Genome Med* 12:12. <https://doi.org/10.1186/s13073-020-0710-2>.
- Andrinopoulou ER, Clancy JP, Szczesniak RD. 2020. Multivariate joint modeling to identify markers of growth and lung function decline that predict cystic fibrosis pulmonary exacerbation onset. *BMC Pulm Med* 20:142. <https://doi.org/10.1186/s12890-020-1177-z>.
- Burke DG, Fouhy F, Harrison MJ, Rea MC, Cotter PD, O'Sullivan O, Stanton C, Hill C, Shanahan F, Plant BJ, Ross RP. 2017. The altered gut microbiota in adults with cystic fibrosis. *BMC Microbiol* 17:58. <https://doi.org/10.1186/s12866-017-0968-8>.
- Price CE, O'Toole GA. 2021. The gut-lung axis in cystic fibrosis. *J Bacteriol* 203:e00311-21. <https://doi.org/10.1128/JB.00311-21>.
- Duytschaever G, Huys G, Bekaert M, Boulanger L, De Boeck K, Vandamme P. 2011. Cross-sectional and longitudinal comparisons of the predominant fecal microbiota compositions of a group of pediatric patients with cystic fibrosis and their healthy siblings. *Appl Environ Microbiol* 77:8015–8024. <https://doi.org/10.1128/AEM.05933-11>.
- Schipa S, Iebba V, Santangelo F, Gagliardi A, De Biase RV, Stamato A, Bertasi S, Lucarelli M, Conte MP, Quattrucci S. 2013. Cystic fibrosis transmembrane conductance regulator (CFTR) allelic variants relate to shifts in faecal microbiota of cystic fibrosis patients. *PLoS One* 8:e61176. <https://doi.org/10.1371/journal.pone.0061176>.
- Antosca KM, Chernikova DA, Price CE, Ruoff KL, Li K, Guill MF, Sontag NR, Morrison HG, Hao S, Drumm ML, MacKenzie TA, Dorman DB, Feenan LM, Williams MA, Dessaint J, Yuan IH, Aldrich BJ, Moulton LA, Ting L, Martinez-del Campo A, Stewart EJ, Karagas MR, O'Toole GA, Madan JC. 2019. Altered stool microbiota of infants with cystic fibrosis shows a reduction in genera associated with immune programming from birth. *J Bacteriol* 201:e00274-19. <https://doi.org/10.1128/JB.00274-19>.
- Knight R, Vrbancic A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciorek T, McCall L-I, McDonald D, Melnik AV, Morton JT, Navas J, Quinn RA, Sanders JG, Swafford AD, Thompson LR, Tripathi A, Xu ZZ, Zaneveld JR, Zhu Q, Caporaso JG, Dorrestein PC. 2018. Best practices for analysing microbiomes. *Nat Rev Microbiol* 16:410–422. <https://doi.org/10.1038/s41579-018-0029-9>.
- Rintala A, Pietilä S, Munukka E, Eröla E, Pursiheimo J-P, Laiho A, Pekkala S, Huovinen P. 2017. Gut microbiota analysis results are highly dependent on the 16S rRNA gene target region, whereas the impact of DNA extraction is minor. *J Biomol Tech* 28:19–30. <https://doi.org/10.1171/jbt.17-2801-003>.
- Cheng K, Ning Z, Zhang X, Li L, Liao B, Mayne J, Figeys D. 2020. MetaLab 2.0 enables accurate post-translational modifications profiling in metaproteomics. *J Am Soc Mass Spectrom* 31:1473–1482. <https://doi.org/10.1021/jasms.0c00083>.
- Zhang X, Li L, Butcher J, Stintzi A, Figeys D. 2019. Advancing functional and translational microbiome research using meta-omics approaches. *Microbiome* 7:154. <https://doi.org/10.1186/s40168-019-0767-6>.
- Zhang X, Figeys D. 2019. Perspective and guidelines for metaproteomics in microbiome studies. *J Proteome Res* 18:2370–2380. <https://doi.org/10.1021/acs.jproteome.9b00054>.
- Mayers MD, Moon C, Stupp GS, Su AI, Wolan DW. 2017. Significant alterations in protein expression from a mouse model of inflammatory bowel disease. *J Proteome Res* 16:1014–1026. <https://doi.org/10.1021/acs.jproteome.6b00938>.
- Zhang X, Li L, Mayne J, Ning Z, Stintzi A, Figeys D. 2018. Assessing the impact of protein extraction methods for human gut metaproteomics. *J Proteomics* 180:120–127. <https://doi.org/10.1016/j.jprot.2017.07.001>.
- Isaac NI, Philippe D, Nicholas A, Raouf D, Eric C. 2019. Metaproteomics of the human gut microbiota: challenges and contributions to other OMICS. *Clin Mass Spect* 14:18–30. <https://doi.org/10.1016/j.clinms.2019.06.001>.
- Debyser G, Mesuere B, Clement L, Van de Weygaert J, Van Hecke P, Duytschaever G, Aerts M, Dawyndt P, De Boeck K, Vandamme P, Devreese B. 2016. Faecal proteomics: a tool to investigate dysbiosis and inflammation in patients with cystic fibrosis. *J Cyst Fibros* 15:242–250. <https://doi.org/10.1016/j.jcf.2015.08.003>.
- Petriz BA, Franco OL. 2017. Metaproteomics as a complementary approach to gut microbiota in health and disease. *Front Chem* 5:4. <https://doi.org/10.3389/fchem.2017.00004>.
- Brown CT, Xiong W, Olm MR, Thomas BC, Baker R, Firek B, Morowitz MJ, Hettich RL, Banfield JF. 2018. Hospitalized premature infants are colonized by related bacterial strains with distinct proteomic profiles. *mBio* 9:e00441-18. <https://doi.org/10.1128/mBio.00441-18>.
- Debyser G, Aerts M, Van Hecke P, Mesuere B, Duytschaever G, Dawyndt P, De Boeck K, Vandamme P, Devreese B. 2019. A method for comprehensive proteomic analysis of human faecal samples to investigate gut dysbiosis in patients with cystic fibrosis. *Adv Exp Med Biol* 1073:137–160. https://doi.org/10.1007/978-3-030-12298-0_6.
- Lee AJ, Einarsson GG, Gilpin DF, Tunney MM. 2020. Multi-omics approaches: the key to improving respiratory health in people with cystic fibrosis? *Front Pharmacol* 11:569821. <https://doi.org/10.3389/fphar.2020.569821>.
- Bevivino A, Bacci G, Drevinek P, Nelson MT, Hoffman L, Mengoni A. 2019. Deciphering the ecology of cystic fibrosis bacterial communities: towards systems-level integration. *Trends Mol Med* 25:1110–1122. <https://doi.org/10.1016/j.molmed.2019.07.008>.
- Rueda-Nieto S, Mondejar-Lopez P, Mira-Escolano MP, Cutillas-Tolín A, Maceda-Roldán LA, Arense-Gonzalo JJ, Palomar-Rodríguez JA. 2022. Analysis of the genotypic profile and its relationship with the clinical manifestations in people with cystic fibrosis: study from a rare disease registry. *Orphanet J Rare Dis* 17:222. <https://doi.org/10.1186/s13023-022-02373-y>.
- Meeker SM, Mears KS, Sangwan N, Brittnacher MJ, Weiss EJ, Treuting PM, Tolley N, Pope CE, Hager KR, Vo AT, Paik J, Frevert CW, Hayden HS, Hoffman LR, Miller SI, Hajjar AM. 2020. CFTR dysregulation drives active selection of the gut microbiome. *PLoS Pathog* 16:e1008251. <https://doi.org/10.1371/journal.ppat.1008251>.
- Martinović A, Cocuzzi R, Arioli S, Mora D. 2020. *Streptococcus thermophilus*: to survive, or not to survive the gastrointestinal tract, that is the question! *Nutrients* 12:2175. <https://doi.org/10.3390/nu12082175>.
- Egan M, Dempsey E, Ryan CA, Ross RP, Stanton C. 2021. The sporobiota of the human gut. *Gut Microbes* 13:1. <https://doi.org/10.1080/19490976.2020.1863134>.
- Zhang X, Chen W, Ning Z, Mayne J, Mack D, Stintzi A, Tian R, Figeys D. 2017. Deep metaproteomics approach for the study of human microbiomes. *Anal Chem* 89:9407–9415. <https://doi.org/10.1021/acs.analchem.7b02224>.

29. García-Durán C, Martínez-López R, Zapico I, Pérez E, Romeu E, Arroyo J, Hernáez ML, Pitarch A, Monteoliva L, Gil C. 2021. Distinct human gut microbial taxonomic signatures uncovered with different sample processing and microbial cell disruption methods for metaproteomic analysis. *Front Microbiol* 12:618566. <https://doi.org/10.3389/fmicb.2021.618566>.
30. Mancabelli L, Milani C, Lugli GA, Fontana F, Turrone F, van Sinderen D, Ventura M. 2020. The impact of primer design on amplicon-based metagenomic profiling accuracy: detailed insights into bifidobacterial community structure. *Microorganisms* 8:131. <https://doi.org/10.3390/microorganisms8010131>.
31. Coffey MJ, Nielsen S, Wemheuer B, Kaakoush NO, Garg M, Needham B, Pickford R, Jaffe A, Thomas T, Ooi CY. 2019. Gut microbiota in children with cystic fibrosis: a taxonomic and functional dysbiosis. *Sci Rep* 9: 18593. <https://doi.org/10.1038/s41598-019-55028-7>.
32. Chua HH, Chou HC, Tung YL, Chiang BL, Liao CC, Liu HH, Ni YH. 2018. Intestinal dysbiosis featuring abundance of *Ruminococcus gnavus* associates with allergic diseases in infants. *Gastroenterology* 154:154–167. <https://doi.org/10.1053/j.gastro.2017.09.006>.
33. Wu H, Rebello O, Crost EH, Owen CD, Walpole S, Bennati-Granier C, Ndeh D, Monaco S, Hicks T, Colville A, Urbanowicz PA, Walsh MA, Angulo J, Spencer DIR, Juge N. 2021. Fucosidases from the human gut symbiont *Ruminococcus gnavus*. *Cell Mol Life Sci* 78:675–693. <https://doi.org/10.1007/s00018-020-03514-x>.
34. Gren C, Spiegelhauer MR, Rotbain EC, Ehmsen BK, Kampmann P, Andersen LP. 2019. *Ruminococcus gnavus* bacteraemia in a patient with multiple haematological malignancies. *Access Microbiol* 1:e000048. <https://doi.org/10.1099/acmi.0.000048>.
35. Crost EH, Tailford LE, Monestier M, Swarbreck D, Henrissat B, Crossman LC, Juge N. 2016. The mucin-degradation strategy of *Ruminococcus gnavus*: the importance of intramolecular trans-sialidases. *Gut Microbes* 7: 302–312. <https://doi.org/10.1080/19490976.2016.1186334>.
36. Azzouz D, Omarbekova A, Heguy A, Schwudke D, Gisch N, Rovin BH, Caricchio R, Buyon JP, Alekseyenko AV, Silverman GJ. 2019. Lupus nephritis is linked to disease-activity associated expansions and immunity to a gut commensal. *Ann Rheum Dis* 78:947–956. <https://doi.org/10.1136/annrheumdis-2018-214856>.
37. Resch A, Rosenstein R, Nerz C, Götz F. 2005. Differential gene expression profiling of *Staphylococcus aureus* cultivated under biofilm and planktonic conditions. *Appl Environ Microbiol* 71:2663–2676. <https://doi.org/10.1128/AEM.71.5.2663-2676.2005>.
38. Fourie KR, Wilson HL. 2020. Understanding GroEL and DnaK stress response proteins as antigens for bacterial diseases. *Vaccines (Basel)* 8:773. <https://doi.org/10.3390/vaccines8040773>.
39. Legrand D, Pierce A, Elass E, Carpentier M, Mariller C, Mazurier J. 2008. Lactoferrin structure and functions. *Adv Exp Med Biol* 606:163–194. https://doi.org/10.1007/978-0-387-74087-4_6.
40. Aratani Y. 2018. Myeloperoxidase: its role for host defense, inflammation, and neutrophil function. *Arch Biochem Biophys* 640:47–52. <https://doi.org/10.1016/j.abb.2018.01.004>.
41. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. 2020. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights* 14: 1177932219899051. <https://doi.org/10.1177/1177932219899051>.
42. Mirzayi C, Renson A, Zohra F, Elsaforay S, Geistlinger L, Kasselman L, Eckenrode K, van de Wijgert J, Loughman A, Marques FZ, STORMS Consortium, Genomic Standards Consortium, Massive Analysis and Quality Control Society, Segata N, Huttenhower C, Dowd JB, Jones HE, Levi Waldron L. 2021. Strengthening The Organization and Reporting of Microbiome Studies (STORMS): a reporting checklist for human microbiome research. *bioRxiv*. <https://doi.org/10.1101/2020.06.24.167353>.
43. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41:e1. <https://doi.org/10.1093/nar/gks808>.
44. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6:90. <https://doi.org/10.1186/s40168-018-0470-z>.
45. Long S, Yang Y, Shen C, Wang Y, Deng A, Qin Q, Qiao L. 2020. Metaproteomics characterizes human gut microbiome function in colorectal cancer. *NPJ Biofilms Microbiomes* 6:14. <https://doi.org/10.1038/s41522-020-0123-4>.
46. Abbondio M, Palomba A, Tanca A, Fraumene C, Pagnozzi D, Serra M, Marongiu F, Laconi E, Uzzau S. 2019. Fecal metaproteomic analysis reveals unique changes of the gut microbiome functions after consumption of sourdough *carasau* bread. *Front Microbiol* 10:1733. <https://doi.org/10.3389/fmicb.2019.01733>.
47. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Pérez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF, Ternent T, Brazma A, Vizcaino JA. 2019. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47:D442–D450. <https://doi.org/10.1093/nar/gky1106>.
48. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T, Juncker AS, Manichanh C, Chen B, Zhang W, Levenez F, Wang J, Xu X, Xiao L, Liang S, Zhang D, Zhang Z, Chen W, Zhao H, Al-Aama JY, Edris S, Yang H, Wang J, Hansen T, Nielsen HB, Brunak S, Kristiansen K, Guarner F, Pedersen O, Doré J, Ehrlich SD, Bork P, Wang J, MetaHIT Consortium. 2014. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 32:834–841. <https://doi.org/10.1038/nbt.2942>.
49. UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47:D506–D515. <https://doi.org/10.1093/nar/gky1049>.
50. Liao B, Ning Z, Cheng K, Zhang X, Li L, Mayne J, Figeys D. 2018. iMetaLab 1.0: a web platform for metaproteomics data analysis. *Bioinformatics* 34: 3954–3956. <https://doi.org/10.1093/bioinformatics/bty466>.