JOURNAL OF PROTEOMICS 97 (2014) 17–26

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

A metaproteomic pipeline to identify newborn mouse gut phylotypes☆,☆☆

Federica Del Chierico[a,b], Andrea Petrucca[a,b,c], Stefano Levi Mortera[d,e], Pamela Vernocchi[a,b,f], Maria M. Rosado[g], Luisa Pieroni[d,e], Rita Carsetti[g], Andrea Urbani[d,e,*,1], Lorenza Putignani[a,b,**,1]

[a]Parasitology Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy
[b]Metagenomics Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy
[c]Department of Diagnostic Science, Sant'Andrea Hospital, Rome, Italy
[d]Department of Experimental Medicine, University of Rome "Tor Vergata", Rome, Italy
[e]IRCCS—Santa Lucia Foundation, Rome, Italy
[f]Interdipartimental Centre for Industrial Research, Alma Mater Studiorum, University of Bologna, Italy
[g]B-cell development Unit and Immunological Diagnosis Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy

ARTICLE INFO

Available online 29 October 2013

Keywords:
Early life gut microbiota phylotypes
Mouse model
Metaproteomic pipeline
MALDI-TOF MS proteomics

ABSTRACT

In order to characterize newborn mouse gut microbiota phylotypes in very early-life stages, an original metaproteomic pipeline, based on LC–MS^2-spectra and Mascot driven NCBI non-redundant repository database interrogation was developed. An original computational analysis assisted in the generation of a taxonomic gut architecture from protein hits to operational taxonomic units (OTUs) and related functional categories. Regardless of the mouse's genetic background, a prevalence of Firmicutes (Lactobacillaceae) and Proteobacteria (Enterobacteriaceae) was observed among the entire Eubacteria taxonomic node. However, a higher abundance of Firmicutes was retrieved for Balb/c gut microbiota compared to $Rag2^{ko}$ mice, the latter was mainly characterized by a Proteobacteria enriched microbiota. The metaproteomic-obtained OTUs were supported, for the identification (ID) of the cultivable bacteria fraction, corroborated by axenic culture-based MALDI-TOF MS IDs. Particularly, functional analysis of $Rag2^{ko}$ mice gut microbiota proteins revealed the presence of abundant glutathione, riboflavin metabolism and pentose phosphate pathway components, possibly related to genetic background.

The metaproteomic pipeline herein presented may represent a useful tool to investigate the highly debated *onset* of the human gut microbiota in the first days of life, when the bacterial composition, despite its very low diversity (complexity), is still very far from an exhaustive description and other complex microbial consortia.

☆☆ This article is part of a Special Issue entitled: Trends in Microbial Proteomics.

* *Correspondence to:* A. Urbani, Department of Experimental Medicine, University of Rome "Tor Vergata", Via Montpellier 00133, Rome, Italy.

** *Correspondence to:* L. Putignani, Parasitology Unit, Metagenomics Unit, Department of Laboratories, Bambino Gesù Children's Hospital, IRCCS, Piazza Sant'Onofrio 4, 00165, Rome, Italy. Tel.: +39 0668592598 2176; fax: +39 0668592218.

E-mail addresses: andrea.urbani@uniroma2.it (A. Urbani), lorenza.putignani@opbg.net (L. Putignani).

[1] Both authors are senior authors.

1874-3919/$ – see front matter
http://dx.doi.org/10.1016/j.jprot.2013.10.025

Biological significance

The manuscript deals with a "frontier" topic regarding the study of the gut microbiota and the application of a metaproteomic pipeline to unveil the complexity of this fascinating ecosystem at the very early stages of life. Indeed during these phases, its diversity is very low but the bacterial content is highly "instable", and the relative balance between mucosal and fecal bacteria starts its dynamics of "fight" to get homeostasis. However, in the neonatal period, especially immediately after birth, a comprehensive description of this microbial eco-organ is still lacking, while it should be mandatory to highlight its first mechanisms of homeostasis and perturbation, while it co-develops with and within the host species.

In order to unravel its low but almost unknown microbial community multiplicity, the newborn mouse gut, characterized by a "very" low complexity, was herein selected as model to design a LC–MS -based shotgun metaproteomic approach, potentially suitable to study onset and shaping[2]in human newborns. A microbiological semi-automatic computational analysis was performed to infer gut phylotypes; such as proof of evidence, related OTUs were compared to axenic-culture-based MALDI-TOF MS IDs showing consistency at family and phyla levels for the bacterial cultivable fraction.

This article is part of a Special Issue entitled: Trends in Microbial Proteomics.

1. Introduction

Complex ecosystems of host and microbial cells are organized into diverse ecological niches (*e.g.*, teeth, mouth, nose and respiratory district, gut, and genito-urinary tract) inside mammalian organisms. In each district different groups of microbes (phylotypes) form communities that change across space (*e.g.*, spatial variability determinants) and time (*e.g.*, age variability determinants) [1] Cohabitation with microbes is the product of a long evolutionary pathway resulting in a relationship based on mutual advantages that in humans is essential for health [2]. The human body together with "its" microbes has been defined as superorganism [3] and because hosting billions of microbes benefits their genes, proteins and metabolites acquiring additional functions and thus amplifying growth and survival tools. The superorganism description needs a new system biology-based approach, employing either metagenomic or metaproteomic strategies, which offers a new holistic paradigm of the human biological ontology [4].

The development of metagenomic studies has arisen from the need to interpret the vast data sets of sequences produced by the microbial genome projects [5], and to link their annotations to the functional counterpart in the context of body habitats. Indeed, in these habitats the bacteria establish a dynamic interplay with the host, generating the microbiome [6]. Metaproteomics, instead, analyzes protein patterns expressed in the ecosystems and thus describes the real-time dynamics of the systems [7]. In this context, a full comprehensive description of human gut phylotypes is essential to interpret gut homeostasis and perturbation [8–11].

In order to unravel gut microbial community multiplicity at the very early stages of life, the mouse gut, which is characterized by a lower gut complexity than the humans [12], may represent an appropriate model to design interpretative pipelines in metaproteomic analyses for studies on neonatal gut microbiota onset and development [13,14]. Indeed, during these stages, microbiota diversity is extremely low and the bacterial content is highly "instable", with a topographic and metabolic balance tremendously depending on the respective mucosal and fecal bacteria counterparts, which starts their fight dynamics to get homeostasis [15]. However, for obvious ethical issues studies on mucosal bacterial contents cannot be performed in neonatal age, and gut topographical-related results can be therefore inferred by employing newborn mouse models. Remarkably, the bacterial complexity and also its contents is particularly affected by the mucosal surface extension [16] hence providing low bacterial content which can actually be related to the very small surfaces of the newborn mice. Indeed, also the pivotal metaproteomic study of Li et al. [17] on the human mucosal luminal interface, performed on healthy adults enrolled in a cancer surveillance program, showed the limited presence of distinct proteins (*i.e.*, 49) associated to biogeographic features, suggesting a possible low diversity index of the mucosal bacterial-associated communities, which can be obviously more restrained in the very early life stages. In fact, to the best of our knowledge up to now, metaproteomic studies on mammalian gut microbiota have been mainly based on the fecal content, either gel-based [18–20], or gel-free [21,22] approaches.

However, sample complexity remains a very significant issue and the few metaproteomic studies performed on rodents gut microbiota have been exploiting either 1D-SDS gel electrophoresis or isoelectrofocusing as the first fractionation step, at the protein or peptide level, prior to LC–MS^2 injection analysis [12–14,23].

As a descriptive approach, we chose a common shotgun proteomics experiment, relying on a monodimensional reverse-phase (RP) chromatography, adjusting the gradient time in order to obtain the maximum spreading of peptides along the whole run time, thus minimizing ion suppression at the ESI source. Tandem MS was performed on a fast scanning 3D-ion trap, coupled to a Biotyper MALDI-TOF MS for the complementary analysis of the cultivable bacterial fraction. The herein study mainly aimed at describing mouse gut enterotypes during very early life stages (*e.g.*, programming phase), when limited microbial content is expected [15]

and metaproteomic analysis can actually improve the mucosal-associated gut microbiota description, still almost unknown and whose cultivable portion can be performed by MALDI-TOF MS. Briefly, an overall of 18 Balb/c and Balb/c-Rag2^{ko} baby mice were sacrificed at 3, 7, and 14 days after birth (three mice for each time point); their intestines were entirely removed and subjected, after suitable treatment, to axenic culture-based MALDI-TOF MS identifications (IDs) and to a metaproteomic workflow consisting of nano HPLC/MS^2 analysis coupled to an automatic bioinformatic algorithm for bacterial operational taxonomic unit (OTU) description. Indeed, our metaproteomic pipeline comprised data processing and Mascot DB search, coupled with an automatic computational procedure to translate the Mascot outputs into mouse gut phylotypes at the level of either bacterial phyla or families. Microbial protein IDs were further complemented by functional categories retrieved by Cluster of Orthologous Group (COG) database interrogation.

Although the number of identified proteins was rather low, a good agreement, in terms of major phylotype (phyla) description, was ascertained by culture-dependent MALDI-TOF MS IDs and metaproteomics, suggesting a link between mouse genetic background (Balb/c and Rag2^{ko} mice) and gut microbiota shaping in the very early stages of life.

2. Materials and methods

2.1. Mice

Balb/c and Balb/c-Rag2^{ko} animals were cared in conventional pathogen-free conditions at the Centro Ricerche Sperimentali, Istituto Regina Elena (Roma, Italy). Three groups, each composed of three baby mice of Balb/c and Balb/c-Rag2^{ko}, were established. All baby mice were kept with their mothers until analysis. The mice were sacrificed at 3, 7 and 14 days after birth; the whole intestine, from the duodenum to the rectal tract, was removed and collected into sterile 1.5 ml Eppendorf tubes. All procedures with animals were performed in compliance with the relevant laws and local institutional guidelines.

2.2. Bacterial recovery from mouse intestine

Bacterial recovery of mouse intestine was performed by following the procedure of Apajalahti et al. [24]. The intestines, from duodenum to the rectum, were mechanically homogenized in 10 ml of Hanks' balanced salt solution (HBSS), using a Stomacher 80 Biomaster (Seward Limited, Worthing, UK) for 1 min at high speed, and intestinal macro-debris was removed by several low speed centrifugation (200 ×*g* for 15 min). Subsequently, cleared supernatants were high-speed serial centrifuged at 30,000 ×*g* for 15 min. The pellets were suspended into 500 μl HBSS and screened by microscopy to evaluate bacterial recovery (90%), hence providing the enriched bacteria suspensions (EBSs) [24].

2.3. Culture-based MALDI-TOF MS bacterial IDs

Ten out of 500 μl EBSs from baby mice were plated either onto elective Columbia (5% sheep blood) agar (COS), PolyViteX blood agar (PVX) and Schaedler Anaerobe blood agar (SCS) plates (bioMérieux, Marcy l'Etoile, France) and incubated at 37 °C for 48–72 h, in aerobic, microaerophilic and anaerobe growth conditions by using generator bags for microaerophilic and anaerobic bacteria (bioMérieux, Marcy l'Etoile, France), respectively. A bacterial cell density was estimated, by streaking serial dilutions of EBSs onto COS, ranging from 10^5 to 10^7 colony forming units/ml for the entire set of starting sample homogenates. Based on morphology and growth conditions, colonies were characterized and re-isolated onto COS, PVX and SCS agar plates in order to proceed with the MALDI-TOF MS-based IDs performed with a Microflex LT mass spectrometer (Bruker Daltonics, GmbH, Bremen, Germany), using Flex Control (version 3.0) and MALDI Biotyper automation control (version 2.0) software. In particular, bacterial cells were directly picked from isolated colonies with a 10 μl tip, smeared in triplicate onto an MSP 96 polished steel target (Bruker DaltonikDaltonics) and air-dried at RT. Each sample was overlaid with 1.5 μl of matrix, which consisted of a saturated solution of α-cyano-4-hydroxy-cinnamic acid (Bruker DaltonikDaltonics) in 50% acetonitrile–2.5% trifluoroacetic acid (Sigma-Aldrich). The matrix/sample was co-crystallized by air drying at RT. Three-hundred shots, in 50-shot steps from different positions of the target spot (automatic mode) were collected from each spectrum (mass range of 2000 to 20,000 Da). Spectral analyses and bacterial IDs were automatically performed with default settings without any user intervention. The generated peak lists were used for matches against the reference library (version 2.0 SR 1; Bruker Daltonics) directly using the integrated pattern-matching algorithm of the MALDI Biotyper automation control. The software provides a log score, and the cut-off score of 2 was used to validate ID at the species level, as recommended by the manufacturer. From all bacterial isolates, characterized at species levels, the complete bacterial lineages were retrieved and the mouse gut microbiota represented at the level of phyla was inferred.

2.4. Metaproteomic analysis pipeline

Two-hundred out of 500 μl EBSs from Balb/c and Rag2^{ko} babies mice, sacrificed at 3, 7 and 14 days, were centrifuged at 10,000 ×*g* for 15 min, resuspended into 200 μl sample buffer (7 M urea, 2 M thiourea, 40 mM Tris base, 4% CHAPS, 50 mM DTT), pre-warmed at 37 °C, sonicated for 20 s (5 ×) and then incubated for 1 h at 37 °C. About 100 μg of proteins were submitted to reduction with 50 mM DTT for 1 h at 37 °C, alkylation with 100 mM iodoacetamide (IAA) for 1 h at RT and finally digested on with 2 μl of 0.5 μg/μl of trypsin at 37 °C.

Tryptic digests were analyzed by nLC–MS on a Proxeon EASY-nLCII (Thermo Fisher Scientific, Milan, Italy) interfaced with an amaZon ETD Ion Trap (Bruker Daltonics). 2 μl (about 400 ng of tryptic digest) from each samples was injected in duplicate (intra-assay replicate) and pre-concentrated for 3 min on a C18-A1 EASY-Column™ (2 cm, 100 μm I.D., 5 μm p.s., Thermo Fisher Scientific) at a flow rate of 5 μl/min. A gradient elution was performed on a C18-Acclaim PepMap (25 cm, 75 μm I.D., 5 μm p.s., Thermo Fisher Scientific), at a flow rate: 0.3 μl/min, T of 20 °C; eluents: A, H_2O + 0.1% HCOOH and B, CH_3CN + 0.1% HCOOH; gradient: from 3 to 30% B in 180 min.

MS data were acquired with an AutoMSn method (Bruker definition for data dependent acquisition) using the ultrascan resolution as scan mode (32,000 m/z sec^{-1}). Three precursor ions were selected for each survey scan, keeping the active exclusion enabled for 30 s after 2 MS^2 scans on the same precursor. Raw data were processed with the Mascot distiller software (version 2.3.2.0, Matrix Science) to generate a peak list for database searching. Protein IDs were performed by Mascot server 2.3 version running on a locally resident cluster using the NCBI non-redundant database version (September 2012). Searching was restricted to bacteria (Eubacteria) as taxonomy entry, setting carbamidomethylation of cysteines as fixed modification and oxidation of methionines as variable modification, allowing one missing cleavage. Maximal error tolerances of 0.4 Da and 0.5 Da were chosen for parent and fragment ions, respectively, according with the low resolution mass analyzer, and providing a false discovery rate (FDR) below 1%, after Mascot Percolator score filtering.

2.5. *Mascot outcome automatization processing*

In the Mascot output, peptides were assembled in protein hits all of which were associated to different groups of homologous proteins shared by distinct OTUs. Only hits with at least one "bold red" peptide and a protein score higher than 39 were considered for the analysis. These data were exported in a comma-separated value (CSV) format and analyzed by two Excel macros, able to automatically convert protein hits in mouse gut phylotypes, hence reported at the level of bacterial phyla and families. The Mascot CSV file was simplified to obtain a final protein hit list table, correlated with their respective taxonomy code number (only "prot_hit_num" and "prot taxa id" columns were taken into consideration). A library of 280195 taxonomy code numbers of microorganisms belonging to Bacteria and Archaea Kingdoms were created and linked with their respective phyla and bacterial family names. A comparison between the taxonomy code numbers present in "prot taxa id" column of Mascot output and those present in the taxonomy code library was performed, thus generating in the Mascot output a list of protein hits associated to their respective phylotypes at phyla and bacterial family levels. The presence of phyla or family redundancies within each protein hit was removed by using the "Remove duplicates" function of Excel 2007 software. The resulting unique phyla and family names were indexed in alphabetical order and enumerated to show the gut microbiota composition (see Supplementary Material, SM1 for the description of manual and automated procedures for conversion of Mascot protein hits into OTUs, SM2, 3 for the operational steps of the manual procedure (*e.g.*, exported CSV-bacterial and phyla counts, 8 sheets; library of 280195 taxonomy code numbers), SM4–6 (phyla macro), SM4, 7–8 (families' macro).

2.6. *Functional annotation*

To characterize the functional properties of protein hits identified by our metaproteomic pipeline the Clusters of Orthologous Group (COG) database was used, probing protein name against gene function groups defined within the COG database (http://www.ncbi.nlm.nih.gov/COG/).

3. Results and discussion

The metaproteomic pipeline was applied to the characterization of gut microbiota in wild-type Balb/c and Balb/c-$Rag2^{ko}$ mice (here nominated as $Rag2^{ko}$), that are expected to have different microbial composition [25]. Recombination activating gene (RAG) deficient $Rag2^{ko}$ mice are severely immunodeficient, because of the RAG-2 gene mutation which hampers the IgA assembling, hence affecting secretory IgA (IgAs) levels in mother milk and B and T cell generations [26]. In the absence of immune defenses, microbial colonization and propagation cannot be driven by the host [25]. The effects of the immune system are instead fully effective in Balb/c mice. In this study, 18 baby mice (namely 1–18) from: i) Balb/c ♀ × $Rag2^{ko}$ ♂ at days 3 (samples 1–3), 7 (samples 4–6) and 14 (samples 7–9); and ii) $Rag2^{ko}$ ♀ × Balb/c ♂ at days 3 (samples 10–12), 7 (samples 13–15) and 14 (samples 16–18) were sacrificed and their entire intestines were processed (Fig. 1A and B). Samples were homogenized and subjected to the bacterial enrichment procedure to obtain EBS fractions (Fig. 2A) [24], for further MALDI-TOF MS-based bacterial IDs and metaproteomic analyses (Fig. 2B and C).

3.1. *Microbiota analysis by axenic culture-based MALDI-TOF MS approach*

In order to determine the early mouse gut microbiota composition, from the streaking on plates of bacterial liquid cultures an overall of 50 distinct individual colonies, characterized by different morphological (*e.g.*, size, shape and color characteristics) and metabolic (*e.g.*, aerobic, anaerobic, microaerophilic conditions and differential growth factor requirements) features (types), were selected, and picked from the plates after a 48–72 hour growth time. The lowest bacterial cell density (10^5 colony forming units, CFU/ml) was obtained for Balb/c and $Rag2^{ko}$ mice at 3 days, probably correlated to the early mouse life stage, whereas the highest (10^7 CFU/ml) was calculated for $Rag2^{ko}$ mice at 14 days. By MALDI-TOF MS Biotyper we identified a total of 14 different species, showing a different distribution linked to the mouse age and genetic background. Each mouse belonging to the same time group presented a similar composition at bacterial species level which resulted in a small variation of the distribution of Proteobacteria and Firmicute levels (Fig. 2B and C). At 3 days we can actually deal with only or prevalently mucosal adherent bacteria in mouse because of the scarceness of the fecal content. Indeed, if we think that a human gastrointestinal (GI) tract from an adult is about 200–300 m^2 [16] and is colonized by $10^{13–14}$ bacteria of 400 different species, it is expectable that the intestine of a 3 days old mouse, which is 1.1 cm^2 (Fig. 1), can host theoretically a number of bacteria which should be 10^7, because of a 6 Log of difference between the mouse gut area at 3 days of life and the adult human gut area. However, the issue to point out is that the number of bacteria is closely related to the topographic complexity of the GI tract: it is extremely simple in the first days of life when the GI tract is very small, with the consequence that in the early days after birth there is very low variety of bacterial species composing the microbiota, as demonstrated by Koenig et al. [15], who have introduced a

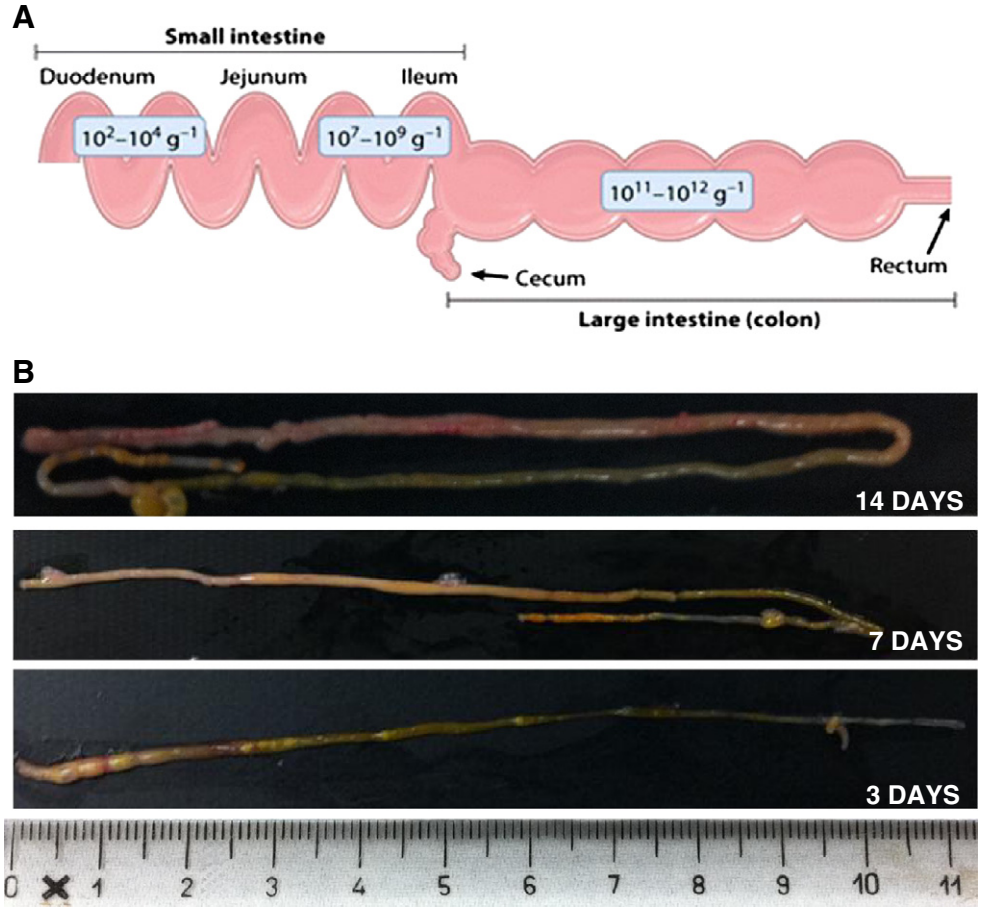

Fig. 1 – The mouse gut anatomy. A schematic representation of a mouse intestine (panel A) and images describing the organ accretion at 3, 7 and 14 days (panel B). The scheme A was modified from Kleerebezem and Vaughan (Annu Rev Microbiol 2009;63:269–90).

parameter of phylogenetic diversity (PD), assessing the gut species multiplicity and hence its complexity in species [15]. Indeed, we can expect that the species present at the very early stage of the mouse life, can be much less than 400, most of them probably not cultivable. In Fig. 3, it is clearly shown that gut microbiota of Balb/c and $Rag2^{ko}$ baby mice were composed exclusively of microbial isolates belonging to Firmicutes and Proteobacteria phyla. In particular, the gut microbiota at 3 days was similar in both types of mice, with a predominance of Firmicutes (95% in Balb/c and 75% in $Rag2^{ko}$ mice) and lower levels of Proteobacteria (Fig. 3). During the succeeding time points (7 and 14 days), the relative abundance of Proteobacteria in Balb/c mice started to increase reaching 50% of the entire gut microbiota at 14 days (Fig. 3). At the same time points, $Rag2^{ko}$ mice showed higher level of Proteobacteria than Balb/c mice, reaching the level of 75% at 14 days (Fig. 3). Regarding the composition at bacterial species level within the Proteobacteria phylum, the $Rag2^{ko}$ baby mice at 3 days of age, presented numerous bacterial isolates belonging to *Pasteurella pneumotropica* (about 20%), while only *Escherichia coli* isolates were identified in the subsequent time points (data not shown). The bacterial core, even if limited to the cultivable bacterial fraction, result is dependent on the genetic background and on the gut microbiota development stages (mouse age). Remarkably, the bacterial cultivable fraction was consistent with that described in the pivotal papers on the onset and development of the mouse gut microbiota [27–29]. Moreover, in these papers, the description of the relative amounts of the main bacteria groups in the first days of mouse life showed that actually, it is limited not only on the bacterial complexity (*e.g.*, only few OTUs) but also on the number of bacterial cells ascribed to each bacterial type (*e.g.*, 10^3 Lactobacilli and Streptococci or 10^1 Enterococci at 2 days).

No bacterial isolates belonging to other phyla different from Proteobacteria and Firmicutes (in particular Bacteroidetes and Actinobacteria) were characterized by our culture-based approach, probably due to their demanding culture condition required for bacterial *in-vitro* growth.

3.2. *Microbiota analysis by metaproteomic–bioinformatic pipeline*

Database searches performed with Mascot 2.3 running on a resident machine returned an overall of 997 protein hits for all mice (compared to 892 hits for the corresponding intra-assay replicates) ranging an average number of 46 ± 7 for Balb/c mice at 3 days (samples 1–3) to 83 ± 2 for $Rag2^{ko}$ mice at

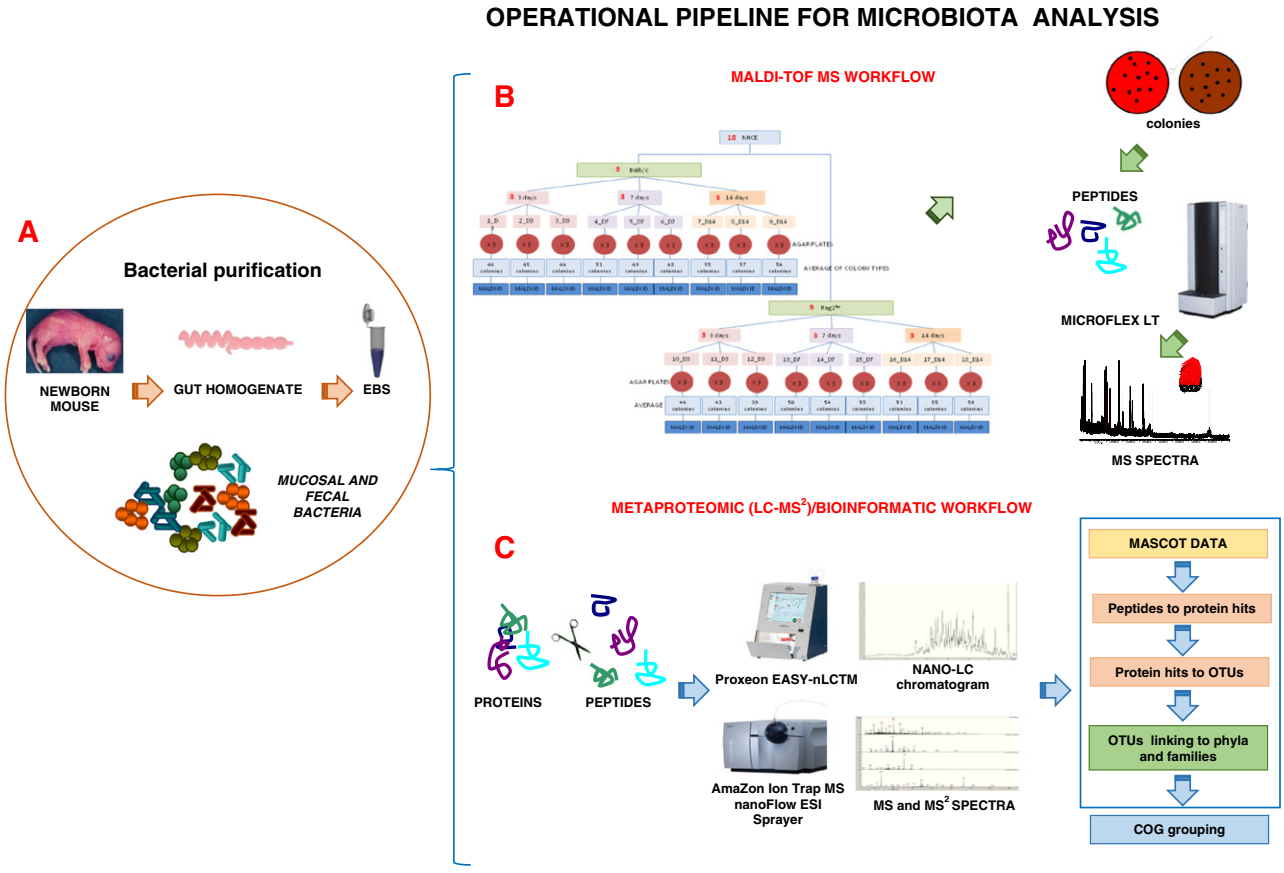

Fig. 2 – Experimental pipeline employed to characterize the mouse gut microbiota. The pipeline consisted of three steps: bacterial enrichment from mouse intestine (panel A); microbial ID assessment by MALDI-TOF MS Biotyper (panel B); LC–MS^2-based metaproteomic/bioinformatic pipeline (panel C).

14 days (sample 16–18) (SM9). Indeed, the lowest number of protein hits was found for Balb/c and Rag2^{ko} mice at 3 days (samples 1–3 and 10–12, 137 and 139 protein hits, respectively;

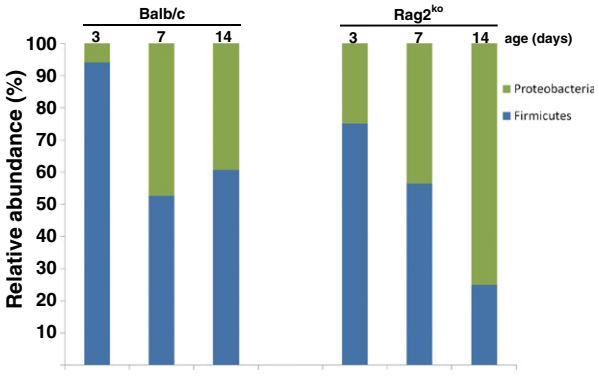

Fig. 3 – Taxonomic classification of Balb/c and Rag2^{ko} gut microbiota characterized by axenic-based MALDI-TOF MS approach. Bar charts showing the relative abundance (%) of bacterial phyla characterized from Balb/c and Rag2^{ko} mice intestine at different time points (3, 7 and 14 days) are presented. Bacterial isolates are described at phyla level. Each bar refers to a group of three mice.

SM9), probably due to the physiology of the intestinal maturation, and the bacterial population dynamics in the early stages of intestinal colonization of mouse gut [30]. A good technical reproducibility was obtained (an overall of 710/997 protein hits were found; 71.2%) ranging from 61.9% in Balb/c mice at 7 days (sample 4) to 91.3% in Rag2^{ko} mice at 7 days (sample 14) (SM9). Even if a low number of protein hits was found, a small variation in their levels was calculated by standard deviation within the three mice of the same time group, thus supporting the validity of the experimental setting presented in this manuscript (SM9). The comparison of mouse protein hits belonging to the same time group (biological replicates) revealed that a consistent group of proteins hits was shared among them (SM9). In particular, the level of overlay was comprised between 59.1% in Balb/c mice at 3 days and 88.7% in Rag2^{ko} mice at 7 days when a protein hit was shared by 2 out of the 3 mice of the same time group (SM9).

An analysis performed using Mascot coupled with a mouse reference proteome returned in an overall of 96 protein hits (SM10), approximately 1/10 of the entire protein catalog obtained with the Mascot Eubacteria database. Fifty-nine out of the 96 protein hits (62%) were related to only 9 proteins, namely 78 kDa glucose-regulated protein, heat shock 70 kDa protein, serum albumin, microsomal triglyceride transfer

protein large subunit, liver fatty acid-binding protein, inactive pancreatic lipase-related protein 1, melanotransferrin, nesprin and titin (SM10).

All protein hits were classified among 17 different phyla (Fig. 4A). Interestingly, the number of inferred species, ranging from 100 to 200 for each mouse genetic background at each time point of the growth (SM11), overcame 1 Log of the number of cultivable species identified by MALDI-TOF MS. High level of protein hits related to Firmicutes and Proteobacteria was found in both mouse offsprings, while only Balb/c mice showed significant levels (>5%) of Bacteroidetes and Actinobacteria. Firmicutes and Proteobacteria protein hits were differentially found in Balb/c and Rag2^{ko} baby mice. At 3 days Balb/c mice presented with 75% of hits belonging to Firmicutes and 20% associated to Proteobacteria, while an inverse distribution was observed in Rag2^{ko} mice (Fig. 4A). In the subsequent time points, the level of protein hits related to Proteobacteria started to increase in both mouse lineages, reaching at 14 days, by 50% of the overall phyla in Balb/c and the astonishing level of 90% in Rag2^{ko} mice (Fig. 4A). In the same manner, at 7 and 14 days the levels of Actinobacteria and Bacteroidetes protein hits belonged exclusively to Balb/c (5%, and 10%, respectively) (Fig. 4A). Similar results were obtained by intra-assay replicates (Fig. S1).

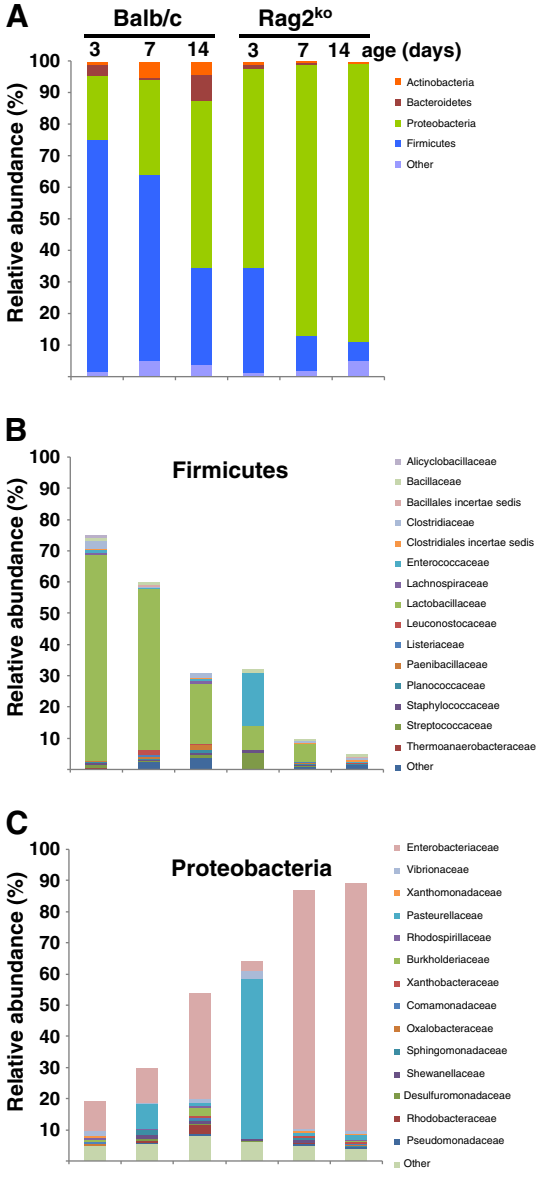

Fig. 4 – Taxonomic classification of Balb/c and Rag2^{ko} gut microbiota characterized by metaproteomic pipeline. Bar charts showing the relative abundance (%) of major bacterial phyla in the gut microbiota of Balb/c and Rag2^{ko} mice, are reported. Each bar represents a group of three mice (panel A). Inset on detailed relative abundance (%) of families in Firmicutes and Proteobacteria phyla (panels B and C, respectively).

To get further information about the composition of mouse gut microbiota, we perform an in-depth taxonomic analysis at the level of bacterial family for all protein hits characterized by the metaproteomic pipeline. With this aim, bacterial family level of protein hits within the most abundant Firmicutes and Proteobacteria phyla was calculated (Fig. 4B and C). In Balb/c mice, for all time points, within the Firmicutes phylum, a predominant number of Lactobacillaceae-related protein hits (>60%) was computed (Fig. 4B). In Rag2^{ko}, protein hits related to Streptococcaceae (20%), Lactobacillaceae (25%) and Listeriaceae (50%) families were found at 3 days (Fig. 4B), while at 7 and 14 days Rag2^{ko} presented a completely different taxonomy with variable levels of protein hits related to Lactobacillaceae (from 50% to 2% at 7 and 14 days, respectively), progressively loosing Streptococcaceae and Listeriaceae (Fig. 4B). Regarding Proteobacteria, the analysis at the family level of all protein hits, showed that the great majority was ascribed to the Enterobacteriaceae family in both mouse lineages and for all time points (with the exception of Rag2^{ko} mice at 3 days of age) (Fig. 4C).

We emphasize that the relative abundance of protein hits related to Enterobacteriaceae family in Rag2^{ko} mice at 7 and 14 days, was greater than 90%, while at the same time points, it did not exceed the 50% in Balb/c mice (Fig. 4C). The bacterial family composition of protein hits belonging to Proteobacteria phylum in Rag2^{ko} mice at 3 days was completely different to that present at 7 and 14 days, with low level of Enterobacteriaceae (<5%) and very high level of Pasteurellaceae (>85%) (Fig. 4C). The different bacterial family composition within Proteobacteria phylum of Rag2^{ko} mice at 3 days was confirmed by culture-based MALDI-TOF MS approach in which a relevant level of bacterial isolates belonging to Pasteurellaceae (20%) and low abundance of Enterobacteriaceae were found (see above). Intriguingly, in Rag2^{ko} mice at 7 and 14 days a complete replacement of Pasteurellaceae to Enterobacteriaceae was observed (see above).

By comparing culture-based MALDI-TOF MS and metaproteomic results for both mice offsprings, opposite trends for Proteobacteria and Firmicute relative abundance were observed through the time-course (Figs. 3 and 4A). However, the Proteobacteria abundance was higher in Rag2^{ko} than in Balb/c mice. Indeed, at 14 days, Proteobacteria in Rag2^{ko} represented the 90% and 75% of the total gut microbiota, by metaproteomics and MALDI-TOF MS IDs, respectively, while at the same time-point Balb/c mice presented significant lower levels (40% and 55%, respectively) (Figs. 3 and 4A). The results herein presented seem to suggest a link between mouse genetic background and gut microbiota shaping, showing an important

increase in Proteobacteria, a partial or total reduction in Firmicutes and Actinobacteria, and Bacteroidetes, respectively, in the case of Rag2^{ko} mice.

Most of the published works on the characterization and variation of human gut microbiota through life stages and during diseases have been approached by metagenomics [15,31,32], and only few by metaproteomic pipelines applied to adult subjects [19,20,22,23]. Since the correct development/maturation of mammal gut microbiota in adults is intimately correlated with programming processes established since the very early life stages, it becomes essential to set up specific procedures to be able to highlight and to characterize gut microbial structure and alteration in childhood, by using appropriate models especially for the very early life stages, when competition between mucosal and fecal bacteria induced the next microbiota shaping. For these reasons, we carried out a metaproteomic pipeline applied to newborn mouse gut samples, despite operational difficulties due to sample scarceness, organ homogenate preparation and low bacterial complexity. The workflow was integrated with an automatic bioinformatic algorithm for gut microbiota OTU generation and the obtained results were compared to axenic culture-based MALDI-TOF MS evidences.

3.3. *Protein classification by COG categories*

Among the 997 Mascot protein hits, only hits belonging to Firmicutes and Proteobacteria major phyla (946/997) were subjected to COG categorization, as shown in Fig. 5. According to the COG nomenclature (http://www.ncbi.nlm.nih.gov/COG/), we grouped similar COG categories into four main classes, namely: *i*) information storage and processing; *ii*) cellular processes and signaling; *iii*) metabolism; and *iv*) poorly characterized. The most prevalent proteins belonged to (*i*) class, mainly including elongation factors (*e.g.*, Tu, Ts, G), and ribosomal proteins (*e.g.*, 30S and 50S). The (*ii*) class is presented with mostly proteins belonging to chaperons (*e.g.*, DnaK, Skp, GroEL and GroES), outer membrane protein A, flagellin, thioredoxin and peroxiredoxin. The (*iii*) class is comprised of proteins belonging to enolase, phosphoglycerate, transaldolase, phosphoglycerate kinase, deoxyribose-phosphate aldolase, L-lactate dehydrogenase, formate dehydrogenase, amino acid biosynthesis and degradation. Finally, the (*iv*) class included proteins with general on unknown function. As shown in Fig. 5, COG groups' comparison in Balb/c and Rag2^{ko} mice, revealed absence of time-dependant significant variation.

To get further functional insights, the 946 Mascot protein hits were screened for the presence of proteins related to specific metabolic pathways, differentially expressed in Balb/c or Rag2^{ko} mice. The results allowed us to characterize, within the Enterobacteriaceae family from Rag2^{ko} mice, recurrent presence of protein hits related to: *i*) peroxiredoxin (EC 1.11.1.15) and thioredoxin (EC 1.17.4.1) linked to glutathione metabolism; *ii*) riboflavin synthase (EC 2.5.1.9) associated to riboflavin metabolism; and *iii*) pentose phosphate pathway related to ribose-phosphate pyrophosphokinase (EC 2.7.6.1), transaldolase (EC 2.2.1.2) and deoxyribose-phosphate aldolase (EC 4.1.2.4).

As reported for human Crohn's disease (CD) [33,34], the Rag2^{ko} functional picture may be associated to a high presence of the family Enterobacteriaceae. Particularly, also Morgan et al, found an increased abundance of genes associated to glutathione, riboflavin metabolism and pentose phosphate pathway [33]. Glutathione is a molecule specifically synthesized by Proteobacteria to maintain gut homeostasis during oxidative stress caused by inflammation [35]. Indeed, oxidized and reduced glutathione conversion forms (riboflavin metabolism) and pentose phosphate pathway (glutathione reduction), are recruited in response to the presence of reactive oxygen and nitrogen metabolites in the gut of CD patients [34,35]. The high level of Enterobacteriaceae in the gut microbiota of both Rag2^{ko} and CD patients, might be the result of a positively driven selection of microbes carrying pathways for oxidative stress response *via* glutathione production and metabolism [33,34]. Clearly the definition of functional ontological classification would need higher dimensionality data sets, and the

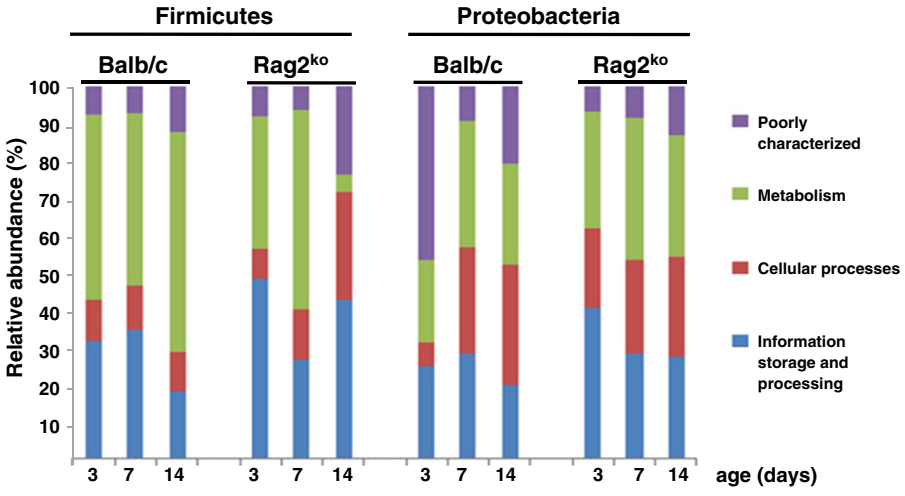

Fig. 5 – COG functional category distribution associated to Firmicutes and Proteobacteria phyla. Bar charts show the COG relative abundance (%) of Mascot protein hits associated to Firmicutes and Proteobacteria phyla from Balb/c and Rag2^{ko} mice at different time points (3, 7 and 14 days). Each bar represents COG-groups retrieved from three mice.

automatization of the OTUs computation herein presented combined to the engineering of mass spectrometers [36–38] will certainly play a central role in the future of metaproteomic workflows.

4. Conclusion

We developed a metaproteomic and computational-combined pipeline which is able to characterize newborn mouse gut microbiota. The design and setup of this procedure are adequately wide-ranging to be employed with specific *ad-hoc* customized databases for further metaproteomic study on other host ecosystems in very early life stages, when mucosal and fecal bacteria play their initial fight for the gut surface conquest, imposing the next microbiota shaping from very low to adult community complexity. We complemented our procedure by comparing gut phylotype charts with those obtained with axenic culture-based MALDI-TOF MS IDs. Intriguingly, agreement between both approaches, was obtained for either different mouse individuals, diverse genetic backgrounds and programming time-course. However, a deep analysis of Mascot protein hits retrieved from public repositories (*e.g.*, NCBI and UniProt), highlighted the presence of bacterial species usually belonging to extreme life ecological niches (*e.g.*, extremophilic bacteria) [39]. For these reasons, we strongly believe that a sensitive microbiological filter, is also required to correctly screen and evaluate "gut-compatible" bacterial OTUs related to proteins characterized by a metaproteomic pipeline.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2013.10.025.

Acknowledgments

This work was supported by Ricerca Corrente (RC 201302P002991 and RC 201302G003050) "The proteomics in microbiology: from single pathogens to systems biology", Bambino Gesù Children's Hospital, IRCCS, to LP, ItPA mobility fellowship 2011 to PV, and Telethon GGP07252, Fondazione Roma 2008, Rete Nazionale di proteomica FIRB RBRN07BMCT to AU.

REFERENCES

[1] Putignani L, Carsetti R, Signore F, Manco M. Additional maternal and nonmaternal factors contribute to microbiota shaping in newborns. Proc Natl Acad Sci U S A 2010;107:E159 [author reply E160].
[2] Kinross JM, Darzi AW, Nicholson JK. Gut microbiome–host interactions in health and disease. Genome Med 2011;3:14.
[3] Nicholson JK, Lindon JC. Systems biology: metabonomics. Nature 2008;455:1054–6.
[4] Dupré J, O'Malley MA. Metagenomics and biological ontology. Stud Hist Philos Biol Biomed Sci 2007;38:834–46.
[5] Relman DA. Microbial genomics and infectious diseases. N Engl J Med 2011;365:347–57.
[6] Pflughoeft KJ, Versalovic J. Human microbiome in health and disease. Annu Rev Pathol 2012;7:99–122.
[7] Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 1998;5:R245–9.
[8] Manco M, Putignani L, Bottazzo GF. Gut microbiota, lipopolysaccharides, and innate immunity in the pathogenesis of obesity and cardiovascular risk. Endocr Rev 2010;31:817–44.
[9] Zoetendal EG, Rajilic-Stojanovic M, de Vos WM. High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. Gut 2008;57:1605–15.
[10] Martins dos Santos V, Müller M, de Vos WM. Systems biology of the gut: the interplay of food, microbiota and host at the mucosal interface. Curr Opin Biotechnol 2010;21:539–50.
[11] Salonen A, de Vos WM, Palva A. Gastrointestinal microbiota in irritable bowel syndrome: present state and perspectives. Microbiology 2010;156:3205–15.
[12] Thompson CL, Hofer MJ, Campbell IL, Holmes AJ. Community dynamics in the mouse gut microbiota: a possible role for IRF9-regulated genes in community homeostasis. PLoS One 2010;5:e10335.
[13] Esworthy RS, Smith DD, Chu FF. A strong impact of genetic background on gut microflora in mice. Int J Inflamm 2010;2010:986046.
[14] Friswell MK, Gika H, Stratford IJ, Theodoridis G, Telfer B, Wilson ID, et al. Site and strain-specific variation in gut microbiota profiles and metabolism in experimental mice. PLoS One 2010;5:e8584.
[15] Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A 2011;108:4578–85.
[16] Hao WL, Lee YK. Microflora of the gastrointestinal tract: a review. Methods Mol Biol 2004;268:491–502.
[17] Li X, LeBlanc J, Truong A, Vuthoori R, Chen SS, Lustgarten JL, et al. A metaproteomic approach to study human-microbial ecosystems at the mucosal luminal interface. PLoS One 2011;6:e26542.
[18] Rooijers K, Kolmeder C, Juste C, Doré J, de Been M, Boeren S, et al. An iterative workflow for mining the human intestinal metaproteome. BMC Genomics 2011;12:6.
[19] Kolmeder CA, de Been M, Nikkilä J, Ritamo I, Mättö J, Valmu L, et al. Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. PLoS One 2012;7:e29913.
[20] Ferrer M, Ruiz A, Lanza F, Haange SB, Oberbach A, Till H, et al. Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure. Environ Microbiol 2013;15:211–26.
[21] Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, et al. Shotgun metaproteomics of the human distal gut microbiota. ISME J 2009;3:179–89.
[22] Erickson AR, Cantarel BL, Lamendella R, Darzi Y, Mongodin EF, Pan C, et al. Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. PLoS One 2012;7:e49138.
[23] Haange SB, Oberbach A, Schlichting N, Hugenholtz F, Smidt H, von Bergen M, et al. Metaproteome analysis and molecular genetics of rat intestinal microbiota reveals section and localization resolved species distribution and enzymatic functionalities. J Proteome Res 2012;11:5406–17.
[24] Apajalahti JH, Särkilahti LK, Mäki BR, Heikkinen JP, Nurminen PH, Holben WE. Effective recovery of bacterial DNA and percent-guanine-plus-cytosine-based analysis of community structure in the gastrointestinal tract of broiler chickens. Appl Environ Microbiol 1998;64:4084–8.
[25] Shulzhenko N, Morgun A, Hsiao W, Battle M, Yao M, Gavrilova O, et al. Crosstalk between B lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut. Nat Med 2011;17:1585–93.

[26] Spanopoulou E. Cellular and molecular analysis of lymphoid development using Rag-deficient mice. Int Rev Immunol 1996;13:257–88.
[27] Schaedler RW, Dubos R, Costello R. The development of the bacterial flora in the gastrointestinal tract of mice. J Exp Med 1965;122:59–66.
[28] Lee A, Gordon J, Lee CJ, Dubos R. The mouse intestinal microflora with emphasis on the strict anaerobes. J Exp Med 1971;133:339–52.
[29] Savage DC, Dubos R, Schaedler RW. The gastrointestinal epithelium and its autochthonous bacteria flora. J Exp Med 1968;127:67–76.
[30] Mackie RI, Sghir A, Gaskins HR. Developmental microbial ecology of the neonatal gastrointestinal tract. Am J Clin Nutr 1999;69:1035S–45S.
[31] Schwartz S, Friedberg I, Ivanov IV, Davidson LA, Goldsby JS, Dahl DB, et al. A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in immune response. Genome Biol 2012;13:R32.
[32] Biagi E, Nylund L, Candela M, Ostan R, Bucci L, Pini E, et al. Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. PLoS One 2012;5:e10667.
[33] Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 2012;13:R79.
[34] Keshavarzian A, Banan A, Farhadi A, Komanduri S, Mutlu E, Zhang Y, et al. Increases in free radicals and cytoskeletal protein oxidation and nitration in the colon of patients with inflammatory bowel disease. Gut 2003;52:720–8.
[35] Sherrill C, Fahey RC. Import and metabolism of glutathione by *Streptococcus mutans*. J Bacteriol 1998;180:1454–9.
[36] Hettich RL, Sharma R, Chourey K, Giannone RJ. Microbial metaproteomics: identifying the repertoire of proteins that microorganisms use to compete and cooperate in complex environmental communities. Curr Opin Microbiol 2012;15:373–80.
[37] Hettich RL, Pan C, Chourey K, Giannone RJ. Metaproteomics: harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. Anal Chem 2013;85:4203–14.
[38] Kolmeder CA, de Vos WM. Metaproteomics of our microbiome — developing insight in function and activity in man and model systems. J Proteomics 2013;24 [S1874-3919(13)00259-5].
[39] Petrucca A, Del Chierico F, Putignani L. Just keep going on description or interpret “life” of gut microbial ecosystems? target on line cell comment (http://www.cell.com/abstract/s0092-8674(12), 00629-0#comments) to gut immune maturation depends on colonization with a host-specific microbiota. Cell 2012;149:1578–93.