environmental microbiology reports

Environmental Microbiology Reports (2016) **8**(3), 399–406 doi:10.1111/1758-2229.12405

Mining metagenomic whole genome sequences revealed subdominant but constant *Lactobacillus* population in the human gut microbiota

Maddalena Rossi,[1]*† Daniel Martínez-Martínez,[2,3,4]† Alberto Amaretti,[1] Alessandro Ulrici,[1] Stefano Raimondi[1] and Andrés Moya[2,3,4]

[1]*Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy.*

[2]*Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Valencia, Spain.*

[3]*Foundation for the Promotion of Health and Biomedical Research in the Valencian Community (FISABIO), Valencia, Spain.*

[4]*Network Research Center for Epidemiology and Public Health (CIBER-ESP), Madrid, Spain.*

Summary

The genus *Lactobacillus* includes over 215 species that colonize plants, foods, sewage and the gastrointestinal tract (GIT) of humans and animals. In the GIT, *Lactobacillus* population can be made by true inhabitants or by bacteria occasionally ingested with fermented or spoiled foods, or with probiotics. This study longitudinally surveyed *Lactobacillus* species and strains in the feces of a healthy subject through whole genome sequencing (WGS) data-mining, in order to identify members of the permanent or transient populations. In three time-points (0, 670 and 700 d), 58 different species were identified, 16 of them being retrieved for the first time in human feces. *L. rhamnosus*, *L. ruminis, L. delbrueckii, L. plantarum, L. casei* and *L. acidophilus* were the most represented, with estimated amounts ranging between 6 and 8 Log (cells g^{-1}), while the other were detected at 4 or 5 Log (cells g^{-1}). 86 *Lactobacillus* strains belonging to 52 species were identified. 43 seemingly occupied the GIT as true residents, since were detected in a time span of almost 2 years in all the three samples or in 2 samples separated by 670 or 700 d. As a whole, a stable community of lactobacilli was disclosed, with wide and understudied biodiversity.

Introduction

The human colon is inhabited by a complex microbial community composed largely of bacteria, whose numbers reaches up to 10^{12} cells per gram of intestinal content, belonging to over thousands species (O'Hara and Shanahan, 2006). The host and associated microbiota form an interrelated ecosystem that has major effects on health or diseases. Differences occurring in microbial composition and in metabolic activities deeply influence the health status, since they could promote a healthy homeostasis or could determine the basis for the onset of pathologies (Sekirov *et al.*, 2010). Most of colonic bacteria are members of a resident population that has a long-term association with the host. Other microorganisms found in the colon can inhabit other sections of the gastrointestinal tract (GIT), such as the mouth or the small intestine, shedding alive to the colon but being unable to replicate in this habitat (Xu and Gordon, 2003). Furthermore, microorganisms ingested with food and water can transit through the upper intestine and reach alive the hindgut, after being challenged by the low pH of the stomach, the digestive enzymes, and the toxicity of bile salts. Lactobacilli are quite versatile in terms of ecosystems, and can fall into all these groups.

The genus *Lactobacillus* includes 217 recognized species enlisted in LPSN database (The List of Prokaryotic Names with Standing in Nomenclature, http://www.bacterio.net, last accessed March 2016) (Parte, 2014). It belongs to the family *Lactobacillaceae* that, with *Enterococcaceae* and *Streptococcaceae*, is included in the order of *Lactobacillales* within to the gram-positive phylum of *Firmicutes*. Despite their wide phylogenetic and functional diversity, lactobacilli are invariably saccharolytic, anaerobic or microaerophilic, aciduric or acidophilic, non-sporulating rods (Hammes and Hertel, 2006). Lactobacilli, enterococci and streptococci are included in the functional group of Lactic Acid Bacteria (LAB), since they catabolize hexoses to lactic acid through obligate or facultative homolactic or heterolactic fermentation, the

Received 12 February, 2016; accepted 8 March, 2016. *For correspondence. E-mail maddalena.rossi@unimore.it; Tel. 39-059-2058589; Fax: 39-059-245156. †These authors contributed equally to the work.

latter yielding also two-carbon products (ethanol and/or acetic acid) and CO_2.

Lactobacilli inhabit a variety of habitats where carbohydrate-based substrates are available, such as plants, plant-derived matrices, silage, fermented foods (e.g. dairy products, fermented dough, milk, vegetables and meats), spoiled foods, organic matrices and sewage (Hammes and Hertel, 2006). They are also found in the commensal microbiota naturally colonizing diverse districts within the body of humans and animals, such as the oral cavity, the GIT and the vagina (O'Donnell *et al*., 2013; Douillard and de Vos, 2014). Moreover, humans have safely been ingesting lactobacilli for centuries in fermented foods and beverages, and more recently, in probiotic products. This makes the role of lactobacilli in the ecology of the intestinal microbiota difficult to determine, since they represent a peculiar microbial group, whose population can be made up by transient members or by true inhabitants of diverse districts of the GIT (Walter, 2008).

Despite the positive impact on health of lactobacilli and the use as probiotics, reports providing a detailed description of *Lactobacillus* population in the human intestinal microbiota, in terms of species and strains, are still rare. Several studies aimed to describe the diversity of *Lactobacillus* species have been performed through a cultivation step, introducing some bias. Bacteria belonging to the genus *Lactobacillus* are expected to be cultivable, then easily isolated and enumerated by traditional culture methods (Hartemink and Rombouts, 1999; Van de Casteele *et al*., 2006). However, unlike lactobacilli in food matrices (e.g. dairy products, meat and vegetables) which can be easily retrieved on selective plates, isolation from human feces is hampered by bifidobacteria, that present similar nutritional and environmental requirements but are much more numerous (Quartieri *et al*., 2016). Investigations based on cultivation of fecal lactobacilli indicated some recurring species such as *L. acidophilus*, *L. brevis*, *L. casei*, *L. crispatus*, *L. delbrueckii*, *L. fermentum*, *L. gasseri*, *L. jensenii*, *L. johnsonii*, *L. mucosae*, *L. paracasei*, *L. plantarum*, *L. reuteri*, *L. rhamnosus*, *L. ruminis*, *L. salivarius* and *L. vaginalis* (Reuter, 2001; Dal Bello *et al*., 2003; Ahrnè *et al*., 2005; Delgado *et al*., 2007; Lönnermatk *et al*., 2012; Öztürk and Meterelliyöz, 2015). Consistently, species-specific amplification and sequencing of 16S rRNA gene confirmed the presence of several of the above-cited species (Heilig *et al*., 2002).

Presently, traditional microbiological methods are overtaken by data provided by next-generation sequencing technologies, such as metagenomic approaches that disclose the relative amounts of the different microorganisms colonizing this habitat (Frank and Pace, 2008). Two different ways are accessible at the moment: 16S rRNA gene sequencing and whole genome sequencing (WGS). 16S rRNA gene sequencing is the most widely exploited for metagenome survey, but in most cases the libraries generated from fecal, colonic or cecal samples of humans fail to retrieve lactobacilli, likely because of their relatively small amounts (Walter, 2008). WGS offers several advantages, such as making available information about the genome of the microorganisms, opening the possibility to analyze their functional capabilities, and providing a picture of the taxonomic composition of microbiota at species- and strain-level, without the bias of 16S copy number variation and polymerase chain reaction amplification. In these latter techniques, it is very difficult to reach taxonomic levels of species or strains, whereas shotgun approaches with good coverage could add statistical power to the taxonomical assignments even in low abundances.

In the present study, a mining approach was applied to WGS data already available (Minot *et al*., 2013), with the aim to longitudinally characterize the fecal lactobacilli of a healthy subject over a two years period.

Results and discussion

Metagenomic WGS sequences were utilized to explore *Lactobacillus* biodiversity of human gut microbiota and to discriminate between true inhabitants and transient bacteria. The data, acquired in a previous longitudinal study aimed to explore the variation of the intestinal virome of a male healthy subject (Minot *et al*., 2013). In our investigation, three time-points were considered and compared (0, 670 and 700 d) in order evaluate short-term (between 670 and 700 d) and long-term variations (between 0 and 670 or 700 d). WGS sequences were analysed using the LMAT-Grand database encompassing complete and partial genome sequences from the NCBI genome database (July 4, 2014), that included genomes from 78 different *Lactobacillus* species, with 326 strains identified (Ames *et al*., 2013). Analysis were carried out through the LMAT pipeline, using the Average Read Score to obtain the relative amounts, normalized against the total reads. For more details on the sequence management and the metagenome processing, please refer to Supporting Information Appendix S1.

The relative amounts of the *Lactobacillales* classifiable in the families of *Enterococcaceae Lactobacillaceae* and *Streptococcaceae* were determined. A special attention was paid to the genus *Lactobacillus* because of the diverse environments that different species can occupy, and for the health promoting properties ascribed to intestinal colonizers. The bacterial concentration of *Lactobacillus* species and strains was roughly extrapolated from the respective relative amount, assuming that the three fecal samples had the same concentration of microorganisms (1×10^{12} cells g^{-1}). In order to take in

account this approximations, only the order of magnitude was considered as reliable and the data were expressed as Log (cells g^{-1}). Doing this assumption, major approximations were introduced for the number of cells per g, which is likely different among the three samples. However, the approximation seems reasonable for the aim of comparing the order of magnitude of bacterial abundance, expressed as the Log of the number of cells per g of feces.

Lactobacillales

WGS led to the identification of a total of 2.1 10^7 OTUs. At the three time-points, the microbiota was dominated by bacteria, which accounted for the 75%, 80% and 74% of total microbes at 0, 670 and 700 d, respectively. The phyla *Firmicutes* and *Bacteroidetes* were always the most represented, whereas the 19%–26% of the bacterial population remained unassigned at taxonomic level (Table 1). *Firmicutes* accounted for 21%–36% of total bacteria, with the class of *Bacilli* ranging between 1% and 3% of them. The order *Lactobacillales* accounted for 47%, 21% and 24% of *Bacilli* at the three time-points, corresponding to 0.08%–0.3% of the whole bacterial population. The majority of *Lactobacillales* (>92%) were ascribable to the families of *Streptococcaceae*, *Lactobacillaceae* and *Enterococcaceae*, in order of abundance (Table 1).

Streptococcaceae were mostly represented by bacteria belonging to the genus *Streptococcus* (89%–97%) that includes both beneficial commensal species, such as *S. thermophilus*, and opportunistic or pathogenic species. Pathogenic species (Krzyściak *et al*., 2013) of *Streptococcus* were not identified in the three fecal samples. The species *S. parasanguinis*, *S. mitis*, *S. infantis* and *S. australis* which are part of the endogenous microbiota of the oral cavity according to the human oral microbiome database (http://www.homd.org/, last accessed March 2016), accounted at most for the 5% of streptococci. 59 to 66 *Streptococcus* species were identified in each sample. Among the health promoting and pro-technological species, *S. thermophilus* ranged from 5% to 27% of streptococci and *S. salivarius* from 4% to 11%. Altogether, *S. thermophilus* and *S. salivarius* represented a rather variable percentage of total bacteria (0.09%, 0.005% and 0.008% at 0, 670 and 700 d respectively).

The bacteria belonging to the family of *Enterococcaceae* were mostly represented by the genus *Enterococcus* (93%–94%), which included 21 species. *E. cecorum*, *E. faecium* and *E. faecalis* were the most abundant enterococci, ranging from 5% to 14%, from 8% to 12% and from 5% to 9% respectively. The species *E. gilvus, E. avium, E. columbae, E. pallens, E. moraviensis, E. durans, E. italicus, E. phoeniculicola, E. caccae, E. saccharolyticus, E. dispar* and *E. haemoperoxidus* were common at the different time points, whereas *E. mundtii, E. malodoratus, E. asini* and *E. villutum* were found at both 670 and 700 d, and *E. casseliflavus* and *E. sulfureus* only at 700 d.

Table 1. Percentage of bacterial WGS reads attributed to the main phyla and to the class, order, and family of the main intestinal LABs.

Phylum/Class/Order/Family	0 d	670 d	700 d
Firmicutes	21	26	36
Bacilli	0.8	0.5	0.7
Lactobacillales	0.4	0.1	0.2
Streptococcaceae	0.3	0.05	0.08
Lactobacillaceae	0.04	0.03	0.03
Enterococcaceae	0.02	0.02	0.03
Bacteroidetes	47	52	34
Proteobacteria	6	4	6
Actinobacteria	1	0.7	1
Unassigned bacteria	21	18	22

The relative amounts were calculated normalizing the Total Read Score (TRS) of each taxa for the sum of all the TRS of the corresponding sample.

The genus *Lactobacillus* covered 94% to 97% of *Lactobacillaceae*, and ranged from 0.03% to 0.04% of bacteria. Excluding unassigned biotypes of the family, the other *Lactobacillaceae* belonged to the genus *Pediococcus*, phylogenetically intermixed with the genus *Lactobacillus*, and in particular to the species *P. claussenii*, *P. acidilactici* and *P. pentosaceus*, the two latter being exploited as starters for fermented foods (Leroy and De Vuyst, 2004). Interestingly, also *P. claussenii*, a common brewery contaminant, was present as a sole biotype in the three samples, being a stable colonizer of the gut of this subject during the 2-year period.

Lactobacillus *species*

Bacteria belonging to the genus *Lactobacillus* ranged from 8.4 to 8.6 Log (cells g^{-1}), with a great biodiversity of detected species, being the OTUs spread among 58 different ones out of the 217 recognized. The species ranged from 4 to 8 Log (cells g^{-1}). The most abundant ones (*L. rhamnosus*, *L. ruminis*, *L. delbrueckii*, *L. plantarum*, *L. casei* and *L. acidophilus*; Table 1) were the same identified in previous studies utilizing culture-based approaches (Reuter, 2001; Delgado *et al*., 2007; Wall *et al*., 2007; Rajilić-Stojanović and de Vos, 2014). Many other, especially among those occurring at 4 or 5 Log (cells g^{-1}), emerged for the first time in human fecal samples. This outcome confirmed the drawbacks associated to isolation of fecal lactobacilli occurring in lower amounts by cultural methods (Quartieri *et al*., 2016).

Most of the taxa identified (e.g. *L. animalis*, *L. acidophilus*, *L. brevis*, *L. buchneri*, *L. casei*, *L. delbrueckii, L. paracasei, L. pentosus, L. plantarum, L. reuteri, L. rhamnosus, L. rossiae, L. ruminis* and *L. sakei*) are not

Species	0 d	670 d	700 d
L. acidipiscis	4.7	5.2	5.5
L. acidophilus	8.1	7	4.6
L. amylolyticus	-	4.9	4.3
L. animalis	-	-	4.8
L. brevis	5.8	5.3	4.4
L. buchneri	5.7	5.5	5.6
L. casei	7.4	6.4	6.7
L. ceti	5.6	5.1	5
L. coleohominis	4.5	4.9	5.2
L. composti	4.7	4.8	4.9
L. coryniformis	-	5.5	5.3
L. crispatus	5.4	5	5.4
L. delbrueckii	6.6	7.2	7.4
L. equi	5.9	5	4.9
L. equicursoris	-	4.9	-
L. farciminis	4.7	4.8	4.2
L. farraginis	5.7	5.4	4.7
L. fermentum	7.2	5.3	5.3
L. florum	-	4.7	4.7
L. fructivorans	5.5	5.3	5.2
L. gasseri	5.9	5.4	4.9
L. gastricus	4.9	-	-
L. gigeriorum	-	4.7	5.1
L. harbinensis	5.4	5.5	5.4
L. helveticus	4.5	6.9	6.4
L. hominis	5.5	5.3	5.5
L. iners	6.3	5.8	6.7
L. jensenii	6.1	5.6	5.8
L. johnsonii	6	5.5	5.5

Species	0 d	670 d	700 d
L. kisonensis	5	4.8	-
L. mali	5.7	5.1	5.5
L. mucosae	6	4.8	4
L. murinus	7.4	6.9	7.1
L. namurensis	4.7	4.2	4.6
L. oris	5.8	6.2	6.3
L. otakiensis	4.9	4.7	-
L. parabrevis	5.4	5.2	5.3
L. paracasei	6.1	5.6	6.3
L. parafarraginis	5.2	5.6	5.6
L. pasteurii	5.6	-	-
L. pentosus	5.7	5	4.8
L. plantarum	7.4	6.7	6.7
L. pobuzihii	5	-	-
L. psittaci	5.2	-	5.3
L. reuteri	6.4	5.3	5.8
L. rhamnosus	8.2	7.9	8.1
L. rossiae	5.2	4.4	5.4
L. ruminis	7.7	8	7.2
L. sakei	5	-	-
L. salivarius	5.9	5.9	5.9
L. sanfranciscensis	8.6	6.1	6.5
L. shenzhenensis	5.5	5.3	5.6
L. suebicus	5.4	4.6	5.3
L. ultunensis	4.5	4.7	-
L. vaginalis	6.1	5.9	5.9
L. versmoldensis	5.6	5.8	5.1
L. vini	6.2	-	5.8
L. zeae	4.9	5.4	5.1

Fig. 1. Heatmap of *Lactobacillus* species in the feces of a healthy subject, at 0, 670 and 700 d. Values are Log (cells g^{-1}). Colours range from the lowest (deepest green) to the highest (deepest red) abundance. - indicates values falling below the detection limit of 3.9 Log (cells g^{-1}).

exclusively human commensals, but are found in a variety of other habitats, such as animals, foods, plants, organic spoiled matter or sewage (Hammes and Hertel, 2006). Interestingly, diverse species occurring in plants or foods (e.g. *L. acidophilus*, *L. casei*, *L. buchneri*, *L. kisonensis*, *L. parafaraginis* and *L. plantarum*), likely ingested with spoiled or fermented aliments, are also enlisted among the bacteria of the human oral microbiome database (http://www.homd.org/). On the other hand, *L. composti, L. farciminis, L. farraginis, L. florum, L. harbinensis, L. namurensis, L. otakiensis, L. parabrevis, L. pasteurii, L. pobuzihii, L. sanfranciscensis, L. shenzhenensis, L. suebicus, L. versmoldensis, L. vini, L. zeae* have been described as members of human fecal microbiota for the first time in the present study. These latter species are generally isolated on other organic matrices, such as plants, sewage, foods, organic matter, but have never been associated to humans or animals (Kröckel *et al*., 2003; Kitahara *et al*., 2005; Miyamoto *et al*., 2005; Rodas *et al*., 2006; Vancanneyt *et al*., 2006; Endo and Okada, 2007a,b; Scheirlinck *et al*., 2007; Watanabe *et al*., 2009; Chen *et al*., 2010; Endo *et al*., 2010; Kim *et al*., 2011; Nam *et al*., 2011a,b; Cousin *et al*., 2013; Zou *et al*., 2013). Conversely, *L. ceti*, *L. equicursoris*, *L. gigeriorum*, *L. equi*, and *L. psittaci* were found in animal samples, but never in human tissues or specimens (Cousin *et al*., 2012; Lawson *et al*., 2001; Morotomi *et al*., 2002; Vela *et al*., 2008; Morita *et al*., 2010). In order to determine which of these species are truly inhabitants of our gut, further insight would be necessary. As a matter of fact, there is always the possibility that the species found in this work could have an external origin as bacteria ingested with food. To address this point, thorough follow-up approaches are necessary and more work has to be done, including efforts in isolating these species or sequencing approaches, to help the evidence of their presence in the GIT as viable bacteria.

Of the 58 *Lactobacillus* species, 43 were retrieved in the three samples, 9 in two and 6 in one (Fig. 1). Among the species occurring in all the samples, *L. rhamnosus* dominated the *Lactobacillus* population at all the time-points (18.98%–56.74%), being the sole with a mean concentration in the magnitude 8 Log (cells g^{-1}). *L. ruminis*, *L. murinus*, *L. delbrueckii* and *L. sanfranciscensis* were in the magnitude of 7 Log (cells g^{-1}) and accounted on average from 4.13% to 18.96% of *Lactobacillus* reads

Strain	0	670 d	700 d
L. acidipiscis KCTC 13900	4.7	5.2	5.5
L. amylolyticus DSM 11664	-	4.9	4.3
L. animalis DSM 20602	-	-	4.8
L. buchneri CD034	-	5.4	5
L. buchneri NRRL B-30929	5.7	4.7	5.4
L. casei 12A	6.1	-	5.8
L. casei Lc-10	5.6	-	-
L. casei LOCK919	-	5.3	5.4
L. casei subsp. *casei* ATCC 393	4.3	4.8	-
L. casei UW4	-	4.8	5.5
L. ceti DSM 22408	5.2	5.1	5
L. coleohominis 101-4-CHN	4.5	4.9	5.2
L. composti JCM 14202	4.7	4.8	4.9
L. crispatus EM-LC1	-	-	5.1
L. delbrueckii subsp. *bulgaricus* 2038	5.7	5.7	5.4
L. delbrueckii subsp. *bulgaricus* ATCC BAA-365	-	5.4	-
L. delbrueckii subsp. *bulgaricus* CNCM I-1519	-	6.1	6
L. delbrueckii subsp. *bulgaricus* CNCM I-1632	4.8	4.9	-
L. delbrueckii subsp. *lactis* CRL581	5.4	-	-
L. equi DPC 6820	5.8	4.4	4.9
L. equicursoris CIP 110162	-	4.9	-
L. farciminis DSM 20184	4.7	4.8	4.2
L. farraginis JCM 14108	5.7	5.4	4.7
L. fermentum 28-3-CHN	5.4	-	-
L. fermentum 3872	5.8	5.3	-
L. fermentum MTCC 8711	5.4	-	-
L. fermentum NB-22	6	-	-
L. fructivorans DSM 20203	5.5	5.3	5.2
L. gasseri K7	5	4.8	4.5
L. gigeriorum CRBIP 24.85	-	4.5	4.6
L. harbinensis DSM 16991	5.4	5.5	5.4
L. helveticus CIRM-BIA 953	-	4.6	-
L. helveticus MTCC 5463	-	5.9	-
L. hominis CRBIP 24.179	5.5	5	5.4
L. iners LEAF 2053A-b	5.4	-	-
L. iners SPIN 2503V10-D	5.4	-	-
L. jensenii 208-1	5.1	-	-
L. johnsonii N6.2	-	4.6	-
L. kisonensis F0435	5	3.9	-
L. mali DSM 20444	5.7	5.1	5.5
L. mucosae LM1	5.9	4.8	4
L. murinus ASF361	7.4	6.9	7.1
L. namurensis str. Chizuka 01	4.7	4.2	4.6

Strain	0 d	670 d	700 d
L. oris PB013-T2-3	5.8	6.2	6.3
L. otakiensis JCM 15040	4.9	-	-
L. parabrevis ATCC 53295	5.4	5.2	5.3
L. paracasei subsp. *paracasei* 8700:2	5.6	-	-
L. paracasei subsp. *paracasei* Lpp120	5.6	-	-
L. paracasei subsp. *paracasei* Lpp122	-	-	5.1
L. paracasei subsp. *paracasei* Lpp126	-	-	5.8
L. paracasei subsp. *paracasei* Lpp219	-	5.2	-
L. paracasei subsp. *paracasei* Lpp229	-	-	5.4
L. paracasei subsp. *paracasei* Lpp41	5.6	-	-
L. paracasei subsp. *paracasei* Lpp43	-	-	5.4
L. parafarraginis F0439	5.2	5.6	5.6
L. pasteurii CRBIP 24.76	5.6	-	-
L. pentosus KCA1	5.7	5	4.8
L. plantarum 4_3	7.2	6.6	6.4
L. plantarum AY01	5.1	-	5
L. plantarum EGD-AQ4	5.2	-	-
L. plantarum subsp. *plantarum* NC8	5.6	-	-
L. pobuzihii E100301	5	-	-
L. psittaci DSM 15354	5.2	-	5.3
L. reuteri I5007	5.4	-	-
L. reuteri lpuph	-	4.5	4.6
L. reuteri mlc3	5.6	-	-
L. rhamnosus 2166	6	-	5.9
L. rhamnosus GG	-	-	5.5
L. rhamnosus HN001	5.4	-	-
L. rhamnosus Lc 705	5.6	-	5.3
L. rhamnosus LMS2-1	6.6	5.7	6.2
L. rhamnosus LOCK908	6	6.3	5.7
L. rossiae DSM 15814	5.2	4.4	5.4
L. ruminis ATCC 25644	7.4	7.2	6.9
L. ruminis ATCC 27782	6.8	7.9	6.4
L. ruminis SPM0211	6.9	6.2	6.3
L. salivarius ATCC 11741	-	4.3	5.3
L. sanfranciscensis TMW 1.1304	8.6	6.1	6.5
L. shenzhenensis LY-73	5.5	5.2	5.6
L. suebicus KCTC 3549	5.4	4.6	5.3
L. ultunensis DSM 16047	4.5	4.7	-
L. vaginalis ATCC 49540	6.1	5.9	5.9
L. versmoldensis KCTC 3814	5.6	5.8	5.1
L. vini JP7.8.9	5.9	-	-
L. zeae KCTC 3804	4.9	5.4	5.1
Lactobacillus sp. ASF360	-	4.3	4.9

Fig. 2. Heatmap of *Lactobacillus* strains in the feces of a healthy subject, at 0, 670 and 700 d. Values are Log (cells g^{-1}). Colours range from the lowest (deepest green) to the highest (deepest red) abundance. - indicates values falling below the detection limit of 3.9 Log (cells g^{-1}).

over the three samples. *L. ruminis*, *L. delbrueckii*, *L. sanfranciscensis* and *L. acidophilus*, identified at all the time-points, presented wide quantitative differences among the samples and were particularly abundant at a sole time-point where peaked up to 8 Log (cells g^{-1}). The vast majority of the recurrent species (27) showed a mean concentration laying in the magnitude of 5 Log (cells g^{-1}), and generally accounted for $<$1% of *Lactobacillus* reads. Seven species (*L. rossiae*, *L. mucosae*, *L. coleohominis*, *L. pentosus*, *L. composti*, *L. farciminis*, *L. namurensis*) were present at concentration in the magnitude of 4 Log (cells g^{-1}), accounting for $<$0.1%. For the 43 species occurring at all the time-points, no correlation was observed between the mean abundance and the variation range over the three samples (Supporting Information Fig. S1A).

Most of the persistent species (33 out of 43), including *L. delbrueckii*, *L. murinus*, *L. rhamnosus* and *L. ruminis*, and also the majority of the less abundant species with mLog (cells g^{-1}) $<$6, exhibited a relevant quantitative stability, with 1 or less magnitudes of difference among the time-points, regardless of their abundance. Only for *L. helveticus*, *L. sanfranciscensis* and *L. acidophilus* ΔLog (cells g^{-1}) was higher than 2, with the latter showing the widest range (3.6 magnitudes) of concentrations over the three time point. Concentration variation over short and long periods was also considered (Supporting Information Fig. S1B), the former being the variation between 670 and 700 d, and the latter as the variation between 0 d and the mean of the two last time points. The vast majority of the species differed by less than 1 magnitude order over both the short and long periods. Only *L. acidophilus* presented short-term and long-term variation both higher than 2 ΔLog (cells g^{-1}), whereas *L. fermentum*, *L. helveticus*, *L. mucosae* and *L. sanfranciscensis* showed high long-term variation, but remained stable over the short period.

Lactobacillus *strains*

WGS data allowed a deeper investigation of bacteria identified, making possible to trace 86 *Lactobacillus* spp. strains in one or more samples. The lactobacilli identified at strain level were the 58.30%, 48.67% and 21.86% of total *Lactobacillus* OTUs, at the three time-points. They belonged to 52 species, the following ones presenting more than one biotype: *L. paracasei* (8 biotypes), *L. rhamnosus* (5), *L. casei* and *L. delbrueckii* (5), *L. fermentum* and *L. plantarum* (4), *L. ruminis* (3), *L. buchneri*, *L. helveticus* and *L. iners* (2). A positive relationship was generally observed between the number of biotypes and the abundance of each species (data not shown). On the other hand, 8 biotypes of *L. paracasei* accounted, on average, for 0.41% of total lactobacilli, whereas *L. sanfranciscensis* and *L. murinus* were represented by a sole biotype, although they results among the most abundant *Lactobacillus* species.

Among the 86 *Lactobacillus* biotypes, 43 strains seemingly occupied the GIT as true residents: 34 occurred in all the three samples, and further 9 were identified both at 0 d and at one of the remainder samples, 670 or 700 d (Fig. 2). Taking into account the recurrent presence of the same strains over more than 22 months, albeit the availability of a small set of data (only 3 points), these 43 strains could be considered as permanent colonizers of the GIT. It should not be excluded that quantitative fluctuations, responsible of concentrations not emerging above the limit of detection of 3.9 Log (cells g^{-1}), may have hindered the identification of low abundance colonizers, such as 9 strains detected at 670 and 700 d, at magnitudes of 4 or 5 Log (cells g^{-1}), but not found at 0 d. Some of the permanent colonizers belong to species generally associated to human gut microbiota (e.g. *L. casei*, *L. crispatus*, *L. delbrueckii*, *L. fermentum*, *L. gasseri*, *L. johnsonii*, *L. paracasei*, *L. plantarum*, *L. reuteri*, *L. rhamnosus*, *L. ruminis*, *L. salivarius* and *L. vaginalis*), while many others belong to species that have never been previously associated to humans (*L. composti, L. farciminis, L. farraginis, L. harbinensis, L. namurensis, L. parabrevis, L. sanfranciscensis, L. shenzhenensis, L. suebicus, L. versmoldensis, L. zeae*), thus modifying the vision of *Lactobacillus* ecology. It is impossible to determine whether these bacteria inhabit the colon and fill the functional niches of this ecosystem, or are members of a bacterial population residing in another part of the GIT and shedding to the hindgut. However, it seems implausible that these strains are transient bacteria, occasionally ingested twice over a period of 2 years.

Most of the persistent strains (30 out of 34) showed a relevant quantitative stability, with 1 or less magnitudes of difference among the time-points, regardless of their mean concentration, and over both the short and the long period (Supporting Information Fig. S1C and D). The exceptions were *L. equi* DPC 6820, *L. mucosae* LM1, and *L. sanfranciscensis* TMW 1.1304, which presented quantitative fluctuations >1 Log (cells g^{-1}) over the long time span, and *L. ruminis* ATCC 27782 which presented quantitative fluctuations >1 Log (cells g^{-1}) over both the long and the short period.

Conclusions

Up to now, human GIT lactobacilli were perceived not only as a marginal population, but also as one of the major transient components of microbiota, originating from exogenous sources (Walter, 2008). Despite the limit of a longitudinal approach on a single subject, the results herein presented indicate the presence of a stable community of lactobacilli, with wide and understudied biodiversity. The low concentration of most of the species suggests a sub-dominant role in the colonic ecosystem. With this new insight, novel questions arise. A major challenge is determining the specific GIT district where this plethora of *Lactobacillus* species replicates and grows, in order to discriminate if they are indigenous resident of the colon, or whether they reach it shedding from upstream sites that they colonize. This can make the difference in terms of microbe-immune system relationship, since in the latter case lactobacilli can actively interact with GALT (gut-associated lymphoid tissue) exerting relevant immunomodulatory properties. Moreover, a recent study on the interaction between gut microbiome and virome highlighted the role of phages and prophages in modulating the bacterial structure and function of the bacterial community, with lytic lifestyles being effective in determining the dynamics of sub-dominant bacteria (Minot *et al.*, 2013; Ogilvie and Jones, 2015). Bacteriophages infecting lactobacilli are numerous, but the knowledge of their biology is still limited to the industrially relevant ones, and their role in shaping the community of intestinal *Lactobacillus* population is not known so far (Mahony and van Sinderen, 2014). As a whole, lactobacilli resulted a stable, relatively abundant, and very biodiverse community within the gut microbiota, but the current status of knowledge on colonic lactobacilli remains a major challenge.

Acknowledgements

The study was funded by Project SAF2012-31187 from MINECO (Ministerio de Economía y Competitividad, Madrid, Spain), PROMETEOII/2014/065 from Generalitat Valenciana (Valencia, Spain), and SIME2013.0627 (Fondazione Cassa di Risparmio di Modena, Italy).

References

Ames, S.K., Hysom, D.A., Gardner, S.N., Lloyd, G.S., Gokhale, M.B., and Allen, J.E. (2013) Scalable metagenomic taxonomy classification using a reference genome database. *Bioinformatics* **29**: 2253–2260.

Ahrné, S., Lönnermark, E., Wold, A.E., Aberg, N., Hesselmar, B., Saalman, R., *et al*. (2005) *Lactobacilli* in the intestinal microbiota of Swedish infants. *Microbes Infect* **7**: 1256–1262.

Chen, Y.S., Miyashita, M., Suzuki, K., Sato, H., Hsu, J.S., and Yanagida F. (2010) *Lactobacillus pobuzihii* sp. nov., isolated from pobuzihi (fermented cummingcordia). *Int J Syst Evol Microbiol* **60**: 1914–1917.

Cousin, S., Gulat-Okalla, M.L., Motreff, L., Gouyette, C., Bouchier, C., Clermont, D., and Bizet, C. (2012) *Lactobacillus gigeriorum* sp. nov., isolated from chicken crop. *Int J Syst Evol Microbiol* **62**: 330–334.

Cousin, S., Motreff, L., Gulat-Okalla, M.L., Gouyette, C., Spröer, C., Schumann, P., *et al*. (2013) *Lactobacillus pasteurii* sp. nov. and *Lactobacillus hominis* sp. nov. *Int J Syst Evol Microbiol* **63**: 53–59.

Dal Bello, F., Walter, J., Hammes, W.P., and Hertel, C. (2003) Increased complexity of the species composition of lactic acid bacteria in human feces revealed by alternative incubation condition. *Microb Ecol* **45**: 455–463.

Delgado, S., Suárez, A., and Mayo, B. (2007) Dominant cultivable *Lactobacillus* species from the feces of healthy adults in northern Spain. *Int Microbiol* **10**: 141–145.

Douillard, F.P., and de Vos, W.M. (2014) Functional genomics of lactic acid bacteria: from food to health. *Microb Cell Fact* **13**: S8.

Endo, A., and Okada, S. (2007a) *Lactobacillus composti* sp. nov., a lactic acid bacterium isolated from a compost of distilled shochu residue. *Int J Syst Evol Microbiol* **57**: 870–872.

Endo, A., and Okada, S. (2007b) *Lactobacillus farraginis* sp. nov. and *Lactobacillus parafarraginis* sp. nov., heterofermentative lactobacilli isolated from a compost of distilled shochu residue. *Int J Syst Evol Microbiol* **57**: 708–712.

Endo, A., Futagawa-Endo, Y., Sakamoto, M., Kitahara, M., and Dicks, L.M. (2010) *Lactobacillus florum* sp. nov., a fructophilic species isolated from flowers. *Int J Syst Evol Microbiol* **60**: 2478–2482.

Frank, D.N., and Pace N.R. (2008) Gastrointestinal microbiology enters the metagenomics era. *Curr Opin Gastroenterol* **24**: 4–10.

Hammes, W.P., and Hertel, C. (2006) The genera *Lactobacillus* and *Carnobacterium*. In The *Prokaryotes*, 3rd edn, Vol. **4**. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., and Stackenbrandt, E. (eds). New York: Springer, pp. 319–402.

Hartemink, R., and Rombouts, F.M. (1999) Comparison of media for the detection of bifidobacteria, lactobacilli and total anaerobes from faecal samples. *J Microbiol Methods* **36**: 181–192.

Heilig, H.G., Zoetendal, E.G., Vaughan, E.E., Marteau, P., Akkermans, A.D., and de Vos, W.M. (2002) Molecular diversity of *Lactobacillus* spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA. *Appl Environ Microbiol* **68**: 114–123.

Kim, D.W., Choi, S.H., Kang, A., Nam, S.H., Kim, D.S., Kim, R.N., *et al*. (2011) Draft genome sequence of *Lactobacillus zeae* KCTC 3804. *J Bacteriol* **193**: 5023.

Kitahara, M., Sakata, S., and Benno, Y. (2005) Biodiversity of *Lactobacillus sanfranciscensis* strains isolated from five sourdoughs. *Lett Appl Microbiol* **40**: 353–357.

Kröckel, L., Schillinger, U., Franz, C.M., Bantleon, A., and Ludwig, W. (2003) *Lactobacillus versmoldensis* sp. nov., isolated from raw fermented sausage. *Int J Syst Evol Microbiol* **53**: 513–517.

Krzyściak, W., Pluskwa, K.K., Jurczak, A., and Kościelniak, D. (2013) The pathogenicity of the *Streptococcus* genus. *Eur J Clin Microbiol Infect Dis* **32**: 1361–1376.

Lawson, P.A., Wacher, C., Hansson, I., Falsen, E., and Collins, M.D. (2001) *Lactobacillus psittaci* sp. nov., isolated from a hyacinth macaw (*Anodorhynchus hyacinthinus*). *Int J Syst Evol Microbiol* **51**: 967–970.

Leroy, F., and De Vuyst, L. (2004) Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends Food Sci Technol* **15**: 67–78.

Lönnermark, E., Nowrouzinan, F., Adlerberth, I., Ahrné, S., Wold, A., and Friman, V. (2012) Oral and faecal lactobacilli and their expression of mannose-specific adhesins in individuals with and without IgA deficiency. *Int J Med Microbiol* **302**: 53–60.

Mahony, J., and van Sinderen, D. (2014) Current taxonomy of phages infecting lactic acid bacteria. *Front Microbiol* **5**: 7.

Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013) Rapid evolution of the human gut virome. *Proc Natl Acad Sci USA* **110**: 12450–12455.

Miyamoto, M., Seto, Y., Hao, D.H., Teshima, T., Sun, Y.B., Kabuki, T., *et al*. (2005) *Lactobacillus harbinensis* sp. nov., consisted of strains isolated from traditional fermented vegetables 'Suan cai' in Harbin, Northeastern China and *Lactobacillus perolens* DSM 12745. *Syst Appl Microbiol* **28**: 688–694.

Morita, H., Shimazu, M., Shiono, H., Toh, H., Nakajima, F., Akita, H., *et al*. (2010) *Lactobacillus equicursoris* sp. nov., isolated from the faeces of a thoroughbred racehorse. *Int J Syst Evol Microbiol* **60**: 109–112.

Morotomi, M., Yuki, N., Kado, Y., Kushiro, A., Shimazaki, T., Watanabe, K., and Yuyama T. (2002) *Lactobacillus equi* sp. nov., a predominant intestinal *Lactobacillus* species of the horse isolated from faeces of healthy horses. *Int J Syst Evol Microbiol* **52**: 211–214.

Nam, S.H., Choi, S.H., Kang, A., Kim, D.W., Kim, R.N., Kim, A., *et al*. (2011a) Genome sequence of *Lactobacillus farciminis* KCTC 3681. *J Bacteriol* **193**: 1790–1791.

Nam, S.H., Choi, S.H., Kang, A., Kim, D.W., Kim, R.N., Kim, D.S., *et al*. (2011b) Genome sequence of *Lactobacillus suebicus* KCTC 3549. *J Bacteriol* **193**: 5532–5533.

O'Donnell, M.M., O'Toole, P.W., and Ross, R.P. (2013) Catabolic flexibility of mammalian-associated lactobacilli. *Microb Cell Fact* **12**: 48.

O'Hara, A.M., and Shanahan, F. (2006) The gut flora as a forgotten organ. *EMBO Rep* **7**: 688–693.

Ogilvie, L.A., and Jones, B.V. (2015) The human gut virome: a multifaceted majority. *Front Microbiol* **6**: 918.

Öztürk, M., and Meterelliyöz, M. (2015) Practical identification of human originated *Lactobacillus* species by amplified ribosomal DNA restriction analysis (ARDRA) for probiotic use. *Mol Biol Rep* **42**: 1323–1332.

Parte, A.C. (2014) LPSN-list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res* **42**: D613–D616.

Quartieri, A., Simone, M., Gozzoli, C., Popovic, M., D'Auria, G., Amaretti, A., *et al*. (2016) Comparison of culture-dependent and independent approaches to characterize fecal bifidobacteria and lactobacilli. *Anaerobe*, **38**: 130–137.

Rajilić-Stojanović, M., and de Vos, W.M. (2014) The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol Rev* **38**: 996–1047.

Reuter, G. (2001) The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession. *Curr Issues Intest Microbiol* **2**: 43–53.

Rodas, A.M., Chenoll, E., Macián, M.C., Ferrer, S., Pardo, I., and Aznar, R. (2006) *Lactobacillus vini* sp. nov., a wine lactic acid bacterium homofermentative for pentoses. *Int J Syst Evol Microbiol* **56**: 513–517.

Scheirlinck, I., Van der Meulen, R., Van Schoor, A., Cleenwerck, I., Huys, G., Vandamme, P., *et al*. (2007) *Lactobacillus namurensis* sp. nov., isolated from a traditional Belgian sourdough. *Int J Syst Evol Microbiol* **57**: 223–227.

Sekirov, I., Russell, S.L., Antunes, L.C., and Finlay, B.B. (2010) Gut microbiota in health and disease. *Physiol Rev* **90**: 859–904.

Vancanneyt, M., Naser, S.M., Engelbeen, K., De Wachter, M., Van der Meulen, R., Cleenwerck, I., *et al*. (2006) Reclassification of *Lactobacillus brevis* strains LMG 11494 and LMG 11984 as *Lactobacillus parabrevis* sp. nov. *Int J Syst Evol Microbiol* **56**: 1553–1557.

Van de Casteele, S., Vanheuverzwijn, T., Ruyssen, T., Van Assche, P., Swings, J., and Huys, G. (2006) Evaluation of culture media for selective enumeration of probiotic strains of lactobacilli and bifidobacteria in combination with yoghurt or cheese starters. *Int Dairy J* **16**: 1470–1476.

Vela, A.I., Fernandez, A., Espinosa de los Monteros, A., Goyache, J., Herraez, P., Tames, B., *et al*. (2008) *Lactobacillus ceti* sp. nov., isolated from beaked whales (*Ziphius cavirostris*). *Int J Syst Evol Microbiol* **58**: 891–894.

Wall, R., Fitzgerald, G., Hussey, S., Ryan, T., Murphy, B., Ross, P., and Stanton, C. (2007) Genomic diversity of cultivable *Lactobacillus* populations residing in the neonatal and adult gastrointestinal tract. *FEMS Microbiol Ecol* **59**: 127–137.

Walter, J. (2008) Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. *Appl Environ Microbiol* **74**: 4985–9496.

Watanabe, K., Fujimoto, J., Tomii, Y., Sasamoto, M., Makino, H., Kudo, Y., and Okada, S. (2009) *Lactobacillus kisonensis* sp. nov., *Lactobacillus otakiensis* sp. nov., *Lactobacillus rapi* sp. nov. and *Lactobacillus sunkii* sp. nov., heterofermentative species isolated from sunki, a traditional Japanese pickle. *Int J Syst Evol Microbiol* **59**: 754–760.

Xu, J., and Gordon, J.I. (2003) Honor thy symbionts. *Proc Natl Acad Sci USA* **100**: 10452–10459.

Zou, Y., Liu, F., Fang, C., Wan, D., Yang, R., Su, Q., *et al*. (2013) *Lactobacillus shenzhenensis* sp. nov., isolated from a fermented dairy beverage. *Int J Syst Evol Microbiol* **63**: 1817–1823.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Panel A and C: variation of the amounts of *Lactobacillus* species (A) and strains (C) within the three time-points, expressed as ΔLog (cells g^{-1}), plotted against their mean abundance, expressed as mLog (cells g^{-1}). Panel B and D: long term variation of the amounts of *Lactobacillus* species (B) and strains (D), plotted against their respective short term variation. Short term variation is calculated as the variation between Log (cells g^{-1}) at 670 and 700 d, long term variation as the variation between Log (cells g^{-1}) at 0 d and the mean of the two last time points.
Experimental procedure.