

Daily follow-up of bacterial communities in the human gut reveals stable composition and host-specific patterns of interaction

Ana Durbán^{1,2}, Juan José Abellán^{1,2}, Nuria Jiménez-Hernández^{1,2}, Amparo Latorre^{1,2} & Andrés Moya^{1,2}

¹Joint Research Unit on Genomics and Health, CSISP, University of Valencia/Cavanilles Institute of Biodiversity and Evolutionary Biology, Valencia, Spain; and ²CIBER Epidemiología y Salud Pública (CIBERESP), Spain

Correspondence: Juan José Abellán, Center for Public Health Research (CSISP) Avinguda Catalunya 21, Valencia 46020, Spain.
Tel.: +34 961 925 917;
fax: +34 961 925 978; e-mail:
abellan_jua@gva.es

Received 15 August 2011; revised 2 March 2012; accepted 12 March 2012.
Final version published online 27 April 2012.

DOI: 10.1111/j.1574-6941.2012.01368.x

Editor: Julian Marchesi

Keywords

faecal bacteria; 16S rRNA; correlation network; within-individual variation.

Abstract

In the last decade, an extensive effort has been made to characterize the human intestinal microbiota by means of SSU rRNA gene sequence and metagenomic analysis. Relatively few studies have followed intestinal bacterial communities over time to assess their stability in the absence of perturbation. In this study, we have monitored the faecal bacteria of three healthy subjects during 15 consecutive days. The global community structure was analysed through SSU rRNA gene sequencing. In agreement with previous studies, we found that the between-subject variation in community structure was larger than within. The composition was fairly stable throughout, although daily fluctuations were detected for all genera and phylotypes at 97% of sequence identity. While the core shared between subjects was very small, each subject harboured a stable high-abundance core composed of a small number of bacterial groups (9% of the phylotypes accounted for between 74% and 93% of the sequences). This may suggest that studies aimed at linking the microbiota composition with disease risk should be limited to the numerically most dominant phylotypes, as the rest appears transient. Networks of potential interactions between co-occurring genera were also subject-specific, even for the same bacterial genus, which might be reflecting host-specific selective pressures and historic events.

Introduction

The human large intestine is colonized by complex microbial communities with several hundreds of species and the highest cell densities of all human-associated microbiota. Functions of the gut microbiota include assistance in nutrient digestion, colonization resistance to pathogens and regulation of host metabolism and immune system (Guarner & Malagelada, 2003; Turnbaugh *et al.*, 2006; Kelly *et al.*, 2007; Neish, 2009; Qin *et al.*, 2010). The gut has been traditionally considered sterile at birth and rapidly colonized by bacteria of maternal origin, but recent studies have shown that meconium, the earliest stools of a newborn, composed of materials ingested during the stay in the uterus, is nonsterile (Jiménez *et al.*, 2008). Over the first years of life, an ecological succession takes place that culminates in the complex

adult pattern (Favier *et al.*, 2002; Palmer *et al.*, 2007). Then, it is generally accepted that, in the absence of perturbation, the adult human gut microbiota is composed of stable climax communities inhabiting the different niches found along and across the intestine.

Many factors are known to influence the structure of the microbiota of adult humans, such as host genotype (Ley *et al.*, 2005; Khachatryan *et al.*, 2008; Benson *et al.*, 2010), disease (Turnbaugh *et al.*, 2006; Qin *et al.*, 2010), diet (Harmsen *et al.*, 2000; De Filippo *et al.*, 2010) and stochastic events such as colonization history (Mulder *et al.*, 2009) and external disturbances (De La Cochetiere *et al.*, 2005; Dethlefsen & Relman, 2010; van Vliet *et al.*, 2009). Modifications of the gut microbiota have also been reported in older people related to physiological changes associated with ageing (Woodmansey, 2007; Claesson *et al.*, 2010). However, little is known about the short- and

long-term effects that these factors may have in the composition of the microbiota. To assess the response of the gut microbiota to external perturbations or its role in pathologic states, it is important to know first the 'normal' temporal dynamics of this ecosystem.

Human faecal communities have been considered stable because the temporal variation found within individuals is smaller than between individuals. The stability has typically been examined with several samples collected at intervals of weeks or months, finding that the microbiota of healthy individuals remains fairly constant over these long time intervals (Franks *et al.*, 1998; Zoetendal *et al.*, 1998; Vanhoutte *et al.*, 2004). However, some variation has been found in the abundance of certain taxa, while others remain more constant (Vanhoutte *et al.*, 2004). Recently, studies of the temporal variation using shorter time periods confirmed the stability of the community composition at a lower phylogenetic level but also revealed a high degree of dynamics, with relatively large fluctuations around the average level (Dethlefsen *et al.*, 2008; Dethlefsen & Relman, 2010; Caporaso *et al.*, 2011).

Ecological interactions between the members of gut communities, like nutritional associations, niche adaptation, growth stimulation, resource competition, and interference mechanisms, also contribute to the shape and stability of these ecosystems. Despite its importance, little is known about the relationships between gut bacteria. A first step for the study of these interactions is searching for patterns of co-occurring bacteria. This can be carried out with cross-sectional studies (see e.g. Arumugam *et al.*, 2011) or with longitudinal ones as well (this study). The latter are helpful to control for the high level of interpersonal variation in the community assembly as the same ecosystem is followed over time.

We are interested in the short-term intrinsic temporal dynamics of human intestinal microbiota, as well as in the potential interactions between the community members. The objectives of the present study were to analyse the daily variation in the community structure of faecal bacteria from three healthy subjects over 2 weeks and to assess correlations between shifts in the relative abundance of specific bacteria to reconstruct potential interaction networks.

Materials and methods

Subjects' characteristics

Three White male volunteers (hereinafter referred to as individuals A, B and C) provided faecal samples daily over a period of fifteen consecutive days. All participants gave prior informed written consent to the study protocol, which was approved by the Ethics Committee of

Hospital Universitario La Fe (Valencia, Spain). A, B and C were 40, 39 and 29 years old, respectively. A was normal weight, B and C were obese class I according to the body mass index (BMI; $30 \leq \text{BMI} < 35$). None had a history of gastrointestinal disease or systemic comorbidities, recent (in the last 3 months) treatment with antibiotics, immunomodulating therapy, antidiarrhoeal medication or laxatives. They followed a Mediterranean diet that remained unchanged throughout the follow-up.

Sample collection and DNA extraction

Faecal samples were self-collected by the volunteers in the morning in tubes containing phosphate-buffered saline (PBS) and stored in their home freezers until its release. Afterwards, samples were stored at $-80\text{ }^{\circ}\text{C}$ until further processing. Before DNA extraction, samples were resuspended in PBS and centrifuged at low speed to remove faecal debris as far as possible. Then, DNA was extracted from supernatants using the QIAamp DNA Stool Mini Kit (QIAGEN) and its protocol for isolation of DNA for pathogen detection.

PCR amplification and sequencing of bacterial 16S rRNA genes

The 16S rRNA genes were amplified using the broad range bacterial primers B8F (5'-AGAGTTTGATCM-TGGCTCAG-3') and B357R (5'-TGCTGCCTCCCGTAG-GAGT-3') modified with the 454 Life Sciences adaptors A and B, respectively, and with barcodes to tag each PCR product. The PCR conditions were 5 min of initial denaturation at $95\text{ }^{\circ}\text{C}$ followed by 20 cycles of denaturation (30 s at $95\text{ }^{\circ}\text{C}$), annealing (30 s at $52\text{ }^{\circ}\text{C}$) and elongation (30 s at $72\text{ }^{\circ}\text{C}$). PCR products were purified by filtration, and equal amounts of the PCR products from different samples were pooled. The mixtures were sent for pyrosequencing with primer A on a 454 sequencer using the GS FLX chemistry for samples of subject A and the GS FLX Titanium chemistry for samples of subjects B and C (Roche). This was due to the change in the sequencing chemistry while the study was ongoing.

Sequence analysis

Sequences with low-quality scores (< 20) and short read lengths (< 200 nt for samples of subject A and < 250 nt for samples of subjects B and C) were removed. The remaining sequences were checked for potential chimeras using the *chimera.slayer* tool incorporated into the MOTHUR software (Schloss *et al.*, 2009). The taxonomic affiliation of the sequences was determined using the *classifier* tool of the Ribosomal Database Project-II (RDP)

with a bootstrap cut-off of 70% (Cole *et al.*, 2007, 2009). Sequences were clustered at 97% of sequence identity using the *cluster* tool of the USEARCH 5.0 package (Edgar, 2010). Input sequences to cluster were previously sorted by decreasing abundance as recommended for 16S sequences. The resulting phylotypes were used to study sample composition at the 'species' level.

Statistical analyses

Exploratory statistical analyses

We assessed the differences between and within subjects according to the bacterial composition using correspondence analysis (CA) and analysis of similarities (ANOSIM) at genus and phylotype levels (Quinn & Keough, 2002; pp. 459 and 484, respectively) as implemented in the VEGAN package (Oksanen *et al.*, 2010) for the statistical environment R (R Development Core Team, 2010). We also computed the Chao1 index of richness and the Shannon's index of biodiversity for each of the 45 samples at both genus and phylotype levels. This was performed with the QIIME pipeline (Caporaso *et al.*, 2010), resampling 1000 times from each daily sample using the smallest sample size of all samples.

Multivariate time series modelling

From a statistical point of view, the data on daily taxa abundances can be regarded as a multivariate time series. The potential interactions between taxa are expected to be reflected in the correlations between taxa, but also some temporal correlation is expected to be present in the data. In other words, if Y_{it} represents the matrix with the number of sequences of taxon $i = 1, \dots, n$ found in the sample collected on day $t = 1, \dots, T$ for a given individual, both rows and columns present correlation structures of different nature. Abundances in a given row are likely to be affected by temporal correlation, whereas values in a specific column may be subject to the correlations generated by the underlying interactions between taxa. To model both correlation structures simultaneously, we applied a Bayesian hierarchical model to the follow-up data for each individual.

Our model specification is as follows. Let $\mathbf{Y}_t = (Y_{1t}, \dots, Y_{nt})'$ be the taxonomic distribution of sequences on day t . Our model first assumes that \mathbf{Y}_t follows a multinomial distribution

$$\mathbf{Y}_t \sim \text{Multinomial}(\boldsymbol{\pi}_t, N_t)$$

where N_t is the total number of sequences on day t and $\boldsymbol{\pi}_t = (\pi_{1t}, \dots, \pi_{nt})'$, π_{it} being the unknown proportion in

which taxon i is present in the community on day t . The proportions π_{it} are in turn decomposed, on the log-odds scale, into

$$\log(\pi_{it}/(1 - \pi_{it})) = \alpha_i + v_{it} + \varepsilon_{it}$$

where α_i is a taxon-specific intercept that picks up the average relative abundance of taxon i over the $T = 15$ days, and v_{it} and ε_{it} are random effects intended to pickup time structured and unstructured variation, respectively. To this end, we chose a normal prior distribution for ε_{it} and a multivariate random walk of order one for \mathbf{v}_t , $t = 1, \dots, T$

$$\mathbf{v}_t | \mathbf{v}_{t-1} \sim \text{MVN}(\mathbf{v}_{t-1}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma}$ is the $n \times n$ variance-covariance matrix between taxa abundances. For convenience, we take $\mathbf{v}_0 = \mathbf{0}_{n \times 1}$. This conditional specification is a particular case of the intrinsic multivariate conditional autoregressive (MCAR) models (Kim *et al.*, 2001; Gelfand & Vounatsou, 2003), for which the full conditional distribution is

$$\mathbf{v}_t | \mathbf{v}_1, \dots, \mathbf{v}_{t-1}, \mathbf{v}_{t+1}, \dots, \mathbf{v}_T \sim \begin{cases} \text{MVN}(\mathbf{v}_{t+1}, \boldsymbol{\Sigma}) & t = 1 \\ \text{MVN}\left(\frac{\mathbf{v}_{t-1} + \mathbf{v}_{t+1}}{2}, \frac{\boldsymbol{\Sigma}}{2}\right) & t = 2, \dots, T - 1 \\ \text{MVN}(\mathbf{v}_{t-1}, \boldsymbol{\Sigma}) & t = T \end{cases}$$

that is, \mathbf{v}_t follows a multivariate normal distribution centred in the average of its temporal neighbours and variance-covariance matrix inversely proportional to the number of neighbours. The joint distribution of $\mathbf{v} = (v_{11}, \dots, v_{n1}, v_{12}, \dots, v_{n2}, \dots, v_{1T}, \dots, v_{nT})'$ is a zero-mean multivariate normal distribution with precision matrix $\boldsymbol{\Omega} = (\mathbf{D} - \mathbf{W}) \otimes \boldsymbol{\Sigma}^{-1}$, where \mathbf{W} is a $T \times T$ matrix with $W_{tt'} = 1$ if time points t and t' are adjacent and $W_{tt'} = 0$ otherwise, \mathbf{D} is a $T \times T$ diagonal matrix with D_{tt} equal to the number of neighbours of time point t (i.e. $D_{11} = D_{TT} = 1$ and $D_{tt} = 2 \forall t = 2, \dots, T - 1$) and \otimes represents the Kronecker product for matrices. The matrix $\mathbf{D} - \mathbf{W}$ is singular, which makes this distribution improper. However, with our choice of \mathbf{W} and \mathbf{D} , $\boldsymbol{\Omega}$ satisfies the so-called symmetry condition that ensures propriety of the posterior. In practice, this impropriety is overcome using the proper full conditionals for v_t and imposing n sum-to-zero constraints. See for example Banerjee *et al.* (2004, pp. 247–251) for further details.

We fitted our model using Markov chain Monte Carlo (MCMC) simulation techniques as implemented in the WINBUGS software (Lunn *et al.*, 2000) and the R2WINBUGS package (Sturtz *et al.*, 2005) for the R statistical

software (R Development Core Team, 2010). We ran two chains with 50 000 iterations, discarded the first 10 000 as burn-in and kept every 40th to reduce autocorrelation in the chains. Therefore, inference for each parameter is based on a thinned sample of size 2000 from its posterior distribution.

Exploring putative interactions between taxa with Graphical Gaussian Networks (GGN)

GGNs (Schäfer & Strimmer, 2005a,b) have been used to recover gene regulation network structures using gene expression data as input. They aim at predicting interaction networks between genes. Here, however, we apply for the first time GGNs to explore patterns of association between taxa using the partial correlations between their abundance profiles. A strong partial correlation between two species is indicative of some form of association. The estimation of that matrix is tricky because typically it is sparse and has large dimensionality. GGNs however allow to estimate efficiently the partial correlation matrix from the variance-covariance matrix. We therefore applied GGNs to the covariance matrix Σ (which measures covariances between taxa abundances on the log-odds scale) obtained with the above Bayesian model to infer a network of potential associations between taxa. This was performed using R package *CORPCOR* (Schäfer *et al.*, 2010). The statistical significance of the estimated partial correlations was performed using the algorithms proposed by Opgen-Rhein & Strimmer (2007). The idea is to model the partial correlations with a mixture of two components. The first component tries to capture the null partial correlations, whereas the second component intends to pick up the sizeable ones. Opgen-Rhein & Strimmer (2007) suggest a method to estimate the mixture components and also a 'local false discovery rate' (LFDR) procedure to assess the statistical significance of each partial correlation. They show that their methods perform very well both in simulations as well as in application to large-scale real expression data in the context of gene association networks. We applied these methods as implemented in the R package *GENE NET* (Schäfer *et al.*, 2009). These analyses were carried out for each individual separately.

The output is a graph, with nodes corresponding to taxa and edges representing statistically significant partial correlations between taxa (taking as such that the probability, $1 - \text{LFDR}$, for the partial correlation to be different from zero is above 0.95). Graphics were generated with the R package *RGRAPHVIZ* (Gentry *et al.*, 2008).

The nonredundant sequences from this study have been deposited in the GenBank database under the accession numbers JN547818–JN555583.

Results

After sequencing all 45 samples, we obtained an average of 1500 sequences per sample for A, 4165 for B and 6510 for C. The average read length was approximately 250 nt in A and approximately 350 nt in B and C. The difference in the read length is the result of the slightly different sequencing technologies employed for subject A and subjects B and C, but in both cases, the variable regions V1 and V2 of the 16S rRNA gene were covered. The comparison between subjects could be affected by the potential bias introduced during the PCR by the different sequencing adaptors or the impact of the read length in the taxonomic assignment. However, we think that the impact on the estimate of the variation between subjects would be small, and none on the analysis within each subject, which is our main objective.

The Chao1 index of richness indicated a good coverage for most samples at genus level (Table S1). The coverage values decreased at phylotype level, with larger differences between observed and expected number of OTUs (Table S2).

Between-subject variation

Most members of the sampled communities belonged to a small number of genera within the Bacteroidetes phylum (61% of the sequences on average in samples from A, approximately 86% in B and C) and the Firmicutes phylum (26% in A, approximately 10% in B and C). The composition of the faecal microbiota of B and C was quite similar from phylum to genus levels. At the genus level, samples from A were dominated by *Alistipes* (23%), *Bacteroides* (22%), several genera of Porphyromonadaceae such as *Barnesiella* (8%) and *Parabacteroides* (2%), and several genera of Ruminococcaceae such as *Faecalibacterium* (4%) and *Oscillibacter* (2%; Fig. 1a). Samples from B and C were dominated by *Prevotella* (76% in B and 72% in C), *Bacteroides* (5% in B and 12% in C) and beta-proteobacteria of the genus *Sutterella* (approximately 3%; Fig. 1b and c).

At a finer taxonomic scale (phylotypes defined at 97% of sequence identity), the faecal microbiota was highly specific to each individual. About half of the phylotypes detected in each subject were exclusive to him, whereas 20% were shared with the other two subjects. The greater similarity between B and C was also seen at this level, as they had more phylotypes in common than each of them with A. Most of the sequences were concentrated in a small fraction of the phylotypes. Thus, 11% of the phylotypes in A and 4% of the phylotypes in B and C had a relative abundance of at least 5%, and all together accounted for 86%, 80% and 89% of their respective

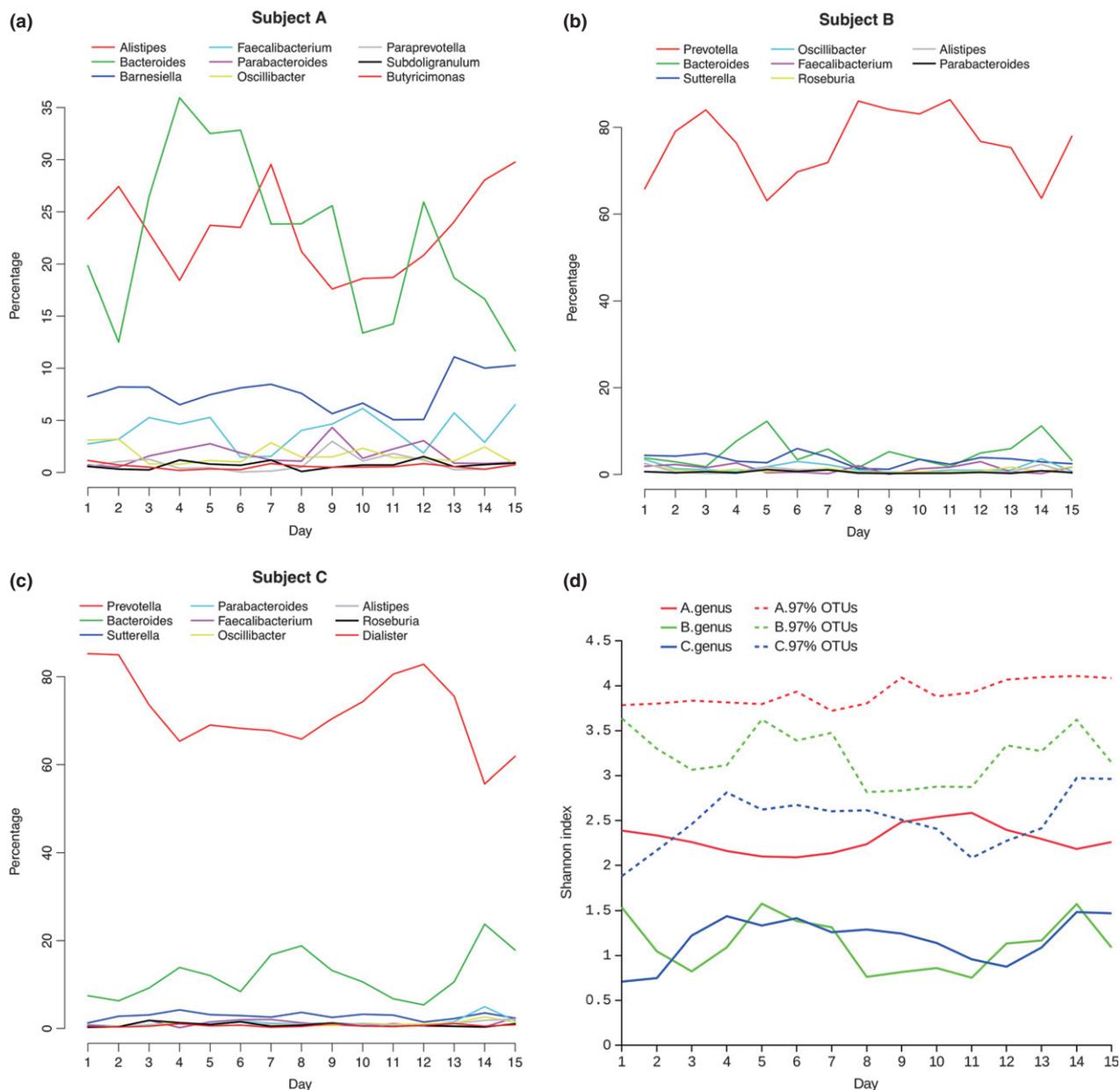


Fig. 1. Daily fluctuations in the relative abundance of the main bacterial genera (average abundance $\geq 5\%$) in subjects A (a), B (b) and C (c). Daily variation in the bacterial diversity (Shannon index) at genus and phylotype levels (d).

number of sequences. The most prevalent genera in A, *Alistipes*, *Bacteroides* and *Barnesiella*, had 9%, 10% and 3% of the total number of phylotypes, whereas only 16% of the phylotypes in B, and 10% in C, belonged to *Prevotella*.

Between-subject differences in the community structure were larger than dissimilarities between samples from the same subject. Therefore, CA plots clearly discriminated the samples of the three subjects at genus and phylotype levels (Fig. 2). The first CA axis separated A from B and

C, and the second separated B from C. The ANOSIM found significant differences between subjects (genus level: $R = 0.686$, $P = 0.001$; phylotype level: $R = 1$, $P = 0.001$). At genus level, the between-subject median rank of distances between samples was 1.5 times that of within A and 4 times that of within B and C. At phylotype level, between-subject variation was 2.6, 4.5 and 8.6 times that of within A, B and C, respectively, thus reflecting even larger differences in microbiota composition between individuals at this finer taxonomic resolution.

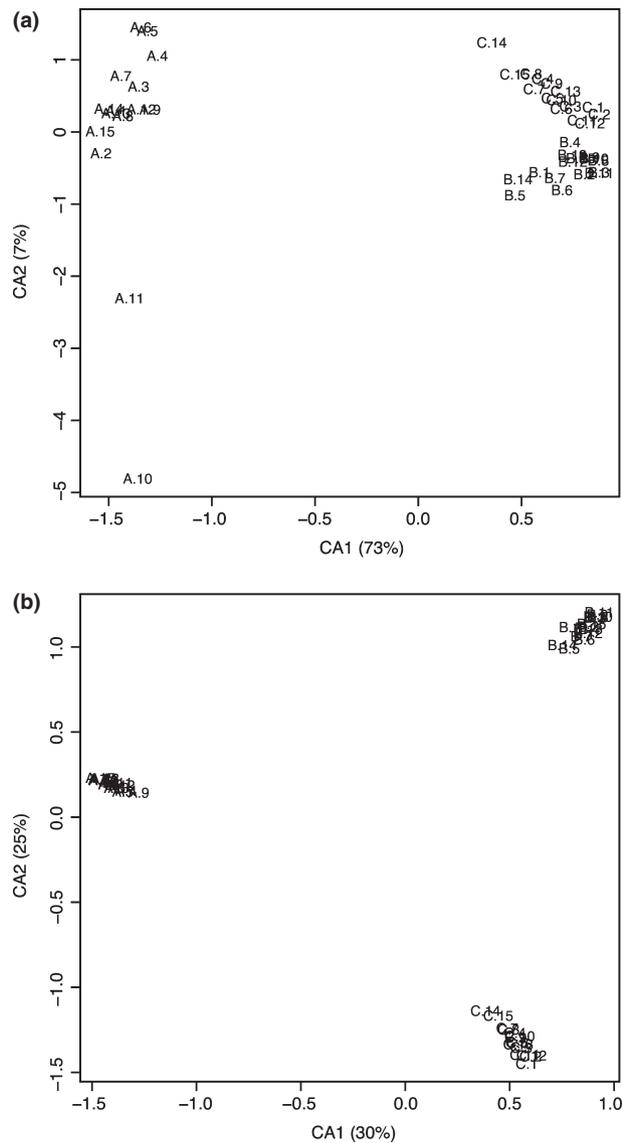


Fig. 2. Correspondence analysis at genus (a) and phylotype (b) levels. Percentages correspond to the fraction of inertia explained by each axis.

Within-subject variation

The communities experienced daily fluctuations in the bacterial diversity according to the Shannon index (Fig. 1d and Tables S1 and S2), although these values varied around a constant level. This is consistent with the fact that the structure of the sampled faecal communities remained quite stable over time. Relative abundances showed daily fluctuations for all genera but no temporal trends were observed (Fig. 1a–c). All genera with an abundance of at least 5‰ were permanent members of the communities, whereas genera with prevalence below that threshold were sometimes lost and recovered,

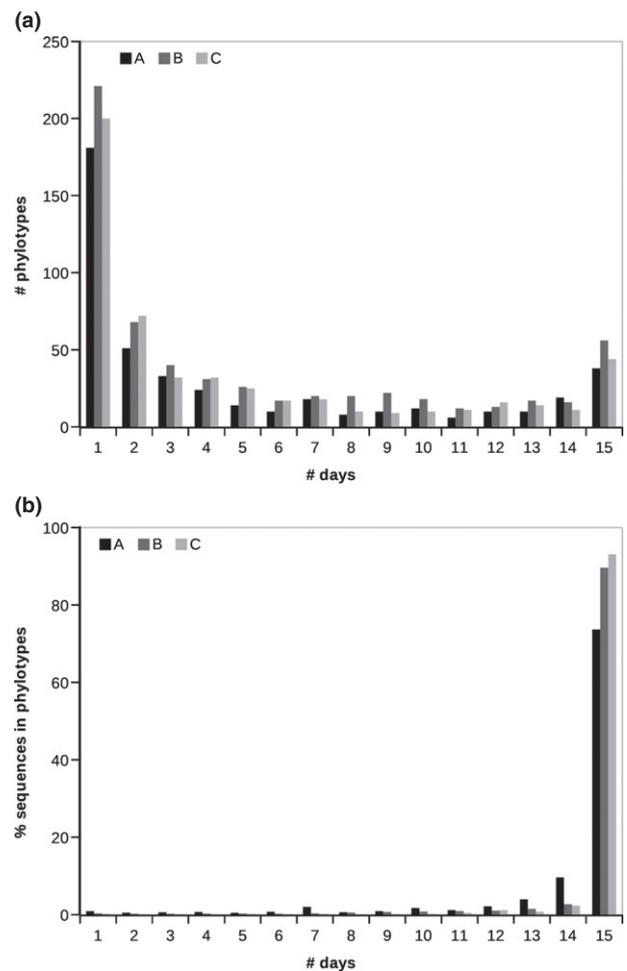


Fig. 3. Community structure at phylotype level in subjects A, B and C. Occurrence of phylotypes during the 15-day study period, that is, number of phylotypes detected on just 1 day, on 2 days, etc. (a). Combined average relative abundance of phylotypes detected on just 1 day, on 2 days, etc., computed over all samples where they occur (b).

although some were constantly present too. The taxa with the highest prevalence showed the lowest relative fluctuations in time. In A, *Alistipes*, *Bacteroides* and *Barnesiella* experienced an average daily relative change of 2.9%, 3.8% and 6.4% in their abundances, respectively, whereas in B and C, the average daily relative variation in *Prevotella* was of approximately 1.8%.

At a subgeneric level, a reduced number of phylotypes were constantly detected throughout within each subject (Fig. 3a, which shows the number of phylotypes detected on just 1 day, on 2 days, etc.). The individual-specific core (phylotypes detected all the days during the study period) comprised 38 phylotypes in A, 56 in B and 44 in C, which corresponded to approximately 9% of the total number of phylotypes within each subject. Conversely, most of the phylotypes detected within each subject were found in a few days only (Fig. 3a). The core phylotypes

belonged to the most prevalent genera and accounted for most of the sequences in the samples (74% of the total number of sequences in A, 90% in B and 93% in C; Fig. 3b, which shows the cumulative percentage of sequences belonging to phylotypes detected on just 1 day, on 2 days, etc.). Again, phylotypes showed fluctuations in their abundance over time, but the trend was constant. Only 0.2% of the phylotypes were detected in all subjects all days during the follow-up, while 2.6% of the phylotypes were simultaneously detected throughout in the more similar subjects B and C. With a less restrictive def-

inition of the bacterial core, we found that 0.5% of the phylotypes was detected in all three subjects in at least 13 out of 15 days.

Correlations between co-occurring genera

Using GGN, we built interaction networks from the statistically significant partial correlations between genera estimated with the Bayesian model. These networks are represented in a graph, where nodes correspond to genera and edges represent interactions (Fig. 4 shows the

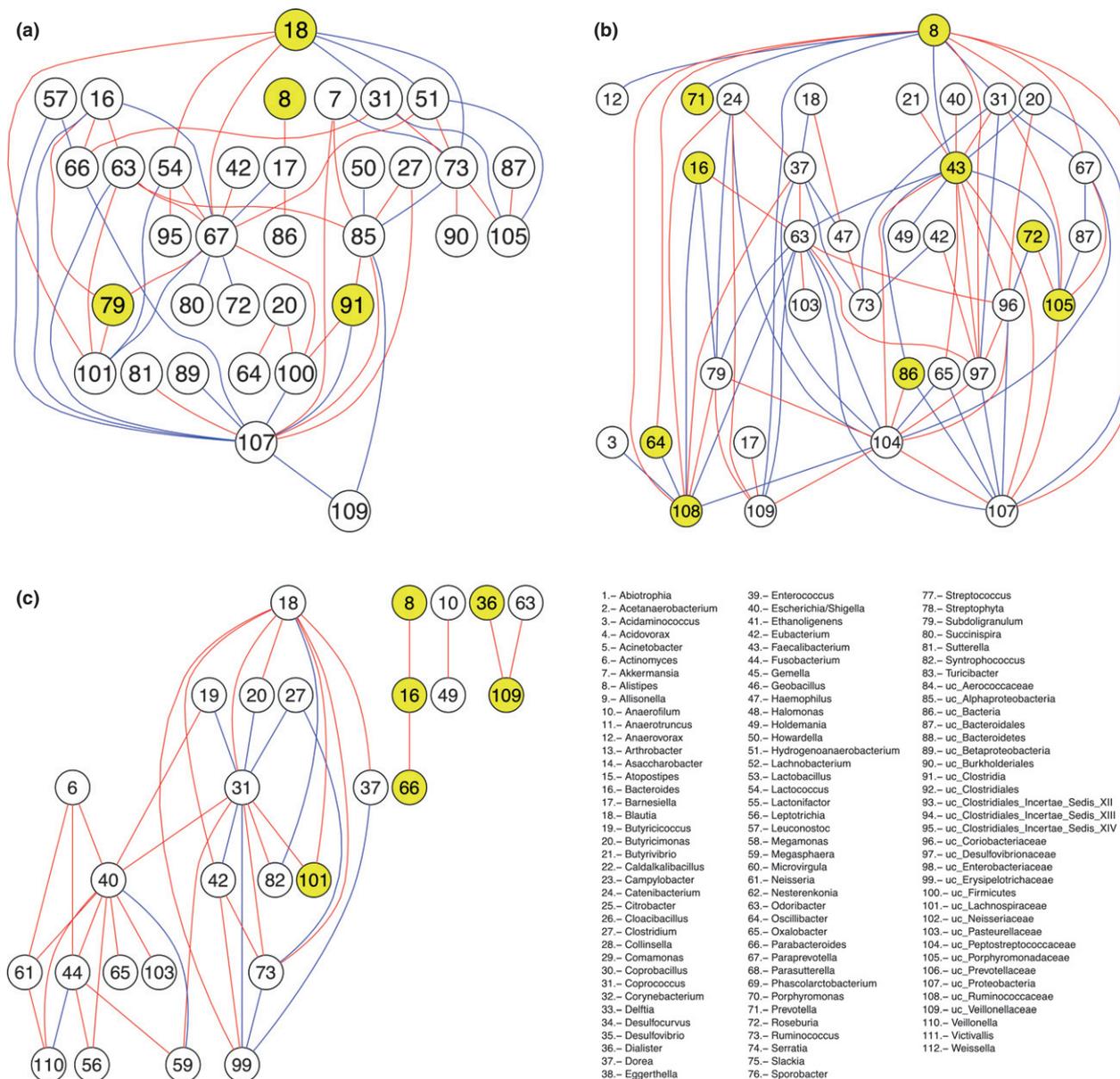


Fig. 4. Gaussian graphical networks representing the interactions between genera within subjects A (a), B (b) and C (c). Red and blue edges represent positive and negative partial correlations, respectively. Yellow nodes are those of genera with an average relative abundance $\geq 5\%$. Genera for which no significant partial correlations were found are not shown in the networks.

subgraphs of the networks including the statistically significant correlations). Most genera showed correlations with a small number of other genera, while a small number of genera were correlated with many. The main components of the faecal communities scarcely correlated with other genera. For instance, *Alistipes* correlated (positively) only with *Barnesiella* in A. *Prevotella* was negatively correlated with *Alistipes* in B, and with no genus in C. *Bacteroides*, the second genus in abundance, was correlated with five, three and two genera in A, B and C, respectively. The most connected genera were *Paraprevotella* in A, *Faecalibacterium*, *Alistipes* and *Odoribacter* in B, and *Coprococcus*, *Escherichia/Shigella* and *Blautia* in C. No significant correlations were found for many of the genera in the samples.

The sign of the correlation did not depend on the degree of phylogenetic relatedness. Both positive and negative correlations were found between closely and distantly related genera. For example, in subject A, *Paraprevotella* (*Bacteroidetes-Bacteroidia-Bacteroidales-Prevotellaceae*) had a positive correlation with *Odoribacter* and a negative correlation with *Barnesiella*, two genera within *Bacteroidetes-Bacteroidia-Bacteroidales-Porphyromonadaceae*. In subject C, *Butyricimonas* (*Bacteroidetes-Bacteroidia-Bacteroidales-Porphyromonadaceae*) had a positive correlation with *Blautia* and a negative correlation with *Coprococcus*, two genera within *Firmicutes-Clostridia-Clostridiales-Lachnospiraceae*.

The patterns of correlations were highly subject-specific, even for B and C, in which the community structure was quite similar. Typically, the same genus correlated with different genera in different individuals. For example, in subject B, *Bacteroides* co-occurred with *Odoribacter*, while in subject C, it co-occurred with *Parabacteroides* and *Alistipes*. Only five pairwise correlations were found in more than one individual: *Bacteroides-Odoribacter* (A and B), *Bacteroides-Parabacteroides* (A and C), *Odoribacter*-unclassified Proteobacteria (A and B), *Coprococcus-Ruminococcus* (A and C) and *Blautia*-unclassified *Lachnospiraceae* (A and C). Furthermore, nine pairwise correlations were positive in one subject but negative in another: *Blautia-Coprococcus*, *Blautia-Dorea*, *Blautia-Ruminococcus*, *Eubacterium-Ruminococcus*, *Coprococcus-Ruminococcus*, *Coprococcus*-unclassified *Porphyromonadaceae*, *Bacteroides-Subdoligranulum*, unclassified *Bacteroidales-unclassified Porphyromonadaceae* and *Odoribacter-unclassified Veillonellaceae*.

We did not find a relationship between the diversity of the community and the complexity of the associated interaction network. Even though the bacterial community of A was the most diverse at genus level (Shannon index 2.3), and the ones of B and C harboured a similar diversity (Shannon index 1.1), the highest number of genera in the network and the largest average number of links per genus were observed in B and the lowest ones in C.

Discussion

In this study, we have monitored the faecal bacterial communities of three subjects during fifteen consecutive days. We analysed the daily variation in the composition of the microbiota and the degree of stability of specific members of the community. The analysis of the time series also revealed the patterns of correlations between the abundance profiles of specific bacterial groups, thus indicating potential cooperative and competitive interactions between them.

In accordance with current knowledge of the human gut microbiota, we found that the faecal microbiota is host-specific and fairly stable in the absence of perturbation, the within-subject variability being much smaller than between subjects (Franks *et al.*, 1998; Zoetendal *et al.*, 1998; Vanhoutte *et al.*, 2004; Dethlefsen & Relman, 2010; Caporaso *et al.*, 2011). The level of prevalence of the predominant members of the faecal communities of our three study subjects was sustained over time, with abundances fluctuating around constant average values. This also held true for less abundant members, although some of them were not persistently found in our sampling. Additionally, a considerable number of bacterial groups were detected at very low abundance in one or few samples across the time series, probably representing transient members of the communities. Similar observations were made at genus and phylotype (97% of identity) levels.

Two previous studies have examined the daily variation of the faecal microbiota using high-throughput sequencing technologies to study the temporal dynamics at subgeneric levels. Dethlefsen & Relman (2010) monitored the response of faecal microbiota to antibiotic perturbation in three subjects and took daily samples in the periods surrounding each antibiotic course. They found that just a few OTUs maintained uniform abundance, even in the antibiotic-free intervals. Caporaso *et al.* (2011) sampled daily three body sites, including faeces, in two healthy subjects for 15 and 6 months, respectively. They reported that only a small fraction of the OTUs within a single body site was present across all or nearly all time points, and from this, the authors conclude that a minimal temporal core exists. In a shorter temporal window, Boeiijk *et al.* (2010) detected morning-afternoon variation in human ileal microbiota that exceeded the fluctuations between samples collected at the same time point of the day. We found that only a small fraction of the 97% OTUs in our samples was systematically present throughout the follow-up (9% of the total number of OTUs within each subject). However, as their combined abundance is high (74%, 90% and 93% in subjects A, B, and C, respectively), we consider that a high-abundance

core faecal microbiota exists also at the subgeneric level. In our view, these evidences are indicative of a dynamic ecosystem with a stable core. These findings may also suggest that studies aimed at linking the prevalence of specific phylotypes with environmental exposures and/or disease risk should be limited to the numerically most dominant phylotypes, as the rest appears transient.

The bacterial communities sampled in this study differed not only on the basis of their composition, but also in the correlation patterns found between their members. Barely, a few of the correlations were shared between subjects. The vast majority of genera usually correlated with others that varied between subjects, even when the community composition was quite similar (subjects B and C). Furthermore, correlations of opposite sign were detected for some pairs of genera in different subjects. This disparity of community assemblies could imply that each individual may be offering somewhat different niches to gut bacteria (regarding pH, temperature, secretions, retention time, etc.) in which the same species can establish different microbial interactions, even though the gut environment is overall similar in all subjects. At the same time, the identity of the interacting bacteria could be affected by host selection of commensal/mutualistic microorganisms, as well as by the order in which they arrive in colonization processes and selection by already established microorganisms (Van den Abbeele *et al.*, 2011). Because of the functional redundancy, several microorganisms can potentially occupy a specific niche within the intestinal habitats. The first ones to arrive can be established and then select for cooperating or nonoverlapping microorganisms, as well as exclude competitors. Related to this, Mulder *et al.* (2009) found a long impact of early life colonization on the intestinal community composition in pigs. Functional studies would be needed to disentangle the biological meaning of the correlations detected between abundance profiles.

We used a Bayesian model to estimate the covariance matrix between relative abundances of taxa (on the log-odds scale), while accounting for the temporal autocorrelation in those. The posterior mean of the covariance matrix Σ (estimated using all the samples of a given individual) was used to estimate the partial correlation matrix, which was in turn the input for the GGN-based methods to detect associations between taxa. Although these methods cope well with sparse matrices of relatively large dimension, some of the statistically significant associations we found involve taxa with low relative abundances, which makes difficult any biological interpretation.

The persistent diversity and individuality of human gut communities can be explained by a combination of factors that vary between individuals, such as host genotype and diet, but also less predictable events, such as the

colonization history during the community assembly and external perturbations with long-term effect on gut microbiota (Dethlefsen *et al.*, 2006), an example of which is antibiotic treatment. Several studies evaluating its effect on gut microbiota have found an important loss of diversity followed by a rapid return to the pretreatment community composition. The ability of these communities to recover their original structure suggests the existence of selective forces shaping them (De La Cochetiere *et al.*, 2005; Dethlefsen & Relman, 2010). Also, the microbiota itself is thought to account for some of its diversity through the modification of intestinal niches and the interactions established between their members (Dethlefsen *et al.*, 2006; Van den Abbeele *et al.*, 2011). Our data suggest that specific microbial interactions are set within each individual, which may be an important factor contributing to the interindividual variability and the temporal stability of the gut microbiota.

Recently, Arumugam *et al.* (2011) identified three main types of microbial assemblies (called enterotypes) in human gut after clustering faecal samples obtained in different studies based on their species composition or gene pools. Incidentally, we noticed that subjects B and C in our study, the obese ones, can be included in the enterotype enriched in *Prevotella*. Looking further into Arumugam's analysis of the pyrosequencing-based 16S rRNA gene samples from Turnbaugh *et al.* (2009), we noticed that 17 of the 20 individuals of the *Prevotella*-enriched enterotype were obese, whereas 87 of the 134 individuals classified in the other enterotypes were so. This gives (using the EPITOOLS R package, Aragon, 2010) a statistically nonsignificant odds ratio OR = 3.04 ($P = 0.1221$) of association between this enterotype and obesity. Adding our three subjects to the corresponding cells (19 obese out of 22 in the *Prevotella* enterotype and 87 of 135 in the other ones), we obtained an OR = 3.47 ($P = 0.0498$), which begins to be more supportive of a potential link between the *Prevotella* enterotype and some type of obesity. Further studies are needed to confirm this marginal finding and to assess whether the link is owing to the *Prevotella* themselves or there are other genera involved. Recently, De Filippo *et al.* (2010) compared the microbiota composition in African and European children and found that the former showed communities rich in *Prevotella*. They argue that perhaps this genus is more efficient in extracting energy from polysaccharide-rich food. In a low-calorie diet, this may be useful to survive, but otherwise, it may lead to obesity.

Recent studies showed that despite the variation in community structure between subjects (Turnbaugh *et al.*, 2009) and the compositional normal fluctuations detected over time and the shifts owing to disturbances (Dethlefsen & Relman, 2010), the overall community function

seems to be maintained, which is indicative of functional redundancy among individual members and consortia within the gut microbiota. Like in the compositional case, fluctuations are likely to be observed when analysing the functionality of the members of the gut microbiota, despite the stability mentioned above. This is because metabolic functions are probably more susceptible to changes because of the dynamic factors such as resource availability, external perturbations, stressors or host physiology at the moment of sampling. Longitudinal studies of expression patterns within a subject under different conditions may help to better understand the contribution of gut microbiota to human nutrition and well-being.

Acknowledgements

This work has been partly funded by the grants BFU2009-04501-E, BFU2009-12895-C02-01 and SAF2009-13032-C02-01 from Ministerio de Ciencia e Innovación, Spain, and Prometeo/2009/092 from Generalitat Valenciana to A.M. A.D. and N.J. are recipients of a fellowship from Instituto de Salud Carlos III, Spain. The authors thank all three volunteers for participating in this study.

References

- Aragon T (2010) *epitools: Epidemiology Tools. R Package Version 0.5-6*. <http://CRAN.R-project.org/package=epitools>
- Arumugam M, Raes J, Pelletier E *et al.* (2011) Enterotypes of the human gut microbiome. *Nature* **473**: 174–180.
- Banerjee S, Carlin B & Gelfand A (2004) *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, Boca Raton, FL.
- Benson AK, Kelly SA, Legge R *et al.* (2010) Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *P Natl Acad Sci USA* **107**: 18933–18938.
- Booijink CCGM, El-Aidy S, Rajilic-Stojanovic M, Heilig HGJ, Troost FJ, Smidt H, Kleerebezem M, De Vos WM & Zoetendal EG (2010) High temporal and inter-individual variation detected in the human ileal microbiota. *Environ Microbiol* **12**: 3213–3227.
- Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.
- Caporaso JG, Lauber CL, Costello EK *et al.* (2011) Moving pictures of the human microbiome. *Genome Biol* **12**: R50.
- Claesson MJ, Cusack S, O'Sullivan O *et al.* (2010) Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *P Natl Acad Sci USA* **108**: 4586–4591.
- Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM & Tiedje JM (2007) The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res* **35**(Database issue): D169–D172.
- Cole JR, Wang Q, Cardenas E *et al.* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**(Database issue): D141–D145.
- De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G & Lionetti P (2010) Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *P Natl Acad Sci USA* **107**: 14691–14696.
- De La Cochetiere MF, Durand T, Lepage P, Bourreille A, Galmiche JP & Dore J (2005) Resilience of the dominant human fecal microbiota upon short-course antibiotic challenge. *J Clin Microbiol* **43**: 5588–5592.
- Dethlefsen L & Relman DA (2010) Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *P Natl Acad Sci USA* **108**: 4554–4561.
- Dethlefsen L, Eckburg PB, Bik EM & Relman DA (2006) Assembly of the human intestinal microbiota. *Trends Ecol Evol* **21**: 517–523.
- Dethlefsen L, Huse S, Sogin ML & Relman DA (2008) The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6**: e280.
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
- Favier CF, Vaughan EE, Vos WMD & Akkermans ADL (2002) Molecular Monitoring of Succession of Bacterial Communities in Human Neonates. *Society* **68**: 219–226.
- Franks AH, Harmsen HJ, Raangs GC, Jansen GJ, Schut F & Welling GW (1998) Variations of bacterial populations in human feces measured by fluorescent *in situ* hybridization with group-specific 16S rRNA-targeted oligonucleotide probes. *Appl Environ Microbiol* **64**: 3336–3345.
- Gelfand A & Vounatsou P (2003) Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4**: 11–25.
- Gentry J, Long L, Gentleman R, Falcon S, Hahne F, Sarkar D & Hansen K (2008) *Rgraphviz: Provides Plotting Capabilities for R graph Objects*. R package version 1.26.0. <http://cran.r-project.org/>
- Guarner F & Malagelada J-R (2003) Gut flora in health and disease. *Lancet* **361**: 512–519.
- Harmsen HJ, Wildeboer-Veloo AC, Raangs GC, Wagendorp AA, Klijn N, Bindels JG & Welling GW (2000) Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods. *J Pediatr Gastroenterol Nutr* **30**: 61–67.
- Jiménez E, Marín ML, Martín R, Odriozola JM, Olivares M, Xaus J, Fernández L & Rodríguez JM (2008) Is meconium from healthy newborns actually sterile? *Res Microbiol* **159**: 187–193.
- Kelly D, King T & Aminov R (2007) Importance of microbial colonization of the gut in early life to the development of immunity. *Mutat Res* **622**: 58–69.

- Khachatryan ZA, Ktsoyan ZA, Manukyan GP, Kelly D, Ghazaryan KA & Aminov RI (2008) Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS One* **3**: e3064.
- Kim H, Sun D & Tsutakawa RK (2001) A bivariate Bayes method for improving the estimates of mortality rates with 2-fold CAR model. *J Am Stat Assoc* **96**: 1506–1521.
- Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD & Gordon JI (2005) Obesity alters gut microbial ecology. *P Natl Acad Sci USA* **102**: 11070–11075.
- Lunn DJ, Thomas A, Best N & Spiegelhalter D (2000) Winbugs – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* **10**: 325–337.
- Mulder IE, Schmidt B, Stokes CR *et al.* (2009) Environmentally-acquired bacteria influence microbial diversity and natural innate immune responses at gut surfaces. *BMC Biol* **7**: 79.
- Neish AS (2009) Microbes in gastrointestinal health and disease. *Gastroenterology* **136**: 65–80.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RB, Simpson GL, Solymos P, Stevens MHH & Wagner H (2010) *vegan: Community Ecology Package*. R package version 1.17-2. <http://cran.r-project.org/>
- Opgen-Rhein R & Strimmer K (2007) From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst Biol* **1**: 37.
- Palmer C, Bik EM, DiGiulio DB, Relman DA & Brown PO (2007) Development of the human infant intestinal microbiota. *PLoS Biol* **5**: e177.
- Qin J, Li R, Raes J *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**: 59–65.
- Quinn GP & Keough MJ (2002) *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge.
- R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Schaefer J, Opgen-Rhein R & Strimmer K (2009) *GeneNet: Modeling and Inferring Gene Networks*. R package version 1.2.4. <http://cran.r-project.org/>
- Schäfer J & Strimmer K (2005a) An empirical bayes approach to inferring large-scale gene association networks. *Bioinformatics* **21**: 754–764.
- Schäfer J & Strimmer K (2005b) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol* **4**: Article32.
- Schäfer J, Opgen-Rhein R & Strimmer K (2010) *corpcor: Efficient Estimation of Covariance and (Partial) Correlation*. R package version 1.5.6. <http://cran.r-project.org/>
- Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Sturtz S, Ligges U & Gelman A (2005) R2winbugs: a package for running winbugs from r. *J Stat Softw* **12**: 1–16.
- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER & Gordon JI (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1031.
- Turnbaugh PJ, Hamady M, Yatsunencko T *et al.* (2009) A core gut microbiome in obese and lean twins. *Nature* **457**: 480–484.
- Van den Abbeele P, Van de Wiele T, Verstraete W & Possemiers S (2011) The host selects mucosal and luminal associations of co-evolved gut microbes: a novel concept. *FEMS Microbiol Rev* **35**: 681–704.
- Vanhoutte T, Huys G, Brandt E & Swings J (2004) Temporal stability analysis of the microbiota in human feces by denaturing gradient gel electrophoresis using universal and group-specific 16S rRNA gene primers. *FEMS Microbiol Ecol* **48**: 437–446.
- van Vliet MJ, Tissing WJE, Dun CAJ, Meessen NEL, Kamps WA, de Bont ESJM & Harmsen HJM (2009) Chemotherapy treatment in pediatric patients with acute myeloid leukemia receiving antimicrobial prophylaxis leads to a relative increase of colonization with potentially pathogenic bacteria in the gut. *Clin Infect Dis* **49**: 262–270.
- Woodmansey EJ (2007) Intestinal bacteria and ageing. *J Appl Microbiol* **102**: 1178–1186.
- Zoetendal EG, Akkermans ADL & Vos WMD (1998) Temperature Gradient Gel Electrophoresis Analysis of 16S rRNA from Human Fecal Samples Reveals Stable and Host-Specific Communities of Active Bacteria. *Appl Environ Microbiol* **64**: 3854–3859.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Number of sequences, observed number of genera, Chao1 richness estimator (using all sequences in the sample), average values of the Chao1 richness estimator and Shannon's index of biodiversity (and the corresponding standard errors) computed for 1000 subsamples of size 590 (the size of the smallest sample).

Table S2. Number of sequences, observed number of Operational Taxonomic Units (OTU, sequences grouped at 97% identity), Chao1 richness estimator (using all sequences in the sample), average values of the Chao1 richness estimator and Shannon's index of biodiversity (and the corresponding standard errors) computed for 1000 subsamples of size 590 (the size of the smallest sample).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.