ORIGINAL ARTICLE

Convergent temporal dynamics of the human infant gut microbiota

Pål Trosvik^{1,2}, Nils Christian Stenseth² and Knut Rudi^{1,3}

¹NOFIMA Mat, Norwegian Food Research Institute, Ås, Norway; ²Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology, University of Oslo, Oslo, Norway and ³Hedmark University College, Hamar, Norway

Temporal dynamics of the human gut microbiota is of fundamental importance for the development of proper gut function and maturation of the immune system. Here we present a description of infant gut ecological dynamics using a combination of nonlinear data modeling and simulations of the early infant gut colonization processes. Principal component analysis of infant microbiota 16S rRNA gene microarray data showed that the main directions of variation were defined by three phylum-specific probes targeting Bacteroides, Proteobacteria and Firmicutes. Nonlinear regression analysis identified several dynamic interactions between these three phyla. Simulations of the early phylum-level colonization process showed the relatively rapid establishment of an equilibrium community after an unstable initial phase. In general, varying the initial composition of phyla in the simulations had little bearing on the final equilibrium. The dynamic interaction model was found to maintain its predictive ability for Proteobacteria and Firmicutes well into the simulation, whereas Bacteroides densities tended to be underestimated, possibly due to host top-down selection for Bacteroides. In accordance with our model, initial perturbation of the microbiota by different mode of delivery (vaginal and C-section) did not affect the later phylum composition in the infants investigated. Considering the predictive ability and convergence of our phylum-level model, we now propose that deterministic bacterial–bacterial interactions are more important for shaping the human infant gut microbiota than previously anticipated.

The ISME Journal (2010) 4, 151–158; doi:10.1038/ismej.2009.96; published online 27 August 2009

Subject Category: microbial population and community ecology

Keywords: microbial ecology; human infant; gut microbiota; population dynamics

Introduction

Although the importance of the human gastrointestinal microbiota is now widely recognized (Stappenbeck *et al.*, 2002; Guarner and Malagelada, 2003; Hooper, 2004; Rakoff-Nahoum *et al.*, 2004; Backhed *et al.*, 2005; Mazmanian *et al.*, 2005; Turnbaugh *et al.*, 2006; Frank *et al.*, 2007; Li *et al.*, 2008), the mechanistic foundations of persistence and dynamics are not well developed (Zoetendal *et al.*, 2006; Ley *et al.*, 2006a; Dethlefsen *et al.*, 2007). One reason for this is the fact that many concepts and methods developed for the ecological study of animals and plants have yet to be implemented in microbial ecology (Prosser *et al.*, 2007). A second complicating factor is the virtual absence of good time-series data from microbial systems such as the human gastrointestinal microbiota. In order to achieve realistic models of systems'

dynamics and biotic/abiotic interactions, temporal population data of adequate resolution are required (Turchin, 2003). Recently, the first relatively extensive time-series data set describing the development of the microbiota in 14 infants became available in the public domain (Palmer *et al.*, 2007). This set was based on ~25 stool samples from each child, collected sequentially, and subsequently analyzed using microarrays measuring relative quantities of 2149 different species and taxonomic groups.

There has been a debate as to what extent the human microbiota is controlled by the host through top-down mechanisms involving, for example, the immune system, relative to microbiota-intrinsic mechanisms involving competitive and cooperative ecological interactions (Dethlefsen *et al.*, 2006; Ley *et al.*, 2006a). In infants, it is reasonable to assume that the latter mechanisms are dominant, as infants do not possess fully developed immune systems (Kelly *et al.*, 2007). As such, it should be feasible to derive realistic dynamic models of the infant microbiota based solely on population time series, within analytical and conceptual frameworks familiar from other ecosystems (Trosvik *et al.*, 2008).

Correspondence: K Rudi, NOFIMA Mat, Norwegian Food Research Institute, Osloveien 1, 1430 Ås, Norway.

E-mail: knut.rudi@nofima.no

Received 1 May 2009; revised 20 July 2009; accepted 20 July 2009; published online 27 August 2009

The biotic factors that regulate populations of organisms are generally thought of as being density dependent, and they often operate in a nonlinear fashion (Turchin, 1995). Statistical models for investigating population dynamic phenomena may be formulated with specified parameters describing the various density-dependent mechanisms (Kendall *et al.*, 1999), be that for the purpose of inference, hypothesis testing or prediction. The implementation of the mechanistic modeling approach does, however, require extensive *a priori* knowledge of the ecological system under study to properly specify the functional form of the individual model terms (Wood, 2001). In cases in which *a priori* information on a system is scarce, a nonparametric modeling approach that requires a minimum of assumptions about the mechanistic workings of the system is often preferable. The generalized additive modeling framework (Hastie and Tibshirani, 1990) offers one such approach. This regression technique incorporates nonparametric smoothing functions for investigation of nonlinear effects without the need for *a priori* assumptions about functional relationships, with the possibility of including fully specified parametric terms in a model. Generalized additive models (GAMs) have been used extensively for analysis of short and noisy ecological time series, providing insight into the dynamics of a range of different systems (Jannicke Moe *et al.*, 2005; Stenseth *et al.*, 2006; Stige *et al.*, 2006).

The aim of this study was to explore and model the intra-bacterial interactions underlying the main dynamic phenomena in the early infant gut, using the GAM framework without any underlying assumptions other than that members of the microbiota somehow interact with one another. The recently published microarray data (Palmer *et al.*, 2007) and quantitative PCR (qPCR) data from the same study formed the empirical foundation of this study. Owing to the sampling scheme, our analysis was focused on data for the first 2 weeks of the infants' lives, as sampling during this period was frequent and regular. This focus should also help minimize any effects of host forcing, which may otherwise represent a confounding factor in our analysis. Using our interaction model, we simulated the early colonization process under a variety of early exposure scenarios.

We present results showing that the main dynamic phenomena in the data can be explained by the three phylum-range (Proteobacteria, Firmicutes and Bacteroides) probes. In the simulations the three phyla were generally able to reach a stable state within 20–40 days, although certain initial conditions were found to delay the establishment of equilibrium by a significant amount of time. In the case of Proteobacteria and Firmicutes, we show that the dynamic model retained good predictive ability after 120 days of simulated growth. For Bacteroides, however, this ability was lost, and the model had a tendency to underestimate the density of this group

at advanced time points, suggesting that factors other than intra-bacterial interactions are complicit in the establishment of Bacteroides in the infant microbiota. We discuss the main dynamic interactions identified, the consequences of these interactions for predicting the colonization process in the human infant gut and the evolutionary implications of the phylum-level interaction model.

Materials and methods

Data

We used data recently published by Palmer *et al.* (2007). These were phylogenetic microarray data based on 2149 16S rRNA gene probes with varying degrees of specificity. The arrays were used to survey the fecal bacterial content of 14 infants from birth until 1 year of age. In particular, for the first 2 weeks each infant was sampled on a nearly daily basis, with subsequent sampling being less frequent and regular. Owing to this sampling scheme, we focused our analysis on the early samples.

Real-time qPCR data were also made available by Palmer *et al.* for the majority of the samples assayed by microarray. These data provided us with a measure of overall bacterial density in the samples.

For the initial principal component analysis (PCA), we used microarray data from all infants corresponding to the first 2 weeks of sampling, representing relative abundance measurements for a total of 161 samples. Before computing the PCA model all probe intensity values were converted to percentages of the total probe intensity for a given sample.

A number of samples were left out of the nonlinear regression modeling. These samples included all those taken from infants 4, 6 and 10, as these individuals had received antibiotic treatment during their first 2 weeks of life. Antibiotics constitute an obvious source of external forcing of the gut microbiota, and as our main concern was bacteria–bacteria interactions, these samples would have represented a confounding factor in our analysis. We observed dramatic dips at certain points in the qPCR data for some of the infants (Supplementary Figure S1). In three cases, these measurements were identical to the sixth decimal, leading us to believe that these extreme values represented a detection threshold. For infant 1, the dip occurred in the middle of the time series, and thus data for this individual was excluded from further analysis in its entirety. In infants 2 and 8, the dips occurred at the beginning of the series, and the extreme measurements were removed from the data along with measurements for any preceding samples (three samples from infant 2 and two samples from infant 8). Finally, the first sample from infants 5, 9 and 12 were omitted, as these were found to constitute high leverage outliers. Such deviant data points would have undue influence in regression modeling, making the estimated model less robust.

To obtain response variables for nonlinear regression, the 10 short series included in the analysis of each bacterial group were log transformed and differenced (see below) separately before concatenation, producing three final series each of 98 data points. Similarly, predictor variables were log transformed before concatenation of lag 1 measurements, producing three series of equal length.

Statistical analysis

Principal component analysis is a data compression and visualization technique that extracts latent variables representing directions of maximum variance in data matrices. This is carried out by decomposing a matrix of samples and variables, $\mathbf{X}$, into a set of orthogonal (uncorrelated) scores, $\mathbf{T}$, representing the main variance components within the samples (infants sampled at different time points) and loadings, and, $\mathbf{P}$, defining the variables (microarray probes) responsible for the main observed variation. The original matrix of samples and variables can then be written as the product of the most dominant scores and their corresponding loadings, $\mathbf{X} = \mathbf{TP}^t + \mathbf{E}$, where $\mathbf{E}$ is a matrix of residuals representing the unexplained variation in $\mathbf{X}$. PCA was carried out using the 'prcomp' function in the R package 'stats'. All variables were zero centered by subtraction of the variable means before computation of the PCA model. For a more thorough account of PCA theory, we refer to Martens and Næs (1989).

Nonlinear regression analysis was carried out by fitting GAMs (Hastie and Tibshirani, 1990). Within the GAM framework a normally distributed response variable, y , is related to a set of explanatory variables, x and z , through the nonparametric smooth functions, s_1 and s_2 , as in $y = \beta + s_1(x) + s_2(z) + \varepsilon$, where β is an intercept representing the response mean and ε is a residual term. The nonparametric functions, s_1 and s_2 , are cubic spline smoothers, that is, piecewise cubic polynomials with the pieces joined at knots specified by the values of x and z . The optimal degree of smoothing is estimated using generalized cross-validation (GCV) (Wood, 2000), which provides a means of balancing goodness-of-fit with model complexity. For a systematic account of the GAM methodology, we refer to (Wood, 2006).

In the case of Proteobacteria, the full GAM would be as follows:

$$\Delta P_t = \beta_P + f_P(P_t) + g_P(B_t) + h_P(F_t) + \varepsilon_{Pt}$$

where ΔP_t is a vector containing the per time-unit changes in log Proteobacterial abundance calculated as $(P_{t+1} - P_t)$. f_P , g_P and h_P are nonparametric smooth functions specifying the effects of log abundances of Proteobacteria (P_t), Bacteroides (B_t) and Firmicutes (F_t), respectively. β_P is an intercept, whereas ε_{Pt} is an error term. Full models for Bacteroides and Firmi-

cutes had equivalent formulations. For each smooth term in each model, the maximum degrees of freedom were set to 4. This puts a restraint on the nonlinearity of the smooth terms, decreasing the chance of overfitting.

The overall model structure of each GAM was also determined according to the GCV criterion (Wood, 2000), with the model structure producing the lowest GCV score generally being preferred. The GAM algorithm provides approximate P -values for assessing the fit to the data of the individual smooth functions. Final models were selected by removing smooth terms with high P -values, inspecting the resulting GCV scores, and omitting the terms whose removal lowered the GCV score (Supplementary Table S1). All GAM computations were carried out using the implementation in the R package 'mgcv' (Wood and Augustin, 2002).

The applicability of the additivity assumption implicit in the GAM model formulation was checked using a Lagrange multiplier test for additivity (Chan *et al.*, 2003), using 300 bootstrap repeats. This technique tests whether the coefficients of higher-order interaction terms within a model equation are significantly different from zero, that is, whether it would be appropriate to include model terms that are products of two or more covariates. As our models featured a maximum of three covariates, we carried out the test for all possible combinations of second-order interaction terms in the case of each model equation.

Simulations of the early colonization process of the infant gut were carried out using various starting densities of Proteobacteria, Firmicutes and Bacteroides according to a factorial scheme (Supplementary Table S3), in which starting abundances of each group were either high (9.6 log units for Bacteroides, 9.7 for Proteobacteria and 9.5 for Firmicutes) or low (4.2 log units for Bacteroides, 6.0 for Proteobacteria and 6.2 for Firmicutes). The starting values were chosen based on the span of the observed abundance distributions of the respective groups (Supplementary Table S3). The rationale for the length of the simulation was the fact that 13 of the 14 babies who were surveyed were sampled after ~120 days (baby 1 being the exception) and provided both microarray and qPCR data for assessing the predictive ability of the model. For each step of the simulation, the bacterial abundance covariates were updated by model predictions from the previous time step, producing the simulated series.

All statistical analyses were carried out using the R software (R Development Core Team, 2007).

Results

Principal component analysis of the microarray relative abundance data for all 14 infants during the first 2 weeks showed that variability in probe intensity could be attributed mainly to abundance

Figure 1 Principal component analysis (PCA) plot for microarray data showing the importance of the phylum probes. The two first principal components are plotted with the proportion of variance explained by each component printed next to the axes labels. The plot clearly shows that most of the variance in the microarray data can be attributed to just three phylum-range probes.

fluctuations in three bacterial phyla (Figure 1); Bacteroides (represented by probe 2.15—FLEXI-BACTER-CYTOPHAGA-BACTEROIDES), Proteobacteria (represented by probe 2.28—PROTEO-BACTERIA) and Firmicutes (represented by probe 2.30—GRAM-POSITIVE BACTERIA). The probes specific for these divisions defined the directions of the two highest eigenvalue principal components, accounting for 80.1% of the variance in the microarray data. Therefore, further analysis was focused on data for these three probes.

Multiplication of the relative abundance estimates of the domains in question by their corresponding qPCR measurements produced 14 short time series (Supplementary Figure S1). Data from 10 of these series were used for GAM modeling (see rationale in Materials and methods section). From the class of full-additive models, the following three were selected according to the GCV criterion (Supplementary Table S2) as described in the Materials and methods section:

$$\begin{aligned}\Delta P_t &= \beta_P + f_P(P_t) + g_P(B_t) + h_P(F_t) + \varepsilon_{Pt} \\ \Delta F_t &= \beta_F + f_F(P_t) + g_F(B_t) + h_F(F_t) + \varepsilon_{Ft} \\ \Delta B_t &= \beta_B + g_B(B_t) + h_B(F_t) + \varepsilon_{Bt}\end{aligned}\quad (1)$$

Inspection of model residuals showed that these were approximately normally distributed (Supplementary Figure S2), homoscedastic (that is, have homogeneous variances; Supplementary Figure S3) and without discernable autocorrelation structure (Supplementary Figure S4). Using the Lagrange multiplier test, we did not identify any statistically

Table 1 Descriptive statistics for Equation 1

	<i>E.d.f.</i>	<i>F</i>	<i>P-value</i>
<i>Proteobacteria</i>			
$f_P(P_t)$	1.90	2.05	0.094
$g_P(B_t)$	1.85	5.55	< 0.001
$h_P(F_t)$	1.00	4.49	0.037
$R^2 = 0.28$			$n = 98$
<i>Firmicutes</i>			
$f_F(P_t)$	2.78	3.40	0.012
$g_F(B_t)$	1.73	3.15	0.018
$h_F(F_t)$	2.89	7.57	< 0.001
$R^2 = 0.37$			$n = 98$
<i>Bacteroides</i>			
$g_B(B_t)$	2.11	7.98	< 0.001
$h_B(F_t)$	1.00	4.65	0.034
$R^2 = 0.25$			$n = 98$

Explained variance, estimated degrees of freedom (E.d.f.), F-statistics and significance of smooth terms included in the final selected models. One term not significant at the 95% confidence level was kept in the model describing Proteobacterial dynamics, as its omission was found to increase the generalized cross-validation score (Supplementary Table S1).

significant second-order interaction terms (Supplementary Table S2), suggesting that the fully additive model formulation specified above was appropriate. Descriptive statistics for the selected models are listed in Table 1, showing the statistical significance of model terms and demonstrating considerable proportions of explained variance in the case of each model. Figure 2 shows the shapes of the smooth terms included in the selected models, that is, the type of effect that each covariate term was found to have on their respective response terms. These plots can be used to identify the nature of the bacteria–bacteria interactions, for example, the strength of competition can be discerned by checking the negative slope of a smooth term. According to the selected interaction model, the three bacterial groups display varying degrees of density-dependent intra-phylum competition (Supplementary Figure 2a, f and g), with the strongest effect seen in the Firmicutes (Figure 2f) and the weakest in the Proteobacteria (Figure 2a). Furthermore, the Proteobacteria are in direct competition with the two other groups (Figures 2a and b), whereas the Firmicutes are inhibited by Bacteroides (Figure 2e), but apparently benefit from the presence of Proteobacteria at moderate densities (Figure 2d). The Bacteroides are in a competitive relationship with the Firmicutes (Figure 2h), but are apparently unaffected by the Proteobacteria. The interaction model (Equation (1), Figure 2) is represented as a network diagram in Figure 3.

Using the interaction model of Equation (1), we simulated the colonization process for 120 time steps (~ 4 months) at various initial compositions of the three phyla. In most of the simulations the three phyla were able to reach an early equilibrium state with the Proteobacteria dominating, Firmicutes occurring at intermediate levels and Bacteroides

Figure 2 Plots of smooth terms from the selected generalized additive models (GAMs) (Equation (1)). The graphs summarize the functional relationship between response variable and covariate terms for each model equation, indicating the nature of the interactions between the three bacterial phyla (that is, degree of competition, cooperation, and nonlinearities in their functional relationships). Covariate values (log abundances; smooth terms on the right-hand side of Equation (1)) are on the abscissae. The ordinate axes indicate the partial additive effect of each covariate term on the response (per time-unit change in log abundance of a phylum; left-hand side of Equation (1)). For example, panel **a** describes the effect of proteobacterial abundance on proteobacterial dynamics (that is, intra-phylum interactions), which can be seen to be nonlinear but predominantly negative (competitive). The interpretation of the nonlinearity in this term would be that, at low abundances, competition within the proteobacterial phylum is not detectable, whereas at higher densities, the Proteobacteria inhibit their own growth. Panels **b** and **c** describe the effect of Bacteroides and Firmicutes log abundances on proteobacterial dynamics, respectively. Both of these terms are strictly negative (competitive), meaning that higher densities of Bacteroides and Firmicutes have more adverse effects of proteobacterial proliferation. The response term of a given function is approximated by the sum of its smooth terms plus an intercept representing the response mean. By convention all smooth terms of a function sum to zero. The shaded areas are 95% confidence bands. The rug plots along the abscissae indicate covariate values. The dots surrounding the regression lines are partial residuals. Panels **a**, **b** and **c** are from the GAM describing Proteobacteria (P) dynamics. Panels **d**, **e** and **f** are for Firmicutes (F). Panels **g** and **h** are for Bacteroides (B).

Figure 3 Network chart showing a model for the phylum-level interactions. The arrows indicate the effect of one group of bacteria on another, with the signs next to the arrows indicating the nature of that effect. Arrows labeled with both plus and minus signs represent interactions described by smooth functions with non-monotonic shapes. Letters in parenthesis (A–H) refer to the panels of Figure 2 where effects are visualized as generalized additive model plots.

occurring at the lowest densities (Figure 4). However, not all of the simulations converged on the same profile at the end of the simulated period. Low starting densities of Bacteroides along with high densities of Proteobacteria resulted in elevated levels of both these groups (7.1 vs 6.8 and 9.2 vs 9.0 log units, respectively) after 120 simulated time steps, regardless of the starting density of Firmicutes (Figures 4a and b), although yielding decreased Firmicutes densities (7.7 vs 8.3 log units) (Figures 4b and c). From these starting points nearly 250 time steps of simulation were needed for the equilibrium state to be reached (data not shown).

Discussion

The study by Palmer *et al.* (2007) showed that the infant microbiota can be extremely unstable, both in

Figure 4 Simulated colonization patterns illustrating the microbiota convergence with age. Gut colonization by *Bacteroides*, *Proteobacteria* and *Firmicutes* simulated for 120 time steps using the estimated interaction model (Equation (1)). The axes indicate log abundances of the different groups. Green circles indicate starting points of the simulations. Each panel shows the progression of the two groups indicated on the axes conditional on either high (black arrows) or low (red arrows) initial densities of the third group, with each arrowhead indicating a separate time step. The blue circle indicates the mean log abundance of the plotted groups as based on measurements of 13 infants sampled after ~4 months. The blue crosshairs indicate two standard errors (s.e.) of the means. The dotted blue lines also indicate two s.e. to facilitate the assessment of prediction accuracy. The centers of the yellow circles represent the predicted final equilibrium densities. (a) Colonization of *Bacteroides* and *Proteobacteria* conditional on *Firmicutes*. (b) Colonization by *Bacteroides* and *Firmicutes* conditional on *Proteobacteria*. (c) Colonization by *Proteobacteria* and *Firmicutes* conditional on *Bacteroides*.

terms of community composition and total bacterial load. In this study we have shown that, even though strong fluctuations may be observed, these dynamic properties may, to a large extent, be explained by a relatively simple model of bacterial interactions.

The main feature of our interaction model was competition, both within and between phyla. These interactions are crucial for determining the long-term structure of the system (Figure 4) by checking the abundance of individuals within each phylum. The *Firmicutes* were found to display the most pronounced intra-phylum competitive effect (Figure 2f), suggesting a role for members of this phylum as founder bacteria with high initial growth rates and reduced competitiveness at high densities. These properties would enable the *Firmicutes* to rapidly colonize vacant niches, although they would not be able to sustain extremely high densities due to competitive pressure. Strong intra-phylum competition could also explain the large fluctuations seen in the initial part of the simulation of *Firmicutes* growth (Figures 4b and c). Interestingly, the intra-phylum effect is quite weak within the *Proteobacteria* (Figure 2a), probably facilitating the high equilibrium densities reached by this phylum both in simulations and in the empirical data (Figures 4a and c). The *Proteobacteria* also display quite smooth trajectories towards the equilibrium state, as well as a quite low standard error for the abundance measurements at 120 days, which may also be explained by lax intra-phylum competition. It is also interesting to note that up to a threshold density of ~8 log units *Proteobacteria* seem to be of benefit to the *Firmicutes* (Figure 2d), although it is not clear how one group of bacteria may facilitate the growth of the other. The remaining interactions in the model are strictly competitive, contributing to phylum-level population regulation and the ultimate stabilization at the equilibrium obtained in the simulated colonization (Figure 4).

In the model simulations, the three groups were generally predicted to reach a stable state roughly

within a month, with the time required to get there, as well as the path taken, determined by initial exposure (Figure 4). The fact that certain initial conditions caused the equilibrium state to be delayed by ~150 days underscores the potential role of initial exposure in early infant microbiota dynamics. In the case of *Proteobacteria* and *Firmicutes*, both the states reached by the end of the simulated period roughly match those observed in the data produced by Palmer *et al.* (2007), with predictions falling within two standard errors of the mean densities as measured by microarray and qPCR (Figure 4). As our dynamic model only included bacteria–bacteria interactions, this result suggests that competition and commensalism among microbes go a long way in explaining the persistence of these two important groups of gut inhabitants during the early phase of life. However, for *Bacteroides* both of the predicted final states represent lower densities than the observed mean minus two standard errors at 120 days (Figures 4a and b). This result suggests that bacteria–bacteria interactions are not adequate for explaining the dynamics of *Bacteroides* in the infant gut microbiota, and that group selection by the host may be taking place.

Clearly, there are potential factors besides the microbiota itself that may influence the bacterial population dynamics of the infant gut. Such factors may be, for example, the presence of bacteriophages or bacteria–host interactions (Dethlefsen *et al.*, 2006). Contingencies of early life events, such as mode of delivery, have also been suggested as determinants of microbiota development (Penders *et al.*, 2006). Surprisingly, however, we did not find statistically significant effects of mode of delivery on microbiota composition in 4-month-old infants as pertaining to the three dominant bacterial phyla (Supplementary Table S4). In fact, our dynamic model was able to account for significant proportions of the observed variation, suggesting that bacteria–bacteria interactions form a key element

in determining the composition of the human gastrointestinal microbiota early in life.

Our results support the notion that development of the human gastrointestinal microbiota is largely deterministic, both because of interactions within the microbiota and host forcing. In our model, the tendency of the microbiota to converge on a common profile may be explained by a small number of simple phylum-level interactions, although the predominance of *Bacteroides* in the more mature profile requires further explanatory factors. Phylum-level convergence is also in accordance with the recent discovery that the phylogeny of the human microbiota has a fan-like structure with low complexity on the deep lineage level, but very high complexity at more shallow levels (Dethlefsen *et al.*, 2007). Furthermore, it has been shown that the human microbiota is linked to the host metabolic phenotype in obese individuals through phylum-level interactions involving the balance between abundances of Firmicutes and *Bacteroides* (Ley *et al.*, 2006b).

According to our model, ecological interactions at the phylum level are structured and deterministic, whereas the study by Palmer *et al.* (2007) showed that dynamics at lower taxonomic levels are seemingly chaotic. These results lend themselves to the interpretation that phylum-level interactions are evolutionary characters that are most likely ancient and conserved, promoting the emergence of a functional community profile. We believe that, as more and better data become available, integration of evolutionary thinking and ecological modeling will be of great importance for understanding the mechanisms underpinning the development of the human gut microbiota.

References

- Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. (2005). Host-bacterial mutualism in the human intestine. *Science* **307**: 1915–1920.
- Chan KS, Kristoffersen AB, Stenseth NC. (2003). Burmann expansion and test for additivity. *Biometrika* **90**: 209–222.
- Dethlefsen L, Eckburg PB, Bik EM, Relman DA. (2006). Assembly of the human intestinal microbiota. *Trends Ecol Evol* **21**: 517–523.
- Dethlefsen L, McFall-Ngai M, Relman DA. (2007). An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* **449**: 811–818.
- Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci USA* **104**: 13780–13785.
- Guarner F, Malagelada JR. (2003). Gut flora in health and disease. *Lancet* **361**: 512–519.
- Hastie TJ, Tibshirani RJ. (1990). *Generalized Additive Models*. Chapman & Hall: London, **XV**. pp 335s.
- Hooper LV. (2004). Bacterial contributions to mammalian gut development. *Trends Microbiol* **12**: 129–134.
- Jannicke Moe S, Kristoffersen AB, Smith RH, Stenseth NC. (2005). From patterns to processes and back: analysing density-dependent responses to an abiotic stressor by statistical and mechanistic modelling. *Proc Biol Sci* **272**: 2133–2142.
- 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.
- Kendall BE, Briggs CJ, Murdoch WW, Turchin P, Ellner SP, McCauley E *et al.* (1999). Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches. *Ecology* **80**: 1789–1805.
- Ley RE, Peterson DA, Gordon JI. (2006a). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**: 837–848.
- Ley RE, Turnbaugh PJ, Klein S, Gordon JI. (2006b). Microbial ecology: human gut microbes associated with obesity. *Nature* **444**: 1022–1023.
- Li M, Wang B, Zhang M, Rantalainen M, Wang S, Zhou H *et al.* (2008). Symbiotic gut microbes modulate human metabolic phenotypes. *Proc Natl Acad Sci USA* **105**: 2117–2122.
- Martens H, Næs T. (1989). *Multivariate Calibration*. Wiley: Chichester, **XVII**. pp 419s.
- Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**: 107–118.
- Palmer C, Bik EM, Digiulio DB, Relman DA, Brown PO. (2007). Development of the human infant intestinal microbiota. *PLoS Biol* **5**: e177.
- Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I *et al.* (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* **118**: 511–521.
- Prosser JI, Bohannan BJ, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP *et al.* (2007). The role of ecological theory in microbial ecology. *Nat Rev Microbiol* **5**: 384–392.
- R Development Core Team (2007). *R is a language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria.
- Rakoff-Nahoum S, Paglino J, Eslami-Varzaneh F, Edberg S, Medzhitov R. (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* **118**: 229–241.
- Stappenbeck TS, Hooper LV, Gordon JI. (2002). Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc Natl Acad Sci USA* **99**: 15451–15455.
- Stenseth NC, Llope M, Anadon R, Ciannelli L, Chan KS, Hjermann DO *et al.* (2006). Seasonal plankton dynamics along a cross-shelf gradient. *Proc R Soc B Biol Sci* **273**: 2831–2838.
- Stige LC, Stave J, Chan KS, Ciannelli L, Pettorelli N, Glantz M *et al.* (2006). The effect of climate variation on agro-pastoral production in Africa. *Proc Natl Acad Sci USA* **103**: 3049–3053.
- Trosvik P, Rudi K, Naes T, Kohler A, Chan KS, Jakobsen KS *et al.* (2008). Characterizing mixed microbial population dynamics using time-series analysis. *ISME J* **2**: 707–715.
- Turchin P. (1995). Population regulation: old arguments and a new synthesis. In: Price PW and Cappuccino N (eds). *Population Dynamics: New Approaches and Synthesis*. Academic Press: San Diego. pp 19–37.
- Turchin P. (2003). *Complex Population Dynamics a Theoretical/Empirical Synthesis*. Princeton University Press: Princeton, NJ. pp 450s.

- 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.
- Wood SN. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *J R Stat Soc Series B Stat Methodol* **62**: 413–428.
- Wood SN. (2001). Partially specified ecological models. *Ecol Monogr* **71**: 1–25.
- Wood SN. (2006). *Generalized Additive Models: an Introduction with R*. Chapman & Hall/CRC: Boca Raton, xvii. pp. 391s.
- Wood SN, Augustin NH. (2002). GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecol Modell* **157**: 157–177.
- Zoetendal EG, Vaughan EE, de Vos WM. (2006). A microbial world within us. *Mol Microbiol* **59**: 1639–1650.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)