The ISME Journal (2008) 2, 707–715

www.nature.com/ismej

ORIGINAL ARTICLE

Characterizing mixed microbial population dynamics using time-series analysis

Pål Trosvik[1,2], Knut Rudi[2,3], Tormod Næs[2,4], Achim Kohler[2], Kung-Sik Chan[5], Kjetill S Jakobsen[1] and Nils C Stenseth[1]

[1]*Department of Biology, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway;* [2]*Matforsk AS, Norwegian Food Research Institute, Norway;* [3]*Department of Natural Sciences, Hedmark University College, Hamar, Norway;* [4]*Department of Mathematics, University of Oslo, Oslo, Norway and* [5]*Department of Statistics and Actuarial Sciences, University of Iowa, Iowa City, IA, USA*

Due to a general shortage of temporal population data, dynamic structures in microbial communities remain largely unexplored. Knowledge of community dynamics is, however, essential for understanding the mechanisms by which microbes interact. Here, we have used a computational approach for quantification of bacteria in multispecies populations, generating data for time-series modeling. Moreover, we have used online FR-IR spectroscopy to monitor the main metabolic processes. The approach enabled us to provide a functional description of the parameters governing the population dynamics in a three-species model bacterial community, demonstrating density-dependent regulation, interspecies competition and even a case of cooperation between two species. Since the field of microbial ecology has yet to embrace many of the concepts and methods developed for the study of ecology of higher plants and animals, the realization that microbial systems can be analyzed within the same conceptual framework as other ecosystems is of fundamental importance.

The ISME Journal (2008) **2,** 707–715; doi:10.1038/ismej.2008.36; published online 3 April 2008
Subject Category: microbial population and community ecology
Keywords: bacterial community; GAM; model system; population dynamics; time series; FT-IR

Introduction

As for all ecological systems, dynamics in bacterial communities are governed by interaction among community members as well as by the environmental conditions (Read and Taylor, 2001; Bell *et al.*, 2005). In general ecological science one often encounters terms such as within and between species competition, cooperation and density dependence (Turchin, 1995; Stenseth *et al.*, 1998; Ciannelli *et al.*, 2005; Moe *et al.*, 2005). Such concepts are much less frequently explored in the context of microbes (Hagen *et al.*, 1982; Bradshaw *et al.*, 1994; Rainey and Rainey, 2003; You *et al.*, 2004; Balagadde *et al.*, 2005). This is partly due to a historical division between the fields of microbiology and general ecology, but the difficulty of quantitative population data collection from complex microbial systems has also been a major factor contributing to the paucity of information on population dynamics in microcosm. Nevertheless, given a convenient sampling scheme, microbial communities as model systems offer certain advantages relative to traditional ecological field systems (Jessup *et al.*, 2004). Large population sizes can provide analytical results of increased statistical robustness, short generation times enable researchers to follow hundreds of generations in a limited amount of time, and controlled laboratory settings ensure experimental tractability. Also, the wealth of knowledge on microbial cell function, along with the fact that microbes are easily amenable to genetic engineering, provides a unique opportunity to investigate the physiological and genetic underpinnings of ecological processes. Indeed, microbial model systems have, in several cases, proved valuable for ecological studies. The pioneering works of Gause during the 1930s, using experimental systems of eukaryotic microorganisms, were instrumental to the development of some of the basic tenets of population ecology (Gause, 1932, 1935), and since then, microbial models have been used to study various aspects of ecology and evolution (Gorden *et al.*, 1969; Dykhuizen and Davies, 1980; Cooper and Lenski, 2000; Rainey and Rainey, 2003).

Even though a substantial amount of ecological research has been carried out using microbial model

Correspondence: NC Stenseth, Department of Biology, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo N-0316, Norway.
E-mail: n.c.stenseth@bio.uio.no
Received 22 November 2007; revised 13 March 2008; accepted 14 March 2008; published online 3 April 2008

systems, functional descriptions of community dynamic structures are all but absent from the literature. For such a description to be formulated, time-series population data, collected on a scale relevant to the organisms under investigation, are required (Paerl and Steppe, 2003; McArthur, 2006). In the work presented here, we have taken a computational approach to quantification of individual species in mixed populations, based on multivariate analysis of bacterial community 16S rDNA sequence electropherograms (Trosvik *et al.*, 2007). A model community consisting of the common gut bacteria *Escherichia coli*, *Lactobacillus salivarius* and *Bacteroides uniformis* was sampled hourly for a period of 28 h, and community composition was determined from the cell samples for each time point. This approach provided us with close interval time-series population data, and subsequent analysis ultimately produced a functional description of the dynamic structure of this simple model system. Furthermore, using online Fourier transform infrared (FT-IR) spectroscopy we were able to follow the entire growth process in terms of main metabolites. Such an analysis has, to our knowledge, not been carried out previously. The study integrates multivariate analysis of molecular genetic—and spectroscopic data with nonlinear statistical modeling of ecological data, highlighting the great potential for a synthesis between traditional ecology, molecular genetics and microbiology in order to advance our knowledge of fundamental population dynamic phenomena in microbial communities.

Materials and methods

Mixed bacterial populations

In order to obtain dynamic community growth data, three species of common gut bacteria were grown in an anaerobic fermentor system. The species used were type strains of *B. uniformis* (ATTC 8492), *E. coli* (ATTC 25922) and *L. salivarius* (DSMZ 20555, ATTC 11741). Prior to inoculation of the fermentor each strain was grown for 24 h in 10 ml tubes of anaerobe basal broth (Oxoid, Basingstoke, UK) under anaerobic conditions. For the initial inoculate, 2 ml of each of the three cultures was transferred to a clean tube, and the fermentor containing 1.5 l of sterile anaerobe basal broth was immediately inoculated.

The fermentor was a ~2 l glass chamber equipped with a flange designed to fit an accompanying glass lid with multiple connection ports. The lid was fitted tightly using a steel clamp. One port was connected to a gas tank containing a mixture of N_2 (80%) and CO_2 (20%). The chamber was continually flushed with this gas during the incubation period in order to obtain anaerobic conditions, as well as to achieve mixing of the culture. A second port was used as a gas outlet to avoid pressure buildup. A third port was opened only during collection of cell samples, which was carried out using 5 ml sterile pipettes. Two ports connected the chamber to the FT-IR measurement cell. The chamber was partially submerged in a stirred water bath at 37 °C in order to obtain normal ambient temperature of the species involved.

The first cell sample was taken 2 h after inoculation, and subsequent samples were collected every hour, for a total of 27 samples. pH values were also measured at the points of sample collection.

Optical density at 600 nm (OD600) was measured in all the fermentation samples using an Ultrospec 3000 UV/Visible spectrophotometer (Pharmacia Biotech, Chalfont St Giles, UK).

FT-IR spectrocopy

Throughout the process, we measured FT-IR spectra of the culture, the main objective being to acquire information on metabolites. We used a two chamber attenuated total reflectance cell fitted with an IR transparent optical crystal (ZnSe), and an Equinox 55 spectrometer (Bruker Optics, Ettlingen, Germany). The measurement system was connected to the growth chamber by a looped plastic tube, with growth medium being pumped through the cell and back into the chamber by a peristaltic pump. Spectra were collected automatically at approximately 12 min intervals.

We analyzed the spectral region corresponding to wavenumbers 3000–2800 cm^{-1}, in which vibrational modes characteristic of fatty acids can be observed (Socrates, 2001). Prior to analysis, the second derivative was calculated before extended multiplicative signal correction (Kohler *et al.*, 2005) was applied. The spectra were analyzed using the software package The Unscrambler (Camo Process, Oslo, Norway).

DNA extraction, PCR amplification and DNA sequencing

Isolation of total DNA was carried out using a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter Inc., Fullerton, CA, USA) and MagPrep Silica particles (Merck, Whitehouse Station, NJ, USA) according to an optimized automated protocol (Skanseng *et al.*, 2006).

PCR amplification of 16S rDNA was carried out using a universally conserved primer set (Nadkarni *et al.*, 2002). The PCR reaction mixture contained 1.25 U of AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA, USA), 200 μM dNTP mix, 1 μl of template DNA with 0.2 μM of each primer and 2.5 mM $MgCl_2$ in a total volume of 25 μl. The amplification profile was a 5 min activation step at 95 °C followed by 25 cycles of 95 °C for 15 s, 50 °C for 15 s and 72 °C for 1 min. The reaction was terminated with a 7 min elongation step at 72 °C.

Cyclic labeling and chain termination for sequencing was carried out using the BigDye v1.1 sequencing

chemistry (Applied Biosystems) and a nested sequencing primer, U515F (Baker *et al.*, 2003), according to instructions supplied by the manufacturer.

The three species involved in the experiment were identical with respect to both PCR and labeling primer sequences.

Excess dye and labeling primers were removed from the labeled DNA fragments using the Montage SEQ96 Sequencing Reaction Cleanup Kit (Millipore Corp., Billerica, MA, USA), and a Biomek 2000 Laboratory Automation Workstation, according to instructions supplied by the manufacturer. Sequencing was carried out using an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems).

Estimation of species composition in fermentation samples

For estimation of relative species abundances in cell samples from the fermentation, we used an approach of direct DNA sequencing and application of a set of partial least squares regression (PLSR) models to mixed sample sequence electropherograms for a 35 nucleotide region (positions 622–656 in *E. coli*) of the 16S rRNA gene (Trosvik *et al.*, 2007). In brief, the regression technique relies on a set of training spectra, **X**, for mixtures of gene fragments of the relevant genetic region, corresponding to a matrix of known mixture ratios, **Y**. **Y** is subsequently modeled as a function of **X** by PLSR, a type of factor analysis that uses the covariance structure between **X** and **Y** in order to find latent variables (PLS components) in **X**. The active use of **Y** in the analysis ensures that the extracted components are optimal for prediction of mixture ratios in new samples of unknown composition (Martens and Næs, 1989). Design, production and pre-processing of PLSR calibration data used in this work have been described previously (Trosvik *et al.*, 2007).

PLSR models, one for quantification of each species in the experiment, were computed with full cross-validation, producing three linear predictors of the form: $y_i = \beta_{i1}x_1 + \ldots + \beta_{i1800}x_{1800}$, where y_i is the proportion (approximate percentage) of either *B. uniformis*, *E. coli* or *L. salivarius*. $\beta_{i1}, \ldots, \beta_{i1800}$ are regression coefficients for prediction of y_i, and $x_1, \ldots, x_{1800}$ are the entries of the relevant spectral vector for a mixed sample. For all the models, the linear predictors were based on three PLS components (Figure 1). PLSR model statistics can be seen in Table 1.

Raw spectral data from the fermentation sample sequences, corresponding to the region used in the calibration were extracted from the raw data files and pre-processed as previously described (Trosvik *et al.*, 2007). The resulting data matrix of dimension (28×1800) was used as input for the PLSR models. Relative abundances of the three bacterial species were calculated for all 28 time points (the initial inoculate being time = 0) providing us with time-series data for the duration of the process.

The species composition in these samples was also checked by quantitative tRFLP analysis (Supplementary Figure S1 and Table S1).

All computations were carried out using the statistics programming interface R (R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria). PLSR was carried out using the R-package 'pls' using the orthogonal scores algorithm for model fitting.

Time-series modeling

The time-series data were analyzed using the R implementation of generalized additive models (GAM) in the mgcv library (Wood, 2000). This is a nonparametric modeling approach which does not presuppose any underlying functional form between response and predictor variables, allowing for effective modeling of nonlinear relationships. The technique fits smooth additive functions to each model covariate where the smooth terms are weighted sums of basis cubic spline functions. The weights are estimated according to the generalized cross-validation (GCV) criterion (Wood, 2000). The response variable is then approximated by the sum

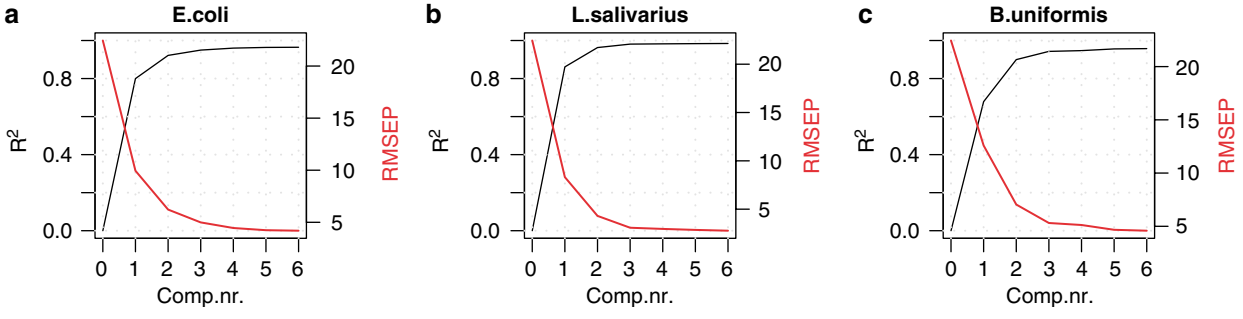

Figure 1 Panels a–c show convergence of partial least squares regression (PLSR) models on three PLS components. The figure shows root mean squared error of prediction (RMSEP, red lines) and squared Pearson's correlations (R^2, black lines) calculated from cross-validation of the three PLSR models. Correlations are between actual mixture ratios and proportions predicted in cross-validation. For each model the RMSEP decreases rapidly while R^2 increases in a similar fashion until they both stabilize upon the inclusion of the third PLS component.

Table 1 Summary statistics demonstrating the good predictive ability of PLSR models for quantification of species composition in fermentation samples

Species	*RMSEP*	*Explained validation variance*	*Explained calibration variance*	*Bias*	*Correlation*
Escherichia coli	4.99	94.9	96.5	−0.03	0.97
Lactobacillus salivarius	3.08	98.1	98.5	0.08	0.99
Bacteroides uniformis	5.31	94.3	95.9	0.04	0.97

Abbreviation: RMSEP, root mean squared error of prediction.
All statistics are based on models incorporating three PLS components (Figure 1). The columns are RMSEP (in percents), explained cross-validation *y* variance (in percents), explained calibration *y* variance (in percents), bias and correlations for each of the three PLSR models. RMSEP values were computed as $\sqrt{(\sum(y-\hat{y})^2/n)}$. Explained *y* variances for validation and calibration were computed as $1-\sum(y-\hat{y})^2/\sum(y-\text{mean}(y))^2$, where $\hat{y}$ are proportion estimates from cross-validation and calibration, respectively. Bias estimates were computed as $\sum(y-\hat{y})/n$, where $\hat{y}$ are proportions estimated from cross-validation. Correlations were computed as $\sum((\hat{y}-\text{mean}(\hat{y}))(y-\text{mean}(y)))/s_{\hat{y}}s_y$, where $s_{\hat{y}}$ and s_y are the standard deviations of predicted and measured values, respectively. For the statistics where it is of relevance, $n=87$.

of the smooth terms plus an intercept representing the response mean. For a systematic account of the GAM methodology we refer to Wood (2006).

For each smooth term, the dimension of the spline basis used is set prior to model fitting. In practice, this is equivalent to setting an upper limit to the degrees of freedom associated with the smooth term. This, in turn, determines the degree of curvature tolerated for the smooth in question. In order to avoid overfitting, we constrained the maximum degrees of freedom to three in the case of each model. The general model as applied to the case of *E. coli* growth is as follows:

$$\Delta E_t = b_e + f_e(E_t) + g_e(B_t) + h_e(L_t) + k_e(\text{pH}_t) + l_e(\text{OD600}_t) + \varepsilon_t$$

where ΔE_t is the per time unit changes in relative bacterial abundance computed as the difference in logarithmic abundance between two consecutive measurements (that is, $\log(E_{t+1})-\log(E_t)$). f_e, g_e, h_e, k_e and l_e are nonparametric smooth functions specifying the effects of *E. coli* (that is, density dependence), *B. uniformis, L. salivarius*, pH and OD600, respectively. By convention the smooth terms are of zero mean over the data. b_e is an intercept and ε_t is the noise term. For the analysis, all the independent variables, save pH, were log transformed. The same formulation was used to specify the models for relative abundance change rates for the other two bacterial species. Data for the initial inoculum were not included in the time-series analysis. The general model formulation above implies the absence of interactions between the covariates. We used a Lagrange multiplier test (Chan *et al.*, 2003) to validate this assumption in the case of each model equation. The test showed no significant interactions (Supplementary Table S2), suggesting that the general, fully additive model formulation used above is appropriate.

To select the optimal model structure of each GAM, we used the GCV criterion (Wood, 2000). This criterion provides a measure of the goodness-of-fit model, balancing prediction accuracy with model complexity, and the GAM structure that minimizes the GCV score is generally preferred. The model fitting algorithm provides approximate *P*-values indicating significance of model smooth terms. Final models were selected by successively removing nonsignificant terms from the full models, thereby minimizing the GCV scores (Supplementary Table S3).

Time-series simulation was carried out using the last observed values of the covariates (relative bacterial abundances and pH measured at 27 h) and predicting the next time-step using the fitted GAMs. For each step of the simulation, covariate values representing bacterial abundances were updated using model predictions from the previous time-step. Also, for each simulated step, errors were sampled randomly, with replacement, from the pertinent residual vectors from Equation (1). In order to maintain the contemporaneous correlation structure, sampled error terms for all three bacterial species always corresponded to the same observed time-step. Throughout the simulation, pH was kept relatively constant (around the final measured value of 5.41), but some variation was allowed by adding a random normal error term to the starting value, for each step, with $\mu=0$ and $\sigma=0.1$. The same simulation was also carried out using starting values for pH within a 1 U range of the final measured value.

Results

During the fermentation experiment, 27 cell samples were collected at hourly intervals (28 samples counting the starter culture). Relative abundances of the three species present in the 28 samples were estimated, providing us with time-series data for the duration of the process. Figure 2a shows the fluctuating population proportions from the starter culture to the final sample collected from the fermentation chamber. Figure 2b shows pH, OD600 and metabolic data measured for the 27 collected samples.

In order to describe the successive dynamic structure of the microbial system under study, we fitted a GAM to each population time series. The

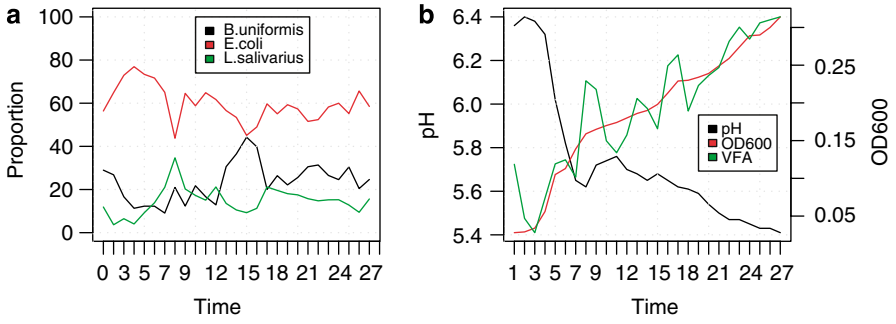

Figure 2 Time-series data collected from the fermentation. (**a**) Proportions (in percents) of the three bacterial species in the initial inoculum (time = 0) and the 27 fermentation samples. (**b**) pH, optical density at 600 nm (OD600) and metabolic data (volatile fatty acids, VFA) for the 27 sampling points. The curve-labeled VFA are scores along the first component of a partial least squares regression (PLSR) model with process references (response variables) and Fourier transform infrared (FT-IR) absorption measurements (explanatory variables), plotted against time. These scores are indicative of changing ratios of methylene/methyl content of metabolites in the culture (see text below and Supplementary Figures S3 and S4).

following models (Equation (1)) were selected from the general class of full additive models according to the GCV criterion (Supplementary Table S3):

$$\begin{aligned}\Delta E_t &= b_e + f_e(E_t) + g_e(B_t) + k_e(\mathrm{pH}_t) + \varepsilon_{et}\\ \Delta L_t &= b_l + g_l(B_t) + h_l(L_t) + k_l(\mathrm{pH}_t) + \varepsilon_{lt}\\ \Delta B_t &= b_b + g_b(B_b) + f_b(E_t) + \varepsilon_{bt}\end{aligned} \qquad (1)$$

Model diagnostics showed that the residuals were white (Supplementary Figure S2) and approximately normally distributed. We assessed the goodness-of-fit of the models by checking the additivity assumption (Chan *et al.*, 2003) for each equation, and there was no evidence of nonadditivity (Supplementary Table S2). Hence, the model provides a good fit to the data. The residuals were, however, contemporaneously, negatively correlated (Supplementary Table S4), that is, community competition is reflected in both the conditional mean and covariance structure.

Descriptive statistics for the three selected models are presented in Table 2, demonstrating highly significant effects of the smooth terms and major proportions of explained variance in the case of each model. Figure 3 summarizes the forms of the functional relationships between the relative proportion change rates and the independent covariates (log-relative abundances or pH) for the three species. The models indicate some interspecies interactions in the system. In the case of *L. salivarius* there is an apparent negative linear relation to increasing proportions of *B. uniformis* (Figure 3d), while *E. coli* and *B. uniformis* have a reciprocal and explicitly nonlinear interaction (Figures 3a and h). All three species display a density-dependent effect of some sort. For *L. salivarius* this effect is near-linear (Figure 3e), whereas for *E. coli* and *B. uniformis* the corresponding negative effects level off at high relative abundances (Figures 3b and g).

E. coli, not unexpectedly, has a linear positive relationship with pH within the experimental range (Figure 3c), increasing toward neutral which is optimal for the growth of this species. *L. salivarius*, being well adapted to an acidic environment, shows the opposite trend with a slightly curved positive relationship corresponding to declining pH, appearing to stabilize around 5.6 (Figure 3f). No significant effect of changing pH was found in the case of *B. uniformis*, indicating robustness to this environmental perturbance. OD600, representing total bacterial density, was not found to have a significant effect in any of the models.

Table 2 Summary statistics for the three selected GAMs (Equation (1) stating the degree of linearity (edf, see below) and significance of smooth terms

	E.d.f.	*F*	P-*value*
Escherichia coli growth			
$f_e(E_t)$	2.12	7.65	<0.01
$g_e(B_t)$	2.28	5.30	<0.01
$k_e(\mathrm{pH}_t)$	1.00	7.30	0.014
$R^2_{adj} = 0.65$			$n = 26$
Lactobacillus salivarius growth			
$h_l(L_t)$	1.99	18.17	<0.01
$g_l(B_t)$	1.00	10.52	<0.01
$k_l(\mathrm{pH}_t)$	2.09	12.95	<0.01
$R^2_{adj} = 0.69$			$n = 26$
Bacteroides uniformis growth			
$g_b(B_b)$	1.97	6.02	<0.01
$f_b(E_t)$	2.27	5.53	<0.01
$R^2_{adj} = 0.50$			$n = 26$

Abbreviations: adj, adjusted; Edf, estimated degrees of freedom.
The first column shows the covariates terms included in the models, i.e. the smooth terms on the right-hand side of Equation (1). E_t is the log-relative abundance of *E. coli* at time *t*. B_t and L_t are the corresponding abundances of *B. uniformis* and *L. salivarius*, respectively. pH_t is the pH measured at time *t*. *f, g, h* and *k* are smooth functions as they apply in each individual case. The column labeled 'edf' contains the smooth functions. They are interpreted as measures of linearity of the smooth functions with 1 being a linear effect and >1 increasingly stronger nonlinear effects. The rightmost column shows the approximate significance of each smooth term. These are based on F-tests using the statistics tabulated in the F column, with degrees of freedom being the estimated rank of each individual smooth term and residual degrees of freedom of each model.

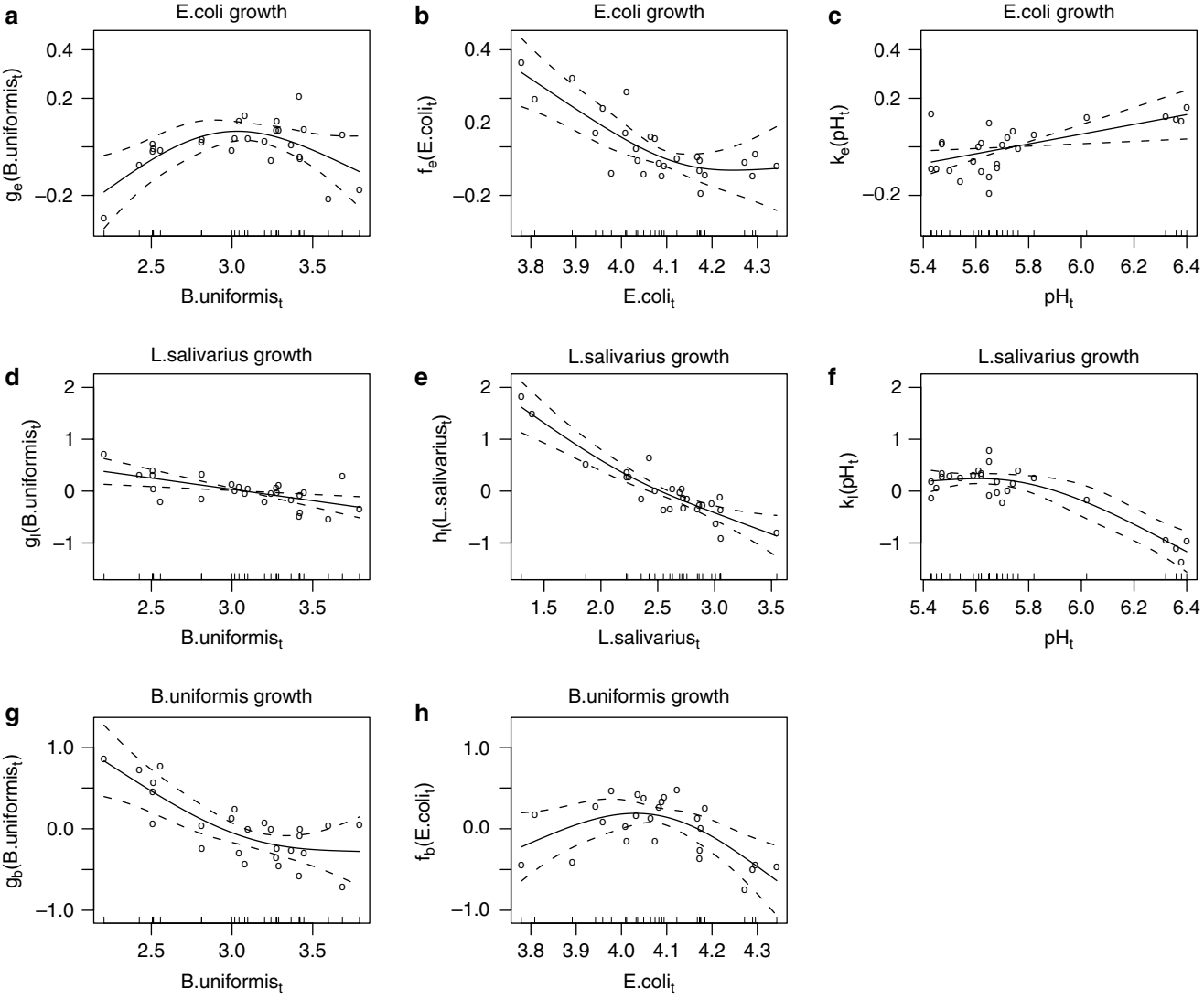

Figure 3 Panels a–h are plots of smooth terms from the generalized additive models (Equation (1)). These summarize the functional relationships between response variables and covariate terms for each model equation. The abscissae give the values of the covariate terms (log-relative abundances and pH). The ordinate axes indicate the partial additive effects of the covariate terms on the response variables (log proportion$_{t+1}$−log proportion$_t$). The predicted values of the responses are given as the sum of the smooth terms plus an intercept term. The broken lines are 95% confidence bands of the fitted curves.

The main metabolites expected in this fermentation are acetate, lactate, propionate and succinate. We thus focused on the FT-IR spectral region corresponding to fatty acid signatures (3000–2800 cm^{-1}). A PLSR model was computed in order to explore correlations between spectral wavenumbers and the process reference variables (bacterial relative abundances, time, pH and OD600). The first model component was found to contain the bulk of relevant information (Figures 2b, Supplementary Figures S4 and S5), explaining 48% and 49% of **X** and **Y** variance, respectively. Along this component, time and OD600 correlate strongly with wavenumbers around 2925 and 2850 cm^{-1} (Supplementary Figure S4). These wavenumbers correspond to the antisymmetric and symmetric C-H stretch modes of methylene groups (Supplementary Figure S5). On the opposite side of the first component axis, pH correlates with wavenumbers around 2960 and 2870 (Supplementary Figure S4). These regions are characteristic of the antisymmetric and symmetric C-H stretch modes of methyl groups (Supplementary Figure S5). The first model component also separates *B. uniformis* and *L. salivarius* (correlating with time, OD600 and methylene vibrations) from *E. coli* (correlating with pH and methyl vibrations). In general, methylene groups are associated with longer fatty acids (the groups being internal in molecules), while methyl groups are associated with shorter chains (being terminal in molecules). Our data clearly indicate a gradual shift toward fatty acids with more methylene groups throughout the

fermentation (Figure 2b). This shift is also concurrent with increasing acidity and optical density in the batch culture.

Discussion

This study of a simple model system demonstrates how bacteria living in communities are affected by each other and by the physical environment surrounding them. While the interaction between *L. salivarius* and *B. uniformis* indicates a scenario of direct competition in terms of the effect of *Bacteroides* on *Lactobacillus* (Figure 3d), the converse effect is not observed. One likely explanation for this is cross-feeding between the two species. While both consume glucose (the primary nutrient of the growth medium), and compete for this nutrient, *Bacteroides* may also ferment lactate (Macy *et al.*, 1978; Schultz and Breznak, 1979), the main metabolic by-product of *L. salivarius*' metabolism. For this nutrient, there would be no competition between the two species, rendering the competitive relationship unreciprocal. This may also explain, to some extent, the fermentation pattern observed in FT-IR spectroscopy, where both *B. uniformis* and *L. salivarius* correlate with progressive formation of methylene-containing fatty acids. In our system these would primarily be propionate and succinate produced by *Bacteroides* fermenting both glucose and lactate.

The interactions between *B. uniformis* and *E. coli* are more complex. In both cases, the smooth functions have a parabolic shape where the covariate species appears to have a positive effect on the response species up to a point where the relationship turns negative (Figures 3a and h). One positive feedback effect of *E. coli* might have on *B. uniformis* growth is that of removing residual oxygen from the ambient growth medium (Hagen *et al.*, 1982). At low *E. coli* abundances this might be beneficial to the strictly anaerobic *Bacteroides* species, but at a certain level of *E. coli* abundance the relationship may shift to one of direct competition, outweighing the benefits of oxygen removal and thus having a negative effect on *B. uniformis* growth. The reverse case is similar, but with a more pronounced positive relationship, leveling off and turning negative at some point of *Bacteroides* abundance. This is likely a result of *Bacteroides'* metabolism feeding back on *E. coli* growth. Excessive amounts of lactate may be toxic to many bacteria, and lactic acid bacteria are known to inhibit growth of *E. coli* (Tomas *et al.*, 2003). *Bacteroides* can, to some extent, remove this deleterious metabolite, fermenting it mainly to propionate, and thus relieve the inhibition. This scenario is also supported by the FT-IR spectroscopic data where *Bacteroides* shows the highest degree of correlation with the transition from nonmethylene (acetate and lactate) to methylene-containing (propionate and succinate) acids. In our system, however, the commensal relationship shifts at some point when direct competition for nutrients becomes the dominant phenomenon.

Altogether, species interactions appear to be an important factor in our system, and metabolic commensalism of the kind observed is most likely a general characteristic of bacterial communities (Bradshaw *et al.*, 1994). Through variable selection according to the stated criteria, we were able to identify, as well as to quantify, significant interactions taking place between different bacteria, and between bacteria and the environment. The selected model formulation (Equation (1)) benefits from the simplifying assumption of additivity (that is, within each model equation the effect of a given covariate is not contingent on changing values of another covariate). This assumption is often rather strong, but it can greatly simplify both model fitting and interpretation. In this particular case, the additive formulation used in Equation (1) was found to be appropriate (Supplementary Table S2). As such, we were able to characterize the isolated effect of each covariate, providing a clear picture of the system's dynamics.

By applying statistical methods traditionally used by animal ecologists (for example, Ciannelli *et al.*, 2005), we estimated the underlying density-dependent effects of all three bacterial species. In the case of *L. salivarius*, this effect is indicative of a simple metabolic feedback mechanism (Figure 3e). As for *E. coli* and *B. uniformis*, the density-dependent effects are nonlinear, leveling off at high proportions (Figures 3b and g). These nonlinearities may be caused by interaction between the two species. Since direct competition for nutrient substrates does not take place before some critical abundance is reached, the negative density dependence becomes the dominant feedback mechanism. Once the point is reached where competition has significant effect on the growth of the two species, the interspecies struggle for nutrients may dissipate the density-dependent effects.

It should be noted that since the population data presented in this study are from a batch culture experiment, our model (Equation (1)) does not necessarily reflect steady-state dynamics. The fact that no new resources were added to the culture during the period of data collection will most likely have had an effect on the level of competition between the bacteria. Thus, ascertaining the degree to which our model is influenced by transitional phase dynamics is difficult. However, when we simulated 1000 h of continued growth with the dynamic model estimated for our system of bacteria (given a growth system with continuous supply of nutrients and dilution), using Equation (1), the process remained quite stable throughout the simulation (Supplementary Figure S3). In the simulation, pH was represented by a constant term plus independent normal errors with zero mean and standard deviation of 0.1. Constant term values for

pH within a 1 U range of the final measured value of 5.41 were tried, producing similar results. Thus, for a relatively stable pH, the simulation indicates that the fitted model (Equation (1)) may reflect a steady state.

It is hardly surprising that bacterial community members compete for growth limiting nutrients. The functional forms of such dynamic relationships, however, have not been thoroughly explored. Infections of, for example, the gastrointestinal system involve complex interactions among the pathogenic species, the resident microflora and the environment (the host), about which relatively little is currently known (Hooper and Gordon, 2001; Eckburg *et al.*, 2005; Nicholson *et al.*, 2005). The long-term dynamics (that is, stability, cyclicity or chaos) of pathogens and commensals will largely be determined by the density-dependent structure of the system. As shown here, density dependence can be heavily influenced by interspecies interactions. Such knowledge is important when it comes to treating and preventing the colonization of human and animal pathogens, and further experimentation will be needed in order to learn more about long-term infection dynamics. Even though *in vitro* systems, in most cases, are far from being representative of natural scenarios, such simplifications of the natural world may serve to increase our understanding of fundamental dynamic properties of microbial consortia, perhaps enabling us to predict the behavior of, and ultimately to manipulate such systems.

References

Baker GC, Smith JJ, Cowan DA. (2003). Review and reanalysis of domain-specific 16S primers. *J Microbiol Methods* **55**: 541–555.

Balagadde FK, You L, Hansen CL, Arnold FH, Quake SR. (2005). Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science* **309**: 137–140.

Bell T, Newman JA, Silverman BW, Turner SL, Lilley AK. (2005). The contribution of species richness and composition to bacterial services. *Nature* **436**: 1157–1160.

Bradshaw DJ, Homer KA, Marsh PD, Beighton D. (1994). Metabolic cooperation in oral microbial communities during growth on mucin. *Microbiology* **140**: 3407–3412.

Chan K, Kristoffersen A, Stenseth N. (2003). Burmann expansion and test for additivity. *Biometrika* **90**: 209–222.

Ciannelli L, Bailey K, Chan K, Belgrano A, Stenseth N. (2005). Climate change causing phase transitions of walleye pollock (*Theragra chalcogramma*) recruitment dynamics. *Proc R Soc Lond B* **272**: 1735–1743.

Cooper VS, Lenski RE. (2000). The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* **407**: 736–739.

Dykhuizen D, Davies M. (1980). An experimental model: bacterial specialists and generalists competing in chemostats. *Ecology* **61**: 1213–1227.

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M *et al.* (2005). Diversity of the human intestinal microbial flora. *Science* **308**: 1635–1638.

Gause GF. (1932). Experimental studies on the struggle for existence: I. Mixed population of two species of yeast. *J Exp Biol* **9**: 389–402.

Gause GF. (1935). Experimental demonstration of Volterra's periodic oscillations in the number of animals. *J Exp Biol* **12**: 44–48.

Gorden RW, Beyers RJ, Odum EP, Eagon RG. (1969). Studies of a simple laboratory microecosystem: bacterial activities in a heterotrophic succession. *Ecology* **50**: 86–100.

Hagen JC, Wood WS, Hashimoto T. (1982). *In vitro* stimulation of *Bacteroides fragilis* growth by *Escherichia coli*. *Eur J Clin Microbiol* **1**: 338–343.

Hooper LV, Gordon JI. (2001). Commensal host-bacterial relationships in the gut. *Science* **292**: 1115–1118.

Jessup CM, Kassen R, Forde SE, Kerr B, Buckling A, Rainey PB *et al.* (2004). Big questions, small worlds: microbial model systems in ecology. *Trends Ecol Evol* **19**: 189–197.

Kohler A, Kirschner C, Oust A, Martens H. (2005). Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform infrared microscopy images of cryo-sections of beef loin. *Appl Spectrosc* **59**: 707–716.

Macy JM, Ljungdahl LG, Gottschalk G. (1978). Pathway of succinate and propionate formation in *Bacteroides fragilis*. *J Bacteriol* **134**: 84–91.

Martens H, Næs T. (1989). *Multivariate Calibration*. Wiley: Chichester.

McArthur JV. (2006). *Microbial Ecology: An Evolutionary Approach*. Elsevier: Amsterdam, vol. 12, pp 416s.

Moe S, Kristoffersen A, Smith R, Stenseth N. (2005). From patterns to processes and back: analysing density-dependent responses to an abiotic stressor by statistical and mechanistic modelling. *Proc R Soc Lond B* **272**: 2133–2142.

Nadkarni MA, Martin FE, Jacques NA, Hunter N. (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* **148**: 257–266.

Nicholson JK, Holmes E, Wilson ID. (2005). Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol* **3**: 431–438.

Paerl HW, Steppe TF. (2003). Scaling up: the next challenge in environmental microbiology. *Environ Microbiol* **5**: 1025–1038.

Rainey PB, Rainey K. (2003). Evolution of cooperation and conflict in experimental bacterial populations. *Nature* **425**: 72–74.

Read AF, Taylor LH. (2001). The ecology of genetically diverse infections. *Science* **292**: 1099–1102.

Schultz JE, Breznak JA. (1979). Cross-feeding of lactate between *Streptococcus-lactis* and *Bacteroides* Sp isolated from termite hindguts. *Appl and Environ Microbiol* **37**: 1206–1210.

Skanseng B, Kaldhusdal M, Rudi K. (2006). Comparison of chicken gut colonisation by the pathogens *Campylobacter jejuni* and *Clostridium perfringens* by real-time quantitative PCR. *Mol Cell Probes* **20**: 269–279.

Socrates G. (2001). *Infrared and Raman Characteristic Group Frequencies: Tables and Charts*, 3rd edn. Wiley: Chichester, vol. 15, pp 347s.

Stenseth NC, Falck W, Chan KS, Bjornstad ON, O'Donoghue M, Tong H *et al.* (1998). From patterns to processes: phase and density dependencies in the

Canadian lynx cycle. *Proc Natl Acad Sci USA* **95**: 15430–15435.

Tomas MSJ, Ocana VS, Wiese B, Nader-Macias ME. (2003). Growth and lactic acid production by vaginal *Lactobacillus acidophilus* CRL 1259, and inhibition of uropathogenic *Escherichia coli*. *J Med Microbiol* **52**: 1117–1124.

Trosvik P, Skanseng B, Jakobsen KS, Stenseth NC, Naes T, Rudi K. (2007). Multivariate analysis of complex DNA sequence electropherograms for high-throughput quantitative analysis of mixed microbial populations. *Appl Environ Microbiol* **73**: 4975–4983.

Turchin P. (1995). Population regulation: old arguments and a new synthesis. In: Price PW, Cappuccino N (eds). *Population Dynamics: New Approaches and Synthesis*. Academic Press: San Diego. pp 19–37.

Wood S. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *J R Stat Soc B* **62**: 413–428.

Wood SN. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC: Boca Raton, vol. 17, pp 391s.

You L, Cox III RS, Weiss R, Arnold FH. (2004). Programmed population control by cell–cell communication and regulated killing. *Nature* **428**: 868–871.

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