Analysis of treatment effects on the microbial ecology of the human intestine

Anna L. Engelbrektson¹, Joshua R. Korzenik², Mary Ellen Sanders³, Brian G. Clement⁴, Gregory Leyer⁵, Todd R. Klaenhammer⁶ & Christopher L. Kitts¹

¹Environmental Biotechnology Institute, California Polytechnic State University, San Luis Obispo, CA, USA; ²Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; ³Dairy and Food Culture Technologies, Centennial, CO, USA; ⁴Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA; ⁵Danisco USA, Madison, WI, USA; and ⁶North Carolina State University, Raleigh, NC, USA

Abstract

A large number of studies have investigated gastrointestinal microbiota and changes in the gastrointestinal community. However, a concern in these studies is how best to assess changes in gastrointestinal community structure. This paper presents two different human trials where the fecal terminal restriction fragment length polymorphism data sets were analyzed to search for treatment effects. Principle components analysis and cluster analysis based on grouped data are compared with analysis of data by subject using distance coefficients. Comparison with baseline within an individual before grouping by treatment provided a clearer indication of treatment effects than did an evaluation of data grouped before analysis. In addition, a large within-subject sample size and multiple baseline samples are necessary to accurately analyze treatment effects.

Keywords
terminal restriction fragment length polymorphism; probiotics; antibiotics; prebiotics; gastrointestinal microbiota.

Introduction

The gut is the largest mucosal surface in the body and the microbial community present can have a large effect on immune function. This community consists of a large number of different species, which differ between the mucosa and the feces (Zoetendal et al., 2002). However, most human studies utilize fecal samples because of their accessibility. The role the intestinal microbiota plays in the protection against infection, especially by potentially pathogenic microorganisms, is well documented (Fuller, 1991).

Antibiotics are known to disrupt the normal intestinal microbiota. They can cause disturbance in normal bowel function, disruption of mucosal integrity, and symptoms including diarrhea, bloating, flatulence, and intestinal pain. Antibiotic therapy can also result in Clostridium difficile colitis, which can cause severe symptoms (Bergogne-Berezin, 2000). Other treatments that affect the gastrointestinal microbiota include the ingestion of probiotics and prebiotics. Prebiotics are defined as live microorganisms administered in adequate amounts to confer a beneficial health effect on the host (FAO/WHO, 2001). Prebiotics are nondigestible substances that when consumed provide a beneficial physiological effect on the host by selectively stimulating the favorable growth or activity of a limited number of indigenous bacteria (Gibson & Roberfroid, 1995). Probiotic treatment can facilitate gut normalization after antibiotic treatment and probiotic bacteria can produce a variety of health benefits including reduction in the intensity and duration of diarrheal illness, improvement in immune system function, alleviation of lactose intolerance, and anticarcinogenic effects (Sanders, 1999; Gill & Guarner, 2004). Probiotics, primarily Lactobacillus and Bifidobacterium species, are found in many dairy foods and supplements. However, when fed to healthy subjects, probiotics only modestly affect fecal microbiota and the strains used do not tend to be permanent colonizers (Chen et al., 1999; Tannock et al., 2000).

A large number of studies have been performed investigating fecal communities and changes in fecal communities. However, it is difficult to assess changes in fecal community structure. A number of different methods have been used: standard plating on selective media (Sullivan et al., 2003; Madden et al., 2005) and different rRNA gene-based
molecular methods including the use of group and species specific primers or probes (Kok et al., 1996; Song et al., 2000; Walter et al., 2001; Matsuki et al., 2002; Silvi et al., 2003), denaturing gradient or temperature gradient gel electrophoresis (DGGE or TGGE) analysis (Simpson et al., 2000; Satokari et al., 2001; Heilig et al., 2002), cloning and sequencing (Wilson & Bitchintong, 1996; Leser et al., 2002), and terminal restriction fragment length polymorphism (TRFLP) analysis (Kaplan et al., 2001; Sakata et al., 2005; Fernberg et al., 2005). PCR methods are generally thought to be superior to culture methods in that many fecal species are not culturable using standard techniques. Cloning can provide very detailed phylogenetic information, but is costly and labor intensive and is thus not realistic for processing large numbers of samples. DGGE and TRFLP are the best methods for rapid high throughput comparison of bacterial communities. However, DGGE primarily provides presence/absence information and is not easily digitized (Forney et al., 2004). TRFLP data are good for characterization of bacterial communities because they can be used to determine both species dominance and species richness within samples and they are automatically digitized. Different phylotypes of bacteria present in each sample can also be tentatively identified by comparison with a database (Kitts, 2001).

TRFLP data are generally analyzed using multivariate statistical techniques such as canonical correspondence analysis (Ayala-del-Rio et al., 2004), principle components analysis (Clement et al., 1998; Kaplan & Kitts, 2004) or cluster analysis (Urukawa et al., 2000; Hiraishi et al., 2000; Kitts, 2001). Distance or similarity coefficients are implicit in all these methods for analysis of TRFLP data. Euclidean distance is commonly used and is an implicit part of principle components analysis, but is not appropriate for data with a large number of zeros (Rees et al., 2004). Jaccard, Dice, Sorensen's and Simpson's coefficients are appropriate for presence/absence analysis but ignore species abundance. Bray–Curtis similarity is superior to other coefficients for analysis of TRFLP data because it can better handle the large amounts of zeros present in TRFLP data and has high statistical power and robustness with species abundance data (Faith et al., 1991; Rees et al., 2004). Culture data can also be treated as multivariate data and Euclidean distance is then an appropriate coefficient of similarity.

Any statistical analysis of fecal community structure needs to take into account that human subjects tend to have individually unique intestinal communities which may or may not be stable over time (Zoetendal et al., 1998; Vaughan et al., 2000; Vanhoutte et al., 2004). This makes analysis of the data difficult because, traditionally, data from multiple subjects are analyzed as a group using either univariate statistics with culture data or principle components analysis/cluster analysis with multivariate (i.e. TRFLP) data. Grouping of fecal community data before analysis can result in loss of statistical significance or false negative results. The approach outlined in this study involves the use of Bray–Curtis similarity to measure an individual’s divergence from baseline after treatment. This can overcome the problem of subject-to-subject variability because subjects are analyzed individually before treatment group differences are assessed. This paper presents two different human trials where the fecal TRFLP data sets were analyzed to search for treatment effects. Analyses based on grouped data are compared with individual data. Culture data from one of the studies were also analyzed.

Methods

Probiotic–prebiotic study design

This randomized, double blind, placebo-controlled, parallel group dietary study comprised two independent blocks of 32 healthy subjects each. The delivery device was 200 mL of pasteurized whole yogurt, consumed twice daily, at least 8 h apart. The probiotics consisted of Lactobacillus rhamnosus strain 271 (Probi AB, Lund, Sweden), Lactobacillus acidophilus strain NCFM (Danisco, Madison, WI), Lactobacillus paracasei ssp. paracasei strain DN114001 (Danone, Paris, France), and Bifidobacterium sp. strain DN B10 175010 (Le Plessis, Robinson, France). The prebiotic was Frutafit®, a type of inulin. All subjects were given control yogurt for the first 3 weeks, and were then randomly assigned to four groups and allocated different experimental yogurts. Group 1 was fed a yogurt with added probiotic cultures [10^9–10^10 colony-forming units (CFU) of each culture per mL]. Group 2 was fed a yogurt with both added probiotic cultures (10^9–10^6 CFU of each culture per mL) and the prebiotic (5% weight in volume, w/v). Group 3 was fed a yogurt supplemented with only the prebiotic (5% w/v). Group 4, the control, was kept on the control yogurt with no additives (all yogurt products were manufactured by Leatherhead Food RA, Surrey, UK). At the end of week 9, all groups were then given the control yogurt for a further 3 weeks. Fecal samples were collected into sterile bags (weeks 3, 9, and 12) and were immediately placed on ice and stored at −20 °C until analysis. By the end of the study, 45 people had complete sample sets from all three sampling times and were used to assess the effects of probiotics and prebiotics on fecal bacterial communities.

Probiotic–antibiotic study design

Healthy individuals were recruited who agreed to a 1-week antibiotic treatment for study purposes only, receiving the broad-spectrum (Gram-positive and Gram-negative) antibiotic Augmentin™ (GlaxoSmithKline, Brentford, London), a mixture of amoxicillin and clavulanic acid, 875 mg
orally twice a day. This antibiotic was selected because of a high rate of antibiotic-associated diarrhea. Patients were then randomly assigned (1:1) to either the placebo or the probiotic test product, which consisted of a capsule containing a dried bacterial preparation of probiotic bacteria in the genera Lactobacillus and Bifidobacterium. The following strains and amounts were fed to individuals in the probiotic group: Bifidobacterium bifidum Bb-02 (5 x 10⁸), Bifidobacterium lactis Bi-04 (5 x 10⁸), B. lactis Bi-07 (5 x 10⁸), Lactobacillus acidophilus NCFM (5 x 10⁸), and Lactobacillus paracasei Lpc-37 (5 x 10⁸) (Danisco). The total dose of probiotic was 2 x 10¹⁰ bid (4 x 10⁹ daily). The other group received a placebo consisting of the same filler used in the bacterial preparation, maltodextran, without the bacteria. The study was conducted over 48 days. Three baseline (no treatment) fecal samples were obtained on days 1, 7, and 14, followed by the 7-day course of Augmentin™. Fecal samples were then collected on days 21, 25, 34, and 48. Probiotic or placebo treatment began on day 14 and continued until day 34.

Forty subjects were recruited with enrollment criteria permitting only patients over 18 years old without significant acute or chronic illnesses. Permitted medications included those that were constant throughout the study if they had no established or suspected impact on gut microbiota. Individuals were excluded if they were pregnant, were breastfeeding, had a penicillin allergy, a history of gastrointestinal illness or had been on any antibiotics in the preceding 4 weeks. Fermented foods or any probiotic preparations were prohibited for 4 weeks before entry into the study and throughout the duration of the study.

Fecal samples were obtained for TRF analysis by adding approximately 1 g of feces to a 2 mL screw-cap tube and freezing at −80 °C until shipment. For culture analyses, 5 g of each fecal sample were placed into 16 mL Cary Blair Transport Medium (Difco, Franklin Lakes, NJ) with indicator (Remel, Lenexa, KS) resulting in a 1:4.2 dilution factor. The sample was then shaken/vortexed briefly to disperse it and frozen at −80 °C until shipment.

**Bacterial culturing (probiotic–antibiotic study)**

Frozen samples received from the clinical study were stored at −80 °C until enumeration, and thawed at 37 °C immediately prior to plating. Duplicate serial dilutions of the samples were prepared (10⁻²–10⁻⁸) in sterile, prereduced 1% yeast extract. In accordance with Summanen et al. (1993), liquid media were boiled for 5 min to drive off dissolved oxygen and used within the same day. Yeast extract diluent was autoclaved for 15 min at 121 °C and allowed to cool prior to being placed in the anaerobic chamber and dispensed into sterile tubes. All fecal samples were thawed, diluted, and plated in an anaerobic chamber (Coy Laborato-

tory Products, Grass Lake, MI), maintained at 37 °C for 3–5 days in an atmosphere of 85% nitrogen, 10% hydrogen, and 5% carbon dioxide.

Dilutions were plated onto duplicate plates. Bifidobacter-
teria were enumerated using BIM-25 (Muñoz & Pares, 1988; Lapiere et al., 1992; Rada et al., 1999), which is a reinforced clostridial agar base containing polymixin B, tetrazolium red, isoacetate, kanamycin, and naladixic acid (Sigma-
Aldrich, St Louis, MO). Lactobacilli were enumerated using LBS Agar (Difco) plus 200 mL L⁻¹ tomato juice from concentrate (Campbell Soup, Camden, NJ) (Rogosa et al., 1951; Sabine & Vaselekos, 1965). Organisms in the Bacteroides fragilis group were enumerated with Bacteroides Bile Escolin Agar (BBE; Difco). Clostridium species were enumerated with Egg Yolk Agar (EYA; Difco). For selection of Clostri-
dium species, aliquots of each dilution were treated at 80 °C for 10 min to kill vegetative cells, leaving spore-formers for enumeration as described by Summanen et al. (1993). Enterobacteriaceae were enumerated with MacConkey Agar (MAC; Difco).

**Creation and normalization of TRFLP data**

In the probiotic–prebiotic study, DNA was isolated according to Clement & Kitts (2000). 16S rRNA gene TRFLP patterns were created and data were normalized following the protocol in Kaplan et al. (2001) with TaqGold® (Applied Biosystems, Foster City, CA) as the polymerase and DpnII (New England Biolabs, Beverly, MA) as the digesting enzyme.

In the probiotic–antibiotic study, samples were extracted in triplicate using the MoBio UltraClean® soil DNA kit (MoBio Laboratories, Carlsbad, CA) following manufacturer’s protocol with the addition of five extra washes with S4. The success of each extraction was determined by measuring the DNA concentration in the extraction product with a Spectramax® spectrophotometer (Molecular Devices, Palo Alto, CA).

Polymerase chain reaction was performed using primers homologous to conserved regions on the bacterial 16S rRNA gene. The reverse primer 536-K2R (5’-GTA TTA CGG CTG CTG G-3’) and the forward primer 46-Ba2F (5’-GCA TTA CAC ATG CAA GTC GA-3’), which was fluorescently labeled with a phosphamide dye (D4; GenSet, La Hoya, CA), were used for each reaction. Reactions of 50 μL were carried out using 1 μL of undiluted extraction product, 5 μL of 10 × Buffer, 3 μL of 10 mM dNTP, 2 μL of 20 g mL⁻¹ BSA, 7 μL of 25 mM MgCl₂, 1 μL of each primer, and 0.3 μL of 5 μL⁻¹ TaqGold® (Applied Biosystems). Reaction temperatures and times were 96 °C for 10 min; 35 cycles of 94 °C for 1 min, 46.5 °C for 1 min, 72 °C for 2 min; and 72 °C for 10 min. All reactions were performed in triplicate and then combined using a MoBio UltraClean® PCR Cleanup Kit.
(MoBio Laboratories) following the manufacturer’s protocol. PCR products were quantified using a fluorometer tuned to the labeling dye.

An enzyme digest was performed on 75 ng of cleaned PCR product using the restriction endonuclease HaeIII (New England Biolabs). Each 40 µL digestion used 75 ng of DNA, 1 U of enzyme, and 4 µL of buffer. The samples were digested for 4 h at 37 °C and inactivated for 20 min at 65 °C. The digestion products were ethanol precipitated and resuspended in 20 µL of formamide and 0.25 µL of CEQ 600 base pair standard. TRFLP profiles were obtained using a Beckman Coulter CEQ8000 DNA analysis system.

Terminal restriction fragment length in nucleotides, and TRF peak area were exported from the CEQ8000 into EXCEL (Microsoft, Seattle, WA). To standardize the data for comparison between samples, the area under each TRF peak was normalized to the total amount of DNA analyzed and expressed as parts per million (ppm). Peaks with an area of less than 5000 ppm (< 0.5% of the total for that sample) were excluded from analysis to reduce noise.

Analysis of TRFLP data

For both studies, normalized TRFLP data sets were transformed by taking the square root of the area under each TRF peak to de-emphasize large TRF peaks while still taking relative abundance into account (Blackwood et al., 2003). Transformed data were analyzed by principal components analysis, cluster analysis with Euclidean distance

\[
\text{similarity} = \frac{100 \times [1 - \frac{1}{\sqrt{\sum_{i,j=60}^{600} \text{Area}_{i,j}^2}}]}{\text{Area}_{i,j}^2 + \text{Area}_{i,j}^d}.
\]

All statistics were performed using Minitab 14 (Minitab Inc., State College, PA) and EXCEL.

Results and discussion

Some recent studies have relied on a simple visual comparison of raw molecular data to show treatment effects (Tannock et al., 2000; Hayashi et al., 2002; Sakamoto et al., 2003). Unfortunately, data analyzed in this way could be quite misleading and may overlook subtle differences in community structure. The large number of variables generated by molecular profile methods makes the use of multivariate statistics essential. Generally, pairwise similarity or distance measures are used and are implicit in both principle components analysis and cluster analysis. Molecular profile data can be analyzed as presence-absence data (Simpson et al., 2000; Donskey et al., 2003; Sakata et al., 2005) or relative abundance data (Kaplan et al., 2001; Wang et al., 2004; Jernberg et al., 2005), although when only presence-absence data are analyzed, information that may be critical to assess treatment effect is lost. Therefore, analyses of treatment effects within individuals have generally involved multivariate data reduction methods such as principle components analysis or cluster analysis (Zoetendal et al., 2002; Wang et al., 2004; Jernberg et al., 2005). Both these methods were tested using data from two studies of treatment effects on human fecal microbiota.

Analysis of the probiotic–prebiotic study TRFLP data

The goal of the probiotic–prebiotic study was to describe the changes in human fecal microbial communities caused by the ingestion of probiotic bacteria and/or prebiotic. The initial analysis followed a standard approach for TRFLP data, where principle components analysis score plots were created for each treatment group (Fig. 1). The samples clearly did not cluster by week, indicating a large amount of subject-to-subject variability. Between-subject variation in TRF patterns was clearly greater than within-subject variability, as indicated by an obvious grouping by subject.

The problem of subject-to-subject variation is compounded with molecular profile data compared to culture data because many more variables are introduced. In some cases, significant results can be obtained when grouping individual data. For example, Sakata et al. (2005) evaluated the effect of infant breast-feeding grouping all subjects together, but the high sample number in the study allowed for some significant results. Both Simpson et al. (2000) and Wang et al. (2004) discussed subject-to-subject variation in human intestinal microbiota. Other investigators (Zoetendal et al., 1998; Vanhoupt et al., 2004) emphasized that individual humans have a unique intestinal microbiota. Therefore, if possible, treatment effects must be evaluated separately for each individual. Since the principle components analysis score plots grouped by treatment could not identify any treatment effect, another common approach was attempted. Principle components analysis score plots and Euclidean distance cluster analysis dendrograms were created for each individual subject. For example, principle components analysis plots for both subject 27 (treated with both probiotics and prebiotics) and subject 10 (control) indicated that weeks 9 and 12 were more similar to each other than to the sample from week 3 (Fig. 2). The dendrograms showed that weeks 9 and 12 were between 20% and 40% similar, whereas week 3 was 3–8% similar to weeks 9 and 12. Unfortunately, there is really no way to evaluate the significance of this apparent difference in microbiota by sample week. In fact, these differences may be entirely random.

It is tempting to categorize subjects’ responses to treatment based on these analyses. A treatment would have an effect if the week 3 sample (pretreatment sample) were different from the other two samples. Although sample-to-sample differences are visible in these individual analyses
(Fig. 1), the problem is how to assess the statistical significance of these differences (i.e. no way to obtain a P-value) and there is very little statistical power because there are more than 200 variables (TRF peaks) and only three samples for each subject. As a result, each sample has a 33% chance of being the most different.

Both principle components analysis and the cluster analysis method used for this study required the use of Euclidean distance, which is not the best choice for calculating distances with molecular profile data (Rees et al., 2004). Although Euclidean distance is the most commonly used in the literature; Rees et al. (2004) argue that Bray–Curtis distances are superior due to the large blocks of zeros present in molecular profile data sets. Bray–Curtis similarity was specifically designed for use with species abundance data sets (Beals, 1984), it is less susceptible to bias introduced by large numbers of zero abundance data, and it takes the relative abundance of each TRF peak into account when comparing two patterns. TRFLP data often contains zero values (a peak present in one pattern that is not seen in another) and can cover as much as 2.5 orders of magnitude variation in TRF peak area. Therefore, Bray–Curtis
similarity is a preferable similarity measure when comparing TRF patterns.

In the probiotic–prebiotic study, Bray–Curtis similarity was calculated between weeks 3 and 9, weeks 3 and 12, and weeks 9 and 12, and interval plots of those values were created for each different treatment group (Fig. 3). With this method, changes due to treatment should be very easy to see; one might expect the similarity between week 3 and week 9 samples to be the least of all three comparisons across all treated subjects.

No within-baseline comparison was possible with this sampling regime, but it was possible to estimate the reproducibility of the TRFLP method with fecal samples. Pairwise comparisons were made of five replicate TRF patterns from the same sample (replication from the DNA extraction step), resulting in an average similarity of 83%. Intriguingly, the variation from week to week in the probiotic treatment group was the same as that seen in the reproducibility experiment (Fig. 3). Therefore, there was no discernible change due to treatment. The week-to-week variation for the prebiotic groups was less than that seen in the reproducibility experiment but all three pairwise similarities were the same, as indicated by overlapping standard error bars. Once again, this indicates there was no change due to treatment. The similarity of week 9 to week 12 control samples was clearly higher than the other pairwise similarities, but this cannot be explained by any particular treatment.

The negative results of this study were most clear after using Bray–Curtis pairwise similarity. This made it possible to visualize the lack of change in fecal bacterial communities given the sampling regime in this study, where no statistical measure with any power could be used. This study also made it clear that a set of baseline samples must be included in investigations of human fecal microbiota.

### Analysis of the antibiotic–probiotic study TRFLP data

The goal of the antibiotic–probiotic study was to observe the effects of probiotic treatment concurrent with antibiotic therapy on fecal communities in humans. As one baseline sample was insufficient for proper analysis of the data in the first probiotic study, we took three baseline samples in this study at days 1, 7, and 14. Principle components analysis score plots of all subjects’ TRFLP data collected over days 1–21 were used to determine the effect of antibiotics on the fecal microbiota. However, subject-to-subject variation obscured the effect of antibiotic treatment (Fig. 4). In addition, PC1 and PC2 when combined only represented a small fraction of the total variation in TRFLP data (16%) and thus may have under-represented any changes induced by antibiotic consumption.

Because of the subject-to-subject variability seen in Fig. 4, principle components analysis and cluster analysis were performed on each subject. It became clear that subjects fell into two major categories: those with stable baseline microbiota (days 1, 7 and 14) and those whose baseline microbiota varied significantly. For example, in subject 50 (Fig. 5) there was a clear antibiotic effect on the fecal bacterial community structure at days 21 and 25 that appeared to be gone by day 34. After the subject stopped taking probiotics, the fecal community changed again (day 48). In subject 42 (Fig. 5) a smaller effect on the fecal bacterial community structure from antibiotic treatment was seen, because variation in the baseline data made it difficult to tell whether the shift at day 21 was significant. Although the antibiotic effect for subject 50 appears obvious in the cluster analysis, there is no way to determine statistical significance and the similarities are actually very low. Also, the large reduction in sample number

---

**Fig. 3.** Bray–Curtis similarity interval plots of probiotic–prebiotic study terminal restriction fragment length profile data; weeks 3–9, weeks 3–12, and weeks 9–12 for each treatment group. Error bars represent one standard error. The large black bar represents the standard error of replicate terminal restriction fragment patterns from a single sample.
when analyzing the data by individual instead of grouping by treatment results in a drastic loss of statistical power, invalidating multivariate hypothesis tests, such as MANOVA.

Principle components analysis and cluster analysis can give a good visual representation of treatment effects in an individual, clearly showing which samples cluster together (Figs 2 and 5). Unfortunately, it is not possible to extract statistical significance from these clusters, which only allows for a discussion of trends. However, the distance measures that are used in principle components analysis and cluster analysis can be used for statistical hypothesis testing. Distance measures are, by definition, pairwise comparisons. Thus, assessment of treatment effects requires comparison of pretreatment samples (baseline) with posttreatment samples. The natural variation in native microbiota can potentially mask treatment effects. Prior studies of stability in intestinal microbiota assert that individuals are generally stable (Zoetendal et al., 1998; Donskey et al., 2003; Vanhoutte et al., 2004). However, as these studies did not quantify stability, it is difficult to identify a subject as having unstable intestinal microbiota. Clearly, a single baseline sample does not allow for an assessment of natural variation. Therefore, the collection of multiple baseline samples per subject is advisable. A comparison of Fig. 3 (one baseline sample per subject in the probiotic–prebiotic study) and Fig. 6 (three baseline samples per subject in the antibiotic–probiotic study) highlights the advantages of having multiple baseline samples.

In the next analysis approach, Bray–Curtis similarity to baseline (days 1–14) was calculated for each day after antibiotic treatment. Similarity within baseline was also calculated for each subject. By comparing within-baseline similarity to similarity from day 21 to baseline it was clear that the antibiotics had a significant effect on fecal microbiota across all subjects (ANOVA, \( P < 0.001 \)). The average similarity of baseline samples compared to the first day after antibiotics (day 21) was 42%, whereas the average similarity within baseline (days 1–14) was 51%. Average similarity to baseline at day 25 increased to 47% and was not significantly different from within baseline at a 95% confidence level.
Figure 6. Average Bray–Curtis similarity from baseline (days 1–14) for each day after treatment using probiotic–antibiotic study terminal restriction fragment length profile data. Error bars represent one standard error. Bray–Curtis similarity in percent is indicated on the left axis.

(P = 0.078). At day 34 the average similarity reached 49% and at day 48 it had decreased to 46%. The significance of the change at day 48 is not clear, especially since it occurred in both the probiotic and placebo treatment groups and thus cannot be solely attributed to the cessation of probiotic ingestion at day 34.

To assess the effect of probiotic ingestion on changes brought about by antibiotic treatment, the Bray–Curtis similarity data were analyzed separately for probiotic and placebo groups (Fig. 6). Both groups exhibited a similar trend toward increased similarity to baseline over days 21 through 34, but the probiotic group exhibited a larger increase in similarity at day 34. The difference between the two groups was not significant when all four postantibiotic treatment days were analyzed (MANOVA, P = 0.135). However, when the anomalous results of day 48 were removed from analysis, a trend revealing a difference could be detected at a 90% confidence level (P = 0.066).

Analysis of the antibiotic–probiotic study culture data

The culture data were first analyzed for antibiotic effects using interval plots as a standard univariate method to look for significant differences with treatment (Fig. 7). However, the large variation in counts (~1 log standard error) across all subjects resulted in a loss of power for statistical analyses and precluded detection of any significant differences (ANOVA). Increasing trends were visible in Bacteroides and enterics at day 21. There was no trend seen for Clostridium, Bifidobacterium, and Lactobacillus. Madden et al. (2005) artificially reduced error by excluding uncountable plates from their analyses. However, this approach is inappropriate since it will bias the analysis, possibly leading to false conclusions. Although analysis of culture data generally involves univariate statistics, multivariate statistics can be applied if multiple media are investigated. When Sullivan et al. (2003) presented univariate analyses of culture data they could only discuss trends, as no statistically significant effects could be reported. However, multivariate statistics may provide a more accurate way to test for treatment effects. Unfortunately, MANOVA of data from all five culture media also showed no statistical difference between bacterial counts before antibiotic treatment compared to the days after antibiotic treatment (MANOVA, P = 0.34). This may be due to the subject-to-subject variation previously seen in the TRFLP data.

Figure 7. Interval plots of log colony-forming units of probiotic–antibiotic study culture data; all subjects, days 1–21. Error bars represent one standard error. Panel a, BBE agar for organisms in the B. fragilis group; panel b, BIM-25 agar for bifidobacteria; panel c, egg yolk agar for Clostridium ssp. spores; panel d, LBS agar for lactobacilli; panel e, MacConkey agar for Enterobacteriaceae.
Euclidean distance was used to compare the average distance of baseline samples (days 1–14) to the first day after antibiotics (day 21) for each subject. Euclidean distance was used in this case because the standard assumptions of normality and nonzero data are not violated with culture data as they are with TRFLP data. When the effect of antibiotics was evaluated on an individual subject basis in the same way as with the TRFLP data, a significant effect was detected (ANOVA, $P = 0.003$). The average pairwise distance at baseline was 3.3 compared to an average distance to baseline of 4.2 after antibiotic treatment. At day 25 the distance to baseline decreased to 3.5, not significantly different from within baseline ($P = 0.56$). At day 34 the average distance to baseline was 3.7 and at day 48 it was 3.5.

To assess the effect of probiotic ingestion on changes brought about by antibiotic treatment, the Euclidean distance data were analyzed separately for probiotic and placebo groups (Fig. 8). Here the culture data showed a much greater difference than the TRFLP data, with the probiotic group maintaining an average distance to baseline of around 3.6–3.7 throughout the study. In contrast, the placebo group showed a large shift from baseline at day 21 and a return to near baseline thereafter. The difference between the two groups was quite significant when all four postantibiotic treatment days were analyzed (MANOVA, $P = 0.004$). When day 48 was removed from analysis there was still a significant difference between groups ($P = 0.046$).

To account for individual variation and still determine which media showed the largest change in fecal microbiota, the average baseline counts for each subject were subtracted from the counts for that subject on each day subsequent to antibiotic treatment. Although these data could not be used to statistically analyze an antibiotic effect, they still reflected the significant difference between the probiotic and placebo treatment groups (MANOVA, $P = 0.049$) over the four postantibiotic treatment days (Fig. 9). Follow-up ANOVAs indicated that a probiotic effect was significant for bifidobacteria ($P = 0.030$) and Enterobacteriaceae ($P = 0.006$), but not significant for the B. fragilis group (ANOVA, $P = 0.104$), Clostridium ($P = 0.601$) or lactobacilli ($P = 0.772$).

**Summary and conclusions**

There are a variety of ways to collect quantitative data for the assessment of treatment effects on intestinal communities,
including both culture and molecular methods. Culturing is a common approach but can give an incomplete picture of the intestinal microbiota as feces contain large numbers of unculturable or difficult to culture organisms. Media choice and sample handling can also skew the data. Molecular profile methods such as DGGE and TRFLP may present a broader view of the intestinal microbiota (Tannock et al., 2000; Donskey et al., 2003; Wang et al., 2004).

Culture or molecular profile data can be grouped by treatment for univariate hypothesis testing as the comparison to baseline implicitly takes into account subject-to-subject variation in intestinal microbiota. A good way to visually represent the pairwise distances is with interval plots (Figs 3, 6, and 8). These interval plots clearly show the significant changes in community structure. For culture data, the use of univariate analyses can be appropriate to identify individual population effects, but only when they are used after an appropriate multivariate analysis. Instead of using the log CFU measurements (Fig. 7), however, the proper univariate method uses the difference between a subject’s own individual baseline and each posttreatment measurement (Fig. 9).

The method of comparison to baseline within an individual before grouping by treatment provides both a good visual representation of the data and a clearer indication of treatment effects on intestinal microbiota than an evaluation of data grouped before analysis. This applies whether the study has a positive result (probiotic–antibiotic study) or a negative result (probiotic–prebiotic study). It is also apparent from analyzing these two data sets that a large within-subject sample size and multiple baseline samples are necessary to accurately analyze treatment effects on intestinal microbiota. In addition, it is important to use Bray–Curtis distances for molecular profile data. The use of Euclidean distances is appropriate for culture data and univariate analysis of culture data is significant only if it follows the appropriate multivariate method.

Acknowledgements

Dairy Management Incorporated funded the probiotic–prebiotic project. The probiotic–antibiotic project was funded by Danisco USA. We would like to thank Rosemary Sano­sky-Dawes from North Carolina State University for the culturing work and Tiffany Glavan from the Environmental Biotechnology Institute for the reproducibility experiment.

References


