COMPARISON OF CLONING AND SEQUENCING VS. TERMINAL RESTRICTION FRAGMENT ANALYSIS OF BACTERIA FROM HUMAN FECES.

David Doherty

and

Danielle Elizondo

October 24th 2001

Abstract

Extracted DNA from the Probiotic fecal sample 62C was chosen for cloning and sequencing. The Probiotic project was a previous study which analyzed human fecal samples before and after the ingestion of probiotic bacteria, bacteria found to be beneficial to human health. The sequenced clones gave a small representation of the organisms that were present in the original sample and their genetic relationship to each other. The sequences were compared to the original Probiotic community terminal restriction fragment (TRF) pattern results for sample 62C. The TRF and sequencing analyses together provided a better understanding of the identification and abundance of the organisms present in the sample than just one method of analysis.

Introduction

In this project, the community DNA samples from the Probiotic Project were used to clone and sequence. The Probiotic project's purpose was to investigate the effect of probiotics in the gut, and to observe the changes on the human physiological system, especially flora in human intestines (1). The project used terminal restriction fragment (TRF) patterns (2) to help examine the organisms present after the consumption of probiotic bacteria in healthy volunteers. The sample we cloned and analyzed, 62C, came from the post-treatment stage of the probiotic project. The purpose of our study was to determine if any of the sequenced clones could have been present in the original TRF sample 62C reading by comparison of the predicted peak locations of the two analysis methods.

There are several problems with community TRF pattern analysis. One problem is that the peaks in the graph could be misleading. Any peak in the TRF pattern could represent more than one organism. The specific number of organisms cannot be determined in a given peak. Also, only those organisms which make at least 1% of the total DNA sample will be present in the analysis. Cloning the extracted DNA gave us the opportunity to confirm the results of the TRF reading.

Materials and Methods

*Amplification, cloning, and screening***.** PCR primers, 46F (GCYTAACAC ATGCAAGTCGA) and K2R (CCAGCAGCCGCGGTAATAC) (2), were used in order to amplify the 16S region of the community DNA samples. A 50 µl reaction was made with 3.0 µl of 10 mM DNTP, 2.0 µl 20ug/ml BSA, 5.0 µl 10x PCR Buffer,1.0 µl 46F,1.0 µl k2R , and 10.0 µl of DNA (1 ng/µl). DN*ase*-free H_20 was used to bring the volume up to 50 µl. *B.subtilis* was used as the positive control in all PCR reactions as well as an open and closed control. The PCR ran for 35 cycles and was set up as follows: initial denaturation of 2 minutes at 94°C, 35 cycles of 94°C for 1.0 minute, 46.5°C for 1.0 minute and 72°C for 2.0 minutes followed by 10 minutes at 72°C. Following PCR, samples were analyzed by electrophoresis on a 1.5% (wt/vol) agarose gel to determine if contamination was present in our PCR product.

Invitrogen's TOPO-TA cloning kit was used and the manufacture's protocol was followed in order to clone PCR products into a plasmid. Fifty µl of Invitrogen competent cells were used and protocol directions were followed with the exception of the amount of ligation reaction used. The entire ligation reaction (10 µl), instead of the suggested volume (2 µl) was used in the transformation reaction to help maximize results. The transformation reaction was kept on ice for 30 minutes and then heat shocked for 30 seconds at 42°C. LB plates coated with 1.60 mg of X-Gal and 50 µg/ml of ampicillin were used to plate out the cells. Plates were then incubated at 35°C for 24 hours for optimum color definition. The colonies were then re-plated in order to make sure the DNA insert was not removed. Fifty white colonies were then picked for plasmid isolation. Three ml of Terrific Broth (TB) with 100 µg/ml of ampicillin was inoculated with one colony and incubated at 35°C for 15 hours. TB instead of the suggested LB was the preferred media to isolate plasmid DNA because TB had more *E. coli* growth.

Sequencing. The plasmids were extracted from overnight cultures using MoBio's mini plasmid prep kit as specified by the manual. The plasmid DNA was analyzed on a 1.5% (wt/vol) electrophoresis agarose gel in order to guarantee that the samples were not lost in the extraction process and that the extraction was successful. The samples were then prepared for sequencing using ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit, with AmpliTaq DNA Polymerase FS (fluorescent sequencing). 1.6 pmol of M13 forward and reverse primers were used along with 4 μl of the BigDye reaction mix and 4 μl of the sample DNA. The reaction was brought up to 10 μl with H_20. This mixture underwent a cyclic incubation as follows: an initial denaturation at 96°C for thirty seconds, forty cycles of 96°C for 10 seconds, 50°C for 10 seconds and 60°C for 240 seconds. After the BigDye reaction, denaturation of the double stranded PCR product was performed for consistent electrophoresis on the ABI Prism 377 (2).

Sequence Analysis. The two sequences (forward and reverse) were aligned using the computer program Seqman (3). The 3.9 Kb plasmid sequence was cut from each sequence result using the Seqman program. The remaining sequence was the original insert. This ensured that only the inserted DNA would be analyzed. A chimeric check was performed using the RNA Ribosomal Database Project II (4). Insert sequences were run through BLAST (5) in order to determine the closest identity of the bacteria from which the DNA was obtained (see Table 1). The highest percentage BLAST result was used in order to determine the origin of the DNA.

Phylogenetic Tree. CLUSTAL and PHYLIP (6) were used to produce a phylogenetic tree. After the maximum likelihood was calculated by the PHYLIP software, the phylogenetic tree program TREEVIEW was used to calculate a dendogram. The tree was then re-organized using the organism *Thermatoga maritima* (6) as the root (see Fig. 1).

TRF Comparison. The sample sequences were compared to sample 62C's TRF results. Using the restriction enzymes *Dpn*II, *Msp*I, and *Hae*III, cut sites were matched and analyzed. Microsoftâ Excel was used to predict the peak positions of the cut sites. Once these were obtained, they were compared to the original sample's peak positions.

Results

Amplification, cloning, and screening A successful gene insertion was determined by an agarose gel. Fifty white colonies were picked from the cloning reaction and plasmid isolation was performed.

Sequencing and Sequencing analysis. All clones from 62C were successfully sequenced and analyzed. Two clones were determined to be chimeric using the Ribosomal Data Base Chimera Check (Table 1). This check was helpful in order to determine sequences coming from more than one organism. The two clones that were chimeric were no longer used for analysis. The remaining clones were then analyzed using BLAST in order to determine the organisms from which the sequences originated.

Table 1. BLAST and Chimeric results

Clone #	Blast result	% Homology	chimeric?
7a	Uncultured rumen bacterium 5C3d-17	94%	No
8a	Uncultured bacterium adhufec202	86%	No
11a	*Bacteriodes* sp.	98%	No
26a	*Clostridium methylpentosum*	90%	No
27a	*Clostridium oroticum* 16S	98%	No
28a	*Ruminococcus lactaris*	92%	No
30a	*Bacteriodes* sp.	99%	No
31a	*Clostridium cellulosi*	88%	No
40a	*Clostridium orbiscindens*	91%	No
41a	Agricultural soil bacterium clone SC-I-18	100%	Yes
42a	*Eubacterium desmolans*	97%	No
43a	Swine feces bacterum F18 16S	97%	Yes
44a	*Ruminococcus torques*	94%	No
45a	*Anaerovorax oderimutans*	85%	No
46a	*Ruminococcus torques*	94%	No
47a	*Ruminococcus torques*	94%	No
48a	*Anaerovorax oderimutans*	85%	No
51a	*Ruminococcus torques*	94%	No
53a	*Anaerovorax oderimutans*	85%	No

Phylogenetic tree. The tree was constructed with the clone sequences, their highest percentage match BLAST result sequences, and sequences from organisms previously identified in a human gut/colon analysis: *Ruminococcus torques*, *Eubacterium rectale*, *Butyrivibrio fibrisolvens*, and *Clostridium viride* (7). The rooted tree displayed the relationship between the clones in terms of their maximum likelihood alignment (Fig 1). It also showed the clone sequences and their BLAST results to be close in relation.

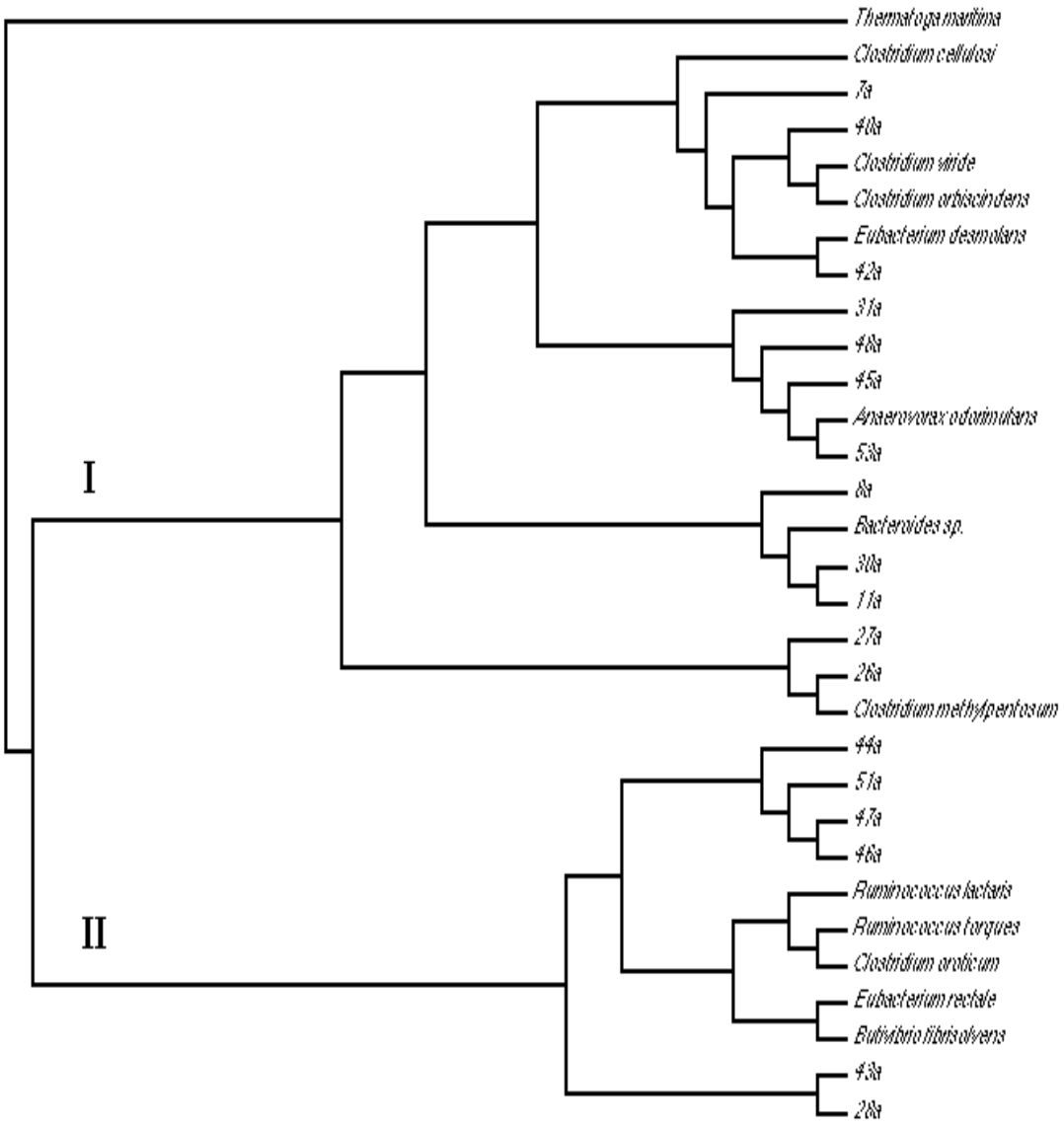

Fig.1: Rooted Phylogenetic Tree

TRF Comparison. A comparison between the predicted cut sites of the sequenced clones and the peak locations 62C from the Probiotic study can be seen in Table 2. The original TRF pattern for 62C with the *Dpn*II cut sites can be seen as Fig. 2. Three enzymes (*Hae*III, *Msp*I, and *Dpn*II) were used in the analysis and 10 out of 17 clones showed up on the original TRF pattern. The remaining 7 samples did not contain the same cuts sites as the TRF pattern and may not have been represented in the original TRF pattern.

Discussion

Sequencing Analysis. The BLAST results were consistent with expected human gut organisms. All matches between the sequences and the BLAST database had a high percentage match (above 90%). Also, all the organisms that were identified through BLAST turned out to be anaerobic. The clones were identified as belonging to only five different genera. *Anaerovorax*, *Clostridium*, and *Ruminococcus* were the three genera that made up the majority of the results.

Table 2. TRF Pattern Analysis

	Predicted TRF peaks			Observed in TRF pattern		
Clone	**DpnII**	**MspI**	**HaeIII**	**DpnII**	**MspI**	**HaeIII**
41A	***18***	**162**	**162**	N/A*	162	161
44A	36	186	281	N/A*	183, 185	278, 280, 281
46A	36	186	281	N/A*	183, 185	278, 280, 281
47A	36	186	281	N/A*	183, 185	278, 280, 281
51A	36	186	281	N/A*	183, 185	278, 280, 281
26A	47	245	***287***	none	242,243,244	264, 266, 267
27A	47	262	265	none	258, 259. 260, 261	261, 262, 264
43A	52	184	235	none	181, 182, 183	232, 233
7A	***239***	249	225	236, 238	246, 247, 248	222, 223, 225
40A	***248***	***258***	199	244, 245	none	198
45A	248	***448***	174	245, 246, 248	443, 444	none
48A	248	***448***	174	245, 246, 248	443, 444	none
53A	248	***448***	174	245, 246, 248	443, 444	none
28A	***251***	186	237	248, 250	185	236
8A	***295***	154	301	293	153, 154	299
11A	***454***	***59***	***225***	451, 454	56, 57	222, 223, 225
30A	***454***	***59***	***225***	451, 454	56, 57	222, 223, 225

* Peak occurs in non-interpretable region

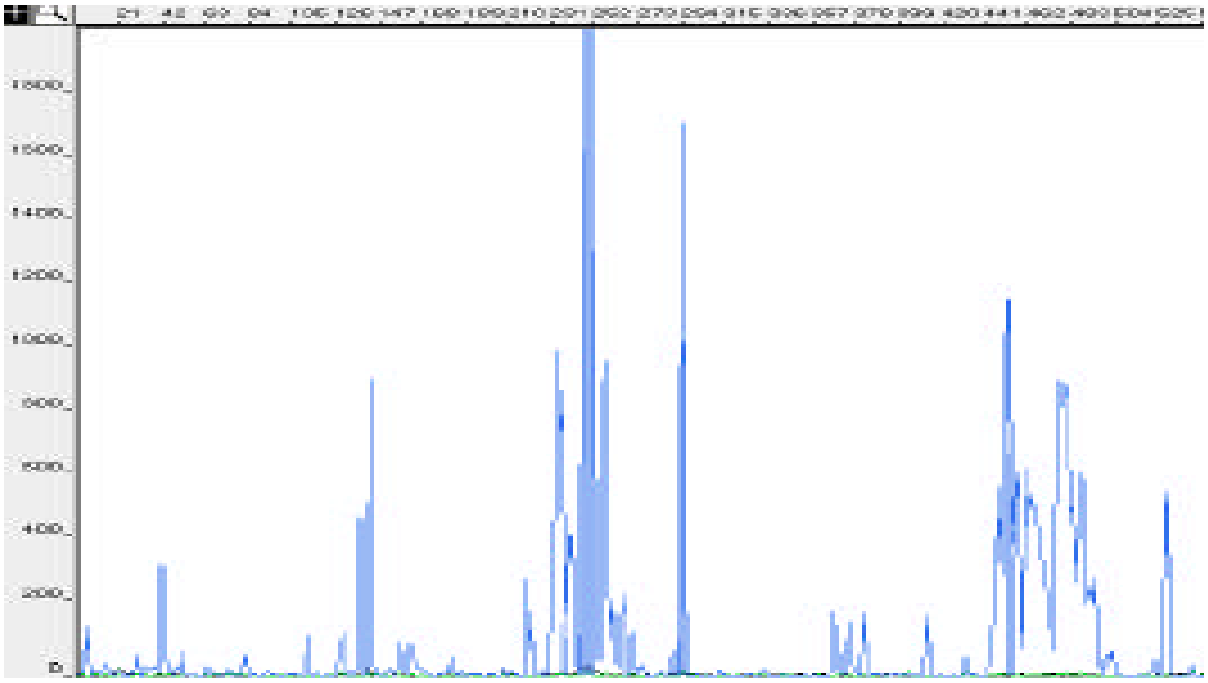

Figure 2: Sample 62C *Dpn*II TRF Pattern. The peaks containing possible clones are shaded.

Phylogenetic Tree. The tree produced two main branches. Branch I contained the largest group of sequenced clones and their respective BLAST matches. Note how the clone sequences and identified organisms are evenly distributed throughout branch I. In branch I, the clone sequences were almost identical to the BLAST results so they clustered together. The distribution in branch II is quite different. The branch II clone sequences were so similar to each other and different from the BLAST matches that the PHYLIP program clustered them with each other rather than with their respective BLAST results. Note the placement of 27a in the tree (branch I) and its BLAST match, *C. oroticum* (branch II). These two are on separate branches because the clone 27a sequence more closely matched another clone than the organism determined by BLAST. Clone 27a was identified as one species of *Clostridium* in BLAST and was placed with another species of *Clostridium* in the tree. BLAST and CLUSTAL align

sequences differently, which most likely accounts for the tree discrepancy. BLAST may use subset sequences to get the best percentage match whereas CLUSTAL uses each entire sequence to align the sequences. Antonia Suau et.al showed that *Clostridum* species were scattered throughout the tree with *Eubacterium* and *Ruminococcus* (7). This is also the case in our tree. *C. oroticum,* clone 27a's BLAST result, was closely associated with *R. torques* as Suau showed. Also note that all but one *Clostridium* species showed up in branch I due to its close association to *E. rectale* and *R. torques* (7). These three organisms clustered more tightly than any others in the tree.

TRF Comparison. An approximate match between all three cut sites of the clones with the original sample was required. Ten of the 17 clones had a close match. Of the 7 clones that were not included in the original TRF, 3 were most likely from the same organism. Since 3 of 17 clones had the same sequence, it is logical to assume this organism was abundant in the original fecal sample. This would be true in any case where two or more clones had the same sequence. It would therefore be expected for this organism to be present in the original TRF analysis. Three possible reasons for this inconsistency could have been: 1) the cut site was not digested by the restriction enzyme, 2) the DNA had a poor primer site, or 3) a sequencing error produced wrongly predicted TRF results. The third reason would be the least likely for clones 45a, 48a, and 53a, since they were most likely from the same organism. Any of the possible error explanations could be associated with clones 26a, 27a, 43a, or 40a.

When comparing the TRF results with the BLAST results, there is a notable similarity. Clones 40a, 45a, 48a and 53a all had the same predicted peak with *Dpn*II. Clone 40a, however, had different peaks with *Hae*III and *Msp*I. These different peak locations are what differentiated

40a from the other three clones which were determined to all be *A. oderimutans* by BLAST. All four clones, however, were on the same branch in the phylogenetic tree.

Conclusions

Sequencing, in addition to TRF analysis, provided a more thorough analysis of the fecal sample DNA. The TRF results gave a general idea of the organisms in abundance, but not the specific identity of the organisms. Sequencing the sample clones gave possible identifications of organisms in the fecal sample in addition to organisms that were missed in the TRF analysis. Picking and sequencing more clones would have given a better understanding of the different clone identifications and varieties in the sample. In addition, a greater number of clones could have helped to explain the discrepancy between the TRF and sequencing predictions of abundant organisms originally contained in the fecal sample.

References

1. BIBRA International Project No. 3310/1. 1999. The Physiological Effects of Daily Consumption of Prebiotic and Probiotic on the Gut Microflora, the Immune System and on the Levels of Serum Lipids in Healthy Volunteers. Report, A29

2. Clement, B.G., L. Kehl, K. DeBord, and C. Kitts. 1998. Terminal restriction fragment patterns (TRFPs), a rapid, PCR based method for the comparison of complex bacterial communities. Jounal of Microbiological Methods. 31:135-142

3. Lasergene. DNASTAR. Lasergene99. MacIntosh. Madison, WI: DNASTAR,1999

4. Ribosomal Database Project II. Michigan State. www.cme.msu.edu/RDP

5. National Center for Biotechnology Information. www.ncbi.nlm.nih.gov/BLAST

6. Page. R.D.M.. 1996. Treeview: An Application to display phylogenetic trees on Personal Computers. *Computer Applications in the Biosciences*. 12:357-358

7. Suau, A., R. Bonnet, M. Sutren, J. Godon, G.R. Gibson, M.D. Collins, and J. Doré. 1999. Direct Analysis of Genes Encoding 16S rRNA from Complex Communities Reveals Many Novel Molecular Species within the Human Gut. *Applied and Environmental Microbiology*. 65:4799-4807