Variation between observed and true Terminal Restriction Fragment length is dependent on true TRF length and purine content

Christopher W. Kaplan, Christopher L. Kitts

Environmental Biotechnology Institute, California Polytechnic State University, San Luis Obispo, CA 93407, USA

Abstract

Terminal Restriction Fragment (TRF) pattern analysis has become a widely used and informative tool for studying microbial communities. Variation between sequence-determined or true TRF length and observed TRF length (TRF drift) has been previously reported and can significantly affect identification of bacterial species using TRF lengths predicted from sequence databases. In this study TRF drift was determined for 21 bacterial species using an ABI 310 Genetic Analyzer. TRF drift was positively correlated with true TRF length and negatively correlated with TRF purine content. This implies that subtle differences in molecular weight, whether from purine content or dye label, can significantly affect the observed TRF length.

Keywords: TRF; T-RFLP; TRFLP

As environmental microbiology has evolved, so have the techniques employed in its analysis. The use of molecular methods to describe microorganisms and the communities they comprise have become commonplace. A recent tool in environmental microbiology is Terminal Restriction Fragment (TRF) pattern analysis (also known as Terminal Restriction Fragment Length Polymorphism, T-RFLP). TRF patterns are produced by amplifying DNA from a bacterial community using PCR with one fluorescently labeled primer and cutting the amplicons with a restriction endonuclease. The terminally labeled fragments are detected by fluorescence after separation via electrophoresis on a denaturing sequencing gel. Organisms in a community are thus differentiated based on sequence variation that results in TRFs of different lengths, which in turn create a pattern unique to that community. The resulting patterns can be used to make inferences about environmental effects on community structure or evaluate community dynamics. Several comprehensive reviews of the TRF method exist which illustrate the utility of this tool (Kitts, 2001; Marsh, 1999). An increasingly popular trend in TRF pattern analysis has been to associate TRF peaks with clones or predicted matches from extensive databases of existing sequences (Braker et al., 2001; Kaplan et al., 2001; Moeseneder et al., 2001; Sakano et al., 2002). Associating sequenced clones or database matches with a TRF peak is problematic since related organisms commonly produce TRFs of the same length, requiring several enzyme digests to resolve community members. To make accurate matches

requires that TRFs in a pattern migrate in such a way that their reported length represents their true length. Discrepancies between sequence-determined TRF length and observed TRF length (TRF drift) have been reported previously with estimates ranging from as little as 1 bp to as much as 7 bp (Kitts, 2001; Kaplan et al., 2001; Liu et al., 1997; Clement et al., 1998; Osborn et al., 2000). In this paper, we evaluated TRF drift in the 16S rDNA region for 26 bacterial strains in an effort to quantify sources of variation and achieve more accurate database matches.

The organisms used in this study were picked from cultures available in our lab based on true TRF length and TRF purine content (Table 1). All organisms were streaked on Tripticase Soy Agar and incubated at optimum temperature and time to provide sufficient growth for DNA extraction. Cells were then scraped from plates and transferred to MoBio® bead lysis tubes (Solana Beach, CA, USA). The protocol given in the MoBio® kit was followed for the extraction process with the following exception: cells were lysed in the Bio 101 FastPrep FP120 (Carlsbad, CA, USA) running at 4.5 m/s for 25 s. The DNA was visualized by agarose gel electrophoresis and quantified by UV spectrophotometry. Amplification of template DNA was performed by using primers 6-FAM labeled 46f (5′-GCYTAACACATGCAAGTCGA), and unlabeled 536r (5′-GTATTACCGCGGCTGCTGG). Reactions were carried out in duplicate with the following reagents in 50 µl reactions: template DNA, 10 ng; 1X Buffer (Applied Biosystems, Foster City, CA, USA); dNTPs, 3×10^{-5} mmols; bovine serum albumin, 4×10^{-2} µg; MgCl$_2$, 1.75×10^{-4} mmols; 46f, 1×10^{-5} mmols; 536r, 1×10^{-5} mmols; *Taq*Gold DNA polymerase (Applied Biosystems), 1.5 U. Reaction temperatures and cycling for samples were as follows: 94 °C for 10 min, 35 cycles of 94 °C for 1 min, 46.5 °C for 1 min, 72 °C for 2 min, followed by 72 °C for 10 min. The products were visualized on a 1.5% agarose gel and any inconsistent or unsuccessful reactions were discarded. To remove primers and concentrate amplicons, the MoBio® PCR Clean-Up kit was utilized according to the protocol included with the kit. The combined amplicons were then quantified by UV spectrophotometry. Restriction enzyme reactions contained 10 ng of labeled DNA, and restriction endonuclease enzyme (*Hha*I, 0.1 U; or *Msp*I, 0.1 U; *Rsa*I, 0.2 U; or *Dpn*II, 0.2 U; or *Hae*III,

0.2 U (New England Biolabs, Beverly, MA, USA) in the manufacturer's recommended reaction buffers. Reactions were digested for 2 h at 37 °C. Samples were ethanol precipitated then dissolved in 9 µl of Hi-DI formamide (Applied Biosystems), with 0.5 µl each of Genescan Rox 500 (Applied Biosystems) and Rox 600 (BioVentures, Murfreesboro, TN, USA) size standards. The DNA was denatured at 95 °C for 5 min and snap-cooled in an ice slurry for 10 min. Samples were run on an ABI Prism™ 310 Genetic Analyzer at 15 kV and 60 °C. TRF sizing was performed on electropherogram output from Genescan™ 3.1.2 software using Local Southern method with heavy smoothing. For DNA sequencing, extracted DNA samples were amplified by PCR as described above except that the forward and reverse primers were replaced with 8df (5′-AGAGTTTGTTCMTGGCT-CAG) and 803r (5′-CTACCAGGGTATCTAATCC). Sequencing reactions (10 µl) contained: DNA, 4 ng; primer, 1.6×10^{-5} mmol; ABI Big Dye (Applied Biosystems), 4 µl. Samples were run on an ABI 377 DNA sequencer and the resulting sequences analyzed in SeqMan™ (DNAStar, Madison, WI, USA). Sequences were analyzed for TRF cut sites of each enzyme used in this study for comparison with TRF pattern data.

TRF data were analyzed using five different analysis methods (2nd order least square, 3rd order least square, local southern, global southern, cubic spline) available with Genescan 3.1.2 software. Different analysis methods produced different standard curves for the internal ladder, thus creating differences in observed TRF length. As previous reports have shown (Osborn et al., 2000), the local southern method produced a standard curve with the least TRF drift (data not shown). To facilitate a statistical analysis, we defined "TRF drift" as the observed TRF length *minus* the true TRF length. Amplicons that did not contain an enzyme cut site, resulting in an uncut amplicon, were not included in this dataset since *Taq* polymerase adds 3′ adenine residues to PCR products resulting in longer fragments than predicted from sequences.

The average TRF drift was approximately -3 bp over the lengths analyzed, with a standard deviation of 1.28 bp. Longer TRFs had larger TRF drift associated with them (Fig. 1). Analysis of electropherogram data suggested that the major source of TRF drift was the differential migration of ladder and sample DNA.

Table 1
True TRF length (TRF), observed TRF drift (TRF drift), predicted TRF drift (Pred. drift), and purine content (%R) for each organism and enzyme combination

Organism	Group	DpnII				HaeII				HhaI				MspI				RsaI			
		TRF	TRF drift	Pred. drift	%R	TRF	TRF drift	Pred. drift	%R	TRF	TRF drift	Pred. drift	%R	TRF	TRF drift	Pred. drift	%R	TRF	TRF drift	Pred. drift	%R
Bacillus licheniformis AY186035	1	266	−3.5	−1.6	56	272	−3.0	−2.0	56	203	−4.4	−2.3	55	107	−4.2	−6.8	56	419	−7.8	−3.6	57
Bacillus licheniformis AY186036	1	266	−3.6	−1.6	56	272	−3.7	−2.0	56	203	−3.9	−2.2	55	107	−4.2	−6.8	56	419	−7.8	−3.6	57
Bacillus sp. AY186037	1	265	−3.3	−1.6	57	271	−3.6	−2.0	57	202	−4.3	−2.2	56	107	−3.7	−6.8	57	418	−7.7	−3.7	58
Bacillus subtilis AY186038	1	265	−2.7	−3.0	57	271	−3.0	−1.7	57	202	−3.6	−2.3	56	107	−3.7	−6.9	57	448	−5.9	−3.7	58
Staphylococcus warneri AY186059	1	192	−1.4	−1.6	57	272	−1.0	−1.7	57	200	−1.6	−1.9	57	117	−3.6	−6.9	58	448	−5.8	−3.6	57
Citrobacter freundii AY186052	2	235	−1.2	−3.0	59	168	−1.4	−2.0	62	335	−1.6	−2.2	59	458	−6.5	−6.7	59	389	−4.8	−3.6	59
Enterobacter aerogenes AY186054	2	233	−1.2	−1.6	59	166	−1.4	−2.0	61	333	−1.7	−2.2	59	456	−6.1	−6.7	59	387	−3.7	−3.6	59
Enterobacter aerogenes AY186051	2	233	−1.3	−1.6	59	166	−1.4	−2.0	61	333	−1.8	−2.2	59	456	−6.6	−6.7	59	387	−4.3	−3.6	59
Enterobacter cloacae AY186050	2	233	−0.9	−1.6	59	166	−1.1	−2.0	62	333	−1.3	−2.3	59	456	−6.9	−6.9	59	387	−4.1	−3.6	59
Escherichia coli AY186040	2	235	−1.3	−1.6	59	168	−1.1	−2.0	61	335	−1.8	−2.2	59	458	−7.0	−6.8	59	389	−4.6	−3.6	58
Escherichia coli AY186041	2	235	−0.8	−1.6	58	168	−0.9	−2.0	61	335	−1.2	−2.2	59	458	−7.2	−6.8	58	389	−4.9	−3.6	58
Klebsiella pneumoniae AY186049	2	233	−1.4	−1.6	59	166	−1.4	−2.0	61	333	−1.7	−2.2	59	456	−7.5	−6.8	59	UC	NA	NA	NA
Salmonella enteritidis AY186056	2	235	−2.0	−1.5	58	168	−1.7	−2.0	61	335	−2.8	−2.3	59	458	−7.3	−6.8	58	389	−5.5	−6.3	58
Salmonella typhimurium AY186057	2	235	−2.0	−1.6	58	168	−1.8	−1.7	61	335	−3.1	−1.7	59	458	−7.5	−2.2	58	389	−4.7	−6.3	58
Salmonella typhimurium AY186058	2	235	−2.1	−1.6	59	168	−1.9	−1.7	61	335	−3.0	−1.7	59	458	−6.8	−2.3	59	389	−4.6	−4.7	58
Serratia marcescens AY186053	2	235	−1.7	−3.0	59	168	−2.0	−1.7	62	335	−2.2	−1.7	59	458	−7.3	−2.6	59	389	−4.7	−4.7	58
Shigella sonnei AY186034	2	235	−1.0	−1.6	59	168	−1.2	−1.7	61	335	−1.5	−1.8	59	458	−7.1	−2.7	58	389	−4.9	−4.7	58
Enterococcus hirae AY186039	3	277	−1.3	−1.6	56	283	−1.3	−1.6	57	180	−1.6	−1.7	57	ND	NA	NA	ND	NA	NA	NA	NA
Lactobacillus acidophilus AY186042	3	157	−1.6	−1.6	60	208	−1.2	−1.6	60	UC	NA	NA	NA	143	−2.8	−2.7	58	UC	NA	NA	NA
Lactobacillus casei AY186043	3	284	−1.9	−2.1	56	290	−2.2	−1.6	56	UC	NA	NA	NA	UC	NA	NA	NA	UC	NA	NA	NA
Lactobacillus johnsonii AY186044	3	288	−1.3	−1.7	57	293	0.0	−1.6	57	UC	NA	NA	NA	151	−2.7	−2.7	58	UC	NA	NA	NA
Lactobacillus murinus AY186045	3	281	−3.0	−1.6	57	209	−1.9	−1.6	56	218	−3.4	−2.3	56	UC	NA	NA	NA	UC	NA	NA	NA
Lactobacillus paracasei AY186046	3	ND	NA	NA	NA	290	−3.0	−2.0	57	NA	NA	NA	NA	UC	NA	NA	NA	UC	NA	NA	NA
Proteus mirabilis AY186055	4	83	−2.1	−1.6	60	287	0.4	−1.6	58	335	0.2	−1.9	59	458	−5.2	−6.8	58	389	−2.4	−3.6	58
Proteus mirabilis AY186047	4	83	−2.2	−1.8	60	287	0.5	−1.7	58	335	0.1	−2.3	59	458	−5.5	−2.7	58	389	−3.0	−4.7	58
Proteus vulgaris AY186048	4	83	−2.1	−1.6	60	287	0.0	−1.7	58	175	−0.4	−1.6	58	458	−5.7	−6.9	58	389	−2.8	−3.7	58

TRF drift is an average of at least three replicates except where indicated by italics.
NA—Not applicable.
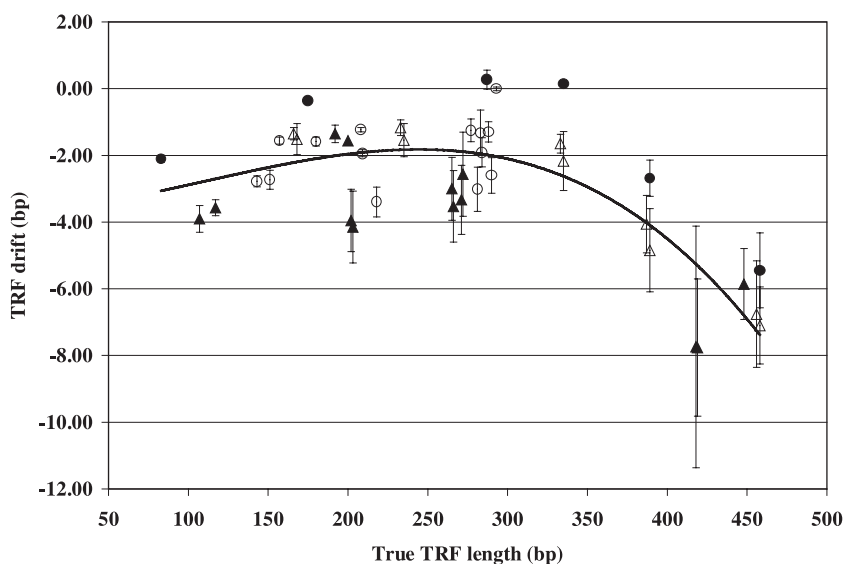ND—Not determined, fragment size <35 bp.
UC—Uncut.

Fig. 1. TRF drift for different phylogenetic groups used in the study. *Proteus* spp. had the least drift of any group at any length. Enterics and *Lactobacillus* spp. had similar TRF drift, while *Bacillus* spp. species had the most drift of any species used in this study. A 3rd order fit is indicated by solid line. *Proteus* spp., closed circle; Enterics, open circle, *Lactobacillus* spp., open triangle; *Bacillus* spp., closed triangle.

Differential migration is the variation between the internal ROX-labeled ladder and the FAM-labeled sample DNA presumably due to the ROX label having 12 more carbon atoms than the 6-FAM label (Applied Biosystems). The effect of this dissimilar migration manifested itself as progressively shorter observed TRFs as retention time in the capillary increased. In fact, fragment analysis software from some manufacturers (e.g. Beckman-Coulter) automatically compensates for differential dye migration. Using our data, this source of variation could also be corrected by using the equations below.

Predicted TRF Drift

$$= -2.24 \times 10^{-7}(\text{Observed TRF length})^3$$
$$+ 8.15 \times 10^{-5}(\text{Observed TRF length})^2$$
$$+ 1.39 \times 10^{-3}(\text{Observed TRF length}) - 3.48$$

$$(1)$$

Adjusted TRF length = Observed TRF length

$$- \text{Predicted TRF Drift} \quad (2)$$

Differential migration only accounted for 65% of the variation in TRF drift as determined by linear regression analysis. Secondary structure was not an additional source of TRF drift since fragment analysis was performed at 60 °C in a denaturing gel matrix. Additional sources of TRF drift were most obvious among organisms with the same true TRF length (Fig. 1). The trend in TRF drift was similar among related bacteria suggesting that sequence composition may affect TRF drift (Table 1). In fact, purine content was negatively correlated with TRF drift (p-value < 0.001). An additional 6% of the variation could be accounted for by incorporating the purine content of TRFs into the analysis. *Proteus* spp. had the least TRF drift at any length (∼ 2 bp), while *Bacillus* spp. had the most TRF drift (∼ 4 bp). Purine content across the entire dataset was 58% (± 2%). TRFs from *Proteus* spp. had an average purine content of 59% (± 1%) while *Bacillus* spp. had an average purine content of 57% (± 1%). A 1% difference in average purine content resulted in a 1-bp shift in average TRF drift for both *Proteus* spp. and *Bacillus* spp. This implies that subtle differences in molecular weight, whether from purine content or dye label, can significantly affect the observed TRF length.

The remaining variation observed in this dataset manifested itself as variation between observed TRF lengths in replicate runs of the same sample (Fig. 1, error bars). The primary cause of this variation was attributed to fluctuations in ambient temperature.

Alarming fluctuations of up to 5 bp were observed, suggesting that this source of variation could have unpredictable effects if a constant lab temperature is not maintained. The amount of TRF drift due to this source of variation may differ on other machines.

Using the equation and recommendations presented here, it is possible to minimize the effects of TRF drift. However, a certain amount of variation between true and observed TRF lengths remains. When matching observed TRF peaks to database predicted TRFs, one should include a window of at least ± 2 bp on the ABI 310 Genetic Analyzer. Other machines may allow for a more conservative window of ± 1 bp. An effective method of narrowing the number of database matches returned when using a large matching window is to employ separate enzyme digests of the same sample (Braker et al., 2001; Kaplan et al., 2001; Moeseneder et al., 2001).

References

Braker, G., Ayala-del-Rio, H.L., Devol, A.H., Fesefeldt, A., Tiedje, J.M., 2001. Community structure of denitrifiers, bacteria, and archaea along redox gradients in Pacific Northwest marine sediments by terminal restriction fragment analysis of amplified nitrite reductase (nirS) and 16S rRNA genes. Appl. Environ. Microbiol. 67, 1893–1901.

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

Kaplan, C.W., Astaire, J.C., Sanders, M.E., Reddy, B.S., Kitts, C.L., 2001. 16S Ribosomal DNA terminal restriction fragment pattern analysis of bacterial communities in feces of rats fed *Lactobacillus acidophilus* NCFM. Appl. Environ. Microbiol. 67, 1935–1939.

Kitts, C.L., 2001. Terminal restriction fragment patterns, a tool for comparing microbial communities and assessing community dynamics. Curr. Issues Intest. Microbiol. 2, 17–25.

Liu, W., Marsh, T.L., Cheng, H., Forney, L., 1997. Characterization of microbial diversity by determining terminal restriction fragments of genes encoding 16S rRNA. Appl. Environ. Microbiol. 63, 4516–4522.

Marsh, T.L., 1999. Terminal restriction fragment (T-RFLP), an emerging method for characterizing diversity among homologous populations of amplification products. Curr. Opin. Microbiol. 2, 323–327.

Moeseneder, M.M., Winter, C., Arrieta, J.M., Herndl, G.J., 2001. Terminal-restriction fragment (T-RFLP) screening of a marine archaeal clone library to determine the different phylotypes. J. Microbiol. Methods 44, 159–172.

Osborn, A.M., Moore, E.R.B., Timmis, K.N., 2000. An evaluation of terminal-restriction fragment (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ. Microbiol. 2, 39–50.

Sakano, Y., Pickering, K.D., Strom, P.F., Kerkhof, L.J., 2002. Spatial distribution of total, ammonia-oxidizing, and denitrifying bacteria in biological wastewater treatment reactors for bioregenerative life support. Appl. Environ. Microbiol. 68, 2285–2293.