

QUANTITATIVE ANALYSIS OF TERMINAL RESTRICTION FRAGMENT LENGTH POLYMORPHISM (T-RFLP) MICROBIAL COMMUNITY PROFILES: PEAK HEIGHT DATA SHOWED TO BE MORE REPRODUCIBLE THAN PEAK AREA

Roberto A. Caffaro-Filho^{1*}; Fabiana Fantinatti-Garboggini²; Lucia R. Durrant¹

¹Universidade Estadual de Campinas, Departamento de Ciência de Alimentos, Faculdade de Engenharia de Alimentos, Campinas, SP, Brasil; ²Universidade Estadual de Campinas, Divisão de Recursos Microbianos, Centro Pluridisciplinar de Pesquisas Químicas, Biológicas e Agrícolas, Paulínia, SP, Brasil

SHORT COMMUNICATION

ABSTRACT

Terminal Restriction Fragment Length Polymorphism (T-RFLP) is a culture-independent fingerprinting method for microbial community analysis. Profiles generated by an automated electrophoresis system can be analysed quantitatively using either peak height or peak area data. Statistical testing demonstrated that peak height data showed to be more reproducible than peak area data.

Key words: Terminal Restriction Fragment Length Polymorphism (T-RFLP); Molecular analysis; Microbial community

Terminal Restriction Fragment Length Polymorphism (T-RFLP) analysis measures the size polymorphism of terminal restriction fragments from a PCR amplified marker, generally the 16S rRNA (10). The primers used for amplification are fluorescently tagged. Subsequent digestion with restriction endonucleases produces different terminal fragments (also called operational taxonomic units) appropriate for sizing on automated electrophoresis systems that provide digital output. T-RFLP analysis of small subunit rRNA genes is currently one of the most powerful methods in microbial ecology for rapidly comparing microbial community structures of environmental samples and for the study of community dynamics and changes in community structure in response to changes in prevailing physicochemical parameters. The data obtained can also be compared with data from *in silico* analyses of sequence databases to infer the potential species composition of samples (5).

Besides high throughput, one of the major advantages of this fingerprinting technique is that even complex T-RFLP profiles data can be satisfactorily quantified by automated electrophoresis systems, making them amenable to analyses

using various statistical methods, such as similarity indices, hierarchical clustering algorithms, ordination methods, and self-organizing maps (5). While qualitative analyses of T-RFLP profiles only take into account the presence/absence of fragments, quantitative analyses can be achieved by considering peak height or peak area as a measure of fragment abundance. These are usually transformed to relative abundance instead of using the raw data.

Each step involved in the molecular analysis of environmental samples is a source of bias (15), which applies to T-RFLP. Thus, reproducibility is crucial since there's likely to be great numbers of samples to be analysed, which would make the practice of replicate analysis for all samples a major drawback. Osborn *et al.* (12) demonstrated that T-RFLP is a robust and reproducible methodology. They also highlighted parameters that can significantly influence variability. Several studies also evaluated the reproducibility of T-RFLP profiles (3,4,8,9). Some of them used peak height data while others used peak area. This choice can also influence variability, something that Osborn *et al.* (12) didn't assess since they only used peak

*Corresponding Author. Mailing address: Microbial Systematics and Physiology Laboratory, Food Sciences Department, Faculty of Food Engineering, State University of Campinas, Campinas, Brazil. Tel.: (19) 3521-2173. E-mail: caffaro.filho@gmail.com

height in their study. In a discussion published in 2003, there was no consensus on whether T-RFLP peak height or area should be analyzed (6). While Using peak heights would down weight longer fragments because of diffusion during electrophoresis, overlapping peaks are not deconvoluted by the software used in the automated sequencer, resulting in an artificial alteration of area based on proximity to other peaks. This discussion comes all the way from the time when only gel electrophoresis was used in T-RFLP analysis. More recently, capillary and multicapillary electrophoresis has been the choice for higher sample throughput and highly precise determination of fragment lengths (11,14). Even then, there's still no consensus. While the use of peak area was recommended (14), some prefer to use peak heights (2,7).

While trying to obtain a better signal from a multicapillary sequencer in the analysis of an activated sludge bacterial community, the same sample was analyzed and further concentrated several times. Peak height and peak area data were retrieved for all the runs. Since the relative abundance of amplicons is the same, no matter what the concentration is, it was possible to assess the variability imposed by the choice of peak height or peak area for further data analysis. Hypothesis testing (F statistic) was used to know if one or the other would show a greater variability. If then, the one showing less variability should always be chosen.

A sample (20 mL) from the mixed liquor of an activated sludge pilot plant was centrifugated and maintained at -20°C until

analysis. The DNA extraction was made according to the method described by Smalla *et al.* (13) using a Mini-BeadBeater (Biospec Products) and DNA was quantified by electrophoresis with ethidium bromide staining. After purification with GFX PCR DNA and gel band purification kit (GE Health Care), the community 16S rRNA genes were amplified by PCR amplification, made in a iCycler thermocycler (Bio-Rad) with a total reaction volume of 50 µL. Each reaction contained 20 pmol of each primer, 200 µM of each deoxiribonucleotide triphosphate, 1,5 mM of MgCl₂ (Invitrogen), 2 U of Taq DNA polymerase (Invitrogen), and 1/10 PCR buffer (10X). After 5 minutes denaturation at 95°C, there were 30 cycles of: denaturation (1 min, 95°C), annealing (1 min, 57°C) and extension (3 min, 72°C), with a final extension step of 7 minutes at 72°C. The primers were 8f (5'-AGAGTTTGA TCCTGGCTCAG-3'), FAM labeled at the 5' end, and 1407r (5'-ACGGGCGGTGTGTACA-3') (Applied Biosystems). These are universal for the domain Bacteria (1). The amplification products were digested with restriction enzyme *Hae*III and loaded on a MegaBace 1000 (Amersham Biosciences) multicapillary automated sequencer with ET 550 size standard and loading solution, on genotyping mode. Four increasing concentrations of the sample were analysed by decreasing sample dilution in the loading solution on the sequencer well.

Two fragments (189 and 208) were detected in all four runs and signals were used on statistical analysis. The results are presented on Table 1. The expected was that their relative abundances in relation to each other should be constant despite different concentrations. This could be measured either with relative peak height or relative peak area. Relative abundance was calculated in percentage by dividing raw peak signal by the total signal.

Results showed at first glance that relative abundances were more variable when considering peak area than when analyzing peak height. Table 2 shows the calculated means, variances (S^2) and standard deviations (in %) of the relative abundances.

ANOVA was used to test variances, assuming a normal distribution for both considered populations (relative height and relative area). The significance level adopted was $\alpha = 0.01$. Statistical testing was made by dividing the variance of relative area by the variance of relative height ($F_0 = S^2_a / S^2_h$). The null hypothesis (equal variances) was rejected.

Thus, relative peak height variance of a fragment is inferior to the relative peak area variance of the same fragment. In other words, peak height data showed to be more reproducible than

Table 1. Raw and relative peak height and area signals for fragments 189 and 208. Raw signals in fluorescence units. Concentrations: 1 > 2 > 3 > 4.

Concentration	Fragment	height	area	Relative height (%)	Relative area (%)
1	189	237	133.5	68.70	63.94
	208	108	75.3	31.30	36.06
2	189	200	151.8	69.44	71.23
	208	88	61.3	30.56	28.77
3	189	143	101.4	70.79	76.13
	208	59	31.8	29.21	23.87
4	189	133	76.6	70.00	68.21
	208	57	35.7	30.00	31.79

Table 2. Statistical values calculated for relative abundances. n=4 for both relative height and area.

Fragment	Relative height mean	Variance (S^2_h)	Standard deviation (%)	Relative area mean	Variance (S^2_a)	Standard deviation (%)
189	69.73	0.78	1.27	69.88	26.32	7.34
208	30.27	1.18	3.59	30.12	39.46	20.85

peak area, and should be preferred in quantitative T-RFLP data analysis unless proved otherwise. This is not a definite conclusion, but rather a call for scientists to give thought to this potential source of variation and not use arbitrarily the data generated in T-RFLP profiles for quantitative analysis. That would be particularly valuable in Brazil where the use of T-RFLP is still crawling but is likely to blossom soon.

RESUMO

Análise quantitativa de perfis de T-RFLP de comunidades microbianas: dados de altura de picos mostraram-se mais reprodutíveis do que os de área

Terminal Restriction Fragment Length Polymorphism (T-RFLP) é um método molecular, independente de cultivo, para análise de comunidades microbianas. Perfis gerados por um sistema automatizado de eletroforese podem ser analisados quantitativamente usando dados de altura ou área dos picos. Os dados de altura mostraram-se mais reprodutíveis do que os de área.

Palavras-chave: Terminal Restriction Fragment Length Polymorphism (T-RFLP); Análise molecular; Comunidades microbianas

REFERENCES

1. Amann, R.I.; Ludwig, W.; Schleifer, K.H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.*, 59: 143-169.
2. Dobretsov, S.; Dahms, H.; Qian, P. (2005). Antibacterial and anti-diatom activity of Hong Kong sponges. *Aquat. Microb. Ecol.*, 38: 191-201.
3. Dollhopf, S.L.; Hashsham, S.A.; Tiedje, J. (2001). Interpreting 16S rDNA T-RFLP data: application of self-organizing maps and principal component analysis to describe community dynamics and convergence. *Microb. Ecol.*, 42: 495-505.
4. Dunbar, J.; Ticknor, L.O.; Kuske, C.R. (2001). Phylogenetic specificity and reproducibility and new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from bacterial communities. *Appl. Environ. Microbiol.*, 67: 190-197.
5. Forney, L.J.; Zhou, X.; Brown, C.J. (2004). Molecular microbial ecology: land of the one-eyed king. *Curr. Opinion Microbiol.*, 7: 210-220.
6. Grant, A.; Ogilvie, L.A. (2003). Terminal restriction fragment length polymorphism data analysis. *Appl. Environ. Microbiol.*, 69: 6342-6343.
7. Hoshino, T.; Terahara, T.; Yamada, K.; Okuda, H.; Suzuki, I.; Tsuneda, S.; Hirata, A.; Inamori, Y. (2006). Long-term monitoring of the succession of a microbial community in activated sludge from a circulation flush toilet as a closed system. *FEMS Microbiol. Ecol.*, 55: 459-470.
8. Lueders, T.; Friedrich, M.W. (2002). Effects of amendment with ferrihydrite and gypsum on the structure and activity of methanogenic populations in rice field soil. *Appl. Environ. Microbiol.*, 68: 2484-2494.
9. Lukow, T.; Dunfield, P.F.; Liesack, W. (2000). Use of the T-RFLP technique to assess spatial and temporal changes in the bacterial community structure within an agricultural soil planted with transgenic and non-transgenic potato plants. *FEMS Microbiol. Ecol.*, 32: 241-247.
10. Marsh, T.L. (1999). Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. *Curr. Opinion Microbiol.*, 2: 323-327.
11. Moeseneder, M.M.; Winter, C.; Herndl, G.J. (2001). Horizontal and vertical complexity of attached and free-living bacteria of the eastern mediterranean sea, determined by 16S rDNA and 16S rRNA fingerprints. *Limnol. Oceanogr.*, 46: 95-107.
12. Osborn, A.M.; Moore, E.R.; Timmis, K.N. (2000). An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environ. Microbiol.*, 2: 39-50.
13. Smalla, K.; Cresswell, N.; Mendonca-Hagler, L.C.; Wolters, A.; van Elsas, J.D. (1993). Rapid DNA extraction protocol from soil for polymerase chain reaction-mediated amplification. *J. Appl. Bacter.*, 74: 78-85.
14. Trotha, R.; Reichl, U.; Thies, F.L.; Sperling, D.; König, W.; König, B. (2002). Adaption of a fragment analysis technique to an automated high-throughput multicapillary electrophoresis device for the precise qualitative and quantitative characterization of microbial communities. *Electrophor.*, 23: 1070-1079.
15. Wintzingerode, F.V.; Göbel, U.B.; Stackbrandt, E. (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.*, 21: 213-229.