Published in Environmental Microbiology 4, issue 11, 634-643, 2002
which should be used for any reference to this work

Minireview

Statistical analysis of denaturing gel electrophoresis (DGE) fingerprinting patterns

N. Fromin,[1]* J. Hamelin,[1] S. Tarnawski,[1] D. Roesti,[1] K. Jourdain-Miserez,[1] N. Forestier,[1] S. Teyssier-Cuvelle,[1] F. Gillet,[2] M. Aragno[1] and P. Rossi[1]

[1]*Microbiology Laboratory, and* [2]*Plant Ecology Laboratory, University of Neuchâtel, PO Box 2, CH-2007 Neuchâtel, Switzerland.*

Summary

Technical developments in molecular biology have found extensive applications in the field of microbial ecology. Among these techniques, fingerprinting methods such as denaturing gel electrophoresis (DGE, including the three options: DGGE, TGGE and TTGE) has been applied to environmental samples over this last decade. Microbial ecologists took advantage of this technique, originally developed for the detection of single mutations, for the analysis of whole bacterial communities. However, until recently, the results of these high quality fingerprinting patterns were restricted to a visual interpretation, neglecting the analytical potential of the method in terms of statistical significance and ecological interpretation. A brief recall is presented here about the principles and limitations of DGE fingerprinting analysis, with an emphasis on the need of standardization of the whole analytical process. The main content focuses on statistical strategies for analysing the gel patterns, from single band examination to the analysis of whole fingerprinting profiles. Applying statistical method make the DGE fingerprinting technique a promising tool. Numerous samples can be analysed simultaneously, permitting the monitoring of microbial communities or simply bacterial groups for which occurrence and relative frequency are affected by any environmental parameter. As previously applied in the fields of plant and animal ecology, the use of statistics provides a significant advantage for the non-ambiguous interpretation of the spatial and temporal functioning of microbial communities.

E-mail nathalie.fromin@unine.ch; Tel. (+41) 32 718 22 30;
Fax (+41) 32 718 22 31.

Fingerprinting techniques applied to microbial communities

Molecular approaches in microbial ecology

A major challenge in the field of microbial ecology is to assess the diversity of the microbial cells present in a defined habitat. Assessing the diversity of microbial communities (in terms of richness and structure) is a way to address how they evolve in their environment. In a more general way, it is a possible means to address the question of the modulation of microbial communities by environmental factors. Phylogenetically meaningful macromolecules, particularly 16S rDNA directly amplified from environmental DNA, are now widely used for such purposes (Ranjard *et al.,* 2000a; O'Donnell *et al.,* 2001; Schäfer and Muyzer, 2001).

However, information collected by these molecular tools quickly revealed the unsuspected complexity of whole bacterial communities (Ward *et al*., 1990). They were shown in turn to be limited in a practical way (O'Donnell *et al.,* 2001). The amount of time and resources needed for the now classical 'cloning-sequencing' technique (which potentially supply an exhaustive description of microbial communities), coupled with the impracticability of complete counts of organisms at present (Dunbar *et al.,* 2002), led to the development of alternative solutions. An original way was to separate polymerase chain reaction (PCR)-amplified fragment pools produced from whole microbial communities by electrophoresis techniques (Table 1). These fingerprinting methods are now widely adopted in the field of bacterial ecology and permit the simultaneous analysis of numerous samples (Ferrari and Hollibaugh, 1999).

DGE fingerprinting of microbial communities

Muyzer *et al.* (1993) first applied denaturing gel electrophoresis (DGE) techniques for the analysis of whole bacterial communities. Denaturing gel electrophoresis allows

Table 1. Fingerprinting methods used for the characterization of microbial communities, with recent publications in the related field.

Amplified ribosomal DNA restriction analysis (ARDRA)	Smit *et al.* (1997); Tiedje *et al.* (1999)
Denaturing gel electrophoresis (DGE)	Muyzer and Smalla (1998)
Ribosomal intergenic spacer analysis (RISA)	Fisher and Triplett (1999); Ranjard *et al.* (2000b)
Single strand conformation polymorphism (SSCP)	Schwieger and Tebbe (1998); Dabert *et al.* (2001)
Terminal restriction fragment length polymorphism (T-RFLP)	Moeseneder *et al.,* (1999); Dollhopf *et al.* (2001)

the separation of small polymerase chain reaction products, commonly up to 400 bp. The separation of the DNA fragments is achieved as a function of their different G + C content and distribution. Consequently, the fingerprinting pattern is built according to the melting behaviour of the sequences along a linear denaturing gradient (Myers *et al.*, 1985). Such a gradient is obtained using either denaturing chemicals for denaturing gradient gel electrophoresis (DGGE) or heat for temperature gradient gel electrophoresis (TGGE) and temporal temperature gradient electrophoresis (TTGE).

The DGE techniques were applied using 16S rDNA fragments to the analysis of bacterial communities in numerous habitats such as soil and rhizosphere (Bruns *et al.*, 1999; Yang and Crowley, 2000; Duineveld *et al.,* 2001; Ibekwe *et al.,* 2001; McCaig *et al.,* 2001; Smalla *et al.,* 2001) and aquatic environments (Murray *et al.*, 1996; 1998; Ferrari and Hollibaugh, 1999; Moeseneder *et al.*, 1999; van der Gucht *et al.,* 2001; Schäfer and Muyzer, 2001; Schäfer *et al.,* 2001). Interestingly, an increasing number of studies based on DGE are carried out on archaeal (Murray *et al.*, 1998; Rölling *et al.,* 2001) or eukaryal communities (van Hannen *et al.*, 1999; van Elsas *et al.,* 2000; Diez *et al.,* 2001; Mohlenhoff *et al.,* 2001).

The sensitivity of DGE analysis can be refined with the targeting of precise (and even non-dominant) taxonomic groups, by using specific PCR primers (Heuer *et al.*, 1997; Nübel *et al.*, 1997; Heilig *et al.,* 2002) or by identifying community members by hybridization of blotted DGE gels with group-specific oligonucleotide probes (Heuer *et al.*, 1999). Other developments were based on the use of 16S rRNA as a target (Felske and Akkermans, 1998; Kowalchuk *et al.*, 1999; Duineveld *et al.,* 2001; Schäfer *et al.,* 2001) to highlight metabolically active populations only. Functional genes (Watanabe *et al.*, 1998; Bruns *et al.*, 1999; Lovell *et al.,* 2000; Fjellbirkeland *et al.,* 2001) or even their transcripts (Wawer *et al.*, 1997) were also analysed, which heralds very interesting prospects in clarifying the functioning of microbial communities.

Guidelines for the interpretation of DGE fingerprinting patterns

Some features of the fingerprinting techniques have to be considered before applying statistics for the analysis of DGE profiles.

In DGE analysis, the generated banding pattern is considered as an 'image' of the whole bacterial community. An individual discrete band refers to an unique 'sequence type' or phylotype (Muyzer *et al.*, 1995; van Hannen *et al.*, 1999), which is treated in turn as a discrete bacterial population. The term population classically refers to a group of bacterial cells present in a specified habitat and belonging to the same species. We are expecting that PCR fragments generated from a single population to display identical electrophoretic mobility in the analysis. This was confirmed by Kowalchuk *et al.* (1997) who showed that co-migrating bands generally corresponded to identical sequences. However, it was shown that rDNA fragments of closely related bacteria are not necessarily resolved (Buchholz-Cleven *et al.*, 1997) or may produce separated bands (Jackson *et al.,* 2001). Moreover, non-related sequences might co-migrate at an identical position (Vallaeys *et al.*, 1997), especially when treating complex community patterns (Kowalchuk *et al.*, 1997; Ben Omar and Ampe, 2000). In this case, the question of the resolution of the gel needs to be addressed. Crowding of the gel has been discussed already and algorithms to assess it were proposed by Nübel *et al.* (1999a). Degenerated primers should be avoided also as one single bacterial strain, or even a single clone, may generate a multiple band pattern (Kowalchuk *et al.*, 1997; Piceno *et al.*, 1999). Some authors have also detected artificial bands when analysing complex DNA templates, probably induced by heteroduplex molecules (Ferris and Ward, 1997). Consequently, care should be taken in assigning a single band to a single bacterial population.

Another assumption for DGE fingerprinting interpretation is that the band intensity is directly related to the density of corresponding bacterial phylotypes within the sample. Results obtained by Murray *et al.* (1996) suggested a relationship between band intensity and relative abundance of the corresponding phylotype in the template DNA mixture. Such an assumption implies that no bias was obtained during the whole extraction–amplification procedure of the bacterial genomes (Muyzer *et al.*, 1993; Wang and Wang, 1997; Garcia-Pichel *et al.,* 2001). The DGE analysis should probably be restricted to samples treated using identical methods, in which DNA extraction and amplification biases are supposed to occur homogeneously. Moreover, it is commonly accepted that the main populations only (those representing more than 0.1–1% of the target organisms in terms of relative proportion) are

displayed in the profiles (Muyzer *et al*., 1993; Murray *et al*., 1996). As a result, all populations present within a habitat do not appear on DGE banding patterns. When assessing the above considerations, the image of the communities which is provided by DGE fingerprinting patterns probably relates more to its structure, i.e. to the relative abundance of the main bacterial populations, than to its total richness (Muyzer and Smalla, 1998). These features and restrictions are nevertheless common to all PCR-based approaches (Lee *et al*., 1996; Fisher and Triplett, 1999; Schäfer and Muyzer, 2001).

The last consideration about this analytical technique is about the reproducibility of the DGE analysis. Reproducibility of sample analysis depends on the upstream analytical steps (from the sampling to the DNA extraction and amplification procedures) as well as the care brought to the DGE gels themselves. A thorough standardization at each level of the experiments results in very high reproducibility. The use of reference patterns, the loading of precise amounts of PCR-amplified fragments and the precision of gel staining are required. As a consequence, identical samples loaded on a single gel display identical patterns (Simpson *et al*., 1999; Schäfer *et al.,* 2001; Yang *et al.,* 2001) and patterns from different gels can be compared with a high degree of confidence. The analysis of large numbers of samples can be exploited for the characterization of the intrinsic variability of the bacterial community structures. This large amount of data can be analysed in turn with statistical tools, which provide a significant advantage for the non-ambiguous interpretation of the observed variability (Morris *et al*., 2002).

Analysis and comparison of DGE community profiles

Denaturing gel electrophoresis techniques have been extensively used to monitor bacterial communities in space and time (Ferris and Ward, 1997; Murray *et al*., 1998; Nübel *et al*., 1999b; van der Gucht *et al.,* 2001) or to evaluate the impact of environmental disturbances (Ibekwe *et al.,* 2001; Müller *et al.,* 2001). The variations between DGE profiles were classically described visually on a single DGE gel by the disappearance, the appearance or the changes in the intensity of selected bands. However, an increasing number of studies propose statistical investigations of DGE banding patterns, which undoubtedly lead to refined results. These advanced analyses are based on a computer-assisted characterization of the banding patterns and the subsequent treatment of the data using a statistical approach.

An example of computer-assisted guideline for the analysis of fingerprinting profiles was proposed by Rademaker *et al*. (1999) using the GelCompar software package (Applied Maths, Kortrijk, Belgium). Briefly, banding patterns were first standardized with a reference pattern included in all gels. Each band was described by its position (*Y,* in pixel on the image file) and its relative intensity in the profile (P_i), which could be calculated by the relative surface of the peak in the profile ($P_i = n_i/N$, where n_i is the surface of the peak *i*, and *N* is the sum of the surfaces for all the peaks within the profile). Using these data various statistical methods can be applied, based either on single band or on whole DGE profile analysis.

Analysis of DGE profiles based on single bands

One way to analyse DGE fingerprinting patterns is to observe the possible changes in the presence/absence or in the variation of intensity of a single band (Murray *et al*., 1996). Putative indicator bands highlighted in this way can be excised from the gels and their sequences analysed using a cloning–sequencing procedure (Kowalchuk *et al*., 1997; Watanabe *et al*., 1998; Ibekwe *et al.,* 2001).

The variation in band presence or intensity can be exploited in two different ways. First of all, the relevance of indicator bands can be evaluated by testing their occurrence in relation with various biological and physicochemical parameters (Widmer *et al.,* 2001) as well as with the presence or absence of other bands in the patterns. In the example shown in Table 2 16S rDNA TTGE banding patterns of 30 raw milk samples were analysed in this way. The occurrence of each TTGE band was tested against qualitative descriptors using a Fisher's exact test and bands found at the positions $Y = 230$ and $Y = 300$ were positively correlated to the cleaning frequency of the milking device and to the hygienic status of the cow tits respectively.

Second, single band analysis can also be used for computing a regression between band intensity (dependent quantitative variable) and an environmental descriptor (independent quantitative variable). In the example

Table 2. Significant correlation ($P < 0.05$, Fisher's exact test) between the presence of a selected band within a gel pattern and a qualitative descriptor. The bands were identified using a cloning-sequencing procedure (P. Rossi, unpublished data).

	Descriptors		
Position of the band (in pixels on Y axis)	Frequency of cleaning of the milking device	Hygienic status of the cow tits before milking	Identification of 16S rDNA fragment (% identity)
Y = 230	$P = 0.0001$	No correlation	*Bacillus* sp. (>95%)
Y = 300	No correlation	$P = 0.004$	*Pseudomonas* sp. (> 95%)

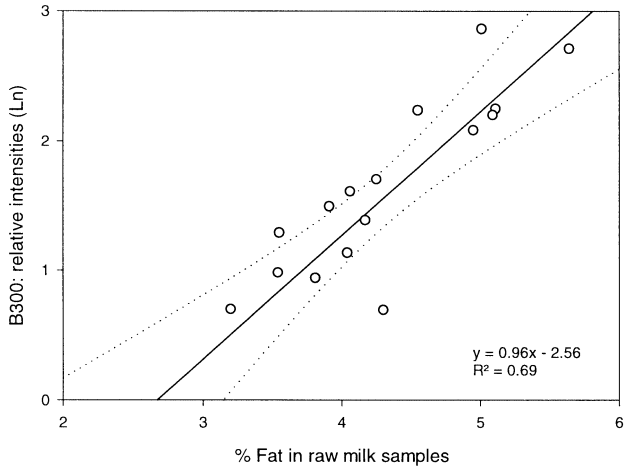

Fig. 1. Regression analysis between the relative intensity (ln) of the band at the $Y = 300$ position and the percentage of fat found in the corresponding raw milk samples (P. Rossi, unpublished data).

given above, the TTGE patterns were analysed by plotting the relative intensities (P_i) of each band versus various physical parameters measured from the same samples. A positive correlation ($R^2 = 0.69$) was found between the relative intensity of the band $Y = 300$ (identified as *Pseudomonas* sp.*;* Table 2) and the per cent of fat measured in the raw milk (Fig. 1).

Whole profile analysis

The second approach for a comparative analysis of DGE patterns is based on the whole set of bands present within the profiles. The total number of bands (called sometimes 'band richness') in each sample pattern is related to the number of dominant phylotypes, and can be used for comparison purposes (Müller *et al.,* 2001; van der Gucht *et al.,* 2001). Comparison of profiles can be refined by taking into account the relative intensity of each band (P_i). Thus, diversity indices, such as Shannon-Weaver and Evenness indices (Nübel *et al*., 1999a; Simpson *et al*., 1999; Kocherginskaya *et al.,* 2001; McCaig *et al.,* 2001; Ogino *et al.,* 2001), can be calculated to describe possible changes in the dominance among phylotypes. An interesting feature is to combine these indices with other sets of environmental data. For instance, Nübel *et al*. (1999a) found a positive linear correlations between Shannon-Weaver indices calculated from both DGE patterns and carotenoid types in oxygenic-phototrophic microbial communities.

Computation of similarity matrix

When considering the presence/absence of the bands, similarities between banding patterns, taken in pairs, can be expressed as a percentage value of a similarity coefficient such as Jaccard (Diez *et al.,* 2001) or Dice (van der Gucht *et al.,* 2001) coefficient, or a distance coefficient such as Euclidean measure (McSpadden Gardener and Lilley, 1997). Other coefficients, such as the Steinhaus coefficient (Fig. 2) or the product moment, also named Pearson correlation coefficient (Rölling *et al.,* 2001; Smalla *et al.,* 2001), allow to take into consideration the relative intensity (P_i) of each band (Legendre and Legendre, 1998; Rademaker *et al*., 1999). As noticed by Murray *et al*. (1998), the use of these similarity coefficients for the calculation of pair-wise levels of similarity between patterns does not require a one-to-one correspondence between the number of bands and the number of

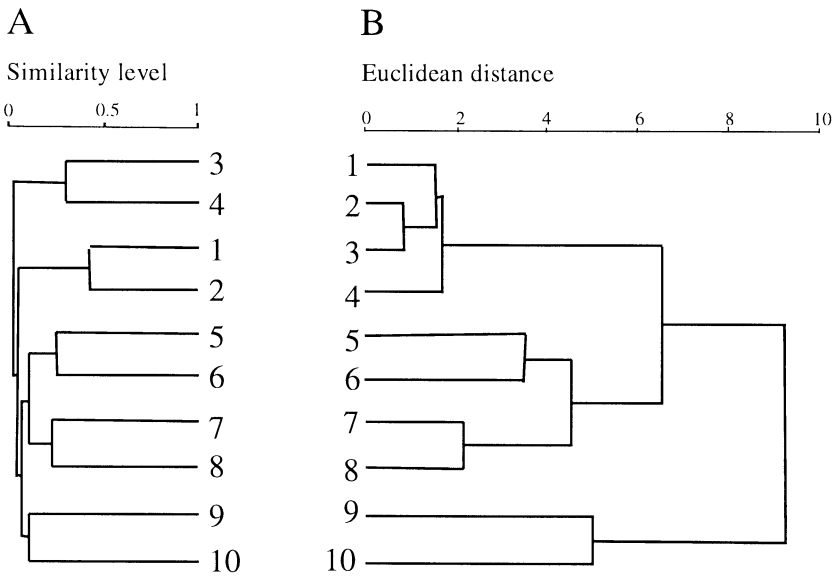

Fig. 2. UPGMA clustering of 10 samples taken along a vertical gradient form the small eutrophic Lake Loclat (Neuchâtel, Switzerland). Samples are ranked by depth: 1, corresponds to the surface; 10, to the bottom of the lake (8.7 m).
A. Clustering according to DGGE data (using Steinhaus coefficient).
B. Clustering according to 23 physical and chemical variables using Euclidean distance (Forestier *et al.,* 2002). Linkage levels were computed using the R package (Casgrain and Legendre, 2001).

sequence types. Similarity or distance matrices can be displayed graphically as a dendrogram, but also give way to clustering and ordination methods.

Clustering techniques

Clustering techniques, such as the unweighted pair group method using arithmetic averages (UPGMA), are applied to the DGE profiling with the aim of identifying the samples which generate similar patterns (Ibekwe *et al.,* 2001; Yang *et al.,* 2001; Boon *et al.,* 2002). One advantage of this presentation is that the coherence of the fingerprinting patterns can be assessed rapidly.

In the example given above (Forestier *et al.,* 2002), 10 samples were taken from a holomictic eutrophic lake along a vertical gradient and were analysed for major ions, organic content, physical parameters and DGGE analysis of 16SrDNA fragment genes. Computation of the DGGE and environmental parameters matrices was carried out using the Euclidean distance and Steinhaus coefficient, respectively, and UPGMA was selected as a clustering method for the presentation of the results. The resulting dendrograms (Fig. 2) showed that the samples were clustered according to the depth of their sampling, in agreement with measured physical and chemical parameters.

Ordination methods

Another way of analysing DGE profiles is to bring out major tendencies of the variance of the samples for the whole set of descriptors using multivariate ordination methods. Legendre and Legendre (1998) provide a excellent review of these methods which are commonly used in the field of ecology. These methods are used for the integration of complex sets of data (i.e. bands in the DGE patterns) into new mathematical variables which can be projected into a few-dimension perspective (reduced space). The major advantage of these methods is to display the whole set of samples on a simple scheme, and to highlight the possible descriptors which are governing their dispersion (ter Braak *et al*., 1995). Of course, true correlation can only be deduced when sufficient amounts of data are provided: the results of proposed statistical analysis should be considered with care, as coincidence or convergence mechanisms cannot be excluded.

Common ordination methods include non-metric multidimensional scaling (NMDS), principal component analysis (PCA), correspondence analysis (CA), canonical variate analysis (CVA) and canonical correspondence analysis (CCA). Several complementary statistical procedures can be applied to analyse DGE data (Yang *et al.,* 2001). Details on the specific underlying theory of each of these methods can be found elsewhere (McSpadden Gardener and Lilley, 1997; Legendre and Legendre, 1998).

Non-metric multidimensional scaling is an ordination method which reduces complex DGE patterns to a point in a two-dimensional space. By connecting the consecutive points, the relative changes in the bacterial community can be visualized. van Hannen *et al*. (1999) proposed to calculate Nei-Li distances from the binary data resulting from DGE profile analysis and to represent these distances using this ordination method. The authors showed that bacterial communities that developed on two distinct detritus substrates differed significantly: the distances calculated between communities from different substrates were greater ($P < 0.05$) than the distances calculated between the replicates for a given substrate. Non-metric multidimensional scaling was used elsewhere for the interpretation of DGE data (Diez *et al.,* 2001; Schäfer *et al.,* 2001). The advantage of NMDS is to represents the objects in two or three dimensions, with dissimilar objects far apart and similar objects close to one another in the ordination space.

Principal component analysis generates new variables, called principal components (linear components of the original variables), which explain the highest dispersion of the samples. This method was often used for the interpretation of DGE community fingerprinting analysis (van der Gucht *et al.,* 2001; Müller *et al.,* 2001; Ogino *et al.,* 2001; Yang *et al.,* 2001). As an example, Müller *et al*. (2001) used PCA to compare 16S rDNA DGGE profiles for bacterial communities present in mercury-contaminated soils. Their investigations showed that the DGGE approach generated more distinctive results than colony morphotyping and substrate utilization. van der Gucht *et al*. (2001) showed that the composition of bacterioplanktonic communities differed between two lakes and during seasons using a PCA applied to presence/absence of bands within 16S rDNA DGGE patterns. Using Spearman's rank correlation, the observed seasonal variations were found to be positively correlated with environmental variables such as temperature, nitrate concentration or microbial biomass. However, PCA is probably not the most suitable statistical approach for analysing DGE patterns, as its underlying model assumes that biological populations have a linear response curve along the axes of ecological variation. Niche theory tells us that populations have ecological preferences. An unimodal (i.e. bell-shaped) response distribution of the different bacterial populations present in a community is probably closer to reality, with more individuals near some optimal environmental values.

Correspondence analysis may be applied to any data table that is dimensionally homogenous. ter Braak (1985) showed that the underlying model was adapted to presence/absence or abundance data tables and consequently, that the analysis was well suited for populations

with unimodal distribution along environmental gradients. Using this statistical analysis, Jourdain-Miserez *et al.* (2001) analysed 16S rDNA gene fragments issued from milk samples on TTGE gels. The results clearly showed different community structures between organic and conventional farming practices (Fig. 3). Correspondence analysis was also used elsewhere for similar approaches (Ibekwe *et al.*, 2001; Yang *et al.*, 2001).

Interpretation of DGE patterns with environmental variables

From our point of view, the greatest opportunity of the statistical analysis of DGE patterns is offered when the community profiles are combined in a joint analysis with environmental data sets. The relevant question here is to know whether the variations observed between different banding patterns could be associated with the variations of measured environmental variables.

McCaig *et al.* (2001) applied multivariate analysis to reduce the original data for grassland DGGE community patterns into six principal components. They showed clear differences between improved and non-improved grassland communities using CVA. This method requires an *a priori* definition of groups and finds linear combinations of variables that maximize the ratio between-group variation to within-group variation.

The 'community matrix' obtained from DGE profiles can be tested also against a second matrix obtained from environmental data sets measured on the same samples. Canonical correspondence analysis is a powerful canonical ordination technique (multivariate direct gradient analysis) allowing the explanation of the structure of a 'species' data table by quantitative environmental descriptors and assuming unimodal distributions of species (ter Braak, 1986). Using this technique, Yang and Crowley (2000) compared bacterial rhizospheric communities associated to barley plants under iron-limiting and iron-sufficient growth conditions. As a result, they showed that about 40% of the variation between microbial communities could be attributed to plant iron nutritional status. Figure 4 presents a CCA of a DGGE analysis carried out on samples taken from the water column of Lake Loclat (Forestier *et al.*, 2002). In this case, five environmental variables were selected according to their high probability of correlation with the samples ($P < 0.05$) using Mantel tests (Mantel, 1967). This test is based on the linear correlation between two distance or similarity matrices obtained from independent data. As shown in the Fig. 4, the redox and the dissolved organic carbon were the variables which influenced mostly the dispersion of the samples. The first five samples taken from the aerobic zone (samples 1–5) are closely related, defining a homogenous bacterial structure. The samples taken from the anaerobic section of the water column (points 6–10) were displayed according to depth indicating a continuum of different bacterial communities.

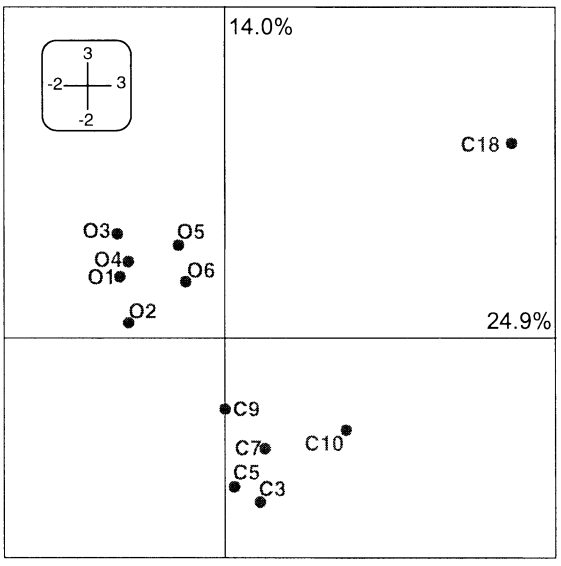

Fig. 3. Correspondence analysis of TTGE community profiles from milk samples from 12 farms (Switzerland). TTGE data of five samples taken on five consecutive days were pooled before being analysed (sum of unconstrained eigen values: 1.74). C, conventional; O, organic farming practices. Numbers refer to the different farms (Jourdain-Miserez *et al.,* 2001).

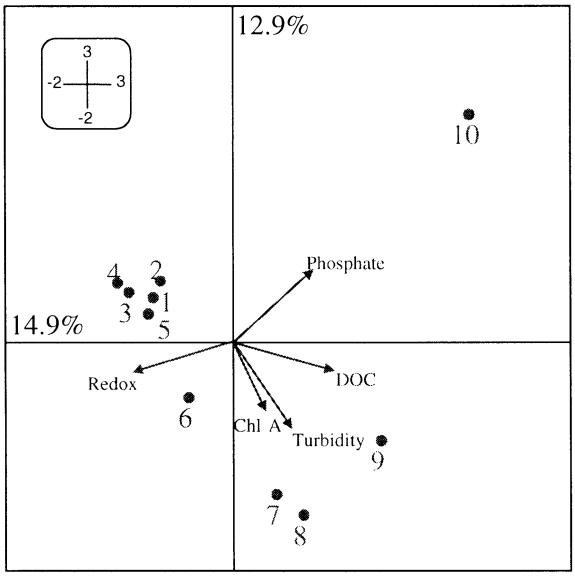

Fig. 4. Canonical correspondence analysis of microbial community patterns generated by 16S rDNA DGGE analysis for 10 water samples (Lake Loclat) ranked by depth (1 for surface to 10 at 8.7 m depth). The total inertia of the matrix was 4.45 and the selected variables explained 59% of the variance of the DGGE data set (sum of canonical eigenvalues: 2.62) (Forestier *et al.,* 2002).

Conclusions

Denaturing gel electrophoresis fingerprinting techniques are very effective methods for the characterization of bacterial community structures. These techniques are convenient for the simultaneous analysis of numerous samples. They are consequently well suited for the monitoring of whole communities, focusing on phylotypes for which the occurrence and/or the relative frequency are affected by any environmental change. As shown above, emphasis should be brought to the standardization of the whole analytical procedure as a means for increasing the reliability of the method and the reproducibility of the patterns. For instance computer-assisted analysis of the profiles should be generalized, escaping the merely qualitative reading of the fingerprinting patterns.

The exploratory aspect of the statistical techniques applied to DGE patterns that we present here is the consequence of statistical developments brought to the field of plant and animal ecology. It is now possible to approach causality in microbial ecology with statistical methods using experimental designs which were impossible to conceive a few years ago, principally because of the time and resources needed for the analysis of high number of samples. Examples provided above showed that it is possible to apply statistical tools to DGE data sets efficiently. The first result is in the validation of the interpretation of the patterns, such as shifts in the microbial community structure or the identification of key-populations which may be affected by changing conditions. Moreover, whole pattern data generated by the DGE analysis can be directly tested for correlation analysis against any single or combination of environmental sets of variables.

However, care should be taken in the choice of the statistical analytical procedure. As shown above, the underlying theoretical model should be carefully assessed before any attempt of application. Some analysis used up to now were probably not well suited to this type of data sets. On the contrary, CA is particularly well suited for abundance data sets, and PCA (normalized using correlation) is perfectly adapted for the analysis of environmental data sets (standardized descriptors).

The complementation of DGE analysis with a statistical approach leads to the definition of new hypotheses and to new prospects in terms of spatial or temporal functioning of microbial systems. Statistical methods reveal putative correlations between different sets of variables. They do not permit, however, conclusions to be drawn regarding the causality of these correlations. Therefore, statistical analyses should not be considered alone, but in a dialectic relationship with an ecological hypothesis. Automated pattern recognition and mechanistic dynamic modeling (combined with field and laboratory experiments) will probably very soon be the future steps in this field. In this sense, it will be conceivable to describe more precisely the relations between the observed diversity of the organisms and their ecological niches, leading to the development of the promising concept of ‘bacterial sociology’.

Acknowledgements

The authors would like to thank Camilla Rusca and Simona Casati, Noémie Matile and Muriel Meier for their technical assistance, Jacqueline Moret and Florian Kohler for their advises in statistics and Patrick Guerin for the English correction of the manuscript.

References

Ben Omar, N., and Ampe, F. (2000) Microbial community dynamics during production of the mexican fermented maize dough pozol. *Appl Environ Microbiol* **66:** 3664–3673.

Boon, N., De Windt, W., Verstraete, W., and Top, E.M. (2002) Evaluation of nested PCR-DGGE (denaturing gradient gel electrophoresis) with group-specific 16S rDNA primers for the analysis of bacterial communities from different wastewater treatment plants. *FEMS Microbiol Ecol* **39:** 101–112.

ter Braak, C.J.F. (1985) Correspondence analysis of incidence and abundance data: properties in terms of unimodal response model. *Biometrics* **41:** 859–873.

ter Braak, C.J.F. (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67:** 1167–1179.

ter Braak, C.J.F., Van Tongeren, O.F.R., and Jongman, R.H.G. (1995) *Data Analysis in Community and Landscape Ecology*. Cambridge, UK: Cambridge University Press.

Bruns, M.A., Stephen, J.R., Kowalchuk, G.A., Prosser, J.I., and Paul, E.A. (1999) Comparative diversity of ammonia oxidizer 16S rRNA gene sequences in native, tilled and successional soils. *Appl Environ Microbiol* **65:** 2994–3000.

Buchholz-Cleven, B.E.E., Rattunde, B., and Straub, K. (1997) Screening for genetic diversity of isolates of anaerobic Fe (II) -oxidizing bacteria using DGGE and whole-cell hybridization. *Syst Appl Microbiol* **20:** 301–309.

Casgrain, P., and Legendre, P. (2001) *The R Package for Multivariate and Spatial Analysis*, Version 4.0d6- User’s Manual. Département de Sciences Biologiques, Université de Montréal. Available, http://www.fas.umontreal.ca/BIOL/legendre/.

Dabert, P., Fleurat-Lessard, A., Mounier, E., Delgenes, J.P., Moletta, R., and Godon, J.J. (2001) Monitoring of the microbial community of a sequencing batch reactor bioaugmented to improve its phosphorus removal capabilities. *Water Sci Technol* **43:** 1–8.

Diez, B., Pedros-Alio, C., Marsh, T.L., and Massana, R. (2001) Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukariotic assemblages and comparison of DGGE with other molecular techniques. *Appl Environ Microbiol* **67:** 2942–2951.

Dollhopf, S.L., Hashsham, S.A., and Tiedje, J.M. (2001) Interpreting 16S rDNA T-RFLP data: application of self-organizing maps and principal component analysis to describe community dynamics and convergence. *Microbial Ecol* **42:** 495–505.

Duineveld, B.M., Kowalchuk, G.A., Keijzer, A., van Elsas, J.D., and van Veen, J.A. (2001) Analysis of bacterial communities in the rhizosphere of chrysanthenum via denaturing gradient gel electrophoresis of PCR-amplified 16S rRNA as well as DNA fragments coding for 16S rRNA. *Appl Environ Microbiol* **67:** 172–178.

Dunbar, J., Barns, S.M., Ticknor, L.O., and Kuske, C.R. (2002) Empirical and theoretical bacterial diversity in four Arizona soils. *Appl Environ Microbiol* **68:** 3035–3045.

van Elsas, J.D., Duarte, G.F., Keijzer-Wolters, A., and Smit, E. (2000) Analysis of the dynamics of fungal communities in soil via fungal-specific PCR of soil DNA followed by denaturing gradient gel electrophoresis. *J Microbiol Meth* **43:** 133–151.

Felske, A., and Akkermans, A.D. (1998) Prominent occurrence of ribosomes from an uncultured bacterium of the Verrucomicrobiales cluster in grassland soils. *Lett Appl Microbiol* **26:** 219–223.

Ferrari, V.C., and Hollibaugh, J.T. (1999) Distribution of microbial assemblages in the central Arctic ocean basin studied by PCR/DGGE: Analysis of a large data set. *Hydrobiologia* **401:** 55–68.

Ferris, M.J., and Ward, D.M. (1997) Seasonal distribution of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **63:** 1375–1381.

Fisher, M.M., and Triplett, E.W. (1999) Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl Environ Microbiol* **65:** 4630–4636.

Fjellbirkeland, A., Torsvik, V., and Øvreås, L. (2001) Methanotrophic diversity in an agricultural soil as evaluated by denaturing gradient gel electrophoresis profiles of *pmoA*, *mxaF* and 16S rDNA sequences. *Antonie Van Leeuwenhoek* **79:** 209–217.

Forestier, N., Steinmann, P., Lazko, E., Aragno, M., and Rossi, P. (2002) Combined statistical analysis of bacterial community structures from an eutrophic lake revealed by DGGE, PLFA and chemical analysis. In *Abstracts Book 61st Annual Assembly of Swiss Society for Microbiology*, Luzern, Switzerland, pp. 43.

Garcia-Pichel, F., López-Córtes, A., and Nübel, U. (2001) Phylogenetic and morphological diversity of cyanobacteria in soil desert crusts from the Colorado Plateau. *Appl Environ Microbiol* **67:** 1902–1910.

van der Gucht, K., Sabbe, K., de Meester, L., Vloemens, N., Zwart, G., Gillis, M., and Vyverman, W. (2001) Contrasting bacterioplankton community composition and seasonal dynamics in two neighbouring hypertrophic freshwater lakes. *Environ Microbiol* **3:** 680–690.

van Hannen, E.J., Zwart, G., van Agterveld, M.P., Gons, H.J., Ebert, J., and Laanbroek, H.J. (1999) Changes in bacterial and eukaryotic community structure after mass lysis of filamentous cyanobacteria associated with viruses. *Appl Environ Microbiol* **65:** 795–801.

Heilig, H.G., Zoedental, E.G., Vaughan, E.E., Marteau, P., Akkermans, A.D., and de Vos, W.M. (2002) Molecular diversity of *Lactobacillus* spp. & other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA. *Appl Environ Microbiol* **68:** 114–123.

Heuer, H., Krsek, M., Baker, P., Smalla, K., and Wellington, E.M. (1997) Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Appl Environ Microbiol* **63:** 3233–3241.

Heuer, H., Hartung, K., Wieland, G., Kramer, I., and Smalla, K. (1999) Polynucleotide probes that target a hypervariable region of 16S rRNA genes to identify bacterial isolates corresponding to bands of community fingerprints. *Appl Environ Microbiol* **65:** 1045–1049.

Ibekwe, A.M., Papiernik, S.K., Gan, J., Yates, S.R., Yang, C.-H., and Crowley, D.E. (2001) Impact of fumigants on soil microbial communities. *Appl Environ Microbiol* **67:** 3245–3257.

Jackson, C.R., Langner, H.W., Donahoe-Christiansen, J., Inskeep, W.P., and McDermott, T.R. (2001) Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring. *Environ Microbiol* **3:** 532–542.

Jourdain-Miserez, K., Aragno, M., and Rossi, P. (2001) Comparison of bacterial community structure in conventional and organic raw milk samples using TTGE analysis. In *Final Programm Abstracts, 9th Int Symp Microbial Ecol*, Amsterdam, Netherlands, FR 044, pp. 176.

Kocherginskaya, S.A., Aminov, R.I., and White, B.A. (2001) Analysis of the rumen bacterial diversity under two different diet conditions using Denaturing Gradient Gel Electrophoresis, random sequencing and statistical ecology approaches. *Anaerobe* **7:** 119–134.

Kowalchuk, G.A., Stephen, J.R., de Boer, W., Prosser, J.I., Embley, T.M., and Woldendorp, J.W. (1997) Analysis of ammonia-oxidizing bacteria of the beta subdivision of the class *Proteobacteria* in coastal sand dunes by denaturing gradient gel electrophoresis and sequencing of PCR-amplified 16S ribosomal DNA fragments. *Appl Environ Microbiol* **63:** 1489–1497.

Kowalchuk, G.A., Naoumenko, Z.S., Derikx, P.J., Felske, A., Stephen, J.R., and Arkhipchenko, I.A. (1999) Molecular analysis of ammonia-oxidizing bacteria of the beta subdivision of the class *Proteobacteria* in compost and composted materials. *Appl Environ Microbiol* **65:** 396–403.

Lee, S.Y., Bollinger, J., Bezdicek, D., and Ogram, A. (1996) Estimation of the abundance of an uncultured soil bacterial strain by a competitive quantitative PCR method. *Appl Environ Microbiol* **62:** 3787–3793.

Legendre, P., and Legendre, L. (1998) Numerical ecology. 2nd English edn. Amsterdam, The Netherlands: Elsevier, pp. 445.

Lovell, C.R., Piceno, Y.M., Quattro, J.M., and Bagwell, C.E. (2000) Molecular analysis of diazotroph diversity in the rhizosphere of the smooth cordgrass, *Spartina alterniflora*. *Appl Environ Microbiol* **66:** 3814–3822.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* **27:** 209–220.

McCaig, A.E., Glover, L.A., and Prosser, J.I. (2001) Numerical analysis of grassland bacterial community structure under different land management regimens by using 16S ribosomal DNA sequence data and denaturing gradient gel electrophoresis banding patterns. *Appl Environ Microbiol* **67:** 4554–4559.

McSpadden Gardener, B.B., and Lilley, A.K. (1997) Application of common statistical tools. In *Modern Soil Microbiology*. van Elsas, J.D., Trevors, J.T and Wellington, E.M.H., (eds). New York: Marcel Dekker, pp. 501–523.

Moeseneder, M.M., Arrieta, J.M., Muyzer, G., Winter, C., and Herndl, G.J. (1999) Optimization of terminal-restriction fragment length polymorphism analysis for complex marine bacterioplankton communities and comparison with denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **65:** 3518–3525.

Mohlenhoff, P., Muller, L., Gorbushina, A.A., and Petersen, K. (2001) Molecular approach to the characterization of fungal communities: methods for DNA extraction, PCR amplification and DGGE analysis of painted art objects. *FEMS Microbiol Lett* **195:** 169–173.

Morris, C.E., Bardin, M., Berge, O., Frey-Klett, P., Fromin, N., Girardin, H., *et al.* (2002) Microbial diversity: approaches to experimental design and hypothesis testing in the primary scientific literature from 1975 to 1999. *Microbiol Mol Biol Rev* (in press).

Müller, A.K., Westergaard, K., Christensen, S., and Sørensen, S.J. (2001) The effect of long-term mercury pollution on the soil microbial community. *FEMS Microbiol Ecol* **36:** 11–19.

Murray, A.E., Hollibaugh, J.T., and Orrego, C. (1996) Phylogenetic compositions of bacterioplankton from two California estuaries compared by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Appl Environ Microbiol* **62:** 2676–2680.

Murray, A.E., Preston, C.M., Massana, R., Taylor, L.T., Blakis, A., Wu, K., and DeLong, E.F. (1998) Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. *Appl Environ Microbiol* **64:** 2585–2595.

Muyzer, G., and Smalla, K. (1998) Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek* **73:** 127–141.

Muyzer, G., de Waal, E.C., and Uitterlinden, A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59:** 695–700.

Muyzer, G., Teske, A., Wirsen, C.O., and Jannasch, H.W. (1995) Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch Microbiol* **164:** 165–172.

Myers, R.M., Fischer, S.G., Lerman, L.S., and Maniatis, T. (1985) Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Res* **13:** 3131–3145.

Nübel, U., Garcia-Pichel, F., and Muyzer, G. (1997) PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl Environ Microbiol* **63:** 3327–3332.

Nübel, U., Garcia-Pichel, F., Kühl, M., and Muyzer, G. (1999a) Quantifying microbial diversity: morphotypes, 16S rRNA genes, and carotenoids of oxygenic phototrophs in microbial mats. *Appl Environ Microbiol* **65:** 422–430.

Nübel, U., Garcia-Pichel, F., Kühl, M., and Muyzer, G. (1999b) Spatial scale and the diversity of benthic cyanobacteria and diatoms in a salina. *Hydrobiologia* **401:** 381–391.

O'Donnell, A.G., Seasman, M., Macrae, A., Waite, I., and Davies, J.T. (2001) Plants and fertilisers as drivers of change in microbial community structure and function in soils. *Plant Soil* **232:** 135–145.

Ogino, A., Koshikawa, H., Nakahara, T., and Uchiyama, H. (2001) Succession of microbial communities during a biostimulation process as evaluated by DGGE and clone library analysis. *J Appl Microbiol* **91:** 625–635.

Piceno, Y.M., Noble, P.A., and Lovell, C.R. (1999) Spatial and temporal assessment of diazotroph assemblage composition in vegetated salt marsh sediments using denaturing gradient gel electrophoresis analysis. *Microb Ecol* **38:** 157–167.

Rademaker, J.L.W., Louws, F.J., Rossbach, U., Vinuesa, P., and De Bruijn, F. (1999) Computer-assisted analysis of molecular fingerprints and database construction. In *Molecular Microbial Ecology Manual*. Akkermans, A.D.L., van Elsas, J.D and De Bruijn, F.J., (eds). Dordrecht: Kluwer Academic Publisher, chapter. 7.1.3.

Ranjard, L., Poly, F., and Nazaret, S. (2000a) Monitoring complex bacterial communities using culture-independent molecular techniques: application to soil environment. *Res Microbiol* **151:** 167–177.

Ranjard, L., Nazaret, S., Gourbière, F., Thioulouse, J., Linet, P., and Richaume, A. (2000b) A soil microscale study to reveal the heterogeneity of Hg (II) impact on indigenous bacteria by quantification of adapted phenotypes and analysis of community DNA fingerprints. *FEMS Microbiol Ecol* **31:** 107–115.

Rölling, W.F.M., van Breukelen, B.M., Braster, M., Lin, B., and van Verseveld, H.W. (2001) Relationships between microbial community structure and hydrochemistry in a landfield leachate-polluted aquifer. *Appl Environ Microbiol* **67:** 4619–4629.

Schäfer, H., and Muyzer, G. (2001) Denaturing gradient gel electrophoresis in marine microbial ecology. *Meth Microbiol* **30:** 425–468.

Schäfer, H., Bernard, L., Courties, C., Lebaron, P., Servais, P., Pukall, R. *et al.* (2001) Microbial community dynamics in Mediterranean nutrient-enriched seawater mesocosms: changes in the genetic diversity of bacterial populations. *FEMS Microbiol Ecol* **34:** 243–253.

Schwieger, F., and Tebbe, C.C. (1998) A new approach to utilize PCR-single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis. *Appl Environ Microbiol* **64:** 4870–4876.

Simpson, J.M., McCracken, V.J., White, B.A., Gaskins, H.R., and Mackie, R.I. (1999) Application of denaturant gradient gel electrophoresis for the analysis of the porcine gastrointestinal microbiota. *J Microbiol Meth* **36:** 167–179.

Smalla, K., Wieland, G., Buchner, A., Zock, A., Parzy, J., Kaiser, S. *et al.* (2001) Bulk and rhizosphere soil bacterial communities studied by denaturing gradient gel electrophoresis: plant-dependent enrichment and seasonal shifts revealed. *Appl Environ Microbiol* **67:** 4742–4751.

Smit, E., Leeflang, P., and Wernars, K. (1997) Detection of shifts in microbial community structure and diversity in soil caused by copper contamination using amplified ribosomal

DNA restriction analysis. *FEMS Microbiol Ecol* **23:** 249–261.

Tiedje, J.M., Asuming-Brempong, S., Nüsslein, K., Marsh, T.L., and Flynn, S.J. (1999) Opening the black box of soil microbial diversity. *Appl Soil Ecol* **13:** 109–122.

Vallaeys, T., Topp, E., Muyzer, G., Macheret, V., Laguerre, G., Rigaud, A., and Soulas, G. (1997) Evaluation of denaturing gradient gel electrophoresis in the detection of 16S rDNA sequence variation in rhizobia and methanotrophs. *FEMS Microbiol Ecol* **24:** 279–285.

Wang, G.C., and Wang, Y. (1997) Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Appl Environ Microbiol* **63:** 4645–4650.

Ward, D.M., Weller, R., and Bateson, M.M. (1990) 16S rRNA sequences reveal numerous uncultured micro-organisms in a natural community. *Nature* **345:** 63–65.

Watanabe, M., Teramoto, M., Futamata, H., and Harayama, S. (1998) Molecular detection, isolation, and physiological characterisation of functionally dominant phenol-degrading bacteria in activated sludge. *Appl Environ Microbiol* **64:** 4396–4402.

Wawer, C., Jetten, M.S., and Muyzer, G. (1997) Genetic diversity and expression of the [NiFe] hydrogenase large-subunit gene of *Desulfovibrio* spp. in environmental samples. *Appl Environ Microbiol* **63:** 4360–4369.

Widmer, F., Fliessbach, A., Laczkó, E., Schulze-Aurich, J., and Zeyer, J. (2001) Assessing soil biological characteristics: a comparison of bulk soil community DNA-, PLFA-, and Biolog™-analysis. *Soil Biol Biochem* **33:** 1029–1036.

Yang, C.H., and Crowley, D.E. (2000) Rhizosphere microbial community structure in relation to root location and plant iron nutritional status. *Appl Environ Microbiol* **66:** 345–351.

Yang, C.H., Crowley, D.E., and Menge, J.A. (2001) 16S rDNA fingerprinting of rhizosphere bacterial communities associated with healthy and Phytophtora infected avocado roots. *FEMS Microbiol Ecol* **35:** 129–136.