Journal of Microbiological Methods 92 (2013) 173–177

Contents lists available at SciVerse ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Note

# A New Standard-Based Polynomial Interpolation (SBPIn) method to address gel-to-gel variability for the comparison of multiple denaturing gradient gel electrophoresis profile matrices

Alexis Valentín-Vargas, Jon Chorover, Raina M. Maier *

*Department of Soil, Water and Environmental Science, 429 Shantz Bldg. #38, 1177 E. Fourth Street, University of Arizona, Tucson, Arizona 85721-0038, USA*

ARTICLE INFO

*Article history:*
Received 12 September 2012
Received in revised form 3 December 2012
Accepted 3 December 2012
Available online 9 December 2012

*Keywords:*
DGGE
Gel-to-gel variations
Microbial ecology
Community fingerprint
Molecular standard

ABSTRACT

The Standard-Based Polynomial Interpolation (SBPIn) method is a new simple three-step protocol proposed to address common gel-to-gel variations for the comparison of sample profiles across multiple DGGE gels. The advantages of this method include no requirement for additional software or modification of the standard DGGE protocol.

© 2013 Elsevier B.V. All rights reserved.

Denaturing Gradient Gel Electrophoresis (DGGE) is a commonly used fingerprinting technique for rapidly assessing changes in the structure and dynamics of complex microbial communities. Since its invention (Muyzer et al., 1993), DGGE has been extensively applied in microbial ecology to study microbes in a wide variety of ecosystems, including: soils (Drees et al., 2006; Nakatsu et al., 2000), water (Schäfer and Muyzer, 2001), air (Li et al., 2010), caves (Legatzki et al., 2012), dust (Maier et al., 2010), mine tailings (Rosario et al., 2007; Solis-Dominguez et al., 2011) extreme environments (Ferris et al., 1996), humans (Favier et al., 2002), and animals (Mrázek et al, 2008). The extensive use of DGGE can be attributed mainly to the fact that it is a relatively simple and affordable technique that has been shown to be highly reproducible and can be used to rapidly analyze large numbers of samples (Neufeld and Mohn, 2005).

One of the drawbacks of DGGE is that a single gel can analyze only a limited number of samples (usually between 16–20 samples) and a straightforward and robust protocol for making meaningful DGGE gel-to-gel comparisons remains elusive. Well-documented limitations include the inconsistency between the denaturing gradients in different gels, as well as differences in the brightness of bands between samples. These issues, particularly the former, pose a major hurdle for the proper alignment and comparison of multiple gels (Nakatsu et al., 2000; Neufeld and Mohn, 2005).

Researchers have proposed the use of custom-made molecular standards to facilitate gel-to-gel comparisons (Gurtner et al., 2000). However, a detailed quantitative method for normalizing and comparing multiple DGGE gels based on custom standards has yet to be established. In fact, often when the use of molecular standards for DGGE normalization is reported, the actual detailed procedure used to normalize the profiles across multiple gels based on those standards is incomplete or not disclosed at all.

Recently, Tourlomousis et al. (2010) published a detailed and comprehensive approach to address gel-to-gel variation in the alignment and comparison of multiple DGGE gels. Their approach was underpinned by the use of several custom molecular standards in each DGGE gel. Using the software Phoretix 1-D (TotalLab, Ltd., Newcastele, UK), they corrected the within-gel distortions by manually calibrating band relative front (Rf) values for all the sample lanes based on a common Rf scale defined for the molecular standards. Rf is a measurement of band position along a lane, relative to the length of that lane (Tourlomousis et al., 2010). Thus, Rf values are calculated by dividing the distance a band has traveled down a lane by the length of the lane, which results in Rf values ranging from 0 to 1 (0 being the top of the lane and 1 being the bottom of the lane). The calibrated Rf matrices, along with the band positions measured in pixels, were then used to align all the profiles onto common axes by piecewise linear interpolation of the raw data using a Mathlab (The Mathworks, Inc., Cambridge, MA) script. This resulted

*Abbreviations:* DGGE, Denaturing Gradient Gel Electrophoresis; (SBPIn), Standard-Based Polynomial Interpolation method; NMDS, Non-Metric Multidimensional Scaling; Rf, Relative Front.

* Corresponding author at: Department of Soil, Water and Environmental Science, 429 Shantz Bldg. #38, 1177 E. Fourth Street, University of Arizona, Tucson, Arizona 85721-0038, USA. Tel.: +1 520 621 1646; fax: +1 520 621 1647.

*E-mail addresses:* valentin@email.arizona.edu (A. Valentín-Vargas), chorover@cals.arizona.edu (J. Chorover), rmaier@ag.arizona.edu (R.M. Maier).

0167-7012/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.mimet.2012.12.001

in a single normalized data matrix across multiple gels that was corrected for gel gradient distortion.

The drawbacks to Tourlomousis' et al. (2010) and other gel-to-gel alignment methods (*e.g.* Neufeld and Mohn, 2005) are the specific requirements for their application that can be difficult and significantly costly to implement. For instance, Tourlomousis' method requires a pixel matrix output that is not available in the Quantity One software, which comes with the D-Code Universal Mutation Detection System (Bio-Rad Laboratories, Hercules, CA) that most laboratories use for DGGE analysis (Green et al., 2009). Thus, the average DGGE user will have to acquire additional software in order to implement the protocol. Another example is the Neufeld and Monh (2005) protocol that not only requires a technical modification of the standard DGGE protocol, but also an expensive piece of specialized equipment in order to implement it.

The objective of this study was to build upon Tourlomousis et al. (2010) and develop an equally robust protocol that would allow comparison of multiple DGGE gels without the need for: (*i*) technical modification of the standard DGGE protocol; (*ii*) specialized equipment; or (*iii*) additional gel-analysis software other than Quantity One or any other software with the capacity of outputting basic Rf matrices. The Standard-Based Polynomial Interpolation (SBPIn) method we propose can be divided into the three major steps described below.

**Step 1 — Creating multiple DGGE gel matrices with normalized Rf values**. The first step for SBPIn requires the matching and normalization of the Rf values for all the bands within a single gel (*i.e.* within-gel normalization). This procedure relies on custom molecular standards that run in triplicate (first, middle and last lanes) on each gel. These custom standards can be created by either mixing amplicons obtained from phylogenetically distinct pure cultures or by mixing amplicons obtained from bands excised directly from a DGGE gel (Legatzki et al., 2012). Here, to illustrate the use of the SBPIn method we used the latter approach, a mix of DGGE amplicons which were chosen from the pool of samples to be analyzed by DGGE following the standard protocol. Random bands, covering as much as possible of the gradient range, were excised from the DGGE gel. Then, the DNA was purified from the polyacrylamide matrix, and re-amplified using the same PCR protocol used for the original amplification. These PCR products were then mixed to make the molecular ladder used as the reference standard. Each standard, as recommended by Tourlomousis et al. (2010), must consist of a minimum of 9 bands to allow good alignment of profiles within and between gels.

Once the DGGE gel is run, an image of the gel is analyzed using Quantity One (or equivalent) software. The standard lanes are defined by the user and Rf values are assigned to each standard band using the Rf values of the middle standard lane as reference. Then, based on the horizontal distribution of the sample bands and the distortions observed, all equivalent sample bands in each lane of a single gel are automatically matched by the software (using a matching algorithm) to form groups known as band types. Each band type is then automatically assigned a normalized Rf value that results from modeling across the gel with respect to the Rf values of the molecular standards. This matching-normalization step requires user input to validate the matches suggested by the software and correct mismatches that arise as a result of intrinsic gel distortions (known as "smiling effect" (Huber and Peduzzi, 2004)) commonly observed across DGGE gels. The result of this first step is multiple presence/absence matrices (one for each gel in a "test group") of Rf values normalized with respect to the same gel molecular standards (see Supplementary Fig. 1). As pointed out by Tourlomousis et al. (2010), this matching step is the backbone for accurate within- and between-gel alignments.

**Step 2 — Interpolation of normalized Rf values: creating a single unified matrix**. The second step involves setting a common normalized Rf scale to allow comparison among all the gel matrices in the test group using a piecewise polynomial interpolation (*i.e.* between-gel normalization). One of the gel matrices in the test group is randomly chosen to be the "reference" matrix. Then, pair-wise interpolations between the reference matrix and each of the other test group gel matrices are performed. For each interpolation, a two-column matrix is constructed with first column containing the Rf values for all of the sample and standard bands from one of the test group matrices and the second column containing the Rf values for the standard bands in the reference gel matrix (*e.g.*, Table 1). By matching the Rf values of the common standard bands from the reference matrix (Table 1, column 2) and the test group matrix (Table 1, column 1), new sample values for the test group matrix can be estimated by interpolation (Table 1, column 3). In other words, the Rf values of the bands in the standard lanes of the reference matrix are used to interpolate the Rf values for all the sample bands in the test group gel matrices being compared. The result is that both matrices (*i.e.* the test group gel for which Rf values were interpolated and the one from which the "reference standard" was obtained) will now share a common Rf scale for all their band types (Table 1). At the end of the interpolation step all the test group matrices will share a common Rf scale and this allows the construction of a single unified matrix that contains all the interpolated Rf types from all of the gels in the test group.

For the interpolation procedure we chose a polynomial approach given that a close examination of our profiles (Fig. 1) showed a marked gel-to-gel variation and the distribution of the bands between the different profiles was not linear. For example, we compared the same molecular standard on 5 different gels and demonstrated that even after aligning the top and bottom bands of the profiles, the distribution of the rest of the bands did not follow a linear distribution (Fig. 1). Thus, we opted for a more complex polynomial interpolation to correct the Rf values of the sample profiles. A regression analysis further confirmed that a polynomial regression was indeed a better fit for the alignment of our data (see Supplementary Fig. 2). From the various types of piecewise polynomial interpolations we chose the natural cubic spline method, mainly because of its lower degree of error as compared to other polynomial interpolation methods (Reinsch, 1967). For convenience, we performed this step using the CSIPOLATE module (Cox, 2009) for

**Table 1**
Example of a basic polynomial interpolation of normalized Rf values based on a reference standard.

| Test group[a] | | Reference standard[b] | | Test group-interpolated[c] |
|---|---|---|---|---|
| **0.309** | | **0.406** | | **0.406** |
| 0.313 | | | | 0.409 |
| 0.319 | | | | 0.414 |
| 0.327 | | | | 0.421 |
| 0.333 | | | | 0.426 |
| **0.338** | | **0.43** | | **0.43** |
| 0.34 | | | | 0.432 |
| 0.347 | | | | 0.437 |
| 0.354 | | | | 0.443 |
| **0.366** | ➔ | **0.453** | ➔ | **0.453** |
| 0.378 | | | | 0.463 |
| 0.385 | | | | 0.469 |
| 0.391 | | | | 0.475 |
| 0.397 | | | | 0.48 |
| 0.404 | | | | 0.487 |
| **0.409** | | **0.492** | | **0.492** |
| 0.412 | | | | 0.495 |
| 0.418 | | | | 0.502 |
| 0.424 | | | | 0.508 |
| **0.427** | | **0.511** | | **0.511** |

[a] Test group: All normalized Rf values from a single test group gel.
[b] Reference Standard: normalized Rf values of the standard bands obtained from the reference gel.
[c] Test group-interpolated: Rf values from the test group interpolated from the Rf values of the reference standard (interpolation based on the Cubic-spline polynomial method). Bolded values indicate Rf values from common standard bands.

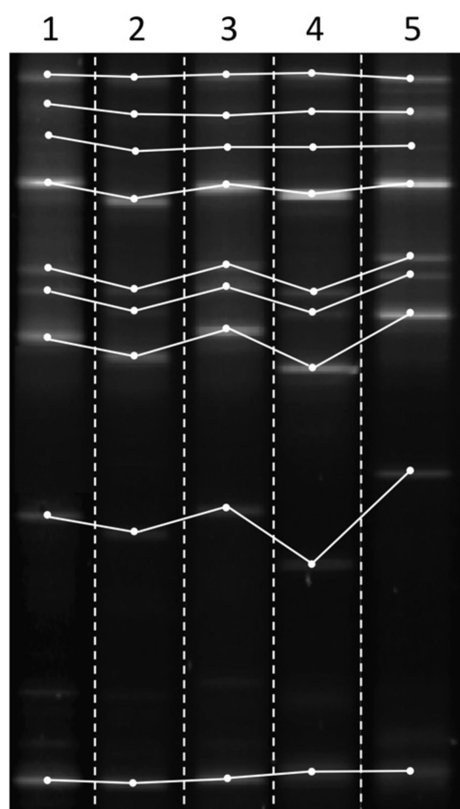


**Fig. 1.** Aligned standard profiles obtained from 5 different DGGE gels showing the nonlinearity of the bands distribution. As can be seen in this figure, only the top and bottom bands of the profiles were well aligned.

the STATA software (StataCorp LP, College Station, TX). However, several easy-to-use Excel-based macros and add-ins for conducting cubic spline interpolations are available free of charge on the Internet (*e.g.* http://www.xlxtrfun.com/XlXtrFun/XlXtrFun.htm).

**Step 3 — Binning of Rf types to create common Rf groups**. The third step of SPBIn consists of binning the interpolated Rf types from all the gels into common Rf groups. This step is done manually and is time-consuming but is necessary because even after interpolating the Rf values of the profiles based on a common Rf standard, it was observed that slight variations persisted between the multiple gels. This observation was verified by selecting 20 random, but prominent, bands present on all the profiles and comparing their interpolated Rf values. Based on this screening procedure we determined that equivalent bands from different gels could vary by up to 0.004 Rf units.

To carry out this third step, the unified matrix created in step 2 is used to manually generate bins (*i.e.* groups of Rf types) based on an adapted triple-criteria approach (Valentín-Vargas et al., 2012): (*i*) each bin may not surpass the maximum allowed per-bin variation of Rf values (*i.e.* $\pm 0.004$); (*ii*) a congruence criterion that prevents the formation of bins that cluster two or more Rf types from the same sample (*i.e.* only allow combination of Rf types that occur on different profiles); and (*iii*) a criterion to favor the selection of bins that include the most Rf types.

Briefly, all possible band types are placed into a single column in a spreadsheet and each band type is identified with the gel(s) in which they are present (*e.g.* band type Rf = 0.250 is present in gels 1, 3 and 4). Then, all possible combinations of bins are manually generated based on the first two criteria (Rf $\pm$ 0.004 and a single bin cannot have two bands from the same gel). Finally, the bins that have best fit with the third criterion are selected to represent that particular set of gels. This binning process is repeated twice, once creating bins in ascending order of Rf value and then creating bins in descending order of Rf value. The bins selected from both trials are used to create a final unified matrix of Rf types that can be subjected to statistical analysis. The Rf value assigned to each Rf type in a bin is an average of the Rf values of the bands within that bin. Note that anyone using the SBPIn method must determine the specific cutoff point for the formation of their bins based on the variability of their own profiles.

**Validation of SPBIn protocol**. The SBPIn protocol was validated using three independent trials from samples collected from an ongoing experiment. Two trials examined bacterial communities and one examined fungal communities. More specifically, the DNA templates used in the validation experiment were extracted from mine tailing samples obtained from a mesocosm-scale (~1.25 $m^3$) greenhouse phytostabilization study (data not published). Here we focus on the microbial community structure changes that occurred in two mesocosm treatments that were studied over the first year of the experiment: 1) mine tailings amended with 15% (w/w) of composted dairy manure and seeded with buffalo grass; and 2) mine tailings amended with 15% compost and seeded with quailbush. Samples were taken at 3, 6, 9, and 12 months to compare the bacterial (buffalo grass and quailbush treatments) and fungal communities (quailbush treatment). For each independent SPBIn trial we used triplicate samples (corresponding to three different mesocosms) obtained at each time point. We performed standard DNA extractions, PCR amplifications and DGGE analysis following previously published protocols (Solís-Domínguez et al., 2011).

To exemplify the SPBIn method, we present results primarily from one of the bacterial community trials, but all trials yielded comparable results even though the DGGE gradients used for the bacterial and fungal communities were different. To test the SPBIn method, triplicate samples from each of the four time points (3, 6, 9, and 12 months) were first analyzed separately on four different DGGE gels. These same four time points were then analyzed together on a fifth DGGE gel. The first four gels were subjected to the SBPIn method and the resulting unified matrix was named as the "Binned" matrix. For comparison purposes, a "Raw" matrix was generated by combining the profiles from these four gels but without subjecting them to the SBPIn method. The fifth gel was subjected to alignment and normalization using only its internal standards and the resulting matrix was named the "Validation" matrix.

All three matrices (*i.e.* Binned, Raw and Validation) were subjected to statistical analysis to assess the effectiveness of the SBPIn method. First, a pairwise Monte Carlo Permutation test (Table 2) was performed to evaluate the differences in community composition between all four time points. As expected, the Validation matrix results showed significant differences between all time points. However, the Raw matrix did not differentiate between treatments for three of the six possible pairs evaluated. After generating the Binned matrix by applying the SBPIn three-step protocol to the Raw data, the results again showed a significant difference among all time points analyzed.

To visually confirm these results, Non-Metric Multi Dimensional Scaling (NMDS) plots were generated for each matrix based on Jaccard

**Table 2**
Pairwise Monte Carlo permutation test to evaluate the differences in bacterial community composition before and after alignment of DGGE profiles.

| Group pair[a] | *P*-value ($p<0.05$)[b] | | |
|---|---|---|---|
| | Raw | Binned | Validation |
| T1–T2 | 0.018 | 0.006 | 0.009 |
| T1–T3 | **0.072** | 0.004 | 0.030 |
| T1–T4 | **0.060** | 0.012 | 0.023 |
| T2–T3 | 0.010 | 0.003 | 0.018 |
| T2–T4 | 0.028 | 0.010 | 0.005 |
| T3–T4 | **0.072** | 0.004 | 0.027 |

[a] T1 = Time 1 (3 months); T2 = Time 2 (6 months); T3 = Time 3 (9 months); T4 = Time (12 months).

[b] *P*-values above the significance level of 0.05 appear bolded and italicized.

distances (Fig. 2A). The NMDS plots verified the permutation test results; the Raw matrix did not show significant differences between the time points but both the Binned and Validation matrices clearly showed differences among the time points. Note that the ellipses surrounding the triplicate samples are not confidence intervals, but rather mark the area within which 95% of points sampled for each group are expected to fall. Thus, the overlap of some ellipses simply indicates that there could be overlapping of some bands (*i.e.* populations) between treatments observed for a particular ordinate dimension.

Three additional NMDS plots were also generated to examine the relationship between the Raw, Binned and Validation matrices (Fig. 2B). For each plot the same triplicate samples within a time point (3, 6, and 9 months) were obtained from the Raw, Binned and Validation matrices and then compared. As can be seen from the

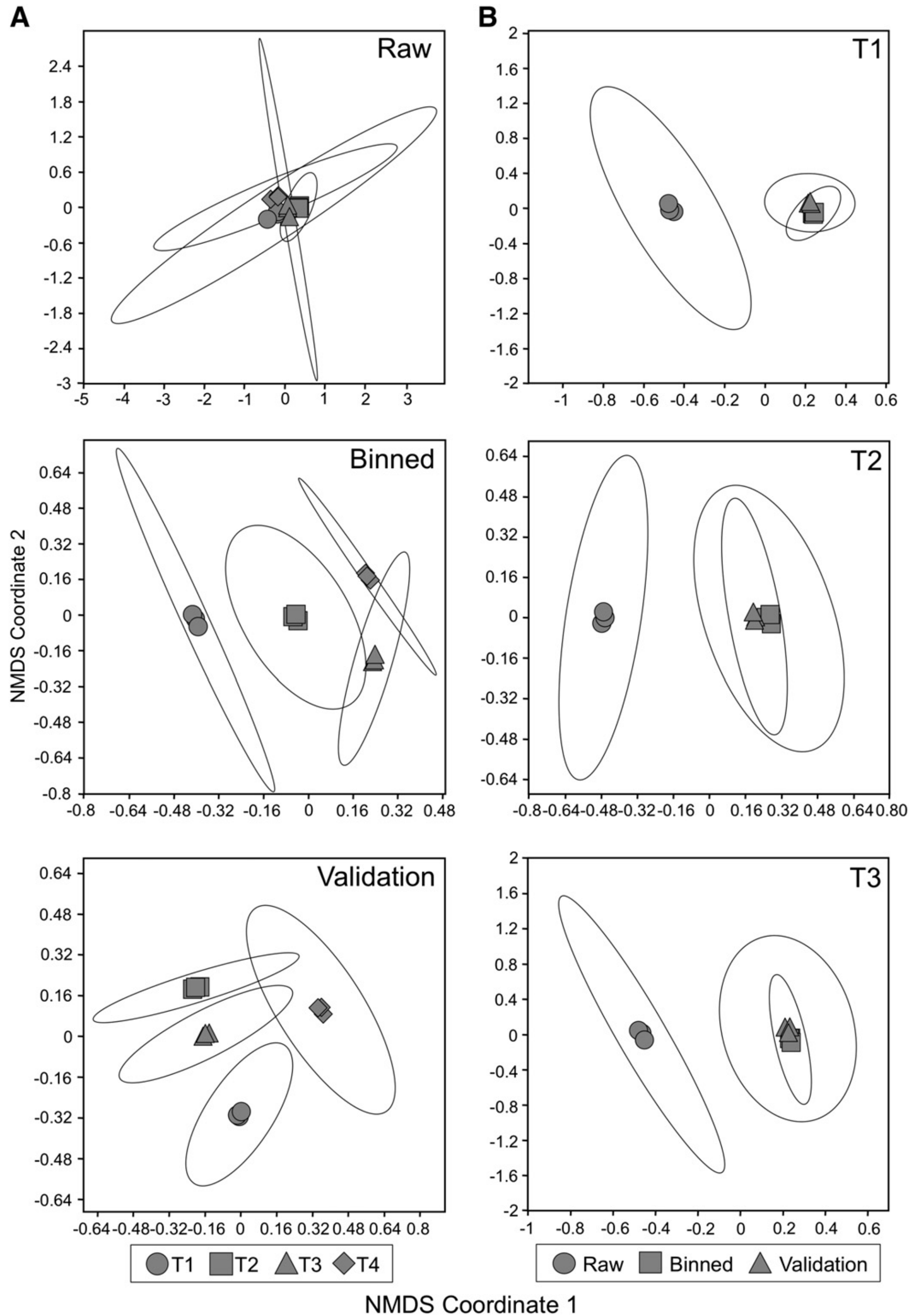


**Fig. 2.** Non-Metric Multidimensional Scaling (NMDS) ordination plots base on Jaccard distances. (A) NMDS plots showing the differences between the bacterial community profiles corresponding to 4 distinctive time points (T1 = Time 1 [3 months]; T2 = Time 2 [6 months]; T3 = Time 3 [9 months]; and T4 = Time 4 [12 months]). Each plot corresponds to a different experimental matrix (*i.e.* Raw, Binned and Validation). (B) NMDS plots showing the differences between the same bacterial community profile obtained from different matrices (*i.e.* Raw, Binned and Validation). Each plot corresponds to a different time point.

plots, the Validation and Binned matrix samples consistently and tightly overlap each other, while the samples from the Raw matrix always cluster apart from the others.

In summary, the novel SBPIn method proposed in this report for the alignment and comparison of DGGE profiles run on multiple gels generates data comparable to that obtained from running the same profiles on a single gel. The method is simple and does not require any additional software (other than a basic gel-analysis software with the capacity to generate Rf value matrices) or any major technical modification of the standard DGGE protocol in order to be performed. Anyone with software capable of generating basic Rf matrices can potentially apply these methods to their study.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.mimet.2012.12.001.

## Acknowledgments

This research was supported by grants from: NIEHS SRP Grant 2 P42 ES04940 and R01 ES17079. Additional financial support for AVV was provided by the American Society for Microbiology Robert D. Watkins fellowship program and the Alfred P. Sloan Foundation. We would like to thank Dr. Robert A. Root for providing the greenhouse samples and Lucia Rodriguez-Freire for her help in running and analyzing the DGGE gels. Thanks to Marian Ortiz and Julia W. Neilson for their valuable comments and suggestions to improve this manuscript.

## References

Cox, N.J., 2009. CSIPOLATE: Stata Module to Perform Cubic Spline Interpolation. Statistical Software Components S457005. Boston College Department of Economics. Available: http://ideas.repec.org/c/boc/bocode/s457005.html. (Accessed Aug 2012).

Drees, K.P., Neilson, J.W., Betancourt, J.L., Quade, J., Henderson, D.A., Pryor, B.M., Maier, R.M., 2006. Bacterial community structure in the hyperarid core of the Atacama Desert, Chile. Appl. Environ. Microbiol. 72, 7902–7908.

Favier, C.F., Vaughan, E.E., De Vos, W.M., Akkermans, A.D., 2002. Molecular monitoring of succession of bacterial communities in human neonates. Appl. Environ. Microbiol. 68, 219–226.

Ferris, M.J., Muyzer, G., Ward, D.M., 1996. Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. Appl. Environ. Microbiol. 62, 340–346.

Green, S.J., Leigh, M.B., Neufeld, J.D., 2009. Denaturing gradient gel electrophoresis (DGGE) for microbial community analysis. In: Timmis, K.N. (Ed.), Microbiology of Hydrocarbons, Oils, Lipids, and Derived Compounds. Springer, Heidelberg, pp. 4137–4158.

Gurtner, C., Heyrman, J., Piñar, G., Lubitz, W., Swings, J., Rölleke, S., 2000. Comparative analyses of the bacterial diversity on two different biodeteriorated wall paintings by DGGE and 16S rDNA sequence analysis. Int. Biodeterior. Biodegrad. 46, 229–239.

Huber, F., Peduzzi, P., 2004. Online tool for analysis of denaturing gradient gel electrophoresis profiles. Appl. Environ. Microbiol. 70, 4390–4392.

Legatzki, A., Ortiz, M., Neilson, J.W., Casavant, R.R., Palmer, M.W., Rasmussen, C., Pryor, B.M., Pierson III, L.S., Maier, R.M., 2012. Factors influencing observed variations in the structure of bacterial communities on calcite formations in Kartchner Caverns, AZ, USA. Geomicrobiol. J. 29, 422–434.

Li, K., Dong, S., Wu, Y., Yao, M., 2010. Comparison of the biological content of air samples collected at ground level and at higher elevation. Aerobiologia 26, 233–244.

Maier, R.M., Palmer, M.W., Andersen, G.L., Halonen, M.J., Josephson, K.C., Maier, R.S., Martinez, F.D., Neilson, J.W., Stern, D.A., Vercelli, D., Wright, A.L., 2010. The bacterial community in household dust: environmental determinants and impact on childhood asthma. Appl. Environ. Microbiol. 76, 2663–2667.

Mrázek, J., Strosová, L., Fliegerová, K., Kott, T., Kopecný, J., 2008. Diversity of insect intestinal microflora. Folia Microbiol. (Praha) 53, 229–233.

Muyzer, G., de Waal, E.C., 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.

Nakatsu, C.H., Torsvik, V., Øvreås, L., 2000. Soil community analysis using DGGE of 16S rDNA polymerase chain reaction products. Soil Sci. Soc. Am. J. 64, 1382–1388.

Neufeld, J.D., Mohn, W.W., 2005. Fluorophore-labeled primers improve the sensitivity, versatility, and normalization of denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. 71, 4893–4896.

Reinsch, C., 1967. Smoothing by spline functions. Numer. Math. 10, 177–183.

Rosario, K., Iverson, S.L., Henderson, D.A., Chartrand, S., McKeon, C., Glenn, E.P., Maier, R.M., 2007. Bacterial community changes during plant establishment at the San Pedro River mine tailings site. J. Environ. Qual. 36, 1249–1259.

Schäfer, H., Muyzer, G., 2001. Denaturing gradient gel electrophoresis in marine microbial ecology. In: Paul, J.H. (Ed.), Methods in Microbiology, vol. 30. Academic Press, London, pp. 425–468.

Solís-Domínguez, F.A., Valentín-Vargas, A., Chorover, J., Maier, R.M., 2011. Effect of arbuscular mycorrhizal fungi on plant biomass and the rhizosphere microbial community structure of mesquite grown in acidic lead/zinc mine tailings. Sci. Total Environ. 409, 1009–1016.

Tourlomousis, P., Kemsley, E.K., Ridgway, K.P., Toscano, M.J., Humphrey, T.J., Narbad, A., 2010. PCR-denaturing gradient gel electrophoresis of complex microbial communities: a two-step approach to address the effect of gel-to-gel variation and allow valid comparisons across a large dataset. Microb. Ecol. 59, 776–786.

Valentín-Vargas, A., Toro-Labrador, G., Massol-Deyá, A.A., 2012. Bacterial community dynamics in full-scale activated sludge bioreactors: operational and ecological factors driving community assembly and performance. PLoS One 7, e42524.