Exploring the links between vis-NIR spectra and bacterial communities structure of soils by PLS2 regression models

R. Joffre,[a,*] F. Gogé,[a] C. Jolivet,[b] N. Saby,[b] S. Dequiedt,[c] M. Lelievre[c] and L. Ranjard[c]

[a]CEFE-CNRS, UMR5175, 1919 Route de Mende, F-34293 Montpellier cefex 5, France. E-mail: richard.joffre@cefe.cnrs.fr
[b]INRA, Unité Infosol, BP 20619, F-45166 Olivet, France
[c]UMR Microbiologie du Sol et de l'Environnement, INRA Université de Bourgogne, BP 86510, F-21065 Dijon Cedex, France

Introduction

The application of visible and near-infrared reflectance spectroscopy (Vis-NIRS) to soil science is increasing rapidly, partly due to the ability of chemometrics to extract relevant biochemical informations from complex spectral signatures dominated by features originating from the mineral matrix.[1] Nevertheless, beyond the quantification of organic matter properties, spectral signatures could be used as metabolic fingerprinting, whose unique features could be related to particular microbial community and/or soil biodiversity parameters. Until now, very few studies have been done to associate Vis-NIRS spectra with molecular approaches of bacterial communities through DNA fingerprinting. A similar approach has been done to investigate relationships between NIR spectra and microbiological changes during fermentation of cocoa.[2] Based on a large soil library, encompassing a wide array of physico-chemical characteristics, climatic factors, land use and historical events, our objective was to explore the links between Vis-NIR spectra and composition of bacterial communities.

Materials and methods

The soils came from the RMQS network soil library (French soil quality monitoring network) representing soils sampled with a 16 × 16 km systematic grid, covering the whole French territory. A NIRSystems Model 6500 spectrophotometer (Foss Analytical) was used to record reflectance spectra of 1614 soils at 2 nm intervals between 400 to 2500 nm. Soil bacterial community structure of these soils was genotyped directly from soil DNA extracts by using a B-ARISA (Bacterial-

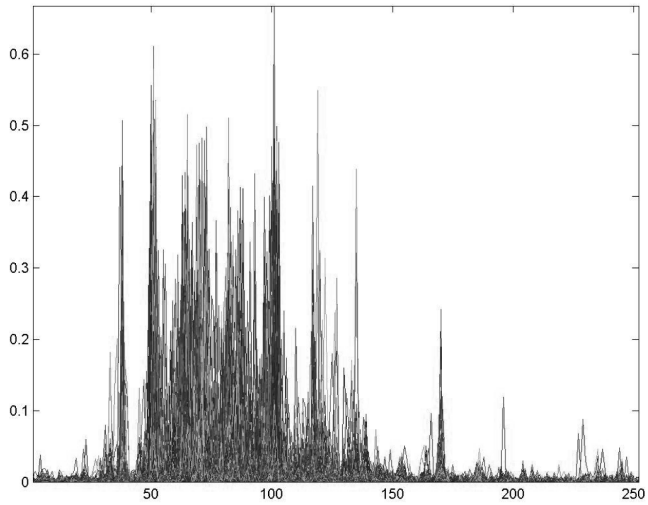

Figure 1. Diversity of the B-RISA profiles.

Automated Ribosomal 3 Intergenic Spacer Analysis) fingerprinting approach, optimized for medium throughput in the platform GenoSol.[3] Complex B-ARISA profiles were obtained from each soil and were compiled in a single matrix (252 variables).

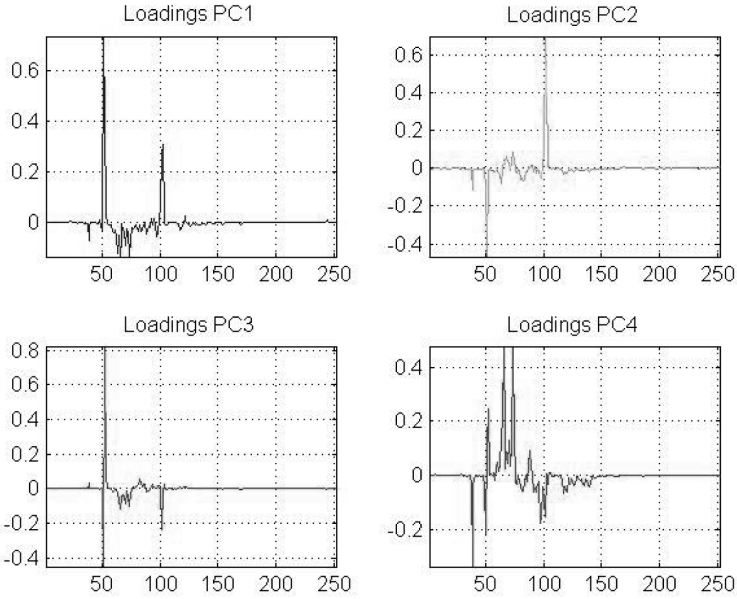

Figure 2. Loadings of 1 to 4 of PCA of B-RISA.

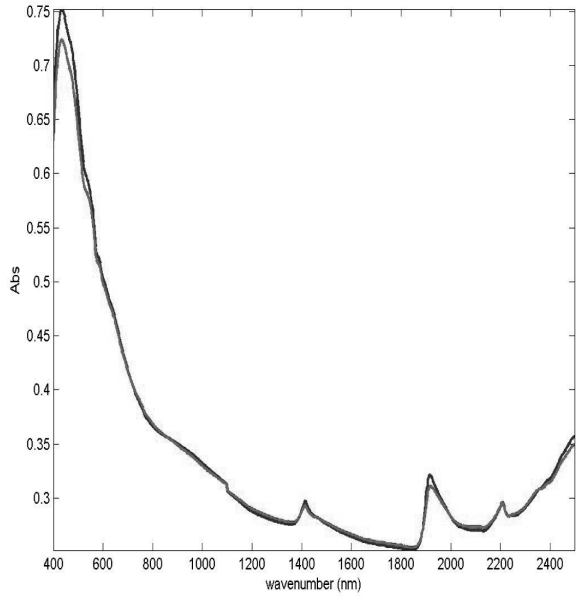

Figure 3. PLS2 modeled (red line) and measured spectra of one soil pertaining to the validation dataset.

Partial Least Squares regression model (PLS2) was next calculated between B-ARISA variables and Vis-NIR spectra. This method maximizes the covariance between the two data sets. Calibration was done over 2/3 (1084 soils) of the samples and validation over 1/3 (533 soils).

Results and discussion

B-RISA signatures offer a rapid culture-independent signature of bacterial community. Nevertheless, it is quite difficult to document the resulting profile precisely, by identifying

Table 1. Cumulated explained variance of the first five component of PLS2 model. X-Block referred to B-ARISA data and Y-Block to vis-NIRS data.

	Explained variance X-BLOCK (%) B-ARISA	Explained variance Y-BLOCK (%) vis-NIR spectra
PC 1	31.6	79.5
PC 2	41.8	92.4
PC 3	46.6	95.3
PC 4	51.1	95.3
PC 5	53.5	95.9

bacterial species with a determined peak. The general hypothesis considers each peak as indicative of the presence of a particular operational taxonomic unit (OTU).

Figure 1 illustrates the diversity and complexity of B-RISA profiles obtained in this study over 1614 soils. The four first components of a principal component analysis of this dataset explained 22%, 11%, 8% and 6% of the total variance, but were mainly due to a few peaks as shown in Figure 2.

The three first components attribute strong weights to two peaks at 51 bps and 101 bps. These peaks correspond to OTUs present in 1412 and 1614 soils i.e. 88% and 93% of samples respectively. It is therefore the relative importance of these OTUs in the structure of bacterial communities that explained their weights in the principal components.

The first component of the PLS2 model explained 31.6% of the B-ARISA and 79.5 of the vis-NIR data (Table 1).

The final model corresponded to the first five components with 53.5% of the B-ARISA variance explaining 95.9% of the vis-NIR data. Figure 3 compared a modeled spectrum (red line) to a measured spectrum (blue line) of the validation dataset, showing a better adjustment in the near infrared region than in the visible one.

70% of the r^2 values between each of the modeled and measured spectra of the validation dataset were higher than 0.9. For the first time, these results indicated significant links between bacterial community fingerprints and vis-NIR spectra from a wide array of soils.

Conclusion

A PLS2 procedure could be used to model Vis-NIR spectra from B-RISA profile. Our results showed that 32% of the B-RISA profile variance corresponded to a large part of the VIS-NIR spectra (79%), considered as a biochemical fingerprint. This could be due to the variation of abundance of a few bacteria species or more exactly to a few OTUs revealed by the distinct height of their peaks (Figure 1). The results of the B-RISA principal component analysis were in agreement with this result, as the first four loadings could be interpreted with very few peaks, corresponding to a reduced number of OTUs (Figure 2). If we considered the Vis-NIR spectra as an indication of the amount of organic matter, we could suppose that the abundance of these major species influence the amount of organic matter inversely. The inverse procedure (reconstructing the B-RISA profiles from the spectral data) did not work well, as only 31% of the variance of B-RISA could be explained by the VIS-NIR absorbance. So, the quality of the organic matter could not explain the richness of the bacterial community. To continue this work, it will be necessary to explore the structure of the B-RISA database intensively, to search some rules of species co-existence or exclusion, and identify OTU peaks better to understand these links. Relationships between the whole spectral dataset (DNA fingerprint and biochemical fingerprint) and the various set of descriptors dealing with pedoclimatic factors, land-use categories, and vegetation community would be intensively pursued through several data-mining approaches. The present results showed that the PLS2 method between the bacterial molecular approach and spectroscopic methods, such as VIS-NIR, seems to offer a potentially interesting tool to link bacterial diversity and activity to soil spectral characteristics.

Acknowledgements

This work was funded through the ECOMIC-RMQS program by the French National research agency (ANR).

References

1. R. Zornoza, C. Guerrero, J. Mataix-Solera, K.M. Scow, V. Arcenegui and J. Mataix-Beneyto, *Soil Biol. Biochem.* **40,** 1923 (2008).
2. D.S. Nielsen, P. Snitkjaer and F. van den Berg, *Int. J. Food Microbiol.* **125,** 133 (2008).
3. L. Ranjard, S. Dequiedt, M. Lelievre, P.A. Maron, C. Mougel, F. Morin and P. Lemanceau, *Environ. Microbiol. Reports* **1,** 7 (2009).