Title:

Impact of farming type, variety and geographical origin on bananas bacterial community.

Authors:

Céline Bigot[a,b,c], Christophe Bugaud[a,b], José Camilo[d], André Kapitan[a,b,e], Didier Montet[a,b] , Jean-Christophe Meile[b, f] .

[a] CIRAD, UMR Qualisud, F-34398 Montpellier, France.

[b] Qualisud, Univ Montpellier, CIRAD, Montpellier SupAgro, Univ d'Avignon, Univ de La Réunion, Montpellier, France.

[c] Danone Research - avenue de la Vauve, 91767 Palaiseau cedex, France.

[d] IDIAF, C/Rafael Augusto Sánchez #89, Ensanche Evaristo Morales, Santo Domingo, Dominican Republic.

[e] GSK, 1330 Rixensart, Belgium

[f] CIRAD, UMR Qualisud, F-97410 Saint-Pierre, Réunion, France.

[*] Corresponding Author: Céline Bigot.

E-mail address: celine.bigot83@hotmail.fr

Postal address: CIRAD-UMR Qualisud, TA B-95/16, 73, rue Jean-François Breton, 34398 Montpellier Cedex 5, France

Tel: +33 (0)6 79 62 20 49

Summary

Organic food products are subjected to high risk of fraud in Europe. Food traceability and authenticity of foods are mainly ensured by administrative means and there is a lack of analytical tools to authenticate organic foods. Hence, we wanted to propose a methodology that could help control and certification bodies for the authentication of organic foods. In this study we demonstrate the robustness of our original approach by comparing bacterial flora of bananas from different farming types, varieties, harvest years and geographical origins. Interestingly, the farming type could be linked to variations in bacterial diversity of bananas even if the geographical origin of bananas had the highest impact. Also, some bacterial groups have demonstrated a higher discriminative power. This opens the perspectives for the development of a food authentication tool based on the use of the natural microbial ecosystem of foods.

Key words: food safety; food quality; traceability; authenticity; organic foods; bacterial flora; fruits; banana.

Highlights

- The comparative analysis of microbial genetic profiles of fruits is an innovative approach in organic food authentication
- The discriminative power of our approach is increased by selecting the most discriminant bacterial groups
- The farming types has a higher impact on the bacterial flora compared to the variety or year of production; and geographical origin of fruits has a higher influence than the farming type

1. Introduction

Banana is the most produced fruit in the world. It is consumed worldwide and is also the most exported, especially the Cavendish variety. In Europe, Spain (Canary Islands) and France (Guadeloupe and Martinique) are the largest banana producers with less than 1 million tons produced per year compared to India which is the world's largest producer with an annual output of about 30 million tons (FAO, 2014). Banana plantations are very vulnerable to fungal diseases, as yellow and black Sigatoka diseases that are respectively caused by *Mycosphaerella musicola* and *Mycosphaerella fijiensi.* The high risk of fungal contamination leads producers to use antifungal treatments on banana trees as well as on fruits during postharvest stages. In addition, the need for farmers to use herbicides to treat their plots against weeds. Nevertheless, organic banana is getting an increasing importance in the European market due to the rise in popularity of organic production mode. This constitutes a challenge for organic producers. Indeed, organic farming is free of synthetic chemical products to favour agrosystems biodiversity, soil biological activity and biological cycles (Definition of the French agency for development and promotion of organic farming, Agence Bio, www.agencebio.org). In this way, this farming type differs from others by the use of controlled inputs free from synthetic chemicals and Genetically Modified Organisms (GMOs). In Europe, organic farming is legislated by regulation CE No 834/2007 and its implementing regulations CE No 889/2008 and CE No 1235/2008. The fight against pests and diseases in organic farming is primarily driven by detailed measurements (Article 12 of Regulation CE No 834/2007 of 28 June 2007). These articles relate mainly to the establishment of cultural and land practices compatible with this production mode as the use of appropriate species and varieties, the establishment of appropriate crop rotations and maintenance of natural enemies' pests. When these measures are not sufficient to protect plants against pests and diseases, the use of plant protection products is allowed (Article 5 of regulation CE No. 889/2008). Products whose active substances are listed

in Annexe II of regulation CE No 889/2008 can be used (for example micro-organisms used in biological control against pests and diseases or substances produced by microorganisms). In this context, farmers retain documentary evidence of the need for the use of these products (Ecocert, http://www.ecocert.fr). It was our interest to show the effect of organic farming practices on food microflora. In organic farming, the soil microflora is strengthened by the presence of self-propagating bacteria or by the addition of biological pest controls. In addition, the use of synthetic chemical pesticides or fertilizers is regulated, and the use of organic fertilizers will enrich bacterial diversity. Thus, we expected that there would be measurable differences in microbial ecosystems between organic and conventional food products. In the last decades, because of food crisis, consumers' expectations of the quality, safety and traceability of foods increased. Robust analytical tools for their authentication are thus needed (Capuano *et al.*, 2012). Recently, some tools were developed to authenticate several organic food products such as Stable Isotope Ratio Analysis (SIRA) or Near Infrared Reflectance Spectroscopy (NIRS) (Capuano *et al.*, 2012; Tres *et al.*, 2012; Sánchez *et al.*, 2013; Laursen *et al.*, 2014). However, there is no analytical tool for the discrimination of organic food from a conventional food bases on molecular microbial ecology approach. The link between microbial diversity and geographical origin was previously demonstrated on imported fish (Le Nguyen *et al.,* 2008; Tatsadjieu *et al.*, 2010) or fruits (El Sheikha *et al.*, 2009; El Sheikha *et al.*, 2011) using a molecular approach, the PCR-DGGE (Polymerase Chain Reaction coupled to Denaturing Gradient Gel Electrophoresis). This method provides a snapshot of the dominant microbial species (bacteria, fungi) that are present on foods, establishing a unique "barcode" for each sample. The barcode is a signature that reflects the number of major microbial phyla in a sample. This microbial composition is also related to the farming type, as previously demonstrated in peaches and nectarines from organic, conventional and sustainable agricultures (Bigot *et al.*, 2015).

100 The present study aimed at validating our approach on organic bananas by comparing the
101 bacterial environment of a wide range of bananas originating from Martinique and Dominican
102 Republic, from different farming types (organic vs. conventional), varieties (Cavendish and
103 CIRAD 925 hybrid) and harvest years (2013 and 2014).
104
105 **2. Materials and methods**
106 *2.1. Banana samplings and treatments applied*
107 All bananas were sampled during harvest periods, with gloves and sterile Whirl-Pak® bags.
108 Organic and conventional fruit batches (Cavendish) analysed for comparison in this study
109 originated from the same variety and geographical origin (between 500 m and 14 km distances,
110 Table 1).
111 *2.1.1. Bananas from Martinique*
112 Three types of bananas were sampled from the same producer (located in the north of
113 Martinique, in Basse-Pointe). Fruits codes were as follows: “925” for CIRAD 925, “CB” for
114 untreated Cavendish and “CC” for conventional treated Cavendish bananas.
115 The untreated Cavendish bananas (CB) were not certified that is why we used the terms
116 “treated” and “untreated” to avoid confusions between these bananas and the conventional ones
117 (CC). The untreated Cavendish bananas (CB) were cultivated on the same plot than CIRAD
118 925 (925) bananas to serve as contamination controls for Sigatoka disease and thus, to test the
119 resistance of CIRAD 925 variety against this fungal disease. Indeed, this variety 925 is a
120 resistant hybrid developed by CIRAD researchers (named CIRAD 925) that does not require
121 treatment against Sigatoka disease. All bananas were sheathed one week after the appearance
122 of the fruit.
123 The field where 925/CB fruits were sampled was located near a forest with no road around. In
124 addition, aerial spraying treatments was not allowed in a restricted area of 100 meters around

this plot The distance between the 2 plots (CIRAD 925/CB and CC) was around 500 m. Among the treatments applied, fungicides were used in the « treated Cavendish » (CC) plot only to fight against the Sigatoka disease (use of Propiconazole and Difenoconazole in solution with paraffin oil by aerial application). A systemic herbicide (whose active molecule is Glufosinate) was also applied twice between March and June 2013 for the treated bananas. The antifungal treatment frequency was every 3 weeks to 6 weeks, depending on weather conditions (wind, humidity, temperature).

About three months after flowering, bananas were sampled in 2013 at the centre of plots during harvest time, which depended on the variety (around 85 days for CIRAD 925 and CC bananas and around 95 days for CB bananas):

- CIRAD 925, 4 bananas per bunch were sampled from a total of 4 bunches (16 fruits),
- Untreated Cavendish (CB), 5 bananas per bunch were sampled from a total of 3 bunches (15 fruits),
- Treated Cavendish (CC), 3 bananas per bunch were sampled from a total of 5 bunches (15 fruits).

No post-harvest treatment was applied for all bananas.

2.1.1. Bananas from Dominican Republic

Bananas originated from the city of Mao, districts of La Caida and Boca de Mao.

After three months in the fields, bananas were sampled during two successive harvest years (on June 13th, 2013 and June 17th, 2014) to compare the effect of harvest time combined with the effect of farming type. All banana samples (n=30) were from Cavendish variety, including half (n=15) of them from organic farming (field located in La Caida, Mao) and the other half (n=15) from conventional farming (field located in Boca de Mao, Mao).

The organic fields were certified by Control Union Certification.

149 Due to the abandonment of organic practices by the farmer from La Caida, in 2014 the organic
150 bananas were sampled from another organic plot belonging to a neighbouring farmer, at
151 approximatively 8 km of distance. The variety of organic bananas sampled in 2014 was the
152 same as those sampled in 2013 (i.e. Cavendish variety) as well as the geographical origin (La
153 Caida, Mao). The only difference between organic bananas from 2013 and 2014 batches was
154 the organic plot (different organic farmers).
155 For each batch,, organic and conventional bananas were sampled from 5 different bunches at 3
156 different levels (upper, central and lower part, for a total of 30 bananas per year of sampling,
157 including the two farming types. A period of 10 days was necessary between the time of
158 sampling, transportation by plane and laboratory analyses at Cirad laboratories in Montpellier,
159 France. Transport was done in a container protected from exposure to direct light (obscurity).
160 No post-harvest treatment was applied. Field chemical treatments were applied to conventional
161 bananas only to fight mainly against Sigatoka disease. The antifungal active molecules were
162 Pyrimethanil, Mancozeb, Spiroxamine, Carbendazim, Triazole, Epoxiconazole,
163 Fenpropimorph and Thiram. As for organic fruit treatments, the active molecules were tea tree
164 oil and Potassium bicarbonate. The fertilizers used were potassium sulfate and ammonium
165 sulfate that were applied on conventional crops and composts were used for organic crop. All
166 bananas were sheathed and insecticides were used in sheath that surrounded the conventional
167 bananas.
168 *2.2. Extraction of total DNA*
169 Total DNA was extracted from the surface of bananas following a methodology to optimize
170 recovery yields of DNA material: 30 mL of sterile peptone water containing 1% Tween 80 were
171 added directly into each bag containing one banana fruit. After 30 min of mixing by rotation,
172 the mixture was transferred to 50 mL Falcon tubes and centrifuged at 3000 g for 5 min. Then,
173 the supernatant was discarded, and the pellet re-suspended with 2 mL of sterile peptone water

containing 1% of Tween 80. Then, we applied the methodology adapted from Masoud *et al.* (2004), Ros Chumillas *et al.* (2007) and El Sheikha *et al.* (2009): 1 mL of the resulting suspension was sampled in Eppendorf tubes containing ~0.3 g of sterile glass beads (Ref G8772, Sigma-Aldrich, France). The mixture was vortexed vigorously for 15 min in a bead beater instrument (Vortex Genie 2, USA) then centrifuged at 12 000 g for 15 min and the supernatant discarded. The cell pellet was resuspended in 300 µL breaking buffer [2 % Triton X-100 (Prolabo, France), 1 % SDS (sodium dodecyl sulfate; Sigma, France), 100 mM NaCl [(Sigma), 10 mM Tris, pH 8.0, 1 mM EDTA, pH 8.0 (Promega, France)]. 100 µL of TE buffer (10 mM Tris–HCl pH 8, 1 mM EDTA; Promega, France), 100 µL of lysozyme solution (25 mg/mL, Eurobio, France) and 100 µL of proteinase K solution (20 mg/mL, Biosolve, Netherlands) were successively added followed by 5 min incubation at room temperature. Samples were vortexed for 1 min and incubated at 42°C for 20 min. Then 50 µL of 20 % SDS were added to each tube, and incubated at 42°C for 10 min. 400 µL of 2 M NaCl, 2% (w/v) CTAB (acetyltrimethylammonium bromid, Merck, Germany) were added to each tube and incubated at 65°C for 10 min. The tubes were vortexed vigorously for 5 min after each addition. The lysates were twice subjected to 700 µL of phenol/chloroform/isoamyl alcohol (25/24/1, v/v/v, Carlo Erba, France), manually mixed and then centrifuged at 12 000 g for 15 min. The aqueous (upper) layer was transferred to a new Eppendorf tube. The residual phenol was removed by adding 600 µL of chloroform/isoamyl alcohol (25:24:1, Carlo Erba, France) and centrifuged at 12 000 g for 10 min. The aqueous (upper) phase was collected and DNA was precipitated by adding 0.1 volume of sodium acetate (3 M, pH 5) followed by one volume of isopropanol and stored at −20°C overnight. After centrifugation at 12 000 g for 30 min, the supernatant was eliminated, DNA pellets were washed with 500 µL of 70% ethanol, and tubes were centrifuged at 12 000 g for 5 min. The ethanol was then discarded and the pellets were air dried at room temperature for several hours (until total evaporation of ethanol). Finally, the

DNA was re-suspended in 50 µL of ultra-pure water and stored at 4°C until analysis. DNA quantities were estimated by electrophoretic migration through a 0.8% agarose gel and by using a UV-spectrophotometer (BioSpec-Nano, Shimadzu). Gels were photographed on a UV transilluminator with the Gel Smart 7.3 system (Clara Vision, France).

2.3. PCR-Denaturing Gradient Gel Electrophoresis (DGGE) analysis

Bacterial 16S variable DNA regions were amplified by PCR by using universal primers to analyse the bacterial ecosystems of bananas. To this end, a fragment of the V3 variable region of the 16S DNA gene was amplified by using the forward primer gc338F (5' CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG GAC TCC TAC GGG AGG CAG CAG 3'; Sigma) and the reverse primer 518R (5' ATT ACC GCG GCT GCT GG 3'), amplifying a 220 pb fragment (Le Nguyen *et al.*, 2008). PCR were performed in a final volume of 50 µL containing 0.1 µM of each primer, all the deoxyribonucleotide triphosphate (dNTPs, Promega) at 200 µM, 5 µL of 10x TopTaq buffer containing 15 mM $MgCl_2$ (Qiagen), 1.25 U of TopTaq DNA polymerase (Qiagen) and 5 µL of DNA extract (≤ 1µg/reaction). The PCR amplifications were carried out as follows: An initial denaturation step at 94°C for 3 min followed by 35 cycles consisting of - DNA denaturation step at 94°C for 30 sec, - primer annealing step at 60°C during 30 sec and - elongation step at 72°C for 1 min and then, a final extension step at 72°C for 10 min. Aliquots (5µL) of PCR products were analysed by electrophoresis through 2% (w/v) agarose gel with TAE 1x buffer (40 mM Tris-HCL, pH 7.4, 20 mM sodium acetate, 1.0 mM Na_2-EDTA), stained with GelRed™ 3x (Phenix Research Product) for 10 min. Gels were photographed and PCR bands signals were estimated by comparing a standard DNA (100 bp mass ladder, Promega). PCR products were then separated by Denaturing Gradient Gel Electrophoresis (DGGE) [(Dcode TM universal mutation detection system, Bio-Rad, USA)], using the procedure first described by Muyzer *et al.* (1993) and improved by Leesing (2005). Samples volumes were adjusted so as to load similar amounts of

PCR amplicons onto 8% (w/v) polyacrylamide gels (acrylamide/N,N-methylene bisacrylamide, 37,5/1, Promega, France) in 1× TAE buffer (40 mM Tris–HCl, pH 7.4, 20 mM sodium acetate, 1.0 mM Na_2-EDTA). Electrophoreses were performed at 60°C, using a denaturing gradient in the 30–60% range (100% corresponding to 7M urea and 40% v/v formamide, Promega, France). The gels were electrophoresed at 20 V for 10 min and then at 80 V for 12h. After electrophoresis, the gels were stained for 30 min with GelRed™ 3x (Phenix Research Product) and then photographed as described above.

2.4. Image and statistical analysis

Individual lanes of gel images were aligned and processed using ImageQuant TL software version 2007 (Amersham Biosciences). This software allows detection, precise measure and record of the relative position of each DNA band. The DGGE banding pattern is considered as an image of all of the major bacterial species in the sample. An individual discrete band refers to a unique "sequence type" or phylotype (Van Hannen *et al*., 1999; Kowalchuk *et al.*, 1997; Muyzer *et al.*, 1996), which is treated as a discrete bacterial species.

2.4.1 Dice similarity coefficient

DGGE fingerprints were manually scored by the presence and the absence of co-migrating bands between lanes, independently of the intensity. Pairwise community similarities were quantified using the Dice similarity coefficient (S_D) (Heyndrickx *et al*., 1996):

$S_D = 2\,Nc / (Na + Nb)$

Where N_a represents the number of bands detected in sample A, N_b the number of bands in sample B, and N_c the number of bands common to both samples. The similarity coefficient was expressed within a range of 0 (completely dissimilar) to 100 (perfect similarity). Dendograms were constructed using the Statistica version 6 software (StatSoft). Significant differences of bacterial communities of bananas were determined by factorial correspondence analysis, using the first two factors that described most of the variation in the data set.

A Cluster Analysis was performed using the similarity matrix to group samples according to their similarity index. The reconstruction method used was group average by using Primer v.6 software (Primer-E Ltd).

2.4.2 Partial least squares discriminant analysis (PLS-DA)

PLS-DA is a regression technique, which maximises the separation between pre-defined classes. The aim is to predict the values of a group of variables Y (dependent variables) from a set of variables X (explanatory variables) (Tenenhaus *et al.*, 2005). In our case, X represents relative quantitative variables with "band percentage", which corresponds to the value measurement of the band's volume divided by the total volume of all the bands in the lane (band volume is defined as the surface of the band in pixel multiplied by the intensity level of each pixel of the band value measured by the Image Quant TL software v2007, Amersham Biosciences). Y represents qualitative variables that are farming types (organic or conventional) or geographical origins (as Martinique or Dominican Republic) or varieties. The quality and significance of the results obtained are represented by R^2Y cum, R^2X cum and Q^2 cum values. R^2Ycum and R^2Xcum are the percent of the variation of all the Y and X explained by the model; Q^2cum is the cumulative percent of the variation of the Y variable predicted by the model, after the last component, according to cross validation. Q^2 tells you the quality of the prediction obtained from PLS regression. The closer to 1 are the values, better is the model.

The model can be improved by selecting the most discriminant variables X through the analysis of the Variable Important in the Projection (VIP). The variables with a VIP score close or equal to 1 is considered as being important for the PLS model (Tenenhaus *et al.*, 1998; Erikson *et al.*, 1999).

In our study, binary classification models were developed (for example 1 for organic and 0 for conventional) and the belonging to one of the classes was predicted by PLS-DA according to the "bacterial band percentage" value. The number of PLS-DA latent components (LV) was

optimised according the percentage of correct classified in cross validation, summarised by the confusion matrix. The confusion matrix allows easy visualisation of correct or incorrect classifications.

The data were processed using XLSTAT's statistical analysis software version 2014.

Therefore, PLS-DA allows to combine variables in the data set to find the maximum correlation between them and the class variable and, thus, the maximum separation among classes (e.g. organic vs. conventional; Martinique vs. Dominican Republic; etc.).

2.4.3 Diversity indices

The definition of microbial diversity refers to two concepts: the richness (number of species) and the relative abundance of different species (dominance / equity). The majority of the indices used to quantify the diversity of a community take into account both aspects. The numerical value of the different indices depends mainly on the weight given to rare and abundant species in the calculation. Previous studies demonstrated the application of diversity indices using the total number of bands (S) present on a DGGE gel and their relative intensity for comparing microbial communities (e.g. Duarte *et al.*, 2012; Nikolcheva *et al.*, 2003). Thus, using diversity indices combined to community structure descriptors (species richness, Shannon and Simpson diversity, dominance and evenness) we can estimate and compare the microbial diversity of a sample according its farming type.

The free software for scientific data analysis Past version 3.04 allowed us to have an estimation of the real diversity of communities present on complex environment.

3. Results and discussion

3.1 Comparison of bacterial communities according to farming types

Bacterial floras associated to bananas from different mode of productions, different varieties (Cavendish vs. CIRAD 925), two successive harvest years (2013 vs. 2014) and geographical origins (Martinique vs. Dominican Republic) were compared. The comparison of these four

parameters allowed to determine to what extend the farming type influenced the bacterial communities associated to bananas.

Sufficient bacterial DNA quantities have been extracted and amplified by PCR to be analysed by DGGE, in opposition to fungal DNA that could not be amplified (not shown).

When comparing the bacterial genetic profiles of bananas from different farming types, a trend of high heterogeneity could be observed whatever the geographical origin and the harvest time (Figures 1A and 1B). Therefore, the inter-fruit variations are important and measurable in quantitative and qualitative ways. PLS-DA analysis was used to interpret the results. This statistical tool allowed to consider both the bacterial richness (number of bacterial species or DNA bands in a sample) and the relative bacterial abundance (intensity of each DNA band of a sample) detected in banana samples by PCR-DGGE. PLS-DA was directly applied from the most discriminant bacterial DNA band data (VIP) because variations obtained between DGGE bacterial profiles were too important to allow a significant classification by using the information provided by all bacterial DNA bands (data not shown). This suggests that the information given by bacterial DNA markers would be hidden by the other bacterial species that would not contribute to sample discrimination. Indeed, the information provided by about twenty bacterial markers (DNA bands) are sufficient to discriminate bananas according to their farming practices (Figures 2 and 3). Figure 2 shows the dispersion of individuals (bacterial DNA markers) and the dispersion of observations (samples from a farming type of a given sampling year). The more variables are close to the circle of correlations, the better they are represented and so, the most important they are for the discriminant analysis. The Figure 3 represents the quality of the statistical analysis and also the most discriminant bacterial DNA markers/bands (or VIP). The histograms obtained in the Figure 3 shows significant results: R^2Y and Q^2cum values are respectively superior to 0.9 and 0.5 on two components. In addition, the results of the confusion matrix indicated that the rate of correct classification was 100% (data

not shown). These observations demonstrate that our model is good and can be used to interpret the results.

The correlation maps illustrated in the Figure 2 allow demonstrating that treated Cavendish (CC) (Figure 2A) and organic bananas (Figure 2B) are negatively correlated with their respective counterparts, i.e. Untreated Cavendish (CB) and "Conv" bananas. It was also observed that some quantitative variables (represented in red with the mention "Bactn") may significantly contribute to a specific class (either organic or conv. / CB or CC). In the case of bananas from Martinique (Figure 2A), 21 "BactN" variables allowed to mainly explain the "CB" class whereas only 3 variables significantly contributed to the "CC" class (Bact23, 29 and 54). The graphic representing the VIP demonstrates the importance of these bacterial DNA markers to discriminate bananas. As for Dominican Republic bananas, the explanative variables are evenly between the classes "Organic" and "Conv" (Figure 2B) and some other variables do not contribute to either of the classes (for examples Bact2, 6, 11, 40 et 41). Therefore, the bananas from Martinique and Dominican Republic can be discriminated according to their farming types, whether certified organic or not. The results obtained demonstrate as well that there are discriminant bacterial DNA markers or species from organic/untreated bananas but also from conventional ones.

The histograms representing the "model quality by number of components" (Figure 3), indicate that results obtained for bananas from Martinique (Figure 3A) are more significant than those obtained for bananas from Dominican Republic (Figure 3B), with Q^2cum values equal to 0,7 and 0,5 respectively. The fact that the effect of two parameters (year and plot in the case of bananas from Dominican Republic) effects were measured in a same dataset comparison (Figures 2B and 3B), the quality of our statistical analysis has been significantly reduced (20%) compared to an analysis taking only into account the effect of the farming practices (data not

348 shown). This means that the year of sampling has an impact on the bacterial ecosystem of
349 bananas but to a lesser extent when compared to the treatments applied to the crops.
350 *3.2 Comparison of bacterial communities according to variety*
351 Thereafter, the variations observed on the bacterial flora due to the farming practices were
352 compared to those that were caused by the variety of bananas (Cavendish vs CIRAD 925) and
353 their geographical origin (Martinique vs Dominican Republic) (Figures 4 and 5). The CB
354 bananas were cultivated on the same field than the variety CIRAD 925 (untreated as well). It
355 was thus possible to measure the sole impact of the variety since the same treatments and
356 geographical origin applied (Figure 4A). Previous studies showed that it was possible to link
357 the microbial ecosystem of foods to their geographical origins (Le Nguyen, 2008; El Sheikha
358 *et al.,* 2009; Tatsadjieu *et al.,* 2010; Dufossé *et al.,* 2013) and their farming types (Bigot *et al.,*
359 2015). It was interesting here to compare the impact of these both parameters on the microbial
360 flora of bananas (Figure 4B) and to be able to determine what is the influence of each parameter.
361 For the study of the "variety effect", the more significant variables were selected by using the
362 analysis of VIP. The VIP with values superior to 1 allowed to improve the statistical results and
363 were retained. The others were excluded from the analysis. Before this selection, the quality of
364 our statistical analysis (Q^2cum) was too low (0.4), even when increasing the number of
365 components (data not shown). This suggests that the quality of the adjustment could be highly
366 variable according to bananas (CIRAD 925 vs Cavendish). It highlighted the need to delete the
367 variables X (Bactn) that were not informative for the discrimination of sample class. Besides,
368 this selection allowed to increase the Q^2cum value by 20% (it increased from 0.4 to 0.6). This
369 selection was not necessary to study the impact of the geographical origin. Indeed, significant
370 results were obtained, with Q^2cum and R^2Y values equal to 0.7 and 0.9 respectively, without
371 having to delete data from non-VIP bacterial DNA bands (Figure 5B). The graphics showing
372 the dispersion of samples (Figure 5) informed that these samples are grouped according to their

373 variety (Figure 5A) or to their geographical origin (Figure 5B). In these graphics, it can also be
374 observed that samples tend to cluster according to their farming type As shown in Figure 5A,
375 where the 3 groups, “925” (CIRAD 925 variety), “CB” and “CC” (Cavendish varieties), are
376 clearly distinct from each other while it is the impact of the variety that is tested. The variety
377 effect leads to important variations on the bacterial flora of bananas from Martinique (R^2X =
378 0,6) because it was possible to distinguish the “Cavendish” and “925” groups (Figure 5A).
379 However, these variations were less important compared to the effect related to the farming
380 type. Indeed, when the data describing the farming type were used as explanative variables, the
381 statistical results were more predictive ($R^2X = 0.7$). The Figure 5B focused on the impact of the
382 geographical origin on the bacterial flora of bananas rather than the farming type. Even if
383 samples tend to be grouped according to their farming type (“CB” and “CC”), the
384 discrimination between samples is mainly driven by their geographical origin. Indeed, as it was
385 mentioned previously, the results obtained are highly significant, with $Q^2cum = 0.7$ and R^2Y =
386 0.9 and this without having to delete non-VIP bacterial DNA bands contrary to what was
387 previously done (Figures 2 and 3). Therefore, the various parameters tested on bananas showed
388 that some of them have greater impact on their bacterial ecosystem. It was demonstrated that
389 the farming practices have more impact than the variety of bananas or the sampling period
390 (Figures 2, 3, 4A and 5A). Interestingly, these observations are in accordance with previous
391 work on peaches and nectarines, the differences observed in the microbial flora were accurate
392 enough to conclude that they resulted mostly from applied treatments compared to the “post-
393 harvest age” and also to the variety of the fruit (Bigot et al., 2015). However, the effect of
394 geographical origin prevails over the farming practices (Figures 4B and 5B). Thereby, if we
395 classify by the order of importance the different parameters that showed an impact on the
396 bacterial flora of bananas we would obtain: 1- Origin geographic; 2- Agricultural practice; 3-
397 Year of production; 4- Variety. Even if studies on other fruits would be necessary to validate

398 the versatility and robustness of our approach, recent works have clearly demonstrated the
399 impact of the mode of production on the bacterial flora and is corroborated with other studies
400 (Ottesen *et al.*, 2009; Leff *et al.*, 2013; Bigot *et al.*, 2015).
401
402 **4. Conclusions and perspectives**
403 The results obtained in this study demonstrate the efficacy of our approach for the
404 discrimination of bananas according to their farming type by using the comparative analysis of
405 their bacterial ecosystem. Data obtained on bananas together with the previous work conducted
406 on peaches and nectarines (Bigot *et al.*, 2015) clearly show the interest of using comparative
407 analyses of microbial genetic profiles as an innovative approach in organic food authentication.
408 In addition, our approach has demonstrated that the variability linked to various parameters
409 such as the year of production (2013 vs 2014), the variety (Cavendish vs CIRAD 925) and the
410 position of bananas on a same bunch (at 3 different levels: upper, central and lower part; data
411 not shown) were less important than the variability related to the treatments (organic/CB vs.
412 conventional/CC). However, it was observed that the geographical origin of bananas exerted an
413 influence on their bacterial ecosystem that is most significant compared to the treatments
414 applied. The development of an analytical tool for authenticating organic products based on the
415 use of the discriminant microbial markers, not only specific to the organic fruits, would
416 potentially be suitable and more powerful than the use of the global microflora.
417 Thus, as a perspective, it would be interesting to identify specifically the bacterial groups
418 ("BactN") to design primers or probes for implementing an authentication tool, such as DNA
419 microarray or NGS for examples. The dosage of these discriminant markers in random samples
420 would demonstrate the validity of our approach.
421 This methodology can be applied to other foodstuffs for both food safety and food quality
422 purposes.

çAcknowledgments

This work was funded by a PhD and DESI grants from Cirad allocated to Céline Bigot.

References

Ampe, F., Ben Omar, N., Moizan, C., Wacher, C., & Guyot, J. P. (1999). Polyphasic study of the spatial distribution of microorganisms in Mexican pozol, a fermented maize dough, demonstrated the need for cultivation-independent methods to investigate traditional fermentations. *Applied and Environmental Microbiology, 65,* 5464–5473.

Bigot, C., Meile, J.C., Kapitan, A., & Montet, D. (2015). Discriminating organic and conventional foods by analysis of their microbial ecology: An application on fruits. *Food Control, 48,* 123–129.

Capuano, E., Boerrigter-Eenling, R., Van der Veer, G., & Van Ruth, S.M. (2012). Analytical authentication of organic products: an overview of markers. *J Sci Food Agric, 1,* 12–28.

Duarte, S., Cassio, F., & Pascoal, C. (2012). Denaturing Gradient Gel Electrophoresis (DGGE) in Microbial Ecology - Insights from Freshwaters. *Gel Electrophoresis - Principles and Basics*, Dr. Sameh Magdeldin (Ed.), InTech, DOI: 10.5772/38177.

Dufossé, L., Donadio, C., Valla, A., Meile, J. C., & Montet, D. (2013). Determination of speciality food salt origin by using 16S rDNA fingerprinting of bacterial communities by PCR-DGGE: an application on marine salts produced in solar salterns from the French Atlantic Ocean. *Food Control, 32,* 644–649.

El Sheikha, A.F., Condur, A., Métayer, I., Le Nguyen, D.D., Loiseau, G., & Montet, D. (2009). Determination of fruit origin by using 26S rDNA fingerprinting of yeast communities by PCR–DGGE: preliminary application to *Physalis* fruits from Egypt. *Yeast, 26,* 567–573.

El Sheikha, A.F., Bouvet, J.M., & Montet, D. (2011). Biological bar code for determining the geographical origin of fruits using 28S rDNA fingerprinting of fungal communities by PCR-DGGE: an application to Shea tree fruits. *Quality Assurance and Safety of Crops & Foods, 3,* 40–47.

453 Eriksson L, Johansson E, Kettaneh-Wold N, Wold S (1999) Introduction to multi- and
454 megavariate data analysis using projection methods (PCA and PLS). Umetrics, Umea.
455 FAO (2014). Banana Market Review and Banana Statistics 2012-2013. Rome, Italy.
456 Heyndrickx, M., Vauterin, L., Vandamme, P., Kersters, K., & De Vos, P. (1996). Applicability
457 of combined amplified ribosomal DNA restriction analysis (ARDRA) patterns in bacterial
458 phylogeny and taxonomy. *Journal of Microbiological Methods, 26,* 247–259.
459 Fung F., Wang H.S. & Menon S. (2018). Food safety in the 21st century. *Biomed*, *41*, 88–95.
460 Kowalchuk, G. A., Stephen, J. R., De Boer, W., Prosser, J. I., Embley, T. M., & Woldendorp,
461 J. W. (1997). Analysis of ammonia-oxidizing bacteria of the beta subdivision of the class
462 Proteobacteria in coastal sand dunes by denaturing gradient gel electrophoresis and sequencing
463 of PCR-amplified 16S ribosomal DNA fragments. *Applied and Environmental Microbiology,*
464 *63,* 1489–1497.
465 Laursen, K.H., Schjoerring, J.K., Kelly, S.D., & Husted, S. (2014). Authentication of
466 organically grown plants – advantages and limitations of atomic spectroscopy for multi-element
467 and stable isotope analysis. *Trends in Analytical Chemistry, 59,* 73–82.
468 Leesing, R. (2005). Identification and validation of specific markers for traceability of
469 aquaculture fish for import/export. *Ph.D. dissertation*, University of Montpellier 2, France.
470 Leff, J.W. and Fierer, N., 2013. Bacterial communities associated with the surfaces of fresh
471 fruits and vegetables. PLoS One 8: e59310.
472 Le Nguyen, D. D., Hanh, H. N., Dijoux, D., Loiseau, G., & Montet, D. (2008). Determination
473 of fish origin by using 16S rDNA fingerprinting of bacterial communities by PCR-DGGE: an
474 application on Pangasius fish from Viet Nam. *Food Control, 19,* 454–460.
475 Leveau, J.H., & Preston, G.M. (2008). Bacterial mycophagy: definition and diagnosis of a
476 unique bacterial-fungal interaction. New Phytol 177: 859–876.

477 Masoud, W., Cesar, L.B., Jespersen, L., & Jakobsen, M. (2004). Yeast involved in the
478 fermentation of *Coffea arabica* in East Africa determined by genotyping and by direct
479 denaturing gradient gel electrophoresis. *Yeast, 21,* 549–556.
480 Muyzer, G., De Waal, E.C., & Uitterlinden, A.G. (1993). Profiling of complex microbial
481 populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-
482 amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology, 59,* 695–700.
483 Muyzer, G., Hottenträger, S., Teske, A., & Wawer, C. (1996). Denaturing Gradient Gel
484 Electrophoresis of PCR-amplified 16S rDNA e a new molecular approach to analyse the genetic
485 diversity of mixed microbial communities. *Molecular Microbial Ecology Manual, 34,* 1–23.
486 Nikolcheva, L.G.; Cockshutt, A.M. & Bärlocher, F. (2003). Determining diversity of
487 freshwater fungi on decomposing leaves: comparison of traditional and molecular approaches.
488 *Applied and Environmental Microbiology, 69,* 2548–2554.
489 Ottesen, A.R., White, J.R., Skaltsas, D.N., Newell, M.J., and Walsh, C.S., 2009. Impact of
490 organic and conventional management on the phyllosphere microbial ecology of an apple crop.
491 J Food Prot. 72: 2321–2325.
492 Sánchez, M.T., Garrido-Varo, A., Guerrero, J.E., & Pérez-Marín, D. (2013). NIRS technology
493 for fast authentication of green asparagus grown under organic and conventional production
494 systems. *Postharvest Biology and Technology, 85,* 116–123.
495 Tatsadjieu, N.L., Maïworé, J., Hadjia, M.B., Loiseau, G., Montet, D., & Mbofung, C.M.F.
496 (2010). Study of the microbial diversity of *Oreochromis niloticus* of three lakes of Cameroon
497 by PCR-DGGE: Application to the determination of the geographical origin. *Food Control, 21,*
498 673–678.
499 Tenenhaus, M., 1998. La régression PLS. Editions Technip, Paris.
500 Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.M. & Lauro, C. (2005). Computational
501 Statistics & Data Analysis, vol. 48, issue 1, pages 159–205.

Computational Statistics & Data Analysis, 2005, vol. 48, issue 1, pages 159-205.

Tres, A., Van Der Veer, G., Perez-Marin, M.D., Van Ruth S.M., & Garrido-Varo, A. (2012). Authentication of Organic Feed by Near-Infrared Spectroscopy Combined with Chemometrics: A Feasibility Study. *J Agric Food Chem., 60,* 8129–33.

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

Food Safety for the 21st Century: Managing HACCP and Food Safety Throughout the Global Supply Chain

Wallace, C.A., Sperber, W.H., & Mortimore, S.E. (2011). Food Safety for the 21st Century: Managing HACCP and Food Safety Throughout the Global Supply Chain. Oxford, United Kingdom. *Wiley-Blackwell (John Wiley and Sons Ltd.),* 2011. 352 pages.

Figure captions

Table 1. Geographical and technical data of the harvesting sites of banana samples from Martinique and Dominican Republic.

Figure 1. DGGE profiles of bacterial rDNA (16S rDNA) of Cavendish bananas from (A) Martinique and (B) Dominican Republic and from different production modes.

Figure 2. Visualization on the first two components of the correlations between the Xs (bacterial DNA bands, "Bactn") and the components, and the Ys (production mode) and the components. (A) Bananas from Martinique and (B) from Dominican Republic.

Figure 3. DGGE data analysis of bacterial flora of bananas from (A) Martinique and (B) Dominican Republic by PLS-DA. The importance of the variables X (VIP or Bactn) is shown in the second component for both.

The histograms below the VIP graphics attest to the quality of the PLS-DA approach in function of the number of components. The quality and significance of the results obtained are represented by R^2Y cum, R^2X cum and Q^2 cum values. X representing the explanatory variables (or bacterial DNA bands), and Y the dependent variables (or the farming type/sampling year/geographical origin/variety of bananas). The closer to 1 the values are, the better the model is.

Figure 4. DGGE profiles of bacterial rDNA (16S rDNA) of bananas from (A) Martinique and (B) Dominican Republic and from different production modes and from two varieties (Cavendish and CIRAD925).

(CB = Untreated Cavendish bananas, CC = Treated Cavendish bananas)

Figure 5. DGGE data analysis by PLS-DA of bananas samples from Martinique and Dominican Republic. (A) Comparison between the effect of treatment and the effect of variety of bananas from Martinique (CB = Untreated Cavendish bananas, CC = Treated Cavendish bananas), (B) Comparison between the effect of treatment and the effect of geographical origin of Cavendish bananas from Martinique Dominican Republic. The histograms below attest to the quality of the PLS-DA approach in function of the number of components: four for the Figure A and three for Figure B.

The quality and significance of the results obtained are represented by R^2Y cum, R^2X cum and Q^2 cum values. X representing the explanatory variables (or bacterial DNA bands), and Y the dependent variables (or the farming type/geographical origin/variety of bananas). More the values are close to 1, better is the model.

Localisation	Type	Sampling (year)	Total (n bananas)	GPS	
Basse-Pointe (Martinique)	CIRAD 925	2013	16	14°50'26.74"N	61° 7'14.90"O
Basse-Pointe (Martinique)	Cavendish “Untreated” (CB)*	2013	15	14°50'26.74"N	61° 7'14.90"O
Basse-Pointe (Martinique)	Cavendish Treated (CC)	2013	15	14°50'41.45"N	61° 7'10.10"O
La Caida, Mao (Dominican Republic)	Cavendish Organic	2013/2014	30	19°35'23.52"N	71°10'20.25"O
Boca de Mao, Mao (Dominican Republic)	Cavendish Conventional	2013/2014	30	19°35'16.46"N	71° 2'40.57"O

*”Untreated” and not “organic” because these bananas were not certified

Table 1.

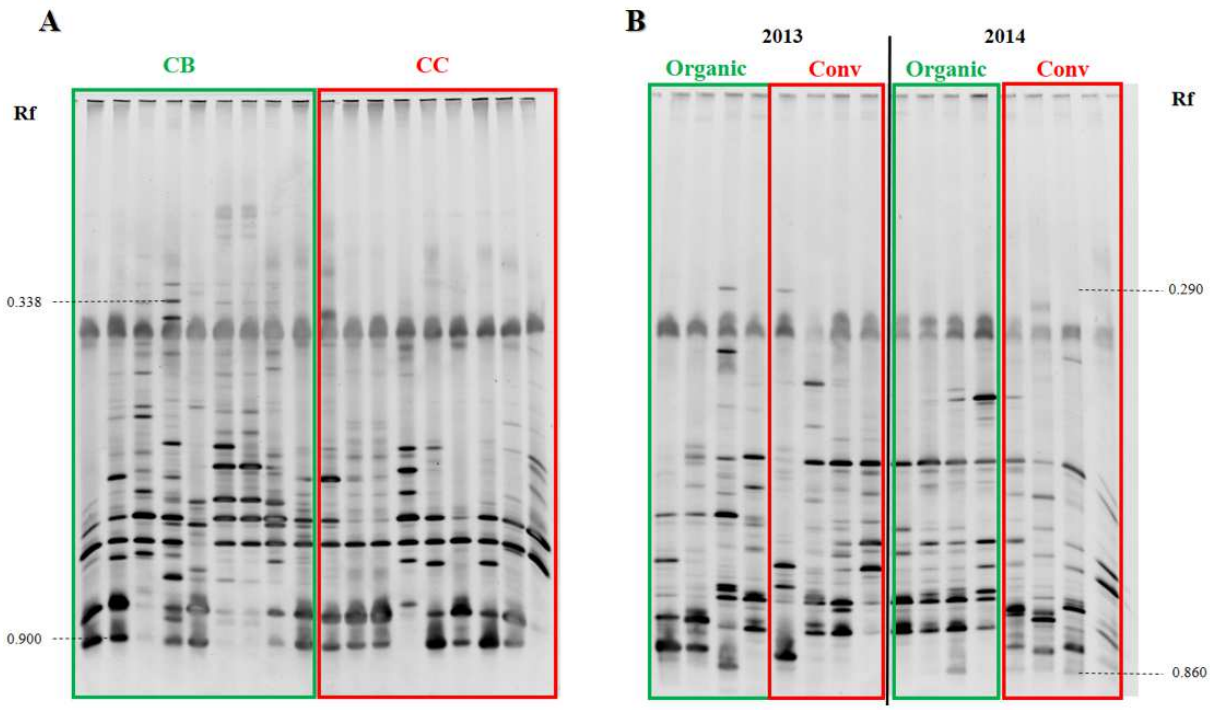
A
CB
CC
Rf
0.338
0.900
B
2013
2014
Organic
Conv
Organic
Conv
Rf
0.290
0.860

Figure 1.

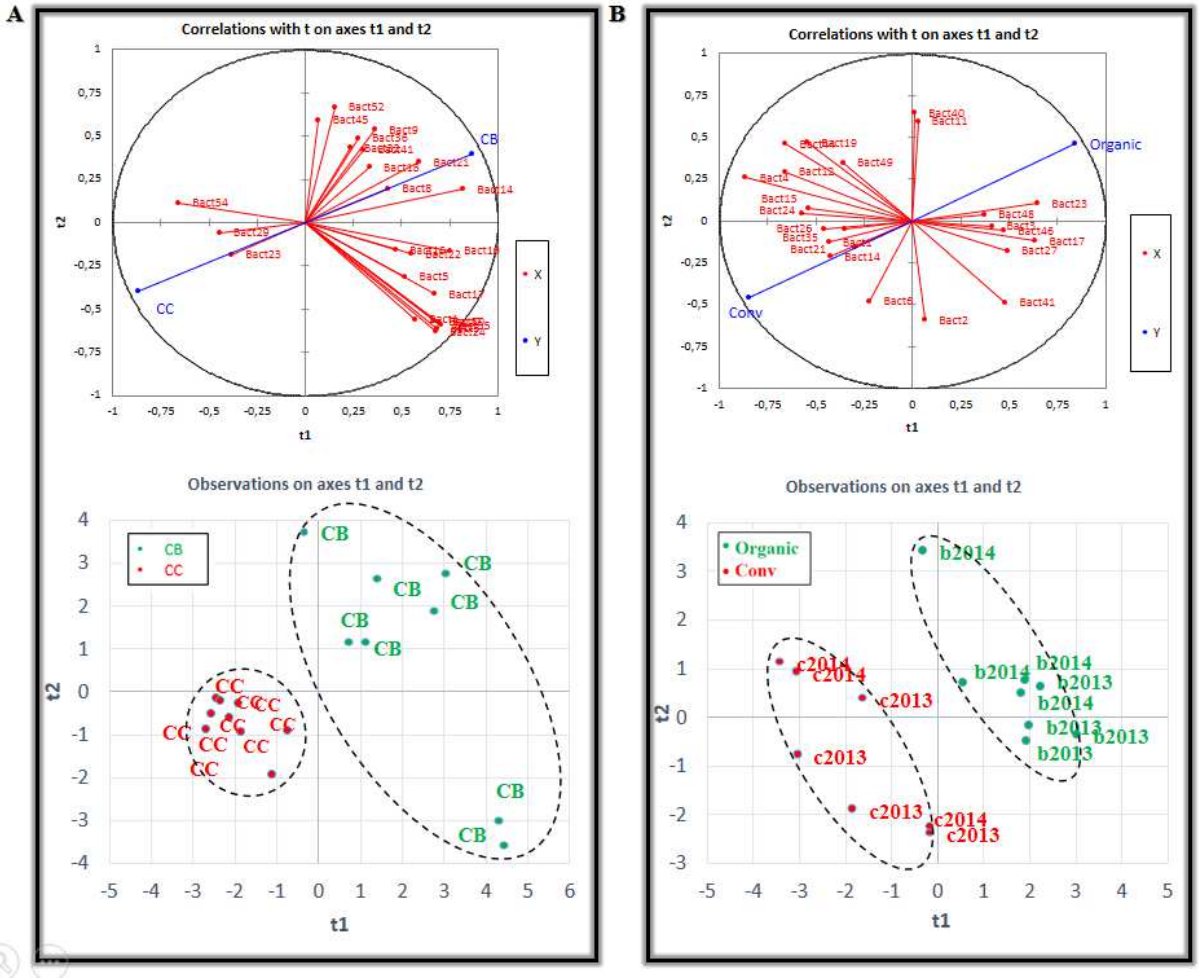
A
Correlations with t on axes t1 and t2
CB
CC
X
Y
t1
t2
Observations on axes t1 and t2
CB
CC
B
Correlations with t on axes t1 and t2
Organic
Conv
X
Y
t1
t2
Observations on axes t1 and t2
Organic
Conv
b2014
b2013
c2014
c2013
t1

Figure 2.

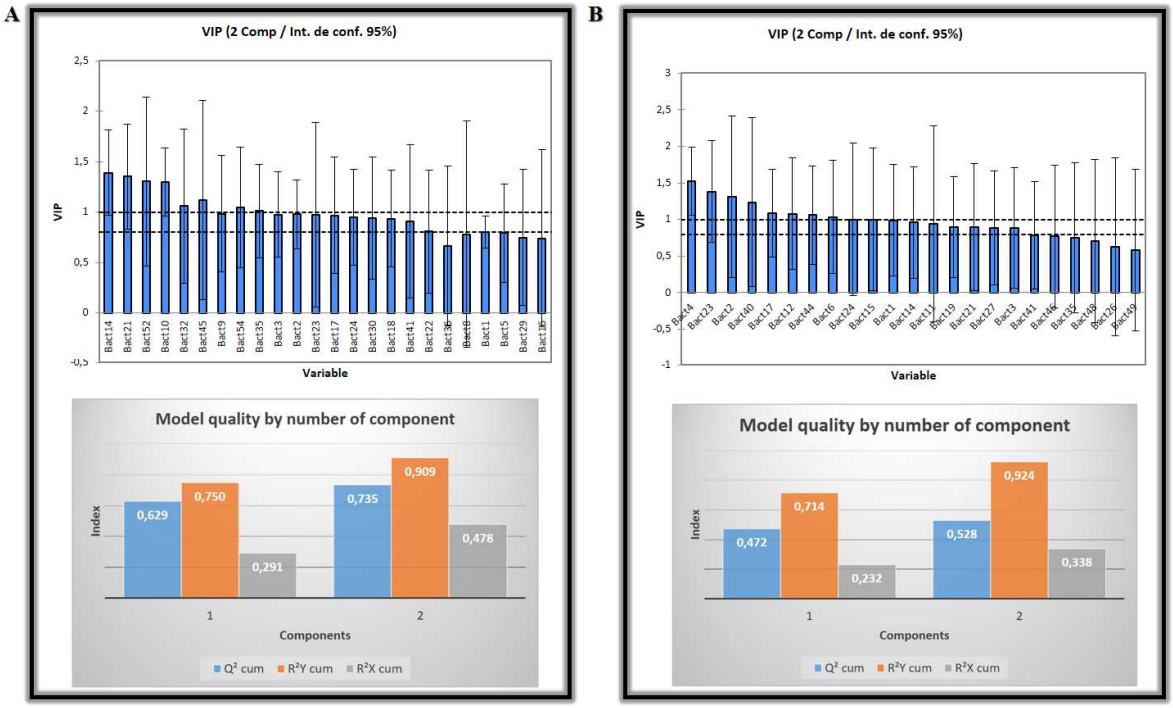
A
VIP (2 Comp / Int. de conf. 95%)
VIP
Variable
Model quality by number of component
Index
0,629
0,750
0,291
0,735
0,909
0,478
Components
Q² cum
R²Y cum
R²X cum
B
VIP (2 Comp / Int. de conf. 95%)
VIP
Variable
Model quality by number of component
Index
0,472
0,714
0,232
0,528
0,924
0,338
Components
Q² cum
R²Y cum
R²X cum

Figure 3.

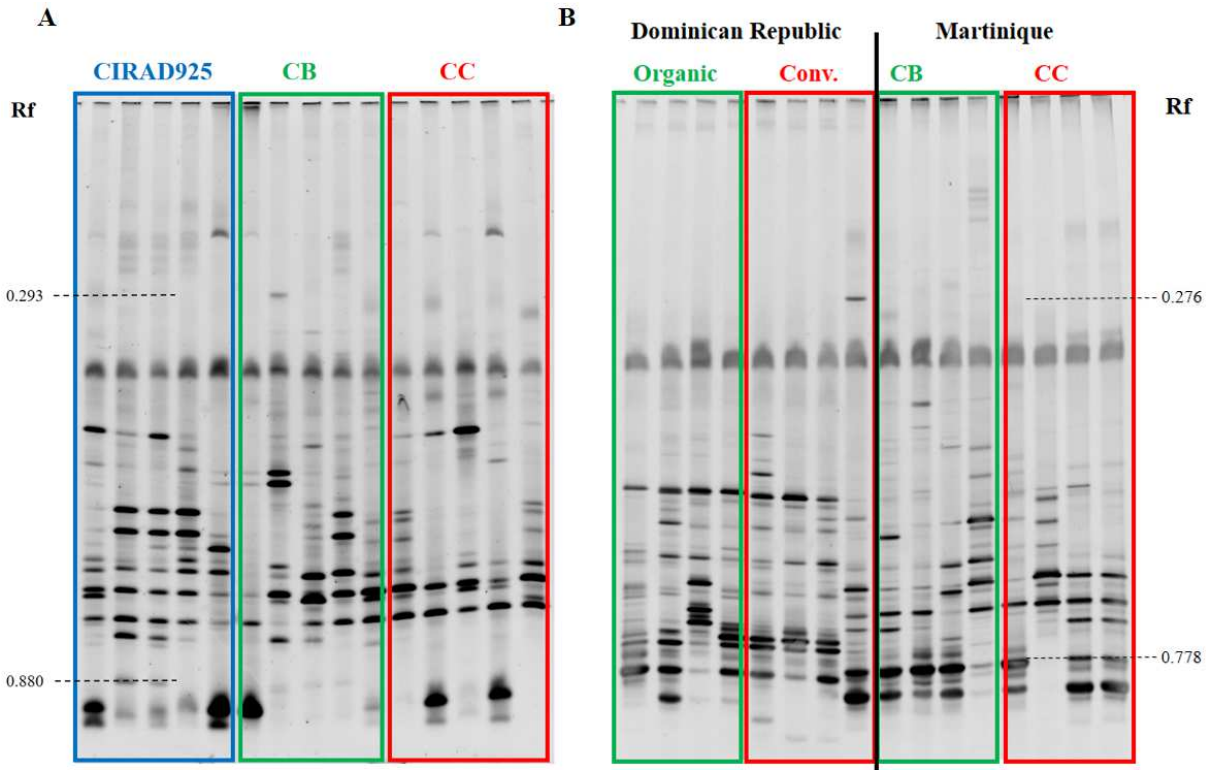
A
CIRAD925
CB
CC
Rf
0.293
0.880
B
Dominican Republic
Martinique
Organic
Conv.
CB
CC
Rf
0.276
0.778

Figure 4.

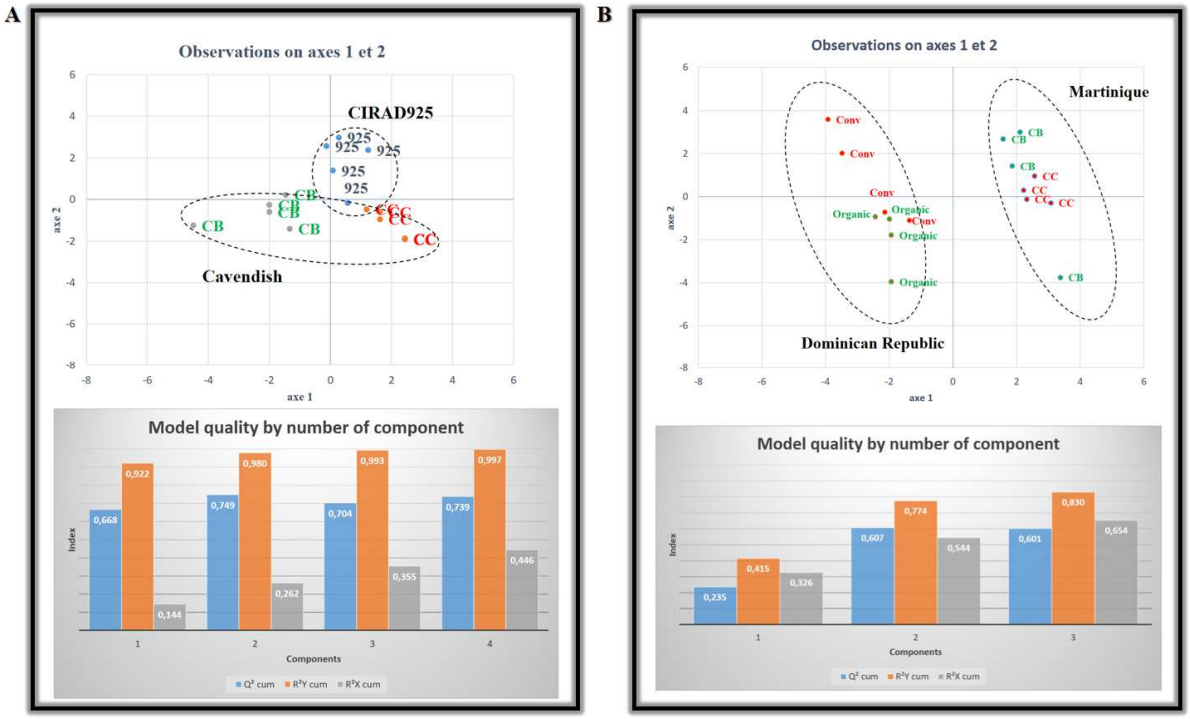
A
Observations on axes 1 et 2
CIRAD925
925
925
925
925
925
CB
CB
CB
CB
CB
CC
CC
CC
CC
Cavendish
axe 2
axe 1
Model quality by number of component
0,668
0,922
0,144
0,749
0,980
0,262
0,704
0,993
0,355
0,739
0,997
0,446
Index
Components
Q² cum
R²Y cum
R²X cum
B
Observations on axes 1 et 2
Martinique
Conv
Conv
Conv
Organic
Organic
Conv
Organic
Organic
CB
CB
CB
CC
CC
CC
CC
CB
Dominican Republic
axe 2
axe 1
Model quality by number of component
0,235
0,415
0,326
0,607
0,774
0,544
0,601
0,830
0,654
Index
Components
Q² cum
R²Y cum
R²X cum

Figure 5.