



Investigating the fermentation of cocoa by correlating Denaturing Gradient Gel Electrophoresis profiles and Near Infrared spectra

Dennis S. Nielsen*, Pia Snitkjaer, Frans van den Berg

Department of Food Science, Faculty of Life Sciences, Centre for Advanced Food Studies (LMC), University of Copenhagen, Denmark

ARTICLE INFO

Article history:

Received 28 September 2007
Received in revised form 6 March 2008
Accepted 24 March 2008

Keywords:

DGGE
NIR spectroscopy
Multivariate data analysis
Cocoa

ABSTRACT

Raw cocoa has an astringent, unpleasant taste and flavour, and has to be fermented, dried and roasted in order to obtain the characteristic cocoa flavour and taste. During the fermentation microbial activity outside the cocoa beans induces biochemical and physical changes inside the beans. The process is complex involving activity of several different groups of microorganisms which bring about numerous biochemical and physical changes inside the beans. Due to the complexity of these processes no thorough investigations of the interactions between the microbial activities on the outside of the beans and the chemical processes inside the beans have been carried out previously.

Recently it has been shown that Denaturing Gradient Gel Electrophoresis (DGGE) offers an efficient tool for monitoring the microbiological changes taking place during the fermentation of cocoa. Near Infrared (NIR) spectroscopy has previously been used to determine various components in cocoa beans, offering a rapid alternative compared to traditional analytical methods for obtaining knowledge about changes in the chemical composition of the cocoa beans during fermentation.

During a number of cocoa fermentations bean samples were taken with 24 h intervals to be dried and analysed by NIR. Cocoa pulp samples taken simultaneously during the same fermentations have previously been characterised using DGGE [Nielsen, D.S., Teniola, O.D., Ban-Koffi, L., Owusu, M., Andersson, T., Holzapfel, W.H. (2007). The microbiology of Ghanaian cocoa fermentations analysed using culture dependent and culture-independent methods. *International Journal of Food Microbiology* 114, 168–186.]. Here we report the first study where microbiological changes during the fermentation determined using DGGE are correlated to changes inside the beans determined by NIR using multivariate data analysis.

Following data pre-processing (baseline correction followed by Co-shift correction or Correlation Optimised Warping) the DGGE spectra were analysed using Principal Component Analysis (PCA). A clear grouping according to fermentation time was seen demonstrating the microbial succession taking place during the fermentation. Subsequently the DGGE spectra were correlated to the NIR spectra using Partial Least Squares regression models (PLS2). Correlations of 0.87 (bacterial derived DGGE spectra) and 0.81 (yeast derived DGGE spectra) were obtained indicating the relationship between the microbial activities in the pulp and the (bio) chemical changes inside the beans. By comparing the X-block loadings of the PLS2 models and the DGGE spectra it was possible to directly link several microbial species with changes in the NIR spectra and consequently also with changes inside the beans.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Cocoa beans are the principal raw material of chocolate. West Africa produces more than two-third of the World's cocoa with Cote d'Ivoire and Ghana alone accounting for 40 and 20% of the World production, respectively (Anon., 2005). Cocoa beans originate as seeds in fruit pods of the tree *Theobroma cacao*, where each fruit pod

contains 30 to 40 beans embedded in a mucilaginous pulp. Raw cocoa has an astringent, unpleasant taste and flavour, and has to be fermented, dried and roasted in order to obtain the characteristic cocoa flavour and taste (Thompson et al., 2001).

The fermentation of cocoa is a spontaneous process. Following opening of the pods the mucilaginous, acidic and sugar rich pulp surrounding the cocoa beans is contaminated with a variety of microorganisms originating from workers hands, containers used for transport, knives, pod surfaces, etc. (Roelofsen, 1958; Thompson et al., 2001; Jespersen et al., 2005). During the fermentation various yeasts, lactic acid bacteria (LAB), acetic acid bacteria (AAB) and possibly *Bacillus* spp. develop in a form of succession carrying out the fermentation (Roelofsen, 1958; Schwan et al., 1995; Thompson et al.,

* Corresponding author. Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark. Tel.: +45 35 33 32 87; fax: +45 35 33 32 14.

E-mail address: dn@life.ku.dk (D.S. Nielsen).

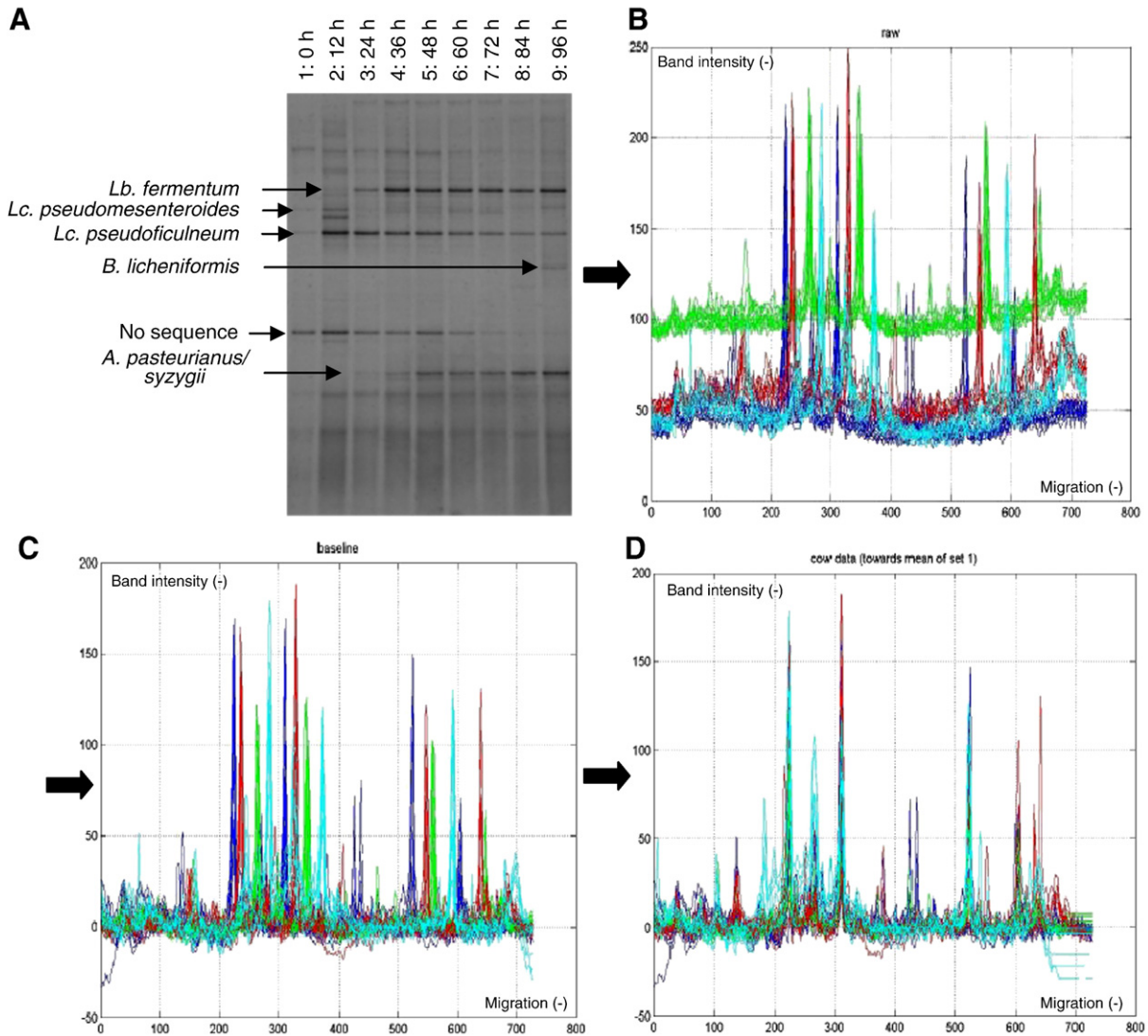


Fig. 1. Data pre-treatment; A) DGGE profiles (35–65% denaturant) representing 16S rRNA gene fragments (bacteria) of cocoa pulp sampled with 12 h intervals during a small heap (50 kg) fermentation. Identity of fragments identified by DNA sequencing indicated, see arrows. *Lb.*: *Lactobacillus*, *Lc.*: *Leuconostoc*, *B.*: *Bacillus*, *A.*: *Acetobacter*. Reprinted with permission from Nielsen et al. (2007); B) plot of raw spectra of the DGGE profiles; C) spectra of DGGE gels after baseline removed; D) spectra of DGGE gels after Correlation Optimized Warping (COW).

2001; Nielsen et al., 2007). The yeasts and LAB primarily metabolise the fermentable pulp sugars to ethanol and lactic acid. Subsequently part of the ethanol is further oxidised to acetic acid via an exothermal process through the activity of AAB. The ethanol and acetic acid cross the testa and penetrate the beans which, in combination with the heat produced, kills the germ and breaks down the cell walls in the bean initiating the processes leading to well fermented beans (Thompson et al., 2001; Schwan and Wheals, 2004; Nielsen et al., 2007). The fresh cocoa bean has an approximate composition of 32–39% water, 30–32% fat, 8–10% proteins, 5–6% polyphenols, 4–6% starch, 4–6% pentosans, 2–3% cellulose, 2–3% sucrose, 1–2% theobromine, 1% acids and 1% caffeine (Lopez and Dimick, 1995). Following breakdown of the cell walls in the bean numerous biochemical processes take place leading to the breakdown of proteins to peptides and amino acids, sucrose to fructose, and glucose and anthocyanins to anthocyanidins and sugars (galactose and arabinose). Later the polyphenols (including the anthocyanidins) are oxidised and polymerize to insoluble high-molecular-weight compounds (tannins). Furthermore there is a net efflux of polyphenols and to a lesser extent theobromine from the beans during the fermentation. Taken together these processes lead to

a colour change in the beans from grey over violet to brown, a reduction of the bitterness and astringency associated with raw cocoa and formation of precursors important for flavour and development during the subsequent drying and roasting processes (Zak and Keeney, 1976; Timbie et al., 1978; Biehl et al., 1982; Lehrian and Patterson, 1983; Wood and Lass, 1985; Kirchoff et al., 1989; Wollgast and Anklam, 2000; Thompson et al., 2001).

It is well established that the microbial activities taking place in the pulp surrounding the beans have a decisive influence on the biochemical processes in the interior of the bean. However, it has never been fully elucidated how the different processes are connected. Due to the large number of microorganisms involved and the complex microbial interactions taking place it is indeed tedious to investigate the microbiology of cocoa fermentations using traditional culture-based methods. Likewise the many complex processes taking place inside the beans would make an investigation of more than just a few biochemical parameters a tremendous task. As a consequence no thorough investigations of the interactions between the microbial activities on the outside of the beans and the chemical processes inside the beans have been carried out.

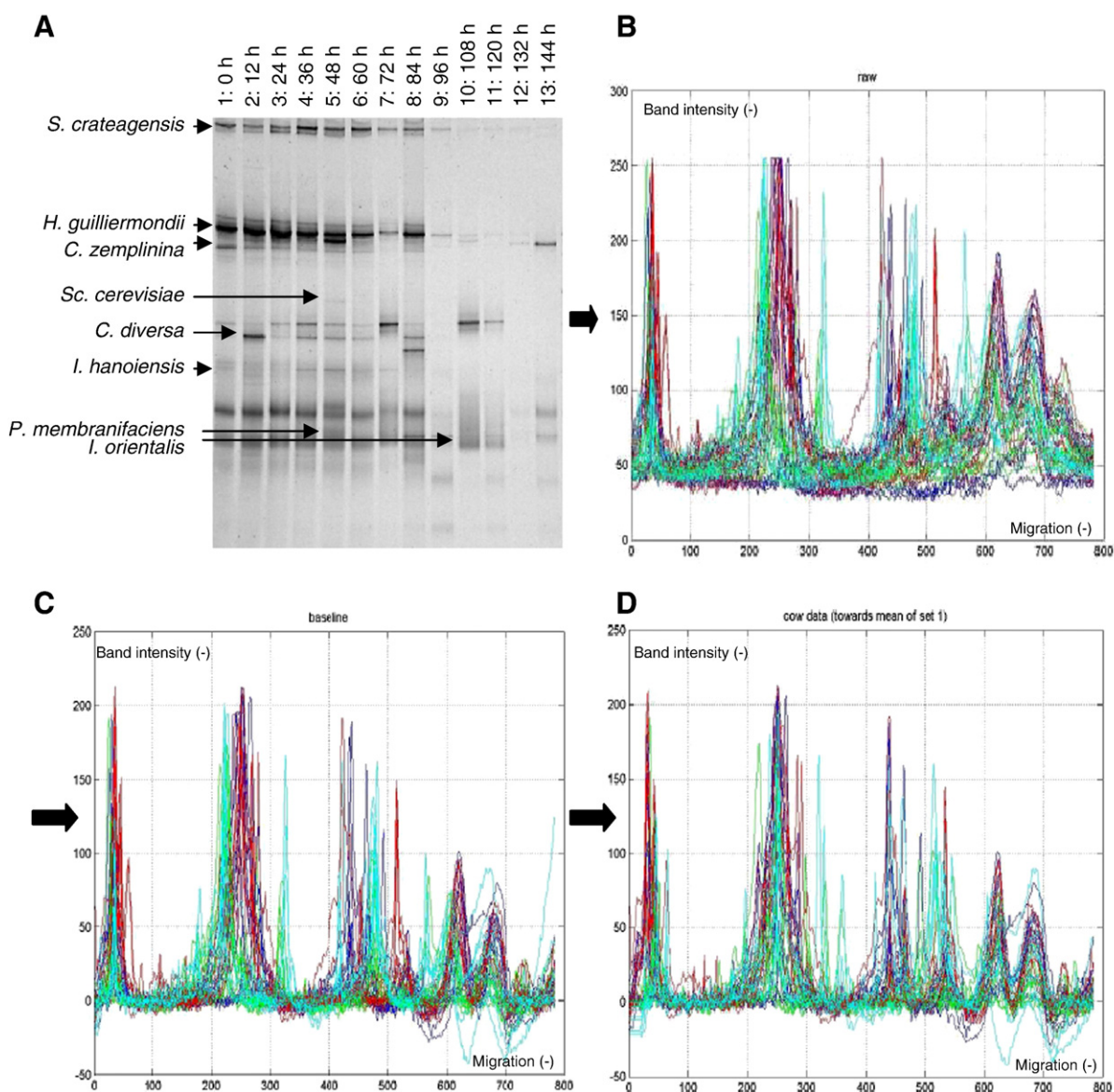


Fig. 2. Data pre-treatment; A) DGGE profiles (35–65% denaturant) representing 26S rRNA gene fragments (yeast) of cocoa pulp sampled with 12 h intervals from the outer parts/top of a large heap fermentation (500 kg). Identity of fragments identified by DNA sequencing indicated, see arrows. *H.*: *Hanseniaspora*, *P.*: *Pichia*, *C.*: *Candida*, *Sc.*: *Saccharomyces*, *I.*: *Issatchenkia*, *S.*: *Saccharomycopsis*. Reprinted with permission from Nielsen et al. (2007); B) plot of raw spectra of the DGGE profiles; C) spectra of DGGE gels after baseline removed; D) spectra of DGGE gels after Correlation Optimized Warping (COW).

Denaturing Gradient Gel Electrophoresis (DGGE) offers a rapid culture-independent alternative to the traditional culture-based microbiological methods. DGGE is based on sequence-specific separation of PCR-derived rRNA gene amplicons in polyacrylamide gels containing a linearly increasing concentration of denaturant (urea and formamide; Muyzer and Smalla, 1998). Recently it has been shown that it is possible to monitor the microbiological changes taking place during the fermentation of cocoa using DGGE; however, the DGGE profiles have never been directly correlated to the changes taking place inside the beans (Nielsen et al., 2005; Nielsen et al., 2007). Various attempts have been made to correlate DGGE profiles to other parameters using multivariate data analysis, especially in soil and water microbiology where DGGE profiles have been correlated to parameters such as heavy metal contamination, crop/land-use management regimes and water quality in lakes using Principal Component Analysis (PCA; Van der Gucht et al., 2001; Sekiguchi et al., 2002; Clegg et al., 2003; Gremion et al., 2004).

Traditionally, the degree of fermentation in cocoa beans is determined using the cut test where the dried beans are cut lengthwise and the internal colour is examined. Grey (slaty) beans are unfermented, purple beans are partly fermented and brown beans are fully fermented. As the cut test is based on visual assessment it is subjective by nature and alternative methods for quality determination of cocoa beans are needed (Rohan, 1963; Lopez, 1984; Lopez and Dimick, 1995). Near Infrared (NIR) spectroscopy has been used to determine and predict fat, proteins, carbohydrates, moisture, sensory quality and proanthocyanidins (polyphenols) in ground cocoa beans and offers a rapid alternative for obtaining knowledge about changes among the chemical constituents of the cocoa beans during fermentation (Kafka et al., 1982; Permanyer and Perez, 1989; Davies et al., 1991; Whitacre et al., 2003). Until now no attempt has been made to correlate microbial profiles and NIR spectra of cocoa beans sampled during the fermentation. Here we report the first research into correlating DGGE based microbial profiles of fermenting cocoa

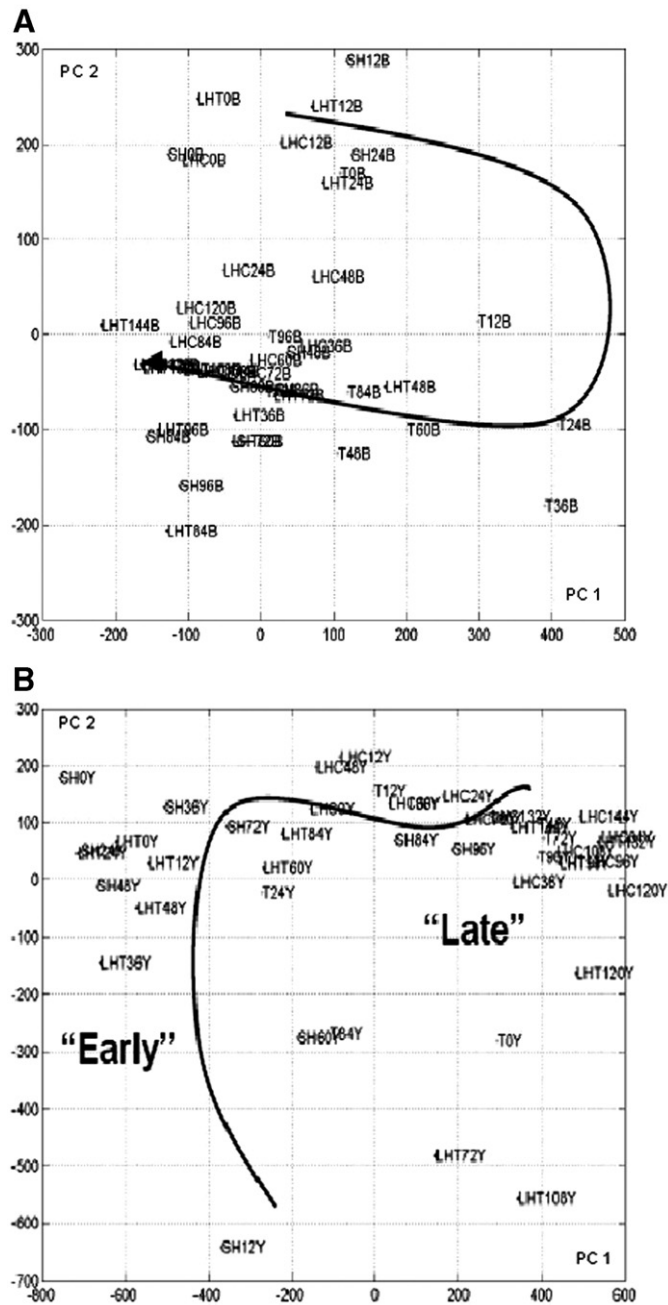


Fig. 3. Principal Component Analysis of the A) bacterial (arrow indicate progress of fermentation) and B) yeast derived DGGE spectra (sampled every 12 h) following Correlation Optimized Warping (COW). Abbreviations: T: Tray fermentation; SH: Small Heap Fermentation; LHT: Large Heap Top fermentation; LHC: Large Heap Centre fermentation. Arrow indicate progress of fermentation.

with NIR spectra with the aim of obtaining a deeper understanding of the interactions between microorganisms outside the beans and the biochemical reactions taking place inside the beans leading to well fermented cocoa.

2. Materials and methods

2.1. Cocoa fermentations

Cocoa pods were harvested by traditional methods (October, ambient temperature during the day 28–30 °C) in Mixed Hybrid Cocoa plantations in Ghana, near New Tafo (Eastern Region). The cocoa beans were fermented either in tray or traditional heaps at the Cocoa

Research Institute of Ghana (CRIG), New Tafo as lined out in Nielsen et al. (2007). In short, the pods were harvested over three days and opened on the fourth day with a cutlass. Following opening of the pods the beans were placed on either plantain leaves (heap fermentations) or in wooden trays (tray fermentations). In the heap fermentations approximately 50 kg (small heap) or 500–700 kg (large heap) of beans were piled on and covered with plantain leaves, and left to ferment. The cocoa beans were fermented for 4 (small heap fermentation, no turning) or six days (large heap fermentations, turning after 48 and 96 h). In the tray fermentations approximately 100 kg of beans was placed in each tray [90 × 120 cm, 10 cm deep, with holes drilled in the base and at the sides to allow aeration and drainage of liquids (sweatings) produced during the fermentations], eight trays were stacked on top of each other, the top tray covered with plantain leaves and left to ferment for four days.

2.2. Sampling

At 12 to 24 h intervals approximately 200 g of beans was sampled into a sterile plastic bag (Nielsen et al., 2007). From the small and large heap fermentations samples were taken approximately 15 cm from the surface. From the large heap fermentation additional samples were taken from the centre of the fermenting mass. From the tray fermentations samples were taken in the centre of the fermenting mass. Each sample was divided into two fractions. From one fraction of the beans pulp was aseptically scraped of the beans and freeze-dried for later DNA extraction and DGGE analysis as described in Nielsen et al. (2007). From the other fraction the pulp was removed from the beans by rubbing with saw dust. Following cleaning the beans were oven-dried (72 h, 70 °C).

2.3. Vis/NIR spectroscopy

Test the beans were frozen (−20 °C) and ground in a coffee grinder (KSM2, Braun, Germany) for 30 s to obtain a fine powder. Using a small ring cup with approximately 5 g of ground cocoa powder and a Spinning Module (NR6506) the beans were measured at room temperature (22 °C) using a Near Infrared spectrometer (NIRSystem 6500, Foss NIRSystem Inc., USA). The Vis/NIR spectra were determined over the range 400–2500 nm in 2-nm intervals, giving a total of 1050 variables, using a split detection system with a silicon detector between 400 and 1100 nm and a lead sulphide detector from 1100 to 2500 nm using back scatter; the reflectance was measured at a 45° angle. The software “NIRS2 ISIMENU” version 3.10 was used for data collection. Each spectrum was the average of 16 scans and all samples were analysed in duplicate. The results were recorded as $\text{Log}1/R$, where R is the Reflectance energy. A pectin standard sample was recorded on each measuring day as a precaution against possible drift of the NIR instrument. The results were analysed by means of multivariate data analysis using Unscrambler version 9.2 (CAMO Process A/S, Norway). A PCA with all replicates was used for detecting outliers among the replicates and after removal of detected outliers the mean value was used for further analysis. The data was mean-centred, and the validation of models was performed with full cross-validation.

2.4. Data pre-treatment of DGGE profiles

Denaturing Gradient Gel Electrophoresis profiles of the samples have been determined in a previous study (Nielsen et al., 2007), where DNA extraction, PCR and DGGE conditions are lined out in detail. In short the bacterial and eukaryotic micro-populations of the fermentation were investigated by analysing 16S and 26S rRNA gene fragments obtained using PCR and universal prokaryotic primers (V3-region of the 16S rRNA gene, primers PRBA338fGC/PRUN518r) and eukaryotic universal primers (D1/D2-region of the 26S rRNA gene, NL1GC and

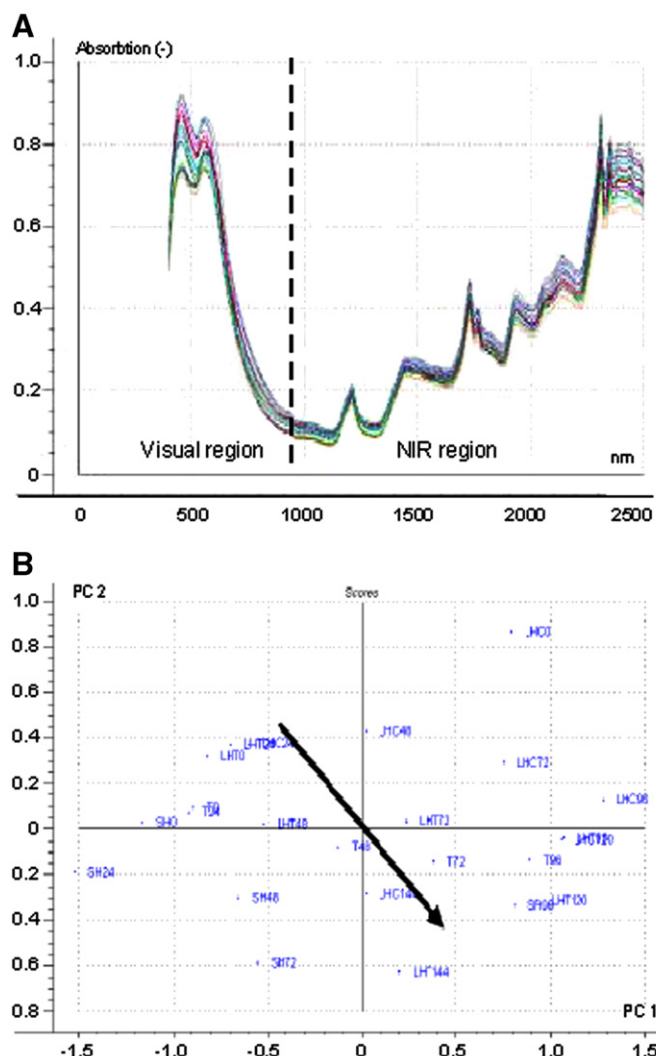


Fig. 4. Vis/NIR spectra of cocoa beans; A) line-plot; B) PCA plot (progress of fermentation tendency line indicated). Abbreviations: T: Tray fermentation; SH: Small Heap Fermentation; LHT: Large Heap Top Fermentation; LHC: Large Heap Centre Fermentation.

LS2), respectively, by DGGE (35–65% denaturant; (Nielsen et al., 2007). The samples for the DGGE spectra have been collected with 12 h intervals. The aim of the present study is to correlate the DGGE spectra with NIR spectra. However, samples for the NIR spectra have only been collected with 24 h intervals. In the pre-processing of the DGGE spectra it was decided to use all DGGE spectra in order to work on a larger number of samples. The DGGE profiles (in the form of image Tiff-files) were converted into numerical values using Matlab v7.2 (Mathworks Inc., USA) and exported into an Excel sheet (Microsoft Inc., USA). The pictures of the gels include a small area outside the gels and some smear in the start and the end of the gels which was corrected for by determining “Start” and “End” (cut off values) manually. Subsequently the baseline of the profiles was determined and removed using Matlab v7.2. A Matlab script fitting a second degree spline with three knots was used for this purpose. By this method the spectra were divided into four equal sections by three “knots”. Between each set of two knots a second order polynomial was fitted to the spectral points with smooths over the knots, and removed from the measured spectra as background. Data points furthest away from this fitted base are gradually removed until 20% of the data entries remain. This last fitting was then used as baseline and subtracted from the signal.

Finally the DGGE profiles were shift corrected using two different Matlab routines: Co-Shift correction and Correlation Optimised Warping (COW; Tomasi et al., 2004). These shift corrections are based on alignment of samples towards one common reference, selected to be one of the samples in the data set, via programming aimed at optimizing the overall correlation between one sample and the reference. The ultimate goal is to make the signals for different measurements more similar by removing small, undesirable day-to-day variations in retention time for the gel chromatograms without disrupting the true information (Tomasi et al., 2004). Co-Shift is a linear pre-processing method where a reference sample and point-wise shifted samples are compared by computing correlation coefficients. The best correction is obtained with by maximum correlation coefficient (Tomasi et al., 2004). Correlation Optimized Warping is a segmented-linear pre-processing method aimed at aligning a sample data vector towards a reference vector by allowing small changes in the segment lengths. The correlation coefficient of each segment is calculated, the correlation coefficients are summed and the best correction found at maximum via a dynamic programming algorithm (Tomasi et al., 2004).

2.5. Data analysis

The data were analysed by means of multivariate data analysis using Unscrambler v9.2 and Matlab v7.2. Principal Component Analysis (PCA) models were determined for all sets of NIR and DGGE spectral data. The DGGE spectra were correlated to the NIR data using PLS2 regression models. Principal Components Analysis finds the best least-squares low rank approximation of a data matrix \mathbf{X} (Martens and Næs, 1989)

$$\mathbf{X} = \mathbf{u}_1 \cdot s_1 \cdot \mathbf{v}_1^T + \mathbf{u}_2 \cdot s_2 \cdot \mathbf{v}_2^T + \dots + \mathbf{E}_X = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T + \mathbf{E}_X$$

$$\text{minimize} \|\mathbf{X} - \mathbf{U} \cdot \text{diagonal}(\mathbf{S}) \cdot \mathbf{V}^T\|^2$$

where the matrix product $\mathbf{U}\mathbf{S}\mathbf{V}^T$ is the singular value decomposition and \mathbf{E}_X is the un-modeled part of \mathbf{X} . In this work \mathbf{X} is assumed column mean-centred before analysis. To eliminate redundancy in the decomposition the singular values are usually included in the object-score matrices

$$\mathbf{X} = \mathbf{t}_1 \cdot \mathbf{p}_1^T + \mathbf{t}_2 \cdot \mathbf{p}_2^T + \dots + \mathbf{E}_X = \mathbf{T} \cdot \mathbf{P}^T + \mathbf{E}_X$$

where object-score vectors fulfill $\mathbf{t}_i \cdot \mathbf{t}_i^T = s_i$ and $\mathbf{t}_i \cdot \mathbf{t}_j^T = 0$, and variable-loadings define criteria $\mathbf{p}_i \cdot \mathbf{p}_i^T = 1$ and $\mathbf{p}_i \cdot \mathbf{p}_j^T = 0$. Hence, the first set of scores and loadings is the best approximation of the original data, and the fraction/percentage explained variance captured from the original data matrix by this first pair expresses how well this approximation succeeded. Similarly, the second pair is the next best approximation. The scores can be seen as new pseudo-values for the objects (samples collected at different time-points of fermentation in this paper) representing their position inside the data set. The loadings show the role of the original variables (migrational behavior in the DGGE profiles/spectra or wavelength axis in the NIR spectra in this paper) in determining the principal components.

In PLS2 regression the aim is to not only explain the variance in \mathbf{X} (the DGGE chromatograms in this paper), but to do this in such a way that \mathbf{X} is correlated as well as possible with variance in a second matrix \mathbf{Y} (NIR spectra in this paper; hence, maximizing the covariance between the two blocks). This can be achieved by solving the following three equations

$$\mathbf{X} = \mathbf{t}_1 \cdot \mathbf{p}_1^T + \mathbf{t}_2 \cdot \mathbf{p}_2^T + \dots + \mathbf{E}_X = \mathbf{T} \cdot \mathbf{P}^T + \mathbf{E}_X$$

$$\mathbf{Y} = \mathbf{u}_1 \cdot \mathbf{q}_1^T + \mathbf{u}_2 \cdot \mathbf{q}_2^T + \dots + \mathbf{E}_Y = \mathbf{U} \cdot \mathbf{Q}^T + \mathbf{E}_Y$$

$$\text{maximize}(\text{covariance}(\mathbf{t}_i, \mathbf{u}_i) | \mathbf{X}\mathbf{w}_i = \mathbf{t}_i, \|\mathbf{w}_i\| = 1)$$

where two sets of object-score vectors are obtained (note that \mathbf{t}_1 in PCA is not the same as for the PLS2 solution, and some of the

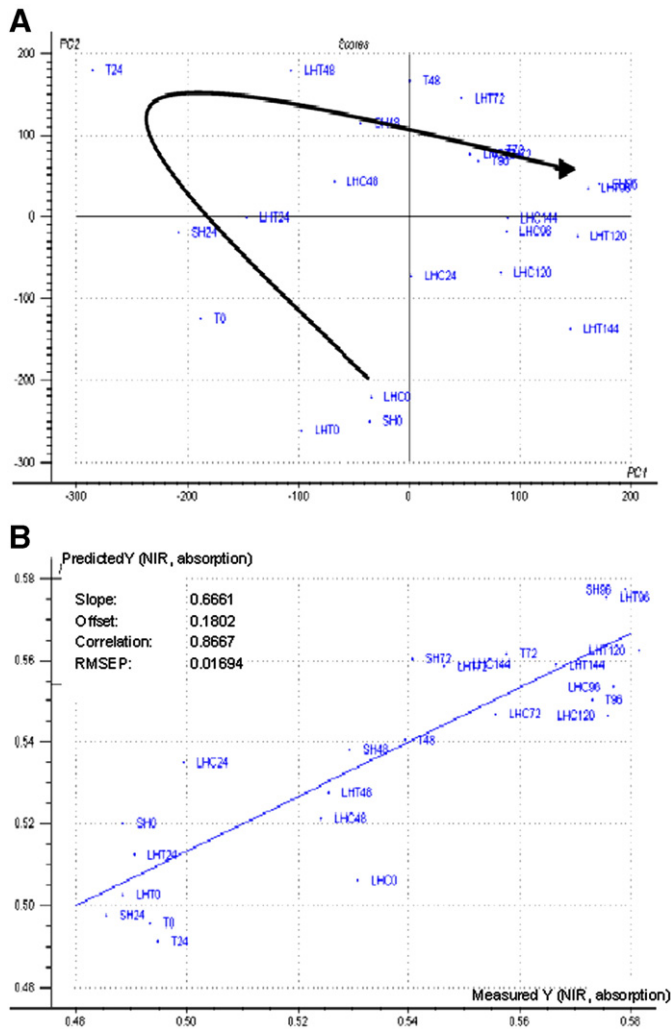


Fig. 5. Bacterial derived DGGE spectra (COW corrected) correlated to Vis/NIR spectra of cocoa beans. A) PC1 vs. PC2 score-plot, trend line (fermentation time) shown; B) predicted vs. measured plot (2 PC's). Abbreviations: T: Tray fermentation; SH: Small Heap Fermentation; LHT: Large Heap Top fermentation; Large Heap Centre fermentation.

orthogonality criteria from PCA are relaxed) and variable-loadings \mathbf{p}_i (for the DGGE block) and \mathbf{q}_i (for the NIR matrix). Hence, the first set of scores and loadings is the best approximation of the original data matrices under the condition or constraint that the covariance between the two object-score vectors is maximized. The auxiliary vectors \mathbf{w}_i are used to compute the final regression model $\mathbf{Y} = \mathbf{X} \cdot \mathbf{B} = \mathbf{X} \cdot \mathbf{W} \cdot (\mathbf{P}^T \cdot \mathbf{W})^{-1}$ (Martens and Næs, 1989). The regression models can be used to “predict” NIR spectra from DGGE signals in e.g. cross-validation. However, in the setting of this paper the PLS2 algorithm should be seen as a multi-block algorithm where two sources of information (DGGE and NIR) are analysed simultaneously in order to utilize and compare the information included in each.

3. Results and discussion

Fermentation is the first post harvest step in the process ultimately leading to cocoa beans with the desired “chocolate” taste and flavour. The process is truly complex, involving a wide range of different microorganisms (Schwan and Wheals, 2004). Denaturing Gradient Gel Electrophoresis offers a rapid alternative to traditional culture-based microbiological methods for investigating the micro-populations involved in the fermentation. Conceptually, each band in the gel represents a single microbial species (Fig. 1). The DGGE method is in

general seen as “semi-quantitative”. However, due to potential PCR-bias, bias in DNA extraction and potential co-migration of bands the gels should be interpreted with care (Giraffa and Neviani, 2001; Prakitchaiwattana et al., 2004; Nielsen et al., 2005). Different primer sets were used for the PCR–DGGE-based analysis of the prokaryotic and eukaryotic organisms involved in the fermentation of cocoa (Nielsen et al., 2007). Consequently two different sets of DGGE profiles were obtained from each fermentation experiment: one set representing the prokaryotic (Fig. 1A) microflora and another set representing the eukaryotic (Fig. 2A) microflora.

3.1. Pre-processing of DGGE spectra

Plotting the raw spectra derived from the bacterial DGGE profiles from the different fermentations together shows that the baseline differs from gel to gel (Figs. 1B and 2B). The baseline differences were compensated for by baseline correction carried out by removing a spline of second order with three knots from the spectra using an in-house algorithm (Figs. 1C and 2C). As the majority of the fragments on the DGGE gels have been previously identified by sequencing (Nielsen et al., 2007) it was possible to unambiguously identify that fragments (peaks) showed undesirable lane to lane and gel to gel variations in

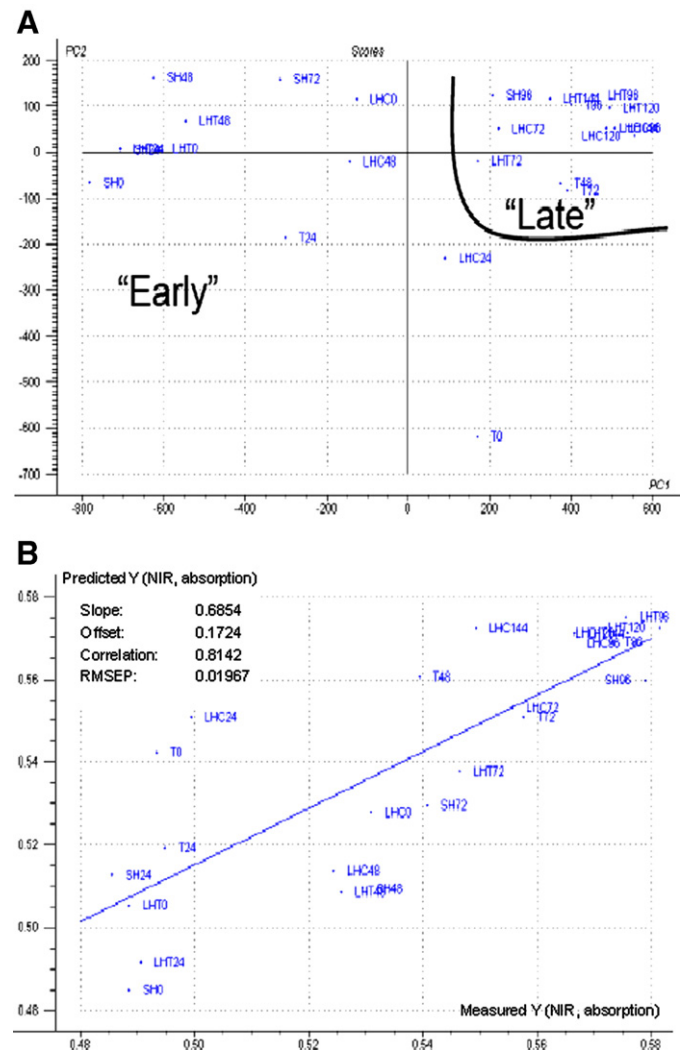


Fig. 6. Yeast derived DGGE spectra (COW corrected) correlated to Vis/NIR spectra of cocoa beans. A) PC1 vs. PC2 score-plot; B) Predicted vs. measured plot (2 PC's). Abbreviations: T: Tray fermentation; SH: Small Heap Fermentation; LHT: Large Heap Top fermentation; Large Heap Centre fermentation.

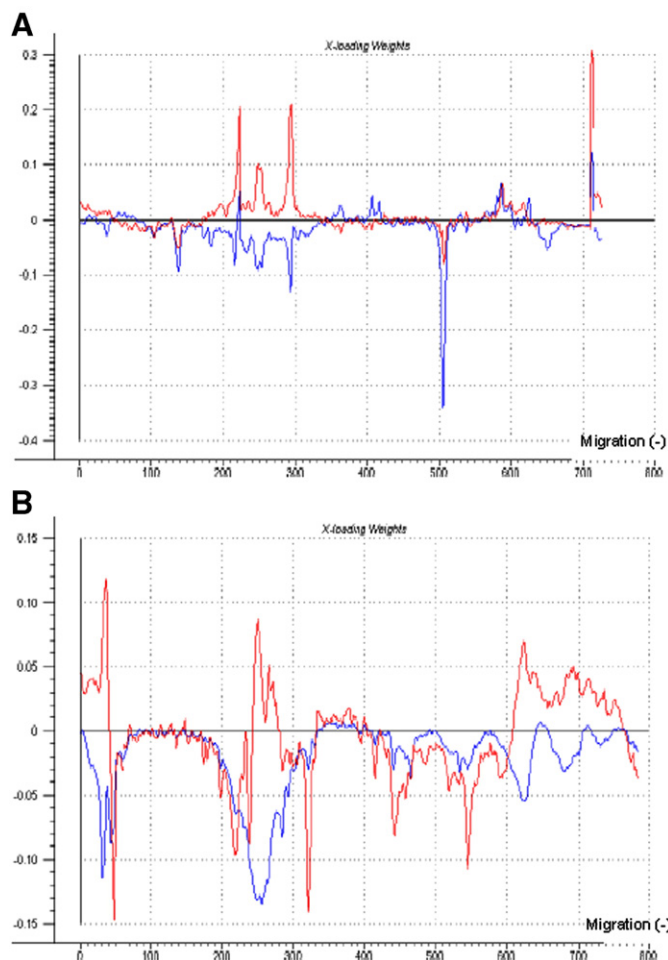


Fig. 7. X-loadings of the PLS2-models, A) bacterial derived DGGE spectra; B) yeast derived DGGE spectra.

their migrational behaviour (shifts). It was attempted to correct the shifting of the peaks using two different warping methods: Correlation Optimized Shifting (Co-Shift) and Correlation Optimized Warping (COW). A better resolution between different fragments (peaks) was obtained using (non-linear) COW as compared to (linear) Co-Shifting (Figs. 1D and 2D; not all results shown). This was especially pronounced for fragments showing almost the same migrational behaviour. Principal Component Analysis (PCA) plots of the COW corrected bacterial and yeast derived DGGE spectra show that the bacterial spectra distribute according to fermentation time (Fig. 3A). For the yeast derived spectra only grouping according to “early” and “late” fermentation is seen (Fig. 3B). However, this is expected, as yeasts are mainly active in the early parts of cocoa fermentations (Schwan and Wheals, 2004).

3.2. Visual/Near Infrared (Vis/NIR) spectra

Spectroscopic methods such as NIR have proven to be valuable tools for rapid determination of chemical constituents in food and feed stuff and for monitoring changes during processing (Cen and He, 2007). Near Infrared spectroscopy provides complex structural information on the samples investigated and has been applied to predict the content of a wide range of constituents in food and feed samples. In cocoa and cocoa products NIR spectroscopy has been used to predict the flavour potential of the beans and quality parameters such as fat, protein and carbohydrate content (Kafka et al., 1982; Permanyer and Perez, 1989; Davies et al., 1991). As seen from the PCA

plot in Fig. 4B the Vis/NIR spectra (Fig. 4A) of cocoa bean samples taken from fermentations carried out using 3 different fermentation methods distribute according to fermentation time with the samples fermented for only a short time distributing in the upper left part of the plot, and the more fermented samples in the lower right part of the plot. The Vis/NIR spectra are as seen from Fig. 4A laying on top of each other and are slightly turned which could indicate a small, almost negligible, scatter effect in the measurements. It was determined not to apply data pre-processing to compensate for this physical effect.

3.3. Correlation of DGGE spectra and Vis/NIR spectra

To correlate the changes taking place inside the beans during fermentation with the microbial activities taking place outside the beans Vis/NIR spectra of dried cocoa beans (Fig. 4) were correlated to the DGGE spectra (Figs. 1 and 2) sampled at the same times and positions during fermentation. A PLS2 regression between the Vis/NIR spectra and baseline plus COW corrected bacterial derived DGGE spectra is seen in Fig. 5. As seen from Fig. 5A a clear grouping according to fermentation time of the samples is observed. The residual variance showed local minimum after 2 PLS Components (PC's) and global minimum after 6 PC's (results not shown). To avoid overfitting the model with only 2 PC's is used. A correlation coefficient of 0.87 (Root Mean Squared Error of Cross-Validation, RMSECV=0.17) is obtained for the model (Fig. 5B). Compared to the PLS2 model based on bacterial derived DGGE spectra a less clear grouping according to fermentation time is seen when the model is based on yeast derived DGGE spectra (Fig. 6A). However, a quite clear grouping into “early” and “late” fermentation is still observed. This is explained by the fact that the major activities of the yeast during cocoa fermentations take place during the first 24–48 h of fermentation and late in the fermentation only few “yeast induced” changes are taking place in the beans (Schwan and Wheals, 2004). The residual variance showed local minimum after 2 PC's which was thus chosen for the PLS2-model. A correlation of 0.81 (RMSECV=0.20) is obtained for the model (Fig. 6B).

A comparison between the X-block loadings of the PLS2-models (Fig. 7A and B) and the DGGE spectra (Figs. 1 and 2) clearly indicates that the three peaks around 200–300 and the peak at 500 in the bacterial derived DGGE spectra are strongly correlated to the changes in the Vis/NIR spectra. Previously we have shown that the peaks around 200–300 represent the lactic acid bacteria *Lactobacillus fermentum*, *Leuconostoc pseudomesenteroides* and *Leuconostoc pseudoficulneum* whereas the peak at 500 represents acetic acid bacteria (AAB; Nielsen et al., 2007). From the X-loadings plot it can be seen that the peaks at 200–300 mainly influences PC1, whereas the peak at 500 representing AAB strongly influences PC2 (Fig. 7). The AAB produces acetic acid and heat which causes the cell walls inside the cocoa beans to collapse inducing the biochemical reactions in the bean leading to well fermented cocoa (Schwan and Wheals, 2004). According to previous findings the AAB reaches high numbers after 24 h of fermentation (Nielsen et al., 2007). As seen from the score-plot (Fig. 5A) the 24 h samples have moved along the second axis (PC2) compared to the zero-hour samples, the only exception being sample LHC0. From a biological point of view this makes good sense, as the LHC samples were collected in the centre of the fermenting mass where the oxygen tension is very low during the first days of fermentation. As growth of AAB is oxygen demanding this severely limits the growth of the AAB during the initial phases of fermentation as was shown by Nielsen et al. (2007). For the yeast derived DGGE spectra it is seen that the strong peaks around 50 and 240, representing *Saccharomycopsis crataegensis* and *Hanseniaspora guilliermondii*, strongly influences the model. From the X-loadings plot it is seen that these two peaks strongly influences PC1 as well as PC2 (Fig. 7). Furthermore peaks around 450, 550 and 600 representing among other *Pichia membranifaciens* influences the

results (Fig. 7). *Saccharomycopsis crataegensis* and *H. guilliermondii* are mainly active in the early phases of the fermentation explaining the separation into “early” and “late” phase as seen in Fig. 6 and discussed above.

4. Conclusion

In this study we show that correlation of microbial DGGE profiles with Vis/NIR spectra could offer a tool for correlating microbial activity with biochemical changes in the beans. Prediction of the NIR spectra from the DGGE spectra gave acceptable correlations taking the highly variable sample material and the limited number of samples into account. From a biological point of view the resulting PLS2 models could be interpreted with specific groups of microorganisms (DGGE spectra) being linked to the development observed in the NIR spectra. In conclusion correlation between microbiological molecular based methods such as DGGE and spectroscopic methods such as Vis/NIR seems to offer a potentially strong tool for unrevealing the connections between microbial activity and changes within the product.

Acknowledgements

The work performed was partly financed through the EU INCO project: “Developing biochemical and molecular markers for determining quality assurance in the primary processing of cocoa in West Africa - COCOQual” (ICA4-CT-2002-10040). The cooperation and invaluable technical assistance of the Cocoa Research Institute of Ghana (Dr. J.S. Takrama) is highly appreciated.

References

- Anon. (2005). Annual Report 2003/2004. International Cocoa Organization, www.icco.org.
- Biehl, B., Wewetzer, C., Passern, D., 1982. Vacular (storage) proteins of cocoa seeds and their degradation during germination and fermentation. *Journal of the Science of Food and Agriculture* 33, 1291–1304.
- Cen, H., He, Y., 2007. Theory and application of Near Infrared Reflectance Spectroscopy in determination of food quality. *Trends in Food Science and Technology* 18, 72–83.
- Clegg, C.D., Lovell, R.D.L., Hobbs, P.J., 2003. The impact of grassland management regime on the community structure of selected bacterial groups in soils. *FEMS Microbiology Ecology* 43, 263–270.
- Davies, A.M.C., Franklin, J.G., Grant, A., Griffiths, N.M., Shepherd, R., Fenwick, G.R., 1991. Prediction of chocolate quality from Near-Infrared spectroscopic measurements of the raw cocoa beans. *Vibrational Spectroscopy* 2, 161–172.
- Giraffa, G., Neviani, E., 2001. DNA-based, culture-independent strategies for evaluating microbial communities in food-associated ecosystems. *International Journal of Food Microbiology* 67, 19–34.
- Gremion, F., Chatzinotas, A., Kaufmann, K., Von Sigler, W., Harms, H., 2004. Impacts of heavy metal contamination and phytoremediation on a microbial community during a twelve-month microcosm experiment. *FEMS Microbiology Ecology* 48, 273–283.
- Jespersen, L., Nielsen, D.S., Hønholt, S., Jakobsen, M., 2005. Occurrence and diversity of yeasts involved in fermentation of West African cocoa beans. *FEMS Yeast Research* 5, 441–453.
- Kafka, K.J., Norris, K.H., Kulcsar, F., Draskovits, I., 1982. Attempts to determine fat, protein and carbohydrate content in cocoa powder by the NIR technique. *Acta Alimentaria* 11, 271–288.
- Kirchhoff, P.-M., Biehl, B., Ziegler-Berghausen, H., Hammor, M., Lieberei, R., 1989. Kinetics of the formation of free amino acids in cocoa seeds during fermentation. *Food Chemistry* 34, 161–179.
- Lehrian, D.W., Patterson, G.R., 1983. Cocoa fermentation. In: Reed, G. (Ed.), *Food and Feed Production with Microorganisms*. Verlag Chemie, Weinheim, pp. 529–575.
- Lopez, A.S., 1984. Limitação da “prova de corte” no controle de qualidade do cacau comercial (In Spanish). *Revista Theobroma* 14, 199–207.
- Lopez, A.S., Dimick, P.S., 1995. Cocoa fermentation. In: Reed, G., Nagodawithana, T.W. (Eds.), *Enzymes, Biomass, Food and Feed*. VCH, Weinheim, pp. 561–577.
- Martens, H., Næs, T., 1989. Multivariate calibration. In: Martens, H., Næs, T. (Eds.), *Methods for Calibration*. John Wiley & Sons, Chichester.
- Muyzer, G., 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.
- Nielsen, D.S., Hønholt, S., Tano-Debrah, K., Jespersen, L., 2005. Yeast populations associated with Ghanaian cocoa fermentations investigated using Denaturing Gradient Gel Electrophoresis (DGGE). *Yeast* 22, 271–284.
- Nielsen, D.S., Teniola, O.D., Ban-Koffi, L., Owusu, M., Andersson, T., Holzapfel, W.H., 2007. The microbiology of Ghanaian cocoa fermentations analysed using culture dependent and culture independent methods. *International Journal of Food Microbiology* 114, 168–186.
- Permany, J.J., Perez, M.L., 1989. Compositional analysis of powdered cocoa products by Near-Infrared Reflectance Spectroscopy. *Journal of Food Science* 54, 768–769.
- Prakitchaiwattana, C.J., Fleet, G.H., Heard, G.M., 2004. Application and evaluation of denaturing gradient gel electrophoresis to analyse the yeast ecology of wine grapes. *FEMS Yeast Research* 4, 865–877.
- Roelofs, P.A., 1958. Fermentation, drying, and storage of cacao beans. *Advances in Food Research* 8, 225–296.
- Rohan, T.A., 1963. *Processing of Raw Cocoa for the Market*. FAO, Rome.
- Schwan, R.F., Wheals, A.E., 2004. The microbiology of cocoa fermentation and its role in chocolate quality. *Critical Reviews in Food Science and Nutrition* 44, 205–222.
- Schwan, R.F., Rose, A.H., Board, R.G., 1995. Microbial fermentation of cocoa beans, with emphasis on enzymatic degradation of the pulp. *Journal of Applied Bacteriology (Supplement)* 79, 965–1075.
- Sekiguchi, H., Watanabe, M., Nakahara, T., Xu, B.H., Uchiyama, H., 2002. Succession of bacterial community structure along the Changjiang River determined by denaturing gradient gel electrophoresis and clone library analysis. *Applied and Environmental Microbiology* 68, 5142–5150.
- Thompson, S.S., Miller, K.B., Lopez, A.S., 2001. Cocoa and coffee. In: Doyle, M.J., Beuchat, L.R., Montville, T.J. (Eds.), *Food Microbiology—Fundamentals and Frontiers*. ASM Press, Washington, D.C., pp. 721–733.
- Timbie, D.J., Sechrist, L., Keeney, P.G., 1978. Application of High-Pressure Liquid Chromatography to the study of variables affecting theobromine and caffeine concentrations in cocoa beans. *Journal of Food Science* 43, 560–565.
- Tomasi, G., van den Berg, F., Andersson, C., 2004. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics* 18, 231–242.
- Van der Gucht, K., Sabbe, K., De Meester, L., Vloemans, N., Zwart, G., Gillis, M., Vyverman, W., 2001. Contrasting bacterioplankton community composition and seasonal dynamics in two neighbouring hypertrophic freshwater lakes. *Environmental Microbiology* 3, 680–690.
- Whitacre, E., Oliver, J., van den Broek, R., van Engelen, P., Kremers, B., van der Horst, B., Stewart, M., Jansen-Beuvinck, A., 2003. Predictive analysis of cocoa procyanidins using near-infrared spectroscopy techniques. *Journal of Food Science* 68, 2618–2622.
- Wollgast, J., Anklam, E., 2000. Review on polyphenols in *Theobroma cacao*: changes in composition during the manufacture of chocolate and methodology for identification and quantification. *Food Research International* 33, 423–447.
- Wood, G.A.R., Lass, R.A., 1985. *Cocoa*. Longman, London.
- Zak, D.L., Keeney, P.G., 1976. Changes in cocoa proteins during ripening of fruit, fermentation, and further processing of cocoa beans. *Journal of Agricultural and Food Chemistry* 24, 483–486.