# Monitoring of Water Spectral Pattern of lactobacilli development as a tool for rapid selection of probiotic candidates

Slavchev, Aleksandar ; Kovacs, Zoltan ; Koshiba, Haruki ; Bazar, Gyorgy ; Pollner, Bernhard ; Krastanov, Albert ; Tsenkova, Roumiana

(Citation)
Journal of Near Infrared Spectroscopy,25(6):423-431

(Issue Date)
2017-10-20

(Resource Type)
journal article

(Version)
Accepted Manuscript

# Monitoring of Water Spectral Patterns of Lactobacilli Development as a Tool for Rapid Selection of Probiotic Candidates

| | |
|---|---|
| Journal: | *Journal of Near Infrared Spectroscopy* |
| Manuscript ID | JNS-17-0025.R2 |
| Manuscript Type: | Original Research Article |
| Date Submitted by the Author: | 14-Oct-2017 |
| Complete List of Authors: | Slavchev, Aleksandar; University of Food Technologies, Microbiology; Kobe University, Graduate School of Agricultural Science, Biomeasurement Technology Laboratory<br>Kovács, Zoltán ; Szent Istvan Egyetem, Department of Physics and Control; Kobe University, Graduate School of Agricultural Science, Biomeasurement Technology Laboratory<br>Koshiba, Haruki; Kobe University, Graduate School of Agricultural Science, Biomeasurement Technology Laboratory<br>Bazar, Gyorgy; Kaposvár University, Faculty of Agricultural and Environmental Sciences, Department of Nutritional Science and Production Technology; Kobe University, Graduate School of Agricultural Science, Biomeasurement Technology Laboratory<br>Pollner, Bernhard; Medical University of Innsbruck, Department for Hygiene and Medical Microbiology; Kobe University, Graduate School of Agricultural Science, Biomeasurement Technology Laboratory<br>Krastanov, Albert; University of Food Technologies, Biotechnology<br>tsenkova, roumiana; Kobe University, Graduate School of Agricultural Science, Biomeasurement Technology Laboratory |
| Keywords: | probiotics, aquaphotomics, near infrared spectroscopy, NIR, chemometrics, lactic acid bacteria |
| Abstract: | Nowadays a quick and inexpensive method, which allows rapid, in vivo comprehensive probiotic bacteria identification, is needed. To elucidate a new concept to evaluate probiotic bacteria, NIR spectroscopy with aquaphotomics were applied to monitor the growth of eight Lactobacillus bulgaricus and one Lactobacillus gasseri bacteria strains. Their resistance to low pH (1.8) in the presence of pepsin and bile were measured and further used as reference data for analysis of the simultaneously acquired spectral data. The acquired spectral data in the region of 1100-1300 nm were subjected to various methods for multivariate data analyses – PCA, LDA, SIMCA and PLSR. The results showed high accuracy of bacteria strains classification according to their resistance and the potential of the tested wavelength region for rapid selection and prediction of some basic phenotypic characteristics of probiotic candidates. Results of the current study also revealed different suitability of each growth stage when using NIR spectra for the classification of bacteria strains. |
| | |

# Monitoring of Water Spectral Patterns of Lactobacilli Development as a Tool for Rapid Selection of Probiotic Candidates

Aleksandar Slavchev[1,2], Zoltan Kovacs[1,3], Haruki Koshiba[1], Gyorgy Bazar[1,4], Bernhard Pollner[1,5], Albert Krastanov[6,7], Roumiana Tsenkova[1*]

[1]Kobe University, Graduate School of Agricultural Science, Biomeasurement Technology Laboratory, Kobe, Japan

[2]University of Food Technologies, Department of Microbiology, Plovdiv, Bulgaria

[3]Szent István University, Faculty of Food Science, Department of Physics and Control, Budapest, Hungary

[4]Kaposvár University, Department of Nutritional Science and Production Technology, Faculty of Agricultural and Environmental Sciences, Kaposvár, Hungary

[5]Department for Hygiene and Medical Microbiology, Medical University of Innsbruck, Innsbruck, Austria

[6]University of Food Technologies, Department Biotechnology, Plovdiv, Bulgaria

[7]Bulgarian Academy of Science, Specialized Academic Unit for Traditional Bulgarian Dairy Products, Plovdiv, Bulgaria

Corresponding author: prof. Roumiana Tsenkova, Kobe University, Faculty of Agriculture,

Rokkodai-cho, Nada-ku, Kobe 657-8501, Kobe, Japan, Tel & Fax: +81 78 803 5911,

e-mail: rtsen@kobe-u.ac.jp

**Abstract**

Nowadays a quick and inexpensive method, which allows rapid, *in vivo* comprehensive probiotic bacteria identification, is needed. To elucidate a new concept to evaluate probiotic bacteria, NIR spectroscopy with aquaphotomics were applied to monitor the growth of eight *Lactobacillus bulgaricus* and one *Lactobacillus gasseri* bacteria strains. Their resistance to low pH (1.8) in the presence of pepsin and bile were measured and further used as reference data for analysis of the simultaneously acquired spectral data. The acquired spectral data in the region of 1100-1300 nm were subjected to various methods for multivariate data analyses – PCA, LDA, SIMCA and PLSR. The results showed high accuracy of bacteria strains classification according to their resistance and the potential of the tested wavelength region for rapid selection and prediction of some basic phenotypic characteristics of probiotic candidates. Results of the current study also revealed different suitability of each growth stage when using NIR spectra for the classification of bacteria strains.

Keywords: probiotics, aquaphotomics, near infrared spectroscopy, NIR, chemometrics, lactic acid bacteria.

## Introduction

Lactic acid bacteria (LABs) are a diverse group of gram-positive, non-sporulent bacteria, anaerobes or facultative anaerobes requiring specific nutrition environment, and whose ultimate metabolite product is lactic acid. Currently, LABs became microorganisms of industrial importance for their fermentation activity, as well as for their health and nutrition benefits. LABs, particularly lactic bacteria, which occupy important niches in the gastro-intestinal tract (GIT) are considered as offering numerous probiotic benefits to general health status and proper physical condition.

Probiotics are live microorganisms which when administered in adequate amounts confer a health benefit on the host. They express their positive effect by balancing of the gastro-intestinal tract microflora, improve the immune response [1], decrease the symptoms of lactose intolerance and allergies in susceptible individuals [2], reduce the risk of cancer [3], and alleviate irritable bowel syndrome and inflammatory bowel diseases [4,5]. The mechanism of probiotic activity has not been established yet, but it probably includes modification of pH levels of the GIT [6], pathogens antagonism through the production of antimicrobials [7], competition for receptor sites [8], nutrients and growth factors, stimulation of immune cells [9] and lactase production [10,11].

The most important probiotic characteristic is the ability of surviving the harsh environment in the upper gastro-intestinal tract. This is because during their passage through the stomach and the small intestine they face an environment with low pH and high concentrations of gall [11]. In the genus *Lactobacillus*, there are strains which differ in resistance to the conditions in the upper gastro-intestinal tract. Strains with high tolerance are of interest for the production of probiotics and

probiotic food. The main issue in this field is the selection of strains which exhibit strong probiotic characteristics. Currently, two main strategies have been applied for the selection of probiotic strains: selection of strains based on the presence of particular genes and *in vitro* examination of strain resistance [12,13]. These methods are time-consuming, require expensive equipment and consumables, and they give uncertain results. Therefore, a quick and inexpensive method, which allows rapid, *in vivo* comprehensive probiotic bacteria selection, is needed [11].

A new approach to this issue has been recently proposed. It is based on using near infrared (NIR) spectroscopy and aquaphotomics for real-time monitoring of growing lactobacilli to classify and discriminate strains possessing probiotic characteristics. Only by analysing their NIR spectra obtained in the time period 11.4-12 h of the cultivation time [11] has it been possible to predict their resistance to conditions similar to those in the upper gastro-intestinal tract.

The main goal of the study reported in this paper was to use data from the region between 1100 and 1300 nm to find the most suitable growth stage for *in vivo* spectral monitoring and discrimination of resistant and non-resistant *Lactobacillus* strains. A further aim was to predict the bile minimal inhibitory concentration and low pH resistance of *Lactobacillus* strains based on their NIR spectra obtained during the cultivation process.

## Materials and Methods

*Bacterial strains*

Five probiotic and four non-probiotic strains (genus *Lactobacillus*) possessing different bile salt

tolerance and ability to resist low pH (pH 1.80) in the presence of pepsin were used: probiotic strains *L. bulgaricus* S06, *L. gasseri* S20 (provided by "Selur Pharma" Ltd. - Bulgaria), *L. bulgaricus* S22, *L. bulgaricus* S11, *L. bulgaricus* S10 and non-probiotic strains - *L. bulgaricus* S09, *L. bulgaricus* Y12, *L. bulgaricus* S03 and *L. bulgaricus* S02. The strains were provided by "Specialized Academic Unit for Traditional Bulgarian Dairy Products within the Bulgarian Academy of Sciences" with exception of *L. bulgaricus* Y12 which was isolated from yoghurt. All microorganisms were freeze-dried and kept at -80 °C.

*Preparation of stock cultures and active bacterial culture*

The strains were cultivated in MRS broth (De Man, Rogosa and Sharpe agar; Merck, Tokyo, Japan) at 37 °C for 24 h. The biomass obtained after centrifugation at 5000 $min^{-1}$ for 5 min was twice washed with PBS (Phosphate-buffered saline) buffer (pH 7.00) and suspended in 15 %w/v glycerol solution to the initial volume and stored at -80 °C for further use.

Tubes containing 1 ml MRS broth were inoculated with 50 µl glycerol suspension of stock culture and cultivated for 18 – 20 h at 37 °C.

*Determination of the optical density of the bacterial cultures*

The optical density was determined by using a micro-plate reader iMark (BioRad, Hercules, California, USA) against MRS broth as blank at $\lambda$ = 665 nm. The sample volume was 150 µl with correction to 1 cm path length. Every optical density is presented as an average value of nine optical densities obtained from three independent samples (tubes or deep well plate wells) measured three

consecutive times.

*Determination of the resistance to low pH value in presence of pepsin*

A modified method of Pitino [13] was used. 750 µl MRS-broth were inoculated with 50 µl of active bacterial culture and cultivated for 18 – 20 h at 37 °C. The culture medium was then centrifuged at 10000 $min^{-1}$ for 5 min, the biomass was washed twice with PBS buffer (pH 7.00). The cells were suspended to the original volume with the low pH buffer (pH 1.8), containing HCl (0.2 M), NaCl (0.08 M), $CaCl_2$ (0.03 mM), and pepsin from porcine gastric mucosa (9000 U $ml^{-1}$; Wako, Japan). After 3 h cultivation at 37 °C the biomass was centrifuged at 10000 $min^{-1}$ for 5 min and washed with PBS buffer and re-suspended to its original volume with PBS buffer. Tubes containing 750 µl MRS broth were inoculated with 50 µl of low-pH treated cells suspensions and cultivated at 37 °C for 24 h. The optical density of the cultures was measured at λ=665 nm at 0h and 24 h. The resistance each strain to low pH in the presence of pepsin was evaluated by cell growth and is presented by the increase in the optical density of the culture medium after 24h cultivation at 37 °C (Yield of biomass after 3 h stay at pH 1.8 and 9000 U $ml^{-1}$ pepsin).

*Determination of bile minimal inhibitory concentration*

MRS broths (750 µl) with double-fold decreasing concentrations of dry bile (Wako, Japan) 0.156-5.000 mg/ml were inoculated with 50 µl active bacterial culture and cultivated at 37 °C for 24 h and the optical density at λ=665 nm was determined.

*Monitoring of the cultures by NIR Spectroscopy*

MRS broth (15 ml) was inoculated with about 0.5 ml active bacterial culture to OD=0.1 (λ=665) and cultivated at 37 °C for 24 h with shaking on a vibratory shaker in a 50 ml centrifuge tube. The NIR transflectance spectra of the culture were acquired in the entire spectral region (400-2500 nm) with 0.5 nm step (4200 data points) at every 4 min by using a FOSS XDS OptiProbe Analyzer attached with immersion type probe (FOSS NIRSystems, Inc., Hoganas, Sweden or Hilleroed, Denmark, recently distributed by Metrohm NIRSystems AG, Herisau, Switzerland). A reference spectrum was taken at the beginning of every measurement series by placing the immersion probe in the dark aperture position of the instrument. The spectra taken in the first 40 min of the cultivation time were discarded and those after 40 min until the scan of 20 h of the monitoring were used for data evaluation. Spectra acquisition was performed with the VISON 3.50 (FOSS NIRSystems, Inc., Hoganas, Sweden) software. After pre-experiments, 0.5 mm layer thickness (set by spacer) was found to be the most appropriate to achieve an applicable signal in the first overtone region of water.

*Spectral Data Analysis*

The water first overtone of combination vibrations at the wavelength range 1100-1300 nm was used for data evaluations. As a first step of spectral pretreatment, smoothing by using a Savitzky-Golay[14] filter with 21 data points and second polynomial order was applied. For eliminating the scattering effect MSC (multiplicative scatter correction) transformation[15] was performed.

Principal Component Analysis[16] (PCA) was used to discover the multidimensional pattern of

variations in the strains low pH and bile resistance scores and the NIR spectral dataset obtained at whole cultivation time (40-1200 min) and to provide orthogonal variables for supervised classifications of linear discriminant analysis using the spectral dataset of different growth stages.

Linear Discriminant Analysis (LDA)[17] was performed to generate supervised qualitative models to identify the groups of bacterial strains. The scores of the first ten principal components (PC) of the PCA models were used in LDA as input variables to build models for the classification of the three main groups of the bacteria strains. LDA models were tested with "one-strain-out" cross-validation (described in more details below) and with independent prediction using the spectra of *L. bulgaricus* S22, *L. bulgaricus* S10 and *L. bulgaricus* Y12.

To find relationship between spectral data and phenotype parameters of bacterial strains (the minimum inhibitory concentration of bile( MIC) and ability to recover after 3 h stay at low pH) we applied Partial Least Squares Regression[15] (PLSR). The PLSR models were evaluated by the coefficient of determination in cross-validation ($R^2_{cv}$) and standard error of cross-validation (SECV). The maximum number of latent variables was determined as $1/10^{th}$ of the number of observation (n) to avoid overfitting. The PLSR models were validated using the "one-strain-out" validation method. The data set was split into training and test sets. The spectral data of 8 strains were used as the training set; and those of one strain left, as the test set. This process of data splitting was repeated 9 times to ensure that the data of all the strains was included in the evaluation set once[17].

For qualitative assessment of strains for low-pH and bile tolerance Soft Independent Modeling of

Class Analogy (SIMCA) analysis was used. Raw spectra were mean cantered. The cross-validation was performed using "one-strain-out" in the same way as is described above.

Aquagrams[18] were calculated in order to show the differences of the absorbance values at the water matrix coordinates (WAMACs) for the group of resistant and non-resistant bacteria. The star-chart displays averaged normalized spectral absorbance values of the groups of resistant and non-resistant bacteria (acquired at the same conditions) cultivated in MRS broth. The spectra of each strain were acquired at 37 °C in the time intervals of 0.3-6.3 h and 6.3-12.3 h of the cultivation time (phases III and IV).

The scripts for PLSR, LDA and aquagram calculation and visualization were written and executed in R-project environment (RStudio Ver. 0.98 and R Ver. 3.0.1, R Foundation for Statistical Computing, Vienna, Austria). The calculation and visualization of PCA and SIMCA were performed with SimcaVer. 13.5 (Umetrics AB, USA; http://umetrics.com/) and Piroette 4 (Infometrix, USA; https://infometrix.com/software/).

## Results and Discussions

*Classification of the strains based on their bile minimal inhibitory concentration* (*MIC*) *and low pH resistance*

Based on the low pH tolerance (Yield of biomass after 3 h stay at pH 1.8 and 9000 U $ml^{-1}$pepsin) and the MIC for bile, all strains were divided in two groups – strains which were capable of growth after

3h stay at pH 1.8 in the presence of 9000 U $ml^{-1}$ pepsin and could survive and grow at high bile concentrations and strains which were not resistant to that environment. *L. bulgaricus* S06 showed medium results and was not included in either of the two groups (Table 1).

Table 1.

Strains varied in their resistance to low pH in the presence of pepsin and bile. These are essential characteristics, which allow strains to survive in the human gastro-intestinal tract. This is due to their adaptation ability and depends on the presence and the expression of some genes, which leads to differences in the levels of some proteins and enzyme activities [19].

PCA results based on the reference data i.e. bile tolerance and pH resistance for each strain are shown in Fig. 1.

Figure 1.

The first principal component (PC1) of this matrix explained 98.3 % of the total variance. Its scores were highly correlated with the probiotic properties of the strain and represented very well the ability of the strains to grow in the presence of bile and to survive at very low pH. On the score plot (Fig. 1) the group of resistant strains is very well separated from the group of non-resistant ones. Loadings of the model (indicated by the arrows) show that both parameters were correlated.

*Determination of Bacterial Growth Phases*

The typical microbial batch culture growth consists of four growth phases: Lag phase, Exponential phase, Stationary phase and Death phase [20]. They are not always clearly defined because of longer

transition periods which may occur due to different factors. In order to separate the growing curves into growth phases a modified model of Gompertz and its fourth derivative were applied [21,22] (Fig. 2).

$$y = A \cdot \exp\left\{-\exp\left[\frac{\mu_m e}{A}(\lambda - t) + 1\right]\right\},$$

where A is the maximal optical density, $\mu_m$ is the maximal specific growth rate, $\lambda$ is the length of the lag-phase and t is the time.

Figure 2.

The fourth derivative of the absorbance at 600 nm changes in the model was used to divide all growth curves into five phases (Fig. 2), which is in accordance with Widdel's model of typical growing bacterial culture [23]. The accurate separation of bacterial growth cycle is essential for further data analyses and allows us to compare strains at their respective growth stage and at equal physiological condition.

*Spectral data analyses*

Near infrared spectra of the growing bacterial cultures changed continuously during the whole cultivation period. In our previous work [11] we showed that the region of 1100 – 1850 nm can be used successfully to classify probiotic lactobacilli and predict some of their essential properties at certain periods during the cultivation period (11.4 – 12h) where the cultures are present in large numbers of similar cells, which are still viable. We have found that water absorbance bands in the first overtone region were the most informative for classification. Here, we are exploring the first overtone of the

combination vibrations of water, in the range 1100 – 1300 nm Figure 3), which has less absorbance and, therefore, it is most suitable when using longer path lengths.

Figure 3.

Raw spectra were MSC transformed to eliminate the scattering effect and baseline differences (Fig. 3), which makes them more suitable for further data analyses. The cultivation time was separated into growing phases by using 4th derivative of the modified Gompertz model and every time period was used to build LDA, SIMCA and PLS regression models.

Principal component analysis of spectral data acquired in the time period of 40-1200 min of cultivation showed that 99.95% of the whole variance was explained by PC1 (Fig. 4a). This variation was observed due to the spectral changes that occurred during the growing time, but it also showed the differences between the groups of resistant and non-resistant strains. All resistant strains were grouped in the right part of the score plot, while the non-resistant ones in the right side this separation were more prominent after Phase II.

Figure 4.

The loading vector (Fig. 4b) of PC1 shows picks at **1120**, **1158** and **1283**nm. These wavelengths are the most important for the separation of the strains and changed the most during the cultivation and show differences in free and protonated water molecules [24].

The spectral data for strains were analyzed by Linear discriminant analysis and Soft Independent Modeling of Class Analogy to test its potential for qualitative assessment of low-pH and bile

tolerance of strains.

The results of the classification matrix showed highest accuracy of prediction at earlier exponential phase (phase III and IV) for both LDA and SIMCA methods (Table 2.).

Best classification of the resistant and non-resistant groups was found in phase IV with LDA where the recognition ability was 98.6% and prediction ability was 100% for both cross-validation and independent prediction (Fig. 4c). However, LDA models built on data of phase III also showed 100% correct prediction ability (Table 2.).

Table 2.

Almost all of the spectra were classified correctly at the second and the third growth phases – 97.7% and 96.5% respectively with SIMCA.

The discriminating power of the SIMCA model (Fig. 4d) built on the spectral data of phase III showed peaks at 1116 nm, **1123** nm, 1147.5 nm, **1158.5** nm, 1184 nm, **1197** nm, 1225 nm, 1264 nm and **1285** nm. Both PCA and SIMCA models showed consistency in the wavelengths with highest importance for the discrimination of resistant and non-resistant *Lactobacillus* strains.

Partial least squire regression models (PLSR) were built to determine relationship between spectral data and the tolerance of strains to low pH and bile. Two models were built – one for prediction of the ability of a strain to survive during 3h stay at pH 1.8 and 9000 U $ml^{-1}$ pepsin (Fig. 5a and Fig. 5c) and another one for the bile minimal inhibitory concentration for a strain (Fig. 5b). The accuracy of the models was evaluated by their $R^2$ of cross validation and SECV (standard error of

cross-validation).

Figure 5.

For the prediction of low pH resistance, the best results were observed at phases IV and V, while for the prediction of bile MIC the most suitable phases were III and IV. The results of PLSR regression models showed close correlation and relatively low error of cross-validation. In both cases models using spectra from phase IV showed the highest accuracy, which confirms our previously obtained results [11].

The wavelengths which were important for the prediction of the low pH resistance of a strain are shown on Fig. 5c. The regression vector showed two groups of wavelengths in the whole region of 1100-1300 nm. The first group repeats wavelengths already discovered by SIMCA modelling (Fig. 4d) wavelengths – 1116, 1148.5, 1183, 1199 and 1225 nm, while the other wavelengths were unique for the regression model – 1134, 1154, 1176, 1190, 1234, 1241, 1255 and 1295 nm.

Another successfully applied aquaphotomics tool, aquagram, was used to examine the water combination band region. It proposes seven specific spectral ranges which are of importance for bacteria characterization. It is a star-chart which contains averaged normalized absorbance values at wavelengths in those regions of interest. These values contain information about water molecular conformations and their respective hydrogen and covalent bonds [11]. In this study we applied aquagrams to the shorter wavelength region of 1100-1300 nm in order to show the differences between non-resistant and resistant strains.

Figure 6.

The aquagrams of the monitored nine *Lactobacillus* strains calculated using the spectral data acquired in Phase III and Phase IV (Fig. 2) are presented on Figure 6. In both phases the group of resistant strains showed higher values at shorter wavelengths, i.e. in the range between 1118.5 and 1198.5 nm and, in contrast, the absorbance values at 1217.5 and 1224.5 nm were higher in the case of *Lactobacillus* S06 and the non-resistant strains.

It is interesting to note that although in both resistant and non-resistant groups we had four different strains the spectral patterns of the strains of the same type were similar while those of the different types are more dissimilar.

The pattern of the moderate strain showed more similarity to the pattern of the non-resistant group, but some differences were observed. In phase III the value of *L. bulgaricus* S06 was higher than the values of the non-resistant group at 1118.5 nm thus it had more similarities with the resistant one. However, in phase IV the overall pattern of *L. bulgaricus* S06 was similar to the one of the non-resistant group but showed higher absorption values between 1113.5 and 1198.5 nm.

The wavelengths presented in the aquagram are characteristic of protonated water molecules presented by the asymmetric OH-stretch vibrational frequencies of $[H^+ \cdot (H_2O)_{4,5}]$ - $H_3O^+$ symmetric stretch and asymmetric stretch of 2$^{nd}$ overtone (at 1160 and 1168 nm respectively)[24]. The band around 1217 nm could refer to 2$^{nd}$ overtone of water[25].

The selection of strains possessing high resistance to conditions similar to those in the human

digestive tract has been studied by many authors using different approaches. The main shortcomings of the classical selection methodology were overcome by using different *omics* approaches, such as proteomics, genomics etc. Recently we showed that the aquaphotomics, which is directly related to system functionality, is a powerful tool and can be used successfully for rapid selection of probiotic lactobacilli [11]. Its non-invasive methodology allows rapid and comprehensive evaluation of potential probiotic candidates and might lead to understanding of new phenomena in microbiology.

## Conclusions

With this work we build on the previous studies by analyzing the whole Lactobacillus cultivation time period (divided into five stages) while having the same goal – to obtain high accuracy of bacteria characterization and to find those characteristic wavelengths which can be used for rapid classification of resistant strains and to assess their resistance to low pH and bile.

We found out that the first overtone of the water combination band in the shorter wavelength region (1100-1300 nm) contains useful information about the resistance of strains, presented by differences in the free water molecules, asymmetric O-H stretches and protonated water bands. These differences are due to the presence of different cell metabolites and other chemical substances that change the water molecular structure analyzed in this study and refer to the phenotype of every group of strains.

Highest accuracy of classification and calibration performance of the PLSR models for prediction of strains resistant to low pH in presence of pepsin and to a high concentration of bile were observed in

the IV growth stage (6.3-12.3h of the cultivation time). This is in agreement with our previous publication [11], but also shows that it is possible to build PLSR models with high $R^2$ and low SECV using spectra obtained at an even earlier time during the cultivation of *Lactobacillus*.

**Acknowledgments**

This work was supported by Japan Society for the Promotion of Science (JSPS), Grant Number L14562 (Aleksandar Slavchev) and by the Janos Bolyai Research Scholarship of the Hungarian Academy of Sciences and by the ÚNKP-17-4 New National Excellence Program of The Ministry of Human Capacities (Zoltan Kovacs). The authors would like to thank Selur Pharma Ltd, Bulgaria for providing two of the bacterial strains.

**References**

1. Parvez S, Malik KA, Ah Kang S, et al. Probiotics and their fermented food products are beneficial for health. *J Appl Microbiol* 2006; 100: 1171–85.
2. Kalliomäki MA, Isolauri E. Probiotics and down-regulation of the allergic response. *Immunol Allergy Clin North Am* 2004; 24: 739–52, viii.
3. Isolauri E. Dietary modification of atopic disease: Use of probiotics in the prevention of atopic dermatitis. *Curr Allergy Asthma Rep* 2004; 4: 270–5.
4. McFarland L V. Meta-analysis of probiotics for the prevention of traveler's diarrhea. *Travel Med Infect Dis* 2007; 5: 97–105.

5. Rastall RA, Gibson GR, Gill HS, et al. Modulation of the microbial ecology of the human colon by probiotics, prebiotics and synbiotics to enhance human health: an overview of enabling science and potential applications. *FEMS Microbiol Ecol* 2005; 52: 145–52.

6. Mack DR, Michail S, Wei S, et al. Probiotics inhibit enteropathogenic E. coli adherence in vitro by inducing intestinal mucin gene expression. *Am J Physiol Gastrointest Liver Physiol* 1999; 276: G941-950.

7. Miraglia del Giudice M, De Luca MG. The role of probiotics in the clinical management of food allergy and atopic dermatitis. *J Clin Gastroenterol* 2004; 38: S84-5.

8. Walker WA. Role of nutrients and bacterial colonization in the development of intestinal host defense. *J Pediatr Gastroenterol Nutr* 2000; 30 Suppl 2: S2-7.

9. Pessi T, Sutas Y, Hurme M, et al. Interleukin-10 generation in atopic children following oral Lactobacillus rhamnosus GG. *Clin Exp Allergy* 2000; 30: 1804–1808.

10. Marteau P, Flourie B, Pochart P, et al. Effect of the microbial lactase (EC 3.2.1.23) activity in yoghurt on the intestinal absorption of lactose: An in vivo study in lactase-deficient humans. *Br J Nutr* 2007; 64: 71.

11. Slavchev A, Kovacs Z, Koshiba H, et al. Monitoring of Water Spectral Pattern Reveals Differences in Probiotics Growth When Used for Rapid Bacteria Selection. *PLoS One* 2015; 10: e0130698.

12. Zotta T, Asterinou K, Rossano R, et al. Effect of inactivation of stress response regulators on

the growth and survival of Streptococcus thermophilus Sfi39. *Int J Food Microbiol* 2009; 129: 211–20.

13. Pitino I, Randazzo CL, Mandalari G, et al. Survival of Lactobacillus rhamnosus strains in the upper gastrointestinal tract. *Food Microbiol* 2010; 27: 1121–7.

14. Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem* 1964; 36: 1627–1639.

15. Naes T, Isaksson T, Fearn T, et al. *A user friendly guide to multivariate calibration and classification*. Chichester, UK: NIR Publications, 2002.

16. Cowe IA, McNicol JW. The use of principal components in the analysis of near-infrared spectra. *Appl Spectrosc* 1985; 39: 257–266.

17. Berrueta LA, Alonso-Salces RM, Héberger K. Supervised pattern recognition in food analysis. *J Chromatogr A* 2007; 1158: 196–214.

18. Kinoshita K, Miyazaki M, Morita H, et al. Spectral pattern of urinary water as a biomarker of estrus in the giant panda. *Sci Rep* 2012; 2: 856.

19. De Angelis M, Gobbetti M. Environmental stress responses in Lactobacillus: A review. *Proteomics* 2004; 4: 106–122.

20. Madigan M, Clark D, Stahl D, et al. *Brock Biology of Microorganisms 13th edition*. 2010.

21. Perni S, Andrew PW, Shama G. Estimating the maximum growth rate from microbial growth curves: definition is everything. *Food Microbiol* 2005; 22: 491–495.

22. Mytilinaios I, Salih M, Schofield HK, et al. Growth curve prediction from optical density data. *Int J Food Microbiol* 2012; 154: 169–76.

23. Widdel F. Theory and measurement of bacterial growth. *Di dalam Grundpraktikum Mikrobiol* 2007; 1–11.

24. Headrick JM, Diken EG, Walters RS, et al. Spectral signatures of hydrated proton vibrations in water clusters. *Science (80- )* 2005; 308: 1765–1769.

25. Siesler HW, Ozaki Y, Kawata S, et al. (eds). *Wiley: Near-Infrared Spectroscopy: Principles, Instruments, Applications*. Weinheim, Germany: Wiley-VCH Verlag GmbH, 2001. Epub ahead of print November 2001. DOI: 10.1002/9783527612666.

## Tables

**Table 1.** Bile minimal inhibitory concentration (MIC) of bile and yields of biomass after low pH stress in presence of pepsin of the strains

| Strains | Group | MIC of Bile, mg/ml | Yield of biomass after 3 h stay at pH 1.80 and 9000 U $ml^{-1}$ pepsin (±SD*) |
|---|---|---|---|
| *L. bulgaricus* S10 | Resistant | 2.500 | 0.126±0.014 |
| *L. bulgaricus* S11 | Resistant | 2.500 | 0.100±0.014 |
| *L. gasseri* S20 | Resistant | 2.500 | 0.114±0.011 |
| *L. bulgaricus* S22 | Resistant | 2.500 | 0.117±0.016 |
| *L. bulgaricus* S06 | Moderate | 1.250 | 0.080±0.010 |
| *L. bulgaricus* Y12 | Non-resistant | 0.625 | 0.041±0.006 |
| *L. bulgaricus* S09 | Non-resistant | 0.156 | 0.029±0.005 |
| *L. bulgaricus* S02 | Non-resistant | 0.156 | 0.006±0.001 |
| *L. bulgaricus* S03 | Non-resistant | 0.313 | 0.005±0.002 |

* - Standard deviations calculated based on three parallel samples scanned three consecutive times

**Table 2.** Average recognition and prediction ability of SIMCA and LDA models built for the classification of resistant and non-resistant *Lactobacillus* strains derived from the spectroscopic data in the range between 1100 and 1300 nm

| | SIMCA | | LDA | | |
|---|---|---|---|---|---|
| | model (%) | CV (%) | model (%) | CV (%) | prediction (%) |
| Phase II | 79.77 | 77.06 | 96.79 | 94.87 | 93.16 |
| Phase III | 98.77 | 97.73 | 97.78 | 100 | 100 |
| Phase IV | 97.53 | 96.46 | 98.61 | 100 | 100 |
| Phase V | 81.13 | 78.85 | 98.92 | 98.29 | 99.71 |

CV – "one-strain-out" cross-validation

prediction – independent prediction using data of *L. bulgaricus* S22, *L. bulgaricus* S10 and *L. bulgaricus* Y12.

## Captions to figures

**Figure 1.** PCA Bi-plot calculated when using only reference data (bile MIC and the yield of biomass after 3h stay at pH 1.8 in presence of pepsin (9000 U $ml^{-1}$), reference data in Table 1) of the *Lactobacillus* strains

**Figure 2.** Bacterial growth curve presented by OD at 600 nm and its 4th derivation of modified Gompertz fitting (*L. bulgaricus* S06) showing the existence of five different growth stages

**Figure 3.** Truncated (1100 – 1300 nm) raw spectra (n=2610) and b) MSC transformed spectra of the analyzed *Lactobacillus* strains, obtained in the time interval of 40 – 1200 min

**Figure 4.** PCA, LDA and SIMCA analyses on the spectral data of the *Lactobacillus* strains, using 1100-1300 nm wavelength interval. a) and b) PCA score plot (n=2610), and loading vectors of PCA model calculated on data obtained in the time period of 40-1200 min, c) Scores of the LDA model built for the classification of resistant and non-resistant *Lactobacillus* strains in growth Phase IV ($n_{model}$=540, $n_{prediction}$=270) d) Discrimination power of SIMCA model built for the classification of resistant and non-resistant *Lactobacillus* strains in growth Phase III (n=405).

**Figure 5.** PLSR models on spectral data obtained in different growth phases of the cultivation process, wavelength interval 1100-1300 nm ($n_{PhaseII}$=351, $n_{PhaseIII}$=405, $n_{PhaseIV}$=810, $n_{PhaseV}$=1044). Accuracy of prediction of a) strains resistant to low pH in presence of pepsin and b) strains MIC of bile. c) Regression vector of the PLSR model built to predict strains resistant to low pH in presence of pepsin in growth Phase IV.

**Figure 6.** Aquagrams on the spectra between 1100 and 1300 nm of *Lactobacillus* strains. Averaged values of normalized absorbance values of the water matrix coordinates for every strain are plotted on each axis. Results were calculated on spectral data obtained a) in growth Phase III (n=405) and b) in growth Phase IV (n=810).

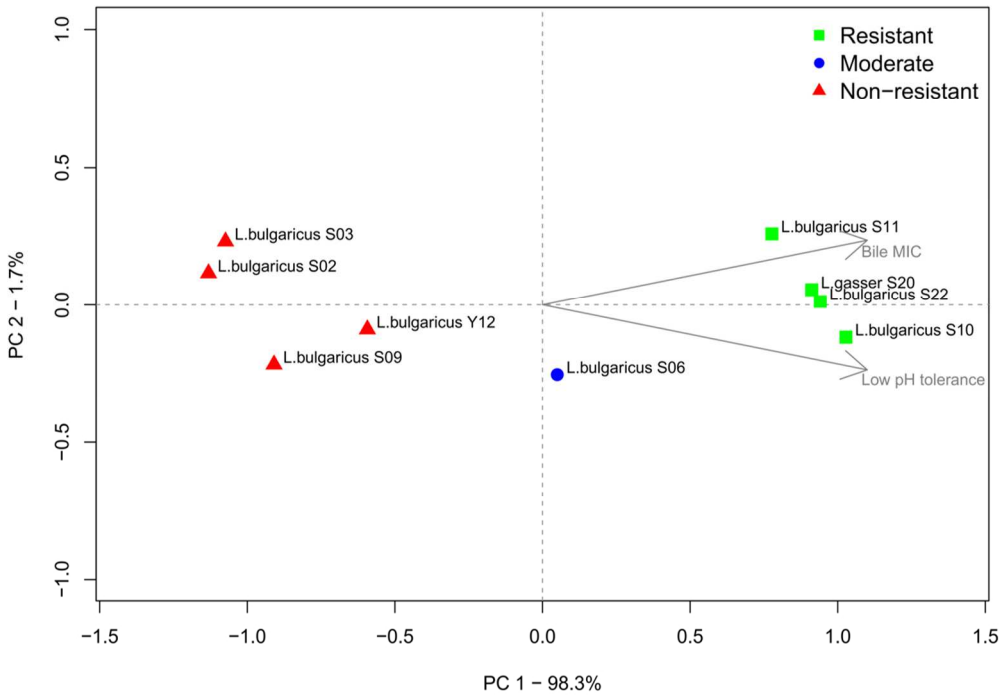


Figure 1. PCA Bi-plot calculated when using only reference data (bile MIC and the yield of biomass after 3h stay at pH 1.8 in presence of pepsin (9000 U ml-1), reference data in Table 1) of the Lactobacillus strains

369x256mm (96 x 96 DPI)

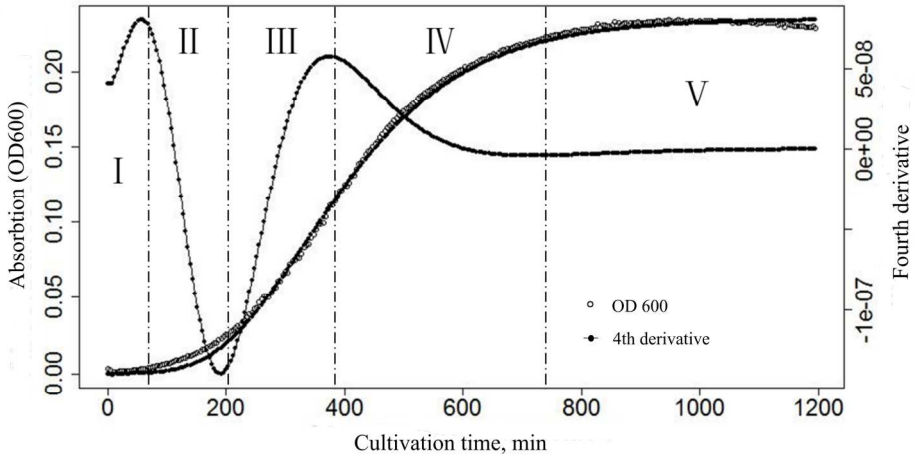


Figure 2. Bacterial growth curve presented by OD at 600 nm and its 4th derivation of modified Gompertz fitting (L. bulgaricus S06) showing the existence of five different growth stages

201x117mm (299 x 299 DPI)

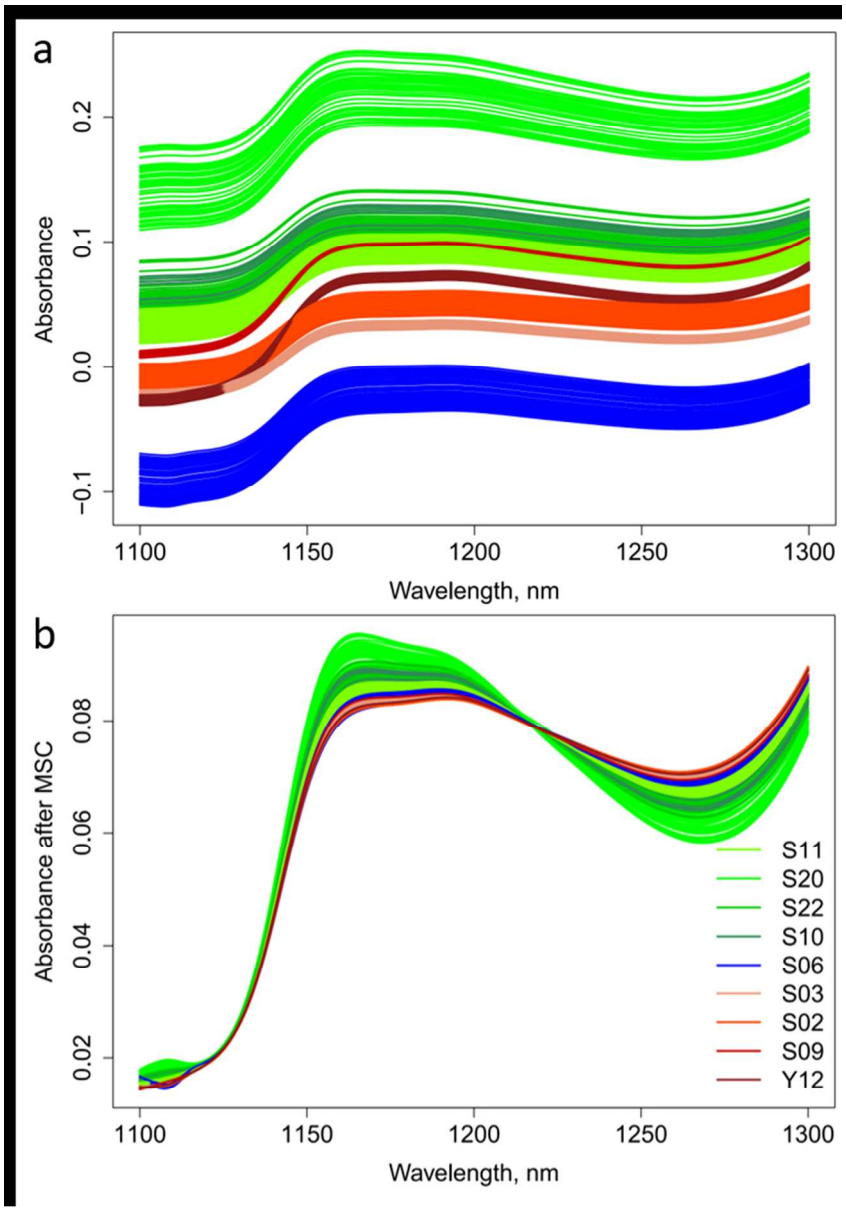


Figure 3. Truncated (1100 – 1300 nm) raw spectra (n=2610) and b) MSC transformed spectra of the analyzed Lactobacillus strains, obtained in the time interval of 40 – 1200 min

199x286mm (96 x 96 DPI)

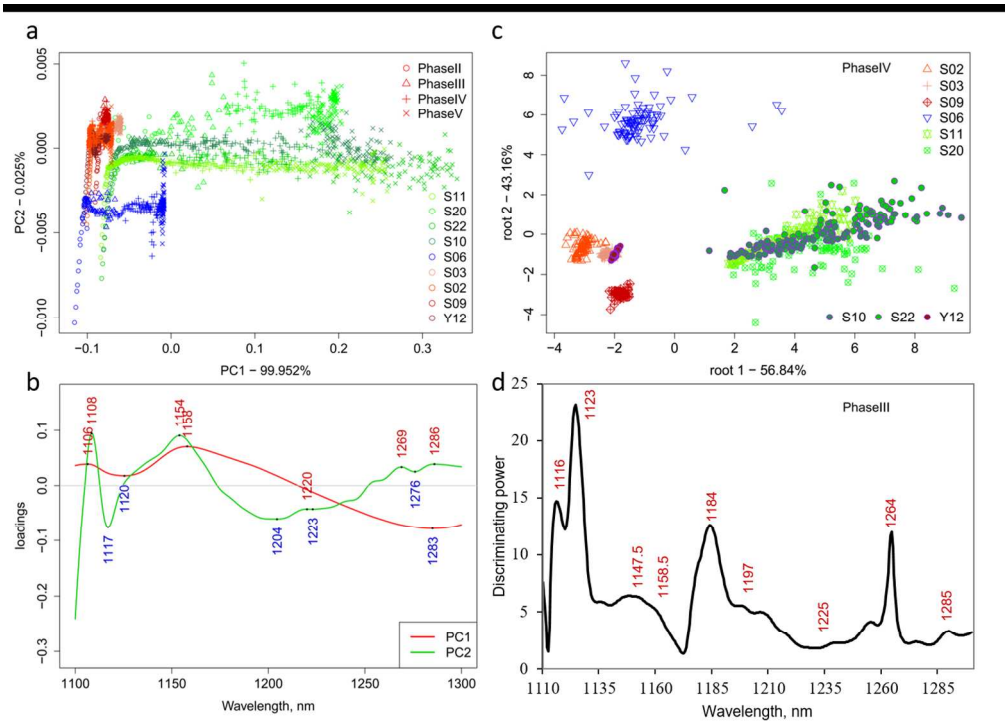


Figure 4. PCA, LDA and SIMCA analyses on the spectral data of the Lactobacillus strains, using 1100-1300 nm wavelength interval. a) and b) PCA score plot (n=2610), and loading vectors of PCA model calculated on data obtained in the time period of 40-1200 min, c) Scores of the LDA model built for the classification of resistant and non-resistant Lactobacillus strains in growth Phase IV (nmodel=540, nprediction=270) d) Discrimination power of SIMCA model built for the classification of resistant and non-resistant Lactobacillus strains in growth Phase III (n=405)

397x291mm (96 x 96 DPI)

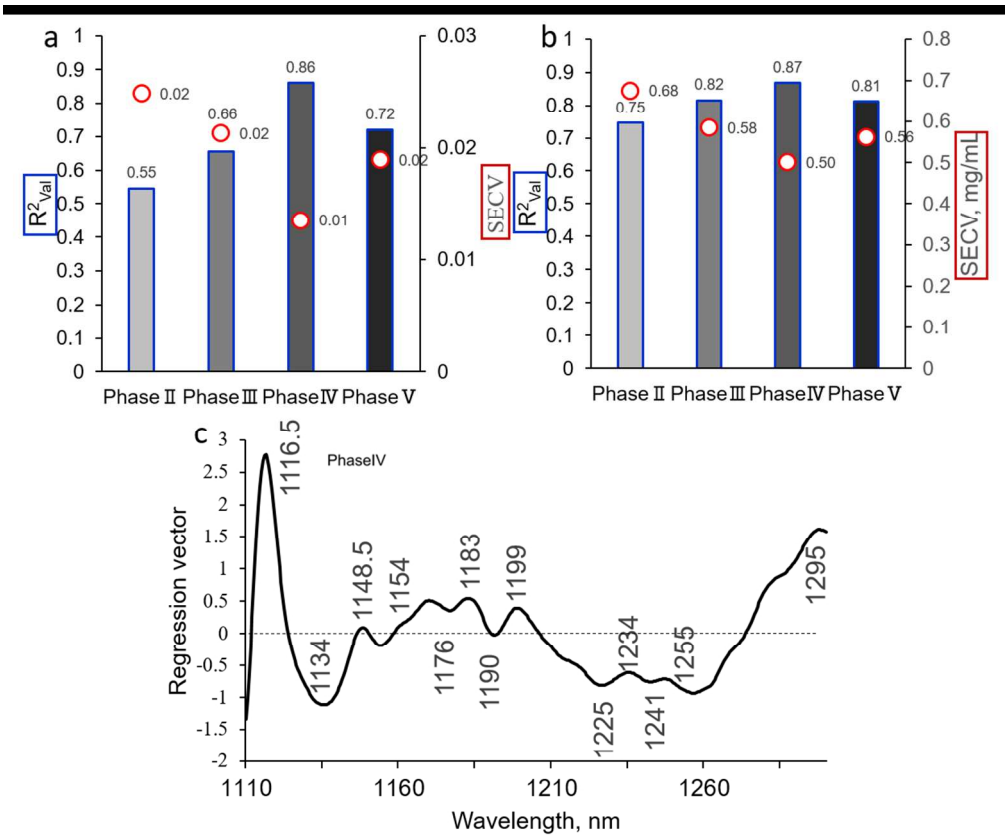


Figure 5. PLSR models on spectral data obtained in different growth phases of the cultivation process, wavelength interval 1100-1300 nm (nPhaseII=351, nPhaseIII=405, nPhaseIV=810, nPhaseV=1044). Accuracy of prediction of a) strains resistant to low pH in presence of pepsin and b) strains MIC of bile. c) Regression vector of the PLSR model built to predict strains resistant to low pH in presence of pepsin in growth Phase IV.

328x279mm (96 x 96 DPI)

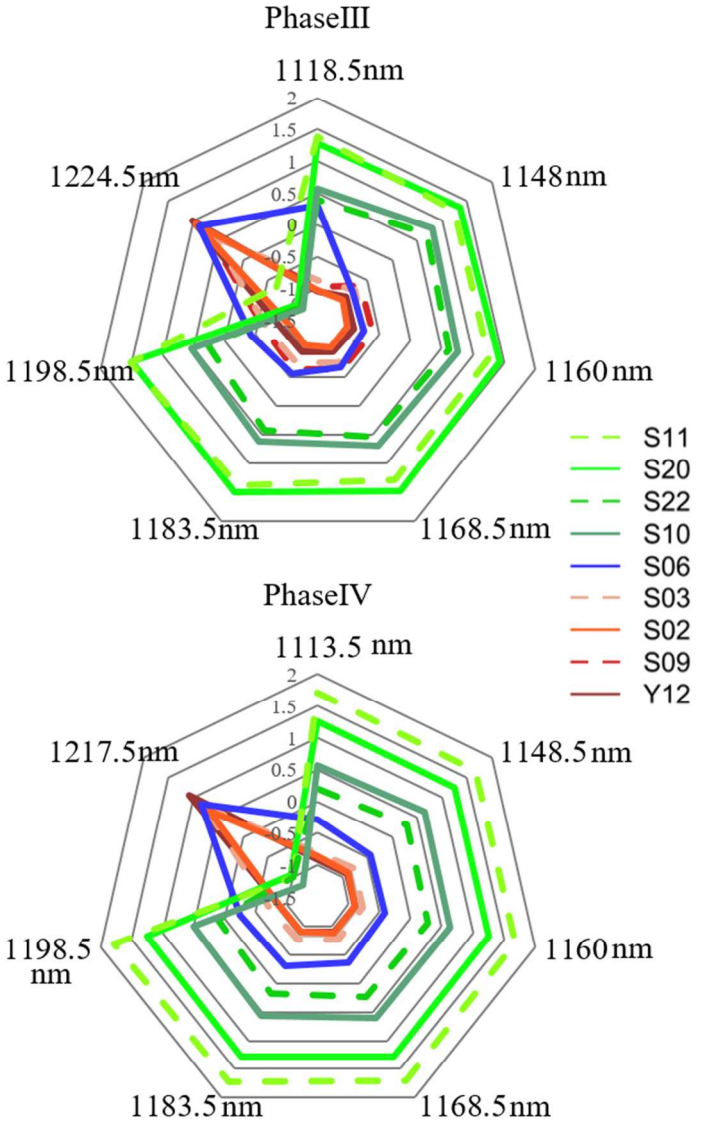


Figure 6. Aquagrams on the spectra between 1100 and 1300 nm of Lactobacillus strains. Averaged values of normalized absorbance values of the water matrix coordinates for every strain are plotted on each axis. Results were calculated on spectral data obtained a) in growth Phase III (n=405) and b) in growth Phase IV (n=810)

198x287mm (96 x 96 DPI)