

## IDENTIFICATION OF MICROORGANISMS BY FOURIER-TRANSFORM INFRARED SPECTROSCOPY

Suntsova AY<sup>1</sup>, Guliev RR<sup>1</sup>, Popov DA<sup>2</sup>, Vostrikova TY<sup>2</sup>, Dubodelov DV<sup>3</sup>, Shchegolikhin AN<sup>1</sup>, Laypanov BK<sup>5</sup>, Pripitnevich TV<sup>3</sup>, Shevelev AB<sup>1,4</sup>✉, Kurochkin IN<sup>1</sup>

<sup>1</sup> Emanuel Institute of Biochemical Physics, Moscow

<sup>2</sup> Bakulev National Medical Research Center of Cardiovascular Surgery, Moscow

<sup>3</sup> Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology, Moscow

<sup>4</sup> Vavilov Institute of General Genetics, Moscow

<sup>5</sup> Skryabin Moscow State Academy of Veterinary Medicine and Biotechnology, Moscow

The need for novel techniques of rapid identification of pathogenic microorganisms arises from the massive spread of drug-resistant nosocomial strains and the emergence of centers for biohazard control. Fourier-transform infrared spectroscopy is a promising alternative to mass spectrometry as it is cost-effective, fast and suitable for field use. The aim of this work was to propose an algorithm for the identification of microorganisms in pure cultures based on the analysis of their Fourier transform infrared spectra. The algorithm is based on the automated principal component analysis of infrared spectra. Unlike its analogues described in the literature, the algorithm is capable of identifying bacteria regardless of the culture medium or growth phase. The training sample included the most prevalent causative agents of infections and sepsis in humans: *Staphylococcus aureus* ( $n = 67$ ), *Enterococcus faecalis* ( $n = 10$ ), *Enterococcus faecium* ( $n = 10$ ), *Klebsiella pneumoniae* ( $n = 10$ ), *Escherichia coli* ( $n = 10$ ), *Serratia marcescens* ( $n = 10$ ), *Enterobacter cloacae* ( $n = 10$ ), *Acinetobacter baumannii* ( $n = 10$ ), *Pseudomonas aeruginosa* ( $n = 10$ ), and *Candida albicans* ( $n = 10$ ). The model we built successfully passed a series of blind tests involving clinical isolates of 10 methicillin-resistant (MRSA) and 10 methicillin-sensitive (MSSA) *Staphylococcus aureus* strains as well as pair mixes of these cultures with clinical isolates of *Pseudomonas aeruginosa*, *Escherichia coli*, and *Klebsiella pneumoniae*.

**Keywords:** spectroscopy, Fourier, IR, principal component analysis, identification, bacteria, yeasts

✉ **Correspondence should be addressed:** Alexei B. Shevelev  
Kosygina 4, Moscow, 119334; shevel\_a@hotmail.com

**Received:** 01.08.2018 **Accepted:** 25.08.2018

**DOI:** 10.24075/brsmu.2018.046

## ИДЕНТИФИКАЦИЯ МИКРООРГАНИЗМОВ С ПОМОЩЬЮ ИНФРАКРАСНЫХ ФУРЬЕ-СПЕКТРОВ

А. Ю. Сунцова<sup>1</sup>, Р. Р. Гулиев<sup>1</sup>, Д. А. Попов<sup>2</sup>, Т. Ю. Вострикова<sup>2</sup>, Д. В. Дубоделов<sup>3</sup>, А. Н. Щеголихин<sup>1</sup>, Б. К. Лайпанов<sup>5</sup>, Т. В. Припутневич<sup>3</sup>, А. Б. Шевелев<sup>1,4</sup>✉, И. Н. Курочкин<sup>1</sup>

<sup>1</sup> Институт биохимической физики имени Н. М. Эмануэля, Москва

<sup>2</sup> Национальный медицинский исследовательский центр сердечно-сосудистой хирургии имени А. Н. Бакулева, Москва

<sup>3</sup> Национальный медицинский исследовательский центр акушерства, гинекологии и перинатологии имени В. И. Кулакова, Москва

<sup>4</sup> Институт общей генетики имени Н. И. Вавилова, Москва

<sup>5</sup> Московская государственная академия ветеринарной медицины и биотехнологии имени К. И. Скрябина, Москва

Актуальность развития быстрых методов идентификации патогенных биологических объектов растет в связи с массовым распространением в лечебных учреждениях микроорганизмов с лекарственной устойчивостью и в связи с развертыванием центров мониторинга биологических угроз. Инфракрасная Фурье-спектроскопия является эффективной альтернативой масс-спектрометрии с точки зрения стоимости и портативности оборудования, экспрессности анализа. Целью работы было описать алгоритм идентификации микроорганизмов в чистых культурах, основанный на анализе колебательных инфракрасных Фурье-спектров культур (Fourier transform infrared, FTIR). Алгоритм основан на автоматизированном анализе спектров методом главных компонент. В отличие от известных в литературе, он позволяет идентифицировать бактерии вне зависимости от стадии роста культуры и состава среды. Обучающая база данных включала наиболее распространенные возбудители гнойно-септических инфекций человека: *Staphylococcus aureus* ( $n = 67$ ), *Enterococcus faecalis* ( $n = 10$ ), *Enterococcus faecium* ( $n = 10$ ), *Klebsiella pneumoniae* ( $n = 10$ ), *Escherichia coli* ( $n = 10$ ), *Serratia marcescens* ( $n = 10$ ), *Enterobacter cloacae* ( $n = 10$ ), *Acinetobacter baumannii* ( $n = 10$ ), *Pseudomonas aeruginosa* ( $n = 10$ ) и *Candida albicans* ( $n = 10$ ). Построенная модель успешно апробирована на серии клинических изолятов *Staphylococcus aureus*: в слепых испытаниях участвовало 10 штаммов с фенотипом лекарственной устойчивости MRSA (methicillin-resistant *Staphylococcus aureus*) и 10 чувствительных штаммов, а также смесь культур этих штаммов с клиническими изолятами *Pseudomonas aeruginosa*, *Escherichia coli* и *Klebsiella pneumoniae*.

**Ключевые слова:** спектроскопия, Фурье, ИК-спектр, метод главных компонент, идентификация, бактерии, дрожжи

✉ **Для корреспонденции:** Шевелев Алексей Борисович  
ул. Косыгина, д. 4, г. Москва, 119334; shevel\_a@hotmail.com

**Статья получена:** 01.08.2018 **Статья принята к печати:** 25.08.2018

**DOI:** 10.24075/vrgmu.2018.046

Identification of microbial species is a routine task for clinical microbiology laboratories. Rapid identification of pathogens in patients with infections or sepsis is essential in prescribing an adequate antibiotic treatment. Efficient therapy for these aggressively progressing conditions is important since they are a common cause of postoperative morbidity and mortality in cardiac surgery [1] and maternal and neonatal death after childbirth [2].

Pathogens can be identified by both traditional microbiological tests and modern techniques now available worldwide, such as matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-ToF MS). The most popular spectrometers are MALDI BioTyper (Bruker; Germany) and Vitek MS (Biomerieux; France). They deliver fast and reliable results but are quite expensive and cannot be afforded by many hospitals. Besides, these devices are too bulky and, therefore, unsuitable for field use by biosafety agencies.

An alternative technique for the identification of microorganisms is Fourier transform infrared (FTIR) spectroscopy. It can determine the chemical composition of a studied culture and identify any type of macromolecule or low molecular weight compound. Similar to MALDI ToF MS, sample preparation for FTIR spectroscopy is easy: the culture sample simply needs to be mounted on a surface transparent to infrared light and left there to dry. Although FTIR spectroscopy ensures a quick diagnosis, its application in clinical microbiology is limited because the FTIR-spectra of a studied culture vary depending on the composition of the growth medium and culture growth phase. The aim of this work was to develop an algorithm for the reliable identification of microorganisms in pure cultures regardless of the growth medium or growth phase based on the analysis of their FTIR spectra.

## METHODS

### Strains of human pathogens

In this work, we used the strains of the most common causative agents of infections and sepsis in humans, including *S. aureus* (20 MRSA and 47 MSSA isolates), *E. faecalis* ( $n = 10$ ), *E. faecium* ( $n = 10$ ), *K. pneumoniae* ( $n = 10$ ), *E. coli* ( $n = 10$ ), *S. marcescens* ( $n = 10$ ), *E. cloacae* ( $n = 10$ ), *A. baumannii* ( $n = 10$ ), *P. aeruginosa* ( $n = 10$ ), *S. epidermidis* ( $n = 10$ ), and *C. albicans* ( $n = 10$ ).

The pathogens were isolated from patients of Bakulev National Medical Research Center of Cardiovascular Surgery and Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology. Isolation and identification were carried out according to the standard technique [3]. Differential diagnostic media included mannitol salt agar selective for staphylococci, Enterococcus agar, Endo agar with fish hydrolysate for culturing gram-negative rods (*Enterobacteriales*, *A. baumannii*, and *P. aeruginosa*), Sabouraud agar and meat-peptone broth supplemented with 1% glucose for *C. albicans*. Confirmatory identification of isolates was performed on the MALDI BioTyper mass spectrometer (Bruker; Germany).

Isolates deposited in the Cryobank were plated onto blood agar plates under aerobic conditions at 37 °C and cultured overnight.

Each isolate was grown in 4 different types of culture media: agarized or liquid, with or without blood. The bloodless media included egg-yolk salt, meat-peptone and Endo broths and Sabouraud agar. For FTIR spectroscopy, the samples were harvested 12, 24 and 48 hours after plating.

To protect the operator of the FTIR spectrometer from the risk of infection and to prepare the cultures for short-term storage, they were inactivated in 70% alcohol before spectroscopy. Thirty microliter aliquots of fresh liquid cultures were collected into polypropylene microtubes, supplemented with 70  $\mu$ l of 96% ethanol and carefully mixed by pipetting. Samples of the cultures grown on solid agar were collected using the inoculation loop and then suspended in 70% aqueous ethanol (30 mg of the biomass per 100  $\mu$ l of the alcohol solution).

### Recording IR spectra of isolates

IR spectra were recorded from the suspensions of pathogen cultures fixed in 70% aqueous ethanol. To prepare individual samples for transmission IR spectroscopy, 5 to 19  $\mu$ l of the suspension were micropipetted onto standard (12.5 mm long and 2 mm thick) ZnSe surfaces (Elektrosteklo; Moscow) and left to dry until complete evaporation of the ethanol (5–15 min). The spectra were recorded by the FTIR spectrometer Spectrum Two (Perkin-Elmer; USA) over the wavenumber range of 4000–600  $\text{cm}^{-1}$  at 4  $\text{cm}^{-1}$  optical resolution and 1  $\text{cm}^{-1}$  digital resolution. The ZnSe surface with an applied pathogen sample was positioned vertically in the Microfocus holder (Perkin-Elmer; USA) and placed in the way of a horizontal probe beam generated by the IR source; 16 individual averaged scans were accumulated for about 2 min. The background spectrum of the spectrometer was recorded under the same conditions but with the clean ZnSe surface and updated before recording the IR spectrum of every new incoming sample.

### Data analysis

After discarding the abnormal spectra, the number of spectra ready for further analysis totalled 347, including 188 spectra obtained from *S. aureus* (39 from these had MRSA phenotype; 48 had MSSA phenotype; and 101 were not characterized in terms of their drug-resistance) and 169 spectra obtained from other pathogens (14 from *A. baumannii*, 32 from *C. albicans*, 8 from *E. cloacae*, 21 from *E. faecalis*, 20 from *E. faecium*, 11 from *E. coli*, 17 from *K. pneumoniae*, 18 from *P. aeruginosa*, 8 from *S. marcescens*, 10 from *S. epidermidis*); 10 spectra represented mixed cultures: 2 were obtained from *S. aureus* (MRSA) + *E. coli*; 2 from *S. aureus* (MSSA) + *E. coli*; 2 from *S. aureus* (MRSA) + *K. pneumoniae*, 2 from *S. aureus* (MSSA) + *K. pneumoniae*, and another 2 from *S. aureus* (MSSA) + *P. aeruginosa*. Fig. 1 illustrates the initial spectra used in the analysis.

Using routine spectroscopy algorithms, the initial spectra were preprocessed for unification; artifacts caused by drifts in the baseline or atmospheric carbon dioxide and water vapor fluctuations were eliminated. Briefly, the initial spectra were normalized to the average transmission, and the first derivative of the envelope was calculated; the relevant wavenumber ranges were narrowed down to 600–1800  $\text{cm}^{-1}$  and 2800–3000  $\text{cm}^{-1}$ . Because the initial spectra were of good quality and did not require any extra smoothing, the derivative was calculated using the symmetric difference formula at two points for numerical differentiation. Fig. 2 shows the preprocessed spectra.

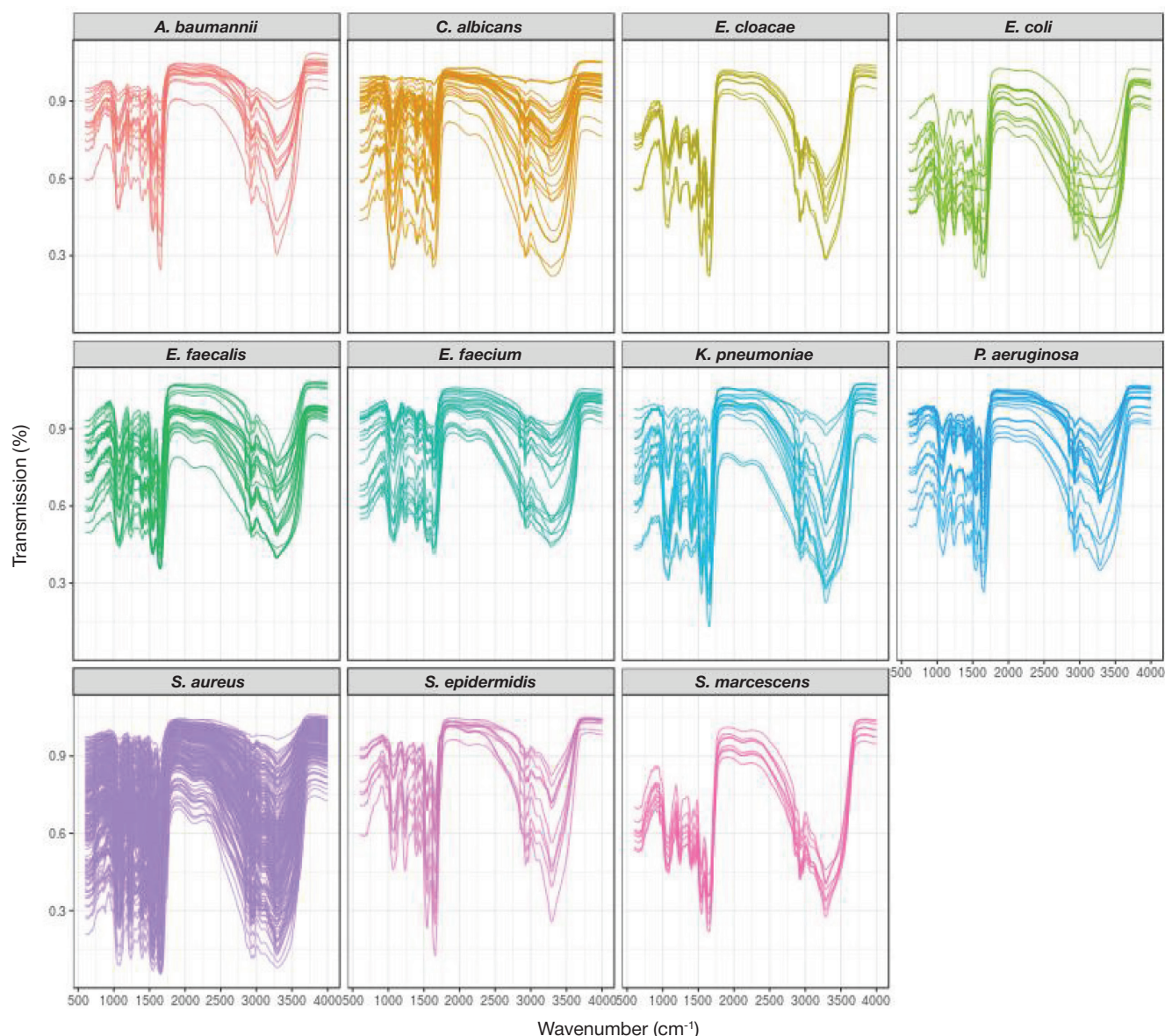
The preprocessed spectra were used to build a mathematical model for the identification of *S. aureus* in a culture sample. Another model capable of discriminating between MRSA and MSSA strains was constructed based on the spectra of MRSA and MSSA phenotypes of *S. aureus*. Both models exploited

the spectra of pure cultures. The spectra of culture mixes were used for model validation.

Both models were built by applying the principal component analysis (PCA) and the linear discriminant analysis (LDA) [4]. LDA [5] allows selecting a line or a hyperplane effectively separating two or more classes of the data. The ratio of between-class to within-class variances shows reliability of classification. However, matrix calculations in LDA do not allow direct application of the spectral data. Datasets must be characterized by a high number of correlated features and regions of poor informative value must be identified. LDA should be preceded by PCA to extract the most informative and uncorrelated spectral data from the dataset. The informative value of the method is assessed by variance: if the latter is low at a given wavenumber, almost all spectra here are expected to behave identically; therefore, such regions cannot provide any valuable information. In PCA, informative and uncorrelated data are extracted by projecting onto corresponding vectors (principal components). In fact, a model constructed with PCA-LDA is a projection of spectral data onto a new vector. In practice, it entails calculation of a linear combination with certain coefficients.

The built models were cross-validated [6]. Cross-validation is a series of blind tests: the initial dataset is randomly partitioned into  $k$  subsets; one of the resulting subsets (the test dataset) is discarded, others  $k - 1$  training sets are analyzed by PCA-LDA. The obtained model predicts the values for the test dataset as if they were initially unknown. This procedure is repeated for each of  $k$  subsets. Once the values predicted for all test subsets are averaged, one can make predictions about new unknown spectra. When splitting the spectra into the subsets, all spectra from the same isolate must fall within one subset only. Otherwise, predicted values will be higher than the actual ones. Cross-validation of our model for *S. aureus* identification was performed at  $k = 20$ , i.e. the total set of spectra was divided into 20 subsets so that the spectra of one and the same isolate always fell into one subset only. For the model discriminating between MSSA and MRSA strains, each strain was represented by an equal number of isolates ( $k = 40$ ). Thus, each subset represented only one isolate.

Preprocessing and the analysis of the obtained spectral data were done in R [7] and the RStudio environment. Spectral data were handled by hyperSpec [8]; the models were built and validated using caret [9] and MASS [10]. Images were created in ggplot2 [11].



**Fig. 1.** The raw Fourier transform infrared spectra of pathogenic bacteria. Left to right and top to bottom are the spectra of *A. baumannii*, *C. albicans*, *E. cloacae*, *E. coli*, *E. faecalis*, *E. faecium*, *K. pneumoniae*, *P. aeruginosa*, *S. aureus*, *S. epidermidis*, and *S. marcescens* expressed in the units of light transmission



## RESULTS

Identification of *S. aureus*

Based on the spectra obtained from 11 species that are the most common causative agents of infections and sepsis in humans (*S. aureus*, *S. epidermidis*, *E. faecalis*, *E. faecium*, *K. pneumoniae*, *E. coli*, *S. marcescens*, *E. cloacae*, *A. baumannii*, *P. aeruginosa*, and *C. albicans*), we built a mathematical model for *S. aureus* identification. The accuracy of the model was assessed by cross-validation on 20 subsets (the spectra of one and the same isolate got into one subset only) and reached 98.4% ( $\pm 4\%$ ). However, further in-depth analysis of cross-validation results revealed that almost all errors arose from the *S. epidermidis* isolate getting into the test sample. This means that at the genus level the model performs well considering the size and composition of the training dataset. The spectra of other staphylococci (two *S. epidermidis* isolates) were too poorly represented in the training dataset to let the model make accurate predictions at the species level. The worst accuracy (81%) observed during cross-validation represented the case when the test dataset included the spectra of *S. epidermidis*

leaving the training set with insufficient data to establish a reliable difference between *S. epidermidis* and *S. aureus*.

From that, one might conclude that these two related species cannot be discriminated using our approach. But the final model that aggregated all obtained data and was built without cross-validation demonstrated an ability to discriminate between these 2 species with 100% accuracy. This ability is visually represented as the projection of the spectral data onto the linear discriminant (the separating axis in the PCA-LDA method; Fig. 3)

This projection is basically a result of multiplication of each spectrum by a coefficient vector: if a preprocessed spectrum is a vector (a set of values), then the linear discriminant is a result of a linear combination. Model tuning is all about the choice of optimal coefficients. Their values for the obtained model are presented as a graph in Fig. 4.

The visual representation of the coefficients serves to roughly interpret the obtained model: the higher is the coefficient expressed in absolute figures, the more significant is the corresponding spectral range. Higher coefficient values expressed in absolute figures (with due account of preprocessing and computation of the derivative) in the zone of

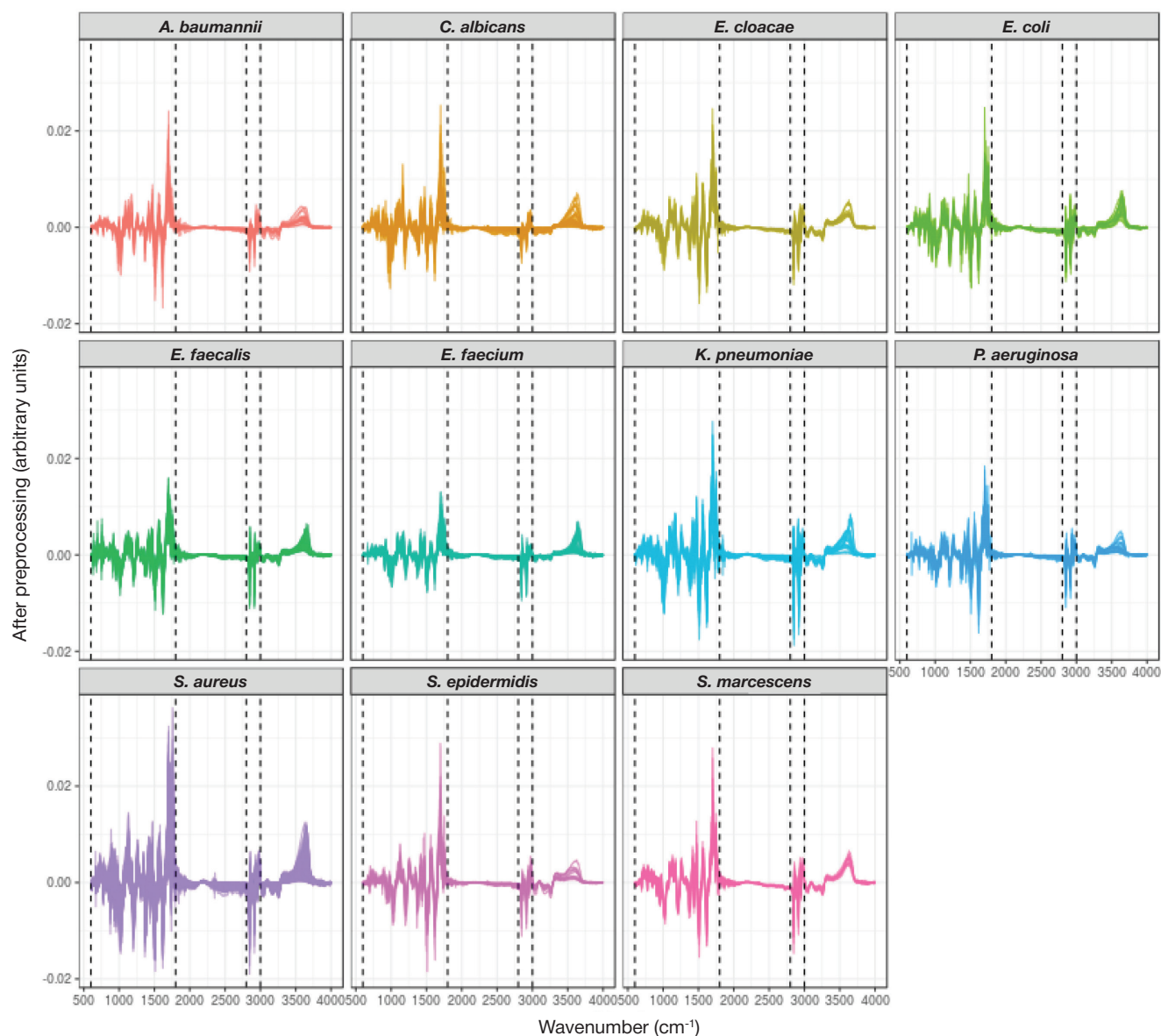
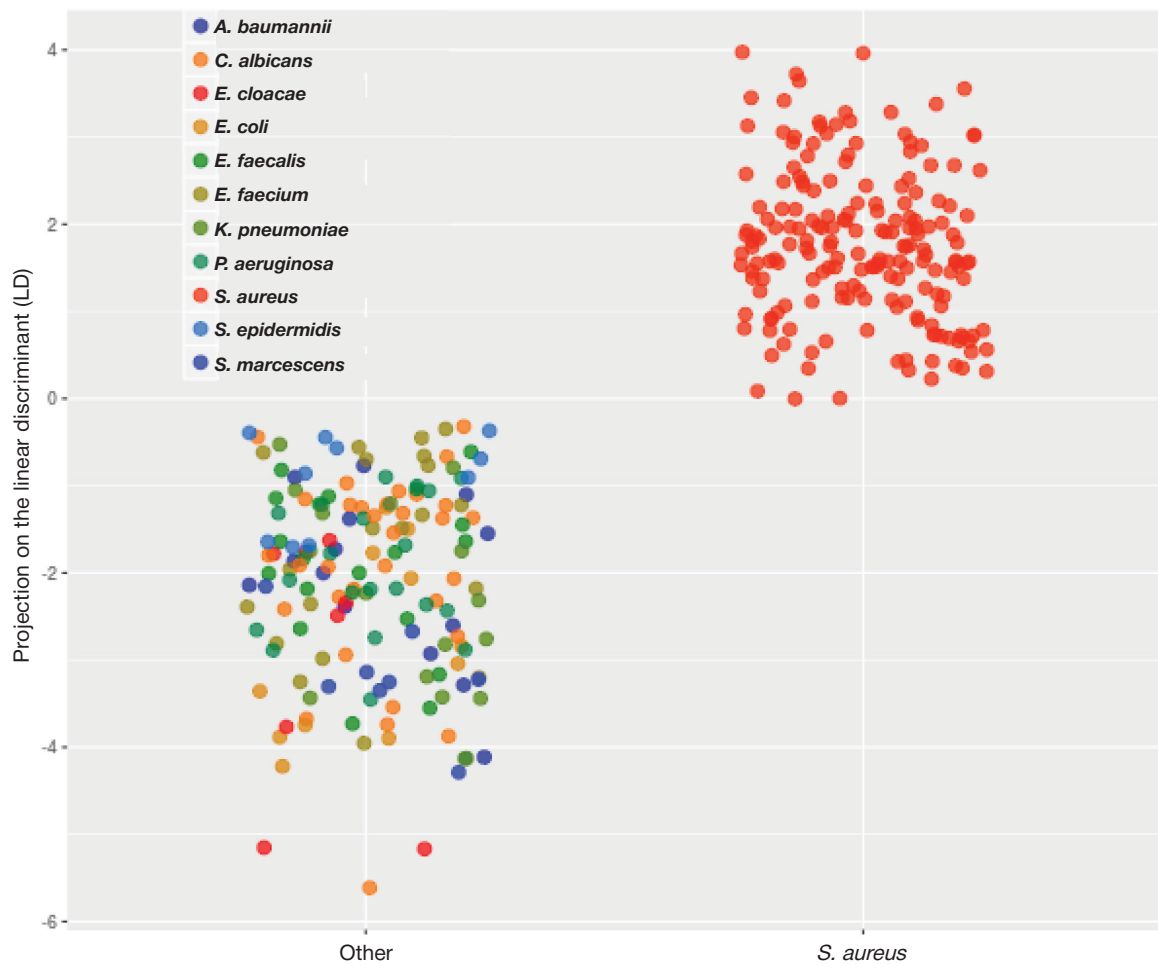
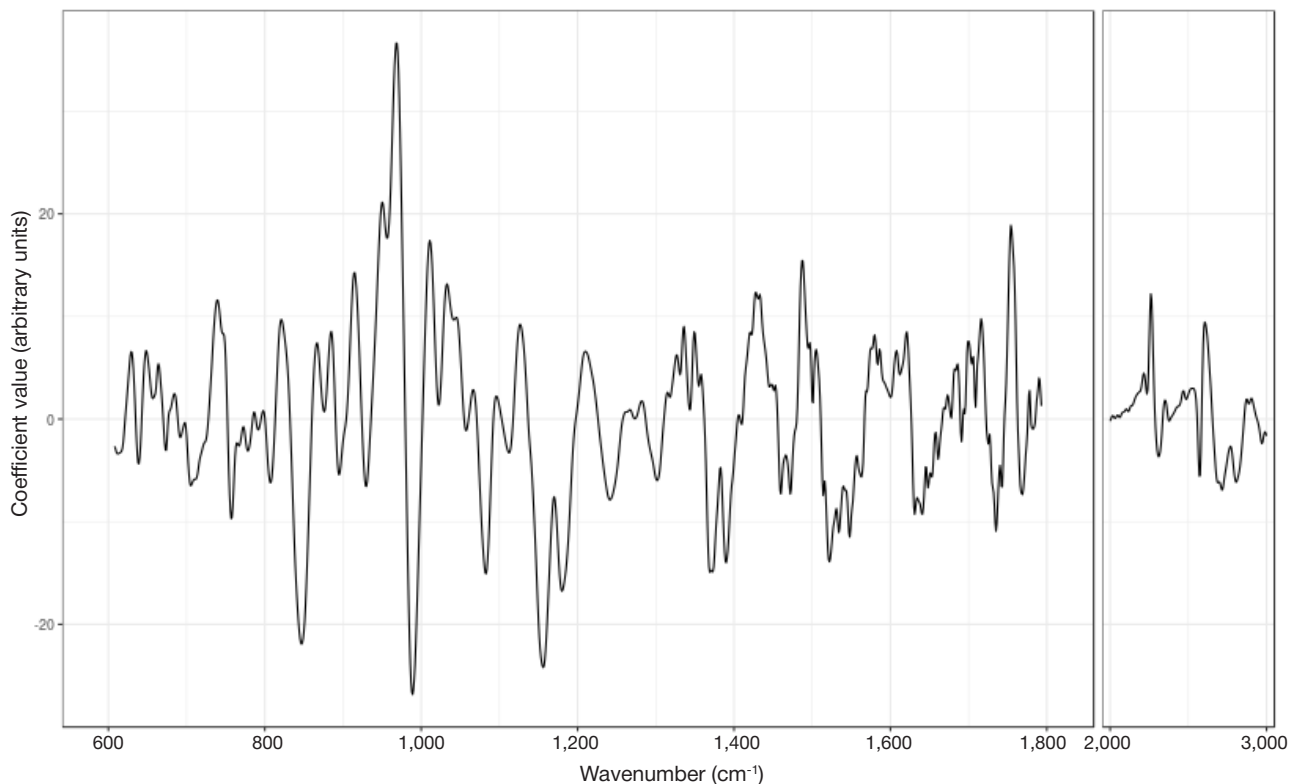


Fig. 2. The Fourier transform infrared spectra after preprocessing: normalization to the average transmission and calculation of the first derivative. Dotted lines show the boundaries of the wavenumber ranges used in the analysis: 600, 1800, 2800, 3000  $\text{cm}^{-1}$



**Fig. 3.** Projection of the initial data onto the linear discriminant (LD) obtained by PCA-LDA. Pathogens are marked by different colors. For the sake of convenience, the spectra of the target pathogen *S. aureus* are separated from others on X-axis. The groups are well-separated given that  $LD > 0$



**Fig. 4.** Linear discriminant coefficients obtained by PCA-LDA. These coefficients were used to obtain the projections in Fig. 3. The coefficients represent the informative value of spectral data. For example, high values of the initial spectrum in the zone of negative coefficients indicate the probability that the studied isolate is not the pathogen of interest

negative coefficients mean that the spectrum is not generated by *S. aureus*, and *vice versa*.

To estimate the feasibility of the proposed approach for clinical microbiology, we studied the ability of our model to identify a target pathogen in mixed cultures. For the analysis, we used 2 methicillin-resistant and 3 methicillin-sensitive *S. aureus* strains. The microorganisms were cultured on blood agar plates for 24 or 48 hours. Then equal volumes of *S. aureus* and gram-negative bacteria (*E. coli*, *K. pneumoniae* or *P. aeruginosa*) cultured on Endo agar were combined. The concentration of bacterial cells per unit volume was not measured. All mixed samples were fixed in alcohol and fed to the final model for identification (see the Table).

In all spectra except 2 representing one and the same sample, the presence of a target pathogen was predicted with high probability, which indicates that the model is reliable and can be used in clinical practice.

### Prediction of methicillin-resistance phenotype in *S. aureus* isolates

Prediction of an methicillin-resistance phenotype in *Staphylococci* is a serious challenge faced by clinical microbiology. In this work we attempted to predict the MRSA/MSSA phenotypes in *S. aureus* isolates based on their FTIR spectra. The classification model was constructed in the same fashion, *i.e.* using PCA and LDA in succession followed by cross-validation.

We failed to achieve the same quality of predictions as with *S. aureus*. The accuracy of the model evaluated by cross-validation was 73%. The projection onto the linear discriminant is shown in Fig. 5. Discrimination here was much worse than for *S. aureus*. Still, 80% of the spectra were identified accurately. This observation leads us to hypothesize a larger size of the training sample could raise the reliability of the identification to the acceptable level.

## DISCUSSION

The first reports of FTIR application for the identification of microorganisms were published in 1991 [12]. The research works that followed were dedicated to the identification of bacteria, such as lactobacilli and agents of foodborne infections, in the environment [13, 14]. A few studies demonstrated that FTIR can be used to identify *Mycobacteria* and *Listeria* [15–17]. In 2011 with the arrival of commercially available spectrometers by Bruker (Germany) and Perkin-Elmer (USA) that reliably identified microorganisms from their FTIR spectra the number of publications on the use of FTIR in microbiology started to grow [18–20]. Research groups were formed outside Germany in Poland [21], the UK [22, 23] and the Netherlands [24]. The Dutch researchers were the first to attempt to identify the causative agents of sepsis in humans and to compare spectral resolutions of different vibrational spectroscopy techniques, including FTIR spectroscopy, Raman spectroscopy, and surface-enhanced Raman spectroscopy (SERS). The authors concluded that FTIR and Raman spectroscopies produced reliable results but were not as sensitive as SERS. In turn, although SERS proved to be a very sensitive technique, its reproducibility was poor.

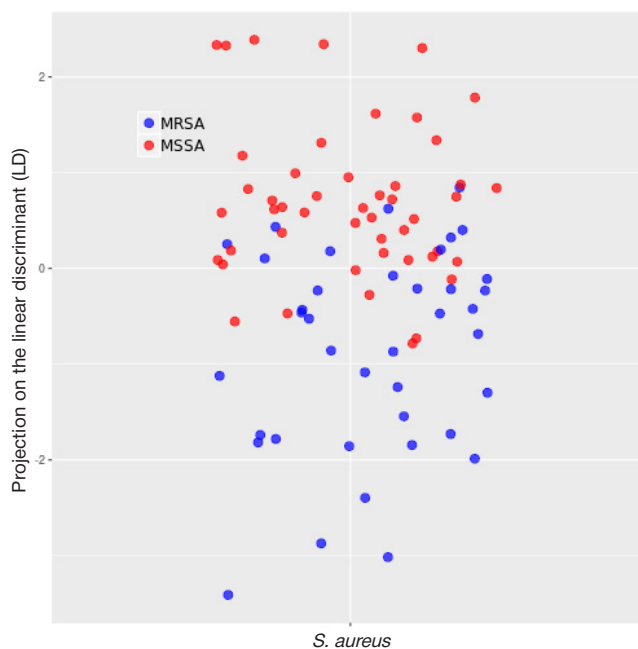
Recently, a lot of research works have been published on the use of FTIR in clinical microbiology [25–28]. The first work listed here compares spectral resolutions of vibrational spectroscopy techniques, including SERS (accuracy of 74.9%), Raman

spectroscopy (accuracy of 97.8%) and FTIR spectroscopy (accuracy of 96.2%), using a number of pathogenic and nonpathogenic bacteria: *P. aeruginosa*, *P. putida*, *E. coli*, *E. faecium*, *Streptomyces lividans*, *B. subtilis*, *B. cereus*, as well as baker's yeast *Saccharomyces cerevisiae*. The last work from the list describes a method for the rapid identification of bacterial microcolonies of 50 to 300  $\mu\text{m}$  in diameter using the state-of-the-art IR-BioTyper spectrometer (Bruker): the colonies are automatically transferred from the agarized culture medium to the  $\text{CaF}_2$  surface; the principal component analysis applied to the obtained spectral data is performed by an artificial neural network (ANN) accessible via the Bruker server.

The findings of those studies suggest that FTIR spectra comprehensively describe the chemical composition of cells, including biopolymers that are building blocks for cell walls and membranes, intracellular DNA, phospholipids, sugars, etc. and therefore ensure a) the reliable discrimination between pathogenic bacterial species; b) the accurate identification of microorganisms at the species level; c) the identification of a phylum the studied isolate belongs to using digital libraries of microbial spectra. Platforms for rapid testing based solely on IR spectroscopy data could provide a quick solution to these

**Table.** Predicted probability of *S. aureus* presence in the sample

Sample ID	Mixt samples	Predicted probability of <i>S. aureus</i> presence in the sample
1	MSSA + <i>K. pneumoniae</i>	96.8%
1	MSSA + <i>K. pneumoniae</i>	95.1%
2	MSSA + <i>E. coli</i>	40.0%
2	MSSA + <i>E. coli</i>	41.5%
3	MRSA + <i>K. pneumoniae</i>	96.4%
3	MRSA + <i>K. pneumoniae</i>	97.0%
4	MRSA + <i>P. aeruginosa</i>	90.6%
4	MRSA + <i>P. aeruginosa</i>	87.7%
5	MRSA + <i>E. coli</i>	82.4%
5	MRSA + <i>E. coli</i>	73.0%



**Fig. 5.** Projection on the linear discriminant separating phenotypes MSSA and MRSA. Projections of MSSA spectra are shown in red; projections of MRSA spectra are shown in blue. Although 100% classification accuracy was not achieved, it was 80% given that  $\text{LD} > 0$

three tasks and assist in optimizing the treatment strategy and adapting it to an individual patient in order to avoid prescription of antibiotics ineffective against the causative pathogen. However, the transition of this approach from the research lab to the clinical setting is obstructed by the absence of an algorithm for automated analysis of microbial FTIR spectra. Such algorithm is expected to identify those components of the spectrum that are determined by the genotype of the strain and not by culturing conditions, such as the growth medium composition, the growth phase, the degree of culture degradation, etc. In all works referred to above the authors sought to standardize culturing conditions, which is quite difficult to achieve in the real clinical setting and is also time consuming.

Such algorithm is proposed in the present work. It allows identification of bacterial species regardless of the growth phase and growth medium composition. We cultured a number of bacterial isolates of *S. aureus*, *E. faecalis*, *E. faecium*, *K. pneumoniae*, *E. coli*, *S. marcescens*, *E. cloacae*, *A. baumannii*, *P. aeruginosa*, *S. epidermidis* and *C. albicans* in different media and for different time periods. Using PCA, we identified the most informative regions of microbial FTIR spectra. The result of the analysis was represented as a system of coefficients that facilitated quick identification of new isolates from their FTIR spectra. The accuracy of the proposed method was assessed by the blind test using pure cultures of *S. aureus* isolates and their paired mixes with *P. aeruginosa*, *E. coli* and *K. pneumoniae*.

The obtained results demonstrate that the proposed algorithm for the analysis of microbial FTIR spectra reliably identifies the presence of *S. aureus* in the culture regardless of the duration of culturing (24 or 48 hours) after being trained

on the sample of 11 pathogens representing different phyla (bacteria and ascomycete yeasts). All samples were inactivated in 70% ethanol before their spectra were recorded. This makes manipulations with virulent pathogens safer and stabilizes the samples until further analysis. The presence of whole blood and admixtures of other microorganisms (gram-negative *E. coli*, *K. pneumoniae* and *P. aeruginosa*) in the sample at concentrations more or less equal to the concentration of *S. aureus* does not affect the ability of the proposed algorithm to identify the pathogen of interest. The model predicts the presence of a methicillin-resistant phenotype (MRSA/MSSA) with 80% accuracy. We hope that our algorithm will be capable of identifying any other pathogen cultured in any media after expanding the training set.

## CONCLUSIONS

We have described a method for creating a database of microbial FTIR spectra and a comparison algorithm suitable for the identification of pathogenic microorganisms that discriminates between the species regardless of the culture growth phase or medium composition. This algorithm can be used in combination with the standard and affordable spectrometer Spectrum Two (Perkin-Elmer; USA). We have tested out algorithm on the clinical isolates of *S. aureus*, which were reliably discriminated from other causative agents of infections, including *E. faecalis*, *E. faecium*, *K. pneumoniae*, *E. coli*, *S. marcescens*, *E. cloacae*, *A. baumannii*, *P. aeruginosa*, and *C. albicans*, taken as pure cultures and pair mixes.

## References

1. Popov DA. Postoperative infectious complications in cardiac surgery. *Annaly hirurgii*. 2013; (5): 15–21.
2. Pripitnevich TV, Ljubasovskaja LA, Dubodelov DV, Melkumjan AR, Igonina EP, Akimkin VG, Degtjarev DN, Suhlih GT. *Jeftektivnaja profilaktika i lechenie ISMP v rodovspomogatel'nyh uchrezhdenijah Rossijskoj Federacii: nereshennye voprosy organizacii i kontrolja*. *Vestnik Roszdravnadzora*. 2017; (4): 34–41.
3. Popov DA, Nadtochej EA. Algorithm of bacteremia diagnostic in cardiosurgical patients in ICU. *Anesteziol Reanimatol*. 2017; 62 (5): 382–7.
4. Shlens J. A tutorial on Principal Components Analysis. April 7, 2014; Version 3.02.
5. Fisher RA. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*. 1936; 7 (2): 179–188.
6. Hastie T, Tibshirani R, Friedman J. 4.3 Linear Discriminant Analysis. *The Elements of Statistical Learning* (2 ed.). New York: Springer; 2009. 763 p.
7. Hastie T, Tibshirani R, Friedman J. 7.10 Cross-Validation. *The Elements of Statistical Learning* (2 ed.). New York: Springer; 2009.
8. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2016. Vienna, Austria. URL <https://www.R-project.org/>
9. Kuhn M. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang and Can Candan. *Caret: Classification and Regression Training*. R package version 6.0-71. 2016. <https://CRAN.R-project.org/package=caret>
10. Venables WN, Ripley BD. *Modern Applied Statistics with S*. New York: Springer; 2002. 495 p.
11. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer; 2009.
12. Naumann D, Helm D, Labischinski H. Microbiological characterizations by FT-IR spectroscopy. *Nature*. 1991; 351 (6321): 81–2.
13. Oust A, Møretør T, Kirschner C, Narvhus JA, Kohler A. Evaluation of the robustness of FT-IR spectra of lactobacilli towards changes in the bacterial growth conditions. *FEMS Microbiol Lett*. 2004 Oct 1; 239 (1): 111–6.
14. Wenning M, Theilmann V, Scherer S. Rapid analysis of two food-borne microbial communities at the species level by Fourier-transform infrared microspectroscopy. *Environ Microbiol*. 2006 May; 8 (5): 848–57.
15. Miguel Gómez MA, Bratos Pérez MA, Martín Gil FJ, Dueñas Díez A, Martín Rodríguez JF, Gutiérrez Rodríguez P, Orduña Domingo A, Rodríguez Torres A. Identification of species of *Brucella* using Fourier transform infrared spectroscopy. *J Microbiol Methods*. 2003 Oct; 55 (1): 121–31.
16. Rebuffo-Scheer CA, Kirschner C, Staemmler M, Naumann D. Rapid species and strain differentiation of non-tuberculous mycobacteria by Fourier-Transform Infrared microspectroscopy. *J Microbiol Methods*. 2007 Feb; 68 (2): 282–90. Epub 2006 Oct 19.
17. Rebuffo-Scheer CA, Dietrich J, Wenning M, Scherer S. Identification of five *Listeria* species based on infrared spectra (FTIR) using macrosamples is superior to a microsample approach. *Anal Bioanal Chem*. 2008 Mar; 390 (6): 1629–35.
18. Schäwe R, Fetzer I, Tönniges A, Härtig C, Geyer W, Harms H, Chatzinotas A. Evaluation of FT-IR spectroscopy as a tool to quantify bacteria in binary mixed cultures. *J Microbiol Methods*. 2011 Aug; 86 (2): 182–7.
19. Wenning M, Scherer S. Identification of microorganisms by FTIR spectroscopy: perspectives and limitations of the method. *Appl Microbiol Biotechnol*. 2013 Aug; 97(16): 7111–20.
20. Wenning M, Breitenwieser F, Konrad R, Huber I, Busch U, Scherer S. Identification and differentiation of food-related bacteria: A comparison of FTIR spectroscopy and MALDI-TOF mass



- spectrometry. *J Microbiol Methods*. 2014 Aug; 103: 44–52.
21. Rygula A, Jekiel K, Szostak-Kot J, Wrobel TP, Baranska M. Application of FT-Raman spectroscopy for in situ detection of microorganisms on the surface of textiles. *J Environ Monit*. 2011 Nov; 13 (11): 2983–7.
  22. Driver T, Bajhaiya AK, Allwood JW, Goodacre R, Pittman JK, Dean AP. Metabolic responses of eukaryotic microalgae to environmental stress limit the ability of FT-IR spectroscopy for species identification. *Algal Res*. 2015 Sep; 11: 148–155.
  23. Alvarez-Ordóñez A, Mouwen DJ, López M, Prieto M. Fourier transform infrared spectroscopy as a tool to characterize molecular composition and stress response in foodborne pathogenic bacteria. *J Microbiol Methods*. 2011 Mar; 84 (3): 369–78.
  24. Maquelin K, Kirschner C, Choo-Smith LP, Ngo-Thi NA, van Vreeswijk T, Stämmmler M, Endtz HP, Bruining HA, Naumann D, Puppels GJ. Prospective study of the performance of vibrational spectroscopies for rapid identification of bacterial and fungal pathogens recovered from blood cultures. *J Clin Microbiol*. 2003 Jan; 41 (1): 324–9.
  25. Muhamadali H, Subaihi A, Mohammadtaheri M, Xu Y, Ellis DI, Ramanathan R, Bansal V, Goodacre R. Rapid, accurate, and comparative differentiation of clinically and industrially relevant microorganisms via multiple vibrational spectroscopic fingerprinting. *Analyst*. 2016 Aug 15; 141 (17): 5127–36.
  26. Lasch P, Stämmmler M, Zhang M, Baranska M, Bosch A, Majzner K. FT-IR Hyperspectral Imaging and Artificial Neural Network Analysis for Identification of Pathogenic Bacteria. *Anal Chem*. 2018 Jul 11.
  27. Bishop CM. *Pattern Recognition and Machine Learning*. New York: Springer; 2006. ISBN-10: 0-387-31073-8.
  28. Beleites C, Sergio V. *HyperSpec: a package to handle hyperspectral data sets in R*, R package version 0.98-20150304. <http://hyperspec.r-forge.r-project.org/>

## Литература

1. Попов Д. А. Послеоперационные инфекционные осложнения в кардиохирургии. *Анналы хирургии*. 2013; (5): 15–21.
2. Припутневич Т. В., Любасовская Л. А., Дубоделов Д. В., Мелкумян А. Р., Игонина Е. П., Акимкин В. Г., Дегтярев Д. Н., Сухих Г. Т. Эффективная профилактика и лечение ИСМП в родовспомогательных учреждениях Российской Федерации: нерешенные вопросы организации и контроля. *Вестник Росздравнадзора*. 2017; (4): 34–41.
3. Попов Д. А., Надточей Е. А. Алгоритм диагностики бактериемии у кардиохирургических больных в орнит. *Анестезиология и реаниматология*. 2017; 62 (5): 382–7.
4. Shlens J. A tutorial on Principal Components Analysis. April 7, 2014; Version 3.02.
5. Fisher RA. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*. 1936; 7 (2): 179–188.
6. Hastie T, Tibshirani R, Friedman J. 4.3 Linear Discriminant Analysis. *The Elements of Statistical Learning* (2 ed.). New York: Springer; 2009. 763 p.
7. Hastie T, Tibshirani R, Friedman J. 7.10 Cross-Validation. *The Elements of Statistical Learning* (2 ed.). New York: Springer; 2009.
8. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, 2016. Vienna, Austria. URL <https://www.R-project.org/>
9. Kuhn M. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang and Can Candan. *Caret: Classification and Regression Training*. R package version 6.0-71. 2016. <https://CRAN.R-project.org/package=caret>
10. Venables WN, Ripley BD. *Modern Applied Statistics with S*. New York: Springer; 2002. 495 p.
11. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer; 2009.
12. Naumann D, Helm D, Labischinski H. Microbiological characterizations by FT-IR spectroscopy. *Nature*. 1991; 351 (6321): 81–2.
13. Oust A, Mørreth T, Kirschner C, Narvhus JA, Kohler A. Evaluation of the robustness of FT-IR spectra of lactobacilli towards changes in the bacterial growth conditions. *FEMS Microbiol Lett*. 2004 Oct 1; 239 (1): 111–6.
14. Wenning M, Theilmann V, Scherer S. Rapid analysis of two foodborne microbial communities at the species level by Fourier-transform infrared microspectroscopy. *Environ Microbiol*. 2006 May; 8 (5): 848–57.
15. Miguel Gómez MA, Bratos Pérez MA, Martín Gil FJ, Dueñas Díez A, Martín Rodríguez JF, Gutiérrez Rodríguez P, Orduña Domingo A, Rodríguez Torres A. Identification of species of *Bruella* using Fourier transform infrared spectroscopy. *J Microbiol Methods*. 2003 Oct; 55 (1): 121–31.
16. Rebuffo-Scheer CA, Kirschner C, Staemmler M, Naumann D. Rapid species and strain differentiation of non-tuberculous mycobacteria by Fourier-Transform Infrared microspectroscopy. *J Microbiol Methods*. 2007 Feb; 68 (2): 282–90. Epub 2006 Oct 19.
17. Rebuffo-Scheer CA, Dietrich J, Wenning M, Scherer S. Identification of five *Listeria* species based on infrared spectra (FTIR) using macrosamples is superior to a microsample approach. *Anal Bioanal Chem*. 2008 Mar; 390 (6): 1629–35.
18. Schäwe R, Fetzer I, Tönniges A, Härtig C, Geyer W, Harms H, Chatzinotas A. Evaluation of FT-IR spectroscopy as a tool to quantify bacteria in binary mixed cultures. *J Microbiol Methods*. 2011 Aug; 86 (2): 182–7.
19. Wenning M, Scherer S. Identification of microorganisms by FTIR spectroscopy: perspectives and limitations of the method. *Appl Microbiol Biotechnol*. 2013 Aug; 97(16): 7111–20.
20. Wenning M, Breitenwieser F, Konrad R, Huber I, Busch U, Scherer S. Identification and differentiation of food-related bacteria: A comparison of FTIR spectroscopy and MALDI-TOF mass spectrometry. *J Microbiol Methods*. 2014 Aug; 103: 44–52.
21. Rygula A, Jekiel K, Szostak-Kot J, Wrobel TP, Baranska M. Application of FT-Raman spectroscopy for in situ detection of microorganisms on the surface of textiles. *J Environ Monit*. 2011 Nov; 13 (11): 2983–7.
22. Driver T, Bajhaiya AK, Allwood JW, Goodacre R, Pittman JK, Dean AP. Metabolic responses of eukaryotic microalgae to environmental stress limit the ability of FT-IR spectroscopy for species identification. *Algal Res*. 2015 Sep; 11: 148–155.
23. Alvarez-Ordóñez A, Mouwen DJ, López M, Prieto M. Fourier transform infrared spectroscopy as a tool to characterize molecular composition and stress response in foodborne pathogenic bacteria. *J Microbiol Methods*. 2011 Mar; 84 (3): 369–78.
24. Maquelin K, Kirschner C, Choo-Smith LP, Ngo-Thi NA, van Vreeswijk T, Stämmmler M, Endtz HP, Bruining HA, Naumann D, Puppels GJ. Prospective study of the performance of vibrational spectroscopies for rapid identification of bacterial and fungal pathogens recovered from blood cultures. *J Clin Microbiol*. 2003 Jan; 41 (1): 324–9.
25. Muhamadali H, Subaihi A, Mohammadtaheri M, Xu Y, Ellis DI, Ramanathan R, Bansal V, Goodacre R. Rapid, accurate, and comparative differentiation of clinically and industrially relevant microorganisms via multiple vibrational spectroscopic fingerprinting. *Analyst*. 2016 Aug 15; 141 (17): 5127–36.
26. Lasch P, Stämmmler M, Zhang M, Baranska M, Bosch A, Majzner K. FT-IR Hyperspectral Imaging and Artificial Neural Network Analysis for Identification of Pathogenic Bacteria. *Anal Chem*. 2018 Jul 11.
27. Bishop CM. *Pattern Recognition and Machine Learning*. New York: Springer; 2006. ISBN-10: 0-387-31073-8.
28. Beleites C, Sergio V. *HyperSpec: a package to handle hyperspectral data sets in R*, R package version 0.98-20150304. <http://hyperspec.r-forge.r-project.org/>