

FTIR Differentiation Based On Genomic DNA for Species Identification of *Shigella* Isolates From Stool Samples

Babak Pakbin

University of Applied Sciences Western Switzerland Valais-Wallis

Leila Zolghadr

Imam Khomeini International University

Shahnaz Rafiei

Imam Khomeini International University

Wolfram Manuel Brück (✉ wolfram.bruck@hevs.ch)

University of Applied Sciences Western Switzerland Valais-Wallis

Thomas B. Brück

Technical University of Munich (TUM)

Research Article

Keywords: FTIR, PCA, HCA, Shigella specie, Species identification, Stool samples

Posted Date: September 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-882528/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Shigellosis is a major public health concern in developing and low-income countries caused by four species of *Shigella* with different intestinal symptoms in patients. There is an apparent need to develop rapid, cost-effective, sensitive and specific methods for differentiation of *Shigella* species to be used in outbreaks and health surveillance systems. We developed a sensitive and specific FTIR based method followed by principal component analysis (PCA) and hierarchical clustering analysis (HCA) assays to differentiate four species of *Shigella* isolates from stool samples. The FTIR based method was validated by isolation and identification of 91 *Shigella* species in clinical isolates using the gold standard methods including culture-based and serologic tests and analyzed them by the developed FTIR assay to evaluate the sensitivity and specificity. In summary, four distinct FTIR spectra associated with four species of *Shigella* were obtained with wide variations in three definite regions including $1800 - 1550 \text{ cm}^{-1}$, $1550 - 1100 \text{ cm}^{-1}$ and $1100 - 800 \text{ cm}^{-1}$, to distinguish these species from each other. We found the FTIR method followed by PCA analysis for differentiation of all species of the *Shigella* isolates from stool samples, with specificity, sensitivity, differentiation error and correct differentiation rate values of 100%, 100%, 0% and 100%, respectively.

Introduction

Shigellosis is a gastrointestinal infectious disease in human caused by invasion of colon, rectum and ileum by different species of *Shigella* strains. This disease is regarded as one of the main health problems around the world especially in children in low-income and developing countries due to the unhygienic conditions of water supplies and improper sanitary systems¹. Regarding the fact that the only host for *Shigella* are humans and three main routes of infection have been reported for *Shigella* species: primarily person-to-person, food and water, and flies with fecal contamination². During the recent decades, up to 164 000 deaths annually are caused by Shigellosis outbreaks worldwide, more than 55 000 of which are among the children up to 5 years of age. Also, more than 60 million out of around 188 million cases of Shigellosis are the cases in children under 5 years old^{1,3}. Shigellosis is caused by four species (Serogroup) of *Shigella* including *S. dysenteriae* (Serigroup A), *S. flexneri* (Serogroup B), *S. boydii* (Serogroup C) and *S. sonnei* (Serogroup D) causing fever, cramp and mild to severe diarrhea depending on the type of the *Shigella* species (Serogroup) in children and adults⁴. Identification and differentiation of four species of *Shigella* are therefore critical in preventing and managing the outbreaks caused by this pathogen^{1,4}.

Culture-based methods, biochemical and serotyping tests are used as gold standards for detection and identification of *Shigella* species in clinical, food and water samples. These methods are highly expensive, time consuming, complex, require skilled personnel and, therefore, are unable to be implemented for investigation of microbial outbreaks⁵. *Shigella* isolates are very sensitive as they are immediately died off while isolated from clinical and food samples because of some environmental conditions such as temperature and pH⁶. However, diagnosis of *Shigella* species are still remains elusive

because of low infectious doses of these pathogens, competition with other commensal bacteria and unsuitable sampling⁷. Molecular techniques in clinical, microbiological and laboratory diagnosis are divided into two main categories: sequencing based such as 16srRNA gene sequencing, next generation sequencing (NGS), etc.) and non-sequencing based methods, such as PCR based methods, matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS), liquid chromatography mass spectrometry (LC-MS), biosensors, Fourier transform infrared (FTIR) spectroscopy, etc⁸. Non-sequencing based molecular methods also fall into two main groups with those associated with gene amplification including all PCR based assays and non-amplification DNA fingerprinting methods namely MALDI-TOF MS, FTIR and Raman spectroscopy techniques⁹. So far, limited assays have been designed and developed to identify and discriminate different species of *Shigella* isolates from each other¹⁰. Some molecular methods including MALDI-TOF MS¹¹, conventional multiplex PCR⁵, LC-MS¹², immunocapture PCR¹³, and NGS¹⁰ techniques have been developed and used for differentiation of four species of *Shigella* isolates from clinical, food and environmental samples. It has also been shown that some common molecular techniques such as MALDI-TOF MS and 16-rRNA gene sequencing were unable to differentiate *Shigella* species from each other as well as from Enteroinvasive *E. coli* strains¹⁴.

FTIR spectroscopy is a relatively cost effective, rapid, convenient and precise analytical technique ,which can be used to reflect the DNA structure and composition¹⁵. FTIR spectroscopy has been used as a powerful tool for species identification and differentiation of eukaryotic and prokaryotic cells based on the genomic DNA characterization¹⁶. Spectral data obtained from FTIR spectroscopy of the DNA templates are so complicated to be analyzed; therefore multivariate statistical and dimension reduction methods such as principal components analysis (PCA), hierarchical clustering analysis (HCA), partial least squares (PLS) and artificial neural networks (ANN) techniques have been used for modeling and interpretation the data¹⁷. Several researchers used and suggested FTIR assay followed by statistical analysis (mostly PCA and HCA assays) as a rapid, simple, relatively cheap, precise, sensitive, specific and convenient method for distinguishing the species of microbial pathogens, isolated from clinical specimens, based on their genomic DNA structural differences¹⁸. So far, this method has not been employed for differentiation of *Shigella* species isolated from clinical samples. Regarding the fact, that Shigellosis and *Shigella* species are now responsible for more than fifty-thousand deaths annually among the children around the world and it is also highly critical and needed to cost-effectively and rapidly differentiate four species of *Shigella* isolates for investigation of Shigellosis outbreaks², the purpose of this study was to design and develop the FTIR assay followed by statistical analysis to identify and differentiate four species of *Shigella* isolates from stool samples.

Results

FTIR spectral data

In this study, we developed an FTIR spectroscopic assay as a DNA barcoding method to differentiate four species of *Shigella* from each other. We implemented this method to identify *Shigella* species isolated

from stool samples to evaluate sensitivity, specificity and efficiency of the assay. Totally, 91 *Shigella* isolates were collected from 1862 stool specimens including 18, 25, 23 and 25, *S. dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei* isolates, respectively; using the culture and biochemical based methods and serologic tests as the gold standard methods to detect and identify the *Shigella* species (Table 1). Presumptive *Shigella* colonies on XLD agar plates were subjected to DNA extraction and species identification using FTIR assay and statistical multivariate or dimension reduction analysis methods. The FTIR spectra of the extracted DNA of the four species of *Shigella* reference strains including *S. dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei* were illustrated in Fig. 1. Four distinctly and significantly different FTIR spectra were observed from DNA of the four species of *Shigella* strains to differentiate these species from each other. Comparison of the FTIR spectra of DNA from different *Shigella* species indicated significant variations in spectral characteristics in three definite regions: 1800 – 1550 cm^{-1} , 1550 – 1100 cm^{-1} and 1100 – 800 cm^{-1} . Also, prominent absorption IR marker bands were observed at 1715, 1689, 1622, 1481, 1436, 1320, 1223, 1175, 1058, 955, 866 and 845 cm^{-1} .

FTIR followed by PCA assay

Dimension reduction and statistical multivariate methods including PCA and HCA assays were used in this study to further differentiate the FTIR spectra of DNA from different species of *Shigella* reference strains and isolates collected from stool samples. Initially, we performed PCA to evaluate the potential of FTIR spectral data for analysis of DNA and differentiate the species of *Shigella*. Prior to analysis of the FTIR spectra by PCA, Bartlett's test of sphericity and KMO test were carried out. The results indicated that the *P* value of Bartlett's test of sphericity was calculated 0.000 (< 0.001) and the KMO was measured 0.94 showing that the FTIR spectral data were significantly suitable for PCA analysis. A 3D score plot was generated by using the first dominant three principal components (PCs) including PC1, PC2 and PC3 which accounted for 40.24, 34.31 and 7.18% of the total variations, respectively (Fig. 2). As it can be seen in Fig. 2, all four species of *Shigella* strains and isolates were identified and distinguished from each other successfully and genomic DNA of different species of *Shigella* isolates were completely discriminated from each other and did not overlap in the plot. 18 out of 18, 25 out of 25, 23 out of 23 and 25 out of 25 species of *S. dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei* isolates, respectively were correctly identified by using the FTIR method followed by PCA assay in this study showing that PCA was able to differentiate and classify the different species of *Shigella* strains.

FTIR followed by HCA assay

The Euclidean distance was used to show the linkage clustering values for hierarchical clustering and calculation of similarity measures among the FTIR spectral data of the genomic DNA of different *Shigella* species. Figure 3 showed the dendrogram form of the HCA results obtained from FTIR spectral data to identify and differentiate four species of *Shigella* isolated from stool samples in this study. We considered 50 and 75% similarity cut-offs; consequently, 4 (A1 - A4) and 7 (H1 – H7) major clades were recognized. In 50% similarity cut-off, *S. sonnei* and *S. boydii* isolates were differentiated into two separate clades (A1 and A2); however, *S. dysenteriae* and *S. flexneri* isolates were grouped in a single group (A4).

On the other hand, in 75% similarity cut-off, *S. dysenteriae* and *S. flexneri* isolates were categorized into two distinct groups (H6 and H7), but four different clades (H1-H4) were recognized for *S. sonnei* and *S. boydii* isolates. In both 50 and 75% similarity cut-offs, one of the *S. boydii* isolates was grouped into a separate single clade (A3 or H5). As a result, whereas the HCA assay was developed to analyze the FTIR spectral data of genomic DNA of *Shigella* species using both 50 and 75% similarity cut-offs in this study, one of the *S. boydii* isolates (the isolate B6, clades A3 or H5) was identified incorrectly. At the present study, 18 out of 18, 25 out of 25, 22 out of 23 and 25 out of 25 species of *S. dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei* isolates respectively were correctly discriminated, respectively; indicating that HCA also has the potential to be used to identify and classify the different species of *Shigella* strains. As shown in Table 2, the results of FTIR spectral data using PCA and HCA assays showed that the identification of four species of *Shigella* isolates from stool samples by FTIR method followed by PCA assay was the best, with specificity, sensitivity, differentiation error and correct differentiation rate values of 100%, 100%, 0% and 100%, respectively for all species of *Shigella* isolates. The specificity, sensitivity, differentiation error and correct differentiation rate values of FTIR analysis with HCA assay were 100%, 95.65%, 1.09% and 98.9%, respectively for identification and differentiation of *S. boydii* isolates.

Discussion

Shigella still remains as a leading cause of mortality and morbidity, and is regarded as the most important communicable of bacterial pathogens causing diarrhea in infants and young children around the world¹⁹. Shigellosis, caused by different species of *Shigella*, occurs usually as occasional outbreaks and sporadic cases in some developed and industrial countries and leads to sporadic and epidemic gastrointestinal diseases in developing and low-income countries²⁰. Four species of *Shigella* cause mild to severe diarrhea in human. Since shigellosis is extremely communicable, it is necessary to develop rapid and cost-effective methods to identify and discriminate different species of *Shigella* from each other in order to control and limit the outbreaks effectively²¹. Notably, classical and conventional methods for identification of pathogens in clinical and environmental samples especially *Shigella* species generally are so time-consuming, expensive and have low specificity and sensitivity²². Hence, in this study, we developed FTIR spectroscopy method followed by PCA and HCA assays which offer sensitivity, specificity, speed and cost-effectivity to differentiate the genomic DNA of *Shigella* species isolated from stool samples and address this problem. However, these methods require specific and expensive equipment and are relatively difficult to perform²³ for differentiation of *Shigella* species.

The spectral data of genomic DNA of *Shigella* species were recorded in the regions between 400 and 4000 cm^{-1} and considered in the range of 1800 – 800 cm^{-1} (a span of 520 wavenumbers) which was previously recommended for characterization of DNA aqueous solutions for analysis by PCA and HCA assays¹⁶. Three distinct regions have been recognized in the FTIR spectra including 1800 – 1500 cm^{-1} , 1500 – 1250 cm^{-1} and 1250 – 800 cm^{-1} . Each section contains specific bands appearing from sugar moiety, sugar phosphate, base sugar, nucleobase, etc²⁴. The FTIR peaks in the region of 1800 – 1500 cm^{-1} originating from vibrations of nucleobase and base pairing interactions were principally assigned to the

stretching vibrations of the C = O, C = N and C = C bonds in the sample¹⁷. The bands at 1715 cm⁻¹ and 1689 cm⁻¹ assigned to stretching vibration of guanines involved in triple helical structures and the stretch of paired guanines, respectively. The strong bands at 1622 cm⁻¹ were essentially due to C = C and C = N ring vibrations of adenine base²⁴. Vibrations localized to sugar-base interactions and giving rise to the marker bands sensitive to backbone conformation, glycosidic bond rotation and sugar puckering modes were observed at the region of 1500 – 1250 cm⁻¹²⁵. The prominent absorption bands at 1481 cm⁻¹, 1436 cm⁻¹ and 1320 cm⁻¹ were due to the adenine ring vibrations, adenine in the Z-form helices and guanine vibration in S-type sugar formation, respectively²⁴. The last region between 800 cm⁻¹ and 1250 cm⁻¹ were due to vibrations along the sugar-phosphate chain and sensitive to the conformation of nucleic acid backbone. B-form double helix of DNA appears at 1223 cm⁻¹. The bands at 1175 cm⁻¹ and 1058 cm⁻¹ also assigned to vibration of sugar phosphate backbone and backbone vibration contributing from the C-O stretch, respectively²⁶. Other prominent absorption bands in the region of 1000 – 800 cm⁻¹, including 955 cm⁻¹, 866 cm⁻¹ and 845 cm⁻¹, were also due to the different nucleic acid N- and S-type of sugar puckering and the sugar phosphate backbone vibrations²⁴.

PCA and HCA assays were used to analyze the FTIR spectral data obtained from the genomic DNA of *Shigella* isolated from stool samples to identify and discriminate the species from each other in this study. KMO measure and the *P* value of sphericity Bartlett's test were observed suitable for PCA analysis²⁷. 3D score plot of PCA model for FTIR spectral data analysis of the DNA of *Shigella* isolates indicated that all species of the isolates can be discriminated from each other correctly by using the PCA analysis at the present study. Dendrogram obtained from HCA analysis of the FTIR spectral data of the DNA of *Shigella* species also showed that FTIR method followed by HCA assay can be used to differentiate all species of *Shigella* from each other. However, regarding the sensitivity and specificity analysis of the assays, we found that that PCA assay was more sensitive, specific and precise for analysis of FTIR spectral data obtained from genomic of *Shigella* isolates from stool samples to discriminate four species of this pathogen from each other in this study. Consequently, it was demonstrated that *S. dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei* isolated from stool samples could be well distinguished by the established FTIR-PCA protocol. FTIR method followed by multivariate statistical or dimension reduction methods have been used to discriminate successfully the genomic DNA of the species of different eucaryotic and procaryotic cells from each other in several researches²⁸. Demir et al. successfully differentiated 12 species of wild wheat from each other using attenuated total reflection FTIR followed by HCA and PCA assays. They reported both analysis methods useful and generally found FTIR method sensitive, low cost and rapid to discriminate the species of eucaryotic cells²⁹. Dinkelacker et al. identified three different species of *Klebsiella* isolates from clinical samples using FTIR, MALDI-TOF and NGS methods. They showed that all of these methods were able to discriminate *Klebsiella* species from each other; however, FTIR showed higher discriminatory power, specificity and sensitivity to recognize three species of *Klebsiella* isolates³⁰. The findings in our and other studies might reflect the fact that the FTIR spectral data shaped by the components of genomic DNA with high diversity among the molecular structures. Han et al. also used FTIR method followed by PCA and PLS assays for species-

specific analysis of the genomic DNA of meat and bone meals. They evaluated these methods to determine the source of 51 meat and bone meal samples. They developed a two-step protocol for distinguishing analysis and found the established method completely (100%) sensitive and specific²⁸. Another study, which was conducted by Potocki et al. indicated that Raman spectra-based and FTIR spectra-based molecular fingerprinting methods can be used effectively with high specificity and sensitivity for identification of different species of the clinical *Candida* isolates³¹. Totally, FTIR spectroscopy is used for analytical chemistry experiments¹⁷. Recently, this method has been used to characterize the biological substances, such as nucleic acids in order to rapidly differentiate and identify different organisms in agricultural and medical sciences³². We have shown for the first time that FTIR-based DNA fingerprinting reflected genomic diversity of 91 *Shigella* isolates collected from stool samples to differentiate the four species of this pathogen correctly. However, using FTIR spectroscopic method for *Shigella* species differentiation may be limited due to the fact that implementing this method requires relatively expensive and specific equipment³⁰. PCR based methods have been proved as the main molecular techniques for identification of *Shigella* species isolated from stool samples; however, presence of the PCR and amplification inhibitors may limit the usefulness and efficacy of PCR method⁵. Regarding the fact that limited studies have investigated and designed practical assays for differentiation of *Shigella* species, biosensor methods have been developed for this purpose^{6,10,17,18}. Even though, it was noted that these assays are so expensive and complicated to be implemented^{7,12,16,27,30}. It is strongly emphasized that there is still a considerable need for developing a rapid, cost-effective, sensitive and specific methods requiring simple equipment to discriminate *Shigella* species in food, water and clinical samples.

Conclusions

In conclusion, a promising FTIR spectroscopic method followed by statistical analysis using PCA and HCA assays was designed and developed to identify and discriminate four species of *Shigella*, isolated from stool samples, based on characterization on genomic DNA at the present study. Totally, 91 *Shigella* strains including 18 *S. dysenteriae*, 25 *S. flexneri*, 23 *S. boydii* and 25 *S. sonnei* were isolated from 1862 stool samples collected from patients with acute diarrhea. FTIR method followed by PCA and HCA assays were used to analyze the DNA extracted from all *Shigella* isolates and reference strains. Four distinct and significantly different FTIR spectra reflecting four species of *Shigella* were obtained with significant variations in three definite regions including 1800 – 1550 cm⁻¹, 1550 – 1100 cm⁻¹ and 1100 – 800 cm⁻¹, to discriminate these species from each other. We found FTIR method followed by PCA assay the best, with specificity, sensitivity, differentiation error and correct differentiation rate values of 100%, 100%, 0% and 100%, respectively for differentiation of all species of the *Shigella* isolates from stool samples. Compared to other molecular techniques, our developed assay is more rapid, relatively cost effective and convenient. Shigellosis is one of the major concerns in public health and it is expected that more rapid, cheap, simple, sensitive and specific methods will be established and developed for

identification and differentiation of *Shigella* species in the future to improve public health surveillance strategies.

Methods

Clinical sample collection and bacterial strains

In this study, A total number of 91 *Shigella* strains including *S. dysenteriae* (n=18), *S. flexneri* (n=25), *S. boydii* (n=23) and *S. sonnei* (n=25), were used. *Shigella* strains were isolated from stool samples collected totally from 1862 patients, female and male aged from 5 to 75 years, with acute diarrhea who admitted to the Milad hospital, Tehran, Iran; between from September 2016 to December 2020. In this study, four species of *Shigella* used as the reference bacterial strains and positive controls including *Sh. dysenteriae* ATCC 13313, *Sh. flexneri* PTCC 1865, *Sh. boydii* ATCC 12030 and *Sh. sonnei* PTCC 1777 were purchased from Pasteur Institute (Pasteur In., Tehran, Iran) in lyophilized form. All lyophilized reference bacterial strains were activated by inoculation in Trypticase Soy Broth (TSB, Merck, Germany) and incubation for 18 h at 37 °C. All strains were subjected to DNA extraction before the FTIR analysis.

Isolation and identification of *Shigella* species from stool samples

Shigella species were isolated from stool samples and differentiated according to the conventional methods as gold standards previously described by Mokhtari et al.³³ and Phiri et al.³⁴ Sterilized disposable inoculation loop of stool samples were directly inoculated on xylose lysine deoxycholate agar (XLD, Merck, Germany) and incubated at 37 °C for 24 h aerobically. Suspected colonies including red ones on XLD agar morphologically resembling *Shigella* were isolated and subjected to biochemical tests. Lysine iron decarboxylase (LIA, Merck, Germany), triple sugar iron (TSI, Merck, Germany), IMViC (Oxoid Ltd., UK), and urease production (Merck, Germany) tests were used to confirm the suspected colonies and identify *Shigella* in stool samples. Genus and species of each presumptive *Shigella* isolate were serologically determined and identified by slide agglutination assay using the commercial *Shigella* genus and species antisera kits (Difco Co., MI, USA), respectively. All *Shigella* isolates were stocked in TSB (Merck, Germany) containing 20% glycerol and kept at -70 °C until further experiments.

DNA extractions

Presumptive *Shigella* isolates on XLD agar from stool samples and the cultured reference *Shigella* strains were subjected to DNA extraction. The genomes of the isolates and strains were extracted using the Sinaclon bacterial DNA extraction commercial kit (Sinaclon Co., Tehran, Iran) according to the manufacturers' instructions. The quantity and quality of the extracted DNA were determined using the NanoDrop 2000 spectrophotometer (ThermoFisher, USA). Also, the final concentrations of all extracted DNA templates were adjusted to 20 ng. μL^{-1} and they were kept at -20 °C until further analysis.

FTIR spectroscopy

Due to the water absorption bands affect significantly the interpretation of FTIR spectra of DNA, the templates (50 μL) were dried and used for FTIR spectroscopy. KBr-FTIR technique previously described by Han et al.²⁸ using an FTIR spectrometer (Magna 550, Madison, USA) over the wavenumber range from 4000 to 400 cm^{-1} was used. The spectra were generated using 64 scans with a resolution of 4 cm^{-1} . Analysis of each DNA template was carried out in triplicates.

Statistical analysis

The FTIR spectral data were analyzed using dimension reduction and multivariate statistical methods by the SPSS version 26.0 software package for Windows (SPSS Inc., Illinois, USA) and the Minitab version 19 software (Minitab Inc., USA). In this study, principal component analysis (PCA) and hierarchical clustering analysis (HCA) methods were used to analyze the FTIR spectral data obtained from the DNA templates. The Bartlett's test of sphericity and Kaiser-Meyer-Olkin test were performed prior to the PCA³⁵. Twenty-five types of the analysis results were selected and analyzed for evaluation of the significant distribution with the infrared intensities and wavenumbers of the *Shigella* isolates spectral data. Differentiation rate of the developed method in this study was evaluated according to specificity, sensitivity, differentiation error and correct differentiation rate of the assay, which were calculated as follows:

$$\text{Specificity} = \text{TN}/(\text{TN}+\text{FP}) \times 100$$

$$\text{Sensitivity} = \text{TP}/(\text{TP}+\text{FN}) \times 100$$

$$\text{Differentiation error} = (\text{FP}+\text{FN})/(\text{TP}+\text{TN}+\text{FP}+\text{FN}) \times 100$$

$$\text{Correct differentiation rate} = 1 - (\text{FP}+\text{FN})/(\text{TP}+\text{TN}+\text{FP}+\text{FN}) \times 100$$

where the true positive (TP) indicates the number of correctly identified species presented in samples, the false positive (FP) indicates the number of false identified species presence in samples, the true negative (TN) indicates the number of correctly identified species not presented in samples and the false negative (FN) indicates the number of false identified species not presented in samples²⁸. All measurements were carried out in triplicates.

Declarations

Ethics approval

The sampling and study protocols were approved by the Ethics Committee of the College of Veterinary Medicine, University of Tehran (IR.UT.REC.1394.108). At the present study, all research was performed in accordance with relevant guidelines/regulations and the Declaration of Helsinki. Also, for all cases, informed consent was obtained from the patients whose stool specimen was included in this study.

Data availability

All raw data included in this study are available from the corresponding author on a reasonable request.

Author contributions

BP, WMB and TBB designed the research; BP, LZ and SR performed all experiments and analysis; BP and WMB wrote the first draft of the manuscript; BP, TBB and WMB finalized and reviewed the manuscript; All authors reviewed the draft and approved the manuscript.

Additional Information

The authors declare that they have no competing interest.

References

1. Kotloff, K. L., Riddle, M. S., Platts-Mills, J. A., Pavlinac, P. & Zaidi, A. K. Shigellosis. *The Lancet* **391**, 801-812 (2018).
2. Pakbin, B. *et al.* Antibiotic susceptibility and genetic relatedness of Shigella species isolated from food and human stool samples in Qazvin, Iran. *BMC research notes* **14**, 1-6 (2021).
3. Bennish, M. L. & Ahmed, S. in *Hunter's Tropical Medicine and Emerging Infectious Diseases* 492-499 (Elsevier, 2020).
4. Puzari, M., Sharma, M. & Chetia, P. Emergence of antibiotic resistant Shigella species: a matter of concern. *Journal of infection and public health* **11**, 451-454 (2018).
5. Kim, H.-J., Ryu, J.-O., Song, J.-Y. & Kim, H.-Y. Multiplex polymerase chain reaction for identification of shigellae and four Shigella species using novel genetic markers screened by comparative genomics. *Foodborne pathogens and disease* **14**, 400-406 (2017).
6. Liu, J. *et al.* Direct detection of Shigella in stool specimens by use of a metagenomic approach. *Journal of clinical microbiology* **56**, e01374-01317 (2018).
7. Elahi, N., Kamali, M., Baghersad, M. H. & Amini, B. A fluorescence Nano-biosensors immobilization on Iron (MNPs) and gold (AuNPs) nanoparticles for detection of Shigella spp. *Materials Science and Engineering: C* **105**, 110113 (2019).
8. Hameed, S., Xie, L. & Ying, Y. Conventional and emerging detection techniques for pathogenic bacteria in food science: A review. *Trends in Food Science & Technology* **81**, 61-73 (2018).
9. Cancino-Padilla, N., Fellenberg, M. A., Franco, W., Ibáñez, R. A. & Pérez, E. V. B. Foodborne bacteria in dairy products: Detection by molecular techniques. *Ciencia e investigación agraria: revista latinoamericana de ciencias de la agricultura* **44**, 215-229 (2017).
10. Smith, A. Using next-generation sequencing to develop a Shigella species threshold and profile faecal samples from suspected diarrhoea cases. *Folia Microbiologica* **66**, 399-410 (2021).
11. Paauw, A. *et al.* Rapid and reliable discrimination between Shigella species and Escherichia coli using MALDI-TOF mass spectrometry. *International Journal of Medical Microbiology* **305**, 446-452 (2015).

12. Everley, R. A., Mott, T. M., Wyatt, S. A., Toney, D. M. & Croley, T. R. Liquid chromatography/mass spectrometry characterization of *Escherichia coli* and *Shigella* species. *Journal of the American Society for Mass Spectrometry* **19**, 1621-1628 (2008).
13. Peng, X., Luo, W., Zhang, J., Wang, S. & Lin, S. Rapid detection of *Shigella* species in environmental sewage by an immunocapture PCR with universal primers. *Applied and Environmental Microbiology* **68**, 2580-2583 (2002).
14. Ragupathi, N. D., Sethuvel, D. M., Inbanathan, F. & Veeraraghavan, B. Accurate differentiation of *Escherichia coli* and *Shigella* serogroups: challenges and strategies. *New microbes and new infections* **21**, 58-62 (2018).
15. Han, Y., Han, L., Yao, Y., Li, Y. & Liu, X. Key factors in FTIR spectroscopic analysis of DNA: the sampling technique, pretreatment temperature and sample concentration. *Analytical Methods* **10**, 2436-2443 (2018).
16. Fredericks, J. D., Bennett, P., Williams, A. & Rogers, K. D. FTIR spectroscopy: a new diagnostic tool to aid DNA analysis from heated bone. *Forensic science international: Genetics* **6**, 375-380 (2012).
17. Rios, T. G. *et al.* FTIR spectroscopy with machine learning: A new approach to animal DNA polymorphism screening. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **261**, 120036 (2021).
18. Pebotuwa, S., Kochan, K., Peleg, A., Wood, B. R. & Heraud, P. Influence of the Sample Preparation Method in Discriminating *Candida* spp. Using ATR-FTIR Spectroscopy. *Molecules* **25**, 1551 (2020).
19. Khalil, I. A. *et al.* Morbidity and mortality due to shigella and enterotoxigenic *Escherichia coli* diarrhoea: the Global Burden of Disease Study 1990–2016. *The Lancet infectious diseases* **18**, 1229-1240 (2018).
20. Kotloff, K. L. *Shigella* infection in children and adults: a formidable foe. *The Lancet Global Health* **5**, e1166-e1167 (2017).
21. Shen, H. *et al.* An outbreak of shigellosis in a Children Welfare Institute caused by a multiple-antibiotic-resistant strain of *Shigella flexneri* 2a. *Journal of infection and public health* **10**, 814-818 (2017).
22. Pakbin, B., Akhondzadeh Basti, A., Khanjari, A., Azimi, L. & Karimi, A. Differentiation of *stx1A* gene for detection of *Escherichia coli* serotype O157: H7 and *Shigella dysenteriae* type 1 in food samples using high resolution melting curve analysis. *Food Science & Nutrition* **8**, 3665-3672 (2020).
23. Zarnowiec, P., Lechowicz, L., Czerwonka, G. & Kaca, W. Fourier transform infrared spectroscopy (FTIR) as a tool for the identification and differentiation of pathogenic bacteria. *Current medicinal chemistry* **22**, 1710-1718 (2015).
24. Banyay, M., Sarkar, M. & Gräslund, A. A library of IR bands of nucleic acids in solution. *Biophysical chemistry* **104**, 477-488 (2003).
25. Mello, M. L. S. & Vidal, B. Changes in the infrared microspectroscopic characteristics of DNA caused by cationic elements, different base richness and single-stranded form. (2012).

26. Doorley, G. W. *et al.* Picosecond Transient Infrared Study of the Ultrafast Deactivation Processes of Electronically Excited B-DNA and Z-DNA Forms of [poly (dG-dC)] 2. *Angewandte Chemie* **121**, 129-133 (2009).
27. Garcia-Santillan, A., Moreno-Garcia, E., Carlos-Castro, J., Zamudio-Abdala, J. H. & Garduno-Trejo, J. Cognitive, affective and behavioral components that explain attitude toward statistics. *Journal of mathematics research* **4**, 8 (2012).
28. Han, Y. *et al.* A novel FTIR discrimination based on genomic DNA for species-specific analysis of meat and bone meal. *Food chemistry* **294**, 526-532 (2019).
29. Demir, P., Onde, S. & Severcan, F. Phylogeny of cultivated and wild wheat species using ATR–FTIR spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **135**, 757-763 (2015).
30. Dinkelacker, A. G. *et al.* Typing and species identification of clinical Klebsiella isolates by Fourier transform infrared spectroscopy and matrix-assisted laser desorption ionization–time of flight mass spectrometry. *Journal of clinical microbiology* **56**, e00843-00818 (2018).
31. Potocki, L. *et al.* FTIR and Raman spectroscopy-based biochemical profiling reflects genomic diversity of clinical Candida isolates that may be useful for diagnosis and targeted therapy of candidiasis. *International journal of molecular sciences* **20**, 988 (2019).
32. Ekruth, J., Gottschalk, C., Ulrich, S., Gareis, M. & Schwaiger, K. Differentiation of *S. chartarum* (Ehrenb.) S. Hughes chemotypes A and S via FT-IR spectroscopy. *Mycopathologia* **185**, 993-1004 (2020).
33. Mokhtari, W., Nsaibia, S., Gharbi, A. & Aouni, M. Real-time PCR using SYBR Green for the detection of *Shigella* spp. in food and stool samples. *Molecular and cellular probes* **27**, 53-59 (2013).
34. Phiri, A. F. *et al.* Burden, Antibiotic Resistance, and Clonality of *Shigella* spp. Implicated in Community-Acquired Acute Diarrhoea in Lilongwe, Malawi. *Tropical medicine and infectious disease* **6**, 63 (2021).
35. Silva, D. L. *et al.* in *Proceedings of the World Congress on Engineering and Computer Science 2014 Vol. I WCECS 2014, San Francisco, CA, October 22.*

Tables

Table 1
Shigella species isolated from stool samples.

Bacterial species	Strain code	Serogroup
<i>S. dysenteriae</i>	MiladH-IDD-2554	A
<i>S. dysenteriae</i>	MiladH-IDD-2556	A
<i>S. dysenteriae</i>	MiladH-IDD-2553	A
<i>S. dysenteriae</i>	MiladH-IDD-2557	A
<i>S. dysenteriae</i>	MiladH-IDD-2513	A
<i>S. dysenteriae</i>	MiladH-IDD-2515	A
<i>S. dysenteriae</i>	MiladH-IDD-2517	A
<i>S. dysenteriae</i>	MiladH-IDD-2551	A
<i>S. dysenteriae</i>	MiladH-IDD-2518	A
<i>S. dysenteriae</i>	MiladH-IDD-2522	A
<i>S. dysenteriae</i>	MiladH-IDD-2529	A
<i>S. dysenteriae</i>	MiladH-IDD-2520	A
<i>S. dysenteriae</i>	MiladH-IDD-2521	A
<i>S. dysenteriae</i>	MiladH-IDD-2526	A
<i>S. dysenteriae</i>	MiladH-IDD-2534	A
<i>S. dysenteriae</i>	MiladH-IDD-2533	A
<i>S. dysenteriae</i>	MiladH-IDD-2539	A
<i>S. dysenteriae</i>	MiladH-IDD-2530	A
<i>S. flexneri</i>	MiladH-IDD-2660	B
<i>S. flexneri</i>	MiladH-IDD-2612	B
<i>S. flexneri</i>	MiladH-IDD-2614	B
<i>S. flexneri</i>	MiladH-IDD-2610	B
<i>S. flexneri</i>	MiladH-IDD-2661	B
<i>S. flexneri</i>	MiladH-IDD-2617	B
<i>S. flexneri</i>	MiladH-IDD-2618	B
<i>S. flexneri</i>	MiladH-IDD-2651	B

Bacterial species	Strain code	Serogroup
<i>S. flexneri</i>	MiladH-IDD-2655	B
<i>S. flexneri</i>	MiladH-IDD-2659	B
<i>S. flexneri</i>	MiladH-IDD-2691	B
<i>S. flexneri</i>	MiladH-IDD-2602	B
<i>S. flexneri</i>	MiladH-IDD-2671	B
<i>S. flexneri</i>	MiladH-IDD-2672	B
<i>S. flexneri</i>	MiladH-IDD-2678	B
<i>S. flexneri</i>	MiladH-IDD-2693	B
<i>S. flexneri</i>	MiladH-IDD-2681	B
<i>S. flexneri</i>	MiladH-IDD-2682	B
<i>S. flexneri</i>	MiladH-IDD-2687	B
<i>S. flexneri</i>	MiladH-IDD-2684	B
<i>S. flexneri</i>	MiladH-IDD-2649	B
<i>S. flexneri</i>	MiladH-IDD-2643	B
<i>S. flexneri</i>	MiladH-IDD-2642	B
<i>S. flexneri</i>	MiladH-IDD-2640	B
<i>S. flexneri</i>	MiladH-IDD-2629	B
<i>S. boydii</i>	MiladH-IDD-2133	C
<i>S. boydii</i>	MiladH-IDD-2112	C
<i>S. boydii</i>	MiladH-IDD-2118	C
<i>S. boydii</i>	MiladH-IDD-2110	C
<i>S. boydii</i>	MiladH-IDD-2114	C
<i>S. boydii</i>	MiladH-IDD-2121	C
<i>S. boydii</i>	MiladH-IDD-2122	C
<i>S. boydii</i>	MiladH-IDD-2129	C
<i>S. boydii</i>	MiladH-IDD-2120	C
<i>S. boydii</i>	MiladH-IDD-2128	C
<i>S. boydii</i>	MiladH-IDD-2141	C

Bacterial species	Strain code	Serogroup
<i>S. boydii</i>	MiladH-IDD-2148	C
<i>S. boydii</i>	MiladH-IDD-2140	C
<i>S. boydii</i>	MiladH-IDD-2143	C
<i>S. boydii</i>	MiladH-IDD-2144	C
<i>S. boydii</i>	MiladH-IDD-2151	C
<i>S. boydii</i>	MiladH-IDD-2156	C
<i>S. boydii</i>	MiladH-IDD-2152	C
<i>S. boydii</i>	MiladH-IDD-2155	C
<i>S. boydii</i>	MiladH-IDD-2164	C
<i>S. boydii</i>	MiladH-IDD-2165	C
<i>S. boydii</i>	MiladH-IDD-2169	C
<i>S. boydii</i>	MiladH-IDD-2171	C
<i>S. sonnei</i>	MiladH-IDD-2812	D
<i>S. sonnei</i>	MiladH-IDD-2810	D
<i>S. sonnei</i>	MiladH-IDD-2811	D
<i>S. sonnei</i>	MiladH-IDD-2818	D
<i>S. sonnei</i>	MiladH-IDD-2817	D
<i>S. sonnei</i>	MiladH-IDD-2823	D
<i>S. sonnei</i>	MiladH-IDD-2821	D
<i>S. sonnei</i>	MiladH-IDD-2822	D
<i>S. sonnei</i>	MiladH-IDD-2828	D
<i>S. sonnei</i>	MiladH-IDD-2827	D
<i>S. sonnei</i>	MiladH-IDD-2825	D
<i>S. sonnei</i>	MiladH-IDD-2834	D
<i>S. sonnei</i>	MiladH-IDD-2835	D
<i>S. sonnei</i>	MiladH-IDD-2837	D
<i>S. sonnei</i>	MiladH-IDD-2839	D
<i>S. sonnei</i>	MiladH-IDD-2841	D

Bacterial species	Strain code	Serogroup
<i>S. sonnei</i>	MiladH-IDD-2845	D
<i>S. sonnei</i>	MiladH-IDD-2844	D
<i>S. sonnei</i>	MiladH-IDD-2848	D
<i>S. sonnei</i>	MiladH-IDD-2849	D
<i>S. sonnei</i>	MiladH-IDD-2851	D
<i>S. sonnei</i>	MiladH-IDD-2856	D
<i>S. sonnei</i>	MiladH-IDD-2858	D
<i>S. sonnei</i>	MiladH-IDD-2861	D
<i>S. sonnei</i>	MiladH-IDD-2864	D

Table 2. Sensitivity and specificity properties of FTIR method to differentiate four species of *Shigella* isolated from stool samples.

	<i>S. dysenteriae</i>		<i>S. flexneri</i>		<i>S. boydii</i>		<i>S. sonnei</i>	
	PCA	HCA	PCA	HCA	PCA	HCA	PCA	HCA
Specificity (%)	100	100	100	100	100	100	100	100
Sensitivity (%)	100	100	100	100	100	95.65	100	100
Differentiation error (%)	0	0	0	0	0	1.09	0	0
Correct differentiation rate (%)	100	100	100	100	100	98.9	100	100

Figures

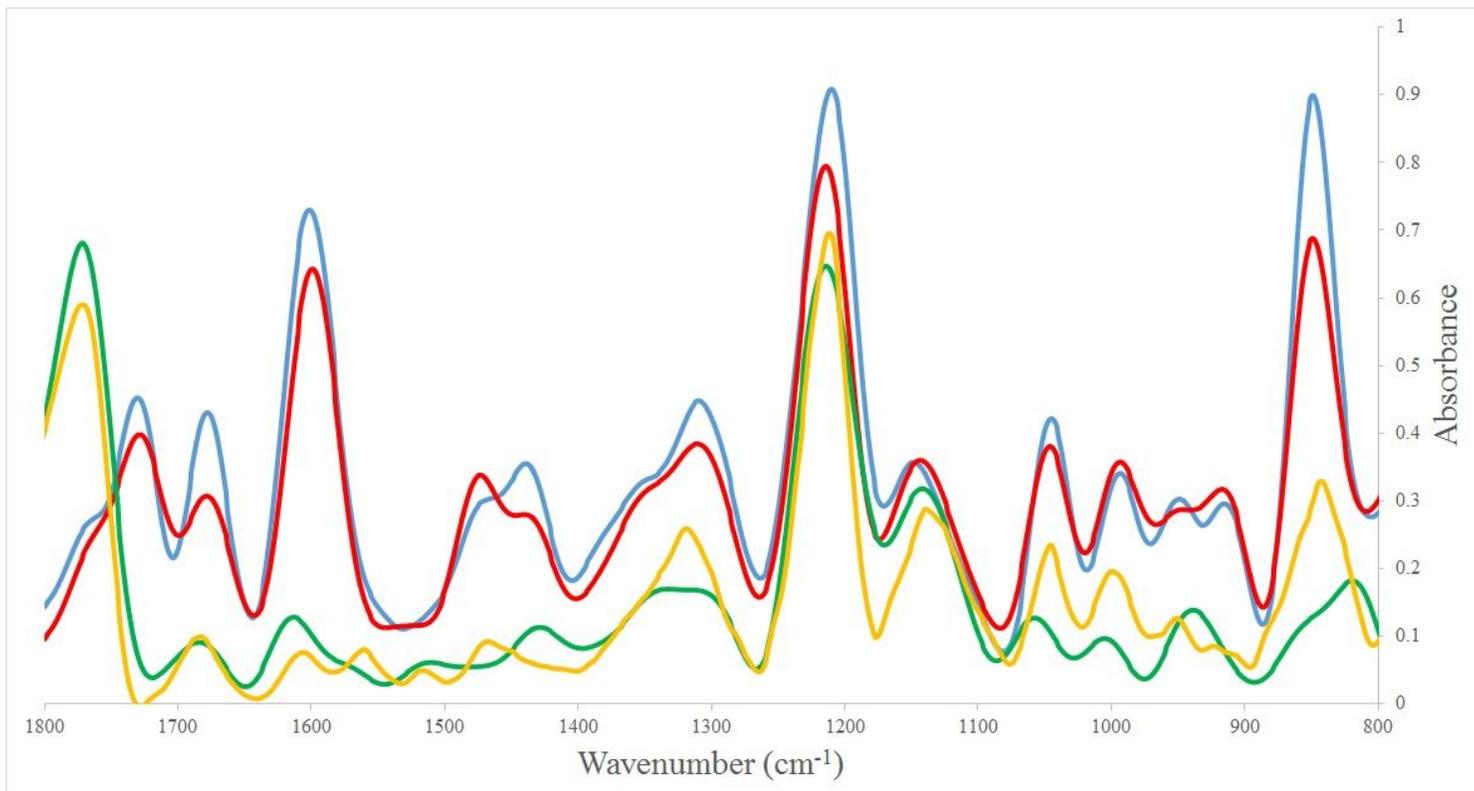


Figure 1

FTIR spectra of *S. dysenteriae* (red line), *S. flexneri* (blue line), *S. boydii* (green line) and *S. sonnei* (yellow line) DNA templates

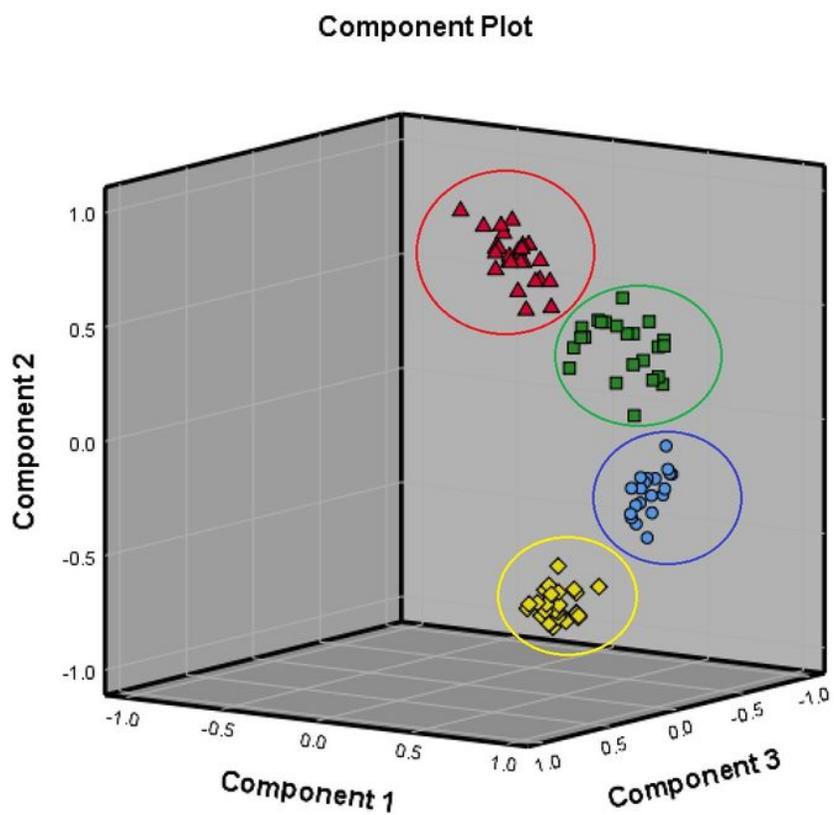


Figure 2

3D PCA score plot of DNA FTIR spectral data of four species of Shigella including *S. dysenteriae* (red triangles), *S. flexneri* (blue circles), *S. boydii* (green squares) and *S. sonnei* (yellow diamonds) isolates from stool samples.

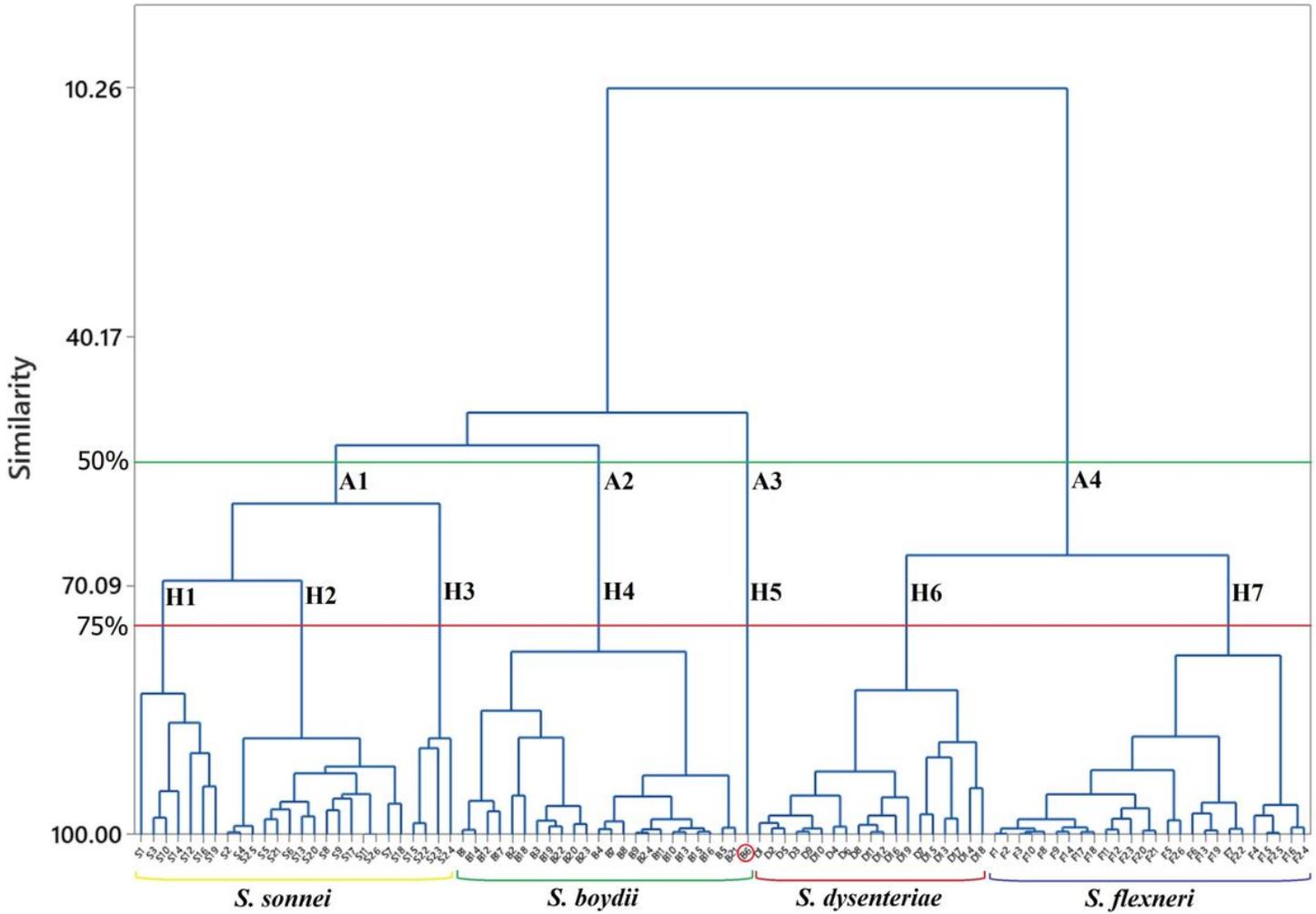


Figure 3

Dendrogram obtained from hierarchical clustering analysis of DNA FTIR spectral data of four species of Shigella including *S. dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei* isolates from stool samples.