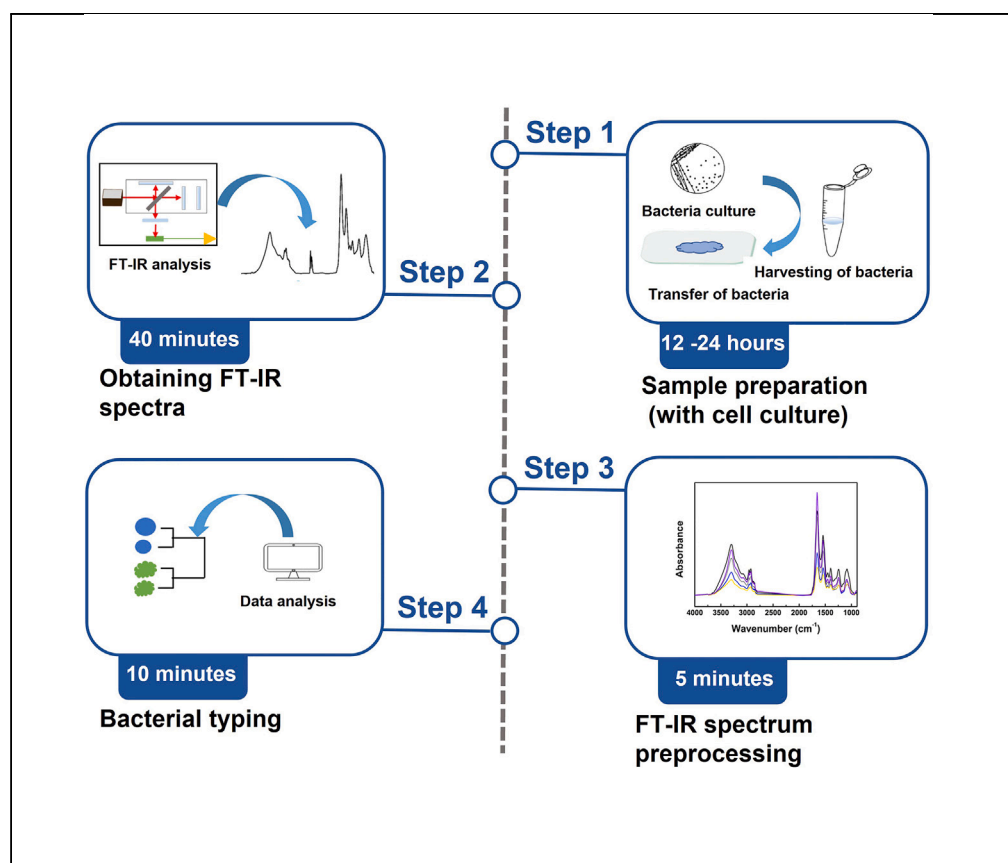


## Protocol

# Protocol for bacterial typing using Fourier transform infrared spectroscopy



Huayan Yang,  
Haimei Shi, Bin  
Feng, ..., Miguel  
Prieto, Fan Yang,  
Shaoning Yu

miguel.prieto@unileon.es  
(M.P.)  
yangfanmt@189.cn (F.Y.)  
yushaoning@nbn.edu.cn  
(S.Y.)

### Highlights

FT-IR-spectroscopy-  
based protocol for  
bacterial typing

Spectral  
manipulations to  
improve spectral  
features

Multivariate analysis  
including machine  
learning scripts to  
analyze microbial IR  
data

The Fourier transform infrared (FT-IR) signals obtained from bacterial samples are specific and reproducible, making FT-IR an efficient tool for bacterial typing at a subspecies level. However, the typing accuracy could be affected by many factors, including sample preparation and spectral acquisition. Here, we present a unified protocol for bacterial typing based on FT-IR spectroscopy. We describe sample preparation from bacterial culture and FT-IR spectrum collection. We then detail FT-IR spectrum preprocessing and multivariate analysis of spectral data for bacterial typing.

**Publisher's note:** Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Yang et al., STAR Protocols 4,  
102223  
June 16, 2023 © 2023 The  
Author(s).  
[https://doi.org/10.1016/  
j.xpro.2023.102223](https://doi.org/10.1016/j.xpro.2023.102223)



## Protocol

## Protocol for bacterial typing using Fourier transform infrared spectroscopy

Huayan Yang,<sup>1,2,6</sup> Haimei Shi,<sup>1,2,6</sup> Bin Feng,<sup>1,2</sup> Li Wang,<sup>3</sup> Liangqiang Chen,<sup>3</sup> Avelino Alvarez-Ordóñez,<sup>4</sup> Li Zhang,<sup>5</sup> Hao Shen,<sup>1,2</sup> Jianhua Zhu,<sup>1</sup> Shouning Yang,<sup>2</sup> Chuanfan Ding,<sup>2</sup> Miguel Prietod,<sup>4,\*</sup> Fan Yang,<sup>3,\*</sup> and Shaoning Yu<sup>1,2,7,8,\*</sup>

<sup>1</sup>Department of Intensive Care Unit, The First Affiliated Hospital of Ningbo University, Ningbo, Zhejiang 315211, China

<sup>2</sup>Institute of Mass Spectrometry, School of Material Science and Chemical Engineering, Ningbo University, Ningbo, Zhejiang 315211, China

<sup>3</sup>Kweichow Moutai Group, Renhuai, Guizhou 564501, China

<sup>4</sup>Institute of Food Science and Technology, University of León, 24071 León, Spain

<sup>5</sup>Department of Laboratory Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing 100730, China

<sup>6</sup>These authors contributed equally

<sup>7</sup>Technical contact: [yushaoning@nbu.edu.cn](mailto:yushaoning@nbu.edu.cn)

<sup>8</sup>Lead contact: [yushaoning@nbu.edu.cn](mailto:yushaoning@nbu.edu.cn)

\*Correspondence: [miguel.prietod@unileon.es](mailto:miguel.prietod@unileon.es) (M.P.), [yangfanmt@189.cn](mailto:yangfanmt@189.cn) (F.Y.), [yushaoning@nbu.edu.cn](mailto:yushaoning@nbu.edu.cn) (S.Y.)  
<https://doi.org/10.1016/j.xpro.2023.102223>

## SUMMARY

The Fourier transform infrared (FT-IR) signals obtained from bacterial samples are specific and reproducible, making FT-IR an efficient tool for bacterial typing at a sub-species level. However, the typing accuracy could be affected by many factors, including sample preparation and spectral acquisition. Here, we present a unified protocol for bacterial typing based on FT-IR spectroscopy. We describe sample preparation from bacterial culture and FT-IR spectrum collection. We then detail FT-IR spectrum preprocessing and multivariate analysis of spectral data for bacterial typing.

## BEFORE YOU BEGIN

Efficient, accurate, and reproducible bacterial typing is highly demand in hospital hygiene management for infection control, epidemiological studies, tracking of bacterial cross-transmission, and monitoring of medical treatment processes.<sup>1–4</sup> Bacterial cells consist of proteins, fatty acids, carbohydrates, nucleic acids, and lipopolysaccharides, and these biochemical signatures can be extracted by FT-IR spectra.<sup>5,6</sup> There is sufficient evidence that the cell surface or some of its components strongly affect the IR spectrum, making bacterial FT-IR spectra presents a specific pattern at species, strain, and/or substrain levels.<sup>7–9</sup> The specific fingerprint-like patterns of bacterial IR signals make FT-IR a non-destructive, low-cost, and high-speed alternative for bacterial typing.<sup>9–14</sup> The IR spectrum reflects the chemical structure of the molecules in the sample. IR absorption bands in the MIR spectrum (4,000–400 cm<sup>−1</sup>) are mainly due to the vibrational modes and can be tentatively assigned to particular functional groups (see Table S1). The frequencies and absorbance intensities of the IR bands are composed of unique broad and complex contours instead of isolated peaks, representing the superposition of all vibrational modes of the biomolecules.<sup>15,16</sup> Naumann et al. recommended dividing the bacterial FT-IR spectrum into five sub-ranges for bacterial identification and typing<sup>17</sup>: (i) the membrane amphiphile region (3000–2800 cm<sup>−1</sup>), (ii) the protein and peptide region (1700–1200 cm<sup>−1</sup>, amide I at ~1650 cm<sup>−1</sup>, amide II at ~1550 cm<sup>−1</sup>, and amide III at ~1260 cm<sup>−1</sup>), (iii) the protein and fatty acid region (“mixed region”; 1500–1200 cm<sup>−1</sup>), (iv) the nucleic acid and polysaccharide region (1200–900 cm<sup>−1</sup>) and (v) the fingerprint region (900–700 cm<sup>−1</sup>). Figure S1 shows a typical bacterial FT-IR spectrum. Different sub-ranges are shown in different colors.



The 1200–900  $\text{cm}^{-1}$  region includes carbohydrates (C-O-O-C symmetric stretching at  $\sim 1155 \text{ cm}^{-1}$ ), phosphate-containing molecules (incl. nucleic acids, wall teichoic acids and lipoteichoic acids; symmetric phosphate stretching at  $\sim 1080 \text{ cm}^{-1}$ ), glycogen (C-O stretching at  $\sim 1030 \text{ cm}^{-1}$ ), and protein phosphorylation ( $\sim 970 \text{ cm}^{-1}$ ). Between 1143–1100  $\text{cm}^{-1}$ , water has a high absorption spectral band, which coincides with the  $\nu_s \text{PO}_2^-$  and  $\nu \text{CO}^-$  of ribonucleic acid RNA (with a small contribution from polysaccharides).<sup>18,19</sup> The degree of hydration of the microbial films exerts an influence in the fingerprint region of the IR spectra and should be controlled. The relative amplitudes in the 1200–900  $\text{cm}^{-1}$  region give these multiple peaks subtle, but reproducible, band shapes for differentiating strains. The typing results based on this region are much better than with any other region or combination of regions.<sup>18</sup> Expanded regions were also suggested and used by another author.<sup>14</sup> For example, the Bruker IR biotyper uses the 1300–800  $\text{cm}^{-1}$  region for bacterial typing.<sup>20</sup>

The whole bacterial FT-IR spectrum is usually complex and the peaks overlap due to contributions from all of the biomolecules present in a bacterial cell.<sup>6,12,17</sup> Various spectral manipulations can be performed to improve the spectral features for spectral interpretation and analysis, as bacterial FT-IR spectra are visually very similar with subtle biochemical differences.<sup>21–23</sup> Before spectra can be studied, they should be processed in order to minimize methodological variability due to sample preparation and spectra acquisition, while simultaneously amplifying the chemically based spectral differences. Spectral normalization, derivative transformation and spectrum window selection of bacterial FT-IR spectra are fundamental steps in spectral preprocessing and can increase the accuracy and reproducibility of bacterial typing.<sup>22</sup>

Previous protocols have provided practical guidance on how to analyze the acquired IR data for the analysis of biological material,<sup>24</sup> such as using IR spectroscopy to distinguish cell types.<sup>25</sup> A tutorial for multivariate classification analysis of vibrational spectroscopy data has also been published.<sup>23</sup> However, the protocol on bacterial typing by FT-IR with specific steps and analytical methods is still limited. Sample preparation, culture conditions, growth period, and spectrum preprocessing can influence the spectral reproducibility and reliability. Moreover, the accurate use of mathematical analysis is essential. Therefore, careful experimental design with unified instrumental settings, sample preparation, spectral acquisition, and data analysis are required to improve the spectral quality and results of bacterial typing. The aim of this protocol is to describe the detailed procedures for bacterial typing based on FT-IR spectroscopy. The experimental steps described in the protocol can be implemented easily, allowing application in microbiology laboratories equipped with a FT-IR spectrometer.

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Bacterial and virus strains</b>		
<i>E. coli</i> (clinically isolated strains)	Peking Union Medical College Hospital	N/A
<i>Pichia kudriavzevii</i>	Kweichow Moutai Group	N/A
<i>Schizosaccharomyces pombe</i>	Kweichow Moutai Group	N/A
<i>Saccharomyces cerevisiae</i>	Kweichow Moutai Group	N/A
<i>Zygosaccharomyces bailii</i>	Kweichow Moutai Group	N/A
<b>Chemicals, peptides, and recombinant proteins</b>		
Ethanol, absolute	Shang hai Titan Scientific Co., Ltd.,	Cat#20210805
Sodium chloride	Shang hai Titan Scientific Co., Ltd.	Cat#20201016
Tryptone soy agar (TSA)	Beijing LandBridge Technology Co., Ltd.	Cat#T8650
Mueller Hinton blood agar plate	Huankai Microbial, Guangzhou, China	Cat#CP0230
LB broth	Sangon Biotech Co., Ltd., Shanghai, China	Cat#H106KA8345
<b>Software and algorithms</b>		
MATLAB <sup>23,26–28</sup>	MathWorks	<a href="http://www.mathworks.com">http://www.mathworks.com</a>
Python	Python Software Foundation	<a href="http://www.python.org">http://www.python.org</a>

(Continued on next page)

### Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
IRootLab	Morais et al. <sup>23</sup>	<a href="http://trevisan.github.io/irootlab/">http://trevisan.github.io/irootlab/</a>
MetaboAnalyst	Dr. Jianguo Xia from McGill University.	<a href="http://www.metaboanalyst.ca">http://www.metaboanalyst.ca</a>
Unscrambler X	CAMO Software	<a href="https://www.camo.com/unscrambler/">https://www.camo.com/unscrambler/</a>
Aspen Unscrambler	AspenTech	<a href="https://www.aspentech.com/">https://www.aspentech.com/</a>
BOMEM GRAMS/32	ABB	<a href="http://new.abb.com/products/measurement-products/analytical">http://new.abb.com/products/measurement-products/analytical</a>
OPUS IR Biotyper Software	Bruker	<a href="http://www.bruker.com/products.html">http://www.bruker.com/products.html</a>
OPUS	Bruker Optics	<a href="http://www.bruker.com/products.html">http://www.bruker.com/products.html</a>
BOMEM GRAMS/32	ABB	<a href="http://new.abb.com/products/measurement-products/analytical">http://new.abb.com/products/measurement-products/analytical</a>
OMNIC	Thermo Fisher Scientific	<a href="http://www.thermoscientific.com/en/products/fourier-transform-infrared-spectroscopy-ftir.html">http://www.thermoscientific.com/en/products/fourier-transform-infrared-spectroscopy-ftir.html</a>
Resolutions Pro	Agilent Technologies	<a href="http://crosslab.chem.agilent.com/">http://crosslab.chem.agilent.com/</a>
Spectrum 10	PerkinElmer	<a href="http://www.perkinelmer.com/">http://www.perkinelmer.com/</a>
LabSolutions IR	Shimadzu	<a href="http://www.shimadzu.com/products/index.html">http://www.shimadzu.com/products/index.html</a>
Spectra Manager	JASCO UK	<a href="http://www.jasco.co.uk/ftir.php">http://www.jasco.co.uk/ftir.php</a>

### Deposited data

Code and example datasets	Yang et al. <sup>29</sup>	<a href="https://doi.org/10.17632/tpkf9kyj7w.1">https://doi.org/10.17632/tpkf9kyj7w.1</a>
---------------------------	---------------------------	---

### Other

Eppendorf tubes	Shanghai Titan Scientific Co., Ltd.	N/A
Petri dish	Shanghai Titan Scientific Co., Ltd.	N/A
Disposable inoculation loops	Shanghai Titan Scientific Co., Ltd.	N/A
Vortexer XW-80A	Shanghai Huxi Instruments Co., Ltd.	N/A
Suitable pipettes for volumes from 10 $\mu$ L to 100 $\mu$ L	Shanghai Titan Scientific Co., Ltd.	N/A
Mask	Shanghai Titan Scientific Co., Ltd.	N/A
Safety glasses	Shanghai Titan Scientific Co., Ltd.	N/A
Pipette tips	Shanghai Titan Scientific Co., Ltd.	N/A
ATL-032R digital shaker	Shanghai Chemstar Instruments Co., Ltd. China	N/A
Centrifuge	Shanghai Boyu Instruments Co., Ltd.	N/A
Autoclave	Shanghai Boxun Industry & commerce Co., Ltd.	N/A
CaF <sub>2</sub> liquid cells	BioTools ( <a href="http://btools.com">http://btools.com</a> )	N/A
Air compressor machine	Taizhou Outstanding Industry And Trade Co., Ltd.	N/A
IR Biotyper	Bruker	N/A
INVENIO/70V/80V spectrometer	Bruker Optics	N/A
MB3000 FT-IR laboratory analyzer	ABB	N/A
Nicolet iS10 FT-IR spectrometer	Thermo Fisher Scientific	N/A
Cary 600 series FT-IR spectrometers	Agilent Technologies	N/A
Frontier spectrometers Spectrum Two IR spectrometers	PerkinElmer	N/A
IRTracer-100	Shimadzu	N/A
FTIR-4000 and FTIR-6000 FT-IR spectrometers	JASCO UK	N/A

## MATERIALS AND EQUIPMENT

**Table 1. Instruments, holders, and corresponding data acquisition software**

Manufacturer	Instrument	Acquisition type	Sample holder	Software	Website
Bruker	IR Biotyper	Transmission	Multiple	OPUS IR Biotyper Software	<a href="http://www.bruker.com/products.html">http://www.bruker.com/products.html</a>
Bruker Optics	INVENIO/70V/80V spectrometer	Transmission/ATR/reflectance	Multiple/single	OPUS	<a href="http://www.bruker.com/products.html">http://www.bruker.com/products.html</a>

(Continued on next page)

**Table 1. Continued**

Manufacturer	Instrument	Acquisition type	Sample holder	Software	Website
ABB	MB3000 FT-IR laboratory analyzer	Transmission	Single	BOMEM GRAMS/32	<a href="http://new.abb.com/products/measurement-products/analytical">http://new.abb.com/products/measurement-products/analytical</a>
Thermo Fisher Scientific	Nicolet iS10 FT-IR spectrometer	Transmission/ATR	Multiple/single	OMNIC	<a href="http://www.thermoscientific.com/en/products/fourier-transform-infrared-spectroscopy-ftir.html">http://www.thermoscientific.com/en/products/fourier-transform-infrared-spectroscopy-ftir.html</a>
Agilent Technologies	Cary 600 series FT-IR spectrometers	Transmission/ATR	Single	Resolutions Pro	<a href="http://crosslab.chem.agilent.com/">http://crosslab.chem.agilent.com/</a>
PerkinElmer	Frontier spectrometers Spectrum Two IR spectrometers	Transmission/ATR/ reflectance	Single	Spectrum 10	<a href="http://www.perkinelmer.com/">http://www.perkinelmer.com/</a>
Shimadzu	IRTracer-100	Transmission/ATR/ reflectance	Single	LabSolutions IR	<a href="http://www.shimadzu.com/products/index.html">http://www.shimadzu.com/products/index.html</a>
JASCO UK	FTIR-4000 and FTIR-6000 FT-IR spectrometers	Transmission	Single	Spectra Manager	<a href="http://www.jasco.co.uk/ftir.php">http://www.jasco.co.uk/ftir.php</a>

## STEP-BY-STEP METHOD DETAILS

IR-spectroscopic measurement modes include traditional absorbance/ transmission (A/T), attenuated total reflection (ATR), and diffuse reflectance mode (Figure 1).<sup>5,16,30</sup> FT-IR spectroscopy equipment: for a list of available commercial instruments, please refer to Table 1. The advantages of A/T mode are its higher signal-to-noise (SNR) ratios and high-throughput in contrast to ATR.<sup>31</sup> The A/T mode is more common and based on the detection of the transmitted IR radiation: the bacterial sample is placed in the path of the IR beam and scanned. In this protocol, transmission mode was employed for bacterial analysis.

## Preparation of bacterial sample

⌚ **Timing:** 12–24 h (for steps 1 to 6)

Many different sample preparation techniques for FT-IR spectroscopy have been published.<sup>15,17,24,32,33</sup> For accurate microbial typing, it is necessary to obtain a single purified colony stemming from a single, individual bacterial cell. Mixed cultures would lead to misidentifications. Agar plates are commonly used for growing bacterial cells, which also allow easy sample preparation. In certain cases, subculturing (on the same medium) may be necessary in order to adapt the organism to the growth medium. In any case, standardization of growing conditions (including media preparation, subculturing, and incubation conditions such as time and temperature, and cell recovery) is critical because any variation introduced in these steps will hamper typing results. Therefore, we recommend comparing the FT-IR signals of strains grown under the same conditions, including media and growth times, in order to achieve reliable results and high typing reproducibility.

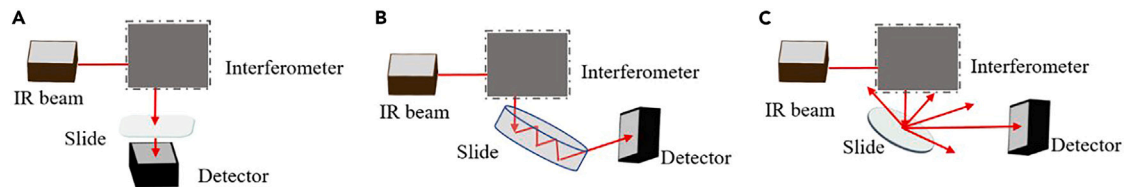
### 1. Bacterial culture

- Streak the isolated strains onto agar plates in a four-quadrant streak pattern by an inoculation loop under sterile conditions

**Note:** Once a specific culture medium is chosen, the protocol should be carried out using the same medium and the same water pH during media preparation.

- To culture clinical bacteria, such as *E. coli*, incubate bacteria for 16 h  $\pm$  30 min at an optimal temperature (35°C–37°C). For food microbiology, a temperature between 25°C and 30°C is often used.

⚠ **CRITICAL:** To achieve data compatibility between laboratories, a simple and reproducible standard is needed for cultivation conditions and experimental procedures. Agar plates



**Figure 1. The main IR radiation mode**

(A) Transmittance and typical bacterial FT-IR spectra, (B) attenuated total reflectance (ATR), (C) reflectance.

allow easy sample preparation and show the best discriminatory power in later typing; therefore, we recommend agar plates. If the strains cannot grow on agar plates or tryptone soy agar (TSA), or need blood or any other nutrition factor, or require incubation for several days to flourish, these strain suspensions should be washed 4–5 times by centrifugation in sterile water to remove culture media.

△ **CRITICAL:** A preliminary experiment (six known strains grown in the same culture conditions) is needed to determine and standardize the growth conditions with the best discriminatory power for typing, which is an essential prerequisite to starting extensive comparative studies. To assess the discriminatory power of the methodology, its robustness and reproducibility, and to differentiate taxonomic from methodological variability, several statistical values (measuring within-replicate variability) should be calculated, by e.g., averaging Pearson's product moment correlation coefficient and expressed as differentiation index (*D*). For a more detailed description of how to (approx. 20 mL), utilize, and assess reproducibility values (*D* value), see the refs.<sup>8,15,34</sup>

**Note:** If concentrations of salts and nutrients in the culture medium differ, it will introduce unwanted variability into the IR spectrum. Therefore, the medium complements must remain consistent in order to obtain reliable results in the multivariate analysis.

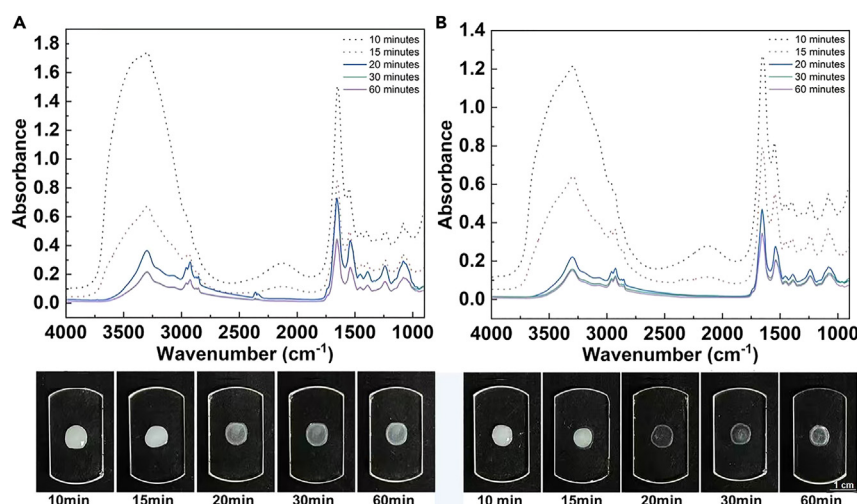
2. Remove the small amount of biomass (10–60 µg dry weight) from the bacterial lawn carefully from confluent colonies on the third quadrant of the agar surface (this technique yielded the best reproducibility) using an incubating loop.<sup>35</sup>

△ **CRITICAL:** For comparative FT-IR spectrometric analysis, only a single colony should be applied. It is important that the bacterial samples that are used are grown under similar conditions, particularly for bacterial analysis on the subspecies level. In any case, the user should avoid analyzing mixed bacterial samples stemming from mixed, impure cultures. The user should also avoid analyzing bacterial samples in the lag phase or that have entered the death phase. For this purpose, strict adherence to standardized growth conditions (incubation time, temperature, humidity, culture medium) is required.

**Note:** Contamination by medium components should be minimized by accurately picking the bacterial clones. Do not transfer agar into the sample, which requires special care when harvesting. Loops should be closed and without sharp ends.

3. After harvesting cells from these cultures by scraping, resuspend approximately 50 µg of microbial biomass in sterilized distilled water or 0.9% NaCl solution and wash three times. Centrifuge the suspension at 11000 × *g* at room temperature (20°C–25°C) for 2 min.

**Optional:** Submerge the bacterial cells in 35% ethanol aqueous solution in the suspension vial, and then remove the bacterial cells from the plastic inoculation loop by flipping. If the biomass



**Figure 2. Comparison of drying effects of different methods**

(A) FTIR spectra of bacteria dried at 40°C for different times, and the corresponding CaF<sub>2</sub> slides are shown below.  
(B) FTIR spectra of bacteria dried under vacuum (25°C, 20 kPa) for different times, and the corresponding CaF<sub>2</sub> slides are shown below.

in the loop is too sticky and will not come off, try to take a larger amount. If this does not work, cut off the loop with the adherent bacterial cells. A metal cylinder (Bruker Part No. 1851760) can be used. Shake the metal cylinder for about 1–2 min to remove the biomass from the loop, transfer the suspension into a new vial, and centrifuge at 11000 × g at room temperature (20°C–25°C) for 2 min.

**Note:** Some bacteria will stay in the loop unsuspended, but this may be compensated for by the mass on and around the loop.

4. After closing the suspension vial, vortex to form a homogeneous suspension.

⚠ **CRITICAL:** Prior to use in bacterial culture, all experimental materials and reagents should be autoclaved and handled under sterile conditions. Use sterile vessels or Petri dishes.

5. Smear 50 µL of bacterial suspension on the surface of a CaF<sub>2</sub> or ZnSe slide.

6. Dry the bacterial suspension at 40°C or under vacuum for 30 min.

**Note:** Thin transparent bacterial films form on the CaF<sub>2</sub> optical substrate. The degree of drying influences the FT-IR spectrum, as shown in Figure 2. Apply the sample as evenly as possible on the imprinted positions of the microtiter plate, i.e., the sample should cover each imprinted position completely and form a homogeneous layer.

⚠ **CRITICAL:** Control and standardize the exposure temperature and exposure time. Any temperature in the range of 25°C–40°C is acceptable. It is recommended that the moisture outside the sample not be visible to the naked eye when drying is completed. Excessive drying may affect the bacterial morphology, which in turn affects the authenticity of the bacterial typing.

**Note:** Be careful when handling the FT-IR slides. Hold the cell at its edges. Do not touch the cell surface, neither the front side nor the rear side. IR-spectroscopy is very sensitive to any contamination and will detect fingerprints or dust. We give a detailed example of sample

preparation and inoculation in the FT-IR equipment in the [supplemental information](#) (Figure S2).

**Note:** The cell should be delivered in a clean, ready-to-use state and packed in a reclosable plastic bag. After use they can be cleaned. For the microtiter plate, if enough free spots are still available, it can be used for another measurement run. You can clean and disinfect the microtiter plate by covering the surface first with 70% (v/v) isopropanol and then with deionized water and carefully rubbing the plate with a sponge or cleaning tissue. To remove any fatty residues, repeat the cleaning with 70% (v/v) isopropanol. To dry the microtiter plate, use clean pressurized air (oil-free) or a clean and dust-free cleaning cloth.

**Note:** If not successful, check [problem 1](#) in [troubleshooting](#) section.

### Equipment preparation

⌚ Timing: 40 min

Transmission mode is recommended for FT-IR spectroscopy collection. For FT-IR spectroscopy, water vapor rotational bands have a low intrinsic half-width and their precise wavenumber position. To collect high-quality FT-IR spectra, the FT-IR instrument should be placed in an enclosed space avoiding disturbance of air flow.

7. Turn on air-conditioning and a liquid desiccant system in the room in which you are doing the experiment. Control the temperature at 25°C.

⚠ **CRITICAL:** Only operate the FT-IR machine in a well-enclosed room; air movement must be minimized. Maintaining a dry and steady test environment is very important to avoid atmosphere changes inside the spectrometer.

8. Turn on the drying equipment for the instrument.

⚠ **CRITICAL:** Dry air or nitrogen should displace water vapor absolutely.

9. Turn on the IR source and the FT-IR spectrometer.

**Note:** The IR instrument should be permanently switched on. Each time after switching on the instrument, you must wait for ~1 h after power-up for the IR source to stabilize.

### Collecting FT-IR spectra

⌚ Timing: 40 min

It should be noted that some or many of the steps included could be performed by the FTIR software automatically and do not need to be executed by the lab operator. More details can be found in previous protocols.<sup>25,36</sup>

10. Set the spectrum collection parameters based on the chosen FT-IR instrument. Set spectrometer resolution to 4 cm<sup>-1</sup> and spectral range to 4,000–500 cm<sup>-1</sup>.

**Note:** To improve the SNR, the scan number should be high enough, always considering time measurement.

**Note:** If not successful, check [problem 2](#) in [troubleshooting](#) section.



11. Collect a background spectrum. The background spectra are used to check the spectral signals that originate from air, moisture (water vapor), and atmospheric changes.

△ **CRITICAL:** If the water vapor is not purged completely by dry air, there will be peaks in the regions of 1900–1750 cm<sup>-1</sup>. If these peaks appear, you should check the tubing and ensure no air leaks are present. You should also check the liquid desiccant system and drying equipment, and then purge for a longer period of time.

△ **CRITICAL:** It is advisable to obtain the single beam reference spectrum through an empty sample-cell system before the single beam sample spectrum is obtained. This can eliminate the contributions from possible impurities in the sample cells and minimize signals arising from CO<sub>2</sub> and water vapor caused by instabilities in the instrument and dry air purging of the system.

**Note:** A background spectrum should be taken if atmospheric changes occur.

**Note:** If not successful, check [problem 3](#) in [troubleshooting](#) section.

12. Collect a spectrum for one type of bacterial sample.

△ **CRITICAL:** To improve reproducibility and use chemometric approaches, the spectral resolution should be constant. During spectrum acquisition, the IR instruments can set a quality control to check the different spectral properties, e.g., absorbance intensity and SNR. Spots with spectra pass this quality test. Failed spectra should not be taken into account for further calculations.

**Note:** The sample cell must be handled with care. Make sure the microbic film does not flake off. Empty areas will result in poor absorption spectra caused by fringing from channel spectra, dispersion of the IR beam by incorrect absorbance from lack of sample.

**Note:** If not successful, check [problem 4](#) in [troubleshooting](#) section.

13. Subtract background spectra.

△ **CRITICAL:** To obtain high-quality bacterial IR spectra, the background spectrum must be carefully and adequately subtracted from the sample spectrum. Most of the IR instruments are equipped with software that can subtract background spectra automatically. Ensure the background spectrum is correctly set.

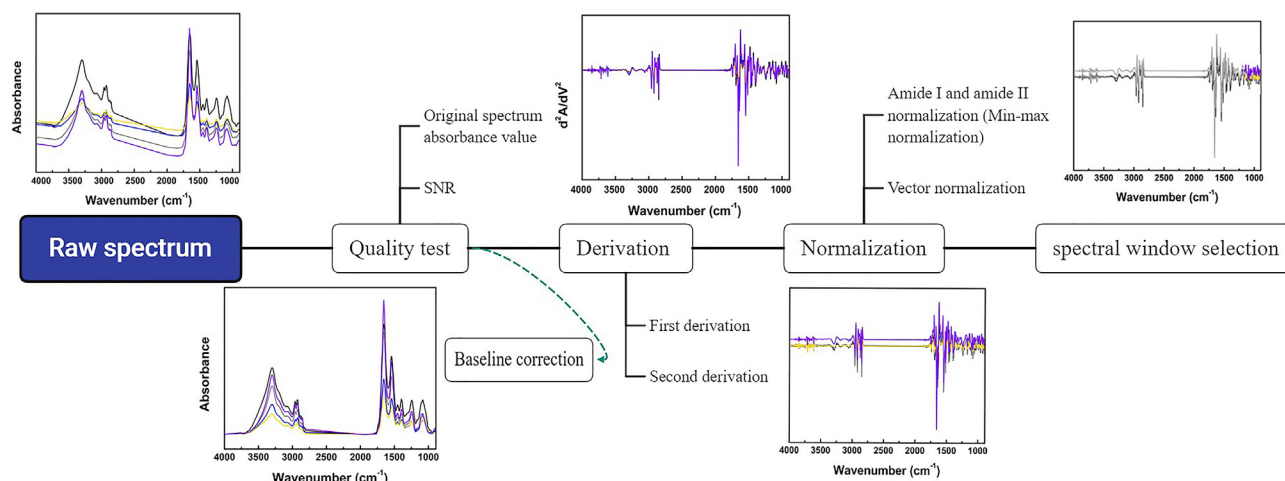
**Note:** If not successful, check [problem 5](#) in [troubleshooting](#) section.

14. Save the raw data on the hard drive or other media.

**Note:** Always save the data immediately to prevent loss if there is a power failure.

15. Clean the cell when finished. Wash the sample cell with ethyl alcohol (~5 mL), and then with an excess of water (~10 mL). Finally, gently wipe the windows with lens cleaning paper to keep them clean.

**Note:** The overriding principle of experimental disposal is that all waste should be decontaminated, autoclaved, or incinerated within the laboratory to guarantee "zero leakage" of infectious biohazards.



**Figure 3. Combination of preprocessing steps**

The raw spectral data are processed using the following sequence of preprocessing steps: (i) quality test of raw spectrum; (ii) signal amplification by first or second derivative; (iv) normalization; (v) spectral window selection. Baseline adjustment is performed to compare absorbance spectra. Baseline correction is not needed if derivatives are calculated but can be a point in the checklist for some cases.

## FT-IR spectrum preprocessing

⌚ Timing: 5 min

Preprocessing of the spectra is a significant issue, as the application and suitability of preprocessing can greatly affect the results of the computational analysis.<sup>37</sup> The preprocessing of spectra can correct variations related to spectrum acquisition, improve the robustness and accuracy of subsequent multivariate analyses, remove outliers, reduce the dimensionality of the data, and increase data interpretability by correcting issues associated with spectral data acquisition.<sup>38</sup> Preprocessing includes a spectral quality test, derivation, normalization, and spectrum window selection.<sup>39</sup> An illustration of the recommended preprocessing workflow based on FT-IR spectra is given in Figure 3. The previous protocols give detailed information about FT-IR spectrum preprocessing.<sup>25,40</sup>

16. Spectral quality test. The spectral quality analysis involves three aspects (Figure 4).
  - a. Check the absorbance value. The difference between the maximum and minimum absorbance values of the original spectrum at 4000–500  $\text{cm}^{-1}$  should be in the range of 0.3 to 2 absorbance units.
  - b. Check the SNR value.
    - i. Evaluated the signal to noise ratio (SNR) from the first derivation spectrum.

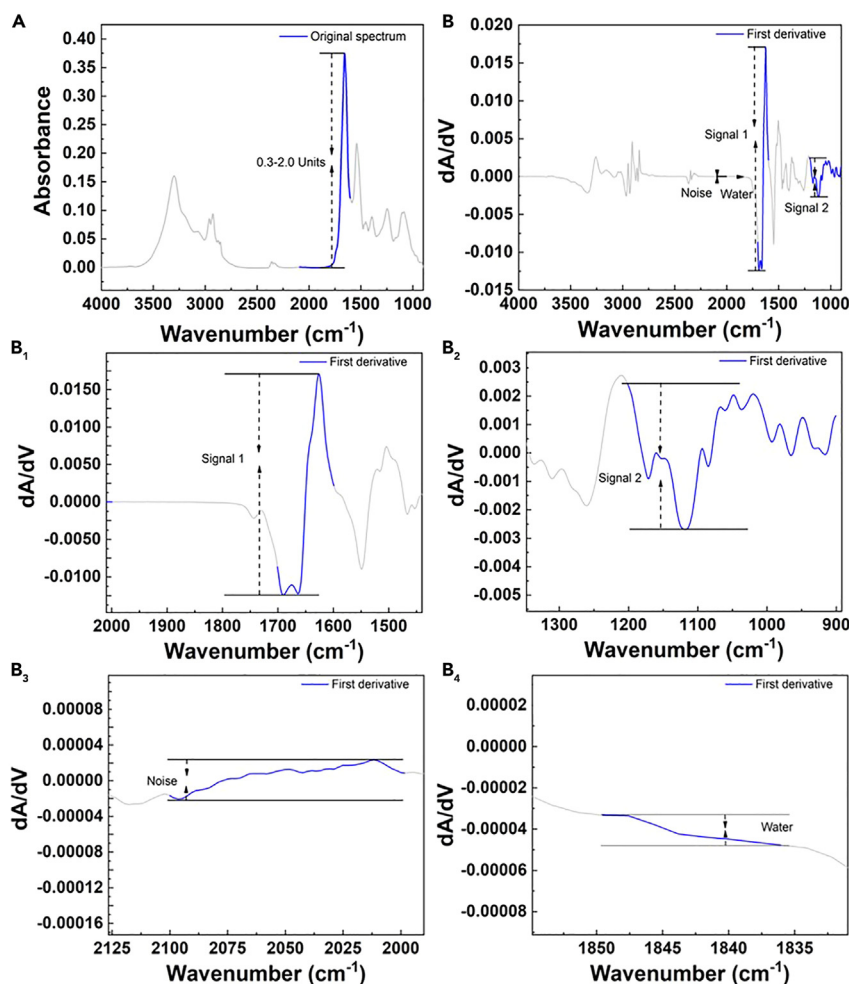
**Note:** As shown in Figure 4B, Signal 1, Signal 2, and Noise are the differences between the maximum and minimum values of the first derivatives at 1700–1600  $\text{cm}^{-1}$ , 1200–900  $\text{cm}^{-1}$ , and 2100–2000  $\text{cm}^{-1}$ , respectively.

- ii. Calculate the noise in any region free of sample absorption bands.

**Note:** the 2100–2000  $\text{cm}^{-1}$  region is designated as noise in this protocol. The value of Noise should not exceed 0.00015 units.

- iii. Calculate the SNR for Signal 1 and Signal 2.

**Note:** These must be greater than 400:1 and 80:1, respectively.<sup>41</sup>



**Figure 4. Strategy for quality testing the bacterial spectrum**

(A) The difference between the maximum and minimum absorbance values of the original spectrum were calculated in the spectral region of 2100–1600  $\text{cm}^{-1}$ .

(B) Signal 1, Signal 2, Noise, and water are the differences between the maximum and minimum values of the first derivatives at 1700–1600  $\text{cm}^{-1}$ , 1200–900  $\text{cm}^{-1}$ , 2100–2000  $\text{cm}^{-1}$ , and 1847–1837  $\text{cm}^{-1}$ , respectively. More details are shown in B<sub>1</sub>, B<sub>2</sub> and B<sub>3</sub>, B<sub>4</sub>.

**Note:** Data not meeting the quality criteria should be excluded from further analysis, and the user should check the sample, instrument, and instrumental setting.

c. Check the water vapor signals.

- Calculate the difference between the maximum and minimum values of the first derivative spectrum in the 1847–1837  $\text{cm}^{-1}$  region as the water vapor signal, which must not exceed  $3 \times 10^{-4}$ .
- Calculate the "signal/water" ratio from Signal 1, Signal 2, and water max values, Signal 1: Water > 100:1; Signal 2: Water > 20:1<sup>41</sup>

**△ CRITICAL:** To ensure the quality of bacterial IR spectra, at least five replicates are required in each class. Ideally, 3 biological (different days) and 3 other experimental (same plate) replicates should be performed for each isolate.

**Note:** If not successful, check [problems 6](#) and [7](#) in [troubleshooting](#) section.

17. Use Savitzky-Golay smoothing to remove spectral noise.
  - a. Try different window sizes from 3 to 21 and observe how the spectra change in shape and how the noise is reduced.
  - b. Use the smallest window that removes a considerable amount of the noise while maintaining the original spectral shape.
  - c. If the Savitzky-Golay method is used in the derivative analysis, the smoothing step can be omitted.
18. Savitsky-Golay derivation.
  - a. From the applications menu, select derivative function (the second-derivative is recommended).
  - b. Use a baseline-corrected seven-point Savitsky-Golay derivative function. Save the second-derivative spectrum.

△ **CRITICAL:** The derivative spectra show the details much better, allowing the differences between different strains to be visualized. In particular, second-derivative spectra (for which ease of visual inspection is often inverted around the frequency axis) are well suited for identifying component bands in a complex spectral region. The sharp and narrow absorption peaks from residual water vapor would be greatly enhanced by the second-derivative analysis and distort the protein amide I bands.

**Note:** Discard points below a threshold, and then fit the remaining points with a straight line.

19. Perform a normalization to correct different scales across spectra (e.g., due to different sample thicknesses when using FT-IR in transmission mode).

△ **CRITICAL:** Vector normalization is recommended. If the spectra are collected on different instruments, the SNRs will be quite different and the application of smoothing and normalization is necessary.

**Note:** As excessive smoothing can build up side lobes and periodic noise, which may be confused with true spectral features, the amount of smoothing should be kept to a minimum.

20. Cutting at interested region. Truncate the second-derivative spectrum to the appropriate sub-range(s) by selecting the ZAP function.
21. Save the results in a comma-separated value or CSV file (.csv), a text file (.txt), or an appropriate, compatible-format file.

### Multivariate analysis of spectral data for bacterial typing

⌚ **Timing:** 10 min (for steps 22 to 27)

Multivariate data analysis can be divided into unsupervised and supervised methods. Unsupervised methods, such as HCA and PCA, are used to describe the associations and patterns among a set of input measures. Such calculations do not necessarily rely on existing spectral databases or a priori knowledge. Supervised methods are machine learning techniques for creating a classification function from training data with a priori knowledge. These methods involve a supervised learning procedure in which models are created that map input objects (spectra) to desired outputs (class assignments). There are three essential steps in the supervised classification analysis: data selection, model construction, and model validation. The data selection step aims to split the dataset into at least two subsets: training and test. In general, the training set accounts for approximately 70% and the test set accounts for approximately 30%. The training set is then used to establish the model, whereas the test set is used to verify the established model. Support vector machine (SVM) and ANNs are recommended in microbial typing analysis.<sup>42–44</sup> In model validation, the receiver

operating characteristic (ROC) curve and confusion matrix are commonly used to evaluate the performance of the model.

22. Create a new folder in your file system. All the preprocessed files should be contained.
23. Assemble the data set.

**Optional:** When using commercial software not supporting multivariate analysis of spectral data, the user can follow the following procedure to assemble the data set.

- a. Execute the following steps in Excel.
  - i. Open the CSV or text file created in the previous step.
  - ii. Next, enter the class labels in this column (take care not to miss any boxes).
  - iii. Save the spreadsheet as a 'Text (tab-delimited)' file.
  - iv. Choose a data analysis procedure appropriate to your analysis goal.

**△ CRITICAL:** Each list file is saved in a .csv or text file (.txt) format containing several columns, with the first column corresponding to wavenumber ( $\text{cm}^{-1}$ ) and the remaining columns corresponding to peak intensities. The replicates should share the same class label. At least five replicates are required in each class.

**Note:** The data (except class labels) must not contain non-numeric values. Missing values or features should be given zeros.

24. Analysis of the spectra by using software. **Option A: MetaboAnalyst.**
  - a. Enter <https://www.metaboanalyst.ca/> in the address bar of the browser and click on the Statistical Analysis button on the corresponding website.
  - b. Upload the files, choose Data Type: spectral bins; Format: samples in columns (unpaired); Data file: .csv.
  - c. Check data integrity to ensure that the data meets the basic requirements for meaningful downstream analysis.
  - d. Low-quality data filtering. Select the default "Interquartile range (IQR)" and click the Process button.
  - e. Select "Normalization by sum" for sample normalization. After normalization, click the Proceed button.

**Note:** MetaboAnalyst is a comprehensive Web application for metabolomic data analysis and interpretation. It provides a variety of data processing and normalization procedures and supports a number of data analysis and data visualization tasks using a range of univariate and multivariate methods, such as PCA, heatmap clustering, and machine learning methods. The software was carefully designed to enable researchers with little statistical or computational background to perform complex data analysis procedures commonly used in metabolomic studies.

- f. Click the Submit button to proceed to the Data Analysis Exploration page. The user can choose the preferred mathematical method.
  - i. Choose one of the major analysis categories (univariate analysis, multivariate analysis, significant feature identification, cluster analysis, and classification & feature selection).
  - ii. Choose one of the subcategories containing individual methods. The most particular approaches useful for clustering are PCA and dendrogram, which are listed under the Multivariate Analysis category, and heatmaps, which is listed under the Cluster Analysis category.
- g. Click Download to generate a .pdf report (Analysis\_Report.pdf).

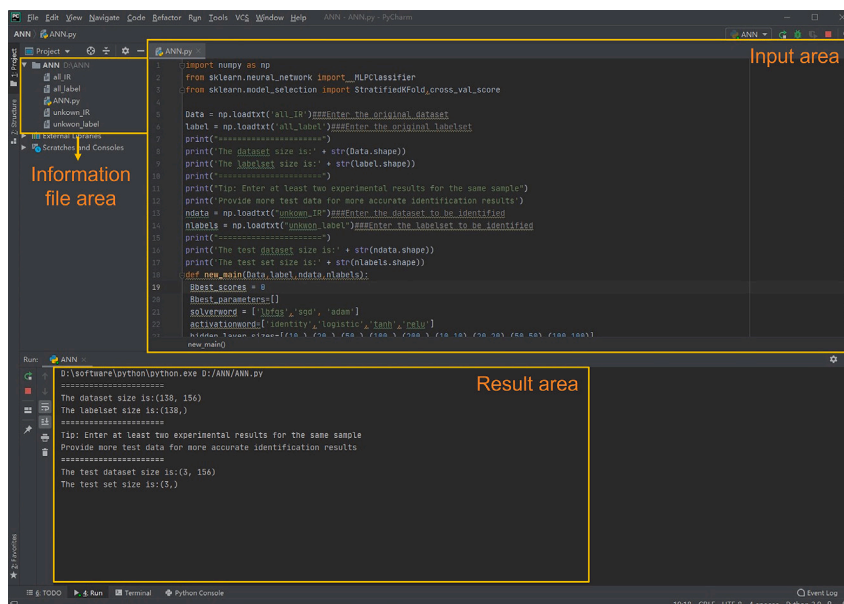


Figure 5. The running interface of the ANN script

△ **CRITICAL:** In general, we recommend using the dendrogram view first.

**Note:** If not successful, check [problems 8](#) and [9](#) in [troubleshooting](#) section.

## 25. Analysis of the spectra by using software. Option B: Home-made tool.

**Note:** Discrimination by FTIR spectroscopy can also be achieved by combining it with supervised learning methods. Here, we also give a feed-forward three-layer ANN method: the information propagates from the input layer, through the hidden layers, to the output layers. During the iterative training process, the connection weights (numbers of inputs and hidden neurons) are automatically adjusted until the global error is minimized.<sup>16</sup> We provide this home-made tool for ANNs in the [supplemental information](#) section ([Data S1](#)).

- Install Python and the related packages.

**Note:** Put the data and script into the same folder (the data set should be in .csv format and compatible with MetaboAnalyst).

- Open the given script. According to the script prompt, enter the name of the "data\_set" and "unknown\_data" in the corresponding location.
- Run program. The the running interface of the script were shown in [Figure 5](#). The ANN model will be trained and validated using the "data\_set" and the "unknown\_data" will be identified by the model.

**Optional:** In commercial software (such as IR biotyper), data analysis and collection are closely related. Usually, after data collection, the software can automatically analyze and provide results without too much intervention.

## 26. Analysis of the spectra by using software. Option C: Commercial software (such as IR biotyper or OPUS) that supports multivariate analysis of spectral data.

- First, set relative information about the strains.

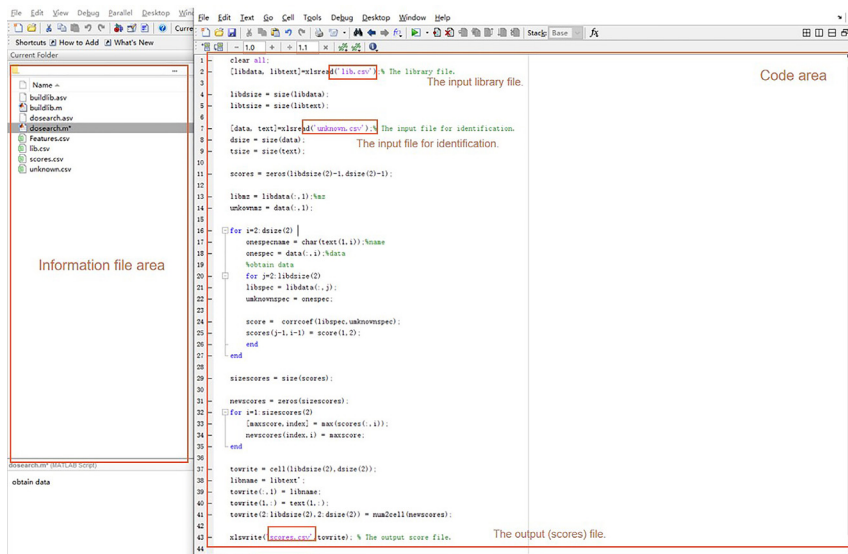
- i. Select or add species name;
- ii. add isolate labels (only one label can be added to isolates);
- iii. input incubation protocols in the software (incubation time, temperature, incubation medium);
- iv. set the preparation protocols. Spectra can be compared only if the same preparation protocol was used. The protocol consists of a name, the suspension medium, such as sterilized distilled water, 0.9% NaCl, or 35% ethanol aqueous solution, and a description. The user can also attach a standard operating procedure (SOP).
- v. acquisition methods, strain typing methods (we always use Euclidean distance metric);
- vi. set exploration methods. Various exploration methods can be chosen in IR Biotyper, and the user can add other algorithms. Generally, Euclidean-ward's algorithm method comprising agglomerative hierarchical clustering using Euclidean distances for distance matrix calculation and Ward's method as the clustering method is almost always used;
- b. Set relative information about the sample plate, the background position, and the test standard position;
- c. Set the Data file path, including IR data directory, IR acquisition control, stage calibration file, and quality test method;
- d. Add the sample position to a table;
- e. Load a target plate into the IR spectrometer;
- f. Go to the test standard position on the IR plate. Measure the standard sample to ensure the instrument is prepared for testing;
- g. Start measuring samples;
- h. For bacterial typing, the spectra with good quality are selected as one project. Select the typing method or add the typing algorithm. Then press 'Run.' The results appear for all FT-IR spectra. For graphical correlations, an average statistical algorithm is implemented in the IR Biotyper software. IR Biotyper can provide visualization tools, such as dendrogram and distance matrix (see [supplemental information, Figures S3 and S4](#)). The user can choose the visualization tools and output the visualization result under specified path.

## 27. Analysis of the spectra by using software. **Optional D: Matlab.**

**Note:** Matlab can also be employed for the genetic algorithm and successive projection algorithm.<sup>23,26–28</sup> Here, we give an example of a Matlab script for library construction and bacterial predication. This Matlab script were provided in the [supplemental information](#) section ([Data S2](#)), and the the running interface of the Matlab script were shown in [Figure 6](#).

- a. Put "Features.csv" (characteristic-feature-data files of a series bacteria) and "buildlib.m" in the same directory.
- b. Run the script in Matlab to construct a reference library file (e.g., "lib.csv").
- c. Set the input and output file names by modifying the corresponding variable in "buildlib.m". The library should include a high range of taxonomic units covering a large scope of taxonomic diversity, similar to that found in the unknown samples. A minimum number of spectra (10) belonging to each taxonomic unit should be included.
- d. The characteristic-feature-data file of unknown bacteria is created (unknown.csv) using the same feature extraction method.
- e. The distance is measured by another Matlab script (dosearch.m). Put "unknown.csv", "lib.csv", and "dosearch.m" in the same directory. Specify the input file ("unknown.csv"), library file ("lib.csv"), and output score file (e.g., "scores.csv") in the script, and then run "dosearch.m" in Matlab to obtain the statistical distance between the unknown spectrum and reference spectra in the library. The scripts are described in the [supplemental information](#).





**Figure 6. The running interface of the Matlab script**

- f. Score and prediction. Similarity scores are recorded in the output file ("scores.csv"). Each score is a metric of the distance between the unknown and reference spectra. A higher score indicates greater similarity, and the highest matching scores and related spectra in the library are selected as the final results among all of the comparisons between the unknown spectrum and the spectra in the library.

**△ CRITICAL:** The score derived from the distance indicates the similarity between the unknown and known bacteria. The similarity between the unknown and known bacteria can be calculated by a distance measure algorithm, most commonly the Pearson correlation coefficient.<sup>45–47</sup> A higher score means a shorter distance and will be obtained when the values in each pair are closer. "

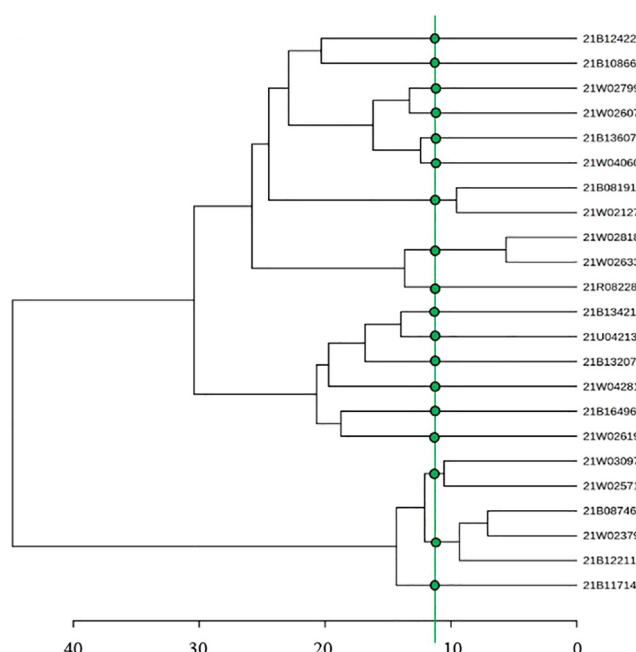
## EXPECTED OUTCOMES

In this subsection, we measured clinical and industrial microbiology from different biological samples following the protocol steps described above. We focus on the application of the protocols introduced above and compared the FT-IR typing results with a molecular biology method to verify this protocol.

### Typing of *E. coli* at the subspecies level

Nosocomial infections are a leading cause of avoidable harm in hospital patients. Microbial typing methods allow the discrimination of diverse organisms of the same species.<sup>48,49</sup> The accurate typing of infections is vital for control of nosocomial infections.<sup>50,51</sup> We analyzed 23 *E. coli* clinical strains from Peking Union Medical College Hospital. Before FT-IR typing, we first performed their MALDI-TOF MS identification to determine bacterial species (Figure S5, Table S2 in the supplemental information) and results confirmed they all belonged to the same species. FT-IR spectroscopy was performed on the 23 *E. coli* strains and analyzed according to this protocol. Briefly, the spectroscopy data were preprocessed by quality testing, Savitzky-Golay second derivative (window = 7 points, second-order polynomial function), and vector normalization (2-norm, length = 1). Finally, the biological fingerprint region (900–1200 cm<sup>-1</sup>) of the second derivative spectrum was selected for multivariate analysis. As shown in Figure 7, 23 *E. coli* strains were classified into 18 clusters in HCA. The second derivative spectra of all 23 *E. coli* strains are shown in Figure S6.





**Figure 7. HCA of the second derivative FT-IR spectra, illustrating the relationship between these strains**

Typing scheme for select *E. coli* strains. The cut-off value defines up to which distance the spectra are considered to be in the same cluster. The green spot indicates that the strains are same clone.

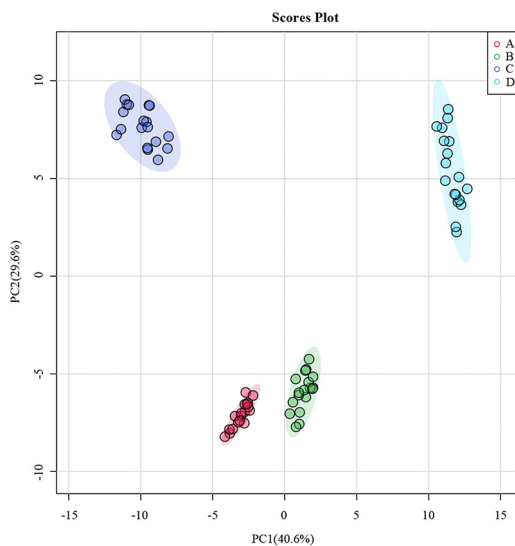
The concordance of the results of the FT-IR typing methods with other typing methods was evaluated. PFGE and WGS are regarded as reference typing methods for bacteria. An online tool was used to evaluate the classification concordance ([www.comparingpartitions.info](http://www.comparingpartitions.info)) of WGS, PFGE, MLST, and FT-IR. The detailed results are provided in the [supplemental information](#) (Table S3, Figures S7 and S8). Simpson's index of diversity (SID) were used to calculate the discriminatory power of each method and the 95% confidence interval (CI) to account for the variability. SID assumes values between 0 and 1, in practice, and should be at least in the range of 0.90 to 0.95 for a typing system to be considered robust.<sup>52</sup> As shown in Table S4, the SID of all typing methods except MLST was greater than 0.9, and PFGE and WGS exhibited higher concordance. FT-IR presented better discriminatory power than MLST (SID 0.976 vs. 0.834) and was comparable to PFGE and WGS.

#### Typing of yeast at species level by the unsupervised method

By using the same procedures, FT-IR was also used for typing of fungi.<sup>53–55</sup> Yeast is a kind of unicellular eukaryotic microorganism with semi permeable cell membrane. Due to its relationship with food flavors and rich nutrients, the typing of yeast is often needed in distilleries and food factories. Four strains of yeast (*Pichia kudriavzevii*, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, *Zygosaccharomyces bailii*) isolated from Kweichow Moutai distilleries were typed by FT-IR spectroscopy (Figures S9, S10, and Figure 8). FT-IR showed good discriminatory power at the species level.

#### Typing of *Saccharomyces* at subspecies level with the supervised ANN method

*Saccharomyces* are important microorganisms used for processing alcoholic beverages and breads.<sup>56,57</sup> We typed 10 strains of *Saccharomyces* isolated from different areas of Kweichow Moutai distilleries by FT-IR spectroscopy and compared the results to the SSCP method, which is regarded as an effective typing method for *Saccharomyces*.<sup>58</sup> The MALDI-TOF MS identification method showed that the 10 strains belong to the same species (*Saccharomyces cerevisiae*, Table S5). The SSCP results (Figure S11) showed that the 10 strains are different at the subspecies level. The FT-IR spectrum (Figure S3) is provided in the [supplemental information](#). ANN-based methods were used to classify these 10 isolates. The classification model was validated by 6-fold



**Figure 8. Two-dimensional projection score plots of four strains of yeast analyzed by PCA**

(A) *Pichia kudriavzevii*, (B) *Saccharomyces cerevisiae*, (C) *Zygosaccharomyces bailii*, (D) *Schizosaccharomyces pombe*.

cross-validation (Figure 9). The results show that the classification accuracy of ANN-based methods was 91.67%. The ROC curve had an average AUC of 0.9792. Therefore, ANN can correctly identify *Saccharomyces* at subspecies level from the FT-IR data.

## LIMITATIONS

The limitations of FT-IR are that only microorganisms that can be grown in culture and have been isolated as pure cultures can be analyzed; reproducible results can only be obtained when the growth conditions (culture medium, cultivation time, and temperature) are controlled and unified; the spectral databases from different laboratories are not compatible unless all conditions (instrument, culture media, etc.) are the same; and only a few highly specialized databases have been developed.<sup>59,60</sup> Many general applications require a broad and comprehensive database, as in clinical, food, or environmental microbiology, as reference databases need to deal with the diversity in species.<sup>61</sup>

FT-IR spectroscopy is particularly recognized as being well suited to bacterial typing below species level. However, the FT-IR spectral pattern is very sensitive to sample preparation, making it difficult to build a globally accessible FT-IR database for bacterial identification.<sup>61</sup> MALDI-TOF mass spectrometry is a more commonly used alternative bacterial identification technique based on specific mass fingerprints of bacterial ribosomal proteins, which are relatively independent of the culture conditions used to grow bacteria.<sup>62</sup> The user can also combine the strength of MALDI-TOF MS and IR for bacterial typing.<sup>1</sup>

## TROUBLESHOOTING

### Problem 1

Insufficient sample purity (Section “[preparation of bacterial sample](#)”, step 6).

#### Potential solution

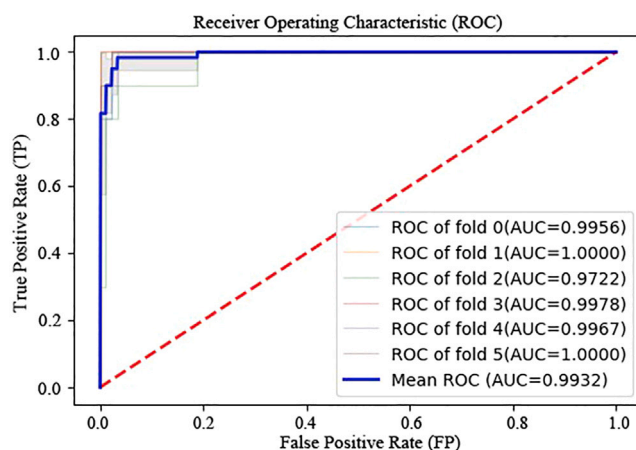
Improve the complex isolation/purification procedure.

### Problem 2

Very low IR absorbance signal or no signal (Section “[collecting FT-IR spectra](#)”, step 10).

#### Potential solution

Check the quantity of biomass. Ensure there are no bubbles or empty cell areas when collecting the sample spectrum.



**Figure 9.** ROC curves for the typing of bacterial samples by 6-fold cross-validation in the ANN-based method at the subspecies level

### Problem 3

Unwanted water vapor band in the spectra (Section “collecting FT-IR spectra”, step 11).

#### Potential solution

Check if the desiccant cartridge in the microplate compartment or the spectrometer compartment is exhausted and needs to be replaced or regenerated. Maintain appropriate atmospheric stability in the lab room.

### Problem 4

Low-quality FT-IR spectra (low absorption values, inadequate signal to noise ratio), (Section “collecting FT-IR spectra”, step 12).

#### Potential solution

(i) Check and standardize amount of cell biomass. (ii) Standardize incubation conditions as reported above. (iii) Resuspend biomass in less volume for chemical treatment.

### Problem 5

Large variability in replicates obtained from the same taxonomic unit (Section “collecting FT-IR spectra”, step 13).

#### Potential solution

Standardize growth conditions (time, temperature, and growth media). (ii) Verify that bacteria are collected during the same growth (log or stationary) phase. (iii) Standardize sample preparation.

### Problem 6

Great variability between independently acquired spectra from the same strain (Section “FT-IR spectrum preprocessing”, step 16, c ).

#### Potential solution

Likely due to: (i) contamination of different clones cultured in liquid culture or grown on the solid media; (ii) the presence of culture medium adhering to the bacterial colony; (iii) one or the same strain of bacteria cultured in different conditions. Ideally at least 3 independently grown replicates, optimally around 5, should be processed and analyzed. Check purity of cultures and repeat cultivation under standardized growth conditions.

### Problem 7

Samples are processed with low reproducibility (Section “[FT-IR spectrum preprocessing](#)”, step 16, c).

### Potential solution

Check if the sample is too thick or too thin in spots. Bubbles in the sample will also increase the variance within the technical replicates. Growth conditions should be standardized. *D* values or similar parameters are used to assess reproducibility and to differentiate taxonomic from methodological variability. If within-sample reproducibility is low, improvement is needed (see text).

### Problem 8

Difficulties for establishing the cut-off for dendrograms (Section “[multivariate analysis of spectral data for bacterial typing](#)”, step 24).

### Potential solution

The user has to decide on the limit for related vs. unrelated. The software can help by taking into account the reproducibility of the replicate measurements. Inclusion of taxonomically related and non-related strains (outliers) should help upscale the dendrogram. The reproducibility can be assessed by calculating the spectral distance/similarity between the same or different strains. *D* values or similar parameters can be used to assess reproducibility.

### Problem 9

Two spectra from the same strain appear to be different (Section “[multivariate analysis of spectral data for bacterial typing](#)”, step 24).

### Potential solution

Check the bacteria culture conditions (medium, time, temperature). Check the way samples have been prepared and analyzed. Use *D* values or similar parameters to assess reproducibility.

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Shaoning Yu ([yushaoning@nbu.edu.cn](mailto:yushaoning@nbu.edu.cn)).

### Materials availability

This protocol does not generate unique reagents.

### Data and code availability

The code and example datasets are available at Mendeley Data (<https://doi.org/10.17632/tpkf9kyj7w.1>).

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xpro.2023.102223>.

## ACKNOWLEDGMENTS

This work was supported by Moutai Group Research and Development Project (no. 2018023) and the National Natural Science Foundation of China (nos. 31470786, U1904196, 82073699, and 21275032).

## AUTHOR CONTRIBUTIONS

S.Y., L.W., M.P., and C.D. are the principal investigators who conceived the idea for the manuscript; H.S., H.Y., B.F., F.Y., L.C., and H.S. provided information regarding FT-IR data; H.Y., H.S., B.F., M.P., L.W., and S.Y. wrote the manuscript; F.Y., H.S., L.Z., A.O., and C.D. contributed to the anticipated

results section and provided material for sample preparation; and S.Y., L.W., and M.P. brought together the text and finalized the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

- Shi, H., Sun, J., Han, R., Ding, C., Hu, F., and Yu, S. (2020). The strategy for correcting interference from water in Fourier transform infrared spectrum based bacterial typing. *Talanta* 208, 120347. <https://doi.org/10.1016/j.talanta.2019.120347>.
- Alvarez-Ordóñez, A., and Prieto, M. (2012). Fourier transform infrared spectroscopy to assist in taxonomy and identification of foodborne microorganisms. In *Fourier Transform Infrared Spectroscopy in Food Microbiology* (Springer US), pp. 19–21. [https://doi.org/10.1007/978-1-4614-3813-7\\_2](https://doi.org/10.1007/978-1-4614-3813-7_2).
- Sabat, A.J., Budimir, A., Nashev, D., Sá-Leão, R., van Dijk, J.M., Laurent, F., Grundmann, H., and Friedrich, A.W.; ESCMID Study Group of Epidemiological Markers ESGEM (2013). Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill.* 18, 17–30. <https://doi.org/10.2807/ese.18.04.20380-en>.
- MacCannell, D. (2013). Bacterial strain typing. *Clin. Lab. Med.* 33, 629–650. <https://doi.org/10.1016/j.cll.2013.03.005>.
- Quintelas, C., Ferreira, E.C., Lopes, J.A., and Sousa, C. (2018). An overview of the evolution of infrared spectroscopy applied to bacterial typing. *Biotechnol. J.* 13, 1700449. <https://doi.org/10.1002/biot.201700449>.
- Novais, Â., Freitas, A.R., Rodrigues, C., and Peixe, L. (2019). Fourier transform infrared spectroscopy: unlocking fundamentals and prospects for bacterial strain typing. *Eur. J. Clin. Microbiol. Infect. Dis.* 38, 427–448. <https://doi.org/10.1007/s10096-018-3431-3>.
- Kim, S., Burgula, Y., Ojanen-Reuhs, T., Cousin, M.A., Reuhs, B.L., and Mauer, L.J. (2006). Differentiation of crude lipopolysaccharides from *Escherichia coli* strains using Fourier transform infrared spectroscopy and chemometrics. *J. Food. Sci.* 71, 57–61. <https://doi.org/10.1111/j.1365-2621.2006.tb08908.x>.
- Mouwen, D., Hörman, A., Korkeala, H., Alvarez-Ordóñez, A., and Prieto, M. (2011). Applying Fourier transform infrared spectroscopy and chemometrics to the characterization and identification of lactic acid bacteria. *Vib. Spectrosc.* 56, 193–201. <https://doi.org/10.1016/j.vibspec.2011.02.008>.
- De Bruyne, S., Speckaert, M.M., and Delanghe, J.R. (2018). Applications of mid-infrared spectroscopy in the clinical laboratory setting. *Crit. Rev. Clin. Lab. Sci.* 55, 1–20. <https://doi.org/10.1016/j.cll.2017.141412>.
- Amiel, C., Curk, M.C., Denis, C., Ingouf, J.C., Reyrolle, J., and Traver, J. (1997). Identification and Classification of Bacteria by Fourier Transform Infrared Spectroscopy (FTIR).
- Kim, S., Kim, H., Reuhs, B.L., and Mauer, L.J. (2006). Differentiation of outer membrane proteins from *Salmonella enterica* serotypes using Fourier transform infrared spectroscopy and chemometrics. *Lett. Appl. Microbiol.* 42, 229–234. <https://doi.org/10.1111/j.1472-765X.2005.01828.x>.
- Burgula, Y., Khali, D., Kim, S., Krishnan, S.S., Cousin, M.A., Gore, J.P., Reuhs, B.L., and Mauer, L.J. (2007). Review of mid-infrared Fourier transform-infrared spectroscopy applications for bacterial detection. *J. Rapid Methods Autom. Microbiol.* 15, 146–175. <https://doi.org/10.1111/j.1745-4581.2007.00078.x>.
- Grunert, T., Wenning, M., Barbagelata, M.S., Fricker, M., Sordelli, D.O., Buzzola, F.R., and Ehling-Schulz, M. (2013). Rapid and reliable identification of *Staphylococcus aureus* capsular serotypes by means of artificial neural network-assisted Fourier transform infrared spectroscopy. *J. Clin. Microbiol.* 51, 2261–2266. <https://doi.org/10.1128/jcm.00581-13>.
- Jöhler, S., Stephan, R., Althaus, D., Ehling-Schulz, M., and Grunert, T. (2016). High-resolution subtyping of *Staphylococcus aureus* strains by means of Fourier-transform infrared spectroscopy. *Syst. Appl. Microbiol.* 39, 189–194. <https://doi.org/10.1016/j.syapm.2016.03.003>.
- Helm, D., Labischinski, H., Schallehn, G., and Naumann, D. (1991). Classification and identification of bacteria by Fourier-transform infrared spectroscopy. *J. Gen. Microbiol.* 137, 69–79. <https://doi.org/10.1099/00221287-137-1-69>.
- Naumann, D. (2006). Infrared spectroscopy in microbiology. In *Encyclopedia of Analytical Chemistry* (John Wiley & Sons). <https://doi.org/10.1002/9780470027318.a0117>.
- Naumann, D., Helm, D., and Labischinski, H. (1991). Microbiological characterizations by FT-IR spectroscopy. *Nature* 351, 81–82. <https://doi.org/10.1038/351081a0>.
- Movasaghi, Z., Rehman, S., and ur Rehman, D.I. (2008). Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* 43, 134–179. <https://doi.org/10.1080/05704920701829043>.
- Song, C.L., and Kazarian, S.G. (2020). Effect of controlled humidity and tissue hydration on colon cancer diagnostic via FTIR spectroscopic imaging. *Anal. Chem.* 92, 9691–9698. <https://doi.org/10.1021/acs.analchem.0c01002>.
- Li, X., Zhu, L., Wang, X., Li, J., and Tang, B. (2022). Evaluation of IR biotyper for lactiplantibacillus plantarum typing and its application potential in probiotic preliminary screening. *Front. Microbiol.* 13, 823120. <https://doi.org/10.3389/fmicb.2022.823120>.
- Davis, R., and Mauer, L.J. (2010). Fourier transform infrared (FT-IR) spectroscopy: a rapid tool for detection and analysis of foodborne pathogenic bacteria. In *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology*, Méndez-Vilas, ed., pp. 1582–1594.
- Trevisan, J., Angelov, P.P., Carmichael, P.L., Scott, A.D., and Martin, F.L. (2012). Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst* 137, 3202–3215. <https://doi.org/10.1039/C2AN16300D>.
- Moraes, C.L.M., Lima, K.M.G., Singh, M., and Martin, F.L. (2020). Tutorial: multivariate classification for vibrational spectroscopy in biological samples. *Nat. Protoc.* 15, 2143–2162. <https://doi.org/10.1038/s41596-020-0322-8>.
- Baker, M.J., Trevisan, J., Bassan, P., Bhargava, R., Butler, H.J., Dorling, K.M., Fielden, P.R., Fogarty, S.W., Fullwood, N.J., Heys, K.A., et al. (2014). Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protoc.* 9, 1771–1791. <https://doi.org/10.1038/nprot.2014.110>.
- Martin, F.L., Kelly, J.G., Llabiani, V., Martin-Hirsch, P.L., Patel, I.I., Trevisan, J., Fullwood, N.J., and Walsh, M.J. (2010). Distinguishing cell types or populations based on the computational analysis of their infrared spectra. *Nat. Protoc.* 5, 1748–1760. <https://doi.org/10.1038/nprot.2010.133>.
- De Gussem, K., De Gelder, J., Vandenabeele, P., and Moens, L. (2009). The biodata toolbox for MATLAB. *Chemometrics and intelligent laboratory systems. Chemometr. Intell. Lab. Syst.* 95, 49–52. <https://doi.org/10.1016/j.chemolab.2008.08.003>.
- Cordella, C.B., and Bertrand, D. (2014). SAISIR: a new general chemometric toolbox. *TRAC Trends Anal. Chem.* 54, 75–82. <https://doi.org/10.1016/j.trac.2013.10.009>.
- Stöckel, S., Kirchhoff, J., Neugebauer, U., Rösch, P., and Popp, J. (2016). The application of Raman spectroscopy for the detection and identification of microorganisms. *J. Raman Spectrosc.* 47, 89–109. <https://doi.org/10.1002/jrs.4844>.
- Huayan, Y., Shaoning, Y., Fan, Y., Miguel, P., Haimei, S., Bin, F., Hao, sh., Shouning, Y., Liangqiang, C., Avelino, A.O., et al. (2023). Dataset for "protocol for bacterial typing using Fourier transform infrared spectroscopy. Mendeley Data 1. <https://doi.org/10.17632/tkpf9kyj7w.1>.
- Griffiths, P., and Haseth, J. (2006). *Fourier Transform Infrared Spectrometry*, Second Edition. <https://doi.org/10.1002/9780470106310.ch9>.

31. Avelino Alvarez-Ordóñez, M.P. (2012). Fourier transform infrared spectroscopy in food microbiology. In *Fourier Transform Infrared Spectroscopy in Food Microbiology* (Springer).
32. Yang, H., Yang, S., Kong, J., Dong, A., and Yu, S. (2015). Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nat. Protoc.* 10, 382–396. <https://doi.org/10.1038/nprot.2015.024>.
33. Isozaki, A., Mikami, H., Hiramatsu, K., Sakuma, S., Kasai, Y., Iino, T., Yamano, T., Yasumoto, A., Oguchi, Y., Suzuki, N., et al. (2019). A practical guide to intelligent image-activated cell sorting. *Nat. Protoc.* 14, 2370–2415. <https://doi.org/10.1038/s41596-019-0183-1>.
34. Mouwen, D.J.M., Weijters, M.J.B.M., Capita, R., Alonso-Calleja, C., and Prieto, M. (2005). Discrimination of enterobacterial repetitive intergenic consensus PCR types of *Campylobacter coli* and *Campylobacter jejuni* by Fourier transform infrared spectroscopy. *Appl. Environ. Microbiol.* 71, 4318–4324. <https://doi.org/10.1128/aem.71.8.4318-4324.2005>.
35. Helm, D., Labischinski, H., and Naumann, D. (1991). Elaboration of a procedure for identification of bacteria using Fourier-Transform IR spectral libraries: a stepwise correlation approach. *J. Microbiol. Methods* 14, 127–142. [https://doi.org/10.1016/0167-7012\(91\)90042-O](https://doi.org/10.1016/0167-7012(91)90042-O).
36. Butler, H.J., Ashton, L., Bird, B., Cinque, G., Curtis, K., Dorney, J., Esmonde-White, K., Fullwood, N.J., Gardner, B., Martin-Hirsch, P.L., et al. (2016). Using Raman spectroscopy to characterize biological materials. *Nat. Protoc.* 11, 664–687. <https://doi.org/10.1038/nprot.2016.036>.
37. Byrne, H.J., Knief, P., Keating, M.E., and Bonnier, F. (2016). Spectral pre and post processing for infrared and Raman spectroscopy of biological tissues and cells. *Chem. Soc. Rev.* 45, 1865–1878. <https://doi.org/10.1039/C5CS00440C>.
38. Lasch, P. (2012). Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. *Chemometr. Intell. Lab. Syst.* 117, 100–114. <https://doi.org/10.1016/j.chemolab.2012.03.011>.
39. Siqueira, L.F., and Lima, K.M. (2016). A decade (2004 – 2014) of FTIR prostate cancer spectroscopy studies: an overview of recent advancements. *Trac-Trend. Anal. Chem.* 82, 208–221. <https://doi.org/10.1016/j.trac.2016.05.028>.
40. Morais, C.L.M., Paraskevaidi, M., Cui, L., Fullwood, N.J., Isabelle, M., Lima, K.M.G., Martin-Hirsch, P.L., Sreedhar, H., Trevisan, J., Walsh, M.J., et al. (2019). Standardization of complex biologically derived spectrochemical datasets. *Nat. Protoc.* 14, 1546–1577. <https://doi.org/10.1038/s41596-019-0150-x>.
41. Naumann, D. (2008). FT-IR spectroscopy of microorganisms at the Robert Koch Institute: experiences gained during a successful project. *Proc. SPIE* 6853, 68530G. <https://doi.org/10.1117/12.761698>.
42. Bosch, A., Miñán, A., Vescina, C., Degrossi, J., Gatti, B., Montanaro, P., Messina, M., Franco, M., Vay, C., Schmitt, J., et al. (2008). Fourier transform infrared spectroscopy for rapid identification of nonfermenting gram-negative bacteria isolated from sputum samples from cystic fibrosis patients. *J. Clin. Microbiol.* 46, 2535–2546. <https://doi.org/10.1128/JCM.02267-07>.
43. Kuhm, A.E., Suter, D., Felleisen, R., and Rau, J. (2009). Identification of *Yersinia enterocolitica* at the species and subspecies levels by Fourier transform infrared spectroscopy. *Appl. Environ. Microbiol.* 75, 5809–5813. <https://doi.org/10.1128/aem.00206-09>.
44. Rebuffo, C.A., Schmitt, J., Wenning, M., von Stetten, F., and Scherer, S. (2006). Reliable and rapid identification of *Listeria monocytogenes* and *Listeria* species by artificial neural network-based Fourier transform infrared spectroscopy. *Appl. Environ. Microbiol.* 72, 994–1000. <https://doi.org/10.1128/aem.72.2.994-1000.2006>.
45. Marley, L., Signolle, J.P., Amiel, C., and Travert, J. (2001). Discrimination, classification, identification of microorganisms using FTIR spectroscopy and chemometrics. *Vib. Spectrosc.* 26, 151–159. [https://doi.org/10.1016/S0924-2031\(01\)00113-8](https://doi.org/10.1016/S0924-2031(01)00113-8).
46. Sandrin, T.R., Goldstein, J.E., and Schumaker, S. (2013). MALDI TOF MS profiling of bacteria at the strain level: a review. *Mass Spectrom. Rev.* 32, 188–217. <https://doi.org/10.1002/mas.21359>.
47. Naumann, D. (2015). Infrared spectroscopy in microbiology. In *Encyclopedia of Analytical Chemistry*, pp. 1–32. <https://doi.org/10.1002/9780470027318.a0117.pub2>.
48. Tartof, S.Y., Solberg, O.D., Manges, A.R., and Riley, L.W. (2005). Analysis of a uropathogenic *Escherichia coli* clonal group by multilocus sequence typing. *J. Clin. Microbiol.* 43, 5860–5864. <https://doi.org/10.1128/jcm.43.12.5860-5864.2005>.
49. Lau, S.H., Reddy, S., Cheesbrough, J., Bolton, F.J., Willshaw, G., Cheasty, T., Fox, A.J., and Upton, M. (2008). Major uropathogenic *Escherichia coli* strain isolated in the northwest of England identified by multilocus sequence typing. *J. Clin. Microbiol.* 46, 1076–1080. <https://doi.org/10.1128/jcm.02065-07>.
50. Das, S., Das, P., and Das, P. (2020). Dynamics and control of multidrug-resistant bacterial infection in hospital with multiple delays. *Commun. Nonlinear Sci. Numer. Simul.* 89, 105279. <https://doi.org/10.1016/j.cnsns.2020.105279>.
51. Monegro, A.F., Muppidi, V., and Regunath, H. (2020). Hospital acquired infections. In *StatPearls* (StatPearls Publishing Copyright © 2020, StatPearls Publishing LLC).
52. Van Belkum, A., Tassios, P.T., Dijkshoorn, L., Haeggman, S., Cookson, B., Fry, N.K., Fussing, V., Green, J., Feil, E., Gerner-Smidt, P., et al. (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin. Microbiol. Infect.* 13 (Suppl 3), 1–46. <https://doi.org/10.1111/j.1469-0691.2007.01786.x>.
53. Fischer, G., Braun, S., Thissen, R., and Dott, W. (2006). FT-IR spectroscopy as a tool for rapid identification and intra-species characterization of airborne filamentous fungi. *J. Microbiol. Methods* 64, 63–77. <https://doi.org/10.1016/j.mimet.2005.04.005>.
54. Tintelnot, K., Haase, G., Seibold, M., Bergmann, F., Staemmler, M., Franz, T., and Naumann, D. (2000). Evaluation of phenotypic markers for selection and identification of *Candida dubliniensis*. *J. Clin. Microbiol.* 38, 1599–1608. <https://doi.org/10.1128/JCM.38.4.1599-1608.2000>.
55. Vatanashenassan, M., Boekhout, T., Mauder, N., Robert, V., Maier, T., Meis, J.F., Berman, J., Then, E., Kostrzewa, M., and Hagen, F. (2020). Evaluation of microsatellite typing, ITS sequencing, AFLP fingerprinting, MALDI-TOF MS, and Fourier-transform infrared spectroscopy analysis of *Candida auris*. *J. Fungi* 6, 146. <https://doi.org/10.3390/jof6030146>.
56. Nandy, S.K., and Srivastava, R.K. (2018). A review on sustainable yeast biotechnological processes and applications. *Microbiol. Res.* 207, 83–90. <https://doi.org/10.1016/j.micres.2017.11.013>.
57. Foschino, R., De Lorenzis, G., Fabrizio, V., Picozzi, C., Imazio, S., Failla, O., and Vigentini, I. (2015). Yeast DNA recovery during the secondary fermentation step of Lombardy sparkling wines produced by Champenoise method. *Eur. Food Res. Technol.* 240, 885–895. <https://doi.org/10.1007/s00217-014-2393-7>.
58. Gasser, R.B., Hu, M., Chilton, N.B., Campbell, B.E., Jex, A.J., Otranto, D., Cafarchia, C., Beveridge, I., and Zhu, X. (2006). Single-strand conformation polymorphism (SSCP) for the analysis of genetic variation. *Nat. Protoc.* 1, 3121–3128. <https://doi.org/10.1038/nprot.2006.485>.
59. Essendoubi, M., Toubas, D., Bouzagou, M., Pinon, J.M., Manfait, M., and Sockalingum, G.D. (2005). Rapid identification of *Candida* species by FT-IR microspectroscopy. *Biochim. Biophys. Acta* 1724, 239–247. <https://doi.org/10.1016/j.bbagen.2005.04.019>.
60. Toubas, D., Essendoubi, M., Adt, I., Pinon, J.M., Manfait, M., and Sockalingum, G.D. (2007). FTIR spectroscopy in medical mycology: applications to the differentiation and typing of *Candida*. *Anal. Bioanal. Chem.* 387, 1729–1737. <https://doi.org/10.1007/s00216-006-1005-1>.
61. Wenning, M., and Scherer, S. (2013). Identification of microorganisms by FTIR spectroscopy: perspectives and limitations of the method. *Appl. Microbiol. Biotechnol.* 97, 7111–7120. <https://doi.org/10.1007/s00253-013-5087-3>.
62. Croxatto, A., Prod'homme, G., and Greub, G. (2012). Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *FEMS Microbiol. Rev.* 36, 380–407. <https://doi.org/10.1111/j.1574-6976.2011.00298>.