frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 12 April 2023
DOI 10.3389/fmicb.2023.1154620

OPEN ACCESS

EDITED BY
Chih-Chieh Chen,
National Sun Yat-sen University,
Taiwan

REVIEWED BY
Yu-Wei Wu,
Taipei Medical University,
Taiwan
Manyun Yang,
Harvard University,
United States

*CORRESPONDENCE
Rafi Ahmad
✉ rafi.ahmad@inn.no

†These authors have contributed equally to this work
‡These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 30 January 2023
ACCEPTED 16 March 2023
PUBLISHED 12 April 2023

CITATION
Ahmad A, Hettiarachchi R, Khezri A, Singh Ahluwalia B, Wadduwage DN and Ahmad R (2023) Highly sensitive quantitative phase microscopy and deep learning aided with whole genome sequencing for rapid detection of infection and antimicrobial resistance.
Front. Microbiol. 14:1154620.
doi: 10.3389/fmicb.2023.1154620

Highly sensitive quantitative phase microscopy and deep learning aided with whole genome sequencing for rapid detection of infection and antimicrobial resistance

Azeem Ahmad[1], Ramith Hettiarachchi[2,3†], Abdolrahman Khezri[4†], Balpreet Singh Ahluwalia[1,5‡], Dushan N. Wadduwage[3‡] and Rafi Ahmad[4,6*‡]

[1]Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway, [2]Department of Electronic and Telecommunication Engineering, University of Moratuwa, Moratuwa, Sri Lanka, [3]Center for Advanced Imaging, Faculty of Arts and Sciences, Harvard University, Cambridge, MA, United States, [4]Department of Biotechnology, Inland Norway University of Applied Sciences, Hamar, Norway, [5]Department of Clinical Science, Intervention and Technology, Karolinska Insitute, Stockholm, Sweden, [6]Institute of Clinical Medicine, Faculty of Health Sciences, UiT—The Arctic University of Norway, Tromsø, Norway

Current state-of-the-art infection and antimicrobial resistance (AMR) diagnostics are based on culture-based methods with a detection time of 48–96h. Therefore, it is essential to develop novel methods that can do real-time diagnoses. Here, we demonstrate that the complimentary use of label-free optical assay with whole-genome sequencing (WGS) can enable rapid diagnosis of infection and AMR. Our assay is based on microscopy methods exploiting label-free, highly sensitive quantitative phase microscopy (QPM) followed by deep convolutional neural networks-based classification. The workflow was benchmarked on 21 clinical isolates from four WHO priority pathogens that were antibiotic susceptibility tested, and their AMR profile was determined by WGS. The proposed optical assay was in good agreement with the WGS characterization. Accurate classification based on the gram staining (100% recall for gram-negative and 83.4% for gram-positive), species (98.6%), and resistant/susceptible type (96.4%), as well as at the individual strain level (100% sensitivity in predicting 19 out of the 21 strains, with an overall accuracy of 95.45%). The results from this initial proof-of-concept study demonstrate the potential of the QPM assay as a rapid and first-stage tool for species, strain-level classification, and the presence or absence of AMR, which WGS can follow up for confirmation. Overall, a combined workflow with QPM and WGS complemented with deep learning data analyses could, in the future, be transformative for detecting and identifying pathogens and characterization of the AMR profile and antibiotic susceptibility.

KEYWORDS

quantitative phase microscopy, whole genome sequencing, machine learning, antibiotic resistance, deep learning, rapid diagnosis, bacterial infection

1. Introduction

Antimicrobial resistance (AMR) is the ability of microorganisms to resist antimicrobial treatments, especially antibiotics. Infections due to antibiotic-resistant bacteria are a threat to modern healthcare. A recent meta-analysis of resistant bacteria burden on human health and well-being has revealed that in 2019 alone, 1.27 million deaths were caused directly by antibiotic-resistant bacteria (ARBs), and 4.95 million deaths were associated with ARBs (Murray et al., 2022). This number has surpassed HIV and malaria. Many (73%) of these deaths are due to infections caused by *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii,* and *Pseudomonas aeruginosa* (Murray et al., 2022), which the WHO identifies as critical and high-priority pathogen groups.

Research has shown that adverse patient outcome directly correlates with the time taken to administer optimal antimicrobial (Kumar et al., 2006). Mortality risk doubles with a 24-h delay in providing appropriate antibiotics in cases of bacteremia (Fraser et al., 2006). Globally only half of the antibiotics are prescribed correctly (Milani et al., 2019). Thus, rapid point-of-care diagnostic tests are a central part of the solution to this problem. Current culture-based methods used to detect and identify agents of infection are inadequate. Incubation times of up to 24–48 h are often necessary to capture the majority of culturable bacteria associated with infection (Taxt et al., 2020). Additional time is required for pathogen identifications (2–4 h) and in case of expected AMR for antibiotic susceptibility testing (AST; 18–24 h). Thus, the time interval from collecting patient samples at the ward until the information is available on antibiotic susceptibility patterns is, in the best case, 2–4 days in the clinical routine (Köser et al., 2012).

There are emerging micro-and nanotechnologies for bacterial identification and AST, including both phenotypic (e.g., microfluidic-based bacterial culture) and molecular (e.g., multiplex PCR, hybridization probes, nanoparticles, synthetic biology, and mass spectrometry) methods (Li et al., 2017). PCR and MS methods used on positive cultures to identify microbes have progressed considerably but still are far from ideal. With PCR, one must decide beforehand what to look for, and MALDI-TOF is expensive (Bacconi et al., 2014; Taxt et al., 2020).

Whole genome sequencing (WGS) can overcome some of these problems since there is no need for targeted primers/probes. Also, with the rise of real-time sequencing and its affordability, WGS has become a potent alternative to time-consuming culture-dependent traditional methods. We have recently demonstrated that using Oxford Nanopore Technologies (ONT) MinION and Flongle sequencing platforms, infectious agents and their resistance profile can be identified within 10 min – 1 h after the start of the sequencing (Harstad et al., 2018; Taxt et al., 2020; Avershina et al., 2022). Additionally, we have used data from direct sequencing of spiked blood cultures for genotype-to-phenotype prediction of resistance toward B-lactams in *E. coli* and *K. pneumoniae*, as fast as 1–8 h from the sequencing start (Avershina et al., 2021). However, around 3–4 h are still required to prepare the sample for sequencing.

Direct identification of the pathogen in biofluid samples will mitigate the above issues. Culture-independent diagnostic tests can also detect dead bacteria when antibiotic therapy has been administered before sampling, which could help detect the pathogen (Bourke et al., 2015). Real-time sequencing, for example, has been successfully implemented in lower respiratory tract infections, urine infections, cerebral spinal fluid, surgical site infections, and orthopedic devices with up to 100% sensitivity and specificity in the pathogen detection (Wang et al., 2020; Whittle et al., 2022). Label-free optical techniques such as quantitative phase microscopy (Kim et al., 2022) and Raman spectroscopy (Ho et al., 2019; Lister et al., 2022) have recently been shown to measure phenotypic and molecular signatures in pathogens at very low concentrations. Oh et al. (2020) utilized optical diffraction tomography to quantitatively analyze the response of *Escherichia coli* and *Bacillus subtilis* to varying concentrations of ampicillin (Oh et al., 2020). Recently, Kim et al. (2022) proposed a microscopy-based framework that employs three-dimensional quantitative phase imaging and artificial neural networks to detect pathogens from a limited number of cells, successfully identifying 19 bacterial species known to cause bloodstream infections (Kim et al., 2022). We have recently reported the multi-excitation Raman spectroscopy (ME-RS) method for the species, resistance, and strain-level classification of pathogens (Lister et al., 2022). In combination with downstream machine learning-based analysis, such methods may open doors to developing new culture-free identification of bacteria strains and their clinically relevant properties, such as wild-type (WT) and non-wild type (NWT) bacteria and provide insight into the AMR profile.

Quantitative phase microscopy (QPM) is a powerful technique due to its non-contact, non-invasive, label-free, and quantitative nature. This makes QPM a suitable candidate in the domain of biomedical imaging applications. QPM can measure various morphological (surface area, volume, sphericity, etc.), bio-physical (dry mass, growth rate, etc.), and statistical measures of morphological (mean, standard deviation, skewness, kurtosis, etc.) parameters related to cells/tissues at nanometric sensitivity (Popescu, 2011; Bhaduri et al., 2014; Dubey et al., 2019; Ahmad et al., 2020). QPM encodes information about the specimens' optical thickness (refractive index × geometrical thickness) in terms of modulated intensity patterns called interferograms. Various optical configurations of QPM exist, such as diffraction phase microscopy (Bhaduri et al., 2014), spatial light interference microscopy (Wang et al., 2011), Linnik interference microscopy (Dubois et al., 2002; Ahmad et al., 2021, 2022), Mirau interference microscopy (Ahmad et al., 2015, 2016, 2019), and Mach-Zehnder interference microscopy (Loehrer et al., 2014).

The spatial phase sensitivity of the QPM system is the most crucial parameter which decides the minimal detectable change in the cell's parameters either as a function of their growth or different categories such as normal and challenged. The phase sensitivity of QPM depends on the type of light source (white light, LEDs, and laser) used to illuminate the specimens for the generation of raw data, i.e., interferograms. These light sources either generate highly sensitive phase images at the cost of reduced temporal resolution (for white light and LEDs) or provide high temporal resolution at the expense of less phase sensitivity (for laser; Ahmad et al., 2021). The pros and cons of these light sources in QPM have been discussed in great detail in Ahmad et al. (2021) and Ahmad et al. (2022). Recently, a temporally high and spatially low coherent light source, also known as a pseudo-thermal light source (PTLS), has been proposed to resolve the issues associated with using the aforementioned light sources in QPM (Choi et al., 2011, 2018; Ahmad et al., 2015, 2016, 2021; Ankit et al., 2020).

The use of PTLS presents an opportunity to distinguish between gram-positive and gram-negative bacteria.

Gram-positive bacterial cell walls typically have a thickness ranging from 20 to 80 nm, while gram-negative bacterial cell walls are generally thinner, ranging from 1.5 to 10 nm (Mai-Prochnow et al., 2016). To distinguish between the structural differences in gram-positive and gram-negative bacteria, a highly sensitive quantitative phase microscopy (QPM) system is needed. PTLS provides an extremely high spatial phase sensitivity of approximately 3.4 mrad, which corresponds to a height measurement sensitivity of 2.05 nm (Ahmad et al., 2021). With this level of sensitivity, it would be possible to pick the differences in cell wall thickness between gram-positive and gram-negative bacteria. It is important to note that QPM only provides the integrated thickness of the cells, which includes the cell membrane, cytoplasm, ribosomes, and nucleoid region. However, our proposed QPM system's high sensitivity has the potential to detect and record the structural changes in the cell wall of gram-negative and gram-positive bacteria as measured in integrated phase maps.

Despite the apparent advantages of QPM, it is still challenging to differentiate the large number of classes of biological samples based on the manually quantified cell parameters mentioned in the previous paragraph. Therefore, the extracted phase images of different categories of the bio-specimens and WGS can be utilized for their classification by employing a powerful computational approach, e.g., machine learning or deep learning. With the larger amount of data generated from QPM and WGS, robust data analysis and correlation methods must be needed to generate clinically relevant results.

Deep learning methods have revolutionized all scientific fields, including bio-imaging, genomics, and computer vision. Deep convolutional neural networks (DCNNs) outperformed all classical methods for vision tasks, such as image segmentation, semantic segmentation, and classification. DCNNs are also increasingly being adopted in microscopy, including in QPM (Jo et al., 2017). Jo et al. (2017) proposed a DCNN to classify quantitative phase images. Their results show that deep learning over conventional machine learning (ML) techniques can improve the classification performance between five species of Bacillus (*B. anthracis, B. thuringiensis, B. cereus, B. atrophaeus,* and *B. subtilis*). Although the network identified *B. anthracis* with high sensitivity (~80.8%), the sensitivity ranged between 13.6 and 70.3% for other species. Kim et al. proposed a CNN-based framework that uses 3D refractive index tomograms to identify 19 types of bacteria and properties such as gram staining and aerobicity (Kim et al., 2022). The authors observed less confusion between bacteria species of the same genus. In a study by Wang et al., they designed a computational live bacteria detection system that periodically captures microscopy images of bacterial growth inside an agar plate (Wang et al., 2020). Within 12 h of incubation, their classification network obtained species-specific sensitivity of ~97.2, ~84.0, and ~98.5% for *E. coli*, *K. aerogenes*, and *K. pneumoniae*, respectively.

This study developed a label-free optical assay to identify bacteria and their clinically relevant culture-free properties. Our assay is based on partially spatially coherent high-sensitive QPM followed by DCNN-based classification. First, various challenges associated with the bacteria phase imaging are resolved, such as the effect of defocusing and background aberration removal for error-free phase map generation. This step is essential as any slight defocusing or aberration will add phase ambiguity, influencing the DCNN-based classification and accuracy. The defocusing correction was applied at a single bacterium level for all the images. The QPM experiments are performed on 21 bacteria strains that were previously genotyped and phenotyped. Our results suggest that the proposed assay can classify 19 of the 21 strains with 100% class-wise sensitivity (recall) at extremely low bacteria concentrations (127 images). Furthermore, at this bacteria concentration level (127 images), the classifier achieved an overall accuracy of 95.45%. For the authenticity of the classification process, we further correlated the interclass similarity between learned QPM features to that of genomic features from WGS data with excellent agreement. The results from this proof-of-concept study show that the developed assay could be tested as a first-pass method in the clinical routine prior to more extensive confirmatory testing with current routine diagnostic methods.

2. Materials and methods

2.1. Bacterial strains and phenotypic assessment

The wild-type/susceptive and phenotypically resistant/non-wild-type strains of Gram-negative *Acinetobacter baumannii* ($n = 6$), *Escherichia coli* ($n = 7$), *Klebsiella pneumoniae* ($n = 5$), and Gram-positive *Staphylococcus aureus* ($n = 2$), *Bacillus subtilis* ($n = 1$) were used for this study. Phenotypic minimum inhibitory concentration (MIC) information on the *E. coli* and *K. pneumoniae* strains was taken from the reference strain collection or has been previously published (Avershina et al., 2021). MIC values of *Acinetobacter* strains were described in Karah et al. (2011), and MIC values for *S. aureus* strains were obtained from (https://www.ccug.se/strain?id=35600 and https://bacdive.dsmz.de/strain/14451). Isolates were classified as susceptible and resistant according to the European Committee on Antimicrobial Susceptibility Testing (EUCAST) Breakpoints v 12.0 (December 2021).

2.2. Sequencing and bacterial genome assembly

The *Acinetobacter* isolates libraries were prepared with the Nextera XT DNA Library preparation kit from Illumina (Cat. No.: FC-131-1096) and were sequenced with the NextSeq 550 instrument (Illumina, United States), using PE mode and a mid-output flow cell (150 cycles). The sequencing reads were first quality-checked using FastQC (v 0.11.8 for Linux; Andrews, 2010). Adapters were removed, and low-quality reads were filtered out using Trimmomatic v2.0.29 (Bolger et al., 2014) integrated with OmicsBox (v 2.2.4 for Linux; Bioinformatics, 2019) with default parameters (window size of 4, required quality threshold of 15, average quality threshold of 25, and minimum length threshold of 36 bp). The WGS data for *E. coli* CCUG 17620, *E. coli* NCTC 13441, *E. coli* K-12, and *S. aureus* NCTC 8325 strains were downloaded from NCBI (BioSample: SAMN0993043, SAMEA2432036, SAMN02604091, and SAMN02604235). Sequencing data for the remaining samples were taken from the European nucleotide archive accession number(s) PRJEB60525 and PRJEB60478 (Taxt et al., 2020; Avershina et al., 2021).

According to our previous data, Unicycler and Flye assemblers were identified as the best short and long-read assemblers (Khezri

et al., 2021a). Therefore, in this study, genome assembly was performed using Unicycler (v0.4.9) and Flye (v2.8.2) for short and long reads, respectively (Wick et al., 2017; Kolmogorov et al., 2019). The assemblies were quality checked using QUAST (v4.6.0 for Linux; Gurevich et al., 2013), in addition to BUSCO (v 5.4.4 on galaxy bioinformatic platform; Simão et al., 2015; Galaxy Community, 2022). On BUSCO, enterobacterales lineage was selected for both *E. coli* and *K. pneumoniae* isolates. While for *S. aureus* and *A. baumannii*, bacillales and pseudomonadales were selected as lineage, respectively. Later, assembled genomes were annotated using Prokka (v1.14.5 for Linux; Seemann, 2014). The whole-genome alignment was performed, and the phylogeny tree was constructed using the CLC Genomics Workbench version 21.0.3 (QIAGEN; QIAGEN, 2022). Using the whole-genome alignment toolbox in CLC, the average nucleotide identity (ANI) and alignment percentage (AP) were obtained for paired comparison between samples. Later the identity score matrix from ANI was used for phylogenic tree construction using the neighbor-joining approach.

2.3. Identification of plasmids and antibacterial resistance genes

FASTA files of *De novo* assembled genomes were used to identify plasmids and antimicrobial resistance genes (AMR). To identify plasmids, the PlasmidFinder online tool (software version: 2.0.1, database version: 2020-07-13; Carattoli et al., 2014), with a minimum identity of 95% and coverage of 60%, was utilized for all isolates, except *A. baumannii*, as PlasmidFinder does not contain information regarding its plasmid replicons. For *A. baumannii*, the plasmids were identified using BLAST search against the PLSDB database (Galata et al., 2019).

For *E. coli*, *K. pneumoniae,* and *S. aureus* isolates, antibacterial resistance genes associated with mobile elements were identified using Resfinder online tool (v 4.1, software version: 2020-10-21, database version: 2020-12-01; Bortolaia et al., 2020). The PointFinder online tool (software version: 2020-10-21, database version: 2019-07-02; Zankari et al., 2017) was used to predict antibacterial resistance genes associated with chromosomal point mutation. The AMR genes in *A. baumannii* isolates were identified using the Resistance Gene Identifier tool in CARD v3.1.4, update version 2021-10-05 (Alcock et al., 2020). Only hits flagged as perfect were considered true resistance genes.

2.4. Culture and fixation of bacterial cells for microscopy

The bacterial strains were cultured from frozen aliquots (−80°C) on blood agar plates, and the cultures were used after 20 h of incubation at 37°C. After overnight culture, the $OD_{600\,nm}$ was measured for each strain, and the cells were collected and re-suspended in 1 mL PBS to reduce the presence of agar components within the samples. The cells were fixed with a ratio of 1:2 2% paraformaldehyde in phosphate sodium buffer (PFA in PBS, pH 7.15) and incubated at room temperature for 1 h. The cells were washed twice and re-suspended in 500 µL PBS and transferred to sterile tubes for shipping for QPM analysis.

2.5. Bacteria immobilization for quantitative phase imaging

Bacteria immobilization is a crucial step for their high-resolution quantitative phase imaging. High-resolution phase recovery requires multiple phase-shifted interferograms; therefore, the bacteria cells should not move during the data acquisition. Due to their small size, bacteria have Brownian motion and quickly change their position and orientation. Therefore, bacteria immobilization is needed so that reliable data can be recorded.

As the Linnik interference microscopy configuration works in the reflection mode, therefore, the bacteria samples are prepared on a reflecting substrate (Si-wafer). The Si-wafer substrate is first incubated with 0.01% poly-L-lysine (PLL) solution for 15–20 min. The excessive PLL solution is removed, and the substrate is washed thoroughly with phosphate-buffered saline (PBS). It formed a thin layer of PLL on the substrate and made it positively charged. The polydimethylsiloxane (PDMS) chamber of 10 mm × 10 mm size with 150 µm thickness is placed on the Si substrate. The bacteria samples were seeded into the PDMS chamber and left for approximately 30 min for their incubation. The electrostatic attraction between the negatively charged bacteria and positively charged PLL solution helped them to adhere to the Si substrate (Allison et al., 2011).

The adherence of some species/strains of the bacteria to the Si substrate was observed to be a little challenging. Different strains of *E. coli* and *K. pneumoniae* were attached straightforwardly on the Si substrate after following the aforementioned protocol. In the case of *S. aureus*, *A. baumannii*, and *Bacillus subtilis*, the attachment was challenging. Several factors, such as the chemical constituents of liquid, morphological features of bacteria, and the incubation time, could affect the immobilization process (Allison et al., 2011). For these classes, the Si substrate is seeded with concentrated sample volume and incubated for an extended period, around 1 h. The sample is gently washed off with PBS to remove mobile bacteria cells. The substrate is then left only with the bacteria cells attached to the substrate. PBS is added to the sample and sealed with # 1.5 cover glass from the top, which enabled the use of a water immersion objective lens for imaging and avoided any air current in the specimen.

2.6. Experimental scheme of quantitative phase microscopy

The experimental scheme of the QPM system is illustrated in Figure 1, which is utilized to acquire phase-shifted interferometric data of the bacterial samples. The working principle of the QPM system is based on a high-resolution Linnik interference microscopy configuration. In the QPM system, a temporally high and spatially low coherent light source, also called a pseudo-thermal light source (PTLS), is utilized due to various advantages over conventional light sources like laser, halogen lamps, and LEDs light sources. PTLS enables high-resolution phase recovery of biological specimens with high spatial phase sensitivity. PTLS is generated when a high temporal coherent laser beam is passed through a rotating diffuser. The output of the rotating diffuser acts as a temporally high and spatially low coherent light source and enables speckle noise-free and a coherent noise-free phase recovery (Ahmad et al., 2021).

In our experimental scheme, a laser light beam at 660 nm wavelength (Cobolt Flamenco) was passed through a rotating diffuser,

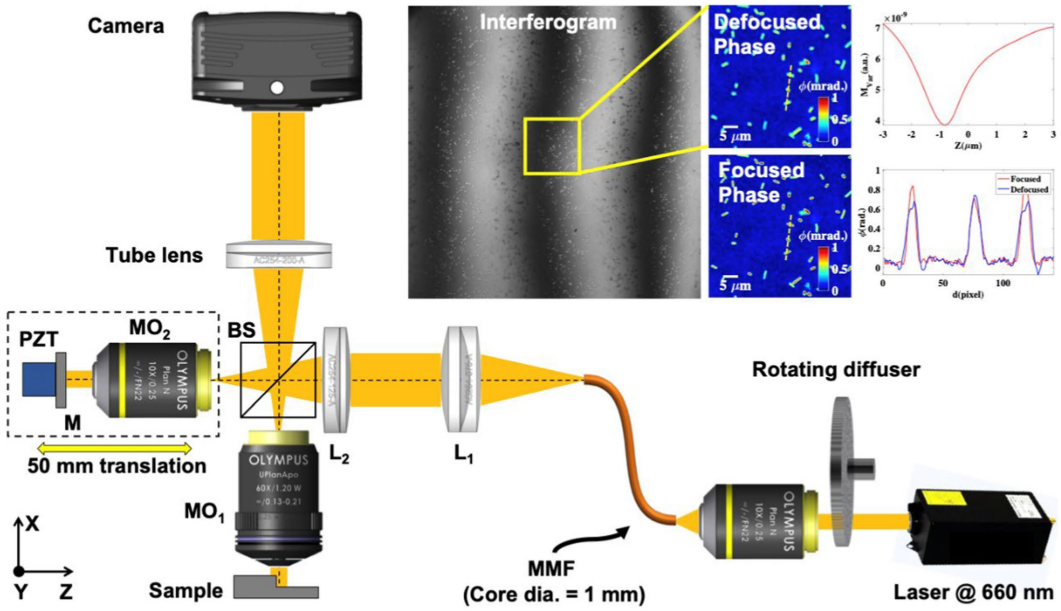

FIGURE 1
Optical schematic diagram of Linnik interferometer-based QPM setup. MO1-2, microscope objectives; BS1-2, beam splitters; L1-2, achromatic doublet lenses; MMF, multi-mode fiber; M, mirror; and PZT, piezo electric transducer. The inset represents the interferogram, reconstructed defocused phase map, numerically focused phase map, and amplitude variance plot of the numerically propagated optical field as a function of the propagation distance. The line profiles of the bacteria phase map before and after focus correction along the yellow dotted lines.

and the output of the diffuser was coupled into a multi-mode fiber (MMF) of a core diameter of 1 mm using an objective lens 10×/0.25NA. The output of MMF was first nearly collimated using lens L_1 and then focused using lens L_2 at the back focal plane of the object arm objective lens (MO_1: 60×/1.2NA) to achieve uniform illumination at the specimen. The focused light beam was split into the object and the reference beam using beam splitter BS. The reference beam was passed through a reference arm objective lens (MO_2: 10×/0.25NA) and reflected from the mirror. The light beam is again collected by the same objective lens and called a reference beam. The objective lens in another arm captured the bacteria sample information, which was recombined with the reference beam at BS and projected at the camera sensor using a tube lens (TL) to form the interferograms. The interferograms were further captured by employing a scientific Hamamatsu CMOS camera. The reference mirror was attached to a nanometric precision piezo stage to introduce a phase shift between the interfering beams to acquire multiple phases shifted interferograms of the specimen for phase recovery. A homemade LabVIEW software program was written and utilized for the required phase stepping in the interferograms and their acquisition using a camera. The reference mirror was also attached to a kinematic mirror mount to control the angle between the object and the reference beam. The total acquisition time of 5–6 phase-shifted interferograms was approximately 600 ms, which can be improved using a high-end computer.

2.7. Quantitative phase microscopy of bacteria and defocus correction

For phase recovery of the sample, multiple phases shifted frames are recorded to achieve diffraction-limited phase recovery. One of the phase-shifted interferograms is depicted in Figure 2A. It can be seen that interferograms suffer from higher-order phase aberration and would influence the recovered phase maps. Multiple phase-shifted interferograms are utilized to recover the complex field information of the bacteria sample using the principal component analysis (PCA) algorithm (Vargas et al., 2011). The full field of view (FOV) amplitude and phase part of the recovered complex field are illustrated in Figures 2B,D. The zoomed views of the regions marked with yellow color boxes in the amplitude and phase parts are depicted in Figures 2C,E, respectively. It can be clearly visualized that the recovered phase map looks defocused, which affected the shape of the bacteria significantly. Therefore, it becomes necessary to perform numerical defocus correction steps to achieve faithfully reconstructed phase maps. The defocus correction algorithm steps are given in the Supplementary material. Figure 2F presents the numerically focused phase image. After following the numerical defocus correction steps, the bacteria phase image looks sharply focused. The amount of defocus is found to be equal to 1 µm by observing the sharpness curve. In addition, defocusing of the bacteria significantly affected the morphological parameters such as area, maximum phase value, volume, etc. Thus, this is a crucial step for the accurate classification of different classes of bacteria using deep learning. Otherwise, there would be a possibility of misclassification of different classes of bacteria samples.

2.8. Data preprocessing and deep learning model architecture

Before employing a deep learning-based classifier, we ensure that there are no images with multiple bacteria co-located together, appearing as a single object, after the segmentation procedure described in Section 3.3. To exclude such images, we employ a heuristic-based approach that considers the area of the object and its

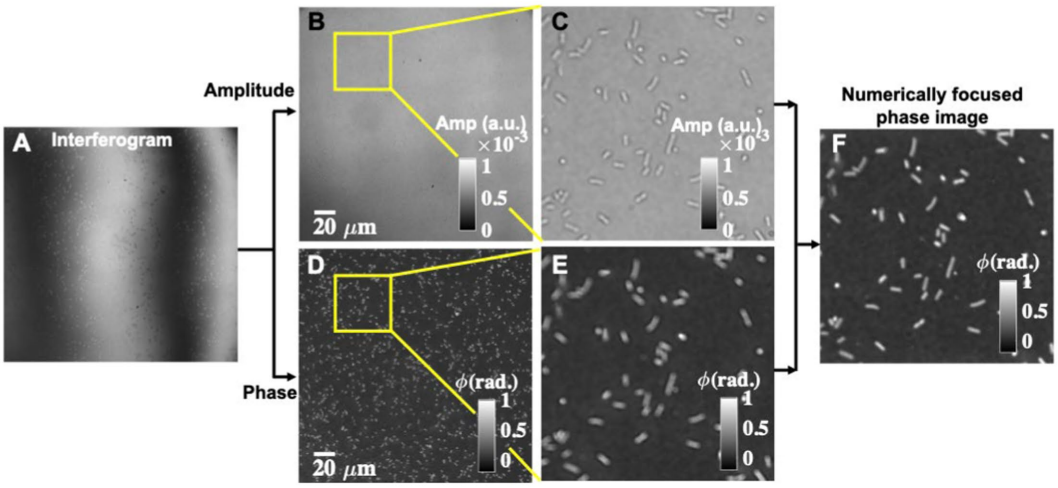

FIGURE 2
Reconstruction of numerically focused phase images using interferometric measurements from the QPM microscope. **(A)** One of the phase-shifted interferograms. **(B,D)** Corresponding amplitude and phase map of the bacteria sample. **(C,E)** Zoomed views of the amplitude and the phase images of the regions marked with yellow color boxes. **(F)** Numerically focused phase map.

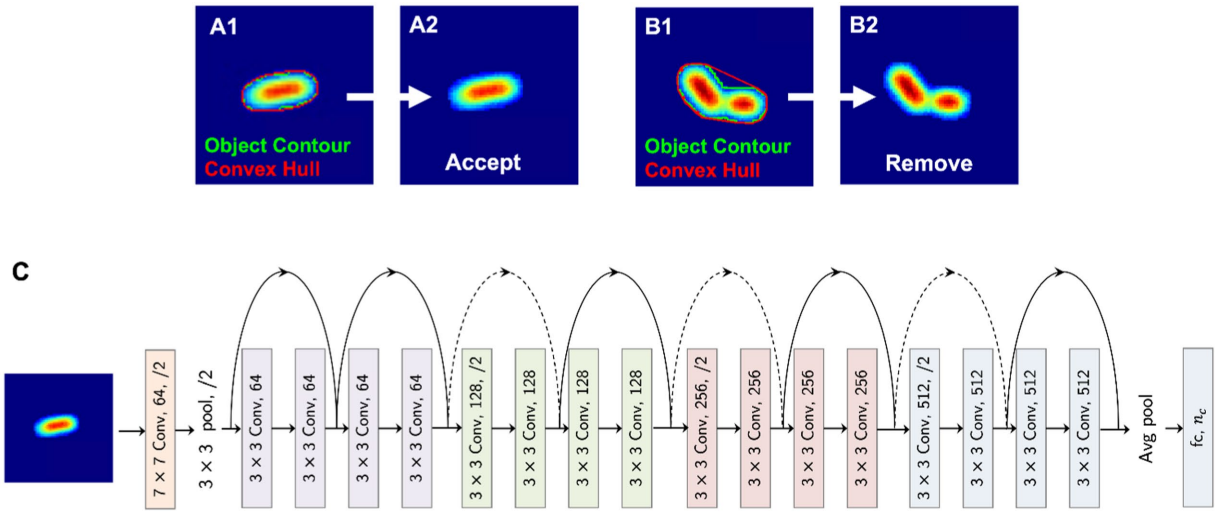

FIGURE 3
Deep neural network pipeline. **(A,B)** A heuristic for detecting co-located bacteria. **(C)** Neural network architecture.

convex hull. As illustrated in Figures 3A,B, we obtain the object contour and the convex hull. By considering the area of these two polygons, we derive a normalized score as,

$$\frac{A_{\text{Boundary}} / A_{\text{Convex Hull}}}{\left(A_{\text{Convex Hull}} - A_{\text{Boundary}}\right)} * 100$$

Based on this score and a selected threshold for the dataset, we remove images that have co-located bacteria (e.g., Figure 3B). After this filtering process, 472,795 images were selected, resized to 224×224, and then split according to the ratios 80, 10, and 10% for training, validation, and testing, respectively. Stratified sampling ensures that identical class distributions are preserved in these three data splits.

We employ the deep residual learning framework introduced by He et al. (2016), which can be used to overcome the performance degradation problem that arises when training deeper networks. The 18-layer residual network (Figure 3C), which was pre-trained on the ImageNet dataset, is used as the initialization for the weights and is fine-tuned on our dataset of bacteria images. The model weights are optimized using the stochastic gradient descent optimizer with a learning rate of 0.01 and a momentum factor of 0.9. The overall model architecture is shown in Figure 3C. Based on our classification requirement (binary vs. multiclass classification), the number of output neurons (n_c) is defined. For both antibiotic resistance prediction and gram stain classification, $n_c = 2$, while for species-level classification and strain classification, the number of output

neurons is 5 and 21 based on the number of output classes, respectively.

3. Results

3.1. Bacterial genome assembly and bioinformatic analyses

Data from whole-genome alignment indicate an average intraspecies nucleotide similarity of 97.4, 93, 99.2, and 96.7% between isolates belonging to *E. coli*, *A. baumannii*, *K. pneumoniae,* and *S. aureus* species (Figure 4). As expected, the only interspecies similarity was between *E. coli* and *K. pneumoniae* isolates (*ca.* 84%). Furthermore, phylogeny analyses showed that all isolates from those four species were clustered together (Supplementary Figure 1). So, although the number of isolates is not large (21 isolates), the selected isolates have a broad molecular level diversity. Genome assembly quality analysis showed comparable results between isolates (Supplementary Table S1).

A wide variety of ARGs conferring resistance toward penicillin, third generation cephalosporins, carbapenems, tetracyclines, and aminoglycosides was identified in the 16 resistant strains (Table 1). All isolates possessed single or multiple plasmids except *A. baumannii* K48-42 and *E. coli* CCUG 17620. Putative plasmids in *E. coli* and *K. pneumoniae* isolates were previously validated using a plasmid-specific assembly approach (Khezri et al., 2021b). In *A. baumannii* K48-48, *E. coli* K-12, *S. aureus* CCUG 35600, and *S. aureus* NCTC 8325 (Table 1).

3.2. Quantitative phase imaging of bacteria samples

To perform quantitative phase microscopy of bacterial samples, the sample is placed under the QPM setup shown in Figure 1 for interferometric recording. The details of the bacteria sample preparation scheme, quantitative phase recovery, and defocus correction steps are provided in the Material and Methods section. The sample is placed on an XY motorized translation stage to acquire phase images of multiple FOVs using a high-resolution water immersion objective lens 60×/1.2NA. More than 50 different FOVs are acquired for each bacterial class to generate large data sets. The number of bacteria present in one FOV was approximately equal to or greater than 300. This way, more than 15,000 bacteria phase images of each class are generated for deep learning training. The recovered phase maps of different types of 21 strains of bacterial samples are exhibited in Figure 5.

3.3. Defocus correction and segmentation at a single bacteria level

It is observed that sample defocus significantly affects the recovered phase profiles of the bacteria samples (Figure 1 in the section Materials and methods); therefore, it needs to be corrected before using bacteria phase images to train deep learning networks. The defocus correction is implemented on a wide FOV. However, there might be different amounts of defocus in different regions of the recovered complex field over the entire FOV of the QPM system due to the presence of higher-order aberrations. The other source of error could be the slight defocus in the raw interferometric data itself due to the inability of the user to judge the best focus during acquisition correctly. Therefore, it is necessary to perform defocus correction at a single bacteria level to avoid any chances of error in the recovered phase maps.

Figure 6 illustrates the steps of the input phase data generation of the bacteria samples for deep learning network training. Firstly, the number of bacteria present in the entire FOV recovered phase maps are counted. Next, the positions of all the bacteria present in the entire FOV are stored and utilized to crop them. The sizes of the cropped images are not kept constant. It is decided by measuring the number of pixels covered by the bacteria cells along both rows and columns using the MATLAB *regionprops* command. The crop area is then decided by increasing 20 pixels along all four sides from the extreme coordinates of the bacteria. This helps to avoid any unwanted cropping of the bacteria region. Before cropping the bacteria phase images, the selection of the coordinates corresponding to each bacterium is made in a way such that the cropped images are of square shape by keeping the number of rows and columns equal. All the bacteria are kept at the center of the cropped phase image for further processing steps, as depicted in Figure 6A.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
E.coli_K12	1		0.00	0.00	0.00	0.00	0.00	0.00	0.00	97.14	97.09	97.11	97.07	99.65	96.74	84.24	84.04	84.02	84.04	84.02	0.00
S.aureus_NCTC_8325	2	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	99.78
Acinetobacter_INN_WT	3	5.20E-3	0.14		84.21	84.13	83.93	84.43	84.62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Acinetobacter_K12_21_301	4	3.17E-3	4.46E-4	8.83		97.85	99.51	97.80	98.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Acinetobacter_K48_42_304	5	3.25E-3	4.54E-4	9.24	85.27		97.90	99.47	97.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Acinetobacter_K55_13_305	6	3.26E-3	4.55E-4	8.73	88.97	85.10		97.81	98.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Acinetobacter_K57-06_307	7	3.26E-3	4.55E-4	9.68	83.88	91.15	84.78		97.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Acinetobacter_K71-71_311	8	3.18E-3	4.44E-4	9.71	84.20	84.81	85.63	85.06		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E. coli_NCTC_13441	9	79.59	0.00	4.86E-3	0.02	0.14	0.02	1.41E-3	0.02		98.59	99.76	98.59	97.04	98.32	84.32	84.05	84.03	85.21	84.12	0.00
E.coli_101	10	81.16	0.00	0.07	0.04	0.06	0.03	0.16	0.14	83.68		98.66	99.03	96.97	98.65	84.43	84.11	84.10	85.02	84.12	0.00
E.coli_102	11	80.71	0.00	4.97E-3	0.05	0.13	9.97E-3	0.12	0.20	92.77	86.38		98.70	97.05	98.38	84.32	84.09	84.08	85.10	84.25	0.00
E.coli_104	12	80.88	0.00	0.07	0.04	0.06	0.03	0.16	0.14	83.69	89.87	85.85		96.93	98.61	84.43	84.14	84.15	84.97	84.13	0.00
E.coli_A2-39	13	83.10	0.00	4.72E-3	0.12	0.24	0.03	0.11	0.24	73.05	75.10	75.05	74.34		96.65	84.51	84.24	84.22	84.54	84.39	0.00
E.coli_CCUG_17620	14	79.75	0.00	3.67E-3	1.34E-3	1.89E-3	1.38E-3	1.38E-3	1.34E-3	83.09	88.23	84.64	85.96	72.87		83.98	83.79	83.81	84.29	83.79	0.00
K.pneumoniae_210	15	16.36	1.73E-4	0.13	0.09	0.12	0.09	0.27	0.24	15.79	16.25	15.97	16.03	15.46	15.14		99.13	99.16	99.27	99.17	0.00
K.pneumoniae_211	16	20.54	0.00	0.08	0.10	0.10	0.12	0.23	0.26	18.83	19.74	19.43	19.38	18.36	18.21	66.19		99.96	99.75	99.23	0.00
K.pneumoniae_212	17	20.72	0.00	0.06	0.06	0.08	0.08	0.22	0.23	19.02	19.87	19.58	19.46	18.69	18.27	65.27	93.76		99.73	99.26	0.00
K.pneumoniae_240	18	19.87	0.00	0.08	0.15	0.18	0.10	0.09	0.16	19.84	20.51	20.28	20.36	18.15	18.35	63.47	88.64	85.56		99.16	0.00
K.pneumoniae_A2-23	19	19.91	0.00	0.08	0.21	0.31	0.20	0.31	0.38	18.45	19.13	19.08	18.90	18.22	17.63	65.70	86.41	83.83	86.76		0.00
S.aureus_CCUG_35600	20	0.00	93.67	0.18	2.84E-5	2.90E-5	2.91E-5	2.91E-5	2.83E-5	0.00	0.00	0.00	0.00	0.00	0.00	1.67E-4	0.00	0.00	0.00	0.00	

FIGURE 4
Whole-genome alignment for the 20 sequenced isolates. The upper and lower comparison gradient shows the percentage of nucleotide similarity and alignment percentage between genomes, respectively.

10.3389/fmicb.2023.1154620

TABLE 1 Overview of bacterial strains' phenotype and genomic background.

Isolate	Gram staining	Phenotype based on AST (antibiotic) wild type (WT)/Non-wild type (NWT)	Plasmid presence	Antibiotic resistance genes
Acinetobacter baumannii K12-21*	Negative	Resistant (ciprofloxacin, Imipenam, gentamicin, and meropenem)—NWT	Yes	OXA (20, 58, 66), sul1, ade (S, L, G, I, K, A, C), Aba (F, Q), abeS, ADC-185, APH(3′)-Ia, and qacEdelta1
Acinetobacter baumannii K48-42*	Negative	Resistant (amikacin, ciprofloxacin, Imipenam, gentamicin, and meropenem)—NWT	No	OXA 69, sul 1, ade (F, G, S, I, K, R), AbaQ, AmvA, catI, mphE, msrE, and cat I
Acinetobacter baumannii K55-13*	Negative	Resistant (ciprofloxacin, gentamicin, and meropenem)—NWT	Yes	OXA (66, 72), sul2, ADC-30, aph(3″)-Ib, aph(6)-Id, tet(B), ant(3″)-IIa, aac(3)-Ia, and aac(6′)-Ip
Acinetobacter baumannii K57-06*	Negative	Resistant (amikacin, ciprofloxacin, Imipenam, gentamicin, and meropenem)—NWT	Yes	OXA (23, 69), sul2, ade (F, G, I, K, R), abeS, APH(3′)-Ia, AbaQ, AmvA, mphE, msrE
Acinetobacter baumannii K71-71*	Negative	Resistant (amikacin, ciprofloxacin, Imipenam, gentamicin, and meropenem)	Yes	OXA (23, 51), sul (1, 2), TEM-1, ade (I, K, L), abeS, APH(3′)-Vla, armA, arr-2, cmlA5, mphE, msrE, PER-7
Acinetobacter (INN)*	Negative	WT	Yes	blaZ, fusB, blaADC-8, blaI, and fosD
Escherichia coli 101	Negative	Resistant (ciprofloxacin)—NWT	Yes	mdf(A), sitABCD
Escherichia coli 102	Negative	Resistant (ampicillin)—NWT	Yes	blaTEM-1B, sul1, mdf(A), mph(A), aac(3)-IId, aadA2, tet(A), dfrA12, qacE, and sitABCD
Escherichia coli 104	Negative	Resistant (ampicillin, ciprofloxacin, gentamicin) – NWT	Yes	mdf(A), sitABCD
Escherichia coli A2-39	Negative	Resistant (cefotaxime, ceftazidime)—NWT	Yes	blaTEM-1B, blaCTX-M-2, sul1, aadA1, dfrA1, tet(A), aac(3)-*VIa*, mdf(A), and qacE
Escherichia coli NCTC 13441	Negative	Resistant (cefotaxime, ceftazidime, and Fluoroquinolone)—NWT	Yes	blaTEM-1B, blaCTX-M-(15, 182), sul1, aadA5, catB3, dfrA17, tet(A), mph(A), qacE, sitABCD, and aac(6′)-Ib-cr
Escherichia coli CCUG 17620	Negative	Susceptible—WT	Yes	No antibiotic resistance genes were found.
Escherichia coli K-12	Negative	Susceptible—WT	No	No antibiotic resistance genes were found.
Klebsiella pneumoniae 210	Negative	Resistant (ciprofloxacin) – NWT	Yes	OqxB, fosA. fosA6
Klebsiella pneumoniae 211	Negative	Resistant (ciprofloxacin, cefotaxime)—NWT	Yes	blaTEM-1B, blaSHV-(61, 155), Oqx (A, B), qnrS1, fosA, floR, aac(3)-IIa, ompK37, and acrR
Klebsiella pneumoniae 212	Negative	Resistant (ciprofloxacin, gentamicin, and cefotaxime)—NWT	Yes	blaTEM-1B, blaSHV-12, aac(3)-IIa, fosA, floR, Oqx (A, B), and acrR
Klebsiella pneumoniae 240	Negative	Resistant (cefotaxime, ceftazidime)—NWT	Yes	blaTEM-1B, blaSHV-(28, 106), aadA1, qacE, Oqx (A, B), sul1, dfrA1, fosA, aac(3)-IIa, ramR, ompK37, and acrR
Klebsiella pneumoniae A2-23	Negative	Resistant (cefotaxime, ceftazidime)—NWT	Yes	blaTEM-1B, blaCTX-M-15, blaSHV-(9, 12, 13, 31, 129, 155, 172), aac(3)-IId, aac(6′)-(Ib3, Ib-cr), aadA1, aph(3′)-Ia, aph(3″)-Ib, aph(6)-Id, catA1, dfrA30, fosA, qacE, Oqx (A, B), sul2, ompK37, and acrR
Staphylococcus aureus CCUG 35600	Positive	Resistant (methicillin, tetracycline, clindamycin, and erythromycin)—NWT	No	blaZ, mecA, erm(A), ant(9)-Ia, tet(K), and fusA
Staphylococcus aureus NCTC 8325	Positive	Susceptible – WT	No	fosB

Phenotype and genotype information from antibiotic susceptibility testing (AST) and whole genome sequencing, respectively. **Acinetobacter* has intrinsic resistance to several antibiotics (Abbott et al., 2013).

Next, the amount of defocus in each cropped bacterium complex field is measured for the correction of defocus from the recovered phase images, if any. The numerically defocus corrected phase images of some of the bacteria samples are presented in Figure 6B. The defocus amount is slightly different in all cropped bacteria phase images. It can be clearly visualized that defocus in the raw interferometric data significantly affected the recovered phase maps of the bacteria cells. Further, numerically focused phase images are utilized to generate corresponding binary images, as illustrated in Figure 6C. To generate the binary images, the thresholding of the phase images is done by

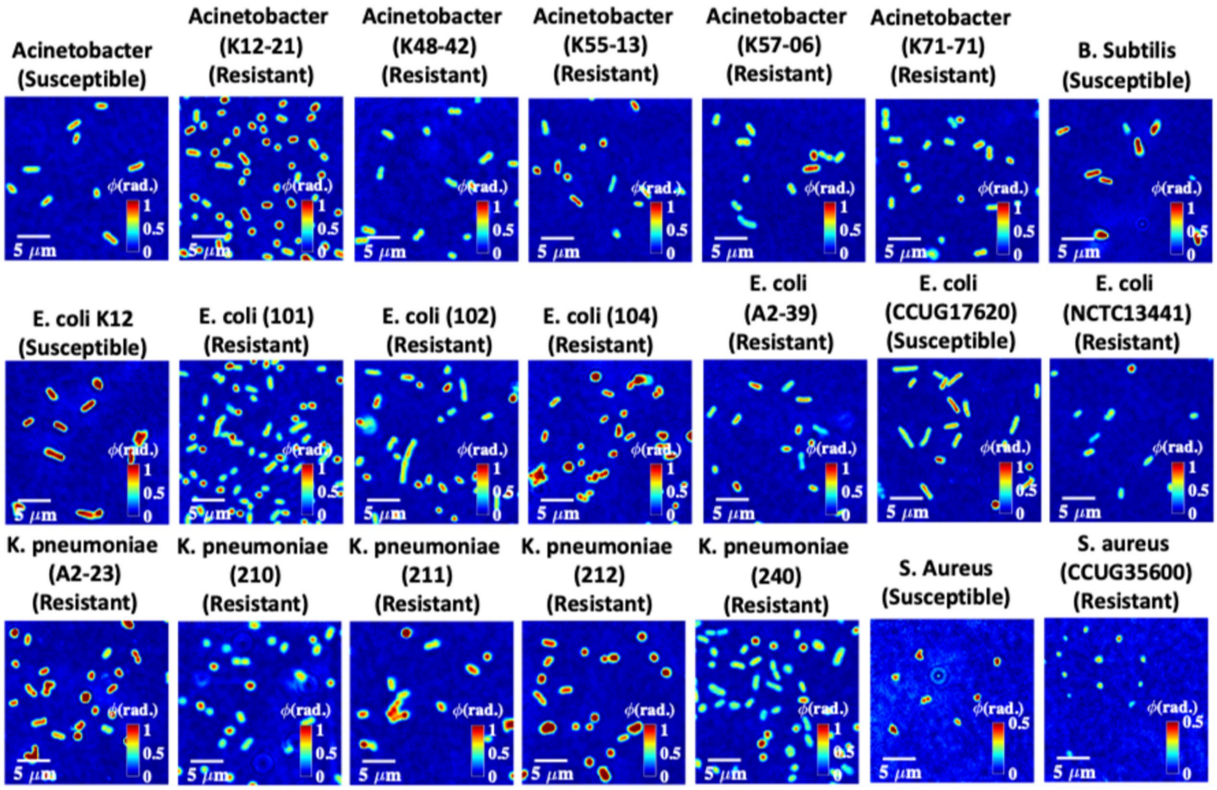

FIGURE 5
Quantitative phase maps of 21 different bacteria strains. The color bars for the bacteria samples are from 0 to 1rad. Note that the color bars for *Staphylococcus aureus* strains are from 0 to 0.5rad.

measuring the system's peak-to-valley phase sensitivity. The threshold phase value is approximately 1.2–1.4 times the peak-to-valley phase error of the system for the bacteria segmentation. This threshold value is found to be an optimum value for the segmentation as the maximum phase values of most of the bacteria are approximately greater than 3–4 times the peak-to-valley phase error. The segmented areas of an image are further dilated by two pixels along all sides to remove any chances of unwanted bacteria phase map cropping. Some of the binary images have more than one isolated bacterium. The binary images are then improved by removing the regions of other bacteria using MATLAB, as shown in Figure 6D. The generated binary image is then multiplied by the recovered phase image to remove the background regions from the phase images, as illustrated in Figure 6E.

3.4. Deep learning for the classification of bacteria samples at the colony and bacteria level

3.4.1. Experiments and evaluation criteria

The model architecture described in Section 2.8 is adopted and trained to measure the performance on classification tasks: (1) Antibiotic resistant/sensitive prediction, (2) Gram stain classification, (3) Species-level classification, and (4) Strain classification. A separate classifier was trained for eight epochs for each task. The trained model measures the performance of bacteria samples from a blind test set. We varied the number of isolated bacteria images (in the form of Figure 3A2) in the sample as $N = 1, 3, 7, 15, 31$, and 63 (also 127 for the strain level classifier) to gauge the minimum number of bacteria that should be imaged for each classification task (please refer to Supplementary Table S2 for more details about the blind test set class distribution). Each image of bacteria was classified independently using the classifier, and individual predictions on these images were aggregated to obtain the prediction per sample (at each N).

We first analyze the antibiotic resistant/sensitive and gram-stain prediction, which are binary classification tasks. Figure 7A shows the results of antibiotic resistance prediction. The positive (resistant class) prediction achieved a recall score of 100% (i.e., the ratio between true positive predictions for a class out of total instances from that class) at $N = 15$, while the non-resistant prediction recall was 96.45% at $N = 31$. In the gram stain prediction task (Figure 7B), the Gram-negative class achieved 100% recall at $N = 7$, while the Gram-positive prediction was 83.43% at $N = 31$.

Second, we present the results for species-level classification (Figure 7C). The model predicted *B. subtilis, K. pneumoniae,* and *S. aureus* classes with 100% sensitivity at $N = 31$. For *A. baumannii,* the sensitivity was 0.957 at $N = 63$. For *E. coli,* the sensitivity was 0.975 at $N = 63$. The confusion matrix for species-level classification at $N = 63$ is shown in Figure 7D. Detailed classification metrics at $N = 63$ are included in Supplementary Tables S2–S5 in the Supplementary material.

Next, Figures 8A,B show the strain level classification results. The model achieved 100% sensitivity in predicting 19 out of the 21 strains at $N = 127$ (Figure 8A). Considering challenging strains, the

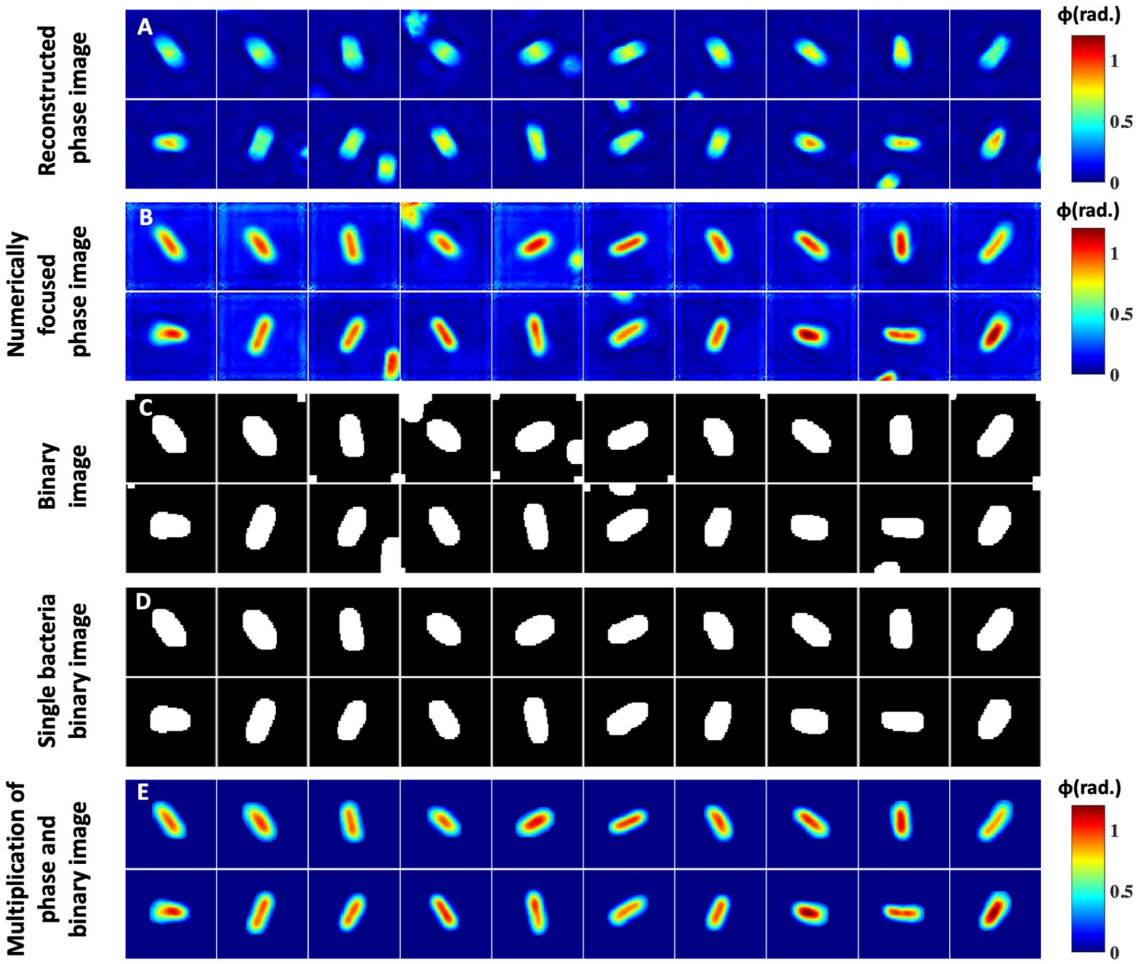

FIGURE 6
Segmentation and postprocessing pipeline for individual bacteria images. **(A)** Reconstructed phase maps of the individual bacteria cells. **(B,C)** Numerically focused phase images and corresponding binary maps. Some phase images have multiple bacteria at the edges, which need to be removed. **(D)** Represents a single bacteria binary mask. **(E)** Multiplication of the numerically focused phase maps and single bacteria binary masks.

classification sensitivity increased with increasing the *N* for *E. coli* 102 and *E. coli* A2-39. At $N = 127$, the sensitivity of *E. coli 102* and *E. coli A2-39* reached 100 and 73%, respectively. However, the classification sensitivity of *A. baumannii* K55-13 remained consistently below 35% with increasing *N*. Figure 8B shows the confusion matrix for strain level classification at $N = 63$. As seen at $N = 63$, a little under 30% of *E. coli* 102 samples were misclassified as *E. coli* CCUG17620. Also, *ca.* 35% of *E. coli* A2-39 samples were misclassified as *E. coli* NCTC13441. Finally, almost 70% of *A. baumannii* K55-13 samples were misclassified as *E. coli* A2-39.

Finally, we investigated the probability of a QPM image from each strain being classified as one of the 21 strains as a similarity measure of QPM features. For each test image from a particular class, the probability of prediction was extracted and averaged over all images from that class. Figure 9B shows the QPM similarity matrix. Figure 9A illustrates the equivalent similarity measure from the genomic analysis; here, the average nucleotide identity (ANI) was used. Shown in red boxes are the similarities between the strains from each species. As expected, ANI was more than 90% similar within the species. Some strains from *Acinetobacter* (K48-42, K55-13, K57-06, and K71-71), *E. coli* (101, 102, 104, and CCUG17620), *K. pneumoniae* (A2-23, 211, and 212), and *S. aureus* (NCTC 8325 and CCUG35600) showed some QPM similarity. Nevertheless, no distinct patterns were observed between the ANI and QPM similarities.

Supplementary Figure 2 shows the training and validation loss curves of the four models trained for each classification task. These models are trained to minimize the cross-entropy loss.

4. Discussion

A significant challenge for clinicians managing infection is to precisely identify those patients who should receive antibiotics (and which ones) and those who should not. This is due to physicians not being able to diagnose patients accurately in real-time, leading to either prescription of antibiotics for viral infections or a prescription of broad-spectrum drugs that should ideally be kept in reserve. We recently established an effective method for direct sequencing for pathogen identification and antibiotic resistance detection (Ahmadi et al., 2023). The current state-of-the-art diagnostic of infection is mainly based on biochemical analysis of the culture of the clinical material. This leads to a detection time of 48–96 h or even longer.

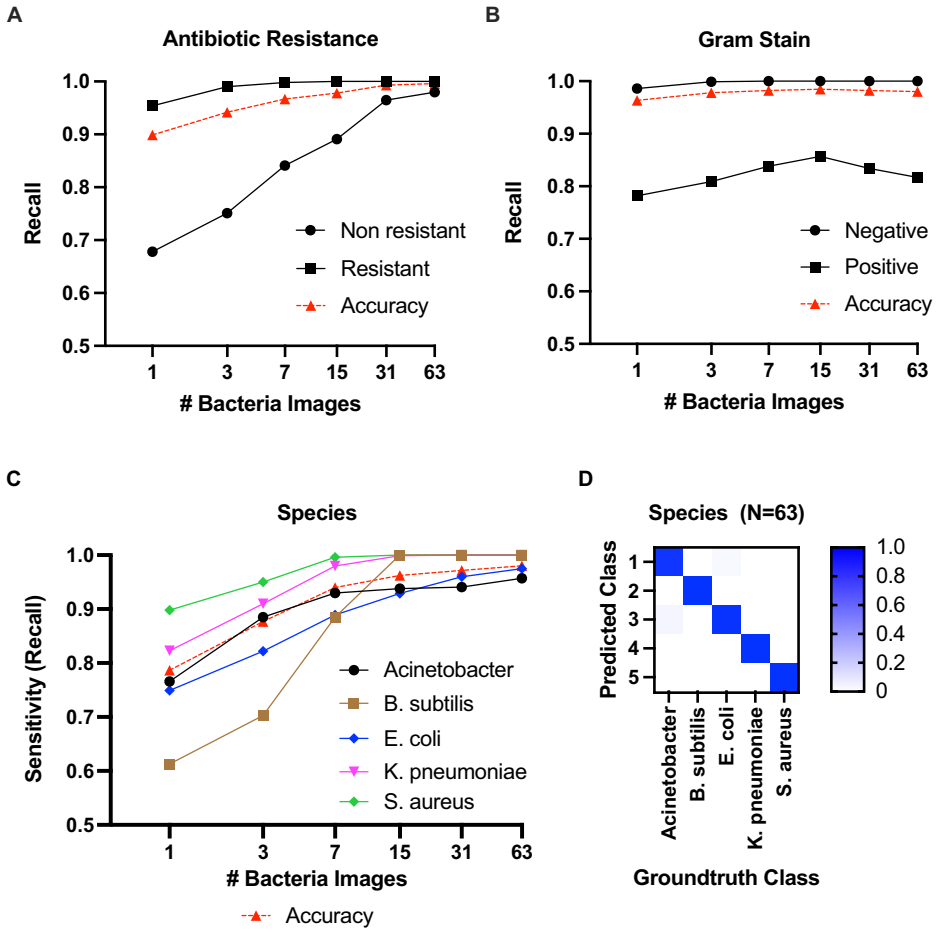

FIGURE 7
Classification results against the number of bacteria images. **(A)** Classification recall scores for antibiotic resistance. **(B)** Classification recall scores for gram +/−ve. **(C)** Species-level classification sensitivity score. **(D)** Confusion matrix for species-level classification with 63 bacteria images in the batch.

In this study, we imaged isolates from 21 unique bacterial strains from five different species, including four WHO priority pathogens (*E. coli*, *A. baumannii*, *K. pneumoniae,* and *S. aureus*) using a PTLS based high-sensitive quantitative phase microscope (QPM). PTLS brings many advantages, such as high-sensitivity, high spatial and temporal resolution, and the use of a high-resolution objective lens in QPM. However, owing to the small size of bacteria and, more importantly, only a minimal change in the phase value between bacterium of different species and strains, it is imperative to pick the minute differences in the phase value. Consequently, post-imaging, several phase errors were removed at the individual bacterium level, such as higher-order phase aberration and defocusing effect on the phase maps of the bacteria samples. These are crucial steps and need to be adopted before using the QPM images of bacteria samples in deep learning, which may mislead its classification outcome. Using a deep neural network, we then classified the QPM images of the individual bacterium at strain and species levels and according to their antibacterial resistance and gram-staining. WGS data (from Illumina and MinION sequencing platforms) and AST data were used to generate the ground truth (strain, species, and resistance profile) to train the DNN.

Our results showed high classification accuracy for antibiotic-resistant/sensitive gram staining and species identification. The network could also classify 19 out of the 21 strains when at least 127 bacterial images were analyzed. One of the misclassifications was of *E. coli* A2-39 samples which were misclassified as *E. coli* NCTC13441, and the other was *A. baumannii* K55-13 samples which were misclassified as *E. coli* A2-39. Among these, the misclassification of *A. baumannii* as *E. coli* is clinically misleading as these are two different species, which would lead to the prescription of the incorrect antibiotic.

Although cell culture still is a gold standard for bacterial identification, results presented here show the potential of the current approach as a new attempt at bacterial identification with minimum dependencies on cell culturing (as a time-consuming approach). It is observed that out of two strains that are either weekly classified or not classified, for one strain, the results may be improved by analyzing more bacteria; nevertheless, for the other strains, the network was incapable of correct classification. The two misclassified classes on our training set were due to their similarity in QPM images, demonstrating the limitations of deep-learning-based methods with limited image information. In the future, to improve the accuracy, it would be useful to apply multi-spectral QPM images to investigate if misclassified bacteria could yield spectral fingerprints in the QPM images. Alternatively, the QPM method could further be complemented with the Raman spectroscopy methods for combining morphological and molecule fingerprinting.

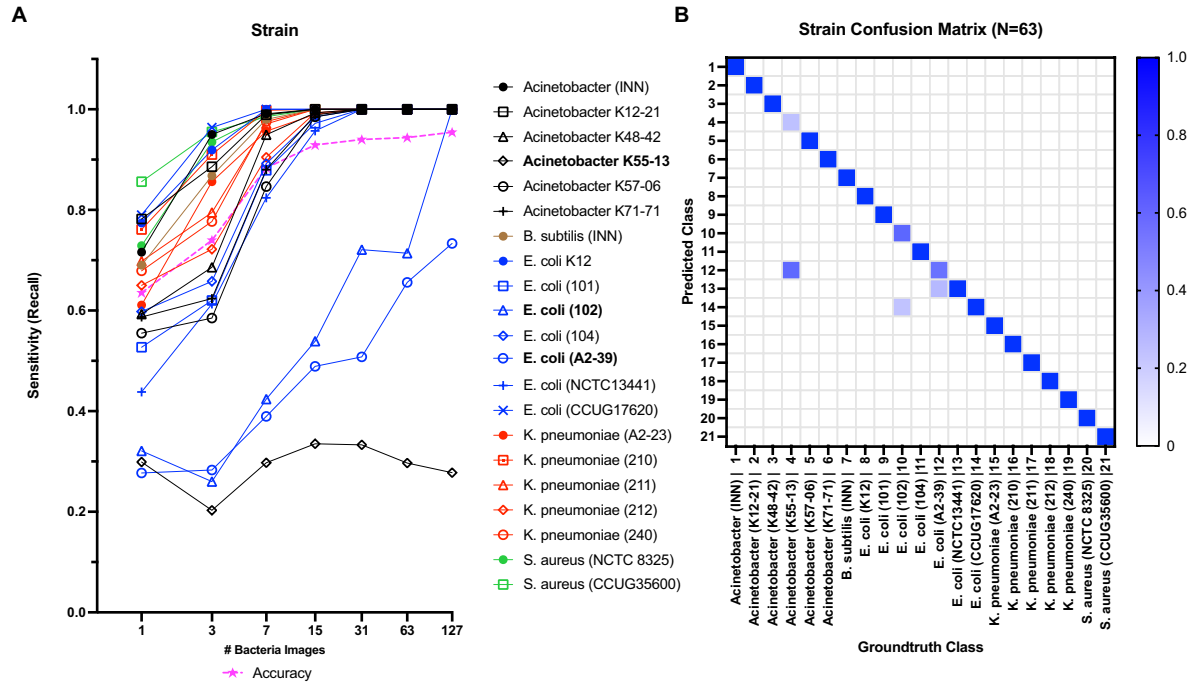

FIGURE 8
(A) Strain level classification sensitivity scores against the number of bacterial QPM images **(B)** Confusion matrix for strain-level classification with 63 bacterial QPM images in the batch. The WGS data provide the ground truth class information.

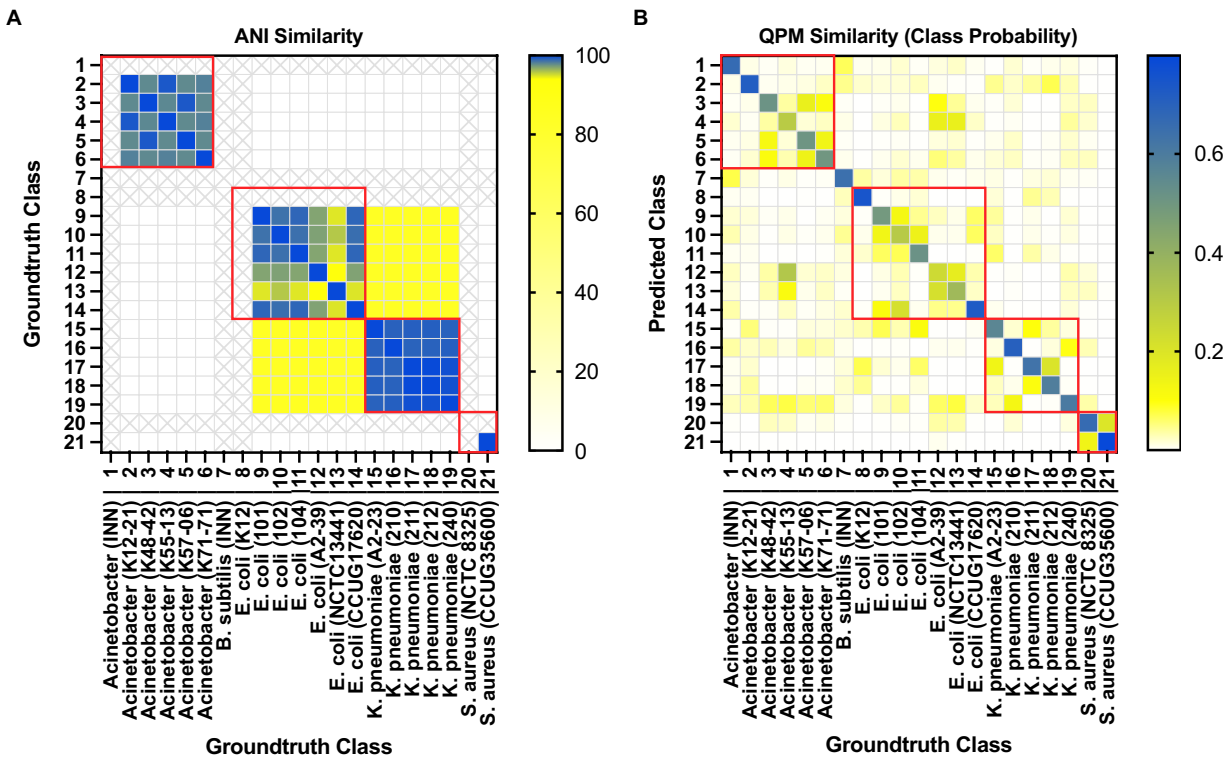

FIGURE 9
(A) Average nucleotide identity (ANI) based similarity between bacteria classes. Classes representing isolates that were not sequenced are shown in "X"s. **(B)** The equivalent QPM-based similarity between bacteria classes. Here the probability of strain level class predictions (averaged over all test instances for each class) was used as the similarity measure.

Our results suggest that differential morphological features from different bacteria strains can be measured using our approach, and they can be used in downstream machine learning pipelines for detection. These findings demonstrate the utility of using QPM and DNN to identify a range of WHO priority pathogens and provide relevant information about microbiological samples, which WGS or conventional microbiological methods can later verify. Our results establish the potential of QPM as a rapid first-stage analytical tool that can complement WGS for phenotype prediction and resistome analysis. Such a workflow can be hugely impactful in handling and preventing the spread of AMR and could lead to future use in clinical microbiology.

In this proof-of-concept study, we have used a combination of enabling technologies using genomics, QPM, and deep learning to rapidly detect pathogens and their resistance profile. Moreover, in addition to the species-level classification, which is frequently studied in the literature, we have analyzed the classification performance under multiple levels of classification tasks, including the presence or absence of resistance, gram staining, and strain-level classification, which has been followed up by WGS data and AST for ground truth information to validate the results. Taken together, all this information is potentially of high clinical importance and can be utilized as a heuristic approach compared to the current state-of-the-art routine solutions. However, to increase the reliability of the method and for any potential future use in clinical microbiology, we need to increase the number of bacterial strains analyzed, and future work should include additional clinically relevant pathogens and AMR profiles. Moreover, bacterial morphological characteristics are affected by the growth environments. It is important that analysis in mixed bacterial samples and in infected tissues or biological fluids rather than isolates grown in rich culture media would provide the reliability of the described approach for potential use for real-time diagnosis.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the epository/repositories and accession number(s) can be found below: https://www.ebi.ac.uk/ena, PRJEB59337. All the raw images and codes have been uploaded to GitHub repository https://github.com/ramithuh/bacteria-classification.

Author contributions

BS and RA planned the study. AA, RH, AK, DW, BS, and RA designed the experiments. AA, AK, RH, DW, and RA wrote the manuscript. AA performed QPM work in discussions with BS. AK performed the bioinformatics analyses in discussions with RA. RH analyzed the data with machine learning in discussions with DW. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the internal strategic research grant 2020 and 2021 from the Inland Norway University of Applied Sciences. BS and AA acknowledge the funding from the Research Council of Norway, Nano-2021 (288565), and BioTek (285571).

Acknowledgments

The authors would like to thank the Genomics Support Center UiT, Tromsø, for sequencing the *Acinetobacter* samples, Stephan A. Frye and Helene Bouras for performing the bacterial cell fixation, Asal Ahmadi for help with some bioinformatics analysis, and Bilal Afzal for his assistance with imaging.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1154620/full#supplementary-material

References

Abbott, I., Cerqueira, G. M., Bhuiyan, S., and Peleg, A. Y. (2013). Carbapenem resistance in Acinetobacter baumannii: laboratory challenges, mechanistic insights and therapeutic strategies. *Expert Rev. Anti-Infect. Ther.* 11, 395–409. doi: 10.1586/eri.13.21

Ahmad, A., Dubey, V., Butola, A., Tinguely, J.-C., Ahluwalia, B. S., and Mehta, D. S. (2020). Sub-nanometer height sensitivity by phase shifting interference microscopy under environmental fluctuations. *Opt. Express* 28, 9340–9358. doi: 10.1364/OE.384259

Ahmad, A., Dubey, V., Jayakumar, N., Habib, A., Butola, A., Nystad, M., et al. (2021). High-throughput spatial sensitive quantitative phase microscopy using low spatial and high temporal coherent illumination. *Sci. Rep.* 11, 1–13. doi: 10.1038/s41598-021-94915-w

Ahmad, A., Dubey, V., Singh, G., Singh, V., and Mehta, D. S. (2016). Quantitative phase imaging of biological cells using spatially low and temporally high coherent light source. *Opt. Lett.* 41, 1554–1557. doi: 10.1364/OL.41.001554

Ahmad, A., Jayakumar, N., and Ahluwalia, B. S. (2022). Demystifying speckle field interference microscopy. *Sci. Rep.* 12:10869. doi: 10.1038/s41598-022-14739-0

Ahmad, A., Kumar, A., Dubey, V., Butola, A., Ahluwalia, B. S., and Mehta, D. S. (2019). Characterization of color cross-talk of CCD detectors and its influence in multispectral quantitative phase imaging. *Opt. Express* 27, 4572–4589. doi: 10.1364/OE.27.004572

Ahmad, A., Srivastava, V., Dubey, V., and Mehta, D. (2015). Ultra-short longitudinal spatial coherence length of laser light with the combined effect of spatial, angular, and temporal diversity. *Appl. Phys. Lett.* 106:093701. doi: 10.1063/1.4913870

Ahmadi, A., Khezri, A., Nørstebø, H., and Ahmad, R. (2023). A culture-, amplification-independent, and rapid method for identification of pathogens and antibiotic resistance profile in bovine mastitis milk. *Front. Microbiol.* 13:1104701. doi: 10.3389/fmicb.2022.1104701

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48, D517–D525. doi: 10.1093/nar/gkz935

Allison, D. P., Sullivan, C. J., Mortensen, N. P., Retterer, S. T., and Doktycz, M. (2011). Bacterial immobilization for imaging by atomic force microscopy. *J. Vis. Exp.* 54:e2880. doi: 10.3791/2880

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data 2010. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ankit, B., Daria, P., Prasad, D. K., Ahmad, A., Anowarul, H., Tinguely, J. C., et al. (2020). High spatially sensitive quantitative phase imaging assisted with deep neural network for classification of human spermatozoa under stressed condition. *Sci. Rep.* 10, 1–12. doi: 10.1038/s41598-020-69857-4

Avershina, E., Frye, S. A., Ali, J., Taxt, A. M., and Ahmad, A. (2022). Ultrafast and cost-effective pathogen identification and resistance gene detection in a clinical setting using Nanopore Flongle sequencing. *Front. Microbiol.* 13:822402. doi: 10.3389/fmicb.2022.822402

Avershina, E., Sharma, P., Taxt, A. M., Singh, H., Frye, S. A., Paul, K., et al. (2021). AMR-Diag: neural network based genotype-to-phenotype prediction of resistance towards β-lactams in Escherichia coli and Klebsiella pneumoniae. *Comput. Struct. Biotechnol. J.* 19, 1896–1906. doi: 10.1016/j.csbj.2021.03.027

Bacconi, A., Richmond, G. S., Baroldi, M. A., Laffler, T. G., Blyn, L. B., Carolan, H. E., et al. (2014). Improved sensitivity for molecular detection of bacterial and Candida infections in blood. *J. Clin. Microbiol.* 52, 3164–3174. doi: 10.1128/JCM.00801-14

Bhaduri, B., Edwards, C., Pham, H., Zhou, R., Nguyen, T. H., Goddard, L. L., et al. (2014). Diffraction phase microscopy: principles and applications in materials and life sciences. *Adv. Opt. Photon.* 6, 57–119. doi: 10.1364/AOP.6.000057

Bioinformatics, B. OmicsBox—Bioinformatics made easy. (2019).

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., et al. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* 75, 3491–3500. doi: 10.1093/jac/dkaa345

Bourke, T. W., McKenna, J. P., Coyle, P. V., Shields, M. D., and Fairley, D. J. (2015). Diagnostic accuracy of loop-mediated isothermal amplification as a near-patient test for meningococcal disease in children: an observational cohort study. *Lancet Infect. Dis.* 15, 552–558. doi: 10.1016/S1473-3099(15)70038-1

Carattoli, A., Zankari, E., García-Fernández, A., Larsen, M. V., Lund, O., Villa, L., et al. (2014). In Silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14

Choi, Y., Hosseini, P., Kang, J. W., Kang, S., Yang, T. D., Hyeon, M. G., et al. (2018). Reflection phase microscopy using spatio-temporal coherence of light. *Optica* 5, 1468–1473. doi: 10.1364/OPTICA.5.001468

Choi, Y., Yang, T. D., Lee, K. J., and Choi, W. (2011). Full-field and single-shot quantitative phase microscopy using dynamic speckle illumination. *Opt. Lett.* 36, 2465–2467. doi: 10.1364/OL.36.002465

Dubey, V., Popova, D., Ahmad, A., Acharya, G., Basnet, P., Mehta, D. S., et al. (2019). Partially spatially coherent digital holographic microscopy and machine learning for quantitative analysis of human spermatozoa under oxidative stress condition. *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-39523-5

Dubois, A., Vabre, L., Boccara, A.-C., and Beaurepaire, E. (2002). High-resolution full-field optical coherence tomography with a Linnik microscope. *Appl. Opt.* 41, 805–812. doi: 10.1364/AO.41.000805

Fraser, A., Paul, M., Almanasreh, N., Tacconelli, E., Frank, U., Cauda, R., et al. (2006). Benefit of appropriate empirical antibiotic treatment: thirty-day mortality and duration of hospital stay. *Am. J. Med.* 119, 970–976. doi: 10.1016/j.amjmed.2006.03.034

Galata, V., Fehlmann, T., Backes, C., and Keller, A. (2019). PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Res.* 47, D195–D202. doi: 10.1093/nar/gky1050

Galaxy Community (2022). The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* 50, W345–W351. doi: 10.1093/nar/gkac247

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Harstad, H, Ahmad, R, and Bredberg, A. Nanopore-based DNA sequencing in clinical microbiology: Preliminary assessment of basic requirements. bioRxiv [Preprint]. (2018). doi: 10.1101/382580

He, K, Zhang, X, Ren, S, and Sun, J, (eds.) "Deep residual learning for image recognition." in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2016).

Ho, C.-S., Jean, N., Hogan, C. A., Blackmon, L., Jeffrey, S. S., Holodniy, M., et al. (2019). Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat. Commun.* 10, 1–8. doi: 10.1038/s41467-019-12898-9

Jo, Y., Park, S., Jung, J., Yoon, J., Joo, H., Kim, M.-h., et al. (2017). Holographic deep learning for rapid optical screening of anthrax spores. *Sci. Adv.* 3:e1700606. doi: 10.1126/sciadv.1700606

Karah, N., Haldorsen, B., Hermansen, N. O., Tveten, Y., Ragnhildstveit, E., Skutlaberg, D. H., et al. (2011). Emergence of OXA-carbapenemase- and 16S rRNA methylase-producing international clones of Acinetobacter baumannii in Norway. *J. Med. Microbiol.* 60, 515–521. doi: 10.1099/jmm.0.028340-0

Khezri, A., Avershina, E., and Ahmad, R. (2021a). Hybrid assembly provides improved resolution of plasmids, antimicrobial resistance genes, and virulence factors in Escherichia coli and Klebsiella pneumoniae clinical isolates. *Microorganisms* 9:2560. doi: 10.3390/microorganisms9122560

Khezri, A., Avershina, E., and Ahmad, R. (2021b). Plasmid identification and plasmid-mediated antimicrobial gene detection in Norwegian isolates. *Microorganisms* 9:52. doi: 10.3390/microorganisms9010052

Kim, G., Ahn, D., Kang, M., Park, J., Ryu, D., Jo, Y., et al. (2022). Rapid species identification of pathogenic bacteria from a minute quantity exploiting three-dimensional quantitative phase imaging and artificial neural network. *Light Sci. Appl.* 11:190. doi: 10.1038/s41377-022-00881-x

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8

Köser, C. U., Ellington, M. J., Cartwright, E. J., Gillespie, S. H., Brown, N. M., Farrington, M., et al. (2012). Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* 8:e1002824. doi: 10.1371/journal.ppat.1002824

Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., et al. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit. Care Med.* 34, 1589–1596. doi: 10.1097/01.CCM.0000217961.75225.E9

Li, Y., Yang, X., and Zhao, W. (2017). Emerging microtechnologies and automated systems for rapid bacterial identification and antibiotic susceptibility testing. *SLAS Technol.* 22, 585–608. doi: 10.1177/2472630317727519

Lister, A, Avershina, E, Ali, J, Devitt, G, Hanrahan, N, Highmore, CJ, et al. Multi-excitation Raman spectroscopy complements whole genome sequencing for rapid detection of bacterial infection and resistance in WHO priority pathogens. bioRxiv [Preprint]. (2022). doi: 10.1101/2022.02.08.479540

Loehrer, M., Botterweck, J., Jahnke, J., Mahlmann, D. M., Gaetgens, J., Oldiges, M., et al. (2014). In vivo assessment by Mach–Zehnder double-beam interferometry of the invasive force exerted by the Asian soybean rust fungus (Phakopsora pachyrhizi). *New Phytol.* 203, 620–631. doi: 10.1111/nph.12784

Mai-Prochnow, A., Clauson, M., Hong, J., and Murphy, A. B. (2016). Gram positive and gram negative bacteria differ in their sensitivity to cold plasma. *Sci. Rep.* 6:38610. doi: 10.1038/srep38610

Milani, R. V., Wilt, J. K., Entwisle, J., Hand, J., Cazabon, P., and Bohan, J. G. (2019). Reducing inappropriate outpatient antibiotic prescribing: normative comparison using unblinded provider reports. *BMJ Open Qual.* 8:e000351. doi: 10.1136/bmjoq-2018-000351

Murray, C. J. L., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., et al. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* 399, 629–655. doi: 10.1016/S0140-6736(21)02724-0

Oh, J., Ryu, J. S., Lee, M., Jung, J., Han, S., Chung, H. J., et al. (2020). Three-dimensional label-free observation of individual bacteria upon antibiotic treatment using optical diffraction tomography. *Biomed. Opt. Express* 11, 1257–1267. doi: 10.1364/BOE.377740

Popescu, G. (2011). *Quantitative Phase Imaging of Cells and Tissues*, McGraw-Hill, New York: McGraw Hill Professional.

QIAGEN (2022). CLC genomics workbench 20.0. Available at: https://digitalinsights.qiagen.com/

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Taxt, A. M., Avershina, E., Frye, S. A., Naseer, U., and Ahmad, R. (2020). Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. *Sci. Rep.* 10, 7622–7611. doi: 10.1038/s41598-020-64616-x

Vargas, J., Quiroga, J. A., and Belenguer, T. (2011). Phase-shifting interferometry based on principal component analysis. *Opt. Lett.* 36, 1326–1328. doi: 10.1364/OL.36.001326

Wang, H., Ceylan Koydemir, H., Qiu, Y., Bai, B., Zhang, Y., Jin, Y., et al. (2020). Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning. *Light Sci. Appl.* 9, 1–17. doi: 10.1038/s41377-020-00358-9

Wang, C.-x., Huang, Z., Fang, W., Zhang, Z., Fang, X., Li, W., et al. (2020). Preliminary assessment of nanopore-based metagenomic sequencing for the diagnosis of prosthetic joint infection. *Int. J. Infect. Dis.* 97, 54–59. doi: 10.1016/j.ijid.2020.05.044

Wang, Z., Millet, L., Mir, M., Ding, H., Unarunotai, S., Rogers, J., et al. (2011). Spatial light interference microscopy (SLIM). *Opt. Express* 19, 1016–1026. doi: 10.1364/OE.19.001016

Whittle, E., Yonkus, J. A., Jeraldo, P., Alva-Ruiz, R., Nelson, H., Kendrick, M. L., et al. (2022). Optimizing Nanopore sequencing for rapid detection of microbial species and antimicrobial resistance in patients at risk of surgical site infections. *mSphere* 7:e00964. doi: 10.1128/msphere.00964-21

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595

Zankari, E., Allesøe, R., Joensen, K. G., Cavaco, L. M., Lund, O., and Aarestrup, F. M. (2017). PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J. Antimicrob. Chemother.* 72, 2764–2768. doi: 10.1093/jac/dkx217