

Microbiological culture methods

Samples for culture were transferred to the laboratory within 30 minutes of collection or processing. Comprehensive cultures were performed for all subjects and specimens, including synovial fluid, periprosthetic tissues, and sonicate fluid. A 0.1 ml aliquot of synovial fluid or concentrated sonicate fluid was directly inoculated onto aerobic and anaerobic sheep blood agar (Becton-Dickinson, Heidelberg, Germany). Aerobic plates were incubated at 37°C in 5% to 7% CO₂ for six days, and anaerobic plates were cultured at 37°C for 14 days. Residual fluids were injected into BacTec Peds Plus/F bottles (Becton-Dickinson) and incubated in a Bactec 9050 instrument (Becton-Dickinson) for six days, and sub-cultured if positive. The periprosthetic tissues were homogenized in brain heart infusion broth (Becton-Dickinson), inoculated onto blood agar, and inoculated/incubated for aerobic and anaerobic culture as indicated for the fluid cultures. Any growth from the synovial fluid samples was considered positive. Growth of 20 or more colony-forming units (CFUs) per plate in sonicate fluid samples was considered positive. A tissue culture was deemed positive when the same organism was isolated from two or more samples. All bacteria isolated were identified using the Vitek 2 system (bioMérieux, Hazelwood, Missouri, USA).

DNA extraction

Bead-beating disruption was used to break the cell wall of coccus and fungi, by placing 1 g of 0.5 mm glass beads (BioSpec, Bartlesville, Oklahoma, USA) mixed with 0.5 ml of specimen into 1.5 ml microcentrifuge tubes (Axygen Scientific, Union City, California, USA). Then, the tube was vortexed vigorously at 2,800 rpm to 3,200 rpm for five minutes. Total genomic DNA was extracted from samples using the TIANamp Micro DNA Kit (DP316; Tiangen Biotech, Beijing, China) according to the manufacturer's instructions. The extracted DNA was then sonicated to generate 200 bp to 300 bp fragments using the Covaris S220 (Covaris, Woburn, Massachusetts, USA).

Microbial genome database

An in-house database was built by BGI. The complete reference genomes of corresponding species were downloaded from the NCBI FTP website.¹ The database has been updated approximately once every six months.

The plasmid sequence and any other non-human-related sequences were removed, keeping only one best genome sequence representing every species based on the assembly results. The alignment index of the reference sequence was built for classifying sequenced reads. The database contains the genomic sequences of 2,700 viruses, 1,494 bacteria, 73 fungi, and 48 parasites that are all related to human diseases.

Table i. An example of the calculation results of the metagenomic next-generation sequencing data.

Variable	Data
Number of total reads	22,341,761
SR	1.12
Genus	<i>Staphylococcus</i>
Species	<i>Staphylococcus aureus</i>
NRSMG	1,730
NRSMS	1,502
SNRSMG	1,549
SNRSMS	1,345
Reference genome length, bp	2,898,306
Coverage rate, %	2.59

NRSMG, number of reads stringently mapped to pathogen in genus-level; NRSMS, number of reads stringently mapped to pathogen in species-level; SNRSMG, standardized number of reads stringently mapped to pathogen in genus-level; SNRSMS, standardized number of reads stringently mapped to pathogen in species-level; SR, standardized ratio.

Bioinformatics pipeline

This process was conducted in the following manner: 1) clean reads of high-quality sequencing data were generated by filtering out the short (< 35 bp), low-quality (Q5 ≤ 0.7), and low-complexity reads; 2) Burrows-Wheeler alignment (BWA)² was utilized in this pipeline for aligner; match sites > 45 nucleotides (nt) and mismatch sites ≤ 2 nt were defined as the mapping quality cut-off points; 3) human host sequences were eliminated by mapping to the human reference genome (hg19); and 4) the remaining sequencing data were aligned to the microbial genome databases, generating four lists separately mapped to viruses, bacteria, fungi, and parasites (Supplementary Figure a). Only bacterial and fungal lists were outputted for further analysis in this study.

Interpretation of metagenomic next-generation sequencing results

The number of raw reads varies among different samples. In order to reach an equalized comparison, the standardized ratio (SR) is defined as the number of total reads/20,000,000. The figure 20,000,000 was set as a target number of reads when sequencing. All the original numbers of reads stringently mapped to pathogen in genus-level (NRSMG) and numbers of reads stringently mapped to pathogen in species-level (NRSMS) were calculated to a standardized number as standardized number of reads stringently mapped to pathogen in genus-level (SNRSMG) and standardized number of reads stringently mapped to pathogen in species-level (SNRSMS). Coverage rate was defined as NRSMS × 50 bp/reference genome length (Supplementary Table i).

Equations illustrating the calculation process of metagenomic next-generation sequencing results are shown below.

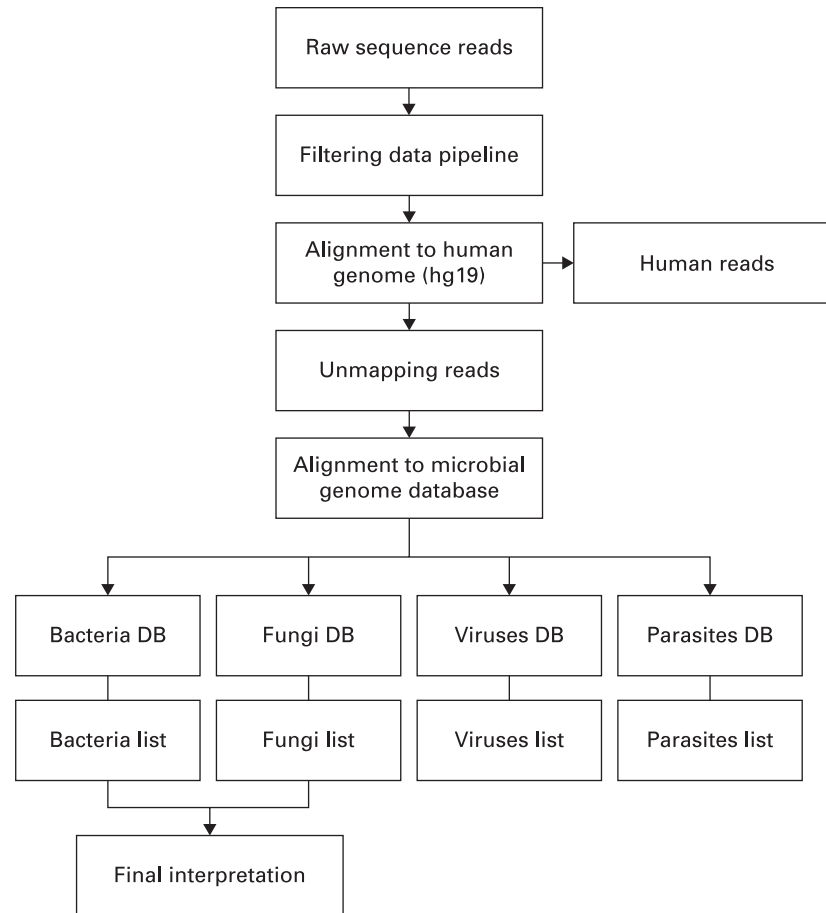


Fig. a

General flowchart of bioinformatic pipeline. DB, database; hg19, human reference genome.

$$SR = \frac{\text{Number of total reads}}{20,000,000}$$

$$SNRSMS = \frac{NRSMS}{SR}$$

$$SNRSMG = \frac{NRSMS}{SR}$$

$$\text{Coverage rate} = \frac{NRSMS \times 50}{\text{Reference genome length}}$$

Optimal thresholds were set up as listed here in order to identify true pathogens: 1) SNRSMG < 3 was considered as insignificant, except for *Mycobacterium tuberculosis* complex (MTC), while any aligned read of MTC was considered positive; 2) *Burkholderia*, *Ralstonia*, and *Delftia* were considered as positive when relative abundance in

genus level $\geq 80\%$, since they were regarded as the most common contamination genera in the lab, and had rarely been cultured or validated by specific polymerase chain reaction (PCR) as pathogens in a microbiology lab; 3) the relative abundance in genus level $\geq 15\%$ was determined as the optimal threshold for bacterial identification, while the relative abundance in genus level was $\geq 30\%$ for fungi, as indicated in the main body; and 4) microbial species whose coverage rate was of the first rank within positive genus and with SNRSMS ≥ 3 was determined as a positive species.

References

1. No authors listed. NCBI Genomes. NCBI. <ftp://ftp.ncbi.nih.gov/genomes/> (date last accessed 27 May 2020).
2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*. 2009;25(14):1754-1760.