Annals of Clinical Microbiology www.acm.or.kr

Review article

Whole-genome sequencing applications for evolution of clinical microbiology

Takashi Takahashi
Laboratory of Infectious Diseases, Graduate School of Infection Control Sciences and Ōmura Satoshi Memorial Institute, Kitasato University, Tokyo, Japan

Abstract

In the present review, we systematically examine the diverse applications of whole-genome sequencing (WGS) and next-generation sequencing (NGS) to elucidate the evolution of clinical microbiology. The review aims to provide novel insight and to improve understanding of the applications of WGS in clinical microbiology laboratories. It is organized into the following sections: (1) the various types of NGS machines; (2) NGS workflows for obtaining genome sequences; (3) comparative genomic analysis; (4) RNA-seq (transcriptome) analysis; (5) genome-based bacterial typing; (6) genome-based antimicrobial resistance (AMR) detection; and (7) identification of integrative and conjugative elements carrying AMR gene(s). Four figures and three tables are provided to illustrate this information. The discussion focuses on WGS applications using several genera of microorganisms (*Streptococcus*, *Enterococcus*, *Staphylococcus*, *Pasteurella*, and *Mycobacterium*). Overall, WGS and related NGS technologies provide innovative clinical microbiology laboratory studies based on high-throughput genomic results for pathogen identification, tracking, and AMR/virulence profiling. In line with the concept of "One Health," human and animal microbiology laboratories should pay careful attention to the drastically dynamic evolution of WGS and related NGS technologies.

Keywords: Microbiology, High-throughput nucleotide sequencing, One health, Whole-genome sequencing

pISSN : 2288-0585
eISSN : 2288-6850

Ann Clin Microbiol 2025 December, 28(4):22
https://doi.org/10.5145/ACM.2025.28.4.3

Correspondence to
Takashi Takahashi
E-mail: taka2si@lisci.kitasato-u.ac.jp

Received: September 16, 2025
Revised: November 03, 2025
Accepted: November 03, 2025

Introduction

Infectious diseases are among the leading causes of human and animal deaths worldwide, especially in developing countries. Bacteria with antimicrobial resistance (AMR) pose a global threat owing to limited antimicrobial treatment regimens. The goal of clinical microbiology laboratories is to provide successful treatment and better host outcomes.

In such laboratories, conventional methods (i.e., culture-based, biochemical, immunological, and molecular procedures) have been widely used for specific pathogen detection. Novel advancements in molecular biology in the 21st century have led to the development of several new diagnostic techniques [1]. Table 1 summarizes the history of molecular diagnostic procedures, descriptions, and their applications: the introduction of "Microarrays," "Metagenomics," and "Metabarcoding" in the 2000s; "Next-generation sequencing (NGS)" and "RNA sequencing (RNA-seq)" in the late 2000s; and "Single-cell sequencing," "Metatranscriptomics," and the "Clustered regularly interspaced short palindromic repeats–Cas9 system" in

recent years. Whole-genome sequencing (WGS) using NGS can overcome the limitations of conventional methods by providing comprehensive genomic data to characterize virulence and AMR features, distinguish closely related strains, and trace outbreak sources, such as in foodborne disease surveillance [2,3]. Table 2 shows the strengths and weaknesses (i.e., principle, applications, testing speed, test sensitivity and specificity, results, and testing costs, including initial and operational charges) of conventional methods. It emphasizes that WGS is a powerful modern tool, while recognizing the practicality and accessibility of traditional methods.

Table 1. Historical timeline of discoveries in molecular diagnostic procedures, with descriptions and applications in the 21st century [1]

Year	Molecular procedures	Description	Applications
2000s	Microarrays	Analysis of gene expression, SNP genotyping, and comparative genomic hybridization	Study of gene expression, detection of genetic variation, and identification of chromosomal abnormalities
2000s	Metagenomics	Comprehensive analysis of entire pathogen populations	Identification of rare and uncultivable pathogens
2000s	Metabarcoding	Comprehensive analysis of pathogen populations based on barcode regions	Identification of rare and uncultivable pathogens
Late 2000s	NGS	Sequencing of entire genomes, transcriptomes, and epigenomes	Identification of genetic variation within and between pathogen populations
Late 2000s	RNA sequencing	Analysis of gene expression and identification of new transcripts	Study of gene regulation and identification of novel genes
Recent years	Single-cell sequencing	Sequencing of individual microbial cells	Analysis of genomic variation at the single-cell level
Recent years	Metatranscriptomics	Analysis of gene expression in pathogen populations	Study of pathogen function and activity in different environments
Recent years	CRISPR-Cas9 system	Targeted genome editing using RNA-guided endonucleases	Study of gene function and development of gene therapeutical approaches

Abbreviations: SNP, single-nucleotide polymorphism; NGS, next-generation sequencing; CRISPR, clustered regularly interspaced short palindromic repeats.

Table 2. Comparison of strengths and weaknesses of traditional approaches and whole-genome sequencing [3]

Item	Traditional approaches	Whole-genome sequencing
Principle	Phenotypic traits, such as culturing, serotyping, biochemical testing, or PCR-based detection	Sequencing the entire genome to identify pathogens and analyze genetic features
Applications	Detection, identification, and enumeration of pathogens	Outbreak tracing, source attribution, evolution study, and functional gene analysis
Testing speed	Time-consuming (days to weeks)	Faster results once sequencing infrastructure is established (hours to days)
Test sensitivity/specificity	Variable and dependent on culture conditions and the detection method applied	High sensitivity/specificity owing to genome analysis
Result output	Qualitative or semi-quantitative results (presence/absence or counts)	Quantitative and comprehensive genetic data (SNPs, resistome, or virulome)
Testing costs (initial and operational costs)	Lower initial and operational costs	High initial cost for WGS equipment; operational costs depend on scale and throughput; these elevated costs may limit some developing countries, or countries with fewer resources, from accessing this technology
Advantages	Cost-effective, well-established, and simple to implement in clinical laboratories	Provision of comprehensive genetic information on antimicrobial resistance and virulence factors
Disadvantages	Limited accuracy in strain differentiation and inability to detect nonculturable organisms. The methods may not detect viable but non-culturable cells or unculturable pathogens	High initial cost requiring advanced infrastructure, expertise, and bioinformatics capabilities, needing high-quality DNA, and generating large datasets that need robust bioinformatics pipelines for analysis

Abbreviations: PCR, polymerase chain reaction; SNP, single-nucleotide polymorphism; WGS, whole-genome sequencing.

Many people, particularly elderly individuals, have companion animals (e.g., dogs and cats) in their homes. Medical hospitals and nursing homes [4,5] have introduced animal-assisted therapy as a mental health service for patients and older individuals. Humans and companion animals are in close contact with their environment. The "One Health" concept [6] is a comprehensive health control strategy for humans, contact animals, and related environments. It states that circulating bacterial communities with virulence factors (VFs) and AMR should be carefully monitored to maintain an environment of total health.

In the present review article, we conducted a systematic search for the diverse applications of WGS using NGS to clarify the evolution of microbiology in human and animal clinical settings in terms of the "One Health" concept. The information described herein may provide novel insight and strengthen the understanding of the applications of WGS for clinical microbiology laboratory personnel.

Various types of NGS machines

Various types of NGS machines are used for long- and short-read sequencing [7]. Oxford Nanopore Technologies (ONTs) offers Minion Nanopore sequencing devices for long-read sequencing. Similarly, PacBio provides Revio and Vega systems for long-read sequencing. Long-read sequencing machines can resolve large single-nucleotide variants and repeat regions, and the ONT series is a portable rapid sequencing kit. However, the sequencing error rate (Phred-type quality score Q20 = error rate 1/100) using ONT devices is not yet at the short-read sequencing error rate level (quality score Q30 = error rate 1/1000). PacBio systems are portable but are also labor-intensive and expensive. In contrast, Illumina provides several Seq systems (e.g., iSeq 100, MiniSeq, MiSeq, MiSeq i100, NextSeq 550, NextSeq 1000 & 2000, and NovaSeq 6000) for short-read sequencing. These systems have high per-base accuracy and account for the majority of currently performed WGS. However, they struggle to resolve large single-nucleotide variants and repeat regions, and their workflows are labor-intensive.

We recently reported the draft genome sequence (accession no. BTGW00000000.1) of a *Streptococcus pyogenes* isolate (*emm*103/sequence type (ST) 1363) from the blood of a woman with peritonitis and streptococcal toxic shock syndrome [8]. Short-read sequencing was performed on a novel DNBSEQ-G400RS platform (MGI-Tech) using DNA Nanoball technology based on circular DNA fragment amplification. DNBSEQ-G400RS and NovaSeq 6000 are equally efficient high-throughput sequencing platforms for investigations that use the "Metabarcoding" method [9]. The main benefit of the DNBSEQ-G400RS is its lower sequencing costs.

NGS workflow for obtaining genome sequences

The NGS workflow for obtaining genome sequences follows two different approaches: culture-independent and culture-dependent. In the culture-independent approach, metagenomic NGS is applied directly from clinical samples, whereas in the culture-dependent approach, WGS is performed on pure culture isolates. Therefore, the two approaches serve fundamentally different diagnostic and epidemiological purposes. Here, we describe the NGS workflow for pure culture isolates to obtain genome sequences for

pathogen identification and epidemiological data [1]. This flow comprises four steps: DNA extraction, library preparation, sequencing, and analysis. The preparation of high-quality DNA samples extracted from infection foci and/or sterile specimens (i.e., blood, cerebrospinal fluid, joint fluid, pleural effusion, and ascites) is very important. High-quality samples contain pure pathogen-derived DNAs, excluding host-derived DNAs. For example, for streptococcal DNA extraction, a 5% sheep blood agar plate is inoculated with blood culture supernatant and aerobically incubated in 5% CO_2 at 35°C for 24 h. A single colony is selected from the plate and grown overnight in Todd–Hewitt broth supplemented with yeast extract. DNA is extracted using a DNeasy Blood & Tissue Kit (Qiagen) after pretreatment with proteinase K with or without lysozyme [8,10,11]. Library preparation is the process of converting the extracted DNA into a sequencing-ready form by attaching platform-specific adapters to DNA fragments. Although adapter ligation is a common principle across sequencing platforms, the adapter structures and ligation methods differ among the Illumina, Oxford Nanopore, and PacBio systems. At the sequencing step, long- and/or short-read sequencing can be selected. Finally, analysis involves the assembly and alignment of the long- and/or short-reads obtained, followed by annotation using tools such as Prokaryotic Genome Annotation Pipeline used by the National Center for Biotechnology Information (NCBI) to identify DNA coding sequences. The obtained genome sequences are subsequently deposited in GenBank, enabling data sharing worldwide. We obtained the complete/circular genome sequences of four *Streptococcus canis* specimens isolated from South Korea dogs using both long- and short-read sequencing, along with hybrid assembly (accession numbers CP053792, CP053793, CP053790, CP053791, CP053789, and CP046521) [12]. For Illumina sequencing, genomic libraries were prepared using the Nextera DNA Flex Library Prep Kit, and sequencing was performed on the Illumina MiSeq platform with a 2×150 bp paired-end protocol. Raw reads were processed using the FASTQ data pre-processing software tool fastp (v. 0.20.0) [13]. For Nanopore sequencing, a DNA library was constructed using the Rapid Sequencing Kit, and sequencing was performed on a MinION flow cell using the standard 48-h sequencing script. Fast5 files were generated by MinION and converted to FASTQ format. The resulting FASTQ reads were assembled using Unicycler. The hybrid assembly of the three strains was performed using Unicycler, with two resulting in a single chromosome and a plasmid. Three types of assemblies (short-, short-/long-, and long-read assemblies), including bridges, were generated, and quality scores were assigned to each bridge, with the most supportive bridge being selected. A complete chromosomal sequence of the vaginal swab isolate *Enterococcus faecalis* (including two complete plasmids) was obtained using a hybrid assembly (accession numbers CP185997, CP185998, and CP185999) [14].

Comparative genomic analysis

A single genome sequence is insufficient to demonstrate how genetic diversity can induce pathogenesis within a bacterial species. Therefore, genome-wide screening for potential vaccine candidates or antimicrobial targets is limited. Comparative genomics have revealed that a bacterial species is best described by its "pan-genome" (from the Greek *pan*, meaning "whole"), which consists of a core genome (genes present in all isolates), an accessory genome (genes present in one or more isolates), and genes that are unique to each isolate [15]. It is common to characterize the full gene complement as the pan-genome of

a closely related group of a single bacterial species using a web-based pipeline [16]. After constructing the pan-genome from WGS data, all coding DNA sequences are clustered into pan-genome orthologous groups (POGs) [17]. The resulting binary matrix, indicating the presence (1) or absence (0) of each POG, facilitates downstream analyses. POGs represent populations of orthologous genes identified across multiple genomes, and are useful for biomarker discovery. Additionally, Venn diagrams illustrate the relationships between shared, partially shared, and unique genes in isolates based on POG analysis [18]. We can gain insight into the spatial distribution of orthologous genes across genomes within a pan-genome. Bacterial metabolic pathways that are uniquely present in one group can be easily identified using pathway enrichment analysis. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database serves as a beneficial resource for such analysis [19]. Using 66 *Streptococcus agalactiae* genome sequences, we previously reported: (i) circular representations of these selected genomes via comparative genome hybridization; (ii) pan- and core-genome prediction curves; (iii) a Venn diagram comparing five representative isolates with capsular polysaccharides Ia, Ib, III, III, and VIII; (iv) a phylogenetic tree based on the POG data; and (v) KEGG pathway IDs, pathway names, and differentially present POG numbers [20]. We have also reported such findings using 20 *S. canis* genome sequences [21]. More recently, we performed a comparative genomic analysis of *Staphylococcus aureus* isolates carrying the staphylococcal cassette chromosome *mec* type V, and estimated the emergence of these clinical isolates in South Korea [22]. Furthermore, to identify homologous gene clusters and generate publication-quality visualizations comparing gene clusters between reference and candidate genome sequences, we used the online CompArative GEne Cluster Analysis Toolbox (CAGECAT) [23], which integrates two components: cblaster (for searches) and clinker (for figures) [24]. Fig. 1 shows a representative clinker image.

In contrast, we identified a *Pasteurella canis*-specific toxin gene through comparative genomic analysis [25]. Specifically, we retrieved the genomes of *P. canis* (n = 10) and *P. multocida* (n = 16) from the NCBI database. The VFanalyzer tool from the VF database was used to predict the VFs of *P. canis* and *P. multocida* [26]. Each genome sequence file (complete/draft genome in GenBank format) of *P. canis* and *P. multocida* was analyzed using the VFanalyzer to identify known or potential VFs, related genes, and corresponding genomic loci including nucleotide positions and sequences. This allowed us to determine putative *P. canis*-specific VFs that were not present in *P. multocida*. Fig. 2 shows the genome structures of *P. canis* (accession number CP085871) and *P. multocida* (accession number CP008918).

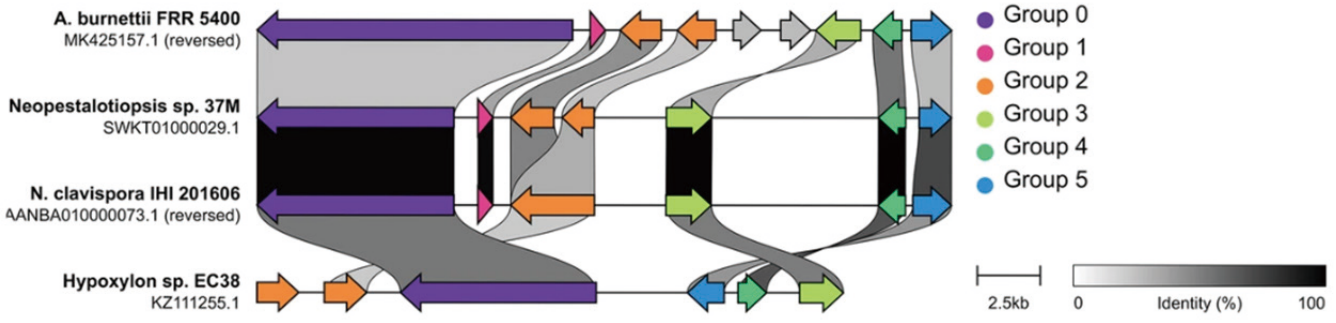

Fig. 1. Clinker visualization. Genes within a gene cluster are color-coded, and identical or similar genes among multiple clusters are connected by links shaded based on sequence identity. Figure cited from [24].

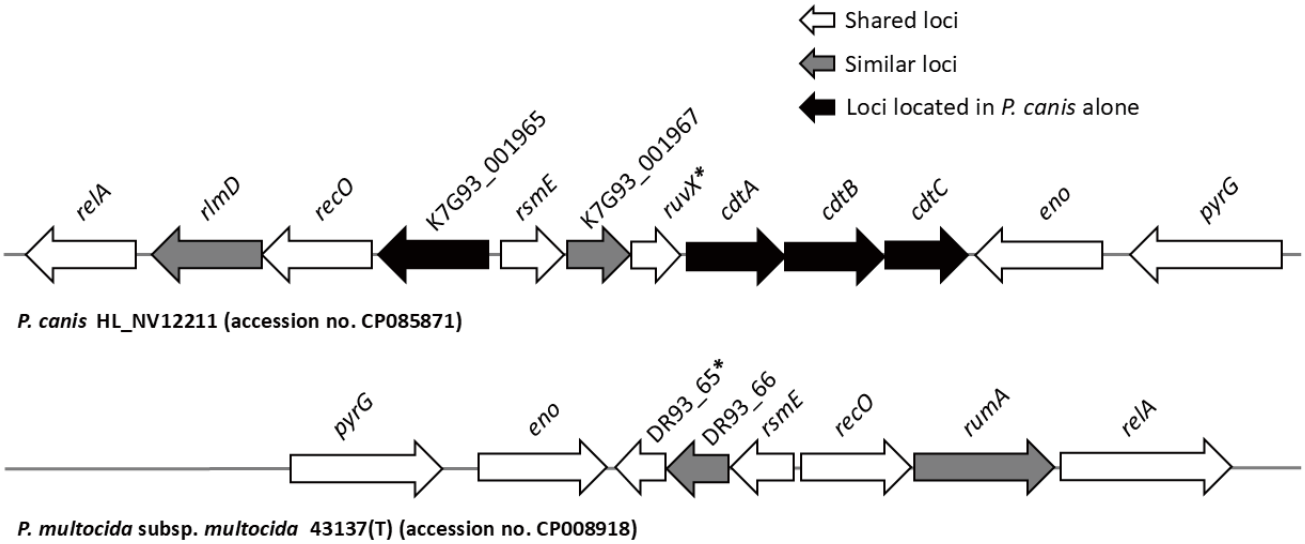

Fig. 2. Genome structure containing the cytolethal distending toxin *(cdt)A–cdtB–cdtC* loci of *Pasteurella canis* and adjacent loci from strain HL_NV12211 (accession number CP085871), compared to the corresponding region of *Pasteurella multocida* subsp. *multocida* ATCC 43137(T) (accession number CP008918). Asterisks indicate putative Holliday junction resolvase. K7G93_001965, K7G93_001967, and DR93_66 represent loci encoding hypothetical proteins. *relA*, GTP diphosphokinase; *rlmD* and *rumA*, 23S rRNA (uracil(1939)-C(5))-methyltransferase; *recO*, DNA repair protein; *rsmE*, 16S rRNA (uracil(1498)-N(3))-methyltransferase; *eno*, phosphopyruvate hydratase; *pyrG*, CTP synthase. Figure cited from [25].

RNA-seq (transcriptome) analysis

RNA-seq, also known as transcriptome analysis, is a method for quantifying gene expression and identifying new transcripts. It is commonly used to investigate gene up/downregulation and identify novel genes (Table 1). Transcriptomic changes to *S. pyogenes* in the inflammatory environment of necrotizing fasciitis have been documented using a mouse model [27]. In addition, RNA-seq revealed that the upregulation of arginine catabolism induces *S. pyogenes* pathogenesis on mouse skin surfaces [28]. *S. canis*, which exhibits pathological features similar to those of *S. pyogenes*, can infect humans who have been in close contact with, or have been bitten by, pet dogs, indicating that skin/soft tissue is an infection entry site. To clarify its pathogenic mechanisms in human cells, we recently determined *S. canis* transcriptomic alterations during the infection of human lineage keratinocytes (HaCat) *in vitro* [29]. Selective removal of human RNA was an important step in this study, and was achieved via differential bead-beating using large beads to lyse keratinocytes while preserving bacterial cells, followed by bacterial lysis using smaller beads for RNA extraction (Fig. 3). RNA was collected at three time-points (baseline, and 2 and 5 h post-inoculation), and RNA integrity was evaluated using a 2100 Bioanalyzer (Agilent Technologies). Total bacterial RNA was treated with the Ribo-Zero Plus rRNA Depletion Kit (Illumina Inc.) to remove rRNA. RNA-seq libraries were generated using the Illumina Stranded Total RNA Prep with RiboZero Plus. Three-stage RNA-seq was performed using a NovaSeq 6000 platform with 100 bp paired-end reads. The reads were quality-trimmed based on the quality scores (Q20/Q30) using the trimming tool in Qiagen CLC Genomics Workbench version 8 (Aarhus). The RNA-seq reads were then mapped to the corresponding genome sequences using the CLC Genomics Workbench, and normalization was conducted using reads per kb per million read values. Comprehensive gene expression analyses that comprised principal component analysis, *k*-means

clustering, and differential gene expression analysis were conducted. The identified differentially expressed genes (DEGs) were categorized according to their functional classifications. RNA-seq produced total read bases ranging from 6.17 to 9.02 Gbp. Both principal component analysis and *k*-means clustering analysis demonstrated inoculation time-dependent clustering. Visualization (i.e., volcano plots and Venn diagrams) revealed that the invasion of keratinocytes by *S. canis* affected the distribution of many genes. Gene ontology (GO) enrichment analysis revealed a dominant downregulation of genes, especially those linked to energy production, conversion/carbohydrate transport, metabolism/amino acid transport, and metabolism/nucleotide transport. Seven of the downregulated DEGs encoded pyrimidine salvage PyrR, pyrimidine biosynthesis PyrB, pyrimidine degradation UraA, orotate phosphoribosyltransferase PyrE, arginine deiminase ArcA, arginine biosynthesis ArgF, and carbamate kinase ArcC. This suggests significant reprogramming of arginine metabolic pathways. However, the upregulated genes were related to transcriptional processes.

Dual RNA-seq facilitates simultaneous monitoring of changes in gene expression in both microbial pathogens and their eukaryotic hosts under specific conditions, such as pathogen–host interactions [30,31]. This method has been applied to elucidate the genetic determinants governing *S. pyogenes*–host interactions in a murine skin infection model [32]. In the present study, dual RNA-seq was employed to assess gene expression changes in both bacterial and host cells at 5/24 h post-infection. The DEGs in *S. pyogenes* were related to metabolic pathways and Rgg2/Rgg3 quorum-sensing pathway activation, whereas those in murine skin were associated with inflammatory responses.

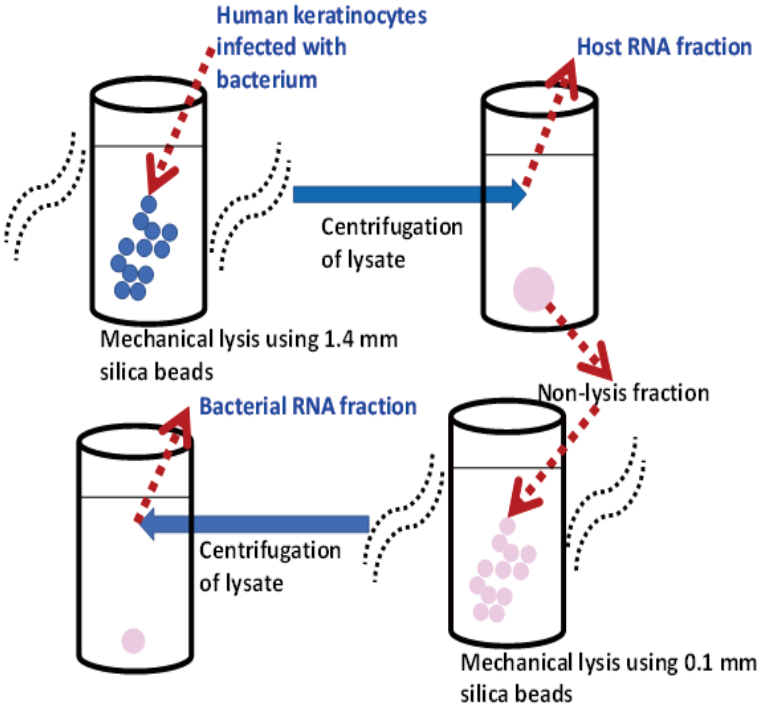

Fig. 3. Bacterial RNA isolation workflow. Mechanical lysis was performed using a MagNA lyser (Roche). Human keratinocytes were lysed with 1.4 mm silica beads (Qbiogene) in RLT lysis buffer (RNeasy Fibrous Tissue Mini Kit; Qiagen), and the human RNA fraction was removed by centrifugation. The pellets were then lysed with 0.1 mm silica beads (Qbiogene) in RLT lysis buffer and centrifuged to obtain the bacterial RNA fraction. Figure cited from [29].

Genome-based bacterial typing

Polymerase chain reaction-based multilocus sequence typing (MLST), introduced in 1998, is an accurate and popular molecular typing approach. It enables standardized, portable, and reproducible bacterial characterization by sequencing seven to eight conserved housekeeping genes [33]. Allelic combinations define sequence types, which are cataloged in curated online databases for standardized comparisons [34]. Thereafter, the advancement and accessibility of genome-based bacterial typing have been addressed.

Currently, genome-based typing methods are classified into two categories: (i) gene-by-gene allelic and (ii) single nucleotide variant (SNV)-based approaches [35]. The gene-by-gene allelic approach remains the gold standard for the classification of bacterial lineages. Core-genome MLST (cgMLST) expands traditional MLST by analyzing hundreds to thousands of conserved loci across the genome, providing higher resolution [7]. Whole-genome MLST (wgMLST) extends this approach to include both core and accessory genes, enabling finer strain differentiation while capturing broader genomic diversity within bacterial populations [7]. Publicly available cgMLST/wgMLST tools, such as chewBBACA (basic local alignment search tool [BLAST] score ratio-based allele calling algorithm), create open portable software that aids in cgMLST schema creation and facilitates these analyses across clinical microbiology laboratories [36].

However, SNV calling provides enhanced discriminatory resolution [7]. Once the alignment files are generated, variant-calling algorithms are used to detect genomic variations, such as single-nucleotide polymorphisms (SNPs) and insertions/deletions (INDELs), in the mapped reads with respect to a reference genome. Typically, these algorithms produce variant call-format files that enable comparisons at the strain level. The performance of different combinations of aligners and variant-calling tools (variant calling pipelines/workflows) has been assessed using short-read sequencing data [37]. Recent studies have demonstrated the viability of ONT-only approaches, with advancements from R9.4.1 to R10.4.1 flow cells (V10 to V14 chemistries) and adoption of improved basecalling models (Guppy to Dorado basecalling) [38].

Genome-based AMR determination

It is predicted that by 2050, AMR will cause approximately 10 million deaths annually and result in global economic losses totaling $1.7 trillion, based on disability-adjusted life years lost [39].

The detection of AMR genes from bacterial genome sequences requires: (i) well-curated and diverse databases of known AMR genes, and (ii) software tools for their identification [40]. Current bioinformatic strategies include: (i) BLAST-based sequence matching to the AMR gene database (nucleotide–nucleotide, protein–protein, translated nucleotide–protein alignments); (ii) combined mapping/alignment and targeted local assembly; and (iii) models identifying homology to existing AMR genes and SNPs within a curated database. Each tool employs one or more AMR reference databases, which catalog AMR determinants that may include combinations of AMR genes, SNPs, and/or INDELs, and may be species-specific or applicable across species. Table 3 presents a list of commonly used AMR databases [40].

However, different tools may produce different outputs, with varying interpretations of the presence/absence of specific AMR mechanisms. Indeed, in an external quality assessment involving nine participating

laboratories, the use of a common reference dataset but different bioinformatics algorithms resulted in high variability in the AMR genes detected [41]. Thus, (i) comprehensive and publicly accessible AMR databases, (ii) clear recommendations on sequencing data quality, and (iii) standardized methods for comparing AMR genotypes and phenotypes are fundamental for the successful implementation of WGS-based antimicrobial susceptibility predictions in clinical microbiology laboratories.

Table 3. Commonly used antimicrobial resistance databases [40]

Database [curator]	Traits
AMRFinderPlus Database [NCBI]	Comprehensive and curated database used by NCBI's AMRFinderPlus software tools Responsible for designation of allele names for new beta-lactamase and tetracycline-resistant genes
Comprehensive Antimicrobial Resistance Database (CARD) [McMaster University]	Comprehensive database of AMR gene sequences and SNPs Includes another database for AMR ontology; used by CARD's RGI tool Integrated with NCBI's AMRFinderPlus database
ResFinder database [Centre for Genomic Epidemiology]	Database for the ResFinder tools, including a web-based graphical user interface, with an emphasis on prediction of AMR phenotypes
Relational Sequencing TB Data Platform (ReSeqTB database)	AMR database for *Mycobacterium tuberculosis*, curated from large global data sets of *Mtb* sequences with phenotypic correlations

Abbreviations: AMR, antimicrobial resistance; NCBI, National Center for Biotechnology Information; SNP, single-nucleotide polymorphism; RGI, Resistance Gene Identifier; *Mtb*, *Mycobacterium tuberculosis*.

Estimation of integrative and conjugative elements (ICEs) carrying AMR gene(s)

Mobile genetic elements (MGEs) play a critical role in horizontal gene transfer between bacteria. The ICEs of MGEs are self-transmissible in microorganisms, and are characterized by both integrative and conjugative features [42]. ICEs are mosaic elements, possessing both bacteriophage-like and plasmid-like features that allow them to integrate and replicate within host cell chromosomes [43]. In addition, ICEs transfer their AMR genes, as well as genes involved in their mobility/regulation/maintenance. Many ICEs have been identified in *Streptococcus* species [44].

We recently characterized ICEs carrying *erm*(B)–*tet*(O) resistance genes in *Streptococcus uberis* genomes (n = 22) isolated from bovine milk in Chiba Prefecture, Japan, using CAGECAT in combination with ICEfinder [45]. We used ResFinder and ICEfinder [46] to identify AMR genes and ICEs. ResFinder detected co-localization *erm*(B)–*tet*(O)–*ant(6)-Ia* on the same contig in all genomes, and ICEfinder detected ICEs belonging to the same contigs containing complete or partial *erm*(B)–*tet*(O)–*ant(6)-Ia* sequences. Comparative genomic analysis using the *S. uberis* NZ01 strain as reference showed that the putative ICE in the UB37 strain was 77,386 bp, identical to that of the other 13 genomes. A similar streptococcal ICE, *Streptococcus suis* ICE*nsui78*–*tet*(O)–*erm*(B), was also identified. Fig. 4 shows the identification of a putative streptococcal ICE resembling *S. uberis* ICEs (UB37/UB68/UB23/UB99) with the *S. uberis* NZ01 reference genome. For ICE characterization in *S. uberis* with genomes, a comparative genomic analysis is needed using ICEfinder, CAGECAT, and other annotation tools (e.g., ICEscreen) [47].

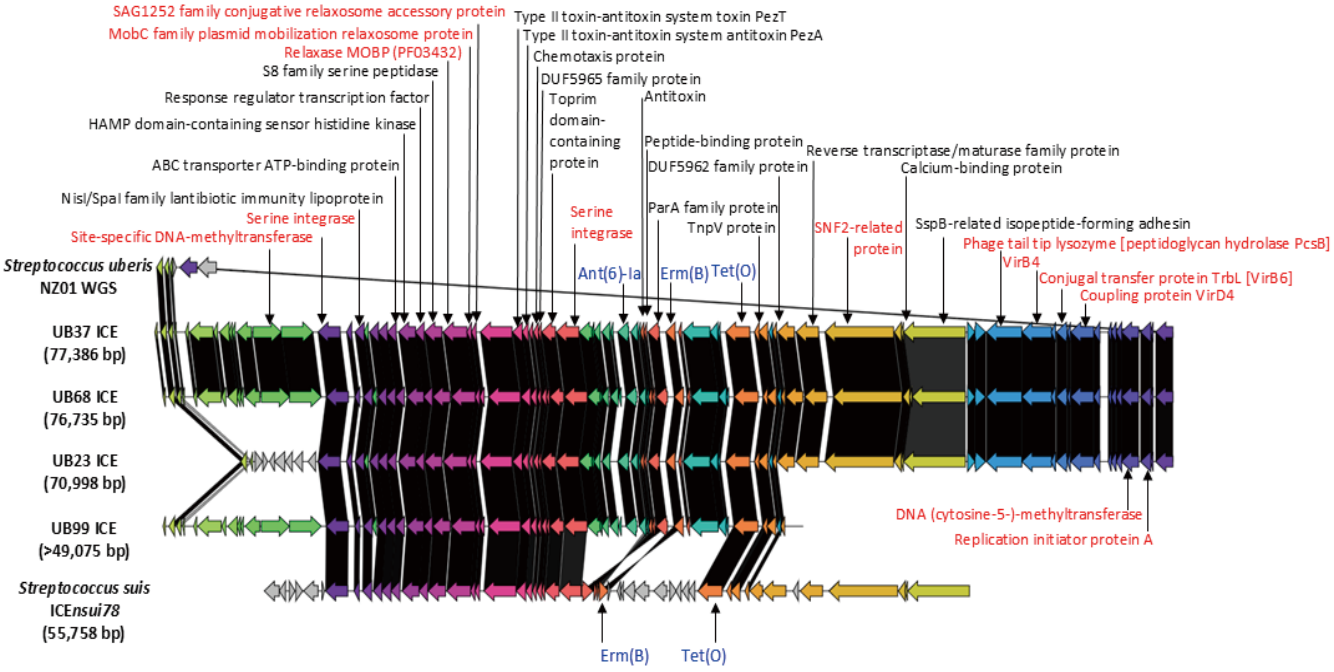

Fig. 4. Putative identification of other streptococcal ICEs resembling *Streptococcus uberis* ICEs (UB37/UB68/UB23/UB99) with the *S. uberis* NZ01 reference genome. Similar or identical products between reference and candidate genomes are indicated by matching colors. Black gradations between corresponding genes indicate percent identity. A similar streptococcal ICE was identified in *Streptococcus suis* strain STC78 (ICE*nsui78*–*tet*(O)–*erm*(B); accession number ON944185), based on conserved product arrangements between UB37/UB68/UB23/UB99 ICEs and ICE*nsui78*. Core products are in red font, and antimicrobial resistance products are in blue font. ICE*UB37* was re-annotated using the Prokaryotic Genomes Annotation Pipeline, and ICE components were identified using ICEscreen. Protein function similarities were inferred using Basic Local Alignment Search Tool x, with the predicted protein functions shown in brackets. Figure is cited from [45]. WGS, whole-genome sequencing; ICE, integrative and conjugative element.

Conclusion

WGS and related NGS are innovative clinical microbiology laboratory techniques based on high-throughput genomic results. They are used for specific pathogen identification, tracking, and AMR/virulence profiling. Although these techniques can involve high testing costs (including initial and operational charges), their benefits outweigh this limitation, thereby solidifying the position of WGS as an important technology in pathogen research. Additionally, WGS data-sharing on websites can promote accessibility among pathogen researchers worldwide and lead to the maintenance of public health. Furthermore, the introduction of portable sequencing machines (i.e., the Minion Nanopore sequencing device and ONT) will be significant for clinical microbiology laboratories. The development and utilization of artificial intelligence to analyze WGS data may further enhance its efficacy. In line with the concept of "One Health," human and animal microbiology laboratories should pay careful attention to the markedly dynamic evolution of WGS and related NGS technologies. Moreover, as single-cell RNA-seq analyses continue to develop, examination and regulation of the roles of microbial communities in clinical and natural environments will be essential [48].

Ethics statement

This was not a human population study. Therefore, institutional review board approval and informed consent were not required.

Conflict of interest

No potential conflicts of interest relevant to this article were reported.

Funding

None.

Data availability

This review article does not involve the generation or analysis of new datasets. All data supporting the findings are derived from previously published studies, which are appropriately cited within the manuscript.

References

1. Naik S, Kashyap D, Deep J, Darwish S, Cross J, Mansoor E, et al. Utilizing next-generation sequencing: advancements in the diagnosis of fungal infections. Diagnostics (Basel) 2024;14:1664.
2. Li W, Cui Q, Bai L, Fu P, Han H, Liu J, et al. Application of whole-genome sequencing in the national molecular tracing network for foodborne disease surveillance in China. Foodborne Pathog Dis 2021;18:538-46.
3. Gomes E, Araújo D, Nogueira T, Oliveira R, Silva S, Oliveira LVN, et al. Advances in whole genome sequencing for foodborne pathogens: implications for clinical infectious disease surveillance and public health. Front Cell Infect Microbiol 2025;15:1593219.
4. Wesenberg S, Mueller C, Nestmann F, Holthoff-Detto V. Effects of an animal-assisted intervention on social behaviour, emotions, and behavioural and psychological symptoms in nursing home residents with dementia. Psychogeriatrics 2019;19:219-27.
5. Thodberg K, Videbech PB, Hansen TGB, Pedersen AB, Christensen JW. Dog visits in nursing homes - increase complexity or keep it simple? A randomised controlled study. PLoS One 2021;16:e0251571.
6. Centers for Disease Control and Prevention. About One Health (updated on 27 June 2025). https://www.cdc.gov/onehealth/index.html [Online] (last visited on 12 September 2025).
7. Shropshire WC, Hanson BM, Shelburne SA. Genome-wide approaches to bacterial strain typing: a history and review of recent methodological advances. Curr Opin Infect Dis 2025;38:329-38.
8. Maeda T, Yoshida H, Abe N, Murakami K, Goto M, Takahashi T. Draft genome sequence of *emm*103/ST1363 *Streptococcus pyogenes* strain AB1, isolated from the blood of a woman with peritonitis and toxic shock syndrome. Microbiol Resour Announc 2024;13:e0102723.

9. Anslan S, Mikryukov V, Armolaitis K, Ankuda J, Lazdina D, Makovskis K, et al. Highly comparable metabarcoding results from MGI-Tech and Illumina sequencing platforms. PeerJ 2021;9:e12254.
10. Yoshida H, Katayama Y, Fukushima Y, Ohtaki H, Ohkusu K, Mizutani T, et al. Draft genome sequence of *Streptococcus canis* clinical strain OT1, isolated from a dog owner with invasive infection without a dog bite in Japan. Microbiol Resour Announc 2019;8:e00770-19.
11. Fukushima Y, Murata Y, Katayama Y, Tsuyuki Y, Yoshida H, Mizutani T, et al. Draft genome sequence of blood-origin *Streptococcus canis* strain FU149, isolated from a dog with necrotizing soft tissue infection. Microbiol Resour Announc 2020;9:e00737-20.
12. Kim JS, Sakaguchi S, Fukushima Y, Yoshida H, Nakano T, Takahashi T. Complete genome sequences of four *Streptococcus canis* strains isolated from dogs in South Korea. Microbiol Resour Announc 2020;9:e00818-20.
13. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884-90.
14. Kolar O, Appleberry H, Wolfe AJ, Kula A, Putonti C. Complete genome sequence of vaginal swab isolate *Enterococcus faecalis* UMB6935B, including two complete plasmids. Microbiol Resour Announc 2025;14:e0036825.
15. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A 2005;102:13950-5.
16. He EM, Chen CW, Guo Y, Hsu MH, Zhang L, Chen HL, et al. The genome of serotype VI *Streptococcus agalactiae* serotype VI and comparative analysis. Gene 2017;597:59-65.
17. EZBioCloud. About Pan-Genome Orthologous Group (POG) (updated on 15 May 2017). https://help.ezbiocloud.net/uncategorized/pan-genome-orthologous-group-pog/ [Online] (last visited on 12 September 2025).
18. Lin G, Chai J, Yuan S, Mai C, Cai L, Murphy RW, et al. VennPainter: a tool for the comparison and identification of candidate genes based on Venn diagrams. PLoS One 2016;11:e0154315.
19. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 2016;44:D457-62.
20. Takahashi T, Lee S, Kim S. Genomic characteristics of *Streptococcus agalactiae* based on the pan-genome orthologous group analysis according to invasiveness and capsular genotype. J Infect Chemother 2021;27:814-9.
21. Kim JM, Fukushima Y, Yoshida H, Kim JS, Takahashi T. Comparative genomic features of *Streptococcus canis* based on pan-genome orthologous group analysis according to sequence type. Jpn J Infect Dis 2022;75:269-76.
22. Takahashi T, Kim H, Kim HS, Kim HS, Song W, Kim JS. Comparative genomic analysis of staphylococcal cassette chromosome *mec* type V *Staphylococcus aureus* strains and estimation of the emergence of SCC*mec* V clinical isolates in Korea. Ann Lab Med 2024;44:47-55.
23. CompArative GEne Cluster Analysis Toolbox (CAGECAT) project. Welcome to CAGECAT. https://cagecat.bioinformatics.nl [Online] (last visited on 12 September 2025).
24. van den Belt M, Gilchrist C, Booth TJ, Chooi YH, Medema MH, Alanjary M. The CompArative GEne Cluster Analysis Toolbox for rapid search and visualisation of homologous gene clusters. BMC Bioinformatics 2023;24:181.
25. Yoshida H, Kim JM, Maeda T, Goto M, Tsuyuki Y, Shibata S, et al. Virulence-associated genome sequences of *Pasteurella canis* and unique toxin gene prevalence of *P. canis* and *Pasteurella multocida* isolated from humans and companion animals. Ann Lab Med 2023;43:263-72.

26. Virulence Factor Database (VFDB) project. Up-to-date knowledge of VFs of various bacterial pathogens. http://www.mgc.ac.cn/VFs/ [Online] (last visited on 12 September 2025).
27. Hirose Y, Yamaguchi M, Okuzaki D, Motooka D, Hamamoto H, Hanada T, et al. *Streptococcus pyogenes* transcriptome changes in the inflammatory environment of necrotizing fasciitis. Appl Environ Microbiol 2019;85:e01428–19.
28. Hirose Y, Yamaguchi M, Sumitomo T, Nakata M, Hanada T, Okuzaki D, et al. *Streptococcus pyogenes* upregulates arginine catabolism to exert its pathogenesis on the skin surface. Cell Rep 2021;34:108924.
29. Yoshida H, Goto M, Tsuyuki Y, Kim JS, Takahashi T. *Streptococcus canis* transcriptomic modifications in host cell entry environments of human keratinocytes. BMC Genomics 2024;25:1028.
30. Westermann AJ, Gorski SA, Vogel J. Dual RNA-seq of pathogen and host. Nat Rev Microbiol 2012;10:618-30.
31. Deb S, Basu J, Choudhary M. An overview of next generation sequencing strategies and genomics tools used for tuberculosis research. J Appl Microbiol 2024;135:lxae174.
32. Wilkening RV, Langouët-Astrié C, Severn MM, Federle MJ, Horswill AR. Identifying genetic determinants of *Streptococcus pyogenes*-host interactions in a murine intact skin infection model. Cell Rep 2023;42:113332.
33. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 1998;95:3140-5.
34. Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 2010;11:595.
35. Uelze L, Grützke J, Borowiak M, Hammerl JA, Juraschek K, Deneke C, et al. Typing methods based on whole genome sequencing data. One Health Outlook 2020;2:3.
36. Silva M, Machado MP, Silva DN, Rossi M, Moran-Gilad J, Santos S, et al. chewBBACA: a complete suite for gene-by-gene schema creation and strain identification. Microb Genom 2018;4:e000166.
37. Seah YM, Stewart MK, Hoogestraat D, Ryder M, Cookson BT, Salipante SJ, et al. *In silico* evaluation of variant calling methods for bacterial whole-genome sequencing assays. J Clin Microbiol 2023;61:e0184222.
38. Hall MB, Wick RR, Judd LM, Nguyen AN, Steinig EJ, Xie O, et al. Benchmarking reveals superiority of deep learning variant callers on bacterial nanopore sequence data. eLife 2024;13:RP98300.
39. Murray CJL, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 2022;399:629-55.
40. Sherry NL, Lee JYH, Giulieri SG, Connor CH, Horan K, Lacey JA, et al. Genomics for antimicrobial resistance-progress and future directions. Antimicrob Agents Chemother 2025;69:e0108224.
41. Doyle RM, O'Sullivan DM, Aller SD, Bruchmann S, Clark T, Coello Pelegrin A, et al. Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: an inter-laboratory study. Microb Genom 2020;6:e000335.
42. Burrus V, Pavlovic G, Decaris B, Guédon G. Conjugative transposons: the tip of the iceberg. Mol Microbiol 2002;46:601-10.

43. Huang J, Liang Y, Guo D, Shang K, Ge L, Kashif J, et al. Comparative genomic analysis of the ICE*Sa*2603 family ICEs and spread of *erm*(B)- and *tet*(O)-carrying transferable 89K-subtype ICEs in swine and bovine isolates in China. Front Microbiol 2016;7:55.
44. Beres SB, Musser JM. Contribution of exogenous genetic elements to the group A *Streptococcus* metagenome. PLoS One 2007;2:e800.
45. Maeda T, Tsuyuki Y, Yoshida H, Goto M, Takahashi T. Characterization of integrative and conjugative elements carrying *erm*(B) and *tet*(O) resistance determinants in *streptococcus uberis* isolates from bovine milk in Chiba prefecture, Japan: CompArative GEne cluster analysis toolbox with ICEfinder. BMC Res Notes 2024;17:377.
46. ICEberg 3.0 project. ICEfinder: Detection of ICE/IME of bacterial genomes. https://tool2-mml.sjtu.edu.cn/ICEberg3/ICEfinder.php [Online] (last visited on 12 September 2025).
47. Lao J, Lacroix T, Guédon G, Coluzzi C, Payot S, Leblond-Bourget N, et al. ICEscreen: a tool to detect Firmicute ICEs and IMEs, isolated or enclosed in composite structures. NAR Genom Bioinform 2022;4:lqac079.
48. Pountain AW and Yanai I. Dissecting microbial communities with single-cell transcriptome analysis. Science 2025;389:eadp6252.