CLINICAL AND LABORATORY STANDARDS INSTITUTE®

2nd Edition

MM18

Interpretive Criteria for Identification of Bacteria and Fungi by Targeted DNA Sequencing

This guideline includes information on sequencing DNA targets of cultured isolates, provides a quantitative metric for perceiving microbial diversity, and can serve as the basis to identify microorganisms. By establishing interpretive criteria for microorganism identification by targeted DNA sequencing, this guideline provides structure to laboratories that identify microorganisms for medical use.

A guideline for global application developed through the Clinical and Laboratory Standards Institute consensus process.

MM18, 2nd ed.
July 2018
Replaces MM18-A

Interpretive Criteria for Identification of Bacteria and Fungi by Targeted DNA Sequencing

Cathy A. Petti, MD
Mary E. Brandt, PhD
Deirdre L. Church, MD, PhD, D(ABMM)
Stefan Emler, MD
Keith Simmon, MS, PhD
Adrian M. Zelazny, PhD, D(ABMM)

Abstract

Clinical and Laboratory Standards Institute guideline MM18—*Interpretive Criteria for Identification of Bacteria and Fungi by Targeted DNA Sequencing* includes information intended for use with molecular diagnostic testing procedures published in CLSI documents MM03[1] and MM09.[2] These guidelines contain information about developing, evaluating, and applying nucleic acid–based testing for infectious diseases in medical laboratories.

Historically, microorganism identification has relied on phenotypic methods and, more recently, matrix-assisted laser desorption/ionization time-of-flight technology. Patient isolates for bacterial and fungal identification may have ambiguous biochemical profiles or mass spectra and cannot be reliably characterized. Laboratories can apply broad-range DNA sequencing for microorganism identification and as a standardized, portable method for data sharing. This guideline includes the most current information for microbial classification by targeted DNA sequencing, with particular emphasis on interpreting and reporting results.

Clinical and Laboratory Standards Institute (CLSI). *Interpretive Criteria for Identification of Bacteria and Fungi by Targeted DNA Sequencing*. 2nd ed. CLSI guideline MM18 (ISBN 978-1-68440-012-6 [Print]; ISBN 978-1-68440-013-3 [Electronic]). Clinical and Laboratory Standards Institute, 950 West Valley Road, Suite 2500, Wayne, Pennsylvania 19087 USA, 2018.

CLINICAL AND LABORATORY STANDARDS INSTITUTE®

Suggested Citation

CLSI. *Interpretive Criteria for Identification of Bacteria and Fungi by Targeted DNA Sequencing.* 2nd ed. CLSI guideline MM18. Wayne, PA: Clinical and Laboratory Standards Institute; 2018.

Previous Editions:
August 2007, April 2008

ISBN 978-1-68440-012-6 (Print)
ISBN 978-1-68440-013-3 (Electronic)
ISSN 1558-6502 (Print)
ISSN 2162-2914 (Electronic) Volume 38, Number 15

Contents

Foreword

Many laboratories use targeted DNA sequencing for bacterial (aerobic, anaerobic, and mycobacteria) and fungal identification, particularly for isolates that are poorly characterized by growth-dependent or matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) methods. Although CLSI document MM09[2] provides important contributions to this field, MM18 specifically focuses on reporting and interpreting results for identifying microorganisms by broad-range targeted DNA (eg, 16S ribosomal RNA, fungal internal transcribed spacer regions) sequencing. As this taxonomy-based method continues to evolve, this guideline provides recommendations for practically applying sequence-based technologies in the medical laboratory. Although the taxonomical classifications are not always clear, a consensus guideline on targeted DNA sequencing provides a systematic and uniform approach for consistently reporting standardized results across all medical laboratories.

Interpretive criteria for defining genus and species have been inconsistent in the literature, often varying with the queried microorganism. Because defining absolute interpretive criteria can be complex and highly nuanced, this guideline establishes recommendations for the systematic approach to classifying bacteria and fungi by broad-range DNA sequencing.

Overview of Changes

This guideline replaces the previous edition of the approved guideline, MM18-A, published in 2008. Several changes were made in this edition, including:

- Reorganized to fit the CLSI quality management system and path of workflow format
- Revised all bacteriology tables (Tables 6 to 14) to reflect current taxonomy and to outline where sequencing of the 16S rRNA gene's V1-V3 region (ie, first ≈500 base pairs) provides genus- and/or species-level identification and where diversity occurs within the entire gene to distinguish each genus and/or species
- Deleted table on bacterial agents of bioterrorism and its introductory text and added discussion of each agent to the group-specific tables (Tables 6 to 14)
- Revised all organism tables to include information on how MALDI-TOF MS may be used to complement sequencing for identification
- Deleted all dendrograms
- Revised organism tables to include emerging, clinically relevant microorganisms
- Updated organism nomenclature

NOTE: The content of this guideline is supported by the CLSI consensus process and does not necessarily reflect the views of any single individual or organization.

Key Words

Bacterial identification, fungal identification, gene sequencing, internal transcribed spacer, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, nucleic acid amplification, primer, 16S rRNA

Interpretive Criteria for Identification of Bacteria and Fungi by Targeted DNA Sequencing

Chapter 1: Introduction

This chapter includes:

- Guideline's scope and applicable exclusions
- Background information pertinent to the guideline's content
- "Note on Terminology" that highlights particular use and/or variation in use of terms and/or definitions
- Terms and definitions used in the guideline
- Abbreviations and acronyms used in the guideline

1.1 Scope

This guideline specifies recommendations for interpreting and reporting results of Sanger-based (dideoxynucleotide chain termination) sequencing of broad-range DNA targets for identifying pure isolates of bacteria, mycobacteria, and fungi from cultured patient isolates. Partial- and full-gene sequencing with 16S ribosomal RNA (rRNA) genes for bacterial and mycobacterial identification, as well as internal transcribed spacer (ITS) regions (ie, ITS-1 and ITS-2) for fungal identification are covered, including alternative DNA targets when appropriate. Although massively parallel (next-generation) sequencing technologies are rapidly emerging, this guideline's scope is limited to Sanger-based targeted DNA sequencing.

To assist the medical laboratory, guidance is provided for:

- Selecting DNA targets and sizes for amplification and sequencing
- Establishing QC parameters for amplification and sequencing
- Measuring sequence quality
- Assessing reference sequences and databases
- Comparing sequences for identification
- Establishing interpretive criteria for identity scores generated by targeted DNA sequencing
- Developing clinically relevant reporting strategies for specific microorganism groups
- Identifying the limitations of targeted DNA sequencing for microbial identification

The intended users of this guideline are medical laboratories and laboratories performing amplification and Sanger-based (dideoxynucleotide chain termination) sequencing of broad-range DNA targets for identifying bacteria, mycobacteria, and fungi from cultured patient isolates.

This guideline does not:

- Include procedures for performing microbial sequencing.
- Include RNA targets for sequencing.

- Provide guidance on definitive taxonomical criteria for microorganism classification or identification methods for novel microorganisms.

- Cover alternative sequencing systems or specific molecular assays designed with these broad-range DNA targets.

- Discuss typing strains for epidemiological purposes.

- Discuss virus or parasite identification.

- Discuss amplification and sequencing directly from patient specimens.

1.2 Background

As a consequence of gene and genome sequence availability, microbial taxonomy has undergone a revolution over the past few decades. Comparing gene sequences from different organisms provides a quantitative metric for recognizing microbial diversity and classifying diverse organisms. Gene sequences also serve as the basis of molecular tools used for sensitive and incisive organism identification. Highly conserved genes, such as rRNA genes, provide information on general organism characteristics based on the properties of their known relatives. Other gene sequences can provide additional detail (eg, for more precise identification to species level or for identifying antimicrobial-resistant bacterial strains).

In contrast to precise identifications based on sequence comparisons, microorganisms traditionally have been identified based on physiological properties exhibited in pure cultures. However, phenotypic properties are often ambiguous, and slowly growing or difficult-to-culture microbes present additional identification barriers and expense. Mass spectrometry–based methods have gained importance for bacterial and fungal identification. Sequence-based methods serve as the gold standard and complement the microbiologist's laboratory practice. Gene sequence–based techniques can be used to detect, identify, and monitor microbes, including poorly known or uncultured pathogens.

1.3 Terminology

1.3.1 A Note on Terminology

CLSI, as a global leader in standardization, is firmly committed to achieving global harmonization whenever possible. Harmonization is a process of recognizing, understanding, and explaining differences while taking steps to achieve worldwide uniformity. CLSI recognizes that medical conventions in the global metrological community have evolved differently in different countries and regions and that legally required use of terms, regional usage, and different consensus timelines are all important considerations in the harmonization process. CLSI recognizes its important role in these efforts, and its consensus process focuses on harmonization of terms to facilitate the global application of standards and guidelines.

The nomenclature used in this guideline, commonly referred to as "IUB codes," was developed by the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biochemistry and Molecular Biology (IUBMB), who established the IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN) and the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) to facilitate the communication of biochemical information.[3]

Chapter 2: Approach to Targeted DNA Sequencing

This chapter includes:

- Discussion of primer design
- Information on controls for DNA preparation (extraction), amplification, and sequencing
- Protocols and instrument requirements
- An overview of sequence data
- Recommendations for reference database selection
- Guidance on comparing sequences for identification

2.1 Primer Design

A sufficient amount of target material is needed for nucleic acid–based sequencing, usually about 3 ng of DNA. Direct bacterial genomic DNA sequencing is difficult. Most laboratories use PCR-based amplification of the region to be sequenced. Generally, amplicons should cover as many variable regions as possible, as well as diversity between species for easier differentiation. Amplicons and sequences should be at least 450 base pairs (bp) in length and well suited for use with current sequencing biochemistries to provide good-quality sequences. For DNA amplification of some organisms, it is useful or necessary to amplify longer DNA fragments and then use multiple sequencing reactions to produce contiguous, quality sequence data covering more variable regions for higher diversity. Sequencing primers can be designed to target a known sequence in the target amplicon or nontemplate sequences generated by tagging the 5′ end of PCR primers. PCR primer selection, amplification parameter optimization, and sequencing primer selection are critical for obtaining good results (see CLSI document MM09[2]).

2.1.1 Bacteria and Mycobacteria Only

Because it was determined that evolutionary relationships could be inferred from comparing parts of a bacterial genome, the use of molecular methods for bacterial identification has increased.[9,10] Primers should be selected to target genome regions that are universal and conserved for a wide bacterial spectrum.

2.1.1.1 16S Ribosomal RNA Gene

Ribosomal RNA (rRNA) genes (5S, 16S, and 23S) and intergenic regions are specific for prokaryotes and are commonly used for taxonomic purposes.[11] These genes:

- Are present in all organisms
- Are considered only weakly affected by horizontal gene transfer
- Contain mosaics of sequence stretches ranging from highly conserved to variable

These characteristics make the 16S rRNA gene the most widely used region for bacterial taxonomy and identification (see Figure 1).

generic public database entries. If the focus is on using type-strain sequences as references for species identification, specialized or commercial databases offering such content should be considered.

2.5.2.2 Sequence Quality

Databases should be curated to ensure that primarily high-quality sequences with correct annotations are present with minimal ambiguous bases (N) and that the gene or regions used for identification are fully covered. Sequences with full gene coverage (> 80%) should be preferred over partial sequences, and sequences with unusual insertions, deletions, or variations should be excluded. All entries should be compared regularly with newly published sequences for that species or strain or updated with the most recent taxonomy.

2.5.2.3 Sequence Redundancy

To reduce the chance of systematic errors in annotations, databases should include multiple entries of the same microorganism from different environments, hosts, and countries and be provided by multiple scientific communities or peer-reviewed articles. Using different sequencing technologies reduces the chance of inherent errors and for bias in the aggregate.

Including multiple sequence entries for a given species, when deposited by different scientific investigators, increases confidence in the accuracy of the sequences and their annotation and reduces the chance of systematic errors, especially in annotation. Furthermore, multiple reference sequences from several sources more adequately represent naturally occurring intraspecies diversity. Finally, entry redundancy should be reasonably limited to avoid the chance that only the most closely related species appear in a match list (eg, BLAST).

2.5.2.4 Periodic Updates

To ensure accurate identification of newly described species and variants and to reflect changes in taxonomy, the reference database used for comparison should be routinely updated, at least every three to six months.

2.5.2.5 Nomenclature and Annotation

Public databases[49,50] share a standardized annotation scheme in which a sequence's annotation includes the species for which the sequence stands (eg, *Staphylococcus aureus*) (see Table 5). Examples are publicly available. [51]

Table 5. Standardized Annotation for Public Databases

Term	Annotation
Unique identifier or accession number	AC
Author(s)	AU
Source of the isolate from which the sequence was generated	SO
Date of last modification and other information	LOCUS
Species	OS
Sequence	ORIGIN

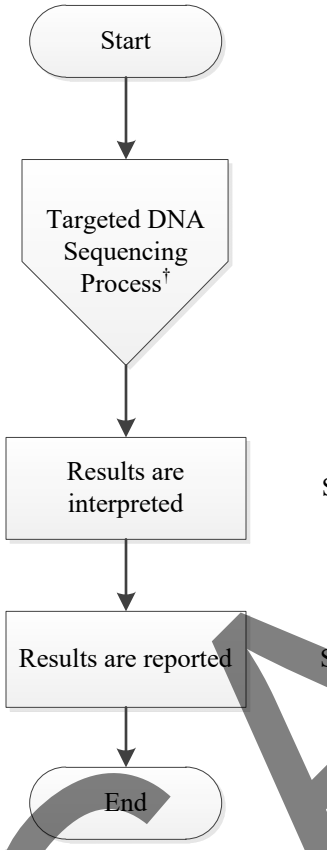
Annotations are generated by the authors who have submitted the entry with its sequence, but there can be errors if the submission has not been peer reviewed. Given the huge amount of sequence data currently being added to the public domain, such detailed verification by human interaction is not possible. More extensively curated and specific databases try to resolve the verification issue by reviewing annotations for specific targets and organisms; however, this time-consuming process can reduce the frequency of updates. Therefore, these databases may lag behind the general public databases. Potential errors in curation may include:

Chapter 3: Targeted DNA Sequencing Interpretation and Reporting Process

This chapter includes:

- Process overview for interpreting and reporting targeted DNA sequencing
- Identity score interpretive criteria for bacteria
- Interpretive criteria for fungal identification
- Suggestions for result reporting

This guideline provides guidance for targeted DNA sequencing examination and postexamination activities. Figure 3 provides an overview of the process discussed in the following subchapters.



* Five basic symbols are used in process flow charts: oval (signifies the beginning or end of a process), arrow (connects process activities), box (designates process activities), diamond (includes a question with alternative "Yes" and "No" responses), pentagon (signifies another process).
† See Chapter 2 and CLSI documents MM06[58] and MM09.[2]
Abbreviation: DNA, deoxyribonucleic acid.

Figure 3. Targeted DNA Sequencing Interpretation and Reporting*

3.1 Identity Score Interpretive Criteria for Bacteria

This subchapter focuses on practically applying broad-range DNA sequencing, with special emphasis on commonly encountered bacteria in a medical laboratory. Bacteria are clustered into related groups, with each table reflecting the group's association within a medical and classical microbiological context. The groups do not necessarily reflect classifications based on strict taxonomical relationships, but current taxonomy has been updated within each group. The most common broad-range DNA target to identify