Molecular Diagnosis of Infection

Sophie Arbefeville, M.D.
Assistant Director, Infectious Disease Diagnostic Laboratory
University of Minnesota Medical Center, Fairview

Disclosure

- None

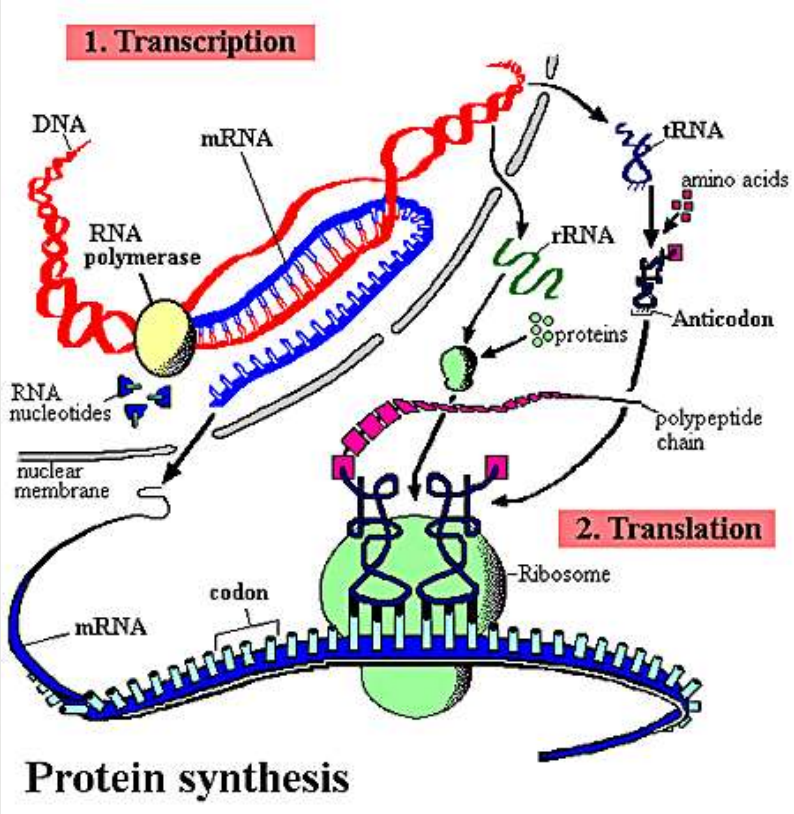
Objectives

- Explain why the 16S rRNA gene was chosen for bacterial identification.
- Describe the mechanism and limits of bacterial 16S rRNA gene sequence analysis.
- Discuss the impact and contribution that the 16S rRNA gene sequence analysis makes to the understanding of clinical microbiology and infectious diseases.
- Discuss the role of Next-Generation Sequencing Technology in microbiology.

Background

- In the 1980s, a new standard for identifying bacteria began to be developed.
- Dr. Woese and others showed that phylogenetic relationships of bacteria, and, indeed, all life-forms, could be determined by comparing a stable part of the genetic code – the ribosomal genes.

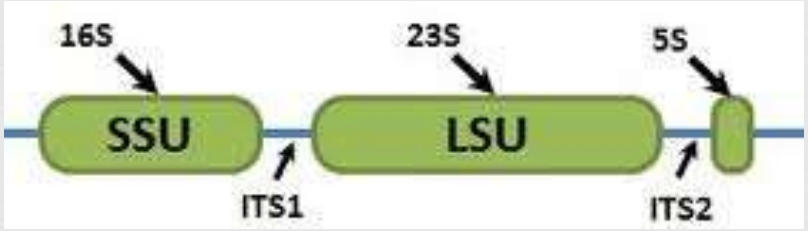
Why use Ribosomal RNA Genes?

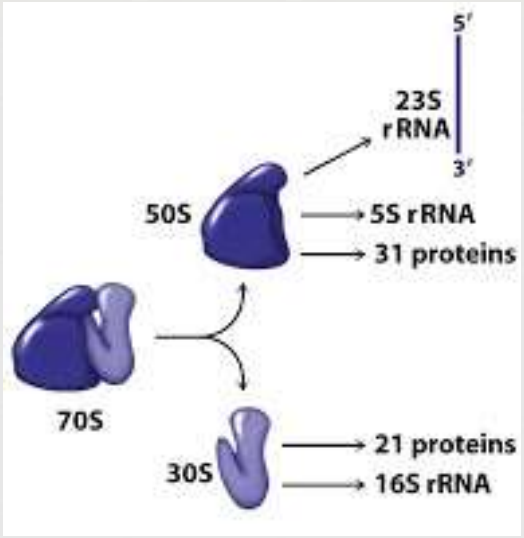


- All organisms, except viruses have them.
- All perform the same function-protein synthesis.
- The high conservation of rRNA genes.
  - Dr. Woese described 16S rRNA gene as a "molecular chronometer" for evolution in bacteria

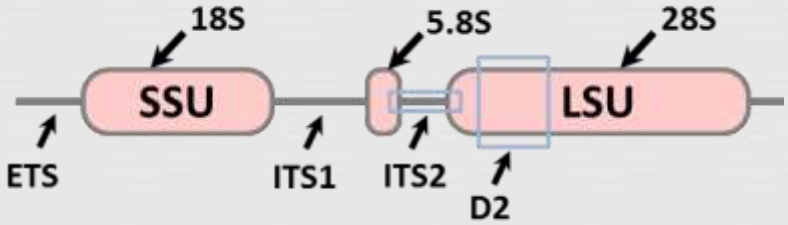
http://www.accessexcellence.org/RC/VL/GG/protein_synthesis.html

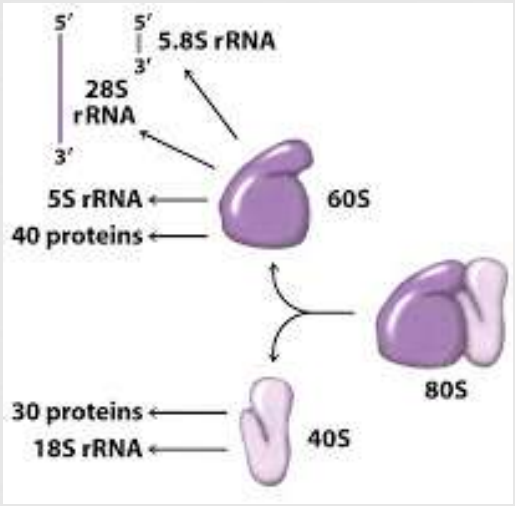
Prokaryote





Eukaryote





Highly conserved rRNA genes

The extraordinary conservation of rRNA genes can be seen in these fragments of the small subunit (16S) rRNA gene sequences from organisms spanning the known diversity of life. Note several areas of identity among these diverse organisms:

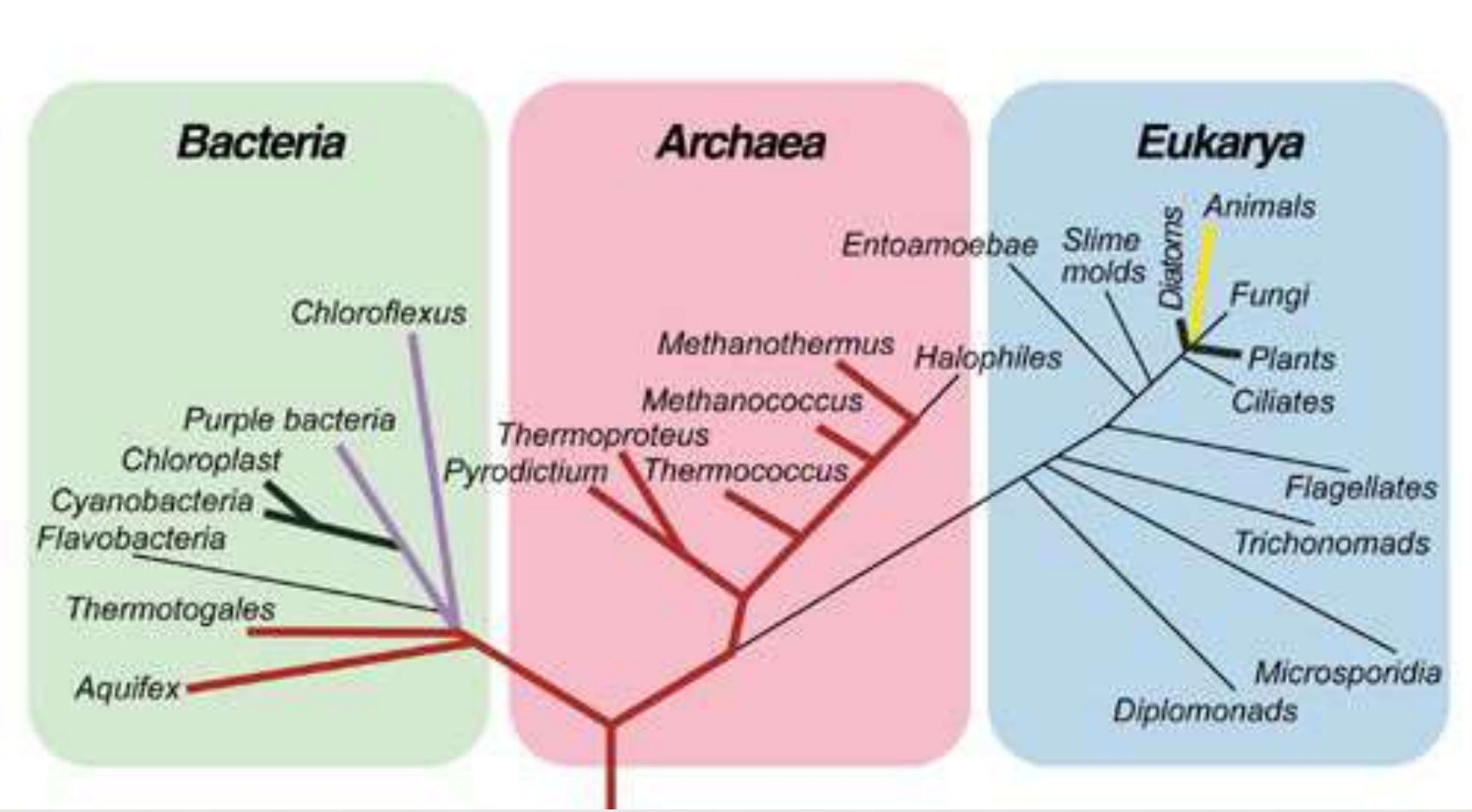
human
GTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTAAAGTTGCTGCAGTTAAAAAG

yeast
GTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTAAAGTTGTTGCAGTTAAAAAG

corn
GTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTTAAAAAG

E. coli
GTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCG

16S/18S Sequencing Redefined Phylogenetic Tree of Life



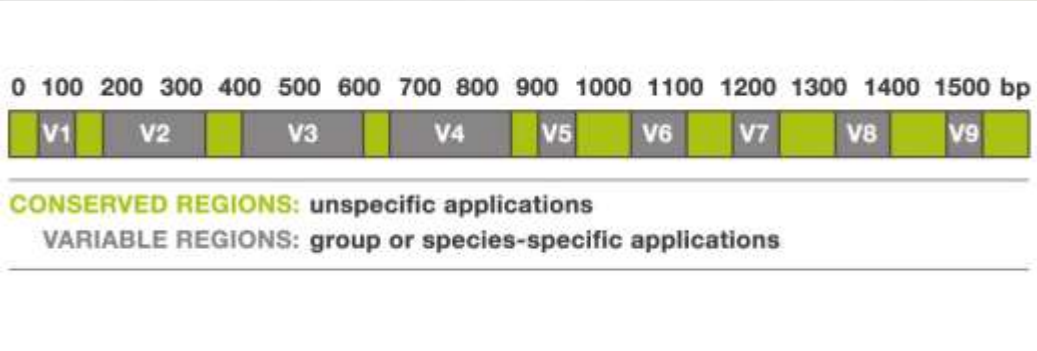
Using these rRNA gene sequences for phylogenetic studies, three domains of life, Archaea, Bacteria and Eukarya, as opposed to traditional classification of living organisms into prokaryotes and eukaryotes only, were described.

What are the characteristics of 16S rRNA gene that make it useful as an analyte for bacterial identification?

- Universal in bacteria
- Highly Conserved Molecule
  - The 16S rRNA gene has a mutation rate which is very close to the rate at which species diverge into new ones, making it ideal for species level identification.
  - The degree of conservation is assumed to result from the importance of the 16S rRNA as a critical component of cell function
  - RNA product is structurally part of the ribosome 30S small subunit
- Contains highly conserved and hypervariable regions
- Extensively studied and represented in databases
  - GenBank, the largest databank of nucleotide sequences, has over 20 million deposited sequences, of which over 90,000 are of 16S rRNA gene
  - 16S specific databases
    - Ribosomal Database Project (RDP)-II
    - Ribosomal Database Project European Molecular Biology Laboratory
    - Ribosomal Differentiation of Medical Microorganisms (RIDOM)
    - ABI MicroSeq Databases

16S rRNA Gene:

- 1.5 kb in length
- Consists of 8 highly conserved regions and 9 variable regions
- 1-15 copies/bacterium depending on the species (*E. coli* has 7 copies of the gene)
- Universal in bacteria
- The 16S rRNA gene is also designated 16S rDNA, and the terms have been used interchangeably: current ASM policy is that "16S rRNA gene" be used.



Baker GC, Smith JJ, Cowan DA. 2003. Review and re-analysis of domain-specific 16S primers. J Microbiol Methods. 55(3):541-55.

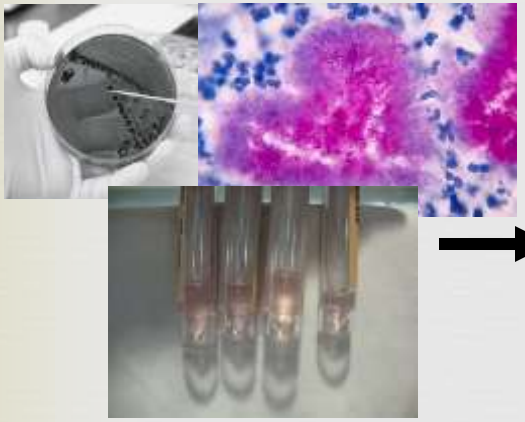
Hypervariable Regions

A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria.

Soumitesh Chakravorty, Danica Helb, Michele Burday, Nancy Connell, and David Alland

Bacterial 16S ribosomal RNA (rRNA) genes contain nine "hypervariable regions" (V1 – V9) that demonstrate considerable sequence diversity among different bacteria. Species-specific sequences within a given hypervariable region constitute useful targets for diagnostic assays and other scientific investigations. No single region can differentiate among all bacteria; therefore, systematic studies that compare the relative advantage of each region for specific diagnostic goals are needed. We characterized V1 - V8 in 110 different bacterial species including common blood borne pathogens, CDC-defined select agents and environmental microflora. Sequence similarity dendrograms were created for hypervariable regions V1 – V8, and for selected combinations of regions or short segments
within individual hypervariable regions that might be appropriate for DNA probing and real-time PCR. We determined that V1 best differentiated among Staphylococcus aureus and coagulase negative *Staphylococcus* sp. V2 and V3 were most suitable for distinguishing all bacterial species to the genus level except for closely related enterobacteriaceae. V2 best distinguished among Mycobacterial species and V3 among Haemophilus species. The 58 nucleotides-long V6 could distinguish among most bacterial species except enterobacteriaceae. V6 was also noteworthy for being able to differentiate among all CDC-defined select agents including *Bacillus anthracis*, which differed from *B. cereus* by a single polymorphism. V4, V5, V7 and V8 were less useful targets for genus or species-specific probes. The hypervariable sequence-specific dendrograms and the "MEGALIGN" files provided online will be highly useful tools for designing specific probes and primers for molecular assays to detect pathogenic bacteria, including select agents.

1) Collect bacteria



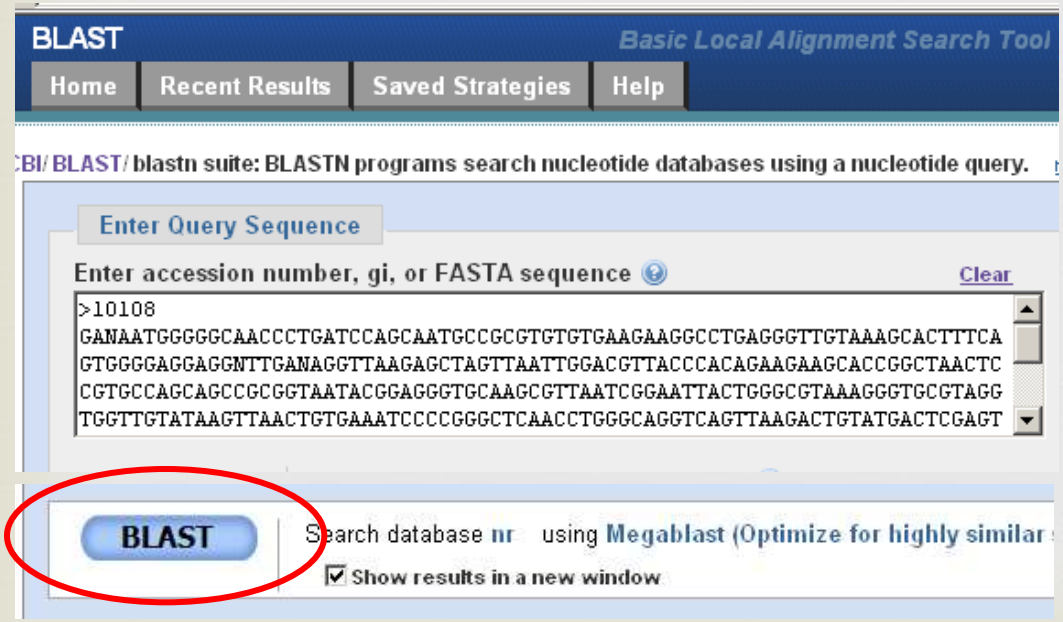
2) Extract DNA

3) PCR the 16S rRNA gene

4) Sequence the products

5) Analyze sequence data

6) Blast the sequence.



Sequence identification is only as good as the sequence database

- Database searching is part of the assay
- A perfect database contains:
  - All clinically relevant organisms
  - Only sequences of high quality
  - Only sequences from well-characterized strains
  - Entries that are fully compliant with current nomenclature

Sequence databases

- Public
  - GenBank
  - Ribosomal Database Project (RDP)
- Private
  - SmartGene
  - MicroSeq

Results

- Scores
  - Bit Score (S)
    - the higher the better
  - Expect (E) Score
    - the lower the better, similar to a probability
- Percent Identity
  - Depends on the length of matching sequence
- Taxonomy Report
  - How many matches to an organism
    - Depends on the amount of sequence in the database
- Are the organisms "type" organisms
  - E.g. are they listed as coming from the American Type Culture Collection or as an ATCC strain?

Classification

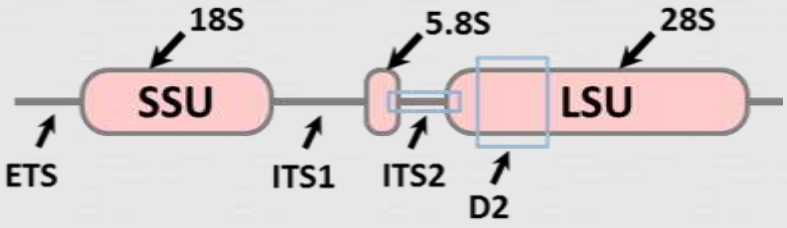
- Family: a group of related genera.
- Genus: a group of related species.
- Species: a group of related strains.
- Type: sets of strain within a species (e.g. biotypes, serotypes).
- Strain: one line or a single isolate of a particular species.

Interpretative Criteria for Identification of Bacteria and Fungi by DNA Target Sequencing; Approved Guideline CLSI: MM18-A

- ≥ 99.0% identity for species identification (with greater than 0.8% separation between different species); report "[Genus and species]"
- ≥97.0% identity for genus identification; consider reporting "[Genus], most closely related to [species]"
- ≥95.0% identity cannot be definitively identified by 16S rRNA gene sequencing; consider reporting "Unable to identify definitively by 16S rRNA gene sequencing, most closely related to [Genus]"

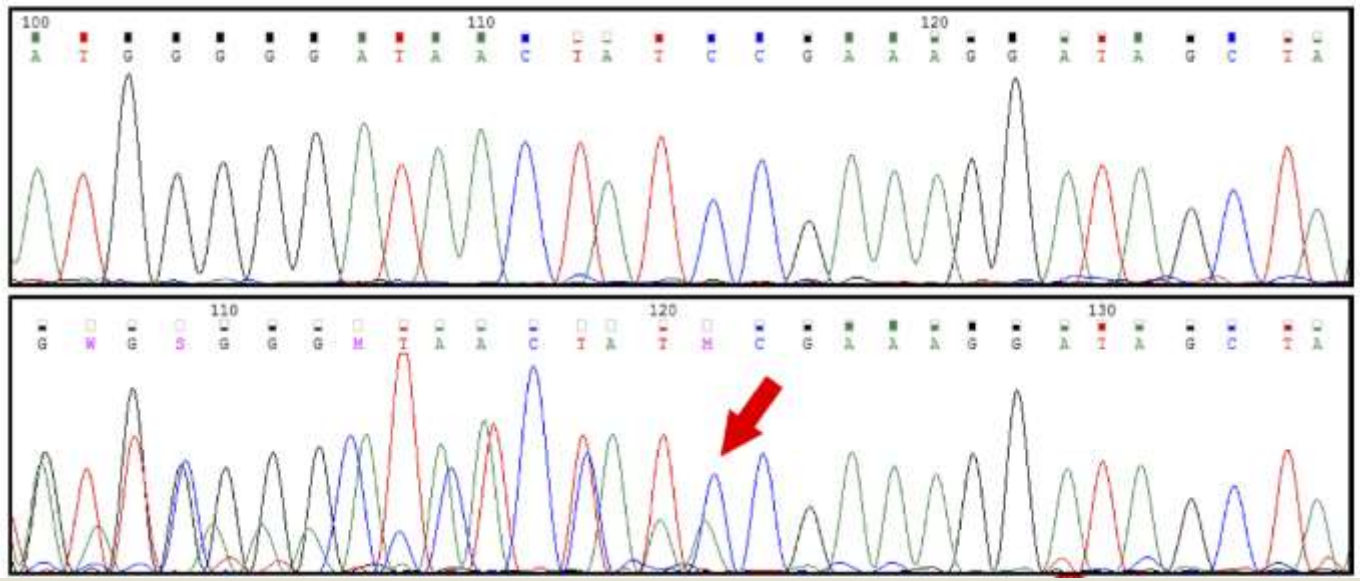
Other genes useful for identification

- Mycobacteria
  - Hsp65
  - rpoB (RNA polymerase)
- Fungi
  - Internal Transcribed Spacer (ITS) 1 and 2 regions
  - 28S rRNA



Limitations

- 16S rRNA gene sequencing using bulk PCR products cannot be applied to polymicrobial specimens: the presence of multiple templates results in superimposed Sanger reads that are generally uninterpretable without using specific software or cloning.



Limitations

- Some organisms are not represented well in the databases
- Some organisms are very similar if not identical i.e. show low phylogenetic resolving power.
  - Definitive species identification is difficult without further phenotypic or molecular testing
  - However, genus identification is usually clear
  - Not adequate usually for epidemiological purposes
- Public databases are not curated
  - Mostly leads to a problem of species assignment when doing the BLAST search

Circulation
JOURNAL OF THE AMERICAN HEART ASSOCIATION

American Heart Association®

Impact of a Molecular Approach to Improve the Microbiological Diagnosis of Infective Heart Valve Endocarditis

Claudia Breitkopf, Dieter Hammel, Hans H. Scheld, Georg Peters and Karsten Becker

Circulation. 2005;111:1415-1421; originally published online March 7, 2005;

- Conclusions
  - Culture-independent molecular methods substantially improve the diagnostic outcome of microbiological examination of excised heart valves.
  - Importantly, this was true not only for fastidious, slow-growing, and/or nonculturable microorganisms but also for easy-to-culture pathogens such as streptococci and staphylococci.
  - Both patient management and empiric antibiotic therapy of IE are likely to benefit from improved knowledge of the spectrum of pathogens now causing IE.

Journal of Clinical Microbiology

Identification of Bacteria in Formalin-Fixed, Paraffin-Embedded Heart Valve Tissue via 16S rRNA Gene Nucleotide Sequencing

Kavita Imrit, Michael Goldfischer, Jie Wang, Jaime Green, Jerome Levine, Joseph Lombardo and Tao Hong
J. Clin. Microbiol. 2006, 44(7):2609. DOI: 10.1128/JCM.00572-06.

- We applied 16S rRNA gene sequencing to identify bacterial species present in formalin-fixed, paraffinembedded heart valve tissue. In 40% (12/30) of the cases, we were able to identify the bacterium to the species-genus level. For more recent cases (<4 years), the success rate was significantly improved, to 70% ($P < 0.001$).

Journal of Clinical Microbiology

Analysis of 525 Samples To Determine the Usefulness of PCR Amplification and Sequencing of the 16S rRNA Gene for Diagnosis of Bone and Joint Infections

Florence Fenollar, Véronique Roux, Andréas Stein, Michel Drancourt and Didier Raoult
J. Clin. Microbiol. 2006, 44(3):1018. DOI: 10.1128/JCM.44.3.1018-1028.2006.

- Overall, the 16S rRNA gene PCR assay followed by sequencing offers several advantages when used to complement culture results, but its use should be restricted.
- We do not recommend performing 16S rRNA gene PCR in the following circumstances:
  - prior to prosthesis implantation
  - in cases where a diagnosis other than infection is established
  - for patients with a positive culture of a single, highly pathogenic bacterium if they are not also at risk for a polymicrobial infection.
- We propose the use of the 16S rRNA gene PCR assay for culture-negative cases when infection is suspected on the basis of clinical signs and symptoms or inflammatory syndrome is present, as highlighted by blood test results and purulent samples.

Next-Generation Sequencing Technology

- Ability to produce a large volume of data in a short period of times
- E. coli (EHEC) outbreak in Europe, researchers were able to quickly generate a high-quality whole genome sequence of the bacterial strain
- Challenges:
  - lack of reference genomes available for most species
  - Whole-genome sequencing must be done *de novo*
  - Short read lengths
- Targeted Sequencing

Analysis of the Cystic Fibrosis Lung Microbiota via Serial Illumina Sequencing of Bacterial 16S rRNA Hypervariable Regions

Heather Maughan[1*☯], Pauline W. Wang[2☯], Julio Diaz Caballero[1], Pauline Fung[2], Yunchen Gong[2], Sylva L. Donaldson[2], Lijie Yuan[2], Shaf Keshavjee[3], Yu Zhang[3,4], Yvonne C. W. Yau[5], Valerie J. Waters[5], D. Elizabeth Tullis[6], David M. Hwang[3,4], David S. Guttman[1,2*]

Table 2. Illumina NGS and CE-Based Sanger Sequencing for Targeted Applications

Parameter	MiSeq System	Sanger Sequencing
Samples in project	96	96
Number of amplicons	12	12
Target panel size	~5 kb	~5 kb
Time for sample prep	< 3 hours	< 3 hours
Sequencing time	1 day	6 days
Price per amplicon*	$1 USD	$4 USD
Project price*	< $2000 USD	> $4500 USD
Coverage depth per amplicon	> 13,000 x	2 x**
On instrument data analysis?	Yes	No

* Excluding PCR amplification.
** Including bidirectional sequencing.

illumina®

References

- Woese C, Fox G. 1977. "Phylogenetic structure of the prokaryotic domain: the primary kingdoms.". Proc Natl Acad Sci USA 74 (11): 5088–90
- Woese C, Kandler O, Wheelis M. 1990. "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.". Proc Natl Acad Sci USA 87 (12): 4576–9
- Baker GC, Smith JJ, Cowan DA. 2003. Review and re-analysis of domain-specific 16S primers. J Microbiol Methods. 55(3):541-55.
- Breitkopf C, Hammel D, Scheld HH, Peters G, Becker K. 2005. Impact of a molecular approach to improve the microbiological diagnosis of infective heart valve endocarditis. Circulation. 111(11): 1415-21
- Fenollar F, Roux V, Stein A, Drancourt M, Raoult D. 2006. Analysis of 525 samples to determine the usefulness of PCR amplification and sequencing of the 16S rRNA gene for diagnosis of bone and joint infections. J Clin Microbiol. 44(3):1018-28
- Imrit K, Goldfischer M, Wang J, Green J, Levine J, Lombardo J, Hong T. 2006. Identification of bacteria in formalin-fixed, paraffin-embedded heart valve tissue via 16S rRNA gene nucleotide sequencing. J Clin Microbiol. 44(7):2609-11
- Maughan H, Wang PW, Diaz Caballero J, Fung P, Gong Y, Donaldson SL, Yuan L, Keshavjee S, Zhang Y, Yau YC, Waters VJ, Tullis DE, Hwang DM, Guttman DS. 2012. Analysis of the cystic fibrosis lung microbiota via serial Illumina sequencing of bacterial 16S rRNA hypervariable regions. PLoS One. 7(10):e45791

Questions ?