Diagnosing an Infectious Disease
Nucleic Acid Testing, Polymerase chain reaction (PCR)

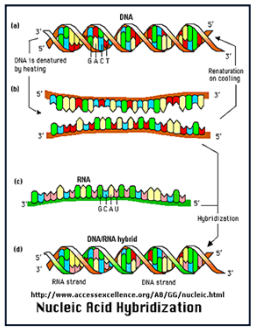
CULTURE INDEPENDENT DIAGNOSTIC TESTING (CIDT)

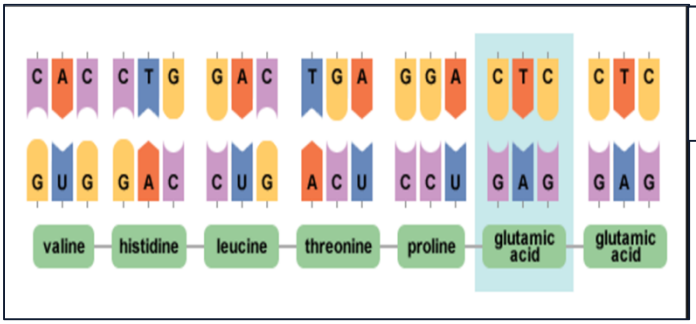
- Look at the pathogens (bacteria, virus, fungus) under a microscope
  - Bacteria often look alike, even when stained
  - Sometimes fairly specific: Syphilis treponema, Gram-negative diplococci in CSF…
- Look for specific antibodies or antigens
  - Slow to develop,
  - Often cross reactions
- **Identification of very specific gene: Polymerase Chain Reaction**

CULTURE

- Very specific
- Bacterial colonies may be missed in a crowded plate
- Viral culture require cell cultures difficult to maintain

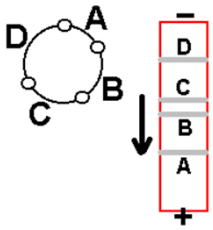
Gene Coding for Proteins





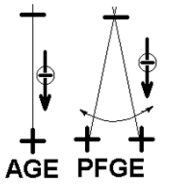
- The genes are used to code for the components of a cell
- Codes used to get amino acids used as building blocks of proteins

Pulse Field Gel Electrophoresis (PFGE)



CUTTING UP DNA IN PIECES

- Plasmid or Chromosomal DNA can be used
- Whole or Fragmented:
- Restriction endonuclease cuts DNA molecules at restriction sites
- Enzymes selected carefully to generate appropriate fragments
- Some are frequent cutters, others not



MIGRATION OF FRAGMENTS IN GELS

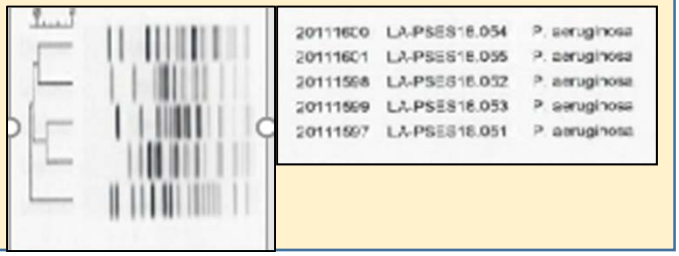
- Polyacrylamide Agar Gel Electrophoresis
- Plain
- Pulsed Field
- ±Transfer to nitrocellulose:
- Restriction Fragment Length Polymorphism (RFLP) analysis with DNA probes or Southern Blot analysis
- DNA fragments hybridized with chemically or radioactively labeled DNA or RNA probe which binds to only a few fragments with complementary sequences.
- RFLP using insertion element IS110 method of choice to type Mtb

- This gel has 7 lanes
- The first and last lanes are the controls used. For this gel, it was *E. faecalis* #ATTC 29212
- Lanes 2-6 are the specimens which are identified by the PFGE # on top of each lane

- PFGE testing is available to sentinel and non-sentinel hospitals
- Must consult with Infectious Disease Epidemiology Section prior to sending in isolates for PFGE

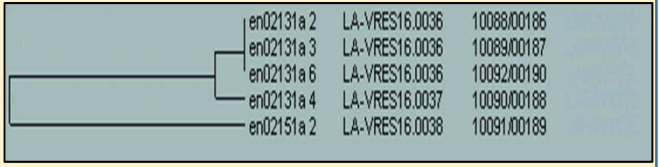
- Typically used only when we have a suspected outbreak/cluster of similar bacterial species; however, it can be utilized for multiple pathogens (i.e. *Salmonella, Campylobacter*, *Serratia marcescens, Pseudomonas aeruginosa)*

- NICU with 5 cases of Pseudomonas infection: Question: Are these cases related? Same source? Staff? Contaminated object?



DENDOGRAM

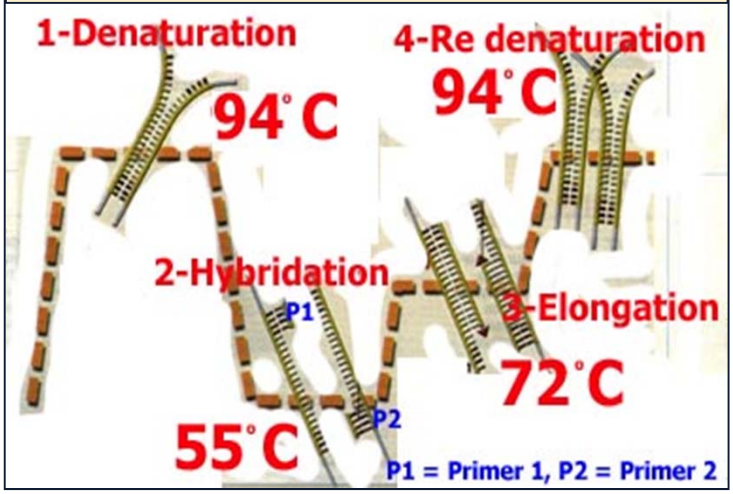
- The dendogram illustrates the relationship between the five isolates
- This example indicates that three isolates (10088, 10089, and 10092) are 100% related; Isolate #10090 is closely related (97%) to the other three isolates
- The fifth isolate (10091) is unrelated to the rest of the VRE isolates



Polymerase Chain Reaction PCR

• **Polymerase chain reaction (PCR)** is a common laboratory technique used to make many copies (millions or billions!) of a particular region of DNA.
• Goal of PCR is to make enough of the target DNA region that it can be analyzed or used in some other way.
• PCR is used in many areas of biology and medicine, including molecular biology research, medical diagnostics, and even some branches of ecology.
• Since the Nucleic Acid sequence of a gene is specific to a pathogen (bacteria, virus, fungus), **PCR can be used to identify pathogens.**
• The **selection of the gene** to used is important since pathogens of the same "family" may have gene similar or fairly similar genes
• PCR requires a **DNA polymerase enzyme** that makes new strands of DNA, to copy strands as templates. The DNA polymerase typically used in PCR is called Taq polymerase from heat-tolerant bacterium (*Thermus aquaticus*).
• *T. aquaticus* lives in hot springs and hydrothermal vents. Its DNA polymerase is very heat-stable and is most active around 70°C. This heat-stability makes Taq polymerase ideal for PCR. As we'll see, high temperature is used repeatedly in PCR to denature the template DNA, or separate its strands.
• Taq polymerase can only make DNA if it's given a **primer, a short sequence of nucleotides that provides a starting point** for DNA synthesis. In a PCR reaction, the experimenter determines the region of DNA that will be copied, or amplified, by the primers chosen.
• PCR primers are short pieces of single-stranded DNA, usually around 202020 nucleotides in length. Two primers are used in each PCR reaction, and they are designed so that they flank the target region (region that should be copied). That is, they are given sequences that will make them bind to opposite strands of the template DNA, just at the edges of the region to be copied. The primers bind to the template by complementary base pairing.

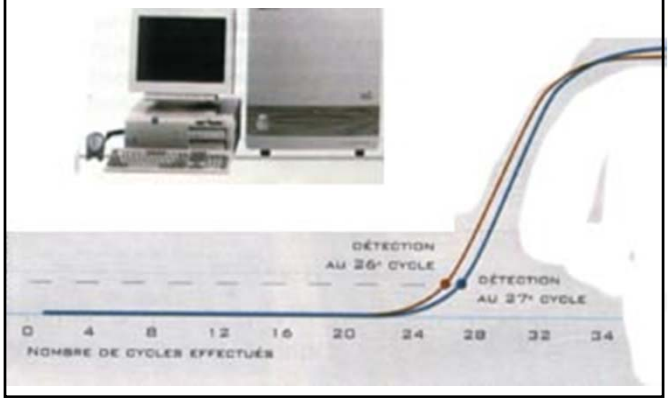
Amplification: 1-Denaturation, 2-Hybridation, 3-Elongation



Real Time PCR

Primer with fluorescent tags as Genome of interest is produced.

Tags at right distance produce desired fluorescence.



The NUMBER OF CYCLES is IMPORTANT

• RT-PCR shows the development of the PCR reaction in time
• If positive reaction appears rapidly, (10 to 20 cycles), it is almost a certain positive reaction confirming the presence of the targeted gene
• If positive reaction appears very late (>40 cycles for example), the result is conclusive (negative)

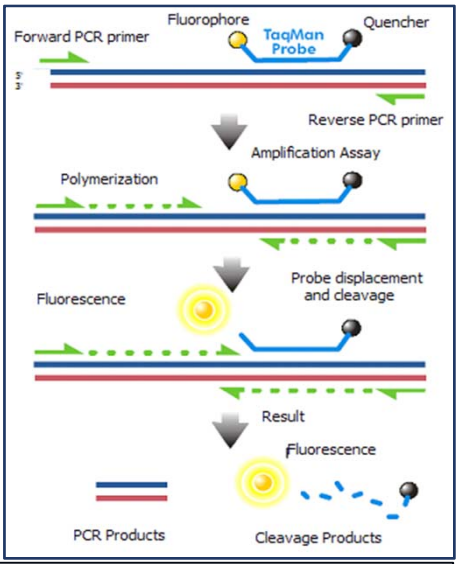
• **The SELECTION OF GENE(S) is IMPORTANT**

• If a single gene is selected, other microbes with similar genes will have a positive reaction
• Need to include a combination of genes that are very specific to the targeted microbe

TAQMAN PROBES

TaqMan probes are designed such that they anneal within a DNA region amplified by a specific set of primers. The probe binds to single stranded DNA. As the *Taq* polymerase extends the primer and synthesizes the nascent strand, the exonuclease activity of the *Taq* polymerase **degrades the probe** that has annealed to the template.

Degradation of the probe **releases the fluorophore** from it and breaks the close proximity to the quencher, thus relieving the quenching effect and allowing fluorescence of the fluorophore. Hence, fluorescence detected in the quantitative PCR thermal cycler is directly proportional to the fluorophore released and the amount of DNA template present in the PCR.



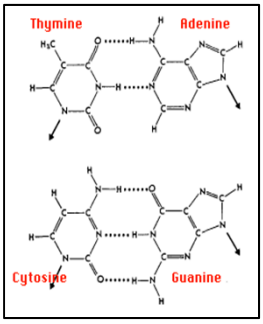
MULTIPLEX ASSAY

• A multiplex assay is a type of test used to measure **multiple analytes in a single run/cycle** of the assay.
• Ability to rapidly process multiple samples in an automated fashion is what characterizes high-throughput techniques.
• Multiplex polymerase chain reaction (Multiplex PCR) refers to the use of polymerase chain reaction to amplify several different DNA sequences simultaneously (as if performing many separate PCR reactions all together in one reaction).
• This process amplifies DNA in samples using multiple primers and a temperature-mediated DNA polymerase in a thermal cycler. The primer design for all primers pairs has to be optimized so that all primer pairs can work at the same annealing temperature during PCR.

Genome Sequencing

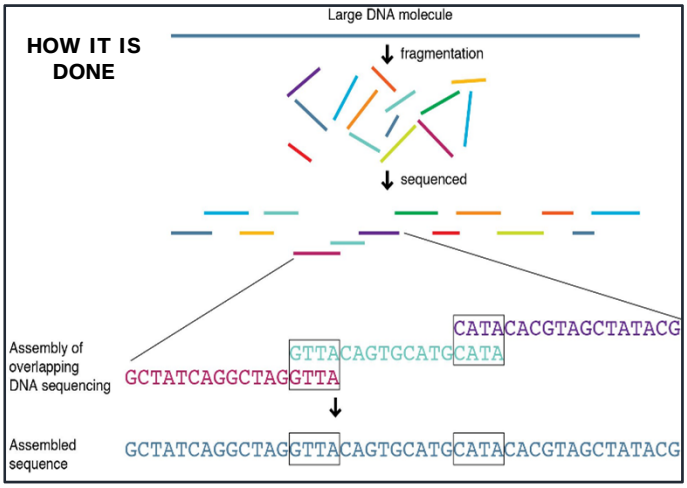
WHAT IS GENOME SEQUENCING?

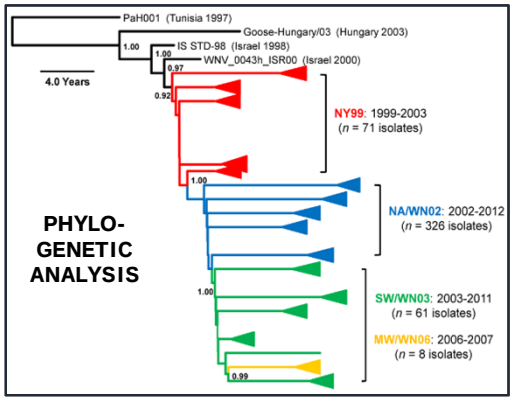
- Genome sequencing is figuring out the order of DNA nucleotides in a genome: The order of A T G C in an organism's DNA.
- Genes account for less than 25% of the DNA in the genome
- Size of human genome estimated around 3,000 Mb (megabase pairs) or 3.0 $\times 10^9$ bp (base pairs)
- Bacterial genomes can range in size from about 130 kbp to over 14 Mbp.

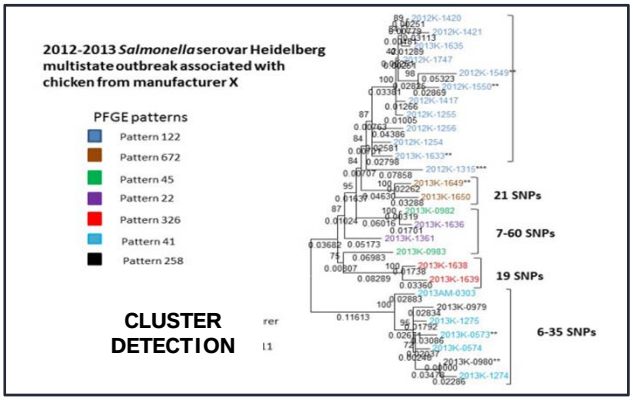


BASE PAIRING

- A with T: the purine adenine (A) always pairs with the pyrimidine thymine (T)
- C with G: the pyrimidine cytosine (C) always pairs with the purine guanine (G)







COMPARISON PFGE/GENOME SEQUENCING	PFGE	WGS
Isolate required	Yes	Yes
Size of outbreaks detected	Large	Large to small
Number of outbreaks detected	Less	More
Possible outbreaks ruled out	Less	More
Interpretation of case, food & environmental matches	If common pattern: weak hypothesis	Always strong hypothesis
Specificity for organisms	High for some only	Very high for all
Ability to evaluate closeness of strains	Low	High
Nature of data	Categorical	Continuous
Match means association	Yes	Yes
Miss match means association less likely, but not impossible	Yes	Yes
Need other data to support conclusion: epidemiology, trace back	Yes	Yes

Gene Sequencing

• Essentially two main surveillance objectives for WGS-based comparative genome analysis, each of which require distinct data analysis and reporting outputs

• **Phylogenetic analysis** which results in the visualization as a tree or network graph, based on measurement of the evolutionary distance between genomes and their hypothesized order of descent from their most recent common ancestor.
• This analysis is used to infer transmission linkage between isolates from different patients or potential infection sources.
• Prediction of clinically and epidemiologically relevant microbial phenotypes in terms of antigenic profile, drug resistance and virulence, including identification of determinants encoded in the accessory genome and mobile genetic elements such as phages, plasmids and transposons (mobilome).

• **Cluster detection**, several analytical strategies have been used to assess genomic distances and apply cluster or phylogenetic analysis methods to group closely-related strains
The three predominant methods currently used in the literature are:
• SNP-calling based phylogenetic reconstruction of assembled draft genomes by alignment to an annotated reference genome; for reference scheme construction followed by gene-by-gene assignment to allelic profile (cgMLST) based on draft de-novo assembled genomes;
• K-mer-based grouping of the closest genome matches by comparing across very short sequences.

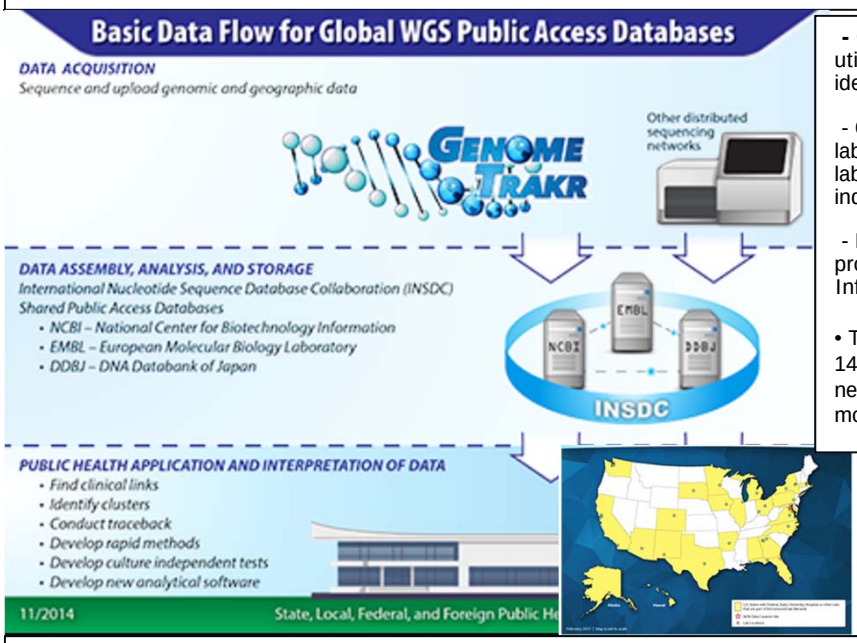
• **Surveillance systems** designed to detect common source outbreaks caused by closely related/identical strains - it is likely that the standard analysis will consist of two steps.
• First, using a gene-by-gene based nomenclature, which enables results to be compared across laboratories and a first clustering to be made.
• This is then followed by SNP analysis to further resolve the phylogenetic structure of identified clusters of isolates with common/closely related cgMLST types. To develop a common nomenclature assignment, an openly accessible database is required for each pathogen species under surveillance. This database should be able to return allele identifiers when provided with a sequence. Minimum database functionality criteria have been defined for the application of WGS in foodborne disease surveillance at EU and global level.

BENEFITS OF GENE SEQUENCING

• Most promising public health benefit may come from pairing **genomic information with geographic location.**
• Applying principles of evolutionary biology to determine the relatedness of pathogens.
• Species genomic information found in one geographic area is different than that of the same species of pathogen from another area.
• Knowing the geographic areas that pathogens are typically associated with can be a powerful tool in tracking down the root source of contamination
• Since 2008, FDA part of an international effort to build a network of Genome sequencing laboratories which upload genomic sequence and pathogen geographic location.
• As the size of the database grows, so will its strength as a tool to help focus and speed investigations into the root cause of illnesses

GENOME SEQUENCING VS PFGE

• Whole genome sequencing reveals the complete DNA make-up of an organism,
⏩enabling to better understand variations both within and between species.
⏩allows to differentiate between organisms with a precision that other technologies do not allow.
⏩PFGE is unable to differentiate some strains of *Salmonella*.



- GenomeTrakr is the first distributed network of labs to utilize whole genome sequencing for pathogen identification.

- Consists of 15 federal labs, 25 state health and university labs, 1 U.S. hospital lab, 2 other labs located in the U.S., 20 labs located outside of the U.S., and collaborations with independent academic researchers.

- Data curation and bioinformatic analyses and support are provided by the National Center for Biotechnology Information (NCBI) at the National Institutes of Health

• The GenomeTrakr network has sequenced more than 146,000 isolates and closed more than 175 genomes. The network is regularly sequencing over 5,000 isolates each month.

• How to translate genomic data into meaningful information for public health decision-making is still incomplete. Current technical limitations of WGS-based typing include the complexity and reproducibility of the sequence data produced
• Sequencing platforms currently used differ in terms of the range of quality control and this may influence the accuracy and inter-laboratory comparability of sequence
• Another potential limitation of WGS typing for some diseases is the lack of backward compatibility with typing systems such as PFGE and MLVA.

Matrix-Assisted Laser Desorption/Ionization (MALDI)

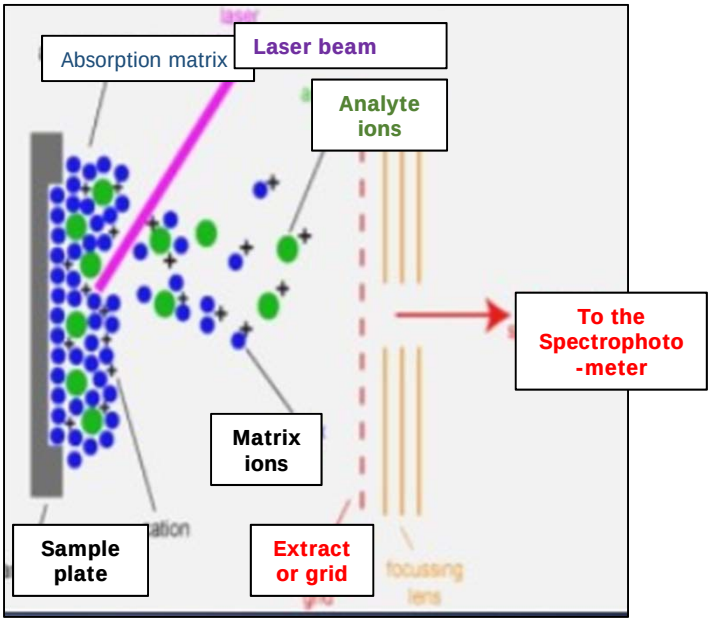
WHAT IS MALDI

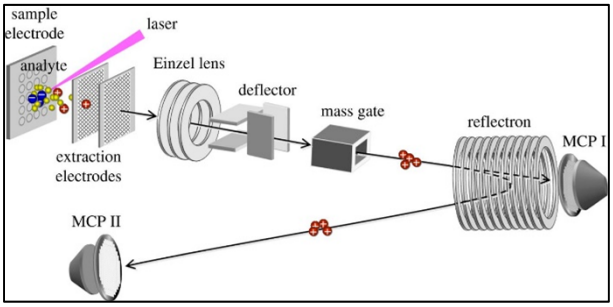
- Ionization technique that uses a laser energy absorbing matrix to create ions from large molecules with minimal fragmentation.
- Amino acid alanine could be ionized easily if it was mixed with amino acid tryptophan & irradiated with pulsed 266nm laser.
- Tryptophan absorbs the laser energy & helps to ionize the non-absorbing alanine.

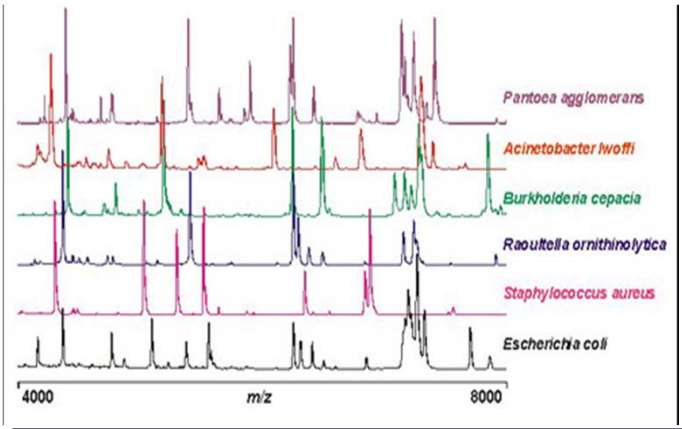
APPLICATIONS - Applied to the analysis of biomolecules (biopolymers such as DNA, proteins, peptides, sugars and large molecules (polymers, dendrimers, other macromolecules).

USES IN MICROBIOLOGY

- It is used for the **identification of microorganisms**. Species diagnosis by this procedure is much faster, more accurate & cheaper than other procedures based on biochemical tests.
- Results are available within 2 hours.
- A **Rapid method to Investigate Blood Stream Infections.** MALDI Sepsityper solution provides a rapid, highly accurate microbial identification directly from a positive blood culture.
- Although it does not provide antimicrobial susceptibility data (with some exception (Direct detection of resistance genes (MecA of MRSA, VRE, CTX-M, KPC, NDM)), it has good potential to guide empirical antimicrobial choice in the treatment of BSIs, yet there remain technical variables that may affect test performance







TECHNIQUE

1-The sample is mixed to a suitable matrix and applied on a metal plate.
2-A pulsed laser irradiates the sample; the matrix absorbs the light.
3-Energy is transferred to the analyte.
3-The analyte is ionized into the gas phase due to the large amount of energy absorbed.
4-A high electrical field accelerate the ions into a flight tube in the mass spectrometer.

LIMITATIONS

- E. coli vs. Shigella – Very closely related and cannot be differentiated
- Streptococcus pneumoniae vs. Streptococcus mitis group – Very closely related, new databases can give a definitive ID – Differentiate by Bile solubility or optichin disk
- Bordetella pertussis vs. Bordetella bronchioseptica – Very closely related and cannot be differentiated
- Stenotrophomonas maltophilia vs. Pseudomonas hibiscola, vs. P.gentculata, vs. P. Beteli: Very closely related and cannot be differentiated
- The Acinetobacter baumanii-calcoaceticus complex (A. baumanii, A. calcoaceticus, A. genospecies 3, A. genospecies: Species differentiation can be difficult. – A. baumanii and A. calcoaceticus can be differentiated; there are several members of the "Genospecies 3" clustering with A. baumanii or A. calcoacteticus, this can lead to "A. genospecies 3"
- Identification of Streptococcal species: The lower yield of valid MALDI-TOF MS results with streptococci and staphylococci might be due: (i) to the close relatedness of the different species of streptococci belonging to the S. mitis group (i.e. S. pneumoniae, S. mitis, S. sanguinis, S. oralis, ...), (ii) to some relatedness of different coagulase negative staphylococci, (iii) to the cell wall composition of Gram-positive bacteria conferring increased resistance to lysis, and (iv) partially to the possible presence of some residual blood proteins I
- Identification of Staphylococcal species For staphylococci, the major goal is to differentiate S. aureus from coagulase negative staphylococci and this may be accurately done on blood 108 culture bacterial pellets using the MALDI-TOF MS.
- Difficulty in identifying S. pneumoniae from other species of the S. mitis group is much more clinically relevant and represents a current limitation of the MALDI-TOF MS. The presence of a capsule may also partially explain the low identification rate of S. pneumoniae, H. influenza and K. pneumoniae. Improved extraction protocols specifically designed for encapsulated bacteria are thus warranted.