Laser-Based Identification of Bacteria using Discriminant Function Analysis

Khadija Sheikh

University of Windsor - Department of Physics - Windsor, Ontario

University of Windsor thinking forward

Motivation

Anthrax found

Seven dead from E. coli contamination

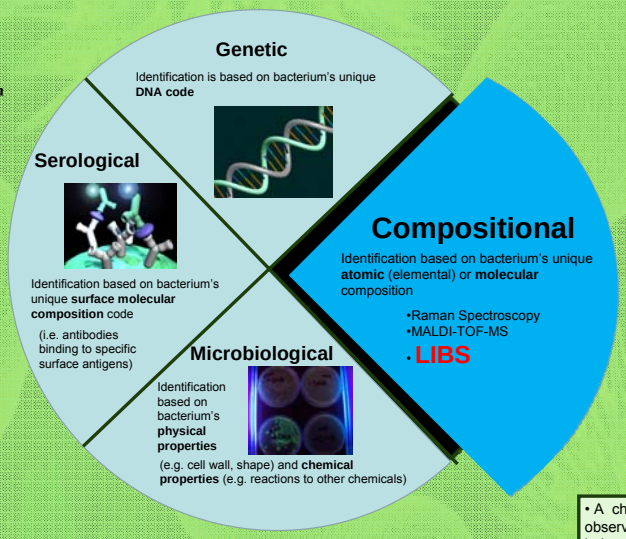
New Strep strain moving

B.C. hospital declares Strep A outbreak

The air we breathe, the food we eat, the people we meet all carry bacterial microorganisms. Some essential for human life, some causing petty illnesses such as **strep throat** or **food poisoning**, and some causing pandemics such as the **"black death"** and **cholera**.

With (1) **emerging threats of bioterrorism**, (2) **ever-present epidemics**, and the increasing threats of (3) **antibiotic-resistant bacteria**, and (4) **food-borne contamination**, methods to rapidly identify pathogenic bacteria (which are essential for human health and safety) will become increasingly important in the future. With the current speed of traditional microbiological techniques of identifying bacteria, there is a critical need for new technologies to rapidly classify them.

How are Bacteria Identified?



Our Chosen Approach: LIBS

Laser-induced breakdown spectroscopy

A laser-based technique that offers a quicker way to detect harmful pathogens in **real-time**.

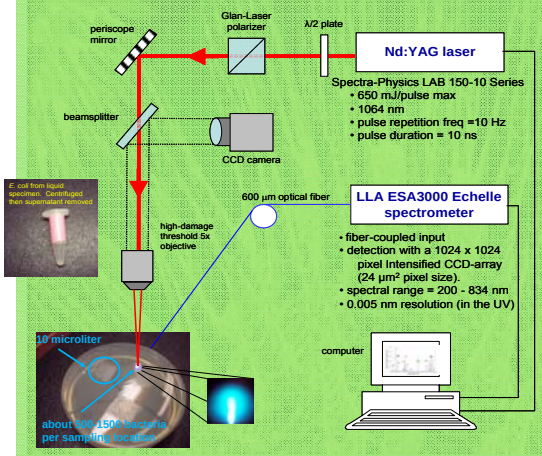
Process

1.An incident pulsed laser deposits its energy in a target sample and atomizes it. A high-temperature micro-plasma results.

2.Light from the spontaneous emission of atoms in the plasma (atomic emission spectroscopy) is collected and dispersed by a spectrometer.

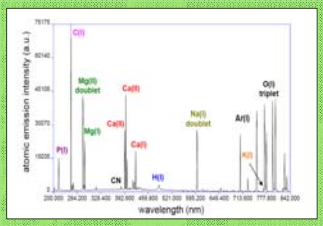
3.The intensity of observed atomic emission peaks correlates to the relative concentrations of the elemental constituents of the original sample .

Experimental Set-Up



Q: How does THIS identify a bacterium?

Element	% of fixed salt fraction
Na	2.6
K	12.9
Ca	9.1
Mg	5.9
P	45.8
S	1.8
Fe	3.4

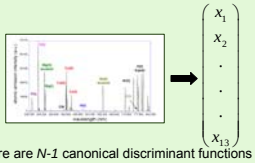


Discriminant Function Analysis

• A chemometric analysis technique used for classifying a set of observations into mutually exclusive groups (based on a set of independent variables)

• Can take the entire spectrum and reduce it to a data point

1. Intensities of emission lines are normalized by the sum of all intensities and separated into *N* groups



2. There are *N-1* canonical discriminant functions constructed:

$(b_0^1, b_1^1, b_2^1, \ldots, b_{13}^1) \rightarrow (b_0^{N-1}, b_1^{N-1}, b_2^{N-1}, \ldots, b_{13}^{N-1})$

3. For each spectrum, *N-1* **discriminant function scores** are calculated:

$$DF^j = b_0^j + \sum_{k=1}^{13} b_k^j x_k$$

4. The discriminant function score for each spectrum is plotted as data point on a graph.

Methodology

1. External-Validation

•A library of **five bacterial genera** with 32 sets of data was tested using **external validation techniques**

•Each of the 32 unique sets of data was withdrawn, individually, from the five-genera model and entered as unclassified data.

•The DFA then assigned these spectra to one of the five genus classes

Leave-One-Out Cross Validation

•A library of five bacterial genera with 32 sets of data was tested using cross validation

•One specimen at a time withheld from the model and than tested in the model (containing data obtained at the same time) yielding higher (but incorrect) accuracies

2. •Previously obtained spectra from five bacterial genera groups (***Escherichia***, ***Enterobacter***, ***Streptococcus***, ***Mycobacterium***, and ***Staphylococcus***) were compiled into a spectral library

•Each spectrum reduced to a spectral fingerprint containing 13 emission line intensities (independent variables)

•These spectral fingerprints were analyzed by a commercial DFA program **(SPSS Inc., SPSS v19.0)** to construct the canonical discriminant functions

•These canonical discriminant functions were tested using external validation techniques and then compared to "leave-one-out" cross-validation tests

Results

External Validation Truth Tables

Escherichia	TRUE	FALSE
Positive	89.97%	4.28%
Negative	95.72%	10.03%

Streptococcus	TRUE	FALSE
Positive	83.82%	2.04%
Negative	97.96%	16.18%

Staphylococcus	TRUE	FALSE
Positive	62.16%	2.55%
Negative	97.45%	37.84%

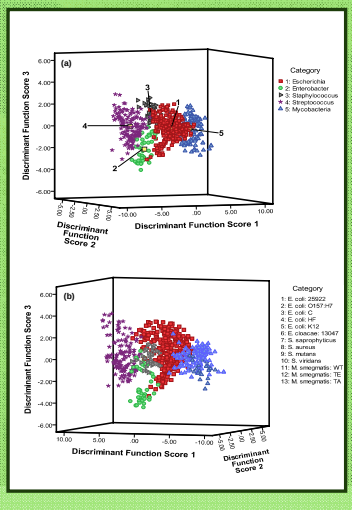
Mycobacterium	TRUE	FALSE
Positive	89.61%	1.27%
Negative	98.73%	10.39%

89.9% of *Escherichia* classified as ***Escherichia***

10.0% of *Escherichia* was classified as **something else**

95.7% of the other classes were **not classified** as *Escherichia*

4.28% was classified as ***Escherichia***



My Project

1. Investigate ways of quantifying the accuracy (sensitivity and specificity) of such a chemometric classification.

2. Utilizing a computerized chemometric technique (discriminant function analysis) investigate whether any repeatable differences can be quantified between bacteria of different genus, species, or strain.

3. Investigate alternate chemometric techniques to see if accuracy can be improved (e.g. PLS-DA)

Conclusions and Future Work

3. •LIBS using DFA is a promising identification technique, but other chemometric techniques must be investigated

•Partial Least Squares Discriminant Analysis constructs predictor variables that maximize the variance between classes

•Classification of a sample is based on a value of 0 and 1 (a value closer to 1 is indicative of the class being closer to the modelled class)

•Future work will compare PLS-DA to DFA to search for improvements in accuracies

Acknowledgements

•The University of Windsor Physics Department

•My advisor: Dr. Steven J. Rehse