A novel genetic algorithm for MALDI ToF mass spectrometry based identification of bacterial species.

E.S. Lisitsyna[1], I.A. Altukhov[2], D.S. Ischenko[2], D.G. Alexeev[2], A.U. Naumov[1], E.N. Ilina[2], V.M. Govorun[2]
[1]LYTECH Co LTD, Moscow, Russia
[2]Federal Institute and Clinical Centre for Physical-Chemical Medicine, Moscow, Russia

26th ECCMID Amsterdam, Netherlands 9 – 12 April 2016
The congress of ESCMID
EV0502

Introduction

MALDI ToF mass spectrometry technique is widely used for bacterial analysis and profiling nowadays. Actual commercial solution is restricted by necessity of preliminary collecting set of referential proteomic profiles. We have developed an original genetic algorithm for identification of bacterial species based on the abundance statistics of orthologous protein groups in MALDI mass spectra and genomic information from the database. The most representative groups in spectra were found to be ribosomal proteins, heat-shock proteins, cold-shock proteins, HU-proteins, elongation factors (fig.1). The proteins encoded genomic sequences have been uploaded from the PATRIC database representing more than 4000 species.

The algorithm compares m/z values of MALDI mass-peaks and a set of theoretical m/z values of proteins of a given organism. The principle of analysis is showed on scheme (fig. 2). Processing MALDI mass spectra and peak picking was carried out via R package "MALDIquant". Identification of orthologous groups between different microbial proteins was done using the software package OrthoMCL. Matching peaks with corresponding proteins according to their masses was performed using ad hoc algorithm developed on a programming language R. The probability of a match between an unidentified species with known MALDI mass spectrum and a species from the genomic database is predicted by a scoring function.

In our study we demonstrated the possibility of applying the genetic algorithm for MALDI ToF MS based identification of bacterial species in laboratory setting.

Methods

In total the 507 isolates of clinically relevant bacteria were involved. They were represented by 54 different species belonged to *Staphylococcus spp., Enterococcus spp., Streptococcus spp., Klebsiella spp., E. coil, Enterobacter spp., Proteus spp., Serratia spp., Acinetobacter spp., Pseudomonas spp., Candida spp., Neisseria spp., Lactobacillus spp., Stenotrophomonas spp., Citrobacter spp., Morganella spp., Bacillus spp.*

Sample preparation for MALDI TOF MS and identification was carried out according to a formic acid/acetonitrile extraction protocol suggested by Bruker Daltonics (Germany).

Mass spectra were obtained by an Autoflex LT instrument (Bruker Daltonics, Germany). Additionally, mass spectra for the same bacterial extracts were aquired on a LaserToF TT (SAI, UK) mass spectrometer.

Species identification was done using the MALDI Biotyper 3.1 software (Bruker Daltonics, Germany). The genetic algorithm was used to analyze mass spectra collected by both Autoflex LT (Bruker Daltonics) and LaserToF TT (SAI) mass spectrometers.

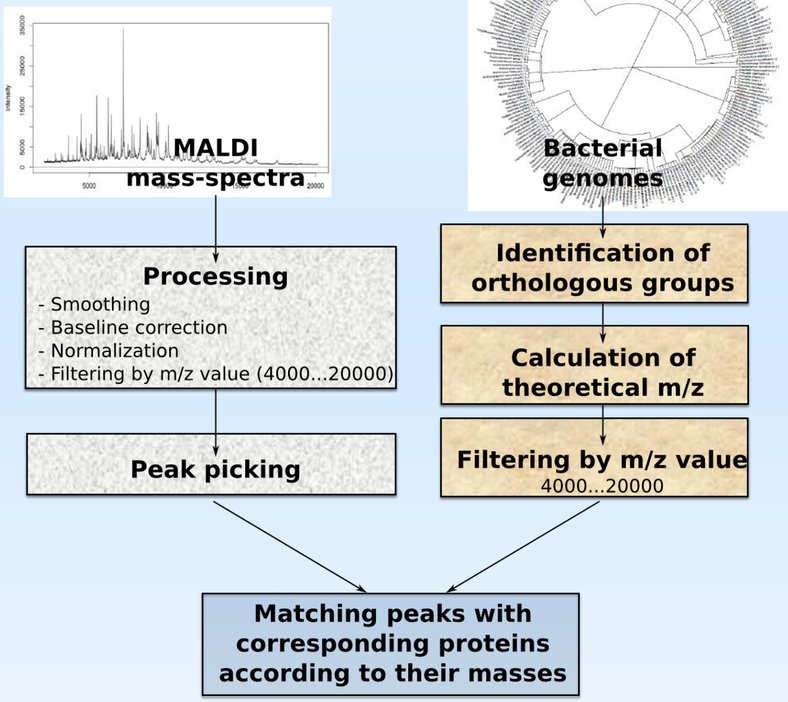

Figure 2. Scheme of MALDI-ToF MS/genetic based bacterial identification.

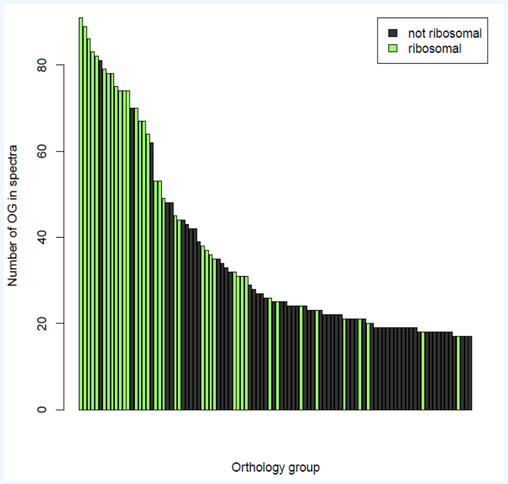

Figure 1. Frequency of occurrence proteins in mass spectra.

Results

- All of 507 bacterial isolates were reliably identified to species level by Biotyper 3.1 software (score higher then 2,0)
- Mass spectra obtained using the Autoflex LT as well as the LaserToF TT (in total 1014 spectra) were analized via the genetic algorithm followed by species identification. In both cases the analysis of mass spectra gave the similar results with suitable score
- To evaluate an accuracy of genetic algorithm the results of Bruker's identification of 507 isolates were used as a reference.
- In case of mass spectra collected by Autoflex LT (Bruker Daltonics) the genetic algorithm correctly identified **95.9%** (486/507) isolates to species level and **98.6 %** (500/507) to genus level.
- For mass spectra obtained by LaserToF TT (SAI) the genetic algorithm showed the identical species identification in **94.1%** (477/507) cases and in **98.2 %** (498/507) there were matches at the genus level.
- The most common mismatches were *Klebsiella oxytoca* vs *Raoultella ornithinolytica*, *Stenotrophomonas maltophilia* vs *Pseudomonas geniculatacan*. It can be explained by close related species and constant changes in modern taxonomy of bacterial species or probable mixed bacterial culture.

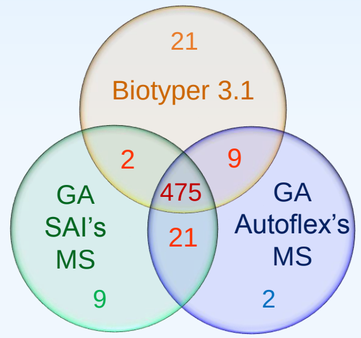

Figure 3. Matches between Biotyper 3.1 and the genetic algorithm (GA) in identification of bacterial species

Conclusions

- We provide the genetic algorithm that identifies a growing range of organisms and, unlike other solutions, requires no additional mass spectrometric dataset.
- The identification based on MALDI-ToF MS and genomic database is unique and allows to acquire more accurate and biologically informative results.
- The genetic algorithm is compatible with different MALDI ToF mass spectrometry platforms remaining high sensitivity and specificity method for bacterial species identification

Corresponding author: Lisitsyna Evgeniia; Lytech Co LTD, 107023 Russian Federation, Moscow, 3a Malaya Semenovskaya Street, email: lisitsinaes@gmail.com