P1669
Abstract (poster session)
Contribution of freely available databases for the accurate classification of emergent pathogens

I. Joao*, C. Paula, L. Jordao (Lisbon, PT)

Objectives: Nontuberculous mycobacteria (NTM) are a heterogeneous group of microorganisms. Their clinical relevance and treatment differs significantly upon NTM species, which makes identification at this level crucial. Additionally, inaccurate diagnosis can lead to therapeutic approaches consistent with *Mycobacterium tuberculosis* infection which are usefulness.

In the present study different molecular methods were used in the identification of NTM. The aim was to evaluate and compare the performance of freely available databases in the accurate identification of NTM.

Methods: Partial sequencing of hsp65 and 16S rRNA genes of 54 strains of NTM. The resulting sequences were compared with the nucleotide sequences from GenBank, RIDOM and EzTaxon databases for 16S rRNA gene. For hsp65 gene GenBank and Web Databases were used.

Results:
Seventy-two percent of NTM were identified to species level by 16SrRNA analysis. The remaining 28% failed this goal. Thirty-three obtained an unique outcome in all databases and 6 isolates (15%) obtained at least one different result

Analysis of hsp65 gene sequences identified unambiguously 76% of the strains, for the remaining 22% the analysis led to the possibility of classification into more than one species. Of all the strains analyzed, this approach failed to identify only 2%.

Conclusion: The obsolescence of databases, including the GenBank and RIDOM, made the identification of emerging pathogens, as NTM, very complex.
If instead of allowing free deposit it was mandatory to update the taxonomy of mycobacteria and asserting the quality of sequences prior to deposit, e.g. in GenBank, several problems would be avoided. Among the shortcomings, we highlight the existence of non-validated "species" names, poorly classified species and even classification limited to genera. As there is no control of sequences submitted to this database, identical sequences could give different percentages of similarity. Furthermore, this database contains an enormous amount of flawed sequences, which can lead to identification errors.
On the other hand controlled databases which contain high quality and properly characterized sequences, have a limited usefulness due to lagging updating (RIDOM), and the absence of sequences other than from reference strains (EzTaxon and Web Database).

Keywords: Nontuberculous mycobacteria (NMT), identification, freely available software