

CODON AND AMINOACID USAGE PATTERNS IN MYCOBACTERIA.

(accompanying file)

By

Scapoli C, Bartolomei E, Di Lorenzo S, Carrieri A, Salvatorelli G,
Rodriguez-Larralde A¹, Barraï I

Department of Biology, University of Ferrara

¹Centro de Medicina Experimental,
Laboratorio de Genetica Humana, IVIC, Caracas, Venezuela

ABSTRACT

Background

Codon usage can be used as a marker for molecular evolution of different organisms. The large number of organisms and of genes sequenced at the present time permits now to study molecular evolution in lower clades such as genera, species, and subspecies. Here, we focus our attention on the genus *Mycobacterium*, in which we examined codon and aminoacid usage in 13 species, and in 12 subspecies from three of the former, for a total of 26,755 sequences and 8,836,513 codons.

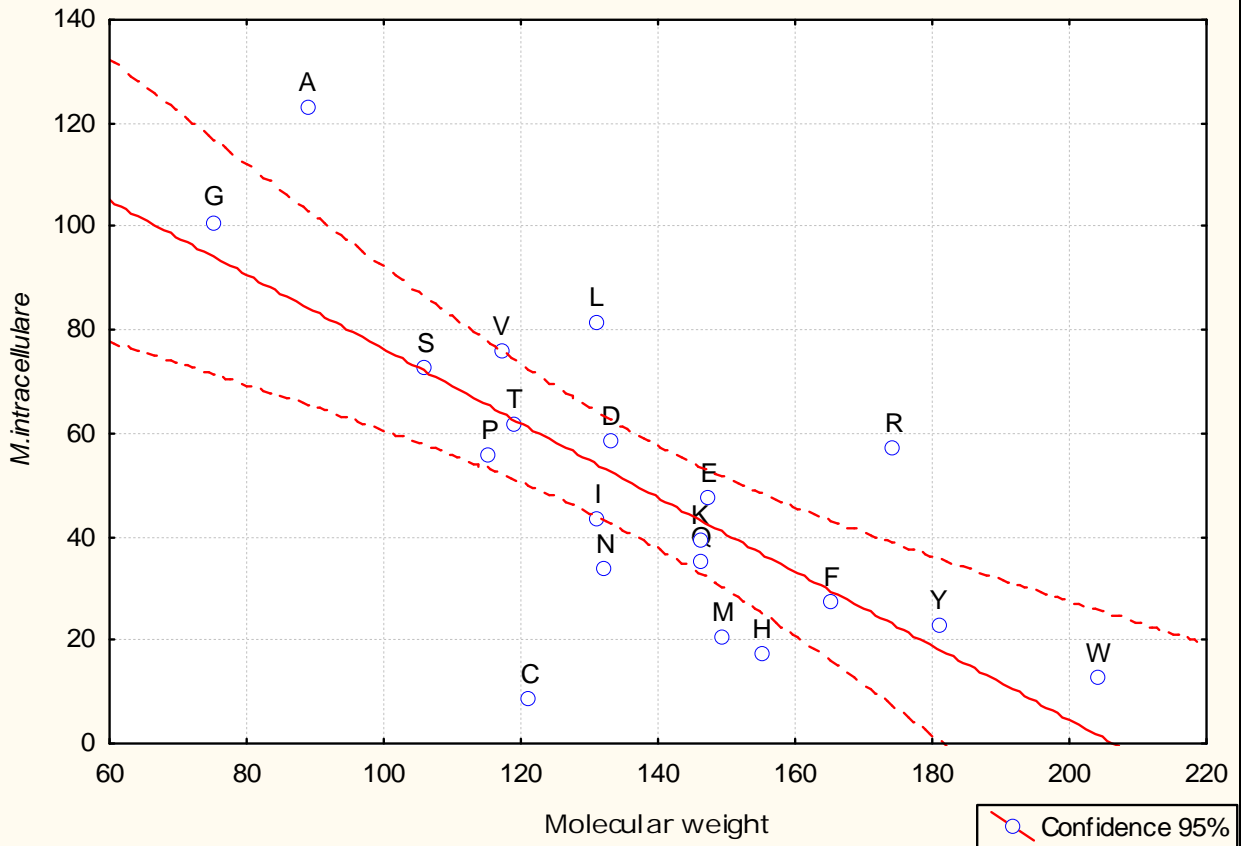
Results

Frequencies of codon and aminoacid usage correlate between species and between subspecies within the genus *Mycobacterium*. The analysis of usage within the genus identifies *M. avium* and *M. bovis* as the closest, and *M. abscessus* and *M. ulcerans* as the most diverse species. In the groups studied, aminoacid molecular weight and codon degeneracy influence correlations between frequencies, while GC content seems the main factor influencing codon effective number.

Conclusion

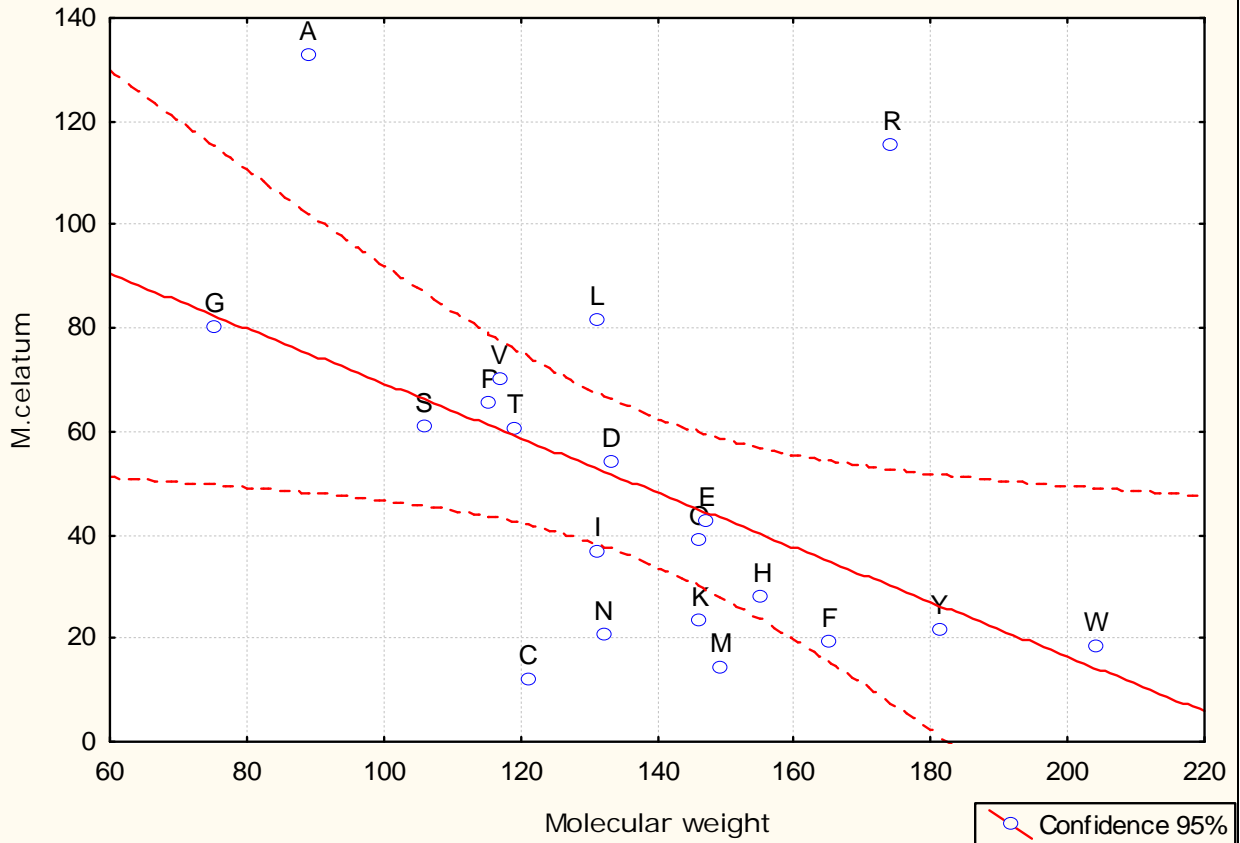
Coding GC content, which is highly correlated with total genomic GC content, seems to be the main factor influencing present synonymous codon usage in the genus *Mycobacterium*. In particular, the GC content at the 3rd base position seems to shape heavily the codon effective number, N_E , giving indication that mutational bias dominates over translational selection in this genus. Evolutionary trees based on codon and aminoacid usage are consistent with traditional phylogenies of species within the genus.

Regression of the aminoacid frequency in *M.intracellulare* on molecular weight
 $r = -0,73$



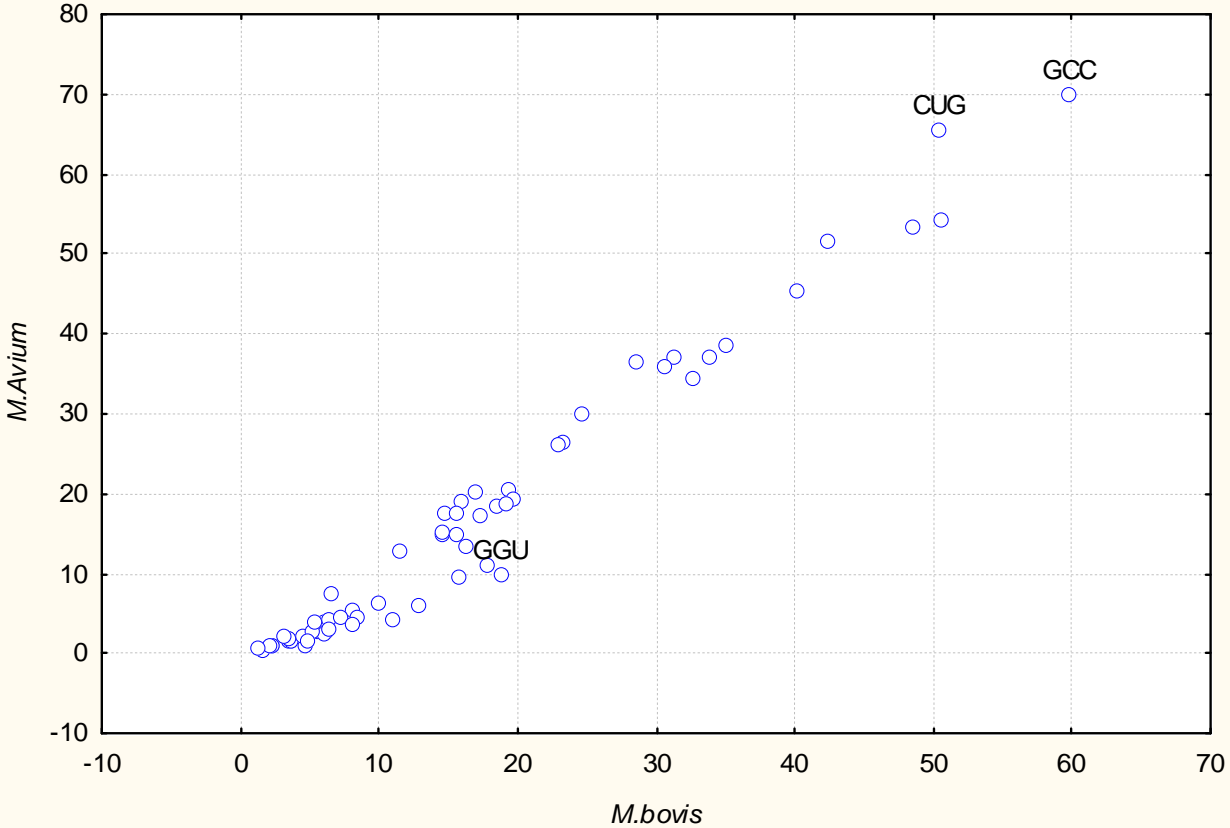
Regression of aminoacid frequency in *M.celatum* on molecular weight

$r = -0,48$



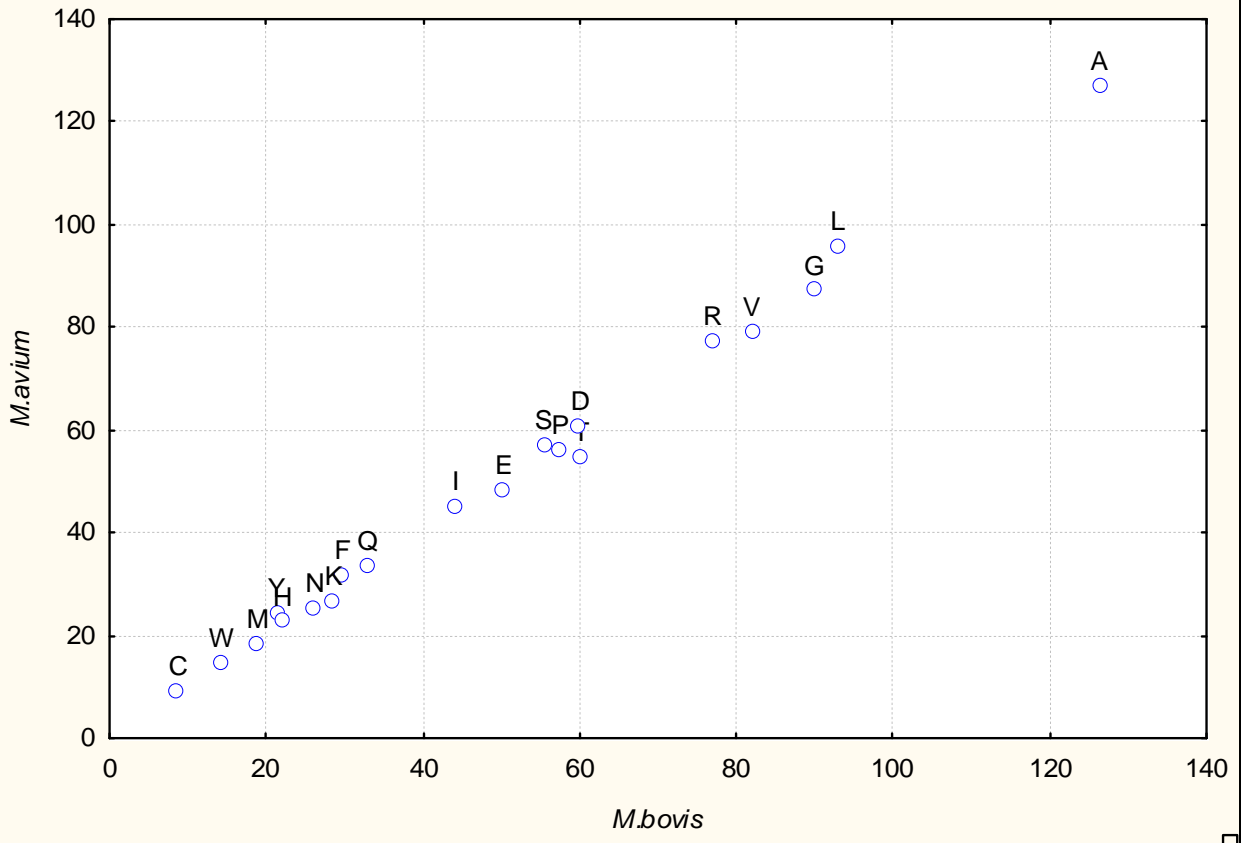
Correlation between the frequency of codons in *M.bovis* and *M.avium*

$r = 0,98$



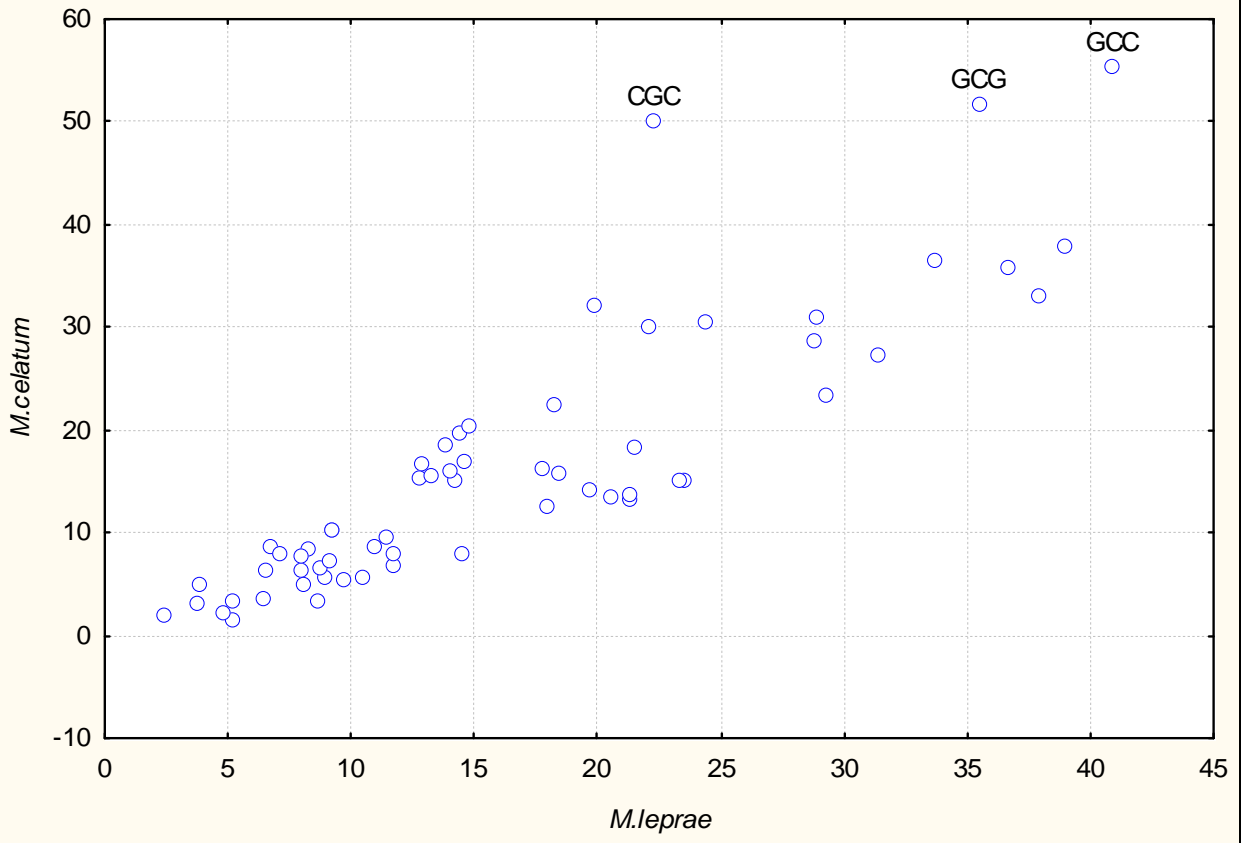
Correlation between the aminoacid frequencies in *M.bovis* and *M.avium*

$r = 0,99$



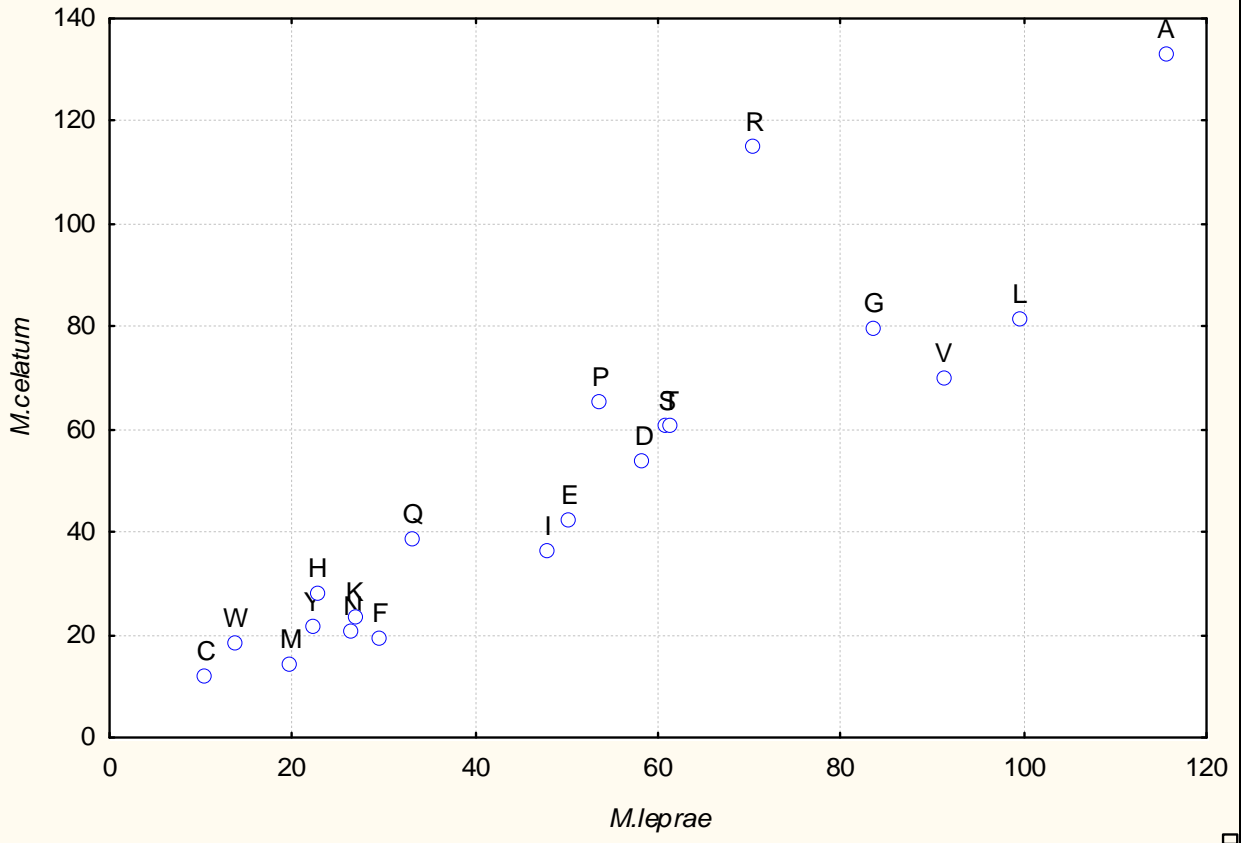
Correlation between the frequency of codons in *M.celatum* and *M.leprae*

$r = 0,87$



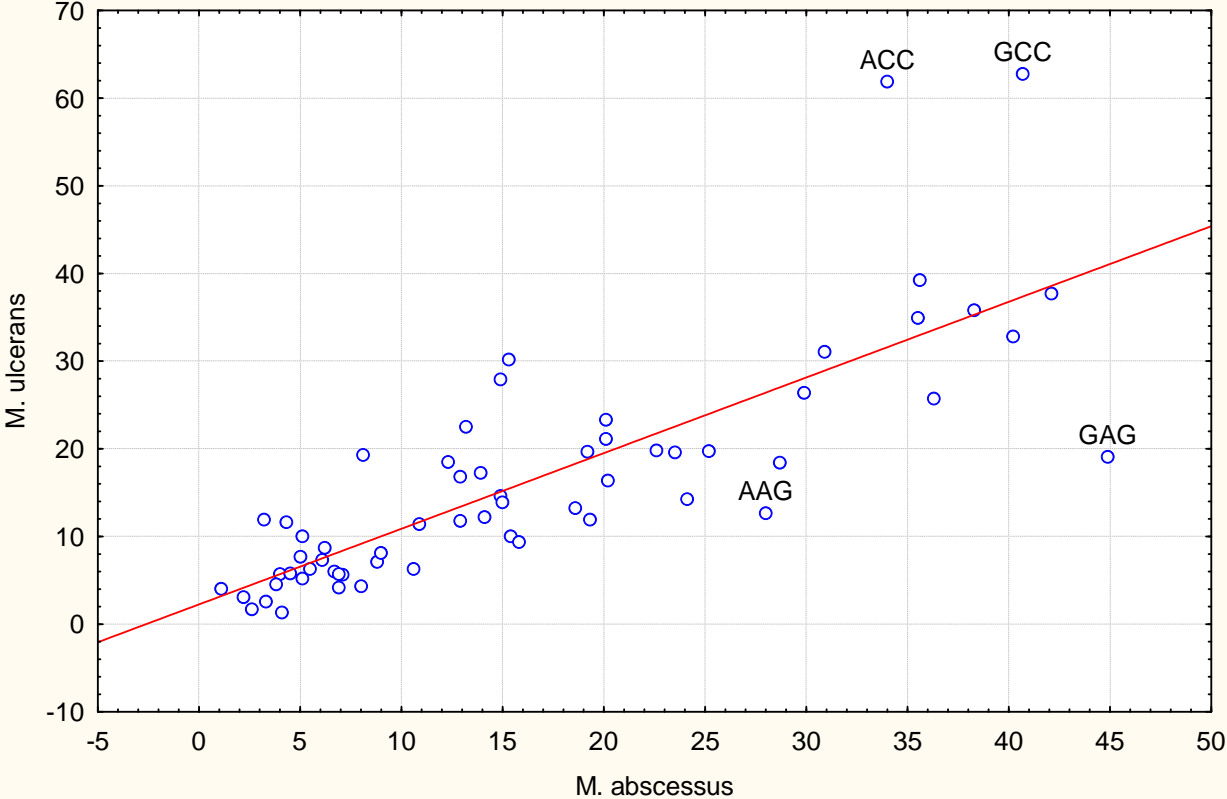
Correlation between the aminoacid frequencies in *M.leprae* and *M.celatum*

$r = 0,91$



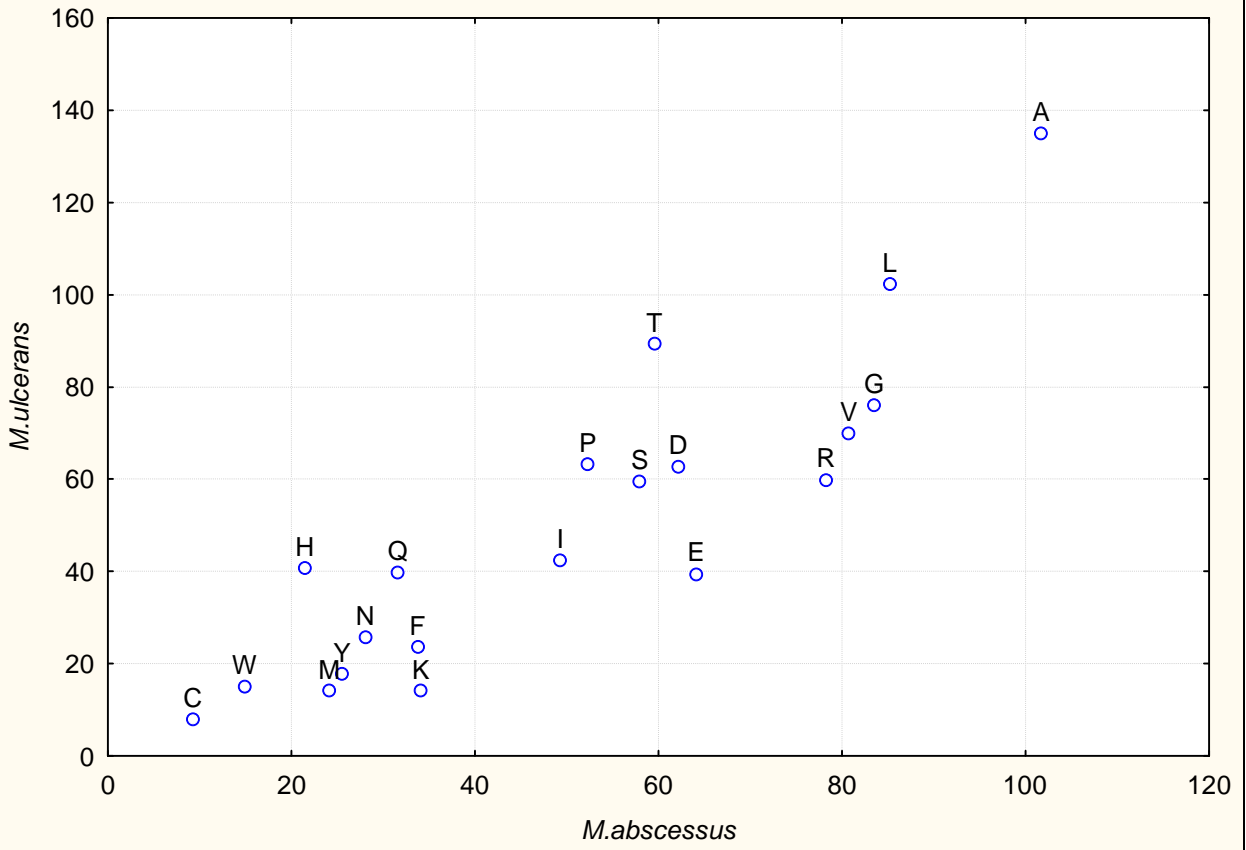
Correlation between the frequency of codons in
M. abscessus and M. ulcerans

$r=0.79$

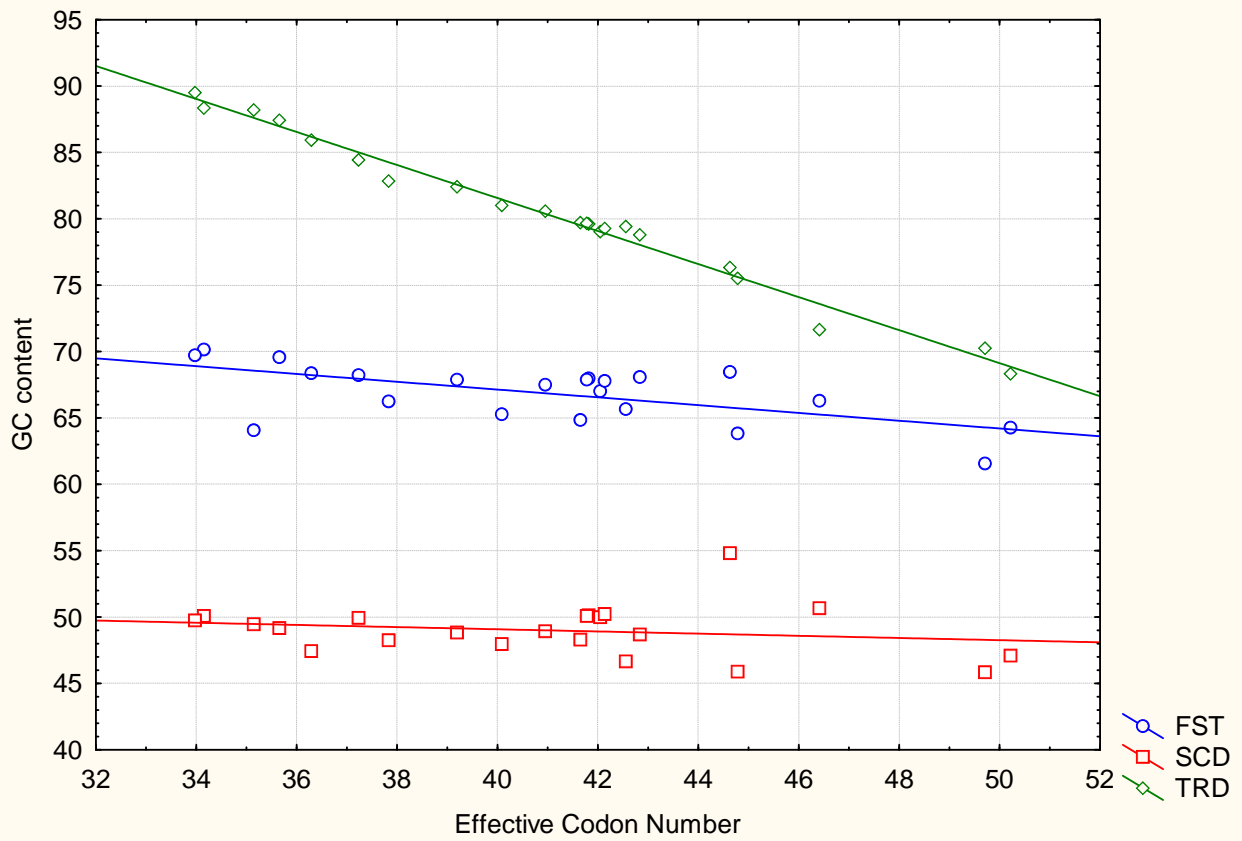


Correlation between the aminoacid frequencies in *M.abscessus* and *M.ulcerans*

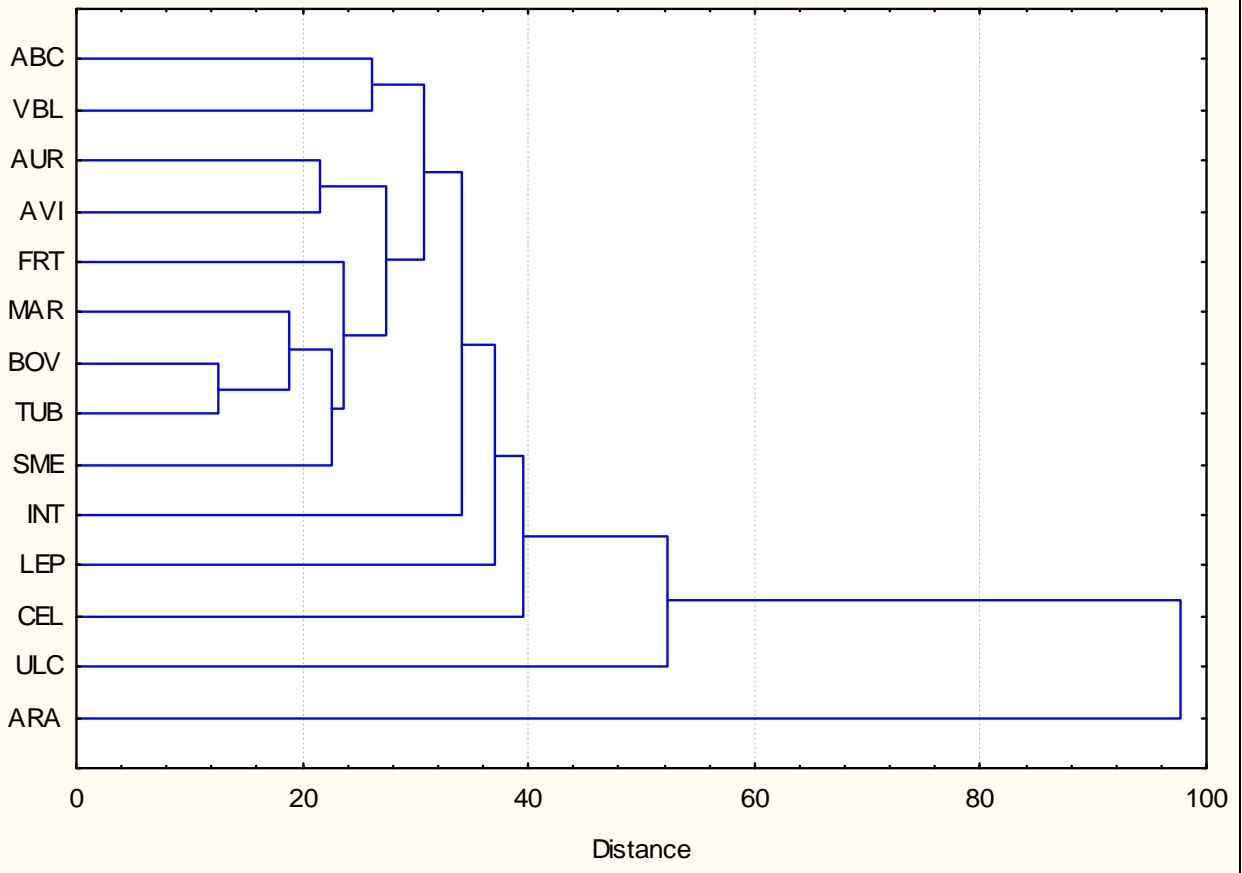
$r = 0,88$



Variation of N_E over the GC content in 1st, 2nd and 3rd letter in 23 species and strains of Mycobacteria



Dendrogram for 13 species of *Mycobacterium* and for *Arabidopsis thaliana*, based on codon usage



Dendrogram for 13 species of *Mycobacterium* and for *Arabis thaliana*, based on amino acid usage

