Analysis of horizontal transfers in the *Mycobacterium tuberculosis* complex in relation with pathogenicity emergence

Jennifer Becq[1], Vania Rosas-Magallanes[2], Brigitte Gicquel[2], Olivier Neyrolles[2] & Patrick Deschavanne[1]

[1] Equipe de Bioinformatique Génomique et Moléculaire, Inserm U726, Université Paris 7, France - [2] Unité de Génétique Mycobactérienne, Institut Pasteur, France

becq@ebgm.jussieu.fr

Purpose

Some pathogenicity islands are transferred horizontally between bacteria species.

What is the importance of horizontal transfers in pathogenicity emergence in *Mycobacterium tuberculosis*-complex, the causative agents of tuberculosis in humans and bovines ?

To address this issue, we :

- searched for horizontal transfers (HT) in MT-complex,
- attributed potential source species for these HTs,
- conducted genomic comparison within *M. tuberculosis* complex and with *M. marinum* (which has a different host specificity).

Abstract

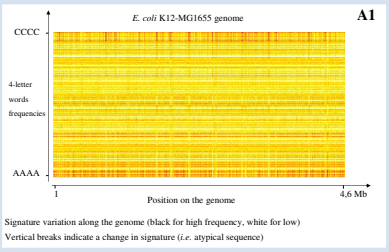
M. tuberculosis and *M. bovis* are the causative agents of tuberculosis in humans and bovines and are parts of the ***M. tuberculosis* complex (MT-complex).** Determination of the complete nucleotide sequences of the ***M. tuberculosis* H37Rv, *M. tuberculosis* CDC1551** and ***M. bovis* AF2122/97** genomes has provided means to compare these genomes and understand their evolution.

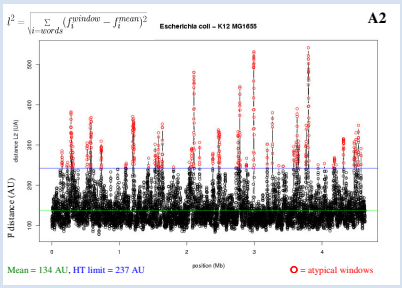
Horizontal gene transfer (HGT) between species is thought to be a factor driving genetic diversity, bearing medical interest when it concerns virulence genes. Because different surrogate methods (based on the fact that foreign sequences to a genome bear unusual sequence characteristics) often fail to identify identical sets of HGTs[1,4], we developed a consensus-based method to predict recent horizontally acquired genes. A gene was considered as a potential HGT when it was detected by at least 2 of the 3 methods used: **abnormal GC1% and GC3%[2], atypical codon usage[3]** and **different genomic signature[1].** Genomic signature can also be used to propose potential donor species when looking for similar signatures in a database of genomic signatures.

About 180 to 225 regions of the studied genomes can be considered as horizontal transfers. Their size ranges from 300 pb to 10 kb (total of ~300 kb per genome) and they each contain 1 to 10 genes. The signature method allowed to attribute potential donors to about one third of the HGT detected. The major donor species are Actinomycetales (mostly Mycobacteria) followed by Rhizobiales (more precisely the Agrobacterium genus).

Comparison with *M. Marinum* should improve our understanding of the emergence of pathology and host specifity in the MT-complex.

Detection of horizontal transfers





Consensus method

A. Genomic signature method[1]

(i) Frequencies of all 4-letter words (*i.e.* the signature) are calculated in slidding windows (size=5 kb, step = 0.5 kb) along the whole genome *(fig. A1)*

(ii) Euclidian (= l2) distance is calculated between each window signature and the whole genome signature

(iii) Windows having a distance higher than a certain threshold[1] is considered as a atypical sequence *(fig. A2)*

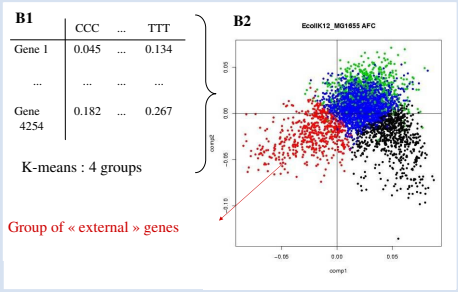
A gene is considered atypical if 2 out of the 3 methods detect it

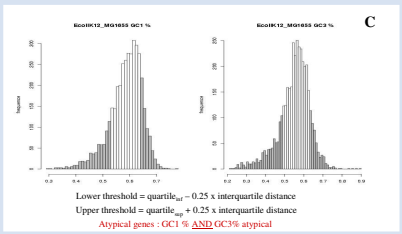
B. Codon usage method[3]

(i) Codon usage frequencies are calculated for each gene over 300 pb *(fig. B1)*

(ii) A Factorial Correspondence Analysis (FCA) of these frequencies is then generated *(fig. B2)*

(iii) A 4-group k-means is conducted in order to differenciate atypical genes from highly or normally expressed genes *(fig. B2)*





C. GC1 % and GC3 % method[2]

(i) G+C % in position 1 and 3 of gene codons is calculated

(ii) Thresholds are calculated as described in figure C

(iii) Genes having atypical GC1% and GC3% are considered as atypical.

Horizontal transfers

atypical regions *(complete genome)*	Number	Total size (kb)	Mean size (kb)	Number of genes	G+C %
M. bovis	187	306 *(4.35 Mb)*	1.6	309 *(3953)*	63.7 *(65.6)*
***M. tub.* CDC1551**	226	327 *(4.4 Mb)*	1.5	374 *(4189)*	64.2 *(65.6)*
***M. tub.* H37Rv**	180	298 *(4.41 Mb)*	1.7	288 *(3999)*	63.3 *(65.6)*

• Atypical sequences represent around 7 % of the genomes. Sequence analyses of these regions support their HT status. Atypical sequences are organised in ~200 « regions » from 300 pb to 10 kb, each containing 1 to 10 genes.

• Even though G+C% of these regions is always slightly lower than the whole genome G+C%, these potential HT are still G+C rich.

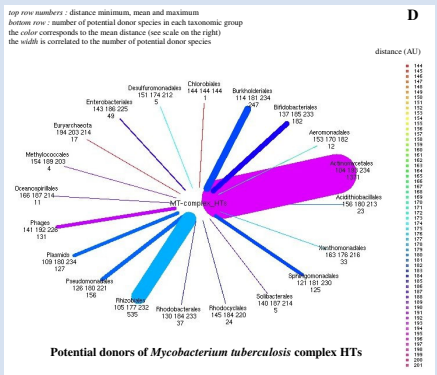
Donor species

• We created a genomic signature database (~25 000 species) using every sequences over 1,5 kb available from genbank.

• Potential « source » species can be attributed to HT islands by looking for the nearest neighbours using euclidian distance[1].

• The potential donor species of all the HT regions have been classified in taxonomic groups and displayed on a circle *(fig. D)*.

• Actinomycetales (especially Mycobacteria) is the biggest pool of donors. Another major origin for HTs is found in Rhizobiales, which is not obvious considering their biotope.



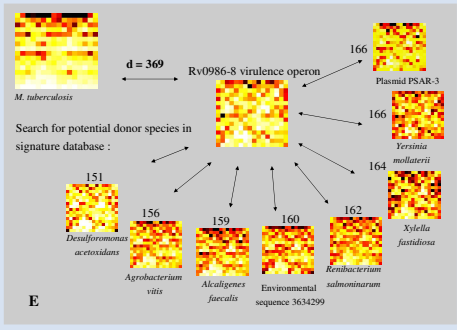
Conclusion and Perspective

HT regions in *Mycobacterium tuberculosis* – complex genomes may not be as rare as previously thought: 7 % of the genomes is considered as horizontally transfered.

Actinomycetales and Rhizobiales species are the most frequent potential sources, suggesting ancient and frequent interaction between Mycobacteria and Rhizobiales.

The virulence operon Rv0986-8, present in all 3 MT-complex genomes, is a pathogenicity island horizontally transferred potentially from a γ-proteobacteria[5].

Comparison with *M. marinum*'s genome (MT-complex closest relative) shows that some identical HT-islands in the 3 genomes could not be found in *M. marinum*. Further investigation should improve our understanding of pathogenic emergence and host specificity of MT-complex species.



• For illustration, potential donor species of a virulence operon have been searched *(fig. E)*.

• Further phylogenic analysis of the 30 closest signatures suggests that this virulence operon is a HT, probably plasmid transferred from a γ-proteobacteria[5].

• This virulence operon has a original source, very few γ-proteobacteria being potential donor species in *fig. D*.

References

[1] C. Dufraigne, B. Fertil, S. Lespinats, A. Giron and P. Deschavanne, **Detection and characterization of horizontal transfers in prokaryotes using genomic signature**. *Nucleic Acids Res.*, 33(1):e6, 2005.

[2] J. Lawrence and H. Ochman, **Molecular archaeology of the *Escherichia coli* genome**. *PNAS*, 95:9413-9417, 1998.

[3] C. Médigue, T. Rouxel, P. Vigier, A. Hénaut and A. Danchin, **Evidence for Horizontal Gene Transfer in *Escherichia coli* speciation**. *J. Mol. Biol.*, 222:851-856, 1991.

[4] M. Ragan, **On surrogate methods for detecting lateral gene transfer**. *FEMS Microbiology Letters*, 201:187-191, 2001.

[5] V. Rosas-Magallanes, P. Deschavanne, L. Quitana-Murci, R. Brosch, B. Gicquel and O. Neyrolles, **Horizontal transfer of a virulence operon in the ancestor of *Mycobacterium tuberculosis***. *J. Mol. Evol.* 23(6):1129-1135, 2006.

Acknowledgements: M. marinum sequence data were produced by the M. marinum Sequencing Group at the Sanger Institute and can be obtained from http://www.sanger.ac.uk/Projects/M_marinum/.