

# Comparative genomics of the genus *Burkholderia*: insights into pathogens evolution

Maida I<sup>(1)</sup>, Perrin E<sup>(1)</sup>, Papaleo MC<sup>(1)</sup>, Emiliani G<sup>(2)</sup>, Fani R<sup>(1)</sup>, Fondi M<sup>(1)</sup>

<sup>(1)</sup> Department of Evolutionary Biology, University of Firenze, Firenze

<sup>(2)</sup> Department of Environmental and Forestry Sciences, University of Firenze, Firenze

## Motivation

One of the main questions of modern microbiology is how a micro-organism adapts to a pathogenic lifestyle. Pathogens biodiversity is a source of inference regarding disease processes and the evolution of pathogens and pathogenesis. Besides, the structure of pathogen populations has important implications for public health interventions, such as vaccination and/or antibiotic usage. The computational analysis of the hundreds of completely sequenced genomes from several bacterial species, as well as from multiple strains of the same species permitted to shed some light on the physiology and pathogenicity mechanisms, also providing insights into the forces driving gene and genome evolution. This was based on the “comparative genomics approach”, which aims at

- i) identifying the differences existing between genomes,
- ii) correlating those differences to biological traits, and
- iii) gaining clues about selective pressures and patterns of gene transfers/losses.

It is also well established that in some cases the genetic determinant for virulence and/or pathogenicity are located on plasmid molecules of different size. Among pathogenic (micro)organisms, members of the genus *Burkholderia* (-proteobacteria) are particularly interesting. Indeed, this genus includes a variety of species occupying different ecological niches such as soil, water, plant and human hosts, with strains able to promote plant growth and/or to degrade pollutants, as well as opportunistic human pathogens (including the so-called *B. cepacia* complex [Bcc]). It is quite interesting the finding that there are indirect evidences that the environment may serve as a reservoir for the appearance of novel strains capable of infecting hosts and responsible for human infections. However, little is known about factors involved in *Burkholderia* virulence. Therefore, the aim of this work was to perform a comparative analysis of all the available completely sequenced *Burkholderia* chromosomes and plasmids in order to gain further insights into different issues, such as functional diversity, genome evolution and pathogenesis-related mechanism(s).

## Methods

All the available *Burkholderia* spp. chromosomes and plasmids were retrieved and used as seeds to probe the functional databases Pfam and COG and virulence factor database LLNL using ad hoc written Perl codes. The evolutionary dynamics of *Burkholderia* plasmids and chromosomes were studied using the software Blast2Network, which permits

- i) the immediate visualization/estimation of evolutionary relationships in a (large) dataset of amino acid or nucleotide sequences, and
- ii) the construction of phylogenetic profiles resembling plasmids similarity (in terms of shared genes) and proteins co-occurrence.

## Results

A deep comparison of the 21 *Burkholderia* available genomes was performed. Data obtained revealed that, in spite of the existence of a large conserved “core” of genes, the analysis of the distribution of functional categories across all the genomes revealed some differences that can be mainly attributed to two functional categories, i.e. DNA transcription and recombination. This finding suggests that both gene expression and genome rearrangements may be responsible for the functional and genomic biodiversity observed in the genus *Burkholderia*. This idea is in agreement with previous observations and with the finding that many *Burkholderia* species harbour extensive arrays of mobile elements that are involved in the above mentioned processes and in the shaping of genomes during evolution. The existence of peculiarities in the protein repertoire of different strains might be due to the adaptation to a specific environment and/or lifestyle. The same analyses were performed also on the whole dataset of *Burkholderia* plasmids. As might be expected, plasmids showed, on average, a number of proteins involved in DNA transposition, integration and replication higher than that found in chromosomes. Interestingly, our analysis showed that they also carry transport and regulation related proteins

with a percentage (5% and 3.7%) similar to the corresponding ones found in chromosomes. Thus, it is possible that plasmids may play an active role in conferring particular biological traits to the cells harbouring them. Furthermore, the building up of similarity networks embedding both Burkholderia plasmids and chromosome provided a detailed insight into the evolutionary relationships existing among them. Lastly, the analysis of virulence factors across the Burkholderia chromosomes and plasmids revealed their variable distribution. Exploring the functional role of the virulence factors retrieved in each genome, might provide useful hints for the understanding of the possible pathogenic mechanisms and underlined the importance of such an in silico approach in the fields of risk control and biotechnological use of bacteria belonging to the genus Burkholderia.

**Contact :** marco.fondi@unifi.it