

**Comparative Mycobacterial Genomics: uncovering the secrets of the genus *Mycobacterium***

<sup>1</sup>Gey van Pittius, N.C., <sup>2</sup>Sampson, S.L., <sup>1</sup>van Helden, P.D., and <sup>1</sup>Warren, R.M.

<sup>1</sup>NRF Centre of Excellence in Biomedical Tuberculosis Research / MRC Centre for Molecular and Cellular Biology, Department of Medical Biochemistry, University of Stellenbosch, Tygerberg, South Africa, <sup>2</sup>Department of Immunology and Infectious Disease, Harvard School of Public Health, Boston, USA

The advent of whole genome sequencing has opened up a new field of science named comparative genomics. By comparing the genomes of the 11 sequenced mycobacterial species and strains, the well-guarded secrets of the genus *Mycobacterium* is slowly being uncovered. One of these secrets involve the duplication of two groups of genes, the PE and PPE multigene families, as well as the ESAT-6 gene cluster regions. The PE and PPE gene families comprise 10% of the coding potential of the genome of *Mycobacterium tuberculosis* and are thought to be involved in antigenic variation due to the highly polymorphic nature of their C-terminal domains. Members are also frequently found within the immunologically-important ESAT-6 cluster regions, which have been shown to be involved in virulence. Previous work have indicated that the expansion of the PE and PPE families is linked to the duplications of the ESAT-6 clusters. This is supported by the absence of a pair of PE and PPE proteins within the ancestral ESAT-6 cluster, indicating that these genes have been integrated and subsequently co-duplicated within the clusters. Using a combination of phylogenetic and comparative genomics techniques, the genomes of *M. avium*, *M. marinum* and *M. smegmatis* were compared to those of members of the *M. tuberculosis* complex, resulting in a model for the expansion of the two gene families. This suggests that the presence of these genes can be used to differentiate between species, which has important implications for diagnostics and the understanding of pathogenesis and evolution in the genus *Mycobacterium*.