Spatial and temporal genomic homogeneity among *Haemophilus influenzae* serotype f

Aida Gonzalez-Diaz[1,2], Miguel Pinto[3], Meritxell Cubero[1,2], Jeroen D. Langereis[4,5], Arie van der Ende[6], Paula Bajanca-Lavado[3], Carmen Ardanuy[1,2], Sara Marti[1,2]

[1]Microbiology Dept. Hospital Universitari Bellvitge. IDIBELL-UB, Barcelona, Spain.
[2]Research Network for Respiratory Diseases (CIBERES), ISCIII, Madrid, Spain.
[3]Department of Infectious Disease, National Institute of Health. Lisbon, Portugal
[4]Section Pediatric Infectious Diseases, Laboratory of Medical Immunology, Radboud Institute for Molecular Life Sciences, Radboudumc, Nijmegen, The Netherlands.
[5]Radboud Center for Infectious Diseases, Radboudumc, Nijmegen, The Netherlands.
[6]Amsterdam UMC, Department of Medical Microbiology, and the Netherlands Reference Laboratory for Bacterial meningitis, Amsterdam, The Netherlands.

Background: *Haemophilus influenzae* is an opportunistic pathogen highly adapted to the human respiratory tract which is often reported as the etiologic agent of infectious diseases. After the introduction of serotype b vaccine, non-typeable *H. influenzae* (NTHi) has become the most frequent cause of respiratory infection, followed in frequency by serotype f strains (Hif). The aim of this study was to analyze the genomic diversity among invasive and colonizing Hif isolates by whole genome sequencing (WGS).

Materials/methods: Thirty-seven Hif isolated from Spain (n=16), The Netherlands (n=18) and Portugal (n=3) between 2004 and 2016 were characterized by WGS. A core-based phylogenetic tree was constructed with Parsnp from the Harvest suite, followed by a core and accessory genome analysis with Roary and roProfile. The MLST was determined *in silico* and single nucleotide analysis was done through Snippy to detect polymorphisms (SNPs) among bacterial genomes. To better understand the phylogenetic diversity in this species, all available genomes on RefSeq were downloaded and included in the analysis.

Results: Among 37 *H. influenzae*, four were collected from the oropharynx of healthy children and the remaining were from adult invasive infection. All isolates belonged to ST124 or a single locus variant. Although all strains were closely related, two major clusters were observed in the phylogenetic tree, one of them contained the four colonizer strains. The estimated core genome was 92% and 12,825 core-SNPs were detected. A total of 1,853 genes were predicted, of them, 1,691 were present in more than 95% of the strains. From 732 *H. influenzae* genomes available on NCBI, 8 were Hia, 56 Hib, 4 Hic, 1 Hid, 20 Hie and 9 Hif and the remaining were NTHi. Duplicated NTHi were removed to construct a phylogenetic tree of 347 *H. influenzae* that revealed a very low core genome (21%) compared to the Hif values (92%).

Conclusions: In contrast to the high genomic heterogeneity described in NTHi, capsulated Hif strains presented very low variability suggesting genomic stability conferred by the presence of the capsule. No genomic variation was observed among countries or the colonizing strains.