ANALYSIS OF VIRAL QUASISPECIES BY NGS TECHNOLOGIES: QUASIFLOW

DÍAZ-MARTÍNEZ L[a,b], SEOANE P[c], CLAROS MG[c] and VIGUERA E[a], GRANDE-PÉREZ A[a,b].

[a] Área de Genética, Universidad de Málaga, 29071, Málaga, Spain. [b] Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora" (IHSM-UMA-CSIC), 29071, Malaga, Spain [c] Departamento de Biología Molecular y Bioquímica, Universidad de Málaga, 29071, Málaga, Spain.

Keywords: Quasispecies, NGS, Evolution.

Objectives: The development of Next Generation Sequencing (NGS) technologies has allowed deep characterization of highly variable sequences such as viral genomes. With respect to RNA and ssDNA viruses, their low replication fidelity generates viral populations consisting of complex mutant spectra termed viral quasispecies. Their study is of special interest as they can be considered a phenotypic reservoir. The analysis of heterogeneous virus populations by computational methods is, however, a difficult task.

Materials and methods: We have developed QuasiFlow, a workflow designed in AutoFlow that takes advantage of NGS technologies to reconstruct quasispecies based in Illumina reads. QuasiFlow characterises and computes several key parameters, such as recombination events, SNPs, transitions, transversions, indels, quasispecies reconstruction, normalized Shannon index, nucleotide diversity and mutation networks. Moreover, it performs a comparative study of the samples comprising correlation, ANOVA and PCA analyses of the previously obtained virus population parameters. Using QuasiFlow we have analysed Illumina MiSeq reads from DNA samples obtained in mixed infections of ssDNA begomovirus in tomato plants amplified by rolling circle amplification. Further, we have extended the use of QuasiFlow to the analysis of the highly variable mitochondrial DNA. For that, we have used DNA Illumina MiSeq reads from 47 human mitochondrial samples from different cell lines obtained from the NCBI SRA database

Results and conclusion: QuasiFlow allowed detection of minor variants of the quasispecies at a frequency of 10^{-4} to 10^{-5} and reconstruction of haplotypes present in the sample. In addition to discover geminivirus variants QuasiFlow uncovered recombinant viruses of two closely-related begomoviruses present in the mixed infected tomatoes. Haplotype reconstruction showed distinct mutant clouds belonging to the two different begomoviruses, but also to the recombinants derived from them. Interestingly, the recombinant haplotypes were the most representative sequences in the mutant spectra. In mitochondria, QuasifFlow generated automatically SNPs, SNP frequencies, indels and analyzed up to 23 variables using PCA analysis and performed a hierarchical clustering of the samples. Our analysis was able to detect pathological variants presented in a frequency lower than 0.08%.

PRESENTACIÓN: ORAL

ÁREAS TEMÁTICAS: Virus emergentes, Evolución vírica y resistencia