## A MOVING-WINDOW GOODNESS-OF-FIT TEST TO DETECT HORIZONTAL GENE TRANSFER

O. Thas<sup>†</sup>, D. De Vleeschauwer, J.P. Ottoy, W. Van Criekinge

Ghent University, Gent, Belgium

<sup>†</sup>E-mail: *olivier.thas@UGent.be* 

Recently, biologists have discovered that for many species the distribution of tetranucleotides remains quite stable over the complete chromosome. Between species, however, the distributions may be different. This distribution is referred to as the sequence signature of a species. Comparing the signature of the complete DNA sequence ("global") with that of a shorter subsequence ("local") is used as a strategy to detect horizontal gene transfer, which is the process in which an organism transfers genetic material to another cell that is not its offspring, This process may e.g. explain resistance against certain infection diseases. We present a goodness-of-fit test based on Pearson statistics to compare the global distribution with local distributions within a sliding window along the DNA sequence. To test the null hypothesis that within all subsequences of a given length the tetranucleodide distributions are equal, we construct a test statistic as a functional of the Pearson chi-squared process. An advantage of the technique is that at the rejection of the null hypothesis, a careful study of the Pearson process gives an indication of where along the chromosome the local distributions are aberrant, which may in turn be caused by a horizontal gene transfer. The method may be altered to make it sensitive to detect horizontal gene transfer from a particular donor species. This is illustrated by applying the method to some synthetic data sets constructed from real genomic data.