

# Wavelets in genome sequence and microarray analysis

Pietro Liò

Computer laboratory, Cambridge University

Wavelet statistics is an important chapter of the long and productive history of interdisciplinary efforts between molecular biology and statistics. Genome, transcriptome and proteome projects have recently provided the field with an abundance of data. This talk is intended to provide an up to date review of some common uses of wavelet methods in molecular biology applications. In particular I will show examples of using wavelets to detect alien portions of DNA, i.e. portions of genomes transferred from another specie, inferring bacteria life-style on the basis of the DNA composition similarities and to combine protein structure information and phylogenetic inference. Increasingly, biological processes are being studied through time series of RNA expression data collected for large numbers of genes. I present implementations of wavelet algorithms applicable to RNA and protein expression data and demonstrate their application to published yeast RNA expression time series. I show wavelet to be superior to the Fourier PVE method. Wavelet functional anova will become a useful method as soon as long and replicated gene expression time series will be available. I document the impact of statistical measurement noise and sample size on the quality of time series data, and present issues related to statistical assessment of considering multiple time series from repeated experiments.

## Reference

Liò, P. (2003). Wavelets in bioinformatics and computational biology: state of art and perspectives. *Bioinformatics Review*, **19**, 2-9.