

TITLE: Digital Signal Processing for Genomic and Metagenomic Sequence Analysis

PROPOSERS: Samaneh Kouchaki, Avraam Tapinos and David L Robertson
Evolution and Genomic Sciences, School of Biological Sciences, University of Manchester.

SCOPE OF THE SESSION: Computational analysis plays an important role in studying the genetic information present in genome sequence data. However, high-throughput (so-called next-generation) sequencing technologies are generating an enormous volume of data outstripping current bioinformatics methods. In the case of metagenomic studies, the data sets contain high numbers of error-prone fragments from bacterial and viral genomes. The species and population diversity in these samples and large data volume are challenging our abilities to analyse these datasets efficiently. The analysis of such big and highly dimensional data requires fast 'intelligent' algorithms, for example, feature selection approaches that reduce dimensionality and enable easier identification of underlying signals and events in biological systems. Comparable analytical challenges are encountered in other data-intensive fields involving signal processing such as time series and image processing, in which dimensionality reduction (i.e., compression), feature extraction, dimension reduction, and data representation methods are routinely used to analyse the data, lessen the computational burden of analyses, and improve results. Due to their role in encoding genetic information and the ordered nature of nucleotides in DNA, nucleotide sequences can be considered as signals. Usually in bioinformatics this information is represented as text, i.e., sequences of the characters A, T, G and C, and represented computationally as strings and sub-strings. However, following numerical representation various signal processing tools can be applied to genetic sequences where dimensionality reduction techniques are used as standard. Such techniques, thus, introduce new feature space considering for example local pattern changes (as routinely used in image processing) for representing genomic sequences. Moreover, compressive algorithms (such as edit script compression, Discrete Wavelet Transforms and the t-Distributed stochastic neighbor embedding) we have shown to lower the computational complexity while having similar results to the original genomic representation. Therefore, the main aim of this session will be to discuss the application of signal processing methods and algorithms to genomic and metagenomic datasets for the realization of 'compressive genomics', the analysis of compressed genetic sequence data.

SHORT BIOGRAPHY OF THE SESSION ORGANIZERS: David Robertson is a research professor at the University of Manchester: <http://www.bioinf.manchester.ac.uk/robertson>. Since 2002: PI, Univ. of Manchester; 1999-2002: Wellcome Trust Research Fellow, Department of Zoology, University of Oxford; 1997-99: ANRS Research Fellow, Marseilles; 1996-97: postdoc, University of Alabama at Birmingham; 1993-96: PhD in Genetics, University of Nottingham; 1992-93: 1st year of PhD, Trinity College Dublin (lab relocated to Nottingham); 1987-91: BSc, Zoology honours, Univ. of Edinburgh.

Samaneh Kouchaki is a postdoc funded by the EU Horizon 2020 Virogenesis project and is applying digital signal processing methods to the virome component of metagenomic sequencing datasets. Qualifications: PhD in computer Science Titles "Tensor based source separation for single and multichannel signals" in Nov 2015 from the University of Surrey; MSc in Artificial Intelligence awarded in Jun 2011 from Shiraz University; BSc in Computer Engineering-Hardware awarded in Jun 2009 from Shiraz University.

Avraam Tapinos is a BBSRC-funded postdoc applying digital signal processing methods to the analysis of metagenomic sequencing datasets. Qualifications: PhD in computer Science Titles "Time series data mining in systems biology" in Feb 2013 from the University of Manchester; MSc in Information Systems Engineering awarded in Dec 2008 from the University of Manchester. BSc in Information Systems awarded in Jun 2007 from Brunel University.

ANTICIPATED NUMBER OF CONTRIBUTING PAPERS: 4/5

Contact: samaneh.kouchaki@manchester.ac.uk; david.robertson@manchester.ac.uk