Fully funded PhD position in Computational Biology
Department of Computer Science and Veterinary Laboratories Agency

Project title: Development of computational methods for linking phenotype to genotype
Supervisors: Dr Alberto Paccanaro (CS-RHUL), Prof Martin Woodward (VLA)

Applications are invited for a joint PhD position at the Department of Computer Science of Royal Holloway, University of London and the Veterinary Laboratories Agency. The PhD position is fully funded by a studentship that provides stipend and fees at UK/EU level.

The successful candidate will work with Alberto Paccanaro (www.cs.rhul.ac.uk/home/alberto) and Martin Woodward on developing novel computational methods for linking phenotype to genotype. This project will run in parallel with the sister project "From Genome to Phenome: Generating and analyzing data to predict Salmonella's phenotype from genomic information using a systems biology approach." with supervisors Martin Woodward and Vincent Jansen.

Candidates should have an equivalent of a UK undergraduate first class degree, or MSc in a biological discipline (e.g. Biology, Systems Biology, Bioinformatics) and a strong background in bioinformatics. Candidates from quantitative disciplines (e.g. Computer Science, Statistics, Engineering) with a strong background in bioinformatics are also strongly encouraged to apply. Previous research experience in bioinformatics/computational biology/systems biology would be an asset for this position.

The successful candidate will have the opportunity to collaborate with members of the newly formed interdisciplinary Centre for Systems and Synthetic Biology (which includes academics of the Computer Science and Biological Sciences Departments at Royal Holloway) and the Veterinary Laboratories Agency.

The Department of Computer Science at Royal Holloway has a leading position in the study of theory and practice of machine learning and in particular the development of the Support Vector learning technique and other kernel-based techniques. It is located on Royal Holloway's pleasant campus in Egham, close to London. Information on the Department may be found at www.cs.rhul.ac.uk

The Veterinary Laboratories Agency (VLA) is internationally recognised as a centre of excellence in veterinary research. The VLA provides a wide range of applied research on farm animal diseases, it is a national and international reference laboratory for a

wide range of farm animal diseases, and it has close links with research institutes and universities worldwide. More information on the VLA can be found at http://www.defra.gov.uk/vla/

The studentship is of 3 years duration. It provides a stipend between £13,250-£15,750 and University fees at UK/EU levels – which means that residents from EU countries can apply. This studentship does not cover the fees for applicants outside the EU.

Informal enquiries may be directed to Alberto Paccanaro (email: alberto@cs.rhul.ac.uk; URL: http://www.cs.rhul.ac.uk/home/alberto; tel: +44 (0)1784 414239) or to Martin Woodward (m.j.woodward@vla.defra.gsi.gov.uk).

Applications should be made on the PhD application form available at (http://www.rhul.ac.uk/Biological-Sciences/Vacancies/SBSVLA09.doc). Please send completed forms and a recent CV by email to Morag Mitchell, secretary to the Training and Postgraduate Education Committee of the VLA at: m.j.mitchell@vla.defra.gsi.gov.uk. Also please also forward all this material to Alberto Paccanaro (alberto@cs.rhul.ac.uk).

See also http://www.rhul.ac.uk/Biological-Sciences/Vacancies/index.html#vla for related information.

We positively welcome applications from all sections of the community.

BRIEF PROJECT DESCRIPTION

Background - Salmonella remains a significant human pathogen with both typhoidal and non-typhoidal diarrhoea diseases claiming high mortality and morbidity world wide. Of concern is the emergence of epidemic strains such as Salmonella enterica serotype Enteritidis PT4 in man as a consequence of spread in egg laying poultry. Other epidemic strains such as S. Typhimurium DT104 and S. Newport have emerged in other food producing animals, in both cited cases in cattle, to cause significant human disease in the developed world. In order to understand the emergence of epidemic strains, it is first essential to have new ways of assessing how genotypic and phenotypic features of a strain impact upon its robustness in the environment. Indeed, for an epidemic to start the host must be susceptible to the emergent pathogen that itself must be fit in a range of environments to enable transmission.

The data sets - The Woodward lab has produced both CGH data and phenotype data (BIOLOG) for Salmonella. The first dataset can be thought of as a binary matrix strains-by-genes comprising some 40 strains measured against some 7000 genes. The second dataset can be thought as a real matrix strains-by-phenotypes in which strains assessed by CGH have been assessed for over 2000 individual phenotypic characters. These phenotypes reflect the metabolic capabilities of the strains that relate directly to genetic content. These data sets are growing as more strains are being assessed.

The Approach - The Paccanaro lab recently participated in the developed of a computational method for quantifying the environmental adaptation of metabolic pathways (Gianoulis et al, *Proc Natl Acad Sci USA.* 2009;106(5):1374-9; also highlighted in *Science*, 323, 5918, 2009). In this work we have been able to quantify the relation between environmental and metabolic features of micro-organisms using two datasets that can be thought as two numerical matrices: the first one, sites-by-environmental features; and the second one sites-by-metabolic features. The paper pioneered an approach that employs correlation and regression to relate multiple, continuously varying factors defining an environment to the extent of particular microbial pathways present in a geographical site. In this project we will be interested in uncovering and quantifying the relation existing between genes-and-phenotypes in Salmonella. We plan therefore to extend the approach by Gianoulis et al. in order to be able to handle Salmonella data and to apply it to the data produced in the Woodward lab for relating genotype and phenotype in Salmonella.