2014 YCCSA SUMMER SCHOLARSHIP PROJECT SUBMISSION

This form is for prospective project supervisors to submit their projects to be included in the YCCSA Summer Scholarships Programme for 2014.

It is the purpose of the Summer School that any projects submitted are interdisciplinary in nature.

Date	*March 5, 2014*
Main Supervisor's Name	*Prof. Seth J. Davis*
Main Supervisor's Department	*Biology*
Co-supervisors' name(s) and Departments	*Prof. Tony Wilkinson; Chemistry* *Dr. Kanchon Dasmahapatra; Biology*
Project Title	*Developing a bio-informatic pipeline to exploit massive genome sequencing data to predict areas of critical function in proteins of undefined activity: A case study in the circadian clock*
Project Description	*The de novo sequencing of genomes has revealed that most encoded polypeptides cannot be assigned function from sequence. Comparative sequencing from related organisms can define clear areas of sub-conservation within one polypeptide, and one can infer that such regions are constrained for function. One can hypothesize that these are thus the "more important" regions within a protein. More recently, genomic re-sequencing programmes of hundreds to thousands of individuals within a species has created a resource of unparalleled capacity to "catch" evolution in action using population-genetic methods. These genomics pipelines have revealed that certain classes of proteins are under fast evolution, others under slow evolution, and finally, that some proteins are evolving in multiple, functional directions. Separately from these approaches, the ability to examine protein secondary and tertiary structure through so-called ab initio approaches has improved to a point where many features of protein folding are often accurately predicted.* ***Here a project is proposed that incorporates and combines comparative genomics, population genetics, and structural modelling*** *as a step to develop a pipeline to predict areas in a protein of un-described activity that are probable for critical function. This fully informatics programme will start with an examination of circadian clock proteins (e.g. ELF3 and TIC) from the model plant Arabidopsis thaliana, where 1000 genomes are available, as these are evolving at a rate likely to be informative for prediction approaches. Successes in this pipeline will then move to similar approaches in the animal circadian system with a focus on insect oscillator proteins. Together a great deal is likely to be learned about how these proteins evolve within and between species, and this will guide future structure-function projects. More broadly, a generic pipe-line could be developed to assist a wide range of functional programmes that work to understand the relationship of protein sequence to protein activity, and that should have broad use in a range of signal transduction studies.*

Required skills	*General understanding of informatics is required. Some capacity to work with "academic" software, and an understanding of scripting. Capacity to write scripts (Perl, Python or R) is helpful, but at a minimum, being able to edit existing scripts is required. Basic previous exposure to DNA analysis programs (sequence alignment and phylogenetics) and an understanding of the nature of protein secondary structure is particularly helpful.*
Project dates	*9 weeks, starting on Monday, 14 July 2014 and finishing on Friday, 12 September 2014.*
Other information	*This project will expose one to several researchers in the training of academic approaches to software use in bio-informatics.*
References	*Herrero, E., Kolmos, E., Bujdoso, N., Yuan, Y., Wang, M., Berns, M., Coupland, G., Saini, R., Jaskolski, M., Webb, A., Gonçalves, J., and Davis, S. J. (2012) EARLY FLOWERING4 Recruitment of EARLY FLOWERING3 in the Nucleus Sustains the Arabidopsis Circadian Clock. Plant Cell 24: 428–443* *Kolmos, E., Schoof, H., Plümer, M., and Davis, S. J. (2008) Structural insights into the function of the core-circadian factor TIMING OF CAB2 EXPRESSION 1 (TOC1). J. Circ. Rhythms 6: 3* *Herrero, E. and Davis, S. J. (2012) Time for a Nuclear Meeting: Protein Trafficking and Chromatin Dynamics Intersect in the Plant Circadian System. Mol. Plant 5: 554-565* *Bujdoso, N. and Davis S. J. (2013) Mathematical modeling of an oscillating gene circuit to unravel the circadian clock network of higher plants Front. Plant Sys. Biol. 4: 1-8* *Anwer, M. U. and Davis, S. J. (2013) An overview of quantitative studies in the Arabidopsis thaliana circadian clock. Sem. in Cell and Dev. Biol. 24: 422-429* *Saini, R., Jaskolski, M., and Davis, S. J. Circadian oscillator proteins across the kingdoms of life BMC Biol. (in press)*

When complete, please email the form to sarah.christmas@york.ac.uk