UNIVERSITY OF CYPRUS
DEPARTMENT OF BIOLOGICAL SCIENCES

The Department of Biological Sciences cordially invites you to the thesis defense of the PhD candidate

Stella Tamana
(Dr. Vasilis Promponas Research Laboratory)

entitled

"COMPOSITIONALLY BIASED REGIONS: FROM STRUCTURAL SIGNATURES TO COMPARATIVE GENOMICS"

Abstract

Compositionally biased regions (CBRs) refer to spans along protein sequences with composition favoring one or a few residue types. CBRs are ubiquitous and are quite often directly related to specific structural patterns, with apparent implications in protein (dys-)function and interactions.

Plasmodium species, the causative agents of malaria, have an unusual high incidence of CBRs owing to their enriched A+T genomes. The extremely biased genomes of malaria parasites induce additional technical challenges for genome sequencing projects, cloning in heterologous vector systems and require special treatment in fundamental steps of comparative genomic analyses. Additionally, the complicate life cycle of *Plasmodium* species classifies these species as very complex organisms to thoroughly study their biology *in vivo*. However, the increased number of fully sequenced *Plasmodium* genomes makes it the ideal taxon for *in silico* studies of the role of CBRs in pathogenicity, their evolutionary behavior and to evaluate strategies for handling CBRs in pan-genome analysis in the present of extreme CBR-content.

The current thesis revolves around three main aspects of CBRs: (i) how CBRs affect the computation of heavily biased pan-genomes, (ii) the role of CBRs in evolutionary behavior of the protein families under study, and (iii) CBR preferences in structural conformations. We used carefully compiled datasets of *Plasmodium* species and non-redundant protein chains retrieved from the PlasmoDB and the Protein Data Bank. Furthermore, we developed novel computational pipelines for CBR detection and masking, sequence comparison/clustering, as well as structural feature computation.

Our results indicate that our view of the plasmodial pan-genome structure is largely dependent on the different strategies used to handle CBRs. We further propose an optimal strategy for comparative genomic analyses based on thorough statistical and biological assessment.

Our comparative genomics data led us to challenge the currently established notion that only 4 out of 8 subunits of the oligosaccharyltransferase (OST) complex –which is considered conserved across eukaryotes– are present in *Plasmodium* species and other protists. Remarkably, the main reason why the unusually short Ost4 protein (36 amino acid residues in yeast) has not been identified so far is the failure of gene-prediction pipelines to detect such a short coding sequence. In fact, based on carefully conducted sequence comparisons we provide unequivocal evidence that all components of the OST complex, with the exception of Swp1/Ribophorin II, can be reliably identified within completely sequenced plasmodial genomes. Importantly our findings are further supported by publicly available EST and RNAseq expression data.

Finally, a reverse engineering approach was followed for mapping sequence and structural signatures of CBRs in experimentally 3D solved protein structures. Specifically, our results portrayed both the structural preferences of CBRs and the sequence features these CBRs possess. We focus on D-/E-rich CBRs, which consistently are sub-clustered across the structure-based clustering, discussing their possible biological roles. Most such protein structures are classified as protein binding, transport proteins or enzymes; notably, some additional classifications, such as toxins, apoptosis and cell adhesion, were also observed.

Thursday, December 6, 2018 at 12:00
Building ΘΕΕ01, Room 020 (Panepistimioupoli Campus)

The presentation is open to the public.