INFORMATION SUPERSTRUCTURE FOR COMPARATIVE ANALYSES OF HOST AND PARASITE AQUAPORINS

Raphael D. Isokpehi1, Nyasha Chambwe1, Jessica M. Murray1, Cynthia D. Jeffries1, Hari H.P. Cohly1, Subhangi Varadharajan2 and Rajendram V. Rajnarayanan3

1Department of Biology, Jackson State University, Jackson MS 39217, USA
2Department of Chemistry, D. G. Vaishnav College, TN, India
3Department of Chemistry, Tougaloo College, Jackson, MS 39174, USA

Abstract: Protozoan parasites contribute significantly to the burden of global infectious diseases. Genome sequencing projects on these organisms combined with high-throughput gene and protein expression experiments are producing wealth of datasets for the discovery of novel drug targets. Water channel proteins collectively termed aquaporins are increasingly recognized as potential drug targets or transport routes for anti/protozoan drugs. Conventional aquaporins are six-transmembrane helix proteins with two Asparagine-Proline-Alanine (NPA) motifs found in the inter-transmembrane loops B and E that form an aqueous pore. There are 13 known human aquaporins (AQP0 to AQP12). Protozoan aquaporins have been shown to transport nutrients and metabolites between host and parasite that are crucial for initiating an infection and survival during parasite’s life cycle. Our objective in this study is to assign a 13-digit binary signature to protozoan aquaporins based on presence or absence of a significant sequence similarity to the human aquaporins. We have developed a computational pipeline that searches for sequence similarity, assigns binary signatures and classifies sequences into groups. The pipeline used the human aquaporin proteins to query over 10,000 protein sequences compiled from multiple databases of the following sequenced protozoan parasites: Cryptosporidium hominis, Cryptosporidium parvum, Entamoeba histolytica, Leishmania major, Plasmodium falciparum, Plasmodium berghei, Plasmodium yoelli, Theileria annulata, Toxoplasma gondii, Trypanosoma brucei, and Trypanosoma cruzi. Our analysis revealed 21 aquaporin-like sequences which classed into 13 of the possible 8192 binary signature groups. Nine groups had only one sequence. The approach correctly classed the aquaporin variant haplotypes annotated in the T. cruzi genome into 3 groups. The last group (no significant similarity to AQP8, AQP11 and AQP12) contained sequences from the rodent plasmodia, T. brucei and L. major. The proteomes of Cryptosporidium, Entamoeba, Theileria did not contain any significant match to the human aquaporins. The human AQP11 did not yield any significant parasite match. In conclusion, by integrating datasets of human and protozoan aquaporins, we have built an information superstructure for the comparative analyses of aquaporins and to support the development of novel anti/protozoan drugs.

Keywords: Aquaporins, Protozoa, Information Science, MEDLINE, Sequence Similarity, Text Mining, Water Transport

Acknowledgements: Mississippi NSF-EPSCoR “Innovations through Computational Sciences” Award (EPS-0556308); Interdisciplinary Undergraduate Training in Biological and Mathematical Sciences (DMS-0531927); and the Research Centers in Minority Institutions (RCMI) – Center for Environmental Health (NIH-NCRR G12RR13459-09)