

Investigation of the Evolution of MutS Protein Family in Animals

EEOB 563 Final Project Outline

Mudith Ekanayake

Introduction

Throughout different stages of the cell cycle, many DNA repair pathways act in the cells allowing the cells to repair the DNA damage. Among these pathways mismatch repair (MMR), base excision repair (BER), nucleotide excision repair (NER), homologous recombination (HR) and non-homologous end joining (NHEJ) are the major pathways that are active in the cells. More importantly DNA mismatch repair (MMR) is responsible for recognizing and repairing erroneous insertion, deletion and mis-incorporation of bases during DNA replication and recombination. In MMR, there are three major steps including mismatch identification, mismatch excision and DNA re-synthesis. Throughout this whole process several proteins are involved and among those proteins, MutS is incorporated in detecting mismatches in the sequences. These proteins are well conserved in prokaryotes, eukaryotes and even in viruses. There are many homologs of MutS including MutS1, MutS2 in bacteria and MSH 1 to 6 in eukaryotes. In this project MSH 2-6 and MutS2 will be selected for the analysis. Since mismatch repair is critical for maintaining genome stability, from this project, evolution of the MutS protein family in animals will be investigated.

Methods

Major source of data for this project will be public databases such as NCBI (<https://www.ncbi.nlm.nih.gov/>) and the first step will be retrieving the data from the databases. Then MAFFT sequence alignment program will be used for aligning the sequences and the alignments will be curated and cleaned using TrimAl or GBlocks. As the next step, model selection will be carried out prior to constructing the phylogeny. Phylogenetic reconstruction will be undertaken under a maximum likelihood framework implemented in RAxML-NG using the GTR model of DNA substitution, and with confidence levels estimated using bootstrap resampling with 1000 replicates. Itol tool will be used for the visualization of the tree. If time permits, PAML (Phylogenetic Analysis by Maximum Likelihood) tool also will be incorporated in the analysis.