Computational investigations into the evolution of mitochondrial genomes

Dissertation, 2015

Abdullah Sahyoun

Mitochondria are organelles present in most eukaryotic cells. They generate most of the cells adenosine triphosphate (ATP) supply which make them essential for cell viability. It is assumed that they are derived from a proteobacterial ancestor as they retain their own, drastically small genome.

The importance in studying mitochondrial genome evolution came from the discovery of a large number of human diseases that are caused by mitochondrial dysfunction (e.g., Parkinson and Alzheimer). Many of these diseases are a result of a mutation in one of the mitochondrial genes or a defective mitochondrial DNA (mtDNA) maintenance, mostly caused by genetic defects in proteins involved in mtDNA replication. In order to explore the diversity and understand the evolution of mitochondrial genomes (mitogenomes) in animals, multiple methods have been developed in this study to deal with two biological problems related to the mitochondrial genome evolution.

A new method for identifying the mitochondrial origins of replication is presented. This method deals with the problem of determining the origins of replication, which despite many previous efforts has remained non-trivial even in the small genomes of animal mitochondria. The replication mechanism is of central interest to understand the evolution of mitochondrial genomes since it allows the duplication of the genetic information.

The extensive work that has been done to study the replication of mitochondrial genomes has generated the assumption of the strand displacement model (SDM) also known as the standard model of replication that is known to leave the mitochondrial H-strand in a single stranded state exposing it to mutation and damage. Later on, other models of replication have been suggested such as the strand coupled bidirectional replication model, its refinement which assumes the bidirectional mode but with a unidirectional start, and the "RNA incorporation throughout the lagging strand" (RITOLS) model proposed as a refinement of the strand displacement model. Based on the observation that the GC-skew is correlated with the distance from the replication origins in the light of the strand displacement model of replication, a new computational method to infer the position of both the heavy strand and the light strand origins from nucleotide skew data has been developed. The method has been applied in a comprehensive survey of deuterostome mitochondria where conserved positions of the replication origins for the vast majority of vertebrates and cephalochordates have been inferred. Deviations from the consensus picture are presumably associated with genome rearrangements.

Additionally, two methods for the identification of tRNA remolding events throughout Metazoa have been developed. Remolding changes the identity of a tRNA by a duplication and a point mutation(s) of the anticodon. This new tRNA takes the identity of another tRNA which is then lost. This can lead to artifacts in the annotation of mitogenomes and thus in studies of mitogenomic evolution. In this work, novel methods are developed to detect tRNA remolding in large-scale data sets. The first method represents an extension of the similarity-based approach to determine remolding candidates with high confidence. This approach uses an extended set of criteria based on both sequence and structural similarities of the tRNAs in conjunction with statistical tests. The second method is a novel phylogeny-based likelihood method which evaluates specific topologies of gene phylogenies of the two tRNA families relevant to a putative remolding event. Both methods have been applied to survey tRNA remolding throughout animal evolution. At least three novel remolding events are identified in addition to the ones previously mentioned in the literature. A detailed analysis of these remoldings showed that many of them are derived ancestral events.