Animal Actin Phylogeny and RNA Secondary Structure Study

Bibhuti Prasad Barik, P.G. Department of Zoology, Khallikote University, Brahmapur, India

ABSTRACT

Animal actin is a diverse and evolutionarily ancient protein. Actin genes and their corresponding protein sequences were used to infer phylogenetic affiliations. The study indicated that several species appear to be polyphyletic and several unrelated species appear to share the same clade. Consensus actin RNA secondary structures showed that the structural features of all forms were quite distinct and different from each other. This observation supports the phylogenetic inference in which similarly named species clustered together based on their lifestyles. Consideration of actin gene geneology and consensus RNA secondary structures could be used as a possible phylogenetic marker among diverse species of the animal kingdom for large scale data analysis. In-silico study revealed variations among the groups. The percentages of long disordered regions in proteins were found to be very high in all forms. Such findings suggest that the complexity and ability to adapt in diverse habitats by species may be due to higher percentage of disordered proteins.

Keywords: Actin, Clade, Phylogeny, Protein Disorder, RNA, Secondary Structure

INTRODUCTION

Conventional actin appears to be ubiquitous in eukaryotes. The occurrence of actin in diverse animal groups is interesting and researchers are curious about their nature of existence. The genes that code for actin are defined as a gene family (Pont et al., 1983). This means that the genetic information of each individual contains instructions that generate actin variants (called isoforms) that possess slightly different functions. This, in turn, means that eukaryotic organisms express different genes that give rise to: α-actin, which is found in contractile structures; β-actin, found at the expanding edge of cells that use the projection of their cellular structures as their means of mobility; and γ-actin, which is found in the filaments of stress fibres (Scott et al., 2012). The evolutionary origin of actin can be traced which have equivalent proteins (Gunning et al., 2015). In addition to the similarities that exist between an organism’s isoforms there is also an evolutionary conservation in the structure and function even between organisms contained in different eukaryotic domains.

The objectives of the present study were:

1. Survey and retrieval of actin gene and protein sequences from world wide database.
2. Phylogenetic study of different animal groups based actin gene sequences.

DOI: 10.4018/IJKDB.2015010104
3. Differentiation of animal phyla based on RNA secondary structure study.
4. Characterization of the actin proteins of different animals based on physicochemical properties of translated proteins and their disorderness.

Materials and Methods

Sequence Retrieval and Taxon Sampling

Actin gene sequences (772256) of diverse animal species available in GenBank database were retrieved using entrez key word search and PERL script. The sequences were filter searched and around 90 sequences were selected referring to specific phyla. These gene sequences were considered for phylogenetic analysis and their corresponding RNA secondary structure were also predicted.

Multiple Sequence Alignment and Phylogenetic Analysis Based on Actin Gene

Multiple sequence alignment (MSA) was performed using CLUSTALW program with default settings. Phylogenetic trees were generated by two different methods (MP and ML) by MEGA 7 (Kumar et al., 2016). All characters were equally weighted and unordered. Alignment gaps were treated as missing data.

RNA Secondary Structure Prediction

The actin gene sequences were used to generate consensus RNA secondary structures using LocARNA version 1.5.2 (Smith et al., 2010). It performed multiple alignments and generated consensus secondary structures of each category using realistic energy model for RNAs.

In-silico Physiochemical Characterization

The physiochemical characterization of the protein sequences was carried out using ProtParam (Gasteiger et al., 2005), tools by Expert Protein Analysis System (ExPASy) web based programs (http://www.expasy.org/tools/protparam.html).

Disorder Prediction

The short and long disorder segments of the actin proteins were predicted with CSpritz web server (Walsh et al., 2011).

Results

Maximum Likelihood Tree

The evolutionary history was inferred by using the Maximum Likelihood method (Figure 1). The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. The analysis involved 90 amino acid sequences. There were a total of 1217 positions in the final dataset.

The maximum likelihood tree depicted ten major clades (Figure 1). The clade I comprised three sub clades. Sub clade I contains ten species of echinoderms. Sub clade II comprised three species of annelids and sub clade III comprised one nematode (O. osterigae) and one mollusc.
Applying Bayesian Networks in the Early Diagnosis of Bulimia and Anorexia Nervosa in Adolescents: Applying Bayesian Networks in Early Diagnosis in Adolescents

New Trends in Graph Mining: Structural and Node-Colored Network Motifs
www.igi-global.com/article/new-trends-graph-mining/40973?camid=4v1a