Phylogenetic analysis of ASCL gene family

Alexandra Yu. Filatova, A. V. Marakhonov

Federal State Budgetary Institution "Research Centre for Medical Genetics" under the Russian Academy of Medical Sciences, Moskvorechie Str., 1, Moscow, Russia, MAACC@yandex.ru

Ancha Baranova,

Federal State Budgetary Institution "Research Centre for Medical Genetics" under the Russian Academy of Medical Sciences, Moskvorechie Str., 1, Moscow, Russia,

School of Systems Biology, David King Hall, MSN 3E1, George Mason University, Fairfax, VA, 22030, abaranov@gmu.edu

Mikhail Yu. Skoblov

Federal State Budgetary Institution "Research Centre for Medical Genetics" under the Russian Academy of Medical Sciences, Moskvorechie Str., 1, Moscow, Russia, mskoblov@generesearch.ru

Most genes have homologous sequences in genome of same species (paralogs) and other species (orthologs). In most cases formation of paralogs is result of duplication followed by functional divergention [1]. Thus it's possible to discover evolution of gene families using bioinformalical approaches.

ASCL (ahaete-scute complex like (Drosophila)) gene family belongs to superfamily of genes encoding transcription factors, which contain HLH (helix-loop-helix)-domain. There are five paralogs in mammalian genomes: *ASCL1*, *ASCL2*, *ASCL3*, *ASCL4*, *ASCL5*. The nucleotide and amino-acid sequences of these genes are close enough, but their expression occurs predominantly in different tissues and different stages of development. Moreover it's known that *ASCL* genes have different physiological functions. For example *ASCL1* plays a role in the neuronal commitment and differentiation and expressed during early development of the nervous system[2], whereas *ASCL2* is required for placental development and expressed only in trophoblast [3]. Thus the formation of ASCL gene family is a result of functional divergence after duplications. And purpose of our study was to characterize ASCL-gene family and analyze its evolution.

For our analysis we used protein and nucleotide sequences of ASCL genes from 35 different animal species. We made alignment, measured phylogenetic distances and built phylogenetic trees. Based on our data, we revealed that ASCL gene family is highly conservative, it has

orthologs in all studied multicellular organisms to begin with Cnidaria. For example, there is 78% of similarity between most conservative bHLH-domains of human protein ASCL4 and Hydra achaete-scute-like protein. The observation that ASCL-gene first appeared in two layers animals can be explained by the fact that ASCL-genes participate in the cell fate determination by mechanism of lateral inhibition [4]. Beside that we observed that ASCL-family divided into two subfamilies: ASCL1,2 and ASCL3,4,5, which are different to each other but similar within subfamily amino-acid sequence of basic region. It's known that basic region of bHLH domain is responsible for recognizing a particular DNA sequences [5]. Perhaps mutations in this region could cause the functional divergence between the ancestral forms of ASCL1, 2 and ASCL3,4,5. In addition, we revealed that subfamily ASCL3,4,5 is closer to common ASCL-ancestor, than ASCL1,2. Although some more complexes organisms don't have annotated ASCL3,4,5 orthologs, but possess ASCL1,2 orthologs.

1. Y. Zhang et al (2003) Functional Divergence after Gene Duplication and Sequence–Structure Relationship: A Case Study of G-Protein Alpha Subunits, Mol DevEevol, 308B:85–96
2. F. Guillemot et al (1993) Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons, Cell, 75(3):463-76.
3. M. Alders et al (1997) The human Achaete-Scute homologue 2 (ASCL2,HASH2) maps to chromosome 11p15.5, close to IGF2 and is expressed in extravillus trophoblasts, Hum Mol Genet., 6(6):859-67.
4. R. Kopan (2009) The canonical Notch signaling pathway: unfolding the activation mechanism, Cell, 137(2):216-33
5. C. Murre et al (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins, Cell, 56(5):777-83.