An Evolutionary Analysis of Membrane-Associated Guanylate Kinase Protein Family

Ryan Carr, Heather Deal, Rebecca Seipelt-Thiemann

Abstract: Gene families come into being through gene and/or genome duplication followed by mutation over time which results in the evolutionarily-related genes having somewhat different nucleotides, amino acids, gene structure, and functions. The membrane-associated guanylate kinase protein family has twelve members in humans: *DLG 1, DLG2, DLG3, DLG4, CASK, MPP1, MPP2, MPP3, MPP4, MPP5, MPP6,* and *MPP7*. This gene/protein family is characterized by the presence of three specific protein domains: PDZ, SH3, and GUK, all of which aid in protein-protein interactions. These proteins are known to interact with cytoskeletal proteins and also are involved in signal transduction. A characteristic member of this family is the DLG3 gene, is responsible for encoding a synapse associated protein (SAP102). The goal of this study was to better understand the evolutionary relationships among the protein/gene family members. To attain this goal, two evolutionary investigations were undertaken. First, phylogenetic trees, which are the traditional method of analysis, were constructed using the amino acids. This analysis indicated evidence for three distinct sub-groups: group A contained *CASK, MPP1, MPP2, MPP6*; group B contained *MPP3, MPP4, MPP5, MPP7*; and group C contained *DLG 1, DLG2, DLG3, DLG4*. Next, the phylogenetic relationship based on the exon structure was undertaken. Briefly, multiple alignments were combined exon boundary information to generate a visual map of similarities and differences in exon structure among the gene family members. This visualization and its comparison to the traditional phylogenetic analysis will be presented.

MEGA 6.0.6 Molecular Evolutionary Genetics Analysis

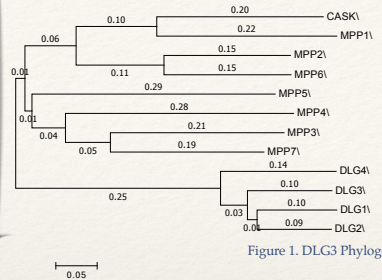

Figure 1. DLG3 Phylogenetic analysis

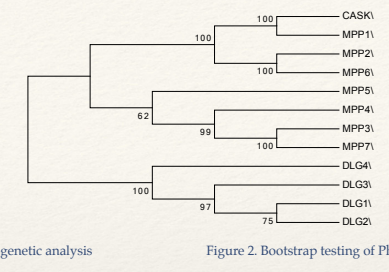

Figure 2. Bootstrap testing of Phylogenetic analysis

Methods: An alignment of the human associated members of Membrane-Associated Guanylate Kinase protein family was researched using EMSEMBL to identify Paralogs *DLG3* (Disks Homolog 3). Twelve gene paralogs were identified: DLG1, DLG2, DLG3, *DLG4*, *MPP1*, *MPP2*, *MPP3*, *MPP4*, *MPP5*, *MPP6*, *MPP7*, and *CASK*. The UCSC genome browser was used to retrieve both the translated and untranslated exons of each gene in FASTA format. Using a Perl script exon sizes were calculated. The data was organized regarding the size and number of exons for each gene. Expasy was used to translate the coding regions of each gene, which formed the basis for further analyses of shared open reading frames. Translated amino acid sequences were aligned and shaded using a Multiple Alignment Tool (Protein Information Resource) and the Box shade server (embnet.org). This alignment was used to identify the regions of amino acid similarity between the amino acids encoding each exon across the gene family and, in combination with the exon sequences, to trace the relationship between exons with shared reading frames. The final step of the gene family analysis was using MEGA software to construct a (type of tree, probably neighbor joining tree?) tree along with a bootstrap testing (X values) to reflect the confidence in these results.

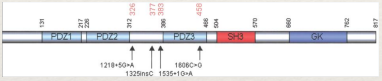

Figure 4. Synaptic Associated Protein 102 (SAP102) showing stop codon introduced

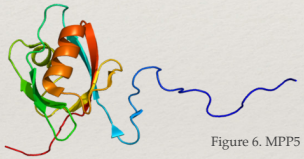

Figure 6. MPP5

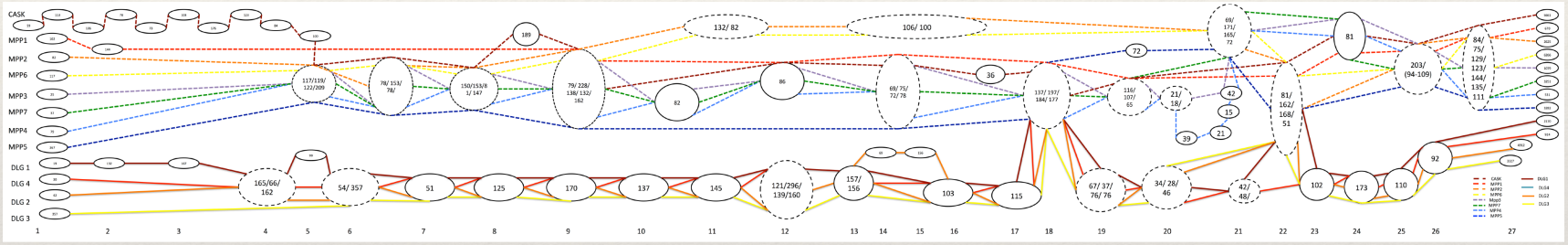

Figure 5. Exon structure analysis. Solid nodes represent exons share exact reading frame size. Dashed nodes represent coding regions that align but contain a variable number amino acids

Figure 3a. Unaligned Amino Acid Sequences

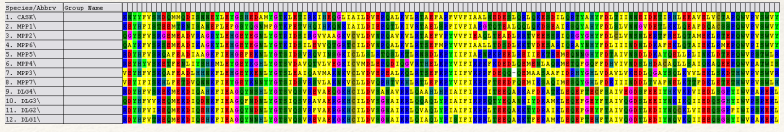

Figure 3b. Aligned Amino Acid Sequences

Conclusions:

Data visualization provides key insights in to the relationships within this gene family. Two large groups were found in both types of analyses, composed of 1) the *MPP* genes and *CASK* and 2) the *DLG* genes. Further, three distinct sub-groups can further be observed in both data sets. These are more evident from the cleaner visualization of phylogenetic analyses, but the exon structure analyses provide deeper information about the relatedness of these genes. In no particular order group A contains *CASK*, *MPP1*, *MPP2*, *MPP6*; group B contains *MPP3*, *MPP4*, *MPP5*, *MPP7*; and group C contains *DLG* 1, *DLG2*, *DLG3*, *DLG4*. Furthermore, the exon structure analyses can suggest how the gene structure changed to generate the current existing genes in the family. For example, *CASK* and *DLG1* likely had exon insertion events near the 5′ end of the gene, whereas, *MPP7* likely had an exon deletion event near the middle of the gene. Investigation of the function and location of expression for these genes show that they also share similarities in function or location of expression. For example, the *DLG* group (C) genes have an expression pattern that is limited to the brain. It would further be interesting to investigate and map the protein domains for these proteins relative to the exons to determine how new domains were inserted or deleted from the gene family. Additionally, it would be interesting to perform the same analyses for this gene family in other species to examine how they are similar and different from the human gene family. The importance of our research involving the Membrane-Associated Guanylate Kinase protein gene family is to observe the relationships among the gene family members to better understand how genes and genomes change.

References:

Nithianantharajah, Jess, Noboru H Komiyama, et al, "Synaptic Scaffold Evolution Generated Components of Vertebrate Cognitive Complexity." *Nature Neuroscience Nat Neurosci* (2012): 16-24. Print.

Kaufman, Liana, Muhammad Ayub, and John B. Vincent. "The Genetic Basis of Non-syndromic Intellectual Disability: A Review." *Journal of Neurodevelopmental Disorders* (2010): 182-209. Print.

"Emory Genetics Lab." Intellectual Disability, DLG3-related: DLG3 Gene Sequencing. Web. 20 Nov. 2015.

Tarpey, Patrick, Josep Parnau, et al, ."Mutations in the DLG3 Gene Cause Nonsyndromic X-Linked Mental Retardation" The *American Journal of Human Genetics* (2010)

"Mega 6.0." Molecular Evolutionary Genetics Analysis. 1 June 2014. Web. 10 Nov. 2015.

"Human Chr21:33,031,597-33,041,570 - UCSC Genome Browser V325." *Human Chr21:33,031,597-33,041,570 - UCSC Genome Browser V325*. Web. 10 Sept. 2015.