Evelyn Zheng

December 1, 2021

Protein: XP_001634797.2

Nematostella vectensis and its gene family, Alpha-adducin

I. Introduction

The gene family alpha-adducin plays an important role in hypertension (Tripodi *et al.*, 1996). Alpha-adducin is found in most tissues, where it is a key component in stabilizing the membrane cytoskeleton. Alpha-adducin also plays an important role in cell signaling, ion transport, mitosis, regulation of cell motility, and cell-cell adhesion. Previous studies have shown that when observing the phosphorylation of alpha-adducin, the alpha-adducin itself appears more frequently in certain tumors, which demonstrates that alpha-adducin plays a key role in evolution (Luo and Shen, 2017). Alpha-adducin and Cnidarian have yet to be explored. The closest study is alpha-adducin-2, which is a second alpha-adducin, on *Nematostella vectensis*.

Nematostella vectensis, a species of the Cnidarian phylum, was selected as they have a nerve net nervous system in both their endodermal and ectodermal tissues (Havrilak *et al.*, 2021). By having a nerve net nervous system, the endodermal and ectodermal tissues are easier to observe through all their stages of development. Alpha-adducin is an important gene in species with a nerve net nervous system as the gene is important to observe in cell-cell adhesion because alpha-adducin appears in the growth of a species' cytoskeleton. Additionally, *Nematostella vectensis* allows for regeneration of their whole body (Amiel *et al.*, 2015). The genes and tissues that were responsible for regeneration is something that has never been explored. *Nematostella*

vectensis is very desirable as they are the very first Cnidarian to be sequenced, with similar abilities to sequence genomes as those of mammals (Reitzel *et al.*, 2007). Their ability to regenerate body parts within days of amputation was deemed extremely beneficial (Putnam *et al.*, 2007).

Alpha-adducin-2 could be found in multiple species, not just *Nematostella vectensis*, making this gene extremely versatile. Alpha-adducin-2 is an actin binding gene. This gene is specifically involved in a species' (in this case, *Nematostella vectensis*) structure and organization of its cytoskeleton, meaning that it helps with the overall shape of the species (Porath *et al.*, 2017). As a result, Alpha-adducin-2 is an important gene in *Nematostella vectensis* as it helps with identification and structure. Alpha-adducin-2 is a function that upregulates during spawning (Porath *et al.*, 2017). This means that Alpha-adducin-2 also aids in recoding, an important function to observe in species that arose due to different gene copies. Overall, Alpha-adducin-2 poses an important role in the overall structure of the species, which is important for evolution as it shows how the overall structure of the species developed over time.

There are limitations to alpha-adducin and its ability to evolve independently in Cnidarians and Bilaterians as it has yet been explored. There was also not enough information on whether copies of alpha-adducin were already present in the Cnidarian-Bilaterian common ancestor. Though, it is suspected that alpha-adducin could be present in the Cnidaria common ancestor as there were studies on alpha-adducin-2 and *Nematostella vectensis*, but that is a question that has yet been answered. As a result, it is hypothesized that alpha-adducin plays a role in evolution.

II. Methods

The alpha-adducin gene of *Nematostella vectensis*, which is an important source for the overall structure of the species, was compared among other species: *N. vectensis* (starlet sea anemone), *B. blecheri* (amphioxus), *A. planci* (crown-of-thorns starfish), *M. yessoensis* (Yesso scallop), *H. sapiens* (human), *H. vulgaris* (fresh-water polyp), and *D. gigantea* (carnation coral).

Homolog alignment

The initial approach is to find the alpha-adducin sequence, its putative homologs, align the homologs, perform statistics on the homolog, and then remove areas that have highly gapped parts. The alpha-adducin sequence was obtained from NCBI by using *blastp* on the protein sequence (Altschul *et al.*, 1990). Then, *awk* was used to filter the initial *blastp* results with all of the homologs, even those with e values above 1×10^{-14}. By setting the e-value to 1×10^{-14}, the accurate hits of the protein homologs can be obtained. *Seqkit* extracts the alpha-adducin proteins from the filtered out homologs (Shen *et al.*, 2016). *Muscle* was able to perform a global sequence alignment (Edgar, 2004). Meanwhile, *T-Coffee* performs statistics on the alpha-adducin alignment, removing positions in the alignment that were more than 50% gapped (Notredame and Heringa, 2000).

Make phylogenetic tree for alpha-adducin

After the gene family homologs for alpha-adducin were aligned, the phylogenetic tree was made. *Iqtree* obtained the unrooted alpha-adducin tree (Nguyen *et al.*, 2015). By using the unrooted tree, the midpoint rooted tree of alpha-adducin was generated using *gotree* (Lemoine and Wang, 2017).

Reconcile species and gene tree

Once the midpoint rooted tree for alpha-adducin has been created, the gene tree will be reconciled with the evolutionary history of the species tree. First, *notung* reconciles the alpha-

adducin tree with the species tree that includes Bilaterian, Cnidaria, Chordata, Echinodermata, and Mollusca (Chen *et al.*, 2000). *Notung root* reroots the alpha-adducin gene tree to minimize the amount of duplications and deletions. *Iqtree* was then used to calculate the number of bootstrap support and tree search for the alpha-adducin gene (Hoang *et al.*, 2018). Lastly, *gotree* reroots the bootstrap support tree into the midpoint root tree (Lemoine and Wang, 2017).

Graph the predicted Pfam domains onto the alpha-adducin tree phylogeny

After the alpha-adducin tree has been reconciled and rerooted to the midpoint root tree, the predicted Pfam domains can be graphed back to the alpha-adducin tree phylogeny. *Iprscan5* analyzes the alpha-adducin gene sequences to the protein signature databases (Jones *et al.*, 2014). The genes were then sorted according to the Pfam database. Finally, *Evolview* (an online web-service) plotted the alpha-adducin gene tree with the Pfam protein domains (Zhang *et al.*, 2012).

The GitHub repository can be found at: https://github.com/evzheng/Nematostella-vectensis-and-Alpha-adducin.git

III. Results

The evolutionary lineage of alpha-adducin in *Nematostella vectensis* was determined through utilizing NCBI blastp to find the homologs of alpha-adducin in the following species: *N. vectensis* (starlet sea anemone), *B. blecheri* (amphioxus), *A. planci* (crown-of-thorns starfish), *M. yessoensis* (Yesso scallop), *H. sapiens* (human), *H. vulgaris* (fresh-water polyp), and *D. gigantea* (carnation coral). Once the homologs were found, they were sorted with the e-value set to 1×10^{-14}. A gene sequence alignment was performed by removing highly gapped regions. Then,

statistics were performed through t-coffee on the homolog, yielding the following results seen in Table 1. The total average identities between all the sequences seem to be low, at 31.88%.

Table 1. T-coffee statistics of alpha-adducin homologs through utilizing NCBI blastp.

Calculations	**Values**
Initial Blast of alpha-adducin Gene Homologs	25
Human alpha-adducin Homolog Hits	3
Alpha-adducin Gene Homologs filtered at 1e-14	16
Standard Deviation (avg. % identity b./t query sequence and all other sequences)	10.04%
Average/Total (avg. % identity amongst all sequences)	31.88%

A gene tree was created in Figure 1 utilizing GoTree, which draws rerooted trees by filtering the putative homologs (through its unrooted tree). The gene tree has its branch lengths proportional to time relative to speciation events. Longer branch length shows more occurring substitutions, which means that it took longer for the next lineage to occur. Figure 1 shows several alpha-adducin paralogs, such as *B. belcheri* and *H. sapiens*, with multiple branches that are short.

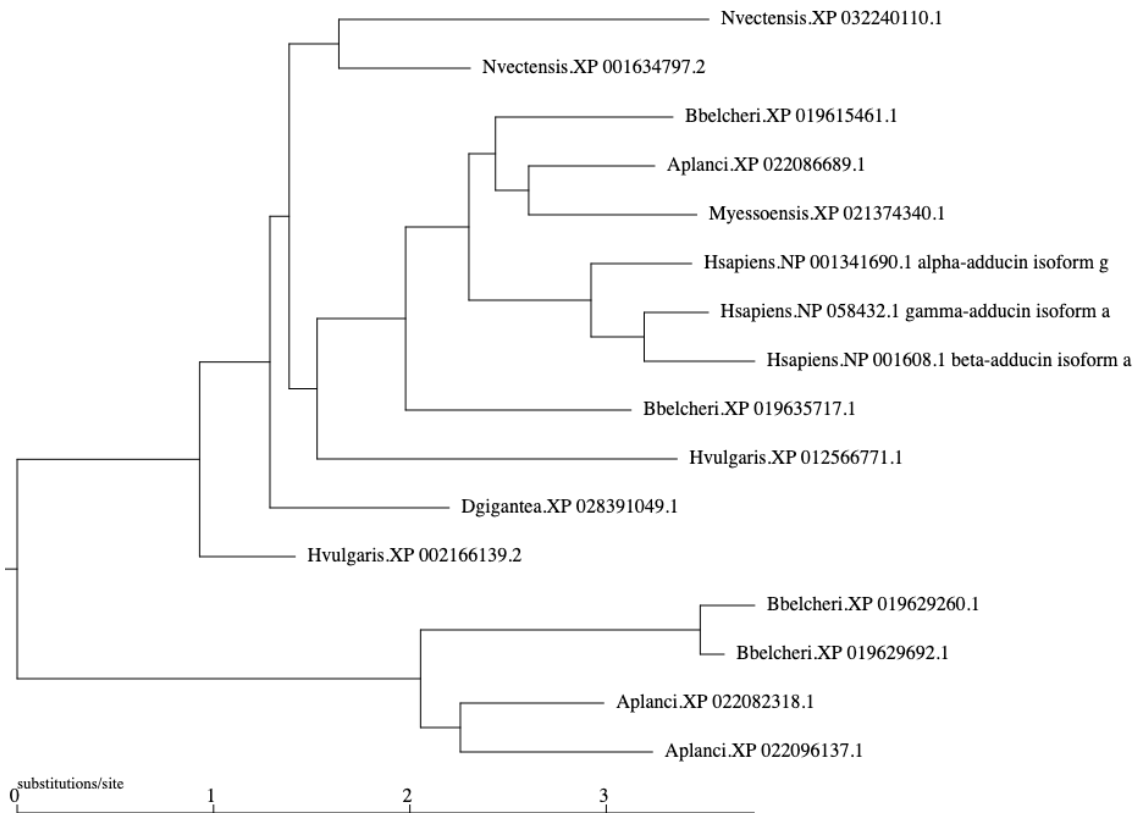

Figure 1. Alpha-adducin gene tree produced by GoTree using alpha-adducin homologs.

Figure 2 displays a reconciled gene and species tree by using Notung 3. Figure 2 includes duplication as well as speciation events that may not be seen in Figure 1. There were a total of 12 duplications and 22 losses in the reconciled gene tree (Figure 2). The score (cost) of the reconciliation between the gene and species tree was calculated to be 40.0. The reconciled gene tree for alpha-adducin is shown, with a red branch marked D as a duplication event and a gray branch marked *LOST as a deletion event (Figure 2). The alpha-adducin gene and species tree shows each duplication often resulting in some kind of loss. There are single-side tree expansion events where the majority of the duplication events occur on one side of the lineage (Figure 2).

There is also a speciation or loss of alpha-adducin homologs in both *Bilateria* and *Cnidaria* species, which is also represented in Figure 1.

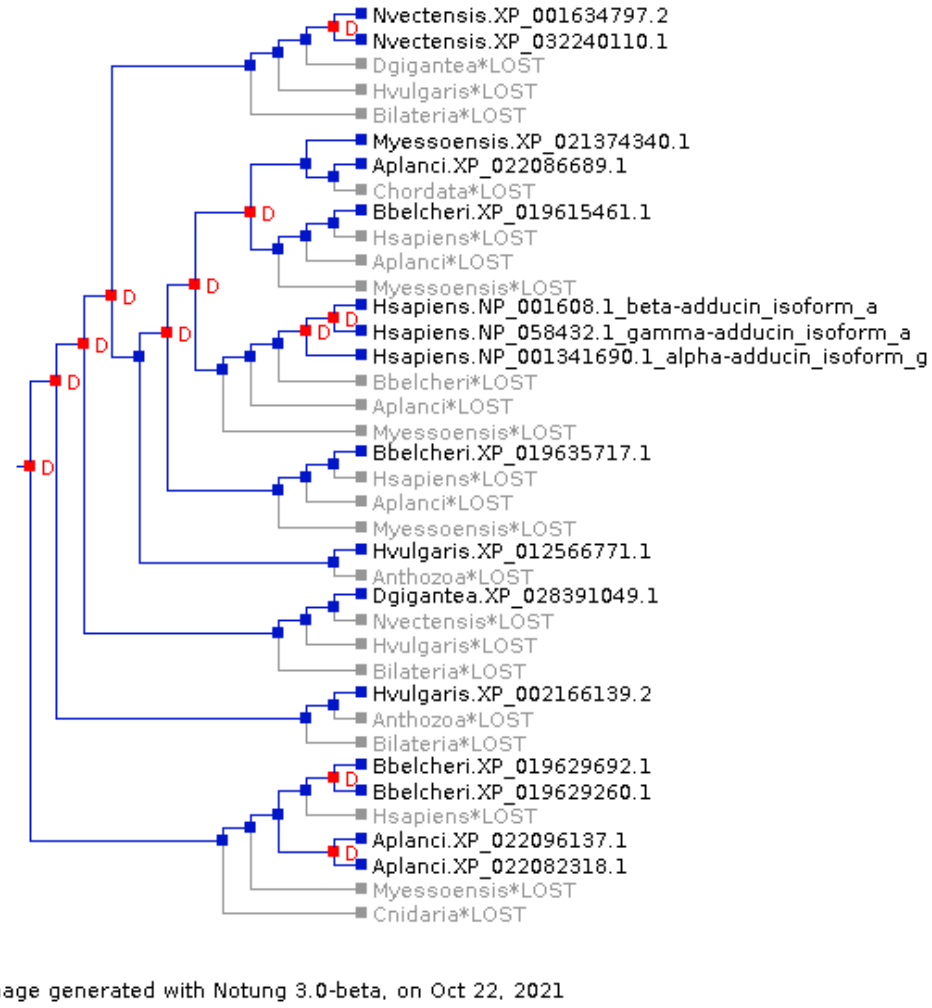

Figure 2. Alpha-adducin gene and species tree produced by Notung 3 using alpha-adducin homologs.

Lastly, the protein domains of alpha-adducin in *Nematostella vectensis* was found through the Pfam database. The alpha-adducin protein domain found was PF00596. This was then mapped onto the alpha-adducin gene and species tree alongside the gene and species that

have the protein PF00596. It has been shown throughout the figures that *Cnidaria* and *Bilateria* have a loss of the alpha-adducin protein.

IV. Discussions

The domain predicted in the Cnidarian gene family is the same domain predicted as the Bilaterian gene for the alpha-adducin gene family. The predictions are similar in that they are the PF00596 domain for both Cnidarian and Bilaterian. The Cnidarian gene copies have all of the domains expected according to the scientific literature about alpha-adducin. Since the tree and protein domain analysis yielded only one protein domain, that protein domain is constantly present across the gene tree. The alpha-adducin gene and species tree shows that each duplication often resulted in some kind of loss (Figure 2). This means that duplication occurs within the same species.

The hypothesis that alpha-adducin plays a role in evolution predicted within the literature could be shown in the results. Alpha-adducin plays an important role in hypertension (Tripodi *et al.*, 1996). Alpha-adducin is also found in most tissues, which is important to stabilize the cytoskeleton membrane. It is also known to help with identification and structure. Alpha-adducin helps especially in recoding, which is extremely important to observe species that arose as a result of different gene copies. There have also been studies showing that alpha-adducin plays an important role in evolution through non-synonymous single nucleotide polymorphisms (Achintya, 2018).

It has been known that *N. vectensis*, *A. planci*, *M. yessoensis*, *H. vulgaris*, and *D. gigantea* are invertebrates as they do not have a cytoskeleton membrane. *B. blecheri* and *H. sapiens* do have a cytoskeleton membrane. This might mean that even though the invertebrates

do not have a cytoskeleton membrane, they do have structure which enables them to move through the current. This is present in a study that shows alpha-adducin helps with the overall shape of the species (Porath *et al.*, 2017). As a result, alpha-adducin appears in both the invertebrates and vertebrates as the protein aids in evolution.

There was not enough external evidence for choosing an outgroup. There was also not enough information to determine if the midpoint rooted phylogeny seemed accurate (Figure 1). The protein function found in the study was not explored; the results displayed evolutionary processes instead of the function of the protein in comparison between species. In the future, if the differences or similarities of the protein function between each species were to be found, it would better help determine alpha-adducin and its relationship with cytoskeleton development. Some future research that could be done is to use another more-commonly researched protein with similarities to alpha-adducin on *Nematostella vectensis*. Perhaps this more-commonly researched protein could help us determine if the alpha-adducin gene, species, and protein homolog tree is as accurate when compared to something that was already predetermined. If the hypothesis of whether alpha-adducin, a protein important in the membrane cytoskeleton, plays an important role in evolution, then alpha-adducin can be compared to more organisms with the actin binding gene and organisms that do not have the actin binding gene to determine a result. This way, there might be a concrete pattern in alpha-adducin's role in evolution. Future research could be conducted on other proteins that are connected to alpha-adducin and its role to stabilize the cytoskeleton membrane.

V. References

ACHINTYA, M.G., 2018. Alpha-adducin nsSNPs affect mRNA secondary structure, protein modification and stability. *Meta Gene* **17(1)**: 153-162.

ALTSCHUL, S.F., W. GISH, W. MILLER, E.W. MYERS & D.J. LIPMAN, 1990. Basic local alignment search tool. *J. Molecular Biol.* **215(3)**: 403-410.

AMIEL, A.R., H.T. JOHNSTON, K. NEDONCELLE, J.F. WARNER, S.FERREIRA & E. RÖTTINGER, 2015. Characterization of morphological and cellular events underlying oral regeneration in the sea anemone, Nematostella vectensis. *International Journal of Molecular Sciences* **16(12)**: 28449-28471.

ARVESTAD, L., 2018. alv: a console-based viewer for molecular sequence alignments. *J. Open Source Softw.* **3(31):** 955. https://doi.org/10.21105/joss.00955.

BATEMAN, A., L. COIN, R. DURBIN, R.D. FINN, V. HOLLICH, S. GRIFFITHS-JONES, A. KHANNA, M. MARSHALL, S. MOXON, E.L. SONNHAMMER & D.J. STUDHOLME, 2004. The Pfam protein families database. *Nucleic Acids Res.* **32(suppl_1)**: D138-D141.

CHEN, K., D. DURAND & M. FARACH-COLTON, 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Computational Biol.* **7(3-4)**: 429-447.

DUCHEMIN, W., G. GENCE, A.M. ARIGON CHIFOLLEAU, L. ARVESTAD, M.S. BANSAL, V. BERRY, B. BOUSSAU, F. CHEVENET, N. COMTE, A.A. DAVÍN & C. DESSIMOZ, 2018. RecPhyloXML: a format for reconciled gene trees. *Bioinformatics* **34(21)**: 3646-3652.

EDGAR, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32(5):** 1792-1797.

HAVRILAK, J.A., L. AL-SHAER, N. BABAN, N. AKINCI & M.J. LAYDEN, 2021. Characterization of the dynamics and variability of neuronal subtype responses during growth, degrowth, and regeneration of Nematostella vectensis. *BMC Biology* **19(104)**: 1-19.

HOANG, D.T., O. CHERNOMOR, A. VON HAESELER, B.Q. MINH & L.S. VINH, 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35(1)**: 518–522.

JONES, P., D. BINNS, H.Y. CHANG, M. FRASER, W. LI, C. MCANULLA, H. MCWILLIAM, J. MASLEN, A. MITCHELL, G. NUKA & S. PESSEAT, 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30(9)**: 1236-1240.

JUNIER, T. & E.M. ZDOBNOV, 2010. The Newick Utilities: High-throughput Phylogenetic tree Processing in the UNIX Shell. *Bioinformatics* **26(13):** 1669–1670. https://doi.org/10.1093/bioinformatics/btq243.

LEMOINE, F. & A. WANG, 2017. Gotree. GitHub repository. https://github.com/evolbioinfo/gotree.

LUO, C & J.Y. SHEN, 2017. Adducin in tumorigenesis and metastasis. *Oncotarget* **8(29)**: 48453-48459.

NGUYEN, L.-T., H.A. SCHMIDT, A. VON HAESELER & B.Q. MINH, 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* **32(1)**: 268-274.

NOTREDAME, C., D.G. HIGGINS, J. HERINGA, 2000. T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment. *J. Mol. Biol.* **302(1)**: 205-217.

PORATH, H.T., A.A. SCHAFFER, P. KANIEWSKA, S. ALON, E. EISENBERG, J. ROSENTHAL, E.Y. LEVANON & O. LEVY, 2017. A-to-I RNA editing in the earliest-diverging eumetazoan phyla. *Molecular Biology and Evolution* **34(8)**: 1890-1901.

PUTNAM, N.H., M. SRIVASTAVA, U. HELLSTEN, B. DIRKS, J. CHAPMAN, A. SALAMOV, A. TERRY, H. SHAPIRO, E. LINDQUIST, V.V. KAPITONOV & J. JURKA, 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317(5834)**: 86-94.

PUTNAM, N.H., M. SRIVASTAVA, U. HELLSTEN, B. DIRKS, J. CHAPMAN, A. SALAMOV, A. TERRY, H. SHAPIRO, E. LINDQUIST, V.V. KAPITONOV, J. JURKA, G. GENIKHOVICH, I.V. GRIGORIEV, S.M. LUCAS, R.E. STEELE, J.R. FINNERTY, U. TECHNAU, M.Q. MARTINDALE & D.S. ROKHSAR, 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317(5834)**: 86-94.

REITZEL, A., P. BURTON, C. KRONE & J. FINNERTY, 2007. Comparison of developmental trajectories in the scarlet sea anemone Nematostella vectensis: embryogenesis, regeneration, and two forms of asexual fission. *Invertebrate Biology* **126(2)**: 99-112.

SHEN, W., S. LE, Y. LI, F. HU, 2016. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLOS ONE.

TRIPODI, G., F. VALTORTA, L. TORIELLI, E. CHIEREGATTI, S. SALARDI, L. TRUSOLINO, A. MENEGON, P. FERRARI, P. MARCHISIO & G. BIANCHI, 1996. Hypertension-associated point mutations in the adducin alpha and beta subunits affect actin cytoskeleton and ion transport. *The American Society for Clinical Investigation* **97(12)**: 2815-2822.

ZHANG, H., S. GAO, M.J. LERCHER, S. HU & W.H. CHEN, 2012. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. Nucleic acids research **40(Web Server issue)**: W569–W572.