Rachel Diao

A13061993

BGGN213

Find a Gene Project Q1-Q4

1. ATG9A (homo sapiens) – autophagy-related 9A, accession [**NP\_001070666.1**](https://www.ncbi.nlm.nih.gov/protein/NP_001070666.1). This protein has been implicated as an important regulator of autophagosome formation.
2. I used NCBI’s tblastn, which searches an amino acid sequence against translated nucleotide databases, to search for matches to the human ATG9A protein accession number from #1 in nematodes (taxid 6231) using the EST (Expressed Sequence Tags) database. I specifically searched against the EST database because I specifically wanted to find matches to the query sequence that were also protein-coding.

Graphical user interface, text, application, email

Description automatically generated

I then selected accession JK285394.1, a 693bp mRNA sequence from *Meloidogyne incognita*, the southern root-knot nematode. This sequence covers 26% of the query sequence and has 33.78% sequence identity, with an e-value of 2e-38.

Text

Description automatically generated

Text

Description automatically generated

Timeline

Description automatically generated with medium confidence

1. Because I used searched the expressed sequence tags (EST) database, there was no need to translate the result because it was already directly matched against protein-coding sequences. The amino acid sequence is below:

>unnamed protein product

LFLAALFWLAHLFRIFCRLVQFSEIRRFFISQLEIADSEMQNLSWAQVVERLCGVQKRLHLIVNREAITP

LDVCQRILRHKNYFVALVYYNILPPKIRLPFLGHVHYLSNGLRFNLEWLLFWGPWSPWKGPYALKDEFKD

PAKLPMIVRQLQRTLTIMAIANLIFFPFVFVYQLLFSFFSYSEHYQRDPNVFGMRKYSNYGREKLRHFNE

MDHELDARLNKS

This is an unnamed protein found in *Meloidogyne incognita* (Eukaryota, Metazoa, Ecdysozoa, Nematoda, Chromadorea, Rhabditida, Tylenchina, Tylenchomorpha, Tylenchoidea, Meloidogynidae, Meloidogyninae, Meloidogyne).

1. I then searched this amino acid sequence against the non-redundant protein sequences database and across all organisms in NCBI’s blastp with the default parameters. The top hit matches to an unnamed protein product in *Meloidogyne enterolobii*, another species of root-knot nematode. This hit has 98.09% identity with an e-value of 3e-149 and covers 94% of the query amino acid sequence. This protein also bears about 60% similarity to homologs of Atg9, the protein that I started with, in other nematodes, as well as to unnamed proteins in other nematodes. Alignment information for a few hits is provided below.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated

Text

Description automatically generatedText

Description automatically generated with medium confidence

1. I performed a CLUSTAL multiple sequence alignment of my novel protein sequence from *Meloidogyne incognita* with human, mouse, rat, C. elegans, fruit fly, and zebrafish Atg9a/Atg9 sequences using MUSCLE. For ease of viewing, only the portion of the alignment where the *M. incognita* novel sequence aligns with the other sequences is shown.

**>sp|Q7Z3C6|ATG9A\_HUMAN Autophagy-related protein 9A OS=Homo sapiens OX=9606 GN=ATG9A PE=1 SV=3**

MAQFDTEYQRLEASYSDSPPGEEDLLVHVAEGSKSPWHHIENLDLFFSRVYNLHQKNGFT

CMLIGEIFELMQFLFVVAFTTFLVSCVDYDILFANKMVNHSLHPTEPVKVTLPDAFLPAQ

VCSARIQENGSLITILVIAGVFWIHRLIKFIYNICCYWEIHSFYLHALRIPMSALPYCTW

QEVQARIVQTQKEHQICIHKRELTELDIYHRILRFQNYMVALVNKSLLPLRFRLPGLGEA

VFFTRGLKYNFELILFWGPGSLFLNEWSLKAEYKRGGQRLELAQRLSNRILWIGIANFLL

CPLILIWQILYAFFSYAEVLKREPGALGARCWSLYGRCYLRHFNELEHELQSRLNRGYKP

ASKYMNCFLSPLLTLLAKNGAFFAGSILAVLIALTIYDEDVLAVEHVLTTVTLLGVTVTV

CRSFIPDQHMVFCPEQLLRVILAHIHYMPDHWQGNAHRSQTRDEFAQLFQYKAVFILEEL

LSPIVTPLILIFCLRPRALEIIDFFRNFTVEVVGVGDTCSFAQMDVRQHGHPQWLSAGQT

EASVYQQAEDGKTELSLMHFAITNPGWQPPRESTAFLGFLKEQVQRDGAAASLAQGGLLP

ENALFTSIQSLQSESEPLSLIANVVAGSSCRGPPLPRDLQGSRHRAEVASALRSFSPLQP

GQAPTGRAHSTMTGSGVDARTASSGSSVWEGQLQSLVLSEYASTEMSLHALYMHQLHKQQ

AQAEPERHVWHRRESDESGESAPDEGGEGARAPQSIPRSASYPCAAPRPGAPETTALHGG

FQRRYGGITDPGTVPRVPSHFSRLPLGGWAEDGQSASRHPEPVPEEGSEDELPPQVHKV

>novel (Meloidogyne incognita)

LFLAALFWLAHLFRIFCRLVQFSEIRRFFISQLEIADSEMQNLSWAQVVERLCGVQKRLH

LIVNREAITPLDVCQRILRHKNYFVALVYYNILPPKIRLPFLGHVHYLSNGLRFNLEWLL

FWGPWSPWKGPYALKDEFKDPAKLPMIVRQLQRTLTIMAIANLIFFPFVFVYQLLFSFFS

YSEHYQRDPNVFGMRKYSNYGREKLRHFNEMDHELDARLNKS

**>tr|Q9TXN6|Q9TXN6\_CAEEL Autophagy-related protein 9 OS=Caenorhabditis elegans OX=6239 GN=atg-9 PE=1 SV=2**

MFNSQSKRAYQQIDDDFDDEVLRNSTCTSRFMQGWGSSTRSLLFGGASNDEQRNLIASSS

SHHSYHDSPAEEPPETHYEQFTATHNHGPPTMASSSQLNSRRWDHVLNLDEFFTHIYEYH

QNGGYLCIVLQKVFSLLQFIFVMSFTTFFTQCVNYQFLFANTNVTSHGTVNQGKRHFGDA

VVDNCPAHISIWMIFAILAAIVYWITRVIKHAQYIMKMSEIQQFYAHELKIADDQLPNLT

WHAIVKRICEAQKKLRLSIHQDNITSIYIYHRILRYKNYMTGMINKRILHPVFDVPFLGP

IAYLPNNLKHEIERILFTSSTSAWTNGPNLREEYKHHEQLDMAAKKMKEDCKFYGFLSLV

LMPLLLPFQIMESFFSLTELIKRRPDGLGMRRYSNFGRYLLRHFNELDHELSARLNRSHI

YAAAYMDQFFSPVLEIAAKNITFIAAAVFGVLTILSAWDEDVLQVEHVITVLTICGIVVL

VCRGMIPDENLVWQPEILMTHVTSELHYLPSTWKGKAHTTGVRHEFDQLFQMKWMFFVLE

LTSPIFTPFVLLFWLRPRCSQLANFFHDYTERVDGLGDVCSFAVMDVGKHGDPKWNHIKE

LKAIVEDQEDQQQAQSVVTSLNRARDGKTELSILHFKTTNPEWQPPKASEKFLRKFRNRL

GQEASMLAPLTSMHLGQQMDRQQQQDQILPPRNILLESVHSIVPTTSGGISASQVAPGRH

PLIGDGLHRIDGPVGNAFQGIQGAKLGSGGVLASLYQEQPRAAESLSNSLRASGVDIDGA

GAEMRINALFLRGLHDESIIHSSSRNYGGTTSSFNMHPTAMQSVFAMPDGFGQPAPAVES

SLIDIETPTYHRESVQRSAAEHKMEHETPEPLAELPDLPGPSSSEHQQRSRLVPNTQHRQ

INEEEEEEEEEDNTPPLSFSS

**>sp|Q68FE2|ATG9A\_MOUSE Autophagy-related protein 9A OS=Mus musculus OX=10090 GN=Atg9a PE=1 SV=2**

MAQFDTEYQRLEASYSDSPPGEEDLLVHVAEGSKSPWHHIENLDLFFSRVYNLHQKNGFT

CMLIGEMFELMQFLFVVAFTTFLVSCVDYDILFANKMVNHSLHPTEPVKVTLPDAFLPAQ

VCSARIQENGSLITILVIAGVFWIHRLIKFIYNICCYWEIHSFYLHALRIPMSALPYCTW

QEVQARIVQTQKEHQICIHKRELTELDIYHRILRFQNYMVALVNKSLLPLRFRLPGLGEV

VFFTRGLKYNFELILFWGPGSLFLNEWSLKAEYKRGGQRLELAQRLSNRILWIGIANFLL

CPLILIWQILYAFFSYAEVLKREPGALGARCWSLYGRCYLRHFNELEHELQSRLNRGYKP

ASKYMNCFLSPLLTLLAKNGAFFAGSILAVLIALTIYDEDVLAVEHVLTTVTLLGVTVTV

CRSFIPDQHMVFCPEQLLRVILAHIHYMPDHWQGNAHRSQTRDEFAQLFQYKAVFILEEL

LSPIVTPLILIFCLRPRALEIIDFFRNFTVEVVGVGDTCSFAQMDVRQHGHPQWLSGGQT

EASVYQQAEDGKTELSLMHFAITNPGWQPPRESTAFLGFLKEQVQRDGAAAGLAQGGLLP

ENALFTSIQSLQSESEPLSLIANVVAGSSCRGPSLSRDLQGSRHRADVASALRSFSPLQP

GAAPQGRVPSTMTGSGVDARTASSGSSVWEGQLQSLVLSEYASTEMSLHALYMHQLHKQQ

TQAEPERHVWHRRESDESGESAPEEGGEGARAPQPIPRSASYPCATPRPGAPETTALHGG

FQRRYGGITDPGTVPRGPSHFSRLPLGGWAEDGQPASRHPEPVPEEGSEDELPPQVHKV

**>sp|Q5FWU3|ATG9A\_RAT Autophagy-related protein 9A OS=Rattus norvegicus OX=10116 GN=Atg9a PE=1 SV=1**

MAQFDTEYQRLEASYSDSPPGEEDLLVHVAEGSKSPWHHIENLDLFFSRVYNLHQKNGFT

CMLIGEIFELMQFLFVVAFTTFLVSCVDYDILFANKMVNHSLHPTEPVKVTLPDAFLPAQ

VCSARIQENGSLITILVIAGVFWIHRLIKFIYNICCYWEIHSFYLHALRIPMSALPYCTW

QEVQARIVQTQKEHQICIHKRELTELDIYHRILRFQNYMVALVNKSLLPLRFRLPGLGEV

VFFTRGLKYNFELILFWGPGSLFLNEWSLKAEYKRGGQRLELAQRLSNRILWIGIANFLL

CPLILIWQILYAFFSYAEVLKREPGALGARCWSLYGRCYLRHFNELEHELQSRLNRGYKP

ASKYMNCFLSPLLTLLAKNGAFFAGSILAVLIALTIYDEDVLAVEHVLTTVTLLGVTVTV

CRSFIPDQHMVFCPEQLLRVILAHIHYMPDHWQGNAHRSQTRDEFAQLFQYKAVFILEEL

LSPIVTPLILIFCLRPRALEIIDFFRNFTVEVVGVGDTCSFAQMDVRQHGHPQWLSGGQT

EASVYQQAEDGKTELSLMHFAITNPGWQPPRESTAFLGFLKEQVQRDGAAAGLAQGGLLP

ENALFTSIQSLQSESEPLSLIANVVAGSSCRGPPLSRDLQGSRHRADVASALRSFSPLQP

GQAPQGRVPSTMTGSGVDARTASSGSSVWEGQLQSLVLSEYASTEMSLHALYMHQLHKQQ

TQAEPERHVWHRRESDESGESAPEEGGEGARAPQPIPRSASYPCATPRPGAPETTALHGG

FQRRYGGITDPGTVPRAPSHFSRLPLGGWAEDGQPASRHPEPVPEEGSEDELPPQVHKV

**>tr|Q961S5|Q961S5\_DROME Autophagy-related protein 9 OS=Drosophila melanogaster OX=7227 GN=Atg9 PE=2 SV=1**

MTAILAHVHYLPSEWRQQAHTTKVRQEFSNFFQFKAGYLLSEIFSPFVTPFVLIFVFRPK

AIELVRFFRTFTVSVRGVGNVCSFAQMDVRKHGNPDWQLTSELEEMTRATAQQPQQEPQQ

QSLAGGKTEMSLLRFTLNNPEWQMPKEAKQFLRGVREHAVGELVQAKTSMVQENPLTNSL

ISFGTMGADYCSIANSVLTAQVTPQQLEISQSLRPGLGPVSGGFPVAASDFRQMLQQNLS

ASVGPLDSMRRLRLSRAEGRLEGPTDTLLYGLCGVDPRVGSTPLNVGVADMCLSALYLHE

LNQQKRQARQSRIDEAEDERPGTSHWPPRPPAAPSADTGFGSRHTVITSKAAESTPLLGS

IRS

**>tr|A2RV24|A2RV24\_DANRE Autophagy-related protein 9 OS=Danio rerio OX=7955 GN=atg9a PE=2 SV=1**

MAHFDTEYQRLEASYSDSPPGEENLLVHVPDGSKSPWHHIENLDLFFQRVYNLHQKNGFT

CMLLGEIFELVQLVFVVAFTVFLANCVDYDILFANKFVNHTDSLKVTLPDAFLPVDVCSA

RIRDSVPVIFILVISGVFWLHRLVKFIYNICCYWEIGSFYINALKISMADLPYFTWQEVQ

ARIVEIQKEHQICIHKKELSELDIYHRILRFKNYMVAMVNKSLLPVRFRLPVLGDTVFYT

RGLKYNFELIFFWGPGSLFENEWSLKSEYKRGGNRLELADRLSSRILWIGIANLLLCPVI

LIWQILYAFFSYTEVVKREPGSLGARCWSLYGRFYLRHFNELDHELMSRLSKGYKASSKY

MNCFMSPLLTVVAKNVAFFAGSILAVLIALTIYDEDVLAVEHVLSSITLLGVCITVCRSF

IPDKHMVFCPEQLLKVILAHIHYMPDHWQGNAHRYETRDEFAQLFQYKAVFILEELLSPV

ITPFILIFCLRRKSLEIIDFFRNFTVDVVGVGDTCSFAQMDVRQHGHPAWMSAGKTEASI

YQQAEDGKTELSLMHFAITNPHWQPPRESTHFISLLKEKVHRDAAVGQQGIIAENAGFTS

THSLHNDSEPRSLIANLLMGPPSLASLHLGREGSINHVSIGVSEGASALRSLSPVSTSLH

LRGSYPSARLPRSDHPAVVAGRGMAGSGTDARTISSGSSAWEGQLTIMILSEYASTEMSI

HALYMHEMHKQQSRGELSRHTWHRQESDESSESVNEDVEAARNFPRSSTFPCTTTSHQEG

AAAQQSGSQRRQGGTSDASSGSFRVQRTPRMAMGGWSEENQTSRHHDPVPEEGSEDELPP

HIHKVT

A picture containing calendar

Description automatically generated

1. I created a phylogenetic tree using the EBI’s online “Simple Phylogeny” tool with a distance-based approach.

Chart

Description automatically generated with medium confidence

1. I saved the alignment file generated through MUSCLE in question #5 as a .clw file, so I used Seaview to convert the file into Fasta format for compatibility with Bio3D in R. I read in the newly converted alignment file using read.fasta(), then I generated a sequence identity matrix using seqidentity() and made a heatmap with heatmap().

Chart, box and whisker chart

Description automatically generated

1. To find a sequence to search against PDB, the protein structure database, I generated a consensus sequence with consensus(). However, there appeared to be too many gaps in this consensus sequence, so I chose to use the original novel protein sequence from *Meloidogyne incognita* instead. I generated a Fasta file for this sequence using Seaview and then read it into R using read.fasta() and searched this sequence against PDB using blast.pdb(). The following are my top hits. 2 of the top 3 hits were from the same group for the same paper (human ATG9A), so I needed to move on to the 4th hit. The first three hits were for human ATG9A, while the 4th was for Atg9 in *S. pombe*.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Technique** | **Resolution** | **Organism** | **E-value** | **Identity** |
| 7JLO | ELECTRON MICROSCOPY | 3.40 | Homo sapiens | 2.66e-48 | 36.937 |
| 6WQZ | ELECTRON MICROSCOPY | 2.80 | Homo sapiens | 1.58e-47 | 36.937 |
| 7D0I | ELECTRON MICROSCOPY | 3.00 | Schizosaccharomyces pombe | 3.07e-21 | 29.565 |

1. I chose to make a VMD visualization of 7JLO, human ATG9A. The protein is trimeric, so I chose to color the protein by chain and used the NewCartoon drawing method to visualize secondary structure. This structure only has 36.937% sequence identity with the novel protein from *M. incognita*, so it’s somewhat similar but the novel protein from *M. incognita* is likely just a much shorter and primitive form of ATG9A or potentially even just consists of just one of these chains. ATG9A was also recently discovered to be a transmembrane protein that bears similarity to lipid scramblases, so this could also be a potential function for the novel protein in *M. incognita*.

A close-up of a bracelet

Description automatically generated with medium confidence

1. I searched my novel protein sequence in ChEMBL and found 14855 targets and 1,458,215 assays. However, none of these hits were very specific to autophagy or other members of the autophagy-related protein (Atg) family. There was one target hit (CHEMBL 4651323) for the lipid bilayer that could be promising due to ATG9A’s and therefore potentially this protein’s nature as a transmembrane protein and potentially its function as a scramblase. ATG9A is also known to cycle between various membrane-bound compartments (endosomes, autophagosomes, Golgi network) in vesicles, so this could provide insight into also understanding this protein’s interactions with various membranes that it is embedded in/travels to/interacts with.