

Rachel Diao
A13061993
BGGN213
Find a Gene Project Q1-Q4

1. ATG9A (homo sapiens) – autophagy-related 9A, accession **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.

The screenshot displays the NCBI BLAST search interface. The 'Enter Query Sequence' section includes a text box with 'ref|NP_001070666.1', a 'Query subrange' section with 'From' and 'To' input fields, and an 'Or, upload file' section with a 'Choose File' button and 'No file chosen' text. Below this is a 'Job Title' field containing 'NP_001070666:autophagy-related protein 9A...' and a checkbox for 'Align two or more sequences'. The 'Choose Search Set' section features a 'Database' dropdown set to 'Expressed sequence tags (est)', an 'Organism' field set to 'Nemata (taxid:6231)' with an 'exclude' checkbox and an 'Add organism' button, and checkboxes for 'Exclude Models (XM/XP)', 'Exclude Uncultured/environmental sample sequences', and 'Limit to Sequences from type material'. There is also an 'Entrez Query' field and a 'BLAST' button. At the bottom, there is a checkbox for 'Show results in a new window'.

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 $2e-38$.

✓	AIAl-aad27f10.b1 Ancylostoma caninum_EST_L3 Activated Ancylostoma caninum cDNA similar to refNP_503178.1... Ancylostoma cani...	186	186	24%	2e-52	44.93%	688	FC555287.1
✓	CPAD-aaa82h01.b1 PB2801_EST_CPAD1 Caenorhabditis brenneri cDNA similar to refNP_503178.1 Putative plas... Caenorhabditis b...	177	177	24%	2e-49	41.35%	672	FF086878.1
✓	HTAB-aab31f06.b1 Heterorhabditis bacteriophora_HTAB2_EST Heterorhabditis bacteriophora cDNA similar to ref... Heterorhabditis b...	160	160	25%	2e-43	40.00%	660	ES740852.1
✓	BJ147598 unpublished oligo-capped cDNA library_C. elegans L1 stage Caenorhabditis elegans cDNA clone yk1253... Caenorhabditis el...	160	160	18%	2e-43	44.77%	667	BJ147598.1
✓	AIaE-aaa18f10.b1 Ancylostoma caninum_EST_Female_pSMART Ancylostoma caninum cDNA similar to gb AAO2... Ancylostoma cani...	157	157	19%	2e-42	46.67%	622	EX561961.1
✓	rq04e11.y1 Heterodera glycines J3 Heterodera glycines cDNA 5' similar to TR-Q9V7N0 Q9V7N0 CG3615 PROTEIN... Heterodera glycin...	154	154	24%	3e-41	37.75%	639	CB377913.1
✓	rq09b11.y1 Heterodera glycines J3 Heterodera glycines cDNA 5' similar to TR-Q9V7N0 Q9V7N0 CG3615 PROTEIN... Heterodera glycin...	150	150	23%	6e-40	37.69%	629	CB378231.1
✓	mij215b23r1.1 Meloidogyne incognita J2 cDNA library Meloidogyne incognita cDNA_mRNA sequence	146	146	26%	2e-38	33.78%	693	JK285394.1
✓	NEFSJ21001_A06_g_037.ab1 Meloidogyne incognita J2 Meloidogyne incognita cDNA_mRNA sequence	145	145	23%	1e-37	38.27%	805	JZ761220.1
✓	rq29e05.y1 Heterodera glycines J3 Heterodera glycines cDNA 5' similar to TR-Q9V7N0 Q9V7N0 CG3615 PROTEIN... Heterodera glycin...	142	142	22%	2e-37	37.84%	593	CB824663.1

mij215b23r1.1 Meloidogyne incognita J2 cDNA library Meloidogyne incognita cDNA, mRNA sequence

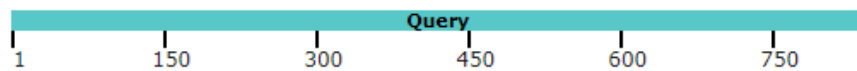
Sequence ID: [JK285394.1](#) Length: 693 Number of Matches: 1

Range 1: 26 to 691 [GenBank](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps	Frame
146 bits(369)	2e-38	Compositional matrix adjust.	82/222(37%)	134/222(60%)	0/222(0%)	+2
Query 136	LVIAGVFIHRLIKFIYNICCYWEIHSFYHLALRIPMSALPYCTWQEVQARIVQTQKEHQ	195				
	L +A +FW+ L + + + EI F++ L I S + +W +V R+ QK					
Sbjct 26	LFLAALFWLAHLFRIFCRLVQFSEIRRRFFISQLEIADSEMQNLSWAQVVERLCGVQKRLH	205				
Query 196	ICIHKRELTELDIYHRLRFQNYMVALVNKS11p1rfr1pg1gEAVFFTRGLKYNFELIL	255				
	+ +++ +T LD+ RILR +NY VALV ++LP + RLP LG + + GL++N E +L					
Sbjct 206	LIVNREAITPLDVCQRILRHKNYFVALVYYNLPKIRLPFLGHVHYLSNGLRFNLEWLL	385				
Query 256	FWGPGSLFLNEWSLKA EYKRGGRLELAQRLSNRILWIGIANFLLCPLILIWQIYAFFS	315				
	FWGP S + ++LK E+K + + ++L + + IAN + P + ++Q+L++FFS					
Sbjct 386	FWGPWSPWKGYPYALKDEFKPAKLPMIVRQLQRTLTIMAIANLIFFPFVQVYQLLFSFFS	565				
Query 316	YAEVLKREPGALGARCWSLYGRCYLRHFNEHEHELQSRLNRG	357				
	Y+E +R+P G R +S YGR LRHFNE++HEL +RLN+					
Sbjct 566	YSEHYQRDPNVFGMRKYSNYGREKLRFHNEMDHELDARLNKS	691				

Distribution of the top 1 Blast Hits on 1 subject sequences



- 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
LFLAALFWLAHLFRIFCRLVQFSEIRRRFFISQLEIADSEMQNLSWAQVVERLCGVQKRLHLIVNREAITP
LDVCQRILRHKNYFVALVYYNLPKIRLPFLGHVHYLSNGLRFNLEWLLFWGPWSPWKGYPYALKDEFKD
PAKLPMIVRQLQRTLTIMAIANLIFFPFVQVYQLLFSFFSYSEHYQRDPNVFGMRKYSNYGREKLRFHNE
MDHELDARLNKS
```

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

- 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.

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

>unnamed protein product
LFLAALFWLAHLFRIFCRLVQFSEIRRFFISQLEIADSEMQLNSWAQVVERLC
GVQKRLHLIVNREAITP
LDVCQRILRHKNYFVALVYNYLPPKIRLPFLGHVHYLSNGLRFNLEWLLFW

Query subrange [?](#)
From
To

Or, upload file
 No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Databases
☒ Standard databases (nr etc.): New ☐ Experimental databases

[Try experimental clustered nr database](#) [?](#)
For more info see [What is clustered nr?](#)

Standard

Database
?

Organism
Optional
 ☐ exclude
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude
Optional
☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Compare
☐ Select to compare standard and experimental database [?](#)

Program Selection

Algorithm
☐ Quick BLASTP (Accelerated protein-protein BLAST)
☒ blastp (protein-protein BLAST)
☐ PSI-BLAST (Position-Specific Iterated BLAST)
☐ PHI-BLAST (Pattern Hit Initiated BLAST)
☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)
Choose a BLAST algorithm [?](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	unnamed protein product [Meloidogyne enterolobii]	Meloidogyne enterolobii	436	436	94%	3e-149	98.09%	520	CAD2198015.1
<input checked="" type="checkbox"/>	unnamed protein product [Meloidogyne enterolobii]	Meloidogyne enterolobii	441	441	94%	4e-146	98.56%	941	CAD2148733.1
<input checked="" type="checkbox"/>	unnamed protein product [Meloidogyne enterolobii]	Meloidogyne enterolobii	437	437	94%	6e-145	97.61%	930	CAD2144253.1
<input checked="" type="checkbox"/>	autophagy protein Apg9 containing protein [Aphelenchus avenae]	Aphelenchus avenae	281	281	94%	2e-85	60.77%	861	KAH7713105.1
<input checked="" type="checkbox"/>	autophagy protein ATG9 [Bursaphelenchus xylophilus]	Bursaphelenchus xylophilus	270	270	94%	6e-82	58.85%	789	QCC87393.1
<input checked="" type="checkbox"/>	unnamed protein product [Bursaphelenchus xylophilus]	Bursaphelenchus xylophilus	270	270	94%	1e-81	58.85%	837	CAD5232042.1
<input checked="" type="checkbox"/>	unnamed protein product [Toxocara canis]	Toxocara canis	266	266	91%	3e-80	57.35%	783	VDM46992.1
<input checked="" type="checkbox"/>	unnamed protein product [Bursaphelenchus okinawaensis]	Bursaphelenchus okinawaensis	259	259	94%	9e-78	58.85%	786	CAD5224214.1
<input checked="" type="checkbox"/>	unnamed protein product [Brugia timori]	Brugia timori	256	256	94%	8e-77	57.42%	760	VDO36727.1
<input checked="" type="checkbox"/>	unnamed protein product [Brugia pahangi]	Brugia pahangi	256	256	94%	2e-76	57.42%	814	VDN92577.1
<input checked="" type="checkbox"/>	Autophagy protein Apg9 containing protein [Brugia malayi]	Brugia malayi	256	256	94%	2e-76	57.42%	813	XP_042937241.1
<input checked="" type="checkbox"/>	unnamed protein product [Litomosoides sigmodontis]	Litomosoides sigmodontis	248	248	94%	5e-76	55.98%	506	VDK84396.1

unnamed protein product [Meloidogyne enterolobii]Sequence ID: [CAD2198015.1](#) Length: 520 Number of Matches: 1Range 1: 18 to 226 [GenPept](#) [Graphics](#)

Score	Expect	Method	Identities	Positives	Gaps
436 bits(1120)	3e-149	Compositional matrix adjust.	205/209(98%)	208/209(99%)	0/209(0%)
Query 14	RIFCRLVQFSEIRRRFISQLEIADSEMQLNSWAQVVERLCGVQKRLHLIVNREAITPLDV				73
Sbjct 18	RIFCRLVQFSEIRRRFISQLEIADSEMQLNSWAQVVERLCGVQKRLHLIVNREAITPLDV				77
Query 74	CQRLRHKNYFVALVYNNILPPKIRLPFLGHVHYLSNGLRFNLEWLLFWGPNSPWKGPYA				133
Sbjct 78	CQRLRHKNYFVALVYNNILPPKIRLPFLGHVHYLSNGLRFNLEWLLFWGPNSPWKGPYA				137
Query 134	LKDEFKDPKALPHIVRQLQRTLTITAIANLIFFPFVYQLLFSFFSYSEHYQRPDPMVFG				193
Sbjct 138	LKDEFKDPKALPHIVRQLQRTLTITAIANLIFFPFVYQLLFSFFSYSEHYQRPDPMVFG				197
Query 194	MRKYSNYGREKLRHFNEMDHEDARLNKS		222		
Sbjct 198	MRKYSNYGREKLRHFNEMDHEDARLNKS		226		

[Download](#) [GenPept](#) [Graphics](#)**unnamed protein product [Meloidogyne enterolobii]**Sequence ID: [CAD2148733.1](#) Length: 941 Number of Matches: 1Range 1: 192 to 400 [GenPept](#) [Graphics](#)

Score	Expect	Method	Identities	Positives	Gaps
441 bits(1133)	4e-146	Compositional matrix adjust.	206/209(99%)	208/209(99%)	0/209(0%)
Query 14	RIFCRLVQFSEIRRRFISQLEIADSEMQLNSWAQVVERLCGVQKRLHLIVNREAITPLDV				73
Sbjct 192	RIFCRLVQFSEIRRRFISQLEIADSEMQLNSWAQVVERLCGVQKRLHLIVNREAITPLDV				251
Query 74	CQRLRHKNYFVALVYNNILPPKIRLPFLGHVHYLSNGLRFNLEWLLFWGPNSPWKGPYA				133
Sbjct 252	CQRLRHKNYFVALVYNNILPPKIRLPFLGHVHYLSNGLRFNLEWLLFWGPNSPWKGPYA				311
Query 134	LKDEFKDPKALPHIVRQLQRTLTITAIANLIFFPFVYQLLFSFFSYSEHYQRPDPMVFG				193
Sbjct 312	LKDEFKDPKALPHIVRQLQRTLTITAIANLIFFPFVYQLLFSFFSYSEHYQRPDPMVFG				371
Query 194	MRKYSNYGREKLRHFNEMDHEDARLNKS		222		
Sbjct 372	MRKYSNYGREKLRHFNEMDHEDARLNKS		400		

autophagy protein Apg9 containing protein [Aphelenchus avenae]Sequence ID: [KAH7713105.1](#) Length: 861 Number of Matches: 1Range 1: 182 to 390 [GenPept](#) [Graphics](#)

Score	Expect	Method	Identities	Positives	Gaps
281 bits(719)	2e-85	Compositional matrix adjust.	127/209(61%)	169/209(80%)	0/209(0%)
Query 14	RIFCRLVQFSEIRRRFISQLEIADSEMQLNSWAQVVERLCGVQKRLHLIVNREAITPLDV				73
Sbjct 182	R+ RL+Q EIR FF + LEI+DSE+ N +W +VV R+C VQ +LHLI+N+E ITPLDV				241
Query 74	CQRLRHKNYFVALVYNNILPPKIRLPFLGHVHYLSNGLRFNLEWLLFWGPNSPWKGPYA				133
Sbjct 242	YQRLRYKNYMIAMVYNNILPPSVHVPFVGSVPVLS5GLRLNLEWLLFWGPNSPWR55YS				301
Query 134	LKDEFKDPKALPHIVRQLQRTLTITAIANLIFFPFVYQLLFSFFSYSEHYQRPDPMVFG				193
Sbjct 302	LKEEVRTGESIDVLAERMRTVVMGIANLVFFPFVFLYQVLFSSFYADLIREPDVFG				361
Query 194	MRKYSNYGREKLRHFNEMDHEDARLNKS		222		
Sbjct 362	TRKYSNYGRKLRHFNEDHEDARLNKS		390		

[Download](#) [GenPept](#) [Graphics](#)**autophagy protein ATG9 [Bursaphelenchus xylophilus]**Sequence ID: [QCC87393.1](#) Length: 789 Number of Matches: 1Range 1: 169 to 377 [GenPept](#) [Graphics](#)

Score	Expect	Method	Identities	Positives	Gaps
270 bits(690)	6e-82	Compositional matrix adjust.	123/209(59%)	163/209(77%)	0/209(0%)
Query 14	RIFCRLVQFSEIRRRFISQLEIADSEMQLNSWAQVVERLCGVQKRLHLIVNREAITPLDV				73
Sbjct 169	R+ RL+Q EI++F+ +L+I D+++ N+ W +VV+R+C Q HLIVN++ I+ LDV				228
Query 74	CQRLRHKNYFVALVYNNILPPKIRLPFLGHVHYLSNGLRFNLEWLLFWGPNSPWKGPYA				133
Sbjct 229	YQRLRYKNYMIGLVNDVLSLEISLPFIGKVDYLS5GLRLNLEWLLFRGPWAPWKGPYA				288
Query 134	LKDEFKDPKALPHIVRQLQRTLTITAIANLIFFPFVYQLLFSFFSYSEHYQRPDPMVFG				193
Sbjct 289	LKDEFKDPESLAAITEQFEKITIYMGIANLVLPFVFLYQVLFSSFYADLVRRDPVFG				348
Query 194	MRKYSNYGREKLRHFNEMDHEDARLNKS		222		
Sbjct 349	SRKYSNYGRKLRHFNEDHEDARLNKS		377		

5. 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
MAQFDTEYQRLEASYSDSPPGEDLLVHVAEGSKSPWHHIENLDLFFSRVYNLHQKNGFT
CMLIGEIFELMQFLFVVAFTTFLVSCVDYDILFANKMVNHSLHPTPEVKVTLPLDAFLPAQ
VCSARIQENGSLTILVIAGVFWIHRLIKFIYNICCYWEIHSFYLHALRIPMSALPYCTW
QEVQARIVQTQKEHQICIHKRELTELDIYHRILRFQNYMVALVNSKSLPLRFRPLPGLGEA
VFTRGLKYNFELILFWGPGSLFLNEWSLKAERYKRGGRLELAQRLSNRILWIGIANFLL
CPILILIWIQILYAFFSYAEVLKREPGALGARCWSLYGRCYLRHFNELEHELQSRNLNGYKP
ASKYMNCFLSPLTLTLANGAFFAGSILAVLIATIIYDEDVLAVEHVLTVTTLGGVTVT
CRSFIPOQHMVFCPEQLLRVILAHIHYPMDHWQGNNAHRSQTRDEFAQLFQYKAVFILEEL
LSPIVTPTLILIFCLRPRALEIIDFRNFTVEVVGVDTCSTFAQMDVRQHGHPQWLSAGQT
EASVYQQAEDGKTELSLMHFAITNPGWQPPRESTAFGLFKEQVQRDGAASLAQGGLLP
ENALFTSIQSLQSESEPLSIANVVGSSCRGPPLPRDLQGSRRHAEVASALRSFSPLPQ
GQAPTGRAHSTMTGSGVDARTASSGSSVWEGQLQSLVLSEYASTEMSLHALYMHQLHKQQ
AQAEPERHVWHRRRESDESSESAPDEGGEGARAPQSI PRSASYPCAAPRPGAPETTALHGG
FQRRYGGITDPGTVPRVPSHFSRLPLGGWAEDQASASRHPEPVPEEGSEDELPPQVHKV
>novel (Meloidogyne incognita)
LFLAALFWLAHLFRIFCRLVQFSEIRRRFISQLEIADSEMQLNSWAQVVERLCGVQKRLH
LIVNREAITPLDVCQIRLRHKNYFVALVYNNILPPKIRLPFLGHVHYLSNGLRFNLEWLL
FWGPNSPWKGPYALKDEFKDPKALPMIVRQLQRTLTITAIANLIFFPFVYQLLFSFFS
YSEHYQRPDPMVFGMRKYSNYGREKLRHFNEMDHEDARLNKS
>tr|Q9TXN6|Q9TXN6_CAEE Autophagy-related protein 9 OS=Caenorhabditis elegans OX=6239 GN=atg-9
PE=1 SV=2
MFNSQSKRAYQQIDDDFDEVLNRNSTCTSRFMQGWGSSSTRSLFLFGASNDEQRNLIASSS
SHHSYHDSFAEPPETHYEQFTATHNHGPPTMASSSQLNSRRWDHVLNLDEFFTHIYEHY
QNGGYLCIVLQKVFSLQFIFVMSFTTFTTQCVNYQFLFANTNVTSHGTNVNQKRRHFGDA
VVDNCPAHISIWMIFAILAAIVYWITRVIKHAQYIMKSEIQQFYAHELKIADDQLPNLT
WHAIVKRICEAQKRLRLSIHQDNITSYIYHRLRYKNYMTGMINKRILHPVFDVPFLGP
IAYLPNNLKHEIERILFTSSTSAWTNGPNLREEYKHHEQLDMAAKMKMEDCKFYGFSLSLV
LMPLLLPQIMESFFSTLELIKRRPDGLGMRRYSNFGRYLLRHFNELDHELSARLNRSHI
YAAAYMDQFFSPVLEIAAKNITFIAAAVFGVLTILSAWDEDVLQVEHVITVLTICGIVVL
VCRGMIPDENLVWQPEILMTHVTSSELHYLPSTWKGKAHTTGVRHEFDQLFQMKWMFFVLE
```

LTSPIFTFPFVLLFWLRPRCSQLANFFHDYTERVDGLGDVCSFAVMDVGKHGDPKWNHIKE
LKAIVEDQEDQQQAQSVVTSLNRRARDGKTELSILHFKTTNPEWQPPKASEKFLRKFRNRL
GQEASMLAPLTSMLGQMDRQQQDDQILPPRNILLESVHSIVPTTSGGISASQVAPGRH
PLIGDGLHRIDGPGVNAFGQIQGAKLGSGGVLASLYQEQPRAAESLSNSLRASGVDDIDGA
GAEMRINALFLRGLHDESIHSSSRNYGGTTSSFNMHPTAMQSVFAMPDGFQGPAPAVES
SLIDIETPTYHRESVQRSAAEHKMEHETPEPLAELPDLPGPSSSEHQQRSRLVPNTQHRQ
INEEEEEEEEDNTPPLSFSS

>sp|Q68FE2|ATG9A_MOUSE Autophagy-related protein 9A OS=Mus musculus OX=10090 GN=Atg9a PE=1 SV=2

MAQFDTEYQRLEASYSDSPPGEEEDLLVHVAEGSKSPWHHIENLDLFFSRVYNLHQKNGFT
CMLIGEMFELMQFLFVVAFTTFLVSCVDYDILFANKMVNHSLHPTEPVKVTLPD AFLPAQ
VCSARIQENGSLITILVIAGVFWIHRLIKFIYNICCYWEIHSFYLHALRIPMSALPYCTW
QEVQARIVQTQKEHQICIKRELTELDIYHRILRFQNYMVALVNKSLPLRFRPLPGLGEV
VFFTRGLKYNFELILFWGPGSLFLNEWSLKA EYKRGGRLELAQRLSNRILWIGIANFLL
CPILILIWQILYAFFSYAEVLKREPGALGARCWSLYGRCYLRHFNELEHEHQRLNRGYKP
ASKYMNCFLSPLLTLLAKNGAFFAGSILAVLIALTIYDEDVLAVEHVLTTVTLLGVTVTV
CRSFI PDQHMVFCPEQLLRVILAHIHYPMDHWQGNHRSTQTRDEFAQLFQYKAVFILEEL
LSPIVTPILILIFCLRPRALEIIDFFRNFTVEVVGVDTC SFAQMDVRQHGHPQWLSGGQT
EASVYQQAEDGKTELSLMHFATNPGWQPPRESTAFLGFLKEQVQRDGAAGLAQGGLLP
ENALFTSIQSLQSESEPLSLIANVVAGSSCRGPSLSRDLQGSRRHRADVASALRSFSPLQP
GAAPQGRVPSTMTGSGVDARTASSGSSVWEGQLQSLVLSEYASTEMSLHALYMHQLHKQQ
TQAEPERHVWHRRESDESSESAP EEGGEGARAPQPIPRSASYPCATPRPGAPETTALHGG
FQRRYGGITDPGTVP RGP SHFSRLPLGWAEDQGPASRHPEVPVEEGSEDELPPQVHKV

>sp|Q5FWU3|ATG9A_RAT Autophagy-related protein 9A OS=Rattus norvegicus OX=10116 GN=Atg9a PE=1 SV=1

MAQFDTEYQRLEASYSDSPPGEEEDLLVHVAEGSKSPWHHIENLDLFFSRVYNLHQKNGFT
CMLIGEIFELMQFLFVVAFTTFLVSCVDYDILFANKMVNHSLHPTEPVKVTLPD AFLPAQ
VCSARIQENGSLITILVIAGVFWIHRLIKFIYNICCYWEIHSFYLHALRIPMSALPYCTW
QEVQARIVQTQKEHQICIKRELTELDIYHRILRFQNYMVALVNKSLPLRFRPLPGLGEV
VFFTRGLKYNFELILFWGPGSLFLNEWSLKA EYKRGGRLELAQRLSNRILWIGIANFLL
CPILILIWQILYAFFSYAEVLKREPGALGARCWSLYGRCYLRHFNELEHEHQRLNRGYKP
ASKYMNCFLSPLLTLLAKNGAFFAGSILAVLIALTIYDEDVLAVEHVLTTVTLLGVTVTV
CRSFI PDQHMVFCPEQLLRVILAHIHYPMDHWQGNHRSTQTRDEFAQLFQYKAVFILEEL
LSPIVTPILILIFCLRPRALEIIDFFRNFTVEVVGVDTC SFAQMDVRQHGHPQWLSGGQT
EASVYQQAEDGKTELSLMHFATNPGWQPPRESTAFLGFLKEQVQRDGAAGLAQGGLLP
ENALFTSIQSLQSESEPLSLIANVVAGSSCRGPPLSRDLQGSRRHRADVASALRSFSPLQP
GQAPQGRVPSTMTGSGVDARTASSGSSVWEGQLQSLVLSEYASTEMSLHALYMHQLHKQQ
TQAEPERHVWHRRESDESSESAP EEGGEGARAPQPIPRSASYPCATPRPGAPETTALHGG
FQRRYGGITDPGTVP RAS SHFSRLPLGWAEDQGPASRHPEVPVEEGSEDELPPQVHKV

>tr|Q961S5|Q961S5_DROME Autophagy-related protein 9 OS=Drosophila melanogaster OX=7227 GN=Atg9 PE=2 SV=1

MTAILAHVHYLPSEWRQQAHTTKVRQEFSNFFQFKAGYLLSEIFSPFVTPFVLI FVFRPK
AIELVRFFRTFTVSVRGVGNVCSFAQMDVRKHGNDWQLTSELEEMTRATAQQPQQEPQQ
QSLAGGKTEMSSLRFTLNNPEWQMPKEAKQFLRGVREHAVGELVQAKTSMVQENPLTNSL
ISFGTMGADYCSIANSVLTAQVTPQQL EISQSLRPGLGPVSGGFVAASDFRQMLQQNLS
ASVGPLDSMRRLRLSRAEGRLEGPTDTLLYGLCGVDPRVGSTPLNVGVADMCLSALYLHE
LNQQRQARQSRIDEAEDERPGTSHWPPRPPAAPSADTGFGRHTVITSKAAESTPLLGSI
IRS

>tr|A2RV24|A2RV24_DANRE Autophagy-related protein 9 OS=Danio rerio OX=7955 GN=atg9a PE=2 SV=1

MAHFDTEYQRLEASYSDSPPGEEENLLVHVPDGSKSPWHHIENLDLFFQRVYNLHQKNGFT
CMLLGEIFELVQLFVVAFTVFLANCVDYDILFANKFVNHTDSLKVTLPDAFLPVDVCSA
RIRDSVPVIFILVISGVFWLHRLVKFIYNICCYWEIGSFYINALKISMADLPYFTWQEVQ
ARIVEIQKEHQICIKKELSELDIYHRILRFKNYMMVAMVNKSLPLVRFRPLPVLGDTVFTY
RGLKYNFELIFFWGPGSLFENEWSL KSEYKRGGNRLELADRLSSRILWIGIANLLCPVI
LIWQILYAFFSYTEVVKREPGSLGARCWSLYGRFYLRHFNELDHELMSRLSKGYKASSKY
MNCFMSPLLTVVAKNVAFFAGSILAVLIALTIYDEDVLAVEHVLSSITLLGVCITVCRSF
IPDKHMFVFCPEQLLKVILAHIHYPMDHWQGNHRYETRDEFAQLFQYKAVFILEELSPV
ITPFILIFCLRRLSLEIIDFFRNFTVDVVGVDTC SFAQMDVRQHGHPAWMSAGKTEASI
YQQAEDGKTELSLMHFATNPHWQPPRESTHFI SLLKEKVHRDAAVGQQGIIAENAGFTS
THSLHNDSEPRSLIANLMLGPPSLASLHLGREGSINHVSIGVSEGASALRSLSPVSTSLH
LRGSYPSARLPRSDHPAVVAGRMAGSGTDARTISSGSSAWEGQLTIMILSEYASTEMSI
HALYMHMHKQSRGELSRTHRQESDESSESVNEDVEAARNFPRSSTFPCTTTS HQEG
AAAQQSGSQRRQGGTSDASSGSRVQRTPRMAMGGWSEENQTSRHHPVPVEEGSEDELFP
HIHKVT

```

Drosophila_melanogaster -----
Danio_rerio              TLPDAFLPVDVCSARIRDSVPVIFILVISGVFWLHRLVKFIYNICCYWEIGSFYINALKI
Homo_sapiens             TLPDAFLPAQVCSARIQENGSLITILVIAGVFWIHLRIKFYINICCYWEIHSFYHLALRI
Mus_musculus             TLPDAFLPAQVCSARIQENGSLITILVIAGVFWIHLRIKFYINICCYWEIHSFYHLALRI
Rattus_norvegicus       TLPDAFLPAQVCSARIQENGSLITILVIAGVFWIHLRIKFYINICCYWEIHSFYHLALRI
Caenorhabditis_elegans  ----HFGDAVVDNCPAHISINMIFAILAAIVYWITRVIKHAQYIMKMSEIQQFYAHELKI
NOVEL_Meloidogyne_incognita -----LFLAALFWLAHLFRIFCRLVQFSEIRFFISQLEI

```

```

Drosophila_melanogaster -----
Danio_rerio              SMADLPYFTWQEVQARIVEIQKEHQICIKKELSELDIYHRILRFKNYMMAMVNKSLLPV
Homo_sapiens             PMSALPYCTWQEVQARIVQTQKEHQICIKKRELTELDIYHRILRFQNYMVALVNKSLLP
Mus_musculus             PMSALPYCTWQEVQARIVQTQKEHQICIKKRELTELDIYHRILRFQNYMVALVNKSLLP
Rattus_norvegicus       PMSALPYCTWQEVQARIVQTQKEHQICIKKRELTELDIYHRILRFQNYMVALVNKSLLP
Caenorhabditis_elegans  ADDQLPNLTHAIVKRICEAQKKLRLSIHQDNITSIYIYHRILRYKNYMTGMINKRIHP
NOVEL_Meloidogyne_incognita ADSEMQLNSMAQVVERLCGVQKRLHLIVNREAITPLDVCQIRLRHKNYFVALVYNYILPP

```

```

Drosophila_melanogaster -----
Danio_rerio              RFRLPVLGDTVIFYTRGLKYNFELIFFWGPGSLFENEWSLKSEYKRGGRLEADRLSSRI
Homo_sapiens             RFRLPGLGEAVFFTRGLKYNFELILFWGPGSLFLNEWSLKAeyKRGGRLELAQRLSNRI
Mus_musculus             RFRLPGLGEVFFTRGLKYNFELILFWGPGSLFLNEWSLKAeyKRGGRLELAQRLSNRI
Rattus_norvegicus       RFRLPGLGEVFFTRGLKYNFELILFWGPGSLFLNEWSLKAeyKRGGRLELAQRLSNRI
Caenorhabditis_elegans  VFDVPFLGPIAYLPNNLKHEIERILFTSSTSANTNGPNLREEYKHHEQLDMAAKMKEDC
NOVEL_Meloidogyne_incognita KIRLPFLGHVHYLSNGLRFNLEILLFWGWPSPWKGPYALKDEFKDPAKLPMIVRQLQRTL

```

```

Drosophila_melanogaster -----
Danio_rerio              LWIGIANLLCPVILIWIQIYAFFSYTEVVKREPGSLGARCHSLYGRFYLRHFNELDH
Homo_sapiens             LWIGIANFLLCPLILIWQIYAFFSYAEVLKREPGALGARCHSLYGRFYLRHFNELDH
Mus_musculus             LWIGIANFLLCPLILIWQIYAFFSYAEVLKREPGALGARCHSLYGRFYLRHFNELDH
Rattus_norvegicus       LWIGIANFLLCPLILIWQIYAFFSYAEVLKREPGALGARCHSLYGRFYLRHFNELDH
Caenorhabditis_elegans  KFYGFLSLVLMPLLLPFQIMESFFSLTELIKRRPDGLMRRYSNFGRYLLRHFNELDH
NOVEL_Meloidogyne_incognita TIMAIANLIFFPFVYVQLLSFFSYSEHYQRDPNVFGMRKYSNYGREKLRHFNEMDH

```

```

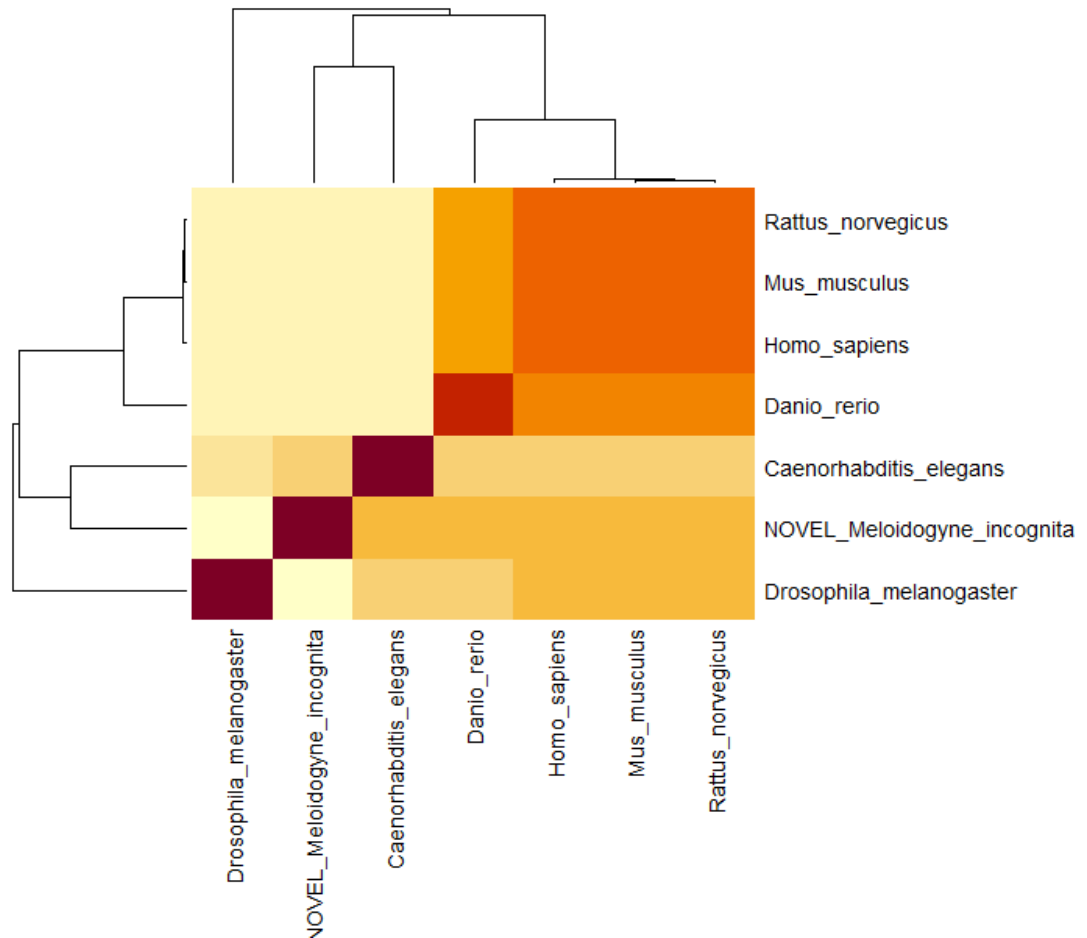
Drosophila_melanogaster -----
Danio_rerio              MSRLSKGYKASSKYMNCFMSPLLTVVAKNVAFFAGSILAVLIALTIYDEVDLAVEHVLSS
Homo_sapiens             QSRLNRGYKPASKYMNCFLSPLLTLLAKNGAFFAGSILAVLIALTIYDEVDLAVEHVLTT
Mus_musculus             QSRLNRGYKPASKYMNCFLSPLLTLLAKNGAFFAGSILAVLIALTIYDEVDLAVEHVLTT
Rattus_norvegicus       QSRLNRGYKPASKYMNCFLSPLLTLLAKNGAFFAGSILAVLIALTIYDEVDLAVEHVLTT
Caenorhabditis_elegans  SARLNRSHYIAAYMDQFFSPVLEIAAKNITFIAAAVFGVLTILSAWDEVDLQVEHVITV
NOVEL_Meloidogyne_incognita DARLNKS-----

```

6. I created a phylogenetic tree using the EBI's online "Simple Phylogeny" tool with a distance-based approach.



7. 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().

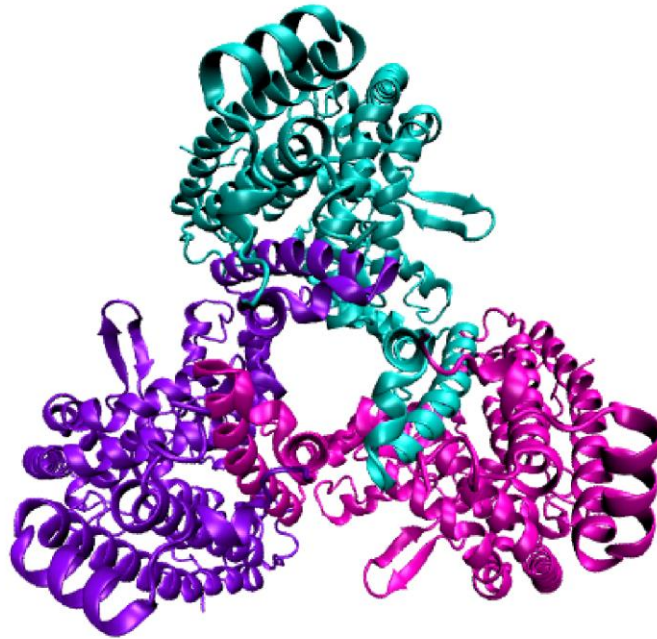


8. To find a sequence to search against PDB, the protein structure database, I generated a consensus sequence with consensus(). However, there appeared to be too many gaps in this consensus sequence, so I chose to use the original novel protein sequence from *Meloidogyne incognita* instead. I generated 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
7DOI	ELECTRON MICROSCOPY	3.00	Schizosaccharomyces pombe	3.07e-21	29.565

9. 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*.



10. 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.