Identification of novel mechanosensory genes in *Caenorhabditis Elegans*

Khalid Yasseen

University of California, Berkeley; Department of Molecular & Cell Biology

Abstract

Caenorhabditis Elegans is a non-parasitic nematode that has been well-established as a model organism. *C. elegans* has been a useful model for mechanosensory responses and has previously identified a gene class of *mec* mutants that produces imperative mechanosensitive proteins. In the present study, a genetic screen was performed to identify additional mechanosensory genes that could produce mechanosensory mutants. After random and unbiased introduction of gene mutations, 8 mutants with mechanosensory phenotypes were chosen and assayed. Of the 8, 1 was shown to be a wild type, while 4 were members of the *mec* gene class and 3 were novel genes: *unc, rol,* and *dpy*. The *unc* mutant was characterized by very low responsiveness (mean touch response index = 8.67, $p < 0.01$) and was homologous to an UNC-119 protein in many other species. The *mec* mutants observed were that of *mec-4, mec-10,* and *mec-7*, though the phenotypes were very similar and only genetic sequencing was able to differentiate them. Furthermore, one *mec* mutant, *mec-7,* was found to have a human orthologous protein in the form of TUBB6 and TUBB8 which elucidates a possible pathway for future work.

Introduction

Caenorhabditis Elegans are free-living, hermaphroditic non-parasitic soil nematode (Hart and Chao, 2009; Strange, 2006). Adult *C. elegans* have a size of ~1 mm in length, a rapid life cycle and a large number of offspring (Hart and Chao, 2009; Brenner, 1974). In 1963, Sydney Brenner recognized the advantages of these characteristics from a molecular biology perspective and began research on *C. elegans* (Brenner, 1974; Strange, 2006). The usefulness of *C. elegans* for genetic studies has been exploited to address a number of biological problems ranging from aging to cell cycle control to synaptic transmission (Brenner, 1974; Strange, 2006). The nervous system of *C. elegans* is composed of only 302 neurons, which is a large advantage for neurobiology studies (Strange, 2006). A further important characteristic of *C. elegans* is its mechanosensitive and chemosensitive interactions with its environment (Strange, 2006; Hart and Chao, 2009). This allows experimenters to perform mechanosensory assays that can directly test the mechanosensory mechanisms in *C. elegans* which can be used to understand mechanosensation in humans. Such an experiment identified five important neurons necessary for mechanosensory responses by killing selected neurons with a laser (Bianchi, 2007). The mechanosensory response was assessed by a "gentle touch response" assay, although other methods have been demonstrated (Chalfie et al., 2013; Shaw et al., 2016). The "gentle touch response" assay consists of scoring responsiveness of worms when they are exposed to a light or gentle stimulus. One of the findings in *C. elegans* is that of mechanosensory mutants that have diminished responsiveness as a result of specific gene mutations. One such mutant class is that of the *mec* gene, which includes the *mec-4* and *mec-10* gene (Shi et al., 2018). The proteins MEC-4 and MEC-10 make up a mechanosensitive sodium channel. In the present study, a gentle touch response assay will be used in a systematic search of more mechanosensory mutants and locate any associated genes.

Methods

Randomly mutated loci were introduced genome-wide in *C. elegans* in an unbiased manner and 8 candidate phenotypes that seem to have some sort of movement and/or mechanosensory defect were identified and used. A "gentle touch assay" was then performed to characterize the touch response of the worms. This assay utilized an eyebrow hair attached to a toothpick which was then used to gently stimulate the *C. elegans* mutants. Responses were assessed by reversals, which included backwards movements, stopping, starting, and accelerating movements. The mutant worms were touched 10 times in this fashion, and their responses were recorded. Stimulation was switched between tail touch and head touch to prevent desensitization and to test for region selective mutants. This 10-touch test was performed on ten worms for each mutant strain. Two controls were present, a negative wild type control and a positive *mec*-10 control. Results were analyzed by calculating a touch response index then performing statistical analysis using Microsoft Excel (Microsoft Corporation, Microsoft Office 365 ProPlus 2016). Averages were calculated for the responses (yielding a 0 – 1 score), then the average of each worms average score was calculated. This average was reported in percentage form (0 – 100) and combined with multiple iterations of this assay. The mutant genes were then mapped and sequenced. Genes were characterized using ApE (A Plasmid Editor), Expasy, and the Basic Local Alignment Search Tool (BLAST).

Results

Responses were scored as a "1" or "0" based on the demonstration of a reversal behavior. Score was averaged to give a "touch response index" on a scale of 0 to 100. **Figure 1** depicts the scoring of a wild type negative control and known mechanosensory positive control, which includes touch response index per worm. Row averages represent possible

sensitization/desensitization measures that could represent a confounding variable. Each column was averaged, and the final touch response index for each worm strain was calculated by averaging the touch response index from each. The final averages of touch response indices for each worm were analyzed, descriptive statistics depicted in **Figure 2.** In **Figure 3**, touch response index as an average percentage (from 0 to 100) is plotted for each unknown mutant strain. A single factor ANOVA was performed which found a significant difference between groups ($p < 0.001$), shown in **Figure 4**. Post-hoc Tukey analysis (**Figure 5**) shows statistical significance in differences between several groups, especially those against wild type (Unknown B, Unknown C, and Unknown F). Genes were mapped and sequenced and were characterized using Expasy and BLAST tools. **Figure 6** (A-H) shows wild-type versus mutant alignment and translation of sequenced gene to identify the mutation. Then the translated protein was run through Expasy to identify important relevant protein motifs. The protein was also processed through BLAST to ascertain related genes in different species. Of those unknowns that were chosen, Unknown H was shown to have no mutation after alignment (**Figure 6. H**) and translation. Unknowns A, B, and D had a nonsense mutation, while C, E, F, and G had missense mutations. Unknown A (**Figure 6. A**) had a C to T substitution at position 836 which gives rise to a stop codon, resulting in a truncated protein. Its average touch response index was 70.16 (σ^2 = 22.23, SE = 9.07, against WT $p = 0.106$). This protein has two known motifs similar to signature tubulin subunits. Unknown B had a C to T substitution at position 308 resulting in a premature stop codon. The average touch response index was 8.67 ($\sigma^2 = 8.38$, SE = 3.42, against WT $p < 0.01$). This protein had no known structural motifs based on the amino acid sequence. For Unknown C, there is a three-nucleotide missense from CTT to GCC at position 2162 which gives rise to an Ala to Leu amino acid substitution. Its average touch response index was 70.33 (σ^2 =

18.86, SE = 6.29, against WT $p < 0.05$). This protein showed similarity to the amiloride-sensitive sodium channel signature motif. For Unknown D, there was a C to T substitution at position 566 resulting in a stop codon. The average touch response index was 82 ($\sigma^2 = 8.06$, SE = 2.69, against WT $p = 0.88$). Upon observation, Unknown D had a starkly unique morphological phenotype which was expressed as a short and fat worm. Unknown E had a G to A substitution at position 213 which results in an Arg to His substitution. The average touch response index was 71.22 ($\sigma^2 = 19.41$, SE = 6.47, against WT $p = 0.055$). Like Unknown D, Unknown E also had a unique phenotype, which was expressed as a rolling movement when stimulated. Unknown worm F had a G to C substitution at position 2027 which results in a Gly to Arg substitution. The average touch response index was 36.11 ($\sigma^2 = 10.14$, SE = 3.38, against WT $p < 0.01$). For Unknown G, there was a C to T substitution at position 315 which resulted in a Phe to Ser substitution. The average touch response index was 70 ($\sigma^2 = 20.86$, SE = 8.52, against WT $p = 0.10$). Lastly, Unknown H had no mismatches between the wild type and mutant versions, however this sequence had an odd open reading frame which resulted in three stop codons. The average touch response index was 78.5 ($\sigma^2 = 20.16$, SE = 8.23, against WT $p = 0.71$). Unknown F and G shared the same amiloride-sensitive sodium channel signature motif with Unknown C according to Expasy.

H/T	1	2	3	4	5	6	7	8	9	10	
Head	1	1	1	1	0	1	1	1	1	1	0.9
Tail	1	1	1	0	1	1	0	1	1	1	0.8
Head	1	0	1	1	1	1	0	1	1	1	0.8
Tail	1	1	1	1	1	1	1	0	1	0	0.8
Head	1	0	1	1	1	1	1	1	0	0	0.7
Tail	1	1	1	1	1	1	1	1	0	1	0.9
Head	0	1	1	1	1	0	0	1	1	1	0.7
Tail	0	0	1	1	1	1	1	1	1	1	0.8
Head	1	1	1	1	0	1	1	1	0	1	0.8
Tail	1	1	1	1	0	1	0	0	1	1	0.7
Avg:	**0.8**	**0.7**	**1**	**0.9**	**0.7**	**0.9**	**0.6**	**0.8**	**0.7**	**0.8**	**0.79**
Head	0	0	1	0	0	1	1	0	0	0	0.3
Tail	0	0	0	0	0	1	0	0	0	1	0.2
Head	0	0	0	1	0	0	0	0	0	0	0.1
Tail	0	1	1	1	0	0	0	0	0	1	0.4
Head	1	0	0	0	0	1	0	0	0	0	0.2
Tail	0	0	0	0	0	1	0	1	1	0	0.3
Head	1	1	0	0	1	0	1	0	0	1	0.5
Tail	0	0	1	1	0	0	1	1	0	0	0.4
Head	0	0	0	0	0	1	0	0	1	0	0.2
Tail	0	0	1	1	1	0	0	0	0	1	0.4
Avg:	**0.2**	**0.2**	**0.4**	**0.4**	**0.2**	**0.5**	**0.3**	**0.2**	**0.2**	**0.4**	**0.3**

Figure 1: Data collection for gentle touch response assay and calculations of touch response index (score 0 – 1) for wild-type and *mec* mutants

Descriptive Statistics	Wildtype Control (touch index)	mec-10 (u20) control	Unknown A	Unknown B	Unknown C	Unknown D	Unknown E	Unknown F	Unknown G	Unknown H
Mean	91.46666667	30.86666667	70.16666667	8.666666667	70.33333333	82	71.22222222	36.11111111	70	78.5
Standard Error	2.218679716	3.441852583	9.07530226	3.422150071	6.287112038	2.687419249	6.471685719	3.380682231	8.516650359	8.233063423
Median	94	29	70	6	78	84	80	36	67.5	83
Mode	94	39	#N/A	#N/A	84	87	80	32	#N/A	#N/A
Standard Deviation	8.592909591	13.33023774	22.2298598	8.382521498	18.86133611	8.062257748	19.41505716	10.14204669	20.8614477	20.16680441
Sample Variance	73.83809524	177.6952381	494.1666667	70.26666667	355.75	65	376.9444444	102.8611111	435.2	406.7
Kurtosis	0.252619441	0.494661121	2.142827513	-1.728546533	3.206674125	1.575908707	-1.650826341	-0.767037275	0.048538789	-1.522708911
Skewness	-1.120199457	0.89633649	-0.758774629	0.645601053	-1.773227358	-1.16844809	-0.787111725	-0.092943112	-0.074348197	-0.535928566
Range	28	46	68	20	58	27	46	31	60	50
Minimum	72	12	32	0	27	65	43	20	39	50
Maximum	100	58	100	20	85	92	89	51	99	100
Sum	1372	463	421	52	633	738	641	325	420	471
Count	15	15	6	6	9	9	9	9	6	6
Confidence Level(95.0%)	4.75859472	7.382039603	23.32880714	8.796916812	14.49810636	6.197199902	14.92373403	7.795867204	21.89274671	21.16376329

Figure 2: Descriptive statistics for average touch response index for all assayed strains

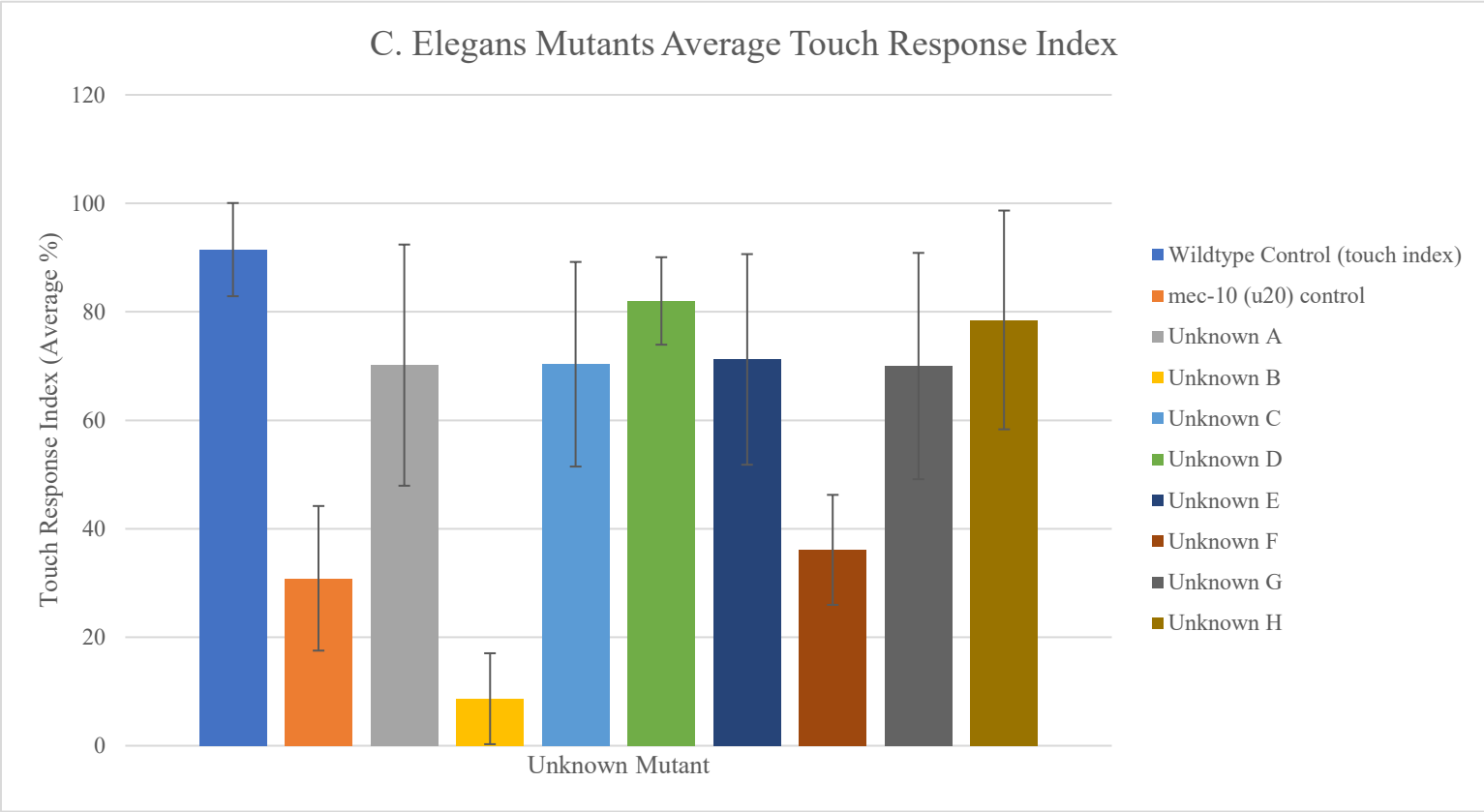

Figure 3: Touch Response Index Averages (percentage 0 – 100 score) plotted against assayed mutants. Error Bars represent standard deviation.

ANOVA: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Wildtype Control (touch index)	15	1372	91.46666667	73.83809524
mec-10 (u20) control	15	463	30.86666667	177.6952381
Unknown A	6	421	70.16666667	494.1666667
Unknown B	6	52	8.666666667	70.26666667
Unknown C	9	633	70.33333333	355.75
Unknown D	9	738	82	65
Unknown E	9	641	71.22222222	376.9444444
Unknown F	9	325	36.11111111	102.8611111
Unknown G	6	420	70	435.2
Unknown H	6	471	78.5	406.7

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	58048.91111	9	6449.879012	29.05747211	1.04028E-21	1.999114806
Within Groups	17757.57778	80	221.9697222			
Total	75806.48889	89				

Figure 4: Single Factor ANOVA results

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence	treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence	treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
WT v. Mec control	15.7533	0.0010053	** p<0.01	Mec-10 v. UnkG	7.69	0.0010053	** p<0.01	UnkC v. UnkD	2.3492	0.7897771	insignificant
WT v. UnkA	4.1856	0.1066835	insignificant	Mec-10 v. UnkH	9.3603	0.0010053	** p<0.01	UnkC v. UnkE	0.179	0.8999947	insignificant
WT v. UnkB	16.2708	0.0010053	** p<0.01	UnkA v. UnkB	10.1112	0.0010053	** p<0.01	UnkC v. UnkF	6.891	0.0010053	** p<0.01
WT v. UnkC	4.7577	0.0369675	* p<0.05	UnkA v. UnkC	0.03	0.8999947	insignificant	UnkC v. UnkG	0.06	0.8999947	insignificant
WT v. UnkD	2.1312	0.8815195	insignificant	UnkA v. UnkD	2.1312	0.8815195	insignificant	UnkC v. UnkH	1.4708	0.8999947	insignificant
WT v. UnkE	4.5576	0.0545331	insignificant	UnkA v. UnkE	0.1901	0.8999947	insignificant	UnkD v. UnkE	2.1702	0.8651015	insignificant
WT v. UnkF	12.4621	0.0010053	** p<0.01	UnkA v. UnkF	6.1335	0.0016308	** p<0.01	UnkD v. UnkF	9.2402	0.0010053	** p<0.01
WT v. UnkG	4.2184	0.1007322	insignificant	UnkA v. UnkG	0.0274	0.8999947	insignificant	UnkD v. UnkG	2.1612	0.8688868	insignificant
WT v. UnkH	2.5481	0.7060937	insignificant	UnkA v. UnkH	1.3701	0.8999947	insignificant	UnkD v. UnkH	0.6304	0.8999947	insignificant
Mec-10 v. UnkA	7.7228	0.0010053	** p<0.01	UnkB v. UnkC	11.1063	0.0010053	** p<0.01	UnkE v. UnkF	7.07	0.0010053	** p<0.01
Mec-10 v. UnkB	4.3625	0.0784946	insignificant	UnkB v. UnkD	13.2075	0.0010053	** p<0.01	UnkE v. UnkG	0.2201	0.8999947	insignificant
Mec-10 v. UnkC	8.885	0.0010053	** p<0.01	UnkB v. UnkE	11.2664	0.0010053	** p<0.01	UnkE v. UnkH	1.3107	0.8999947	insignificant
Mec-10 v. UnkD	11.5115	0.0010053	** p<0.01	UnkB v. UnkF	4.9428	0.0253352	* p<0.05	UnkF v. UnkG	6.1035	0.0017603	** p<0.01
Mec-10 v. UnkE	9.0852	0.0010053	** p<0.01	UnkB v. UnkG	10.0838	0.0010053	** p<0.01	UnkF v. UnkH	7.6343	0.0010053	** p<0.01
Mec-10 v. UnkF	1.1807	0.8999947	insignificant	UnkB v. UnkH	11.4813	0.0010053	** p<0.01	UnkG v. UnkH	1.3975	0.8999947	insignificant

Figure 5: Post-hoc Tukey Analysis of ANOVA results

Unknown A

```
801>gccaggattcgcaccactgactagcagaagcaatcagcagtatcgtgccattactgtccctgagctgacccaacaatgtttcgacgcaaagaacatgatg>900
801>gccaggattcgcaccactgactagcagaagcaattagcagtatcgtgccattactgtccctgagctgacccaacaatgtttcgacgcaaagaacatgatg>900
```

```
Translation 441 a.a. MW=49261.85000000005

MetArgGluIleValHisIleGlnAlaGlyGlnCysGlyAsnGlnIleGlySerLysPhe
TrpGluValIleSerAspGluHisGlyIleAspProSerGlyGlnTyrValGlyAspSer
AspLeuGlnLeuGluArgIleAsnValTyrTyrAsnGluAlaGlySerAsnLysTyrVal
ProArgAlaValLeuValAspLeuGluProGlyThrMetAspSerValArgSerGlyPro
PheGlyGlnLeuPheArgProAspAsnTyrValPheGlyGlnSerGlyAlaGlyAsnAsn
TrpAlaLysGlyHisTyrThrGluGlyAlaGluLeuValAspAsnValLeuAspValVal
ArgLysGluAlaGluSerThrAspCysLeuGlnGlyPheGlnLeuThrHisSerLeuGly
GlyGlyThrGlySerGlyMetGlyThrLeuLeuIleSerLysIleArgGluGluTyrPro
AspArgIleMetAsnThrPheSerValValProSerProLysValSerAspThrValVal
GluProTyrAsnAlaThrLeuSerValHisGlnLeuValGluAsnThrAspSerThrPhe
CysIleAspAsnGluAlaLeuTyrAspIleCysPheArgThrLeuLysLeuThrThrPro
ThrTyrGlyAspLeuAsnHisLeuValSerAlaThrMetSerGlyValThrThrCysLeu
ArgPheProGlyGlnLeuAsnAlaAspLeuArgLysLeuAlaValAsnMetValProPhe
ProArgLeuHisPhePheMetProGlyPheAlaProLeuThrSerArgSerAsnGlnGln
TyrArgAlaIleThrValProGluLeuThrGlnGlnCysPheAspAlaLysAsnMetMet
AlaAlaCysAspProArgHisGlyArgTyrLeuThrAlaAlaAlaIlePheArgGlyArg
MetSerMetLysGluValAspGluGlnMetLeuAsnIleGlnAsnLysAsnSerSerTyr
PheValAspTrpIleProAsnAsnValLysThrAlaValCysAspIleProProArgGly
LeuLysMetSerAlaThrPheIleGlyAsnSerThrAlaIleGlnGluLeuPheLysArg
IleSerGluGlnPheThrAlaMetPheArgArgLysAlaPheLeuHisTrpTyrThrGly
GluGlyMetAspGluMetGluPheThrGluAlaGluSerAsnMetAsnAspLeuValSer
GluTyrGlnGlnTyrGlnGluAlaAlaAlaAspGluAspAlaAlaGluAlaPheAspGly
GluEnd
```

```
Translation 278 a.a. MW=30581.259999999984
???(163 extra codons after stop)
MetArgGluIleValHisIleGlnAlaGlyGlnCysGlyAsnGlnIleGlySerLysPhe
TrpGluValIleSerAspGluHisGlyIleAspProSerGlyGlnTyrValGlyAspSer
AspLeuGlnLeuGluArgIleAsnValTyrTyrAsnGluAlaGlySerAsnLysTyrVal
ProArgAlaValLeuValAspLeuGluProGlyThrMetAspSerValArgSerGlyPro
PheGlyGlnLeuPheArgProAspAsnTyrValPheGlyGlnSerGlyAlaGlyAsnAsn
TrpAlaLysGlyHisTyrThrGluGlyAlaGluLeuValAspAsnValLeuAspValVal
ArgLysGluAlaGluSerThrAspCysLeuGlnGlyPheGlnLeuThrHisSerLeuGly
GlyGlyThrGlySerGlyMetGlyThrLeuLeuIleSerLysIleArgGluGluTyrPro
AspArgIleMetAsnThrPheSerValValProSerProLysValSerAspThrValVal
GluProTyrAsnAlaThrLeuSerValHisGlnLeuValGluAsnThrAspSerThrPhe
CysIleAspAsnGluAlaLeuTyrAspIleCysPheArgThrLeuLysLeuThrThrPro
ThrTyrGlyAspLeuAsnHisLeuValSerAlaThrMetSerGlyValThrThrCysLeu
ArgPheProGlyGlnLeuAsnAlaAspLeuArgLysLeuAlaValAsnMetValProPhe
ProArgLeuHisPhePheMetProGlyPheAlaProLeuThrSerArgSerAsnEndGln
TyrArgAlaIleThrValProGluLeuThrGlnGlnCysPheAspAlaLysAsnMetMet
AlaAlaCysAspProArgHisGlyArgTyrLeuThrAlaAlaAlaIlePheArgGlyArg
MetSerMetLysGluValAspGluGlnMetLeuAsnIleGlnAsnLysAsnSerSerTyr
PheValAspTrpIleProAsnAsnValLysThrAlaValCysAspIleProProArgGly
LeuLysMetSerAlaThrPheIleGlyAsnSerThrAlaIleGlnGluLeuPheLysArg
IleSerGluGlnPheThrAlaMetPheArgArgLysAlaPheLeuHisTrpTyrThrGly
GluGlyMetAspGluMetGluPheThrGluAlaGluSerAsnMetAsnAspLeuValSer
GluTyrGlnGlnTyrGlnGluAlaAlaAlaAspGluAspAlaAlaGluAlaPheAspGly
GluEnd
```

PS00228 **TUBULIN_B_AUTOREG** *Tubulin-beta mRNA autoregulation signal* :

1 - 4: [confidence level: (0)] MREI

PS00227 **TUBULIN** *Tubulin subunits alpha, beta, and gamma signature* :

140 - 146: [confidence level: (0)] GGGTGSG

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☑	PREDICTED: Odobenus rosmarus divergens tubulin beta-2B chain-like (LOC101362359), transcript variant X2, mRNA	805	805	95%	0.0	90.00%	XM_004408404.2
☑	PREDICTED: Leptonychotes weddellii tubulin beta-2B chain-like (LOC102725609), mRNA	805	805	95%	0.0	90.00%	XM_006735691.2
☑	PREDICTED: Erinaceus europaeus tubulin beta-2B chain-like (LOC103112717), transcript variant X1, mRNA	805	805	95%	0.0	90.00%	XM_007522196.2
☑	PREDICTED: Peromyscus leucopus tubulin beta 4B class IVb (Tubb4b), mRNA	804	804	95%	0.0	90.24%	XM_028868716.1
☑	PREDICTED: Mesocricetus auratus tubulin beta 4B class IVb (Tubb4b), mRNA	804	804	95%	0.0	90.24%	XM_005083713.3
☑	PREDICTED: Calidris pugnax tubulin beta-1 chain (LOC106887380), mRNA	804	804	95%	0.0	90.00%	XM_014940844.1
☑	Bos taurus tubulin, beta 2C, mRNA (cDNA clone MGC:128183 IMAGE:7898040), complete cds	804	804	95%	0.0	90.24%	BC105181.1
☑	PREDICTED: Anas platyrhynchos tubulin beta-1 chain (LOC101802187), transcript variant X2, mRNA	803	803	95%	0.0	90.00%	XM_027451707.1
☑	PREDICTED: Capra hircus tubulin beta 4B class IVb (TUBB4B), mRNA	803	803	95%	0.0	90.24%	XM_018056236.1
☑	PREDICTED: Callithrix jacchus tubulin beta 4B class IVb (TUBB4B), mRNA	803	803	95%	0.0	90.24%	XM_009006050.2

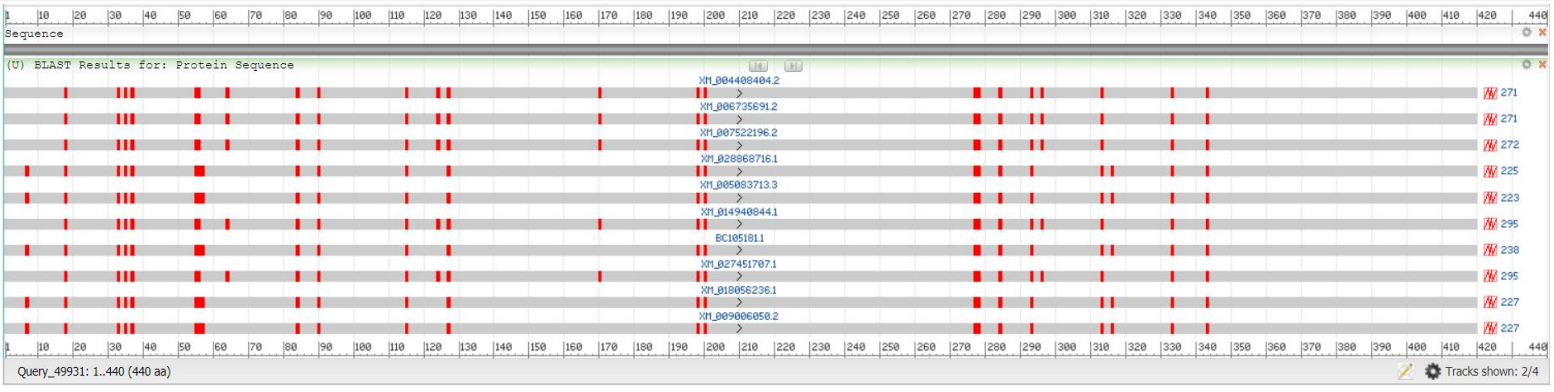

Figure 6a: Alignment with ApE of Unknown A wild type to mutant. ApE translated mutant and wildtype sequences with noted differences. Expasy results with two similar motif sequences of tubulin subunits. BLAST results of first 10 out of 100 similar species with graphical representation.

Unknown B

```
301>caggcgcaagccgaatcggcaagatatgtccgatatcgatttgcgccgaattttctgaaattaaagacggtcggcgcgacggtcgagttcaaggtcggcg>400
301>caggcgtaagccgaatcggcaagatatgtccgatatcgatttgcgccgaattttctgaaattaaagacggtcggcgcgacggtcgagttcaaggtcggcg>400
```

```
Translation 219 a.a. MW=25264.350000000006

MetLysAlaGluGlnGlnGlnGlnSerIleAlaProGlySerAlaThrPheProSerGln
MetProArgProProProValThrGluGlnAlaIleThrThrGluAlaGluLeuLeuAla
LysAsnGlnIleThrProAsnAspValLeuAlaLeuProGlyIleThrGlnGlyPheLeu
CysSerProSerAlaAsnValTyrAsnIleGluPheThrLysPheGlnIleArgAspLeu
AspThrGluHisValLeuPheGluIleAlaLysProGluAsnGluThrGluGluAsnLeu
GlnAlaGlnAlaGluSerAlaArgTyrValArgTyrArgPheAlaProAsnPheLeuLys
LeuLysThrValGlyAlaThrValGluPheLysValGlyAspValProIleThrHisPhe
ArgMetIleGluArgHisPhePheLysAspArgLeuLeuLysCysPheAspPheGluPhe
GlyPheCysMetProAsnSerArgAsnAsnCysGluHisIleTyrGluPheProGlnLeu
SerGlnGlnLeuMetAspAspMetIleAsnAsnProAsnGluThrArgSerAspSerPhe
TyrPheValGluAsnLysLeuValMetHisAsnLysAlaAspTyrSerTyrAspAlaEnd
```

```
Translation 102 a.a. MW=11269.080000000005
???(117 extra codons after stop)
MetLysAlaGluGlnGlnGlnGlnSerIleAlaProGlySerAlaThrPheProSerGln
MetProArgProProProValThrGluGlnAlaIleThrThrGluAlaGluLeuLeuAla
LysAsnGlnIleThrProAsnAspValLeuAlaLeuProGlyIleThrGlnGlyPheLeu
CysSerProSerAlaAsnValTyrAsnIleGluPheThrLysPheGlnIleArgAspLeu
AspThrGluHisValLeuPheGluIleAlaLysProGluAsnGluThrGluGluAsnLeu
GlnAlaEndAlaGluSerAlaArgTyrValArgTyrArgPheAlaProAsnPheLeuLys
LeuLysThrValGlyAlaThrValGluPheLysValGlyAspValProIleThrHisPhe
ArgMetIleGluArgHisPhePheLysAspArgLeuLeuLysCysPheAspPheGluPhe
GlyPheCysMetProAsnSerArgAsnAsnCysGluHisIleTyrGluPheProGlnLeu
SerGlnGlnLeuMetAspAspMetIleAsnAsnProAsnGluThrArgSerAspSerPhe
TyrPheValGluAsnLysLeuValMetHisAsnLysAlaAspTyrSerTyrAspAlaEnd
```

Hits for all PROSITE (release 2019_10) motifs on sequence USERSEQ1 :

no hit!

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☑	PREDICTED: Chanos chanos protein unc-119 homolog A-like (LOC115819047), mRNA	259	259	85%	5e-84	63.83%	XM_030782585.1
☑	PREDICTED: Salarias fasciatus protein unc-119 homolog B-like (LOC115389186), mRNA	260	260	94%	2e-83	57.14%	XM_030092644.1
☑	PREDICTED: Sphaeramia orbicularis protein unc-119 homolog B-like (LOC115418820), transcript variant X2, mRNA	260	260	94%	2e-83	57.89%	XM_030133310.1
☑	PREDICTED: Sphaeramia orbicularis protein unc-119 homolog B-like (LOC115418820), transcript variant X1, mRNA	259	259	94%	3e-83	57.62%	XM_030133309.1
☑	PREDICTED: Cottoperca gobio protein unc-119 homolog B-like (LOC115008288), mRNA	261	261	94%	3e-83	58.57%	XM_029431789.1
☑	PREDICTED: Cimex lectularius protein unc-119 (LOC106664059), transcript variant X1, mRNA	261	261	87%	5e-83	62.83%	XM_014389433.2
☑	PREDICTED: Oreochromis niloticus protein unc-119 homolog B (LOC100700719), transcript variant X4, mRNA	261	261	89%	5e-83	60.80%	XM_003441559.5
☑	PREDICTED: Metaseiulus occidentalis protein unc-119 homolog B (LOC100901037), mRNA	261	261	98%	9e-83	57.40%	XM_003746796.2
☑	PREDICTED: Anolis carolinensis unc-119 lipid binding chaperone B (unc119b), mRNA	256	256	84%	1e-82	62.83%	XM_003226010.3
☑	PREDICTED: Protobothrops mucrosquamatus unc-119 lipid binding chaperone B (UNC119B), mRNA	255	255	84%	1e-82	61.78%	XM_015823035.1

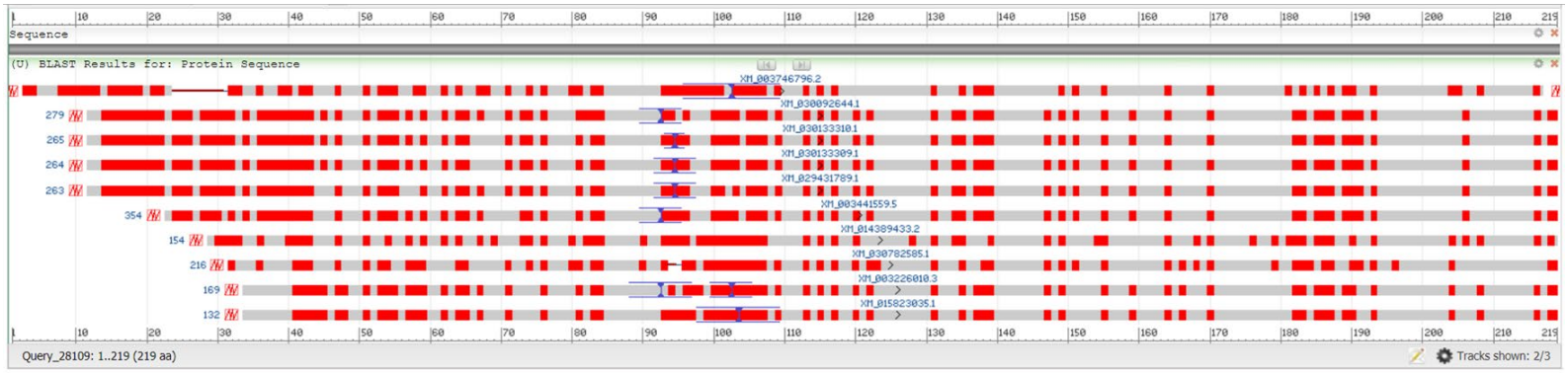

Figure 6b: Alignment with ApE of Unknown B wild type to mutant. ApE translated mutant and wildtype sequences with noted differences. Expasy results shows no similar protein sequences. BLAST results of first 10 out of 100 similar species with graphical representation.

Unknown C

2101>actgaatcagaggcttatgggtttgtcaacttgctagccgattttggtggacaactcggtctttggtgcggaatatccttccttacctgttgcgaatttg>2200
2101>actgaatcagaggcttatgggtttgtcaacttgctagccgattttggtggacaactcggtgccttggtgcggaatatccttccttacctgttgcgaatttg>2200

Translation 768 a.a. MW=87154.75999999992

MetSerTrpMetGlnAsnLeuLysAsnTyrGlnHisLeuArgAspProSerGluTyrMet
SerGlnValTyrGlyAspProLeuAlaTyrLeuGlnGluThrThrLysPheValThrGlu
ArgGluTyrTyrGluAspPheGlyTyrGlyGluCysPheAsnSerThrGluSerGluVal
GlnCysGluLeuIleThrGlyGluPheAspProLysLeuLeuProTyrAspLysArgLeu
AlaTrpHisPheLysGluPheCysTyrLysThrSerAlaHisGlyIleProMetIleGly
GluAlaProAsnValTyrTyrArgAlaValTrpValValLeuPheLeuGlyCysMetIle
MetLeuTyrLeuAsnAlaGlnSerValLeuAspLysTyrAsnArgAsnGluLysIleVal
AspIleGlnLeuLysPheAspThrAlaProPheProAlaIleThrLeuCysAsnLeuAsn
ProTyrLysAlaSerLeuAlaThrSerValAspLeuValLysArgThrLeuSerAlaPhe
AspGlyAlaMetGlyLysAlaGlyGlyAsnLysAspHisGluGluGluArgGluValVal
ThrGluProProThrThrProAlaProThrThrLysProAlaArgArgArgGlyLysArg
AspLeuSerGlyAlaPhePheGluProGlyPheAlaArgCysLeuCysGlySerGlnGly
SerSerGluGlnGluAspLysAspGluGluLysGluGluLeuLeuGluThrThrThr
LysLysValPheAsnIleAsnAspAlaAspGluGluTrpAspGlyMetGluGluTyrAsp
AsnGluHisTyrGluAsnTyrAspValGluAlaThrThrGlyMetAsnMetMetGluGlu
CysGlnSerGluArgThrLysPheAspGluProThrGlyPheAspAspArgCysIleCys
AlaPheAspArgSerThrHisAspAlaTrpProCysPheLeuAsnGlyThrTrpGluThr
ThrGluCysAspThrCysAsnGluHisAlaPheCysThrLysAspAsnLysThrAlaLys
GlyHisArgSerProCysIleCysAlaProSerArgPheCysValAlaTyrAsnGlyLys
ThrProProIleGluIleTrpThrTyrLeuGlnGlyGlyThrProThrGluAspProAsn
PheLeuGluAlaMetGlyPheGlnGlyMetThrAspGluValAlaIleValThrLysAla
LysGluAsnIleMetPheAlaMetAlaThrLeuSerMetGlnAspArgGluArgLeuSer
ThrThrLysArgGluLeuValHisLysCysSerPheAsnGlyLysAlaCysAspIleGlu
AlaAspPheLeuThrHisIleAspProAlaPheGlySerCysPheThrPheAsnHisAsn
ArgThrValAsnLeuThrSerIleArgAlaGlyProMetTyrGlyLeuArgMetLeuVal
TyrValAsnAlaSerAspTyrMetProThrThrGluAlaThrGlyValArgLeuThrIle
HisAspLysGluAspPheProPheProAspThrPheGlyTyrSerAlaProThrGlyTyr
ValSerSerPheGlyLeuArgLeuArgLysMetSerArgLeuProAlaProTyrGlyAsp
CysValProAspGlyLysThrSerAspTyrIleTyrSerAsnTyrGluTyrSerValGlu
GlyCysTyrArgSerCysPheGlnGlnLeuValLeuLysGluCysArgCysGlyAspPro
ArgPheProValProGluAsnAlaArgHisCysAspAlaAlaAspProIleAlaArgLys
CysLeuAspAlaArgMetAsnAspLeuGlyGlyLeuHisGlySerPheArgCysArgCys
GlnGlnProCysArgGlnSerIleTyrSerValThrTyrSerProAlaLysTrpProSer
LeuSerLeuGlnIleGlnLeuGlySerCysAsnGlyThrAlaValGluCysAsnLysHis
TyrLysGluAsnGlyAlaMetValGluValPheTyrGluGlnLeuAsnPheGluMetLeu
ThrGluSerGluAlaTyrGlyPheValAsnLeuLeuAlaAspPheGlyGlyGlnLeuGly
AlaTrpCysGlyIleSerPheLeuThrCysCysGluPheValPheLeuPheLeuGluThr
AlaTyrMetSerAlaGluHisAsnTyrSerLeuTyrLysLysLysLysAlaGluLysAla
LysLysIleAlaSerGlySerPheEnd

Translation 768 a.a. MW=87196.82999999993

MetSerTrpMetGlnAsnLeuLysAsnTyrGlnHisLeuArgAspProSerGluTyrMet
SerGlnValTyrGlyAspProLeuAlaTyrLeuGlnGluThrThrLysPheValThrGlu
ArgGluTyrTyrGluAspPheGlyTyrGlyGluCysPheAsnSerThrGluSerGluVal
GlnCysGluLeuIleThrGlyGluPheAspProLysLeuLeuProTyrAspLysArgLeu
AlaTrpHisPheLysGluPheCysTyrLysThrSerAlaHisGlyIleProMetIleGly
GluAlaProAsnValTyrTyrArgAlaValTrpValValLeuPheLeuGlyCysMetIle
MetLeuTyrLeuAsnAlaGlnSerValLeuAspLysTyrAsnArgAsnGluLysIleVal
AspIleGlnLeuLysPheAspThrAlaProPheProAlaIleThrLeuCysAsnLeuAsn
ProTyrLysAlaSerLeuAlaThrSerValAspLeuValLysArgThrLeuSerAlaPhe
AspGlyAlaMetGlyLysAlaGlyGlyAsnLysAspHisGluGluGluArgGluValVal
ThrGluProProThrThrProAlaProThrThrLysProAlaArgArgArgGlyLysArg
AspLeuSerGlyAlaPhePheGluProGlyPheAlaArgCysLeuCysGlySerGlnGly
SerSerGluGlnGluAspLysAspGluGluLysGluGluLeuLeuGluThrThrThr
LysLysValPheAsnIleAsnAspAlaAspGluGluTrpAspGlyMetGluGluTyrAsp
AsnGluHisTyrGluAsnTyrAspValGluAlaThrThrGlyMetAsnMetMetGluGlu
CysGlnSerGluArgThrLysPheAspGluProThrGlyPheAspAspArgCysIleCys
AlaPheAspArgSerThrHisAspAlaTrpProCysPheLeuAsnGlyThrTrpGluThr
ThrGluCysAspThrCysAsnGluHisAlaPheCysThrLysAspAsnLysThrAlaLys
GlyHisArgSerProCysIleCysAlaProSerArgPheCysValAlaTyrAsnGlyLys
ThrProProIleGluIleTrpThrTyrLeuGlnGlyGlyThrProThrGluAspProAsn
PheLeuGluAlaMetGlyPheGlnGlyMetThrAspGluValAlaIleValThrLysAla
LysGluAsnIleMetPheAlaMetAlaThrLeuSerMetGlnAspArgGluArgLeuSer
ThrThrLysArgGluLeuValHisLysCysSerPheAsnGlyLysAlaCysAspIleGlu
AlaAspPheLeuThrHisIleAspProAlaPheGlySerCysPheThrPheAsnHisAsn
ArgThrValAsnLeuThrSerIleArgAlaGlyProMetTyrGlyLeuArgMetLeuVal
TyrValAsnAlaSerAspTyrMetProThrThrGluAlaThrGlyValArgLeuThrIle
HisAspLysGluAspPheProPheProAspThrPheGlyTyrSerAlaProThrGlyTyr
ValSerSerPheGlyLeuArgLeuArgLysMetSerArgLeuProAlaProTyrGlyAsp
CysValProAspGlyLysThrSerAspTyrIleTyrSerAsnTyrGluTyrSerValGlu
GlyCysTyrArgSerCysPheGlnGlnLeuValLeuLysGluCysArgCysGlyAspPro
ArgPheProValProGluAsnAlaArgHisCysAspAlaAlaAspProIleAlaArgLys
CysLeuAspAlaArgMetAsnAspLeuGlyGlyLeuHisGlySerPheArgCysArgCys
GlnGlnProCysArgGlnSerIleTyrSerValThrTyrSerProAlaLysTrpProSer
LeuSerLeuGlnIleGlnLeuGlySerCysAsnGlyThrAlaValGluCysAsnLysHis
TyrLysGluAsnGlyAlaMetValGluValPheTyrGluGlnLeuAsnPheGluMetLeu
ThrGluSerGluAlaTyrGlyPheValAsnLeuLeuAlaAspPheGlyGlyGlnLeuGly
LeuTrpCysGlyIleSerPheLeuThrCysCysGluPheValPheLeuPheLeuGluThr
AlaTyrMetSerAlaGluHisAsnTyrSerLeuTyrLysLysLysLysAlaGluLysAla
LysLysIleAlaSerGlySerPheEnd

USERSEQ1 (768 aa)

PS01206 **ASC** *Amiloride-sensitive sodium channels signature* :

577 - 597: [confidence level: (0)] YsvEgCyrsCfQqlVLkeCrC

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Helobdella robusta hypothetical protein partial mRNA	176	176	40%	5e-44	31.83%	XM_009027326.1
PREDICTED: Ornithorhynchus anatinus sodium channel epithelial 1 delta subunit (SCNN1D), mRNA	170	226	52%	2e-40	30.89%	XM_029065944.1
Xenopus laevis sodium channel, nonvoltage-gated 1, gamma (scnn1g), mRNA	169	216	52%	1e-39	31.48%	NM_001096984.1
PREDICTED: Pocillopora damicornis amiloride-sensitive sodium channel subunit beta-like (LOC113682090), mRNA	163	163	43%	2e-39	32.29%	XM_027199240.1
Xenopus laevis epithelial sodium channel, gamma subunit, mRNA (cDNA clone MGC:197106 IMAGE:9093761), complete cds	167	211	54%	4e-39	31.08%	BC170379.1
Xenopus laevis epithelial sodium channel, gamma subunit, mRNA (cDNA clone MGC:197110 IMAGE:9093765), complete cds	167	211	54%	4e-39	31.08%	BC170383.1
Xenopus laevis sodium channel, non voltage gated 1 gamma subunit L homeolog (scnn1g.L), mRNA	167	211	54%	5e-39	31.08%	NM_001085654.1
PREDICTED: Acropora millepora amiloride-sensitive sodium channel subunit alpha-like (LOC114954005), transcript variant X3, mRNA	167	218	55%	5e-39	32.07%	XM_029330468.1
Xenopus laevis sodium channel, non voltage gated 1 gamma subunit S homeolog (scnn1g.S), mRNA	167	214	52%	6e-39	31.48%	NM_001085662.1
PREDICTED: Xenopus tropicalis sodium channel, non voltage gated 1 gamma subunit (scnn1g), mRNA	167	215	54%	8e-39	31.38%	XM_004917989.3

Figure 6c: Alignment with ApE of Unknown C wild type to mutant. ApE translated mutant and wildtype sequences with noted differences. Expasy results shows protein motif of amiloride-sensitive sodium channels signature. BLAST results of first 10 out of 100 similar species with graphical representation.

Unknown D

501>tggaccagatggaattccaggagttccaggagttgacggagaagatgctgatgatgccaaggctcagactcaacaatacgatggatgcttcacttgccca>600
501>tggaccagatggaattccaggagttccaggagttgacggagaagatgctgatgatgccaaggctcagactcaacaatacgatggatgcttcacttgccca>600

Translation 352 a.a. MW=34547.57999999993

MetSerValPheAlaGlyTyrAlaAlaCysThrLeuGlyAlaValSerMetLeuLeuCys
ValSerLeuValProGlnValTyrGlnGlnValSerMetLeuArgAspGluLeuThrThr
GluMetGluAlaTrpArgLeuGluSerAspGlnIleTyrMetAspMetGlnLysPheGly
ArgValArgArgGlnAlaGlyGlyTyrGlyGlyTyrGlyGlyTyrGlySerGlyProSer
GlyProSerGlyProSerGlyProHisGlyGlyPheProGlyGlyProGlnGlyHisPhe
ProGlyAsnThrGlySerSerAsnThrProThrLeuProGlyValIleGlyValProPro
SerValThrGlyHisProGlyGlySerProIleAsnProAspGlySerProSerAlaGly
ProGlyAspLysCysAsnCysAsnThrGluAsnSerCysProAlaGlyProAlaGlyPro
LysGlyThrProGlyHisAspGlyProAspGlyIleProGlyValProGlyValAspGly
GluAspAlaAspAspAlaLysAlaGlnThrGlnGlnTyrAspGlyCysPheThrCysPro
AlaGlyProGlnGlyProProGlySerGlnGlyLysProGlyAlaArgGlyMetArgGly
AlaArgGlyGlnAlaAlaMetProGlyArgAspGlySerProGlyMetProGlySerLeu
GlyProIleGlyProProGlyAlaAlaGlyGluGluGlyProThrGlyGluProGlyAla
AspValGluHisGlnIleGlyLeuProGlyAlaLysGlyThrProGlyAlaProGlyGlu
SerGlyAspGlnGlyGluGlnGlyAspArgGlyAlaThrGlyIleAlaGlyProProGly
GluArgGlyProGlnGlyGluLysGlyAspAspGlyProAsnGlyAlaAlaGlySerPro
GlyGluGluGlyGluProGlyGlnAspAlaGlnTyrCysProCysProGlnArgAsnThr
AsnAlaAlaValSerGlyAsnGlnGlyTyrArgAsnEnd

Translation 188 a.a. MW=18742.18999999999
???(164 extra codons after stop)
MetSerValPheAlaGlyTyrAlaAlaCysThrLeuGlyAlaValSerMetLeuLeuCys
ValSerLeuValProGlnValTyrGlnGlnValSerMetLeuArgAspGluLeuThrThr
GluMetGluAlaTrpArgLeuGluSerAspGlnIleTyrMetAspMetGlnLysPheGly
ArgValArgArgGlnAlaGlyGlyTyrGlyGlyTyrGlyGlyTyrGlySerGlyProSer
GlyProSerGlyProSerGlyProHisGlyGlyPheProGlyGlyProGlnGlyHisPhe
ProGlyAsnThrGlySerSerAsnThrProThrLeuProGlyValIleGlyValProPro
SerValThrGlyHisProGlyGlySerProIleAsnProAspGlySerProSerAlaGly
ProGlyAspLysCysAsnCysAsnThrGluAsnSerCysProAlaGlyProAlaGlyPro
LysGlyThrProGlyHisAspGlyProAspGlyIleProGlyValProGlyValAspGly
GluAspAlaAspAspAlaLysAlaEndThrGlnGlnTyrAspGlyCysPheThrCysPro
AlaGlyProGlnGlyProProGlySerGlnGlyLysProGlyAlaArgGlyMetArgGly
AlaArgGlyGlnAlaAlaMetProGlyArgAspGlySerProGlyMetProGlySerLeu
GlyProIleGlyProProGlyAlaAlaGlyGluGluGlyProThrGlyGluProGlyAla
AspValGluHisGlnIleGlyLeuProGlyAlaLysGlyThrProGlyAlaProGlyGlu
SerGlyAspGlnGlyGluGlnGlyAspArgGlyAlaThrGlyIleAlaGlyProProGly
GluArgGlyProGlnGlyGluLysGlyAspAspGlyProAsnGlyAlaAlaGlySerPro
GlyGluGluGlyGluProGlyGlnAspAlaGlnTyrCysProCysProGlnArgAsnThr
AsnAlaAlaValSerGlyAsnGlnGlyTyrArgAsnEnd

Hits for all PROSITE (release 2019_10) motifs on sequence USERSEQ1 :

no hit!

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☑	PREDICTED: Nanorana parkeri scavenger receptor class A member 5 (SCARA5), mRNA	60.1	60.1	12%	6e-06	65.91%	XM_018568477.1
☑	PREDICTED: Ficedula albicollis scavenger receptor class A member 5 (SCARA5), transcript variant X1, mRNA	57.0	57.0	12%	6e-05	65.91%	XM_005062106.1
☑	PREDICTED: Ficedula albicollis scavenger receptor class A member 5 (SCARA5), transcript variant X2, mRNA	56.2	56.2	12%	8e-05	65.91%	XM_016305326.1
☑	PREDICTED: Vulpes vulpes scavenger receptor class A member 5 (SCARA5), mRNA	56.2	56.2	12%	1e-04	61.36%	XM_026007873.1
☑	PREDICTED: Canis lupus familiaris scavenger receptor class A member 5 (SCARA5), transcript variant X2, mRNA	56.2	56.2	12%	1e-04	61.36%	XM_543223.6
☑	PREDICTED: Canis lupus familiaris scavenger receptor class A member 5 (SCARA5), transcript variant X1, mRNA	56.2	56.2	12%	1e-04	61.36%	XM_005635672.3
☑	PREDICTED: Mustela putorius furo scavenger receptor class A, member 5 (SCARA5), transcript variant X3, mRNA	55.5	55.5	12%	2e-04	61.36%	XM_004775130.2
☑	PREDICTED: Mustela putorius furo scavenger receptor class A, member 5 (SCARA5), transcript variant X2, mRNA	55.5	55.5	12%	2e-04	61.36%	XM_013047162.1
☑	PREDICTED: Mustela putorius furo scavenger receptor class A, member 5 (SCARA5), transcript variant X1, mRNA	55.5	55.5	12%	2e-04	61.36%	XM_004775129.2
☑	PREDICTED: Enhydra lutris kenyoni scavenger receptor class A member 5 (LOC111155460), mRNA	55.1	55.1	12%	2e-04	61.36%	XM_022515626.1

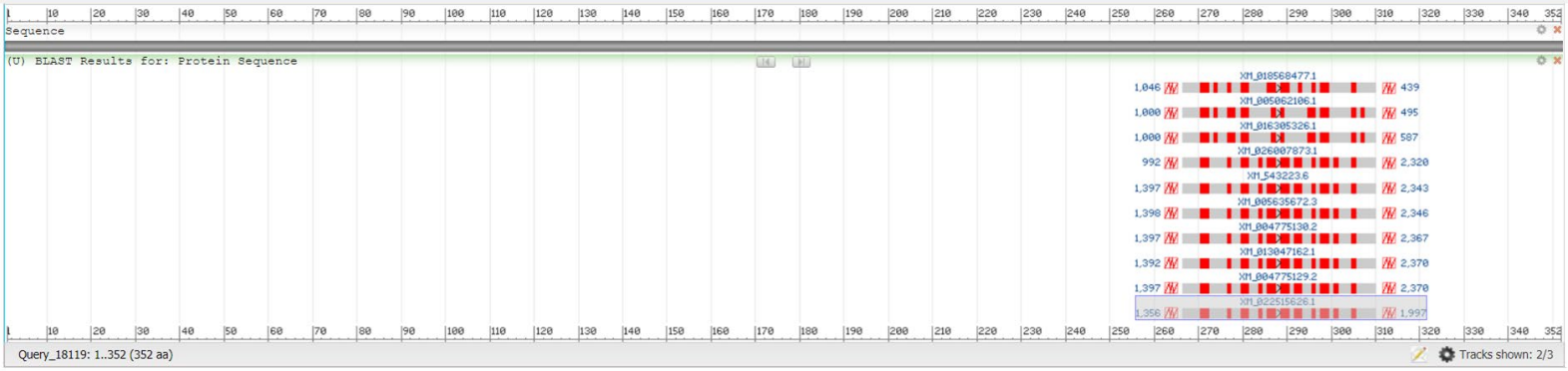

Figure 6d: Alignment with ApE of Unknown D wild type to mutant. ApE translated mutant and wildtype sequences with noted differences. Expasy results shows no similar protein motifs. BLAST results of first 10 out of 100 similar species with graphical representation.

Unknown E

```
201>ccgtgtgagacatcaacaatatggaggatatggagccactggtgttcagccaccagcaccaactccaaacccatatggaggatatggagcaagccagcca>300
201>ccgtgtgagacgtcaacaatatggaggatatggagccactggtgttcagccaccagcaccaactccaaacccatatggaggatatggagcaagccagcca>300
```

```
Translation 348 a.a. MW=34761.95999999993

MetThrLeuThrThrAlaThrSerGlyAlaIleValPheSerGlyAlaThrLeuLeuVal
SerLeuPheAlaAlaAlaSerLeuTyrSerGlnValSerAsnIleTrpAsnGluLeuAsp
AlaGluIleAlaAsnPheArgSerLeuThrGluAspMetTrpValAspMetValLysLeu
GlyAlaGlyThrAlaSerAsnArgValArgArgGlnGlnTyrGlyGlyTyrGlyAlaThr
GlyValGlnProProAlaProThrProAsnProTyrGlyGlyTyrGlyAlaSerGlnPro
AlaProProGluLysPheProAspGlyIleProAsnGlyGlyAsnGlnProLysPhePro
GlyGlyGlyPheProAspGlyProPheProAsnGlyGlyGlyProArgGlyGlyAsnGln
CysGlnCysThrValGluAsnSerCysProProGlyProAlaGlyProGluGlyGluGlu
GlyProAspGlyHisAspGlyGlnAspGlyValProGlyPheAspGlyLysAspAlaGlu
AspValGlnAsnThrProProThrGlyCysPheThrCysProGlnGlyProLeuGlyPro
GlnGlyProAsnGlyAlaProGlyLeuArgGlyMetArgGlyAlaArgGlyGlnProGly
ArgProGlyArgAspGlyAsnProGlyMetProGlyAspCysGlyProProGlyAlaPro
GlySerAspGlyLysProGlySerProGlyGlyLysGlyAspAspGlyGluArgProLeu
GlyArgProGlyProArgGlyProProGlyGluAlaGlyProGluGlyProGlnGlyPro
ThrGlyArgAspAlaTyrProGlyGlnSerGlyProGlnGlyGluProGlyLeuGlnGly
TyrGlyGlyAlaAlaGlyGluAspGlyProGluGlyProProGlyAlaProGlyLeuPro
GlyLysAspAlaGluTyrCysLysCysProGlyArgGluGlyAspAlaGlyArgSerAla
ArgArgHisArgLysPheGlnLeuEnd
```

```
Translation 348 a.a. MW=34742.90999999992

MetThrLeuThrThrAlaThrSerGlyAlaIleValPheSerGlyAlaThrLeuLeuVal
SerLeuPheAlaAlaAlaSerLeuTyrSerGlnValSerAsnIleTrpAsnGluLeuAsp
AlaGluIleAlaAsnPheArgSerLeuThrGluAspMetTrpValAspMetValLysLeu
GlyAlaGlyThrAlaSerAsnArgValArgHisGlnGlnTyrGlyGlyTyrGlyAlaThr
GlyValGlnProProAlaProThrProAsnProTyrGlyGlyTyrGlyAlaSerGlnPro
AlaProProGluLysPheProAspGlyIleProAsnGlyGlyAsnGlnProLysPhePro
GlyGlyGlyPheProAspGlyProPheProAsnGlyGlyGlyProArgGlyGlyAsnGln
CysGlnCysThrValGluAsnSerCysProProGlyProAlaGlyProGluGlyGluGlu
GlyProAspGlyHisAspGlyGlnAspGlyValProGlyPheAspGlyLysAspAlaGlu
AspValGlnAsnThrProProThrGlyCysPheThrCysProGlnGlyProLeuGlyPro
GlnGlyProAsnGlyAlaProGlyLeuArgGlyMetArgGlyAlaArgGlyGlnProGly
ArgProGlyArgAspGlyAsnProGlyMetProGlyAspCysGlyProProGlyAlaPro
GlySerAspGlyLysProGlySerProGlyGlyLysGlyAspAspGlyGluArgProLeu
GlyArgProGlyProArgGlyProProGlyGluAlaGlyProGluGlyProGlnGlyPro
ThrGlyArgAspAlaTyrProGlyGlnSerGlyProGlnGlyGluProGlyLeuGlnGly
TyrGlyGlyAlaAlaGlyGluAspGlyProGluGlyProProGlyAlaProGlyLeuPro
GlyLysAspAlaGluTyrCysLysCysProGlyArgGluGlyAspAlaGlyArgSerAla
ArgArgHisArgLysPheGlnLeuEnd
```

Hits for all PROSITE (release 2019_10) motifs on sequence USERSEQ1 :

no hit!

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☑	Loa loa nematode cuticle collagen domain-containing protein partial mRNA	62.0	62.0	22%	7e-07	36.14%	XM_003143129.1
☑	Trichinella spiralis cuticle collagen rol-6 (Tsp_02838) mRNA, complete cds	47.0	47.0	26%	0.060	28.71%	XM_003379120.1
☑	Trichinella spiralis cuticle collagen 39 (Tsp_03443) mRNA, complete cds	43.1	43.1	18%	0.99	25.37%	XM_003379743.1
☑	Loa loa hypothetical protein partial mRNA	40.0	40.0	16%	9.8	31.58%	XM_003135557.1

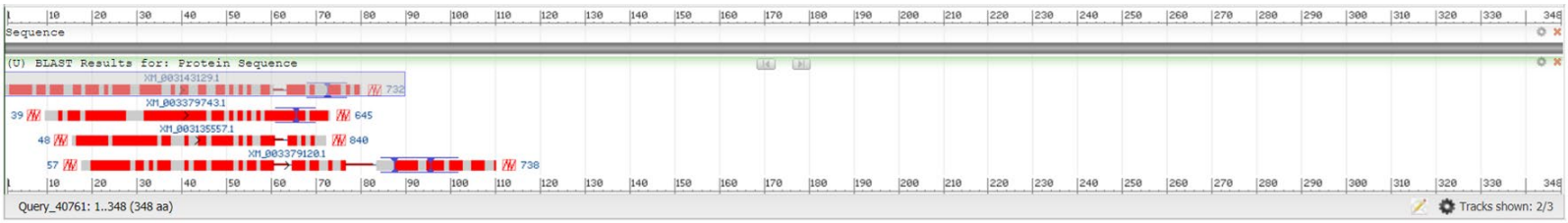

Figure 6e: Alignment with ApE of Unknown E wild type to mutant. ApE translated mutant and wildtype sequences with noted differences. Expasy results shows no similar protein motifs. BLAST reveals 4 similar species with graphical representation.

Unknown F

2001>tattgtaaaaatgatggctgattttggaggacaccttggactttggtcaggagtttctgttatgacttgttgtgagtttgtgtgtctagcctttgagctg>2100
2001>tattgtaaaaatgatggctgattttcgaggacaccttggactttggtcaggagtttctgttatgacttgttgtgagtttgtgtgtctagcctttgagctg>2100

Translation 724 a.a. MW=82384.29000000004

MetAsnArgAsnProArgMetSerLysPheGlnProAsnProArgSerArgSerArgPhe
GlnAspGluThrAspLeuArgSerLeuArgSerPheLysThrAspPheSerAsnTyrLeu
AlaSerAspThrAsnPheLeuAsnValAlaGluIleMetThrSerTyrAlaTyrGlyGlu
SerAsnAsnAlaHisGluLysGluIleGlnCysAspLeuLeuThrGluAsnGlyGlyIle
GluIleAspProThrArgLeuSerTyrArgGluArgIleArgTrpHisLeuGlnGlnPhe
CysTyrLysThrSerSerHisGlyIleProMetLeuGlyGlnAlaProAsnSerLeuTyr
ArgAlaAlaTrpValPheLeuLeuLeuIleCysAlaIleGlnPheIleAsnGlnAlaVal
AlaValIleGlnLysTyrGlnLysMetAspLysIleThrAspIleGlnLeuLysPheAsp
ThrAlaProPheProAlaIleThrLeuCysAsnLeuAsnProTyrLysAspSerValIle
ArgSerHisAspSerIleSerLysIleLeuGlyValPheLysSerValMetLysLysAla
GlyAspSerSerSerGluAlaLeuGluGluGluGluGluThrGluTyrAspMetAsnGly
IleThrIleGlnAlaLysArgLysLysArgGlyAlaGlyGluLysGlyThrPheGluPro
AlaAsnSerAlaCysGluCysAspGluGluAspGlySerAsnGluCysGluGluArgSer
ThrGluLysProSerGlyAspAsnAspMetCysIleCysAlaPheAspArgGlnThrAsn
AspAlaTrpProCysHisArgLysGluGlnTrpThrAsnThrThrCysGlnThrCysAsp
GluHisTyrLeuCysSerLysLysAlaLysLysGlyThrLysArgSerGluLeuLysLys
GluProCysIleCysGluSerLysGlyLeuPheCysIleLysHisGluHisAlaAlaMet
ValLeuAsnLeuTrpGluTyrPheGlyAspSerGluAspPheSerGluIleSerThrGlu
GluArgGluAlaLeuGlyPheGlyAsnMetThrAspGluValAlaIleValThrLysAla
LysGluAsnIleIlePheAlaMetSerAlaLeuSerGluGluGlnArgIleLeuMetSer
GlnAlaLysHisAsnLeuIleHisLysCysSerPheAsnGlyLysProCysAspIleAsp
GlnAspPheGluLeuValAlaAspProThrPheGlyAsnCysPheValPheAsnHisAsp
ArgGluIlePheLysSerSerValArgAlaGlyProGlnTyrGlyLeuArgValMetLeu
PheValAsnAlaSerAspTyrLeuProThrSerGluAlaValGlyIleArgLeuThrIle
HisAspLysAspAspPheProPheProAspThrPheGlyTyrSerAlaProThrGlyTyr
IleSerSerPheGlyMetArgMetLysLysMetSerArgLeuProAlaProTyrGlyAsp
CysValGluAspGlyAlaThrSerAsnTyrIleTyrLysGlyTyrAlaTyrSerThrGlu
GlyCysTyrArgThrCysPheGlnGluLeuIleIleAspArgCysGlyCysSerAspPro
ArgPheProSerIleGlyGlyValGlnProCysGlnValPheAsnLysAsnHisArgGlu
CysLeuGluLysHisThrHisGlnIleGlyGluIleHisGlySerPheLysCysArgCys
GlnGlnProCysAsnGlnThrIleTyrThrThrSerTyrSerGluAlaIleTrpProSer
GlnAlaLeuAsnIleSerLeuGlyGlnCysGluLysGluAlaGluGluCysAsnGluGlu
TyrLysGluAsnAlaAlaMetLeuGluValPheTyrGluAlaLeuAsnPheGluValLeu
SerGluSerGluAlaTyrGlyIleValLysMetMetAlaAspPheGlyGlyHisLeuGly
LeuTrpSerGlyValSerValMetThrCysCysGluPheValCysLeuAlaPheGluLeu
IleTyrMetAlaIleAlaHisHisIleAsnGlnGlnArgIleArgArgArgGluAsnAla
AlaAsnGluTyrEnd

USERSEQ1 (724 aa)

Translation 724 a.a. MW=82483.43000000004

MetAsnArgAsnProArgMetSerLysPheGlnProAsnProArgSerArgSerArgPhe
GlnAspGluThrAspLeuArgSerLeuArgSerPheLysThrAspPheSerAsnTyrLeu
AlaSerAspThrAsnPheLeuAsnValAlaGluIleMetThrSerTyrAlaTyrGlyGlu
SerAsnAsnAlaHisGluLysGluIleGlnCysAspLeuLeuThrGluAsnGlyGlyIle
GluIleAspProThrArgLeuSerTyrArgGluArgIleArgTrpHisLeuGlnGlnPhe
CysTyrLysThrSerSerHisGlyIleProMetLeuGlyGlnAlaProAsnSerLeuTyr
ArgAlaAlaTrpValPheLeuLeuLeuIleCysAlaIleGlnPheIleAsnGlnAlaVal
AlaValIleGlnLysTyrGlnLysMetAspLysIleThrAspIleGlnLeuLysPheAsp
ThrAlaProPheProAlaIleThrLeuCysAsnLeuAsnProTyrLysAspSerValIle
ArgSerHisAspSerIleSerLysIleLeuGlyValPheLysSerValMetLysLysAla
GlyAspSerSerSerGluAlaLeuGluGluGluGluGluThrGluTyrAspMetAsnGly
IleThrIleGlnAlaLysArgLysLysArgGlyAlaGlyGluLysGlyThrPheGluPro
AlaAsnSerAlaCysGluCysAspGluGluAspGlySerAsnGluCysGluGluArgSer
ThrGluLysProSerGlyAspAsnAspMetCysIleCysAlaPheAspArgGlnThrAsn
AspAlaTrpProCysHisArgLysGluGlnTrpThrAsnThrThrCysGlnThrCysAsp
GluHisTyrLeuCysSerLysLysAlaLysLysGlyThrLysArgSerGluLeuLysLys
GluProCysIleCysGluSerLysGlyLeuPheCysIleLysHisGluHisAlaAlaMet
ValLeuAsnLeuTrpGluTyrPheGlyAspSerGluAspPheSerGluIleSerThrGlu
GluArgGluAlaLeuGlyPheGlyAsnMetThrAspGluValAlaIleValThrLysAla
LysGluAsnIleIlePheAlaMetSerAlaLeuSerGluGluGlnArgIleLeuMetSer
GlnAlaLysHisAsnLeuIleHisLysCysSerPheAsnGlyLysProCysAspIleAsp
GlnAspPheGluLeuValAlaAspProThrPheGlyAsnCysPheValPheAsnHisAsp
ArgGluIlePheLysSerSerValArgAlaGlyProGlnTyrGlyLeuArgValMetLeu
PheValAsnAlaSerAspTyrLeuProThrSerGluAlaValGlyIleArgLeuThrIle
HisAspLysAspAspPheProPheProAspThrPheGlyTyrSerAlaProThrGlyTyr
IleSerSerPheGlyMetArgMetLysLysMetSerArgLeuProAlaProTyrGlyAsp
CysValGluAspGlyAlaThrSerAsnTyrIleTyrLysGlyTyrAlaTyrSerThrGlu
GlyCysTyrArgThrCysPheGlnGluLeuIleIleAspArgCysGlyCysSerAspPro
ArgPheProSerIleGlyGlyValGlnProCysGlnValPheAsnLysAsnHisArgGlu
CysLeuGluLysHisThrHisGlnIleGlyGluIleHisGlySerPheLysCysArgCys
GlnGlnProCysAsnGlnThrIleTyrThrThrSerTyrSerGluAlaIleTrpProSer
GlnAlaLeuAsnIleSerLeuGlyGlnCysGluLysGluAlaGluGluCysAsnGluGlu
TyrLysGluAsnAlaAlaMetLeuGluValPheTyrGluAlaLeuAsnPheGluValLeu
SerGluSerGluAlaTyrGlyIleValLysMetMetAlaAspPheArgGlyHisLeuGly
LeuTrpSerGlyValSerValMetThrCysCysGluPheValCysLeuAlaPheGluLeu
IleTyrMetAlaIleAlaHisHisIleAsnGlnGlnArgIleArgArgArgGluAsnAla
AlaAsnGluTyrEnd

PS01206 **ASC** *Amiloride-sensitive sodium channels signature* :

537 - 557: [confidence level: (0)] YstEgCyrtCfQelIIdrCgC

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
PREDICTED: Saccoglossus kowalevskii amiloride-sensitive sodium channel subunit alpha-like (LOC102802084), mRNA	179	223	56%	2e-44	30.56%	XM_006813281.1
PREDICTED: Orbicella faveolata degenerin deg-1-like (LOC110059196), mRNA	174	220	60%	2e-41	32.09%	XM_020765895.1
PREDICTED: Corapipo altera sodium channel epithelial 1 beta subunit (SCNN1B), mRNA	169	169	87%	7e-40	23.07%	XM_027668002.1
PREDICTED: Acanthaster planci degenerin mec-10-like (LOC110987987), mRNA	168	228	60%	2e-39	31.56%	XM_022251205.1
PREDICTED: Branchiostoma belcheri degenerin deg-1-like (LOC109466016), transcript variant X1, mRNA	164	220	58%	6e-39	29.19%	XM_019763538.1
PREDICTED: Saccoglossus kowalevskii degenerin mec-10-like (LOC102803894), partial mRNA	156	156	38%	9e-39	31.36%	XM_006818585.1
PREDICTED: Strongylocentrotus purpuratus amiloride-sensitive sodium channel subunit beta-like (LOC105441174), mRNA	164	216	57%	6e-38	30.64%	XM_011672057.2
PREDICTED: Latimeria chalumnae acid sensing (proton gated) ion channel family member 4 (ASIC4), mRNA	159	159	43%	8e-38	29.41%	XM_005995780.1
PREDICTED: Lingula anatina degenerin deg-1-like (LOC106181583), mRNA	161	161	45%	1e-37	28.02%	XM_013566016.1
PREDICTED: Branchiostoma belcheri degenerin deg-1-like (LOC109466016), transcript variant X2, mRNA	159	215	58%	3e-37	28.90%	XM_019763539.1

Figure 6f: Alignment with ApE of Unknown F wild type to mutant. ApE translated mutant and wildtype sequences with noted differences. Expasy results shows amiloride-sensitive sodium channel motif signature. BLAST results of first 10 out of 100 similar species with graphical representation.

Unknown G

```
301>tgctataaaacatctagtcatggtattccaatgcttgggcaagctccaaacagtttatacagagctgcatgggtattcttgcttctaatctgtgctatac>400
301>tgctataaaacatctagtcatggtattccaatgcttgggcaagctccaaacagtttatacagagctgcatgggtattcttgcttctaatctgtgctatac>400
```

```
Translation 724 a.a. MW=82384.29000000004

MetAsnArgAsnProArgMetSerLysPheGlnProAsnProArgSerArgSerArgPhe
GlnAspGluThrAspLeuArgSerLeuArgSerPheLysThrAspPheSerAsnTyrLeu
AlaSerAspThrAsnPheLeuAsnValAlaGluIleMetThrSerTyrAlaTyrGlyGlu
SerAsnAsnAlaHisGluLysGluIleGlnCysAspLeuLeuThrGluAsnGlyGlyIle
GluIleAspProThrArgLeuSerTyrArgGluArgIleArgTrpHisLeuGlnGlnPhe
CysTyrLysThrSerSerHisGlyIleProMetLeuGlyGlnAlaProAsnSerLeuTyr
ArgAlaAlaTrpValPheLeuLeuLeuIleCysAlaIleGlnPheIleAsnGlnAlaVal
AlaValIleGlnLysTyrGlnLysMetAspLysIleThrAspIleGlnLeuLysPheAsp
ThrAlaProPheProAlaIleThrLeuCysAsnLeuAsnProTyrLysAspSerValIle
ArgSerHisAspSerIleSerLysIleLeuGlyValPheLysSerValMetLysLysAla
GlyAspSerSerSerGluAlaLeuGluGluGluGluGluThrGluTyrAspMetAsnGly
IleThrIleGlnAlaLysArgLysLysArgGlyAlaGlyGluLysGlyThrPheGluPro
AlaAsnSerAlaCysGluCysAspGluGluAspGlySerAsnGluCysGluGluArgSer
ThrGluLysProSerGlyAspAsnAspMetCysIleCysAlaPheAspArgGlnThrAsn
AspAlaTrpProCysHisArgLysGluGlnTrpThrAsnThrThrCysGlnThrCysAsp
GluHisTyrLeuCysSerLysLysAlaLysLysGlyThrLysArgSerGluLeuLysLys
GluProCysIleCysGluSerLysGlyLeuPheCysIleLysHisGluHisAlaAlaMet
ValLeuAsnLeuTrpGluTyrPheGlyAspSerGluAspPheSerGluIleSerThrGlu
GluArgGluAlaLeuGlyPheGlyAsnMetThrAspGluValAlaIleValThrLysAla
LysGluAsnIleIlePheAlaMetSerAlaLeuSerGluGluGlnArgIleLeuMetSer
GlnAlaLysHisAsnLeuIleHisLysCysSerPheAsnGlyLysProCysAspIleAsp
GlnAspPheGluLeuValAlaAspProThrPheGlyAsnCysPheValPheAsnHisAsp
ArgGluIlePheLysSerSerValArgAlaGlyProGlnTyrGlyLeuArgValMetLeu
PheValAsnAlaSerAspTyrLeuProThrSerGluAlaValGlyIleArgLeuThrIle
HisAspLysAspAspPheProPheProAspThrPheGlyTyrSerAlaProThrGlyTyr
IleSerSerPheGlyMetArgMetLysLysMetSerArgLeuProAlaProTyrGlyAsp
CysValGluAspGlyAlaThrSerAsnTyrIleTyrLysGlyTyrAlaTyrSerThrGlu
GlyCysTyrArgThrCysPheGlnGluLeuIleIleAspArgCysGlyCysSerAspPro
ArgPheProSerIleGlyGlyValGlnProCysGlnValPheAsnLysAsnHisArgGlu
CysLeuGluLysHisThrHisGlnIleGlyGluIleHisGlySerPheLysCysArgCys
GlnGlnProCysAsnGlnThrIleTyrThrThrSerTyrSerGluAlaIleTrpProSer
GlnAlaLeuAsnIleSerLeuGlyGlnCysGluLysGluAlaGluGluCysAsnGluGlu
TyrLysGluAsnAlaAlaMetLeuGluValPheTyrGluAlaLeuAsnPheGluValLeu
SerGluSerGluAlaTyrGlyIleValLysMetMetAlaAspPheGlyGlyHisLeuGly
LeuTrpSerGlyValSerValMetThrCysCysGluPheValCysLeuAlaPheGluLeu
IleTyrMetAlaIleAlaHisHisIleAsnGlnGlnArgIleArgArgArgGluAsnAla
AlaAsnGluTyrEnd
```

```
Translation 724 a.a. MW=82444.39000000003

MetAsnArgAsnProArgMetSerLysPheGlnProAsnProArgSerArgSerArgPhe
GlnAspGluThrAspLeuArgSerLeuArgSerPheLysThrAspPheSerAsnTyrLeu
AlaSerAspThrAsnPheLeuAsnValAlaGluIleMetThrSerTyrAlaTyrGlyGlu
SerAsnAsnAlaHisGluLysGluIleGlnCysAspLeuLeuThrGluAsnGlyGlyIle
GluIleAspProThrArgLeuSerTyrArgGluArgIleArgTrpHisLeuGlnGlnPhe
CysTyrLysThrPheSerHisGlyIleProMetLeuGlyGlnAlaProAsnSerLeuTyr
ArgAlaAlaTrpValPheLeuLeuLeuIleCysAlaIleGlnPheIleAsnGlnAlaVal
AlaValIleGlnLysTyrGlnLysMetAspLysIleThrAspIleGlnLeuLysPheAsp
ThrAlaProPheProAlaIleThrLeuCysAsnLeuAsnProTyrLysAspSerValIle
ArgSerHisAspSerIleSerLysIleLeuGlyValPheLysSerValMetLysLysAla
GlyAspSerSerSerGluAlaLeuGluGluGluGluGluThrGluTyrAspMetAsnGly
IleThrIleGlnAlaLysArgLysLysArgGlyAlaGlyGluLysGlyThrPheGluPro
AlaAsnSerAlaCysGluCysAspGluGluAspGlySerAsnGluCysGluGluArgSer
ThrGluLysProSerGlyAspAsnAspMetCysIleCysAlaPheAspArgGlnThrAsn
AspAlaTrpProCysHisArgLysGluGlnTrpThrAsnThrThrCysGlnThrCysAsp
GluHisTyrLeuCysSerLysLysAlaLysLysGlyThrLysArgSerGluLeuLysLys
GluProCysIleCysGluSerLysGlyLeuPheCysIleLysHisGluHisAlaAlaMet
ValLeuAsnLeuTrpGluTyrPheGlyAspSerGluAspPheSerGluIleSerThrGlu
GluArgGluAlaLeuGlyPheGlyAsnMetThrAspGluValAlaIleValThrLysAla
LysGluAsnIleIlePheAlaMetSerAlaLeuSerGluGluGlnArgIleLeuMetSer
GlnAlaLysHisAsnLeuIleHisLysCysSerPheAsnGlyLysProCysAspIleAsp
GlnAspPheGluLeuValAlaAspProThrPheGlyAsnCysPheValPheAsnHisAsp
ArgGluIlePheLysSerSerValArgAlaGlyProGlnTyrGlyLeuArgValMetLeu
PheValAsnAlaSerAspTyrLeuProThrSerGluAlaValGlyIleArgLeuThrIle
HisAspLysAspAspPheProPheProAspThrPheGlyTyrSerAlaProThrGlyTyr
IleSerSerPheGlyMetArgMetLysLysMetSerArgLeuProAlaProTyrGlyAsp
CysValGluAspGlyAlaThrSerAsnTyrIleTyrLysGlyTyrAlaTyrSerThrGlu
GlyCysTyrArgThrCysPheGlnGluLeuIleIleAspArgCysGlyCysSerAspPro
ArgPheProSerIleGlyGlyValGlnProCysGlnValPheAsnLysAsnHisArgGlu
CysLeuGluLysHisThrHisGlnIleGlyGluIleHisGlySerPheLysCysArgCys
GlnGlnProCysAsnGlnThrIleTyrThrThrSerTyrSerGluAlaIleTrpProSer
GlnAlaLeuAsnIleSerLeuGlyGlnCysGluLysGluAlaGluGluCysAsnGluGlu
TyrLysGluAsnAlaAlaMetLeuGluValPheTyrGluAlaLeuAsnPheGluValLeu
SerGluSerGluAlaTyrGlyIleValLysMetMetAlaAspPheGlyGlyHisLeuGly
LeuTrpSerGlyValSerValMetThrCysCysGluPheValCysLeuAlaPheGluLeu
IleTyrMetAlaIleAlaHisHisIleAsnGlnGlnArgIleArgArgArgGluAsnAla
AlaAsnGluTyrEnd
```

USERSEQ1 (724 aa)

PS01206 **ASC** *Amiloride-sensitive sodium channels signature* :

537 - 557: [confidence level: (0)] YstEgCyrtCfQelIIdrCgC

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Caenorhabditis elegans Degenerin mec-10 (mec-10), mRNA	1425	1425	100%	0.0	94.34%	NM_077037.5
Caenorhabditis elegans Degenerin mec-4 (mec-4), mRNA	526	706	85%	1e-173	53.81%	NM_078311.6
Caenorhabditis elegans Degenerin del-1 (del-1), partial mRNA	432	432	84%	1e-138	37.94%	NM_001368546.1
Caenorhabditis elegans Degenerin del-1 (del-1), partial mRNA	432	432	84%	2e-138	37.94%	NM_001368544.1
Caenorhabditis elegans Degenerin del-1 (del-1), partial mRNA	432	432	84%	2e-138	37.94%	NM_001368545.1
Brugia pahangi genome assembly, scaffold: BPAG_scaffold0000033	110	628	77%	8e-59	65.28%	LK964192.1
Thelazia callipaeda genome assembly, scaffold: TCLT_scaffold0000184	110	642	77%	1e-58	75.41%	LK979744.1
Syphacia muris strain Valencia genome assembly, scaffold: SMUV_scaffold0000215	104	630	65%	8e-54	70.97%	LK996259.1
Helobdella robusta hypothetical protein partial mRNA	184	184	47%	4e-47	29.00%	XM_009027326.1
Heligmosomoides polygyrus genome assembly, scaffold: HPBE_scaffold0003058	115	894	80%	2e-44	53.39%	LL191440.1

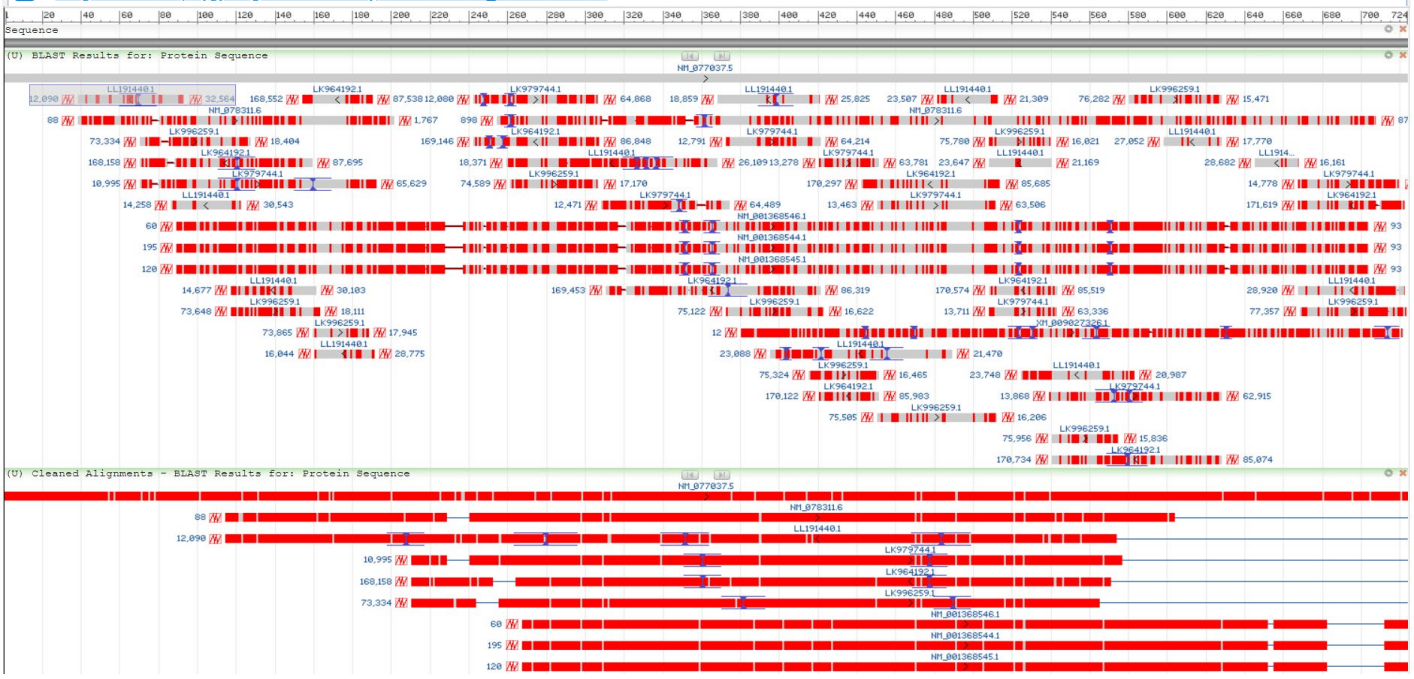

Figure 6g: Alignment with ApE of Unknown G wild type to mutant. ApE translated mutant and wildtype sequences with noted differences. Expasy results shows amiloride-sensitive sodium channel motif signature. BLAST results of first 10 out of 100 similar species with graphical representation.

Unknown H

```
H.wt.ape from 1 to 1050
Alignment to
H.mut.ape--  Matches:1050; Mismatches:0; Gaps:0; Unattempted:0
Translation 277 a.a. MW=32698.799999999945
???(72 extra codons after stop)
ValSerValHisLeuLeuGlnLeuArgHisProProThrAspHisHisMetAspGlnThr
MetThrIleThrValGluSerArgArgLeuLeuGluAspArgIleHisLeuAsnSerLeu
ProValHisGlnLeuArgThrAlaValArgIleThrGluHisHisCysThrIleHisSer
HisLeuGlnLysLeuAspIleGlnGlnIleLeuValTyrArgThrIleHisLeuSerGln
ThrIleLeuGlyAsnGlnArgIleLeuLysThrThrAlaArgHisIleSerGlnGlnAsp
HisGlnThrLeuHisHisLeuValAsnLeuLeuAlaAspThrIleGlnMetGlyArgLeu
HisArgHisArgLeuGluPheThrIleLeuValValArgThrIleMetValGlnGlyIle
GlnArgIleLysTyrGlnGlnArgHisGlnPheGlnAspAsnLysGluPheHisArgLeu
LysLeuGluIleHisGlnIleGlnLeuIleProValAsnHisArgAspGlnCysLeuGln
HisGlnMetAspIleLeuArgGlnProLeuArgIleIleHisLeuLeuAsnIleProGln
GluLysGlyValHisThrLeuAspLeuAspLeuArgAspLeuValGlnGluLeuAspGln
GluAlaIleHisValIleAsnValLeuLeuSerArgIleIleAlaHisGlnAspLeuGln
AspLeuValValAspGlnGlyLeuGlnAspSerGlnAspLysMetGlyHisValGlyPhe
GluValSerMetValValIleLeuValPheAsnHisLeuLeuMetIleGlnEndLeuAla
ValTyrSerAlaGlnLeuAspHisLeuGluAsnAlaAspGlnThrValHisGlnGluCys
GlnGluLysMetValLeuMetAlaSerLysGluEndThrAspLysMetValAsnGlnGlu
HisGlnGlyHisLeuAspIleMetGluEndMetValLeuArgGluHisLeuGluSerLeu
AspTyrLeuGluGluThrAlaAsnHisAla
```

Hits for all PROSITE (release 2019_10) motifs on sequence USERSEQ1 :

no hit!

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Caenorhabditis elegans Cuticle collagen bli-1 (bli-1), mRNA	535	535	100%	1e-180	100.00%	NM_063910.5

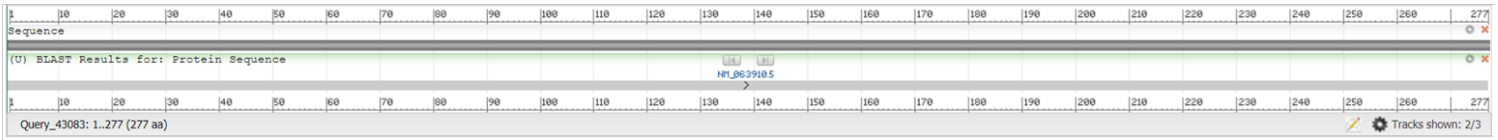

Figure 6h: Alignment with ApE of Unknown H wild type to mutant. ApE translated wildtype sequence with STOP codons highlighted. Expasy results show no similar sequence motifs. BLAST results of one similar species and graphical representation shown.

Discussion

This experiment aimed to elucidate possible genes and their functions in *C. elegans* by exploiting unique mechanosensory characteristics. This was done by introducing random unbiased mutations and choosing mechanosensory typed phenotypes for testing. Then, worms were assayed using a gentle touch response assay to determine a touch response index. Based on the results, there were four genes characterized by this screening. One gene is the *mec* gene that had 4 phenotypes, the *unc* gene, *dpy* gene, and *rol* gene that had one phenotype each. One chosen phenotype was a wildtype worm. Unknown B was the most prominent phenotype, with very little observed

responses and an average touch response index of 8.67, meaning a responsiveness of less than 10%. This phenotype was linked to the *unc* gene which was illustrated by the UNC-119 homologs in other species through tBLAST. Furthermore, the touch response index of Unknown B was decreased significantly in every post-hoc comparison except when compared against the *mec-10* positive control. However, this worm mutant showed more homology to UNC-119 than to MEC or related proteins. This *unc* gene results in a truncated UNC protein which causes major mechanosensory deficiencies. Another clear phenotype was represented by Unknown D, with a touch response index of 82, but clearly observed morphological change. These worms were observed to be shorter than wildtype worms with a stubbier, fatter appearance. There was no statistical significance in post-hoc comparisons to wild-type, and some mutants had significantly decreased touch response indices compared to Unknown D. The gene associated with this phenotype is *dpy* and is most homologous to a SCARA5 protein in other species. The mutant form of this gene results in a nonsense mutation and truncated DPY protein which causes severe morphological changes. Another phenotypically clear mutant was that of Unknown E, which was found to be a *rol* gene mutant. This mutant had an average touch response index of 71.22, with no statistical significance against wildtype and unknown D. This mutation is a missense that causes an amino acid change from Arg to His, and gives a phenotype associated with rolling movements. As explained, this mutant had almost wild type mechanosensory responses, but its movements were characterized by a clear rolling pattern upon stimulation in stark contrast to the sinusoidal movements typical of *C. elegans*. When this protein sequence was run through tBLAST, 4 homologs emerged, one called the Loa Loa nematode cuticle collagen, and another was ROL-6 in *Trichinella spiralis*. This and the observations seen gave evidence for the implication of the *rol* gene in this particular mechanosensory mutant. Unknown H had a phenotype similar to that of

wild type, with a touch response index of 78.5, which was not statistically significant in comparison to wild type, unknown D, and unknown E. Furthermore, when sequenced and aligned using ApE, there were no mismatches or mutations in the mutant sequence and when run through tBLAST the results illustrated a known *C. elegans* cuticle collagen protein. This supports the conclusion that Unknown H was a wild type worm. The remaining worms, Unknown A, C, F and G were versions of a *mec* mutant. Unknown F, with a touch response index of 36.11, was significantly lower compared to every group except for the known *mec-10* control mutant. This mutant resulted in a Gly to Arg missense mutation when sequences were analyzed and translated. When the amino acid sequence was analyzed through tBLAST there were numerous homologs to MEC-10 in other species. This suggests that the gene responsible for the phenotype in Unknown F was that of *mec-10*. One of the key results in this mutant that was not relevant in the analysis of the aforementioned mutants was that of the Expasy motif search, that yielded a sequence similarity to the amiloride-sensitive sodium channels signature motifs. This motif was also implicated in the sequences of Unknown C and Unknown G's protein. However, one of the most contradictory evidences is that of the statistical analysis which shows statistical significance in the difference between *mec-10* control mutants v. Unknown C and *mec-10* v. Unknown G. Unknown G is also not statistically different when compared to wild-type (although Unknown C is). This supports a hypothesis of a *mec* mutant that is region specific. After sequencing and aligning the mutant to wild-type sequences, it is clear that Unknown C and Unknown G are mechanosensory mutants due to the missense substitutions that result from the mutations (Unknown C results in an Ala to Leu; Unknown G results in a Ser to Phe). Looking further into the data, it was noticed that the Unknown G had a tBLAST result that showed homology to *mec-4* and *mec-10*, which suggests a varied *mec* phenotype. This likely is consistent with a *mec* mutation that affects the head or tail region,

although some of the data shows no difference in the average touch response index for head stimulations versus tail stimulations. Nonetheless, it can be confirmed that Unknown G is a mutant of a *mec* gene, and further experimentation will be required to elucidate the specific type. For Unknown C, it is more difficult to elucidate the specific *mec* mutant although it is clear that Unknown C is a *mec* mutant. It shares the same motif as two known *mec* mutants, as well as having a statistically significant difference in average touch response index compared to wild type (mean = 70.33, $p < 0.05$). However, tBLAST results are inconclusive as they point to a homolog of an SCNN1D protein, which is a sodium channel in other species. It is likely that this is a *mec* mutant that effects the tail region, as some average touch indices of just the tail region showed just slightly lower responsiveness, although this may not prove to be statistically significant. The last mutant, Unknown A, has a touch response index average of 70.16, with no statistical difference against wild type and other mutants except for Unknown F and *mec-10* control. The mutant substitution results in a nonsense mutation based on ApE alignment and translation. Expasy showed a similar motif to tubulin subunits, and tBLAST analysis revealed homologs and orthologs to tubulin in other species. This is connected to a known *mec* protein known as MEC-7 which is orthologous to TUBB6 and TUBB8. This is evidence that Unknown A is likely a mutant of *mec-7*.

Despite the results of the study, there are a fair share of limitations. One of the biggest surprises came in the form of the statistical analysis which showed insignificant differences in a few relevant comparisons. For instance, despite clear evidence that Unknown G contained a mutant *mec-10* or *mec-4*, the statistics showed an insignificant difference between this group and wild type. Furthermore, Unknown G was statistically significant compared to the *mec-10* positive control, which should be insignificant if Unknown G's identity as a *mec-10* or *mec-4* is to be confirmed with statistics. This may represent an error in the methodology versus error in the

conclusions, due to the other evidences that support the conclusions. A possible argument could be that of a desensitization consequence of multiple stimulations back to back with not enough time in between. However, most of the statistics suggest increased mechanosensory response which would not align with a desensitization argument. It is also possible that false positives were recorded, meaning mechanosensory responses that were not considered reversals were recorded as "1" rather than "0". This would explain some of the statistical inconsistencies and align with the expected results of a *mec* mutant. However, it is also the case that the gentle response assay itself was performed incorrectly, since it is a very sensitive assay. The gentle response assay could portray a high mechanosensory response if performers unintentionally performed a harsh touch. Even if a harsh touch occurred 20% of the time, that could reduce the touch response index enough to be statistically significant (i.e. Mean = 70.66 with 20% decrease would be 56.528). For the most part, the methodological errors are simple but can have a huge difference on the statistics. Nonetheless, the preponderance of the evidence supports the conclusions made above and the statistics, when aligned with the conclusions, only solidify the conclusion more.

This study is a genetic screen of mechanosensory deficient *C. elegans* mutant phenotypes that identified four gene classes: *mec, rol, dpy,* and *unc.* This gives great insight into mechanosensation in *C. elegans* and provides a starting point for understanding mechanotransduction and mechanosensation from a neurobiological perspective. Another important discovery is a potential human application when considering the orthologous *mec-7* gene to the TUBB6 and TUBB8 proteins in humans. These screens rely heavily on careful conduction of the procedure and even a few false positives can have dramatic effects on the statistical analyses. However, the gene classes discovered through ApE alignment, Expasy search, and tBLAST analysis allowed sufficient evidence to locate and conclude the identities of the 8 phenotypes.

References

Bianchi, L. 2007. Mechanotransduction: Touch and feel at the molecular level as modeled in Caenorhabditis elegans. *Mol. Neurobiol.* 36:254–271. doi:10.1007/s12035-007-8009-5.

Brenner, S. 1974. THE GENETICS OF CAENORHABDITIS ELEGANS. *Genetics*. 77.

Chalfie, M., A.C. Hart, C.H. Rankin, and M.B. Goodman. 2013. Assaying mechanosensation. *WormBook*. doi:10.1895/wormbook.1.172.1.

Hart, A.C., and M.Y. Chao. 2009. From odors to behaviors in caenorhabditis elegans. *In* The Neurobiology of Olfaction. CRC Press. 1–33.

Shaw, M., M. Elmi, V. Pawar, and M.A. Srinivasan. 2016. Investigation of mechanosensation in C elegans using light field calcium imaging. *Biomed. Opt. Express*. 7:2877. doi:10.1364/boe.7.002877.

Shi, S., S.M. Mutchler, B.M. Blobner, O.B. Kashlan, and T.R. Kleyman. 2018. Pore-lining residues of mec-4 and mec-10 channel subunits tune the caenorhabditis elegans degenerin channel's response to shear stress. *J. Biol. Chem.* 293:10757–10766. doi:10.1074/jbc.RA118.002499.

Strange, K. 2006. An overview of C. elegans biology. *Methods Mol. Biol.* 351:1–11. doi:10.1385/1-59745-151-7:1.