I acknowledge that this paper is a product of my original work, and that I understand every concept discussed within.

Signed,

Daniel Zoleikhaeian

Characterization and Cloning of Strongylocentrotus Purpuratus Gene Sp20946 for Eventual In Situ Hybridization Experiments

Introduction

Strongylocentrotus purpuratus has been a popular model organism in developmental biology for the past 150 years. Due to its ease of propagation under laboratory conditions, rapid embryogenesis, and transparent embryos, and well-characterized genome S. purpuratus is a choice model organism for studying gene expression during development (Adonin et al., 2021).

In this study, S. purpuratus gene sequence Sp20946 was studied. Using NCBI's Basic Local Alignment Search Tool (BLAST), protein families (Pfam) database search, phylogenetic analysis, and multiple sequence alignment, the gene sequence was identified as S. purpuratus myoneurin, a protein widely present across both deuterostome and protostome species. Within S. purpuratus myoneurin, a chromo domain of the polycomb group (PcG) chromodomain protein polycomb (Pc) and a CBX7-C motif were identified. Chromo domain proteins are involved in packaging DNA into heterochromatin-like structures to achieve gene silencing (Jones et al., 2000). The CBX7-C domain associates with C-terminus of PcG protein RING1B to help form Polycomb repression complex 1 (PRC1) binds to chromatin with H3K27 and H3K9 trimethylation (H3K27Me3 and H3K9Me3) to ultimately induce transcriptional repression (Wang et al., 2010). Hence, it is likely that S. purpuratus myoneurin is responsible for gene

silencing.

The objective of subsequent experiments was to clone an Sp20946 target sequence into a bacterial plasmid for later study of Sp20946 expression patterns in S. purpuratus embryos via in situ hybridization. Based on gel electrophoresis of PCR products, amplification of the Sp20946 target sequence and that of control gene SpGataE were successful. Presence of LacZ-disrupted colonies on culture plates and subsequent sequencing of plasmid isolates confirmed that Sp20946 ligation into the plasmid vector and uptake of the recombinant plasmid were successful. For the studied SpGataE insert colony, sequencing of the inserted region revealed unsuccessful ligation of the SpGataE insert into the plasmid. This finding, alongside overall low transformation efficiencies and ligation efficiencies, encourage repetition of ligation and transformation steps and/or selection of a different SpGataE colony for future expansion. Nonetheless, presence of the intended Sp20946 insert in the studied Sp20946 recombinant-plasmid-transformed colony showed that transformation and ligation were successful for Sp20946. Based on this finding, the analyzed colony could be used for subsequent study of Sp20946 expression via in situ hybridization.

Methods

Unknown Gene Identification

To identify unknown sea urchin gene Sp20946, a Basic Alignment Search Tool (BLAST) search was conducted on the raw nucleotide sequence (blastn) and the translated nucleotide (blastx) sequence using NCBI nucleotide collection and non-redundant protein sequences databases. BLAST search was followed by a search in the protein families database (Pfam) of the amino acid sequence corresponding to the longest open reading frame of gene Sp20946 as given

by translation of Sp20946 nucleotide query sequence by ExPasy.

After obtaining a myoneurin as the putative gene identity from BLAST and Pfam searches, myoneurin mRNA sequences from a wide range of species were collected for use in phylogenetic analysis by protein sequence alignment. S. purpuratus myoneurin transcript variants were also included in a separate phylogenetic tree. Finally, multiple sequence alignment was performed to identify stretches of conserved amino acids between the translated Sp20946 gene and myoneurin proteins from protostome and other deuterostome species.

Polymerase Chain Reaction (PCR)

PCR was performed using gene-specific primers specific for unknown gene Sp20946 and control gene SpGataE to amplify known segments within each gene. Successful PCR amplification was confirmed through electrophoresis of PCR product on an agarose gel and calculation of band length using a standard curve. Gel electrophoresis and band length calculation was also performed on purified PCR products to ensure that the integrity of the amplified sequences was not compromised by the purification process. Concentration of purified SpGataE and Sp20946 PCR products was determined using a Nanodrop spectrophotometer.

Plasmid Recombination

To create recombinant plasmids using Sp20946 and SpGataE inserts, T4 DNA ligase was used to create recombinants between pGEM-T plasmid vectors and the aforementioned sea urchin gene inserts. Ligation tubes containing a 1:3 molar ratio of plasmid DNA to insert DNA, T4 DNA ligase, 2X ligase buffer, and water were incubated to allow for ligation.

Bacterial Transformation

Agar culture plates for bacterial transformation were prepared using luria broth, ampicillin, isopropylthiogalactoside (IPTG), and X-gal. Plates were subsequently incubated overnight. Next, mid-log phase DH5-α bacteria were treated with $CaCl_2$ solution to promote binding of plasmid DNA to bacterial cell surfaces.

The competent bacterial cells were then added to culture tubes containing no DNA, 1 ng pBSK control plasmid, 10 ng pBSK control plasmid, 100 ng pBSK control plasmid, 5 µL GataE ligation, or 5 µL Sp20946 ligation (Table 1). Tubes were incubated on ice followed by heat shock to facilitate uptake of foregin plasmids. Then, luria broth was added to each tube, and each tube was incubated in the shaker. After resuspending the bacteria on-ice, bacterial solution from each tube was added to the respective agar culture plates, which were subsequently incubated.

Transformation efficiency was assessed by counting the number of blue colonies found on each of the pBSK-treated agar plates divided by the nanograms of pBSK DNA added. Ligation efficiency was assessed by counting the number of white colonies on each of SpGataE and Sp20946-treated plates.

Insert Sequencing

To confirm the identity of the inserts in select white colonies, white colonies on SpGataE and Sp20946 culture plates were sampled and expanded. Then, plasmid DNA for each sample colony was isolated. Plasmid isolate DNA concentration from one colony each of SpGataE transformed bacteria and Sp20946 transformed bacteria was characterized via a Nanodrop spectrophotometer, and the same plasmid isolate DNA samples were sent for Sanger sequencing by Laragen. Sequencing only occurred for the location in the plasmid past the T7 primer, which

would capture the presence or absence of the intended insert.

Sequencing data retrieved from Laragen was cleaned by base editing through the Chromas program. Edited insert sequences exported and aligned against known Sp20946 and SpGataE sequences via BLAST. In the case of the SpGataE plasmid isolate insert sequence, further analysis of the sequenced data was accomplished through a BLAST nucleotide search of the retrieved sequence using the blastn search algorithm.

Results

Unknown Gene Identification

Nucleotide BLAST search (BLASTn) in the NCBI nucleotide collection database returned Strongylocentrotus purpuratus myoneurin (LOC100893972) mRNA as the top alignment between the unknown sequence and all sequences in the database. S. purpurpatus myoneurin achieved a max score of 3508, 99% query cover, 100% percent identity, and an E-value of 0 (Figure 1a). Translated nucleotide BLAST (BLASTx) in the protein database similarly returned S. purpuratus myoneurin as the top match for the query Sp20946 sequence. S. purpuratus myoneurin returned a max score of 641, 69% query cover, 100% percent identity, and an E-value of 0 (Figure 1b.)

BLAST searches were followed by a search in the Pfam database of the longest amino acid sequence yielded by translation of the Sp20946 nucleotide sequence. Domains within the translated Sp20946 sequence matched to the Chromo domain and the CBX7-C motif with E-values of 2.8e-14 and 2.3e-13, respectively. The translated sequence also matched to proteins of unknown function with E-values of 2200, 0.00066, and 730 (Figure 2).

Next, phylogenetic analysis on myoneurin mRNA sequences from a variety of

protostome and deuterostome species conveyed that Sp20946 was as closely related to myoneurin from arthropod species as it was to myoneurin from chordate species, though bootstrap values were exceptionally low for some arthropod-containing clades (Figure 3a). An additional phylogenetic tree that included transcript variants of S. purpuratus myoneurin showed closer grouping with myoneurin mRNA from chordate species. However, this tree also exhibited low bootstrap values and close grouping between the splice variants and some arthropod myoneurin mRNA sequences (Figure 3b).

Multiple sequence alignment between translated Sp20946 and myoneurin amino acid sequences from select protostome and deuterostome species showed similarity between Sp20946 and the selected myoneurin amino acid sequences (Figure 4).

Analysis of PCR Product

Gel electrophoresis of PCR amplified products showed a 548 base pair band in the lane corresponding to amplified SpGataE, and a 697 base pair band in the lane corresponding to amplified Sp20946. Intensity of the SpGataE band was much greater than that of the Sp20946 band (Figure 5).

After the PCR products were purified, SpGataE PCR product had a concentration of 65.5 ng/µL with an A260/A280 of 1.84, and Sp20946 PCR product had a concentration of 29.7 ng/µL with an A260/A280 of 1.83 (Table 2). Gel electrophoresis of purified PCR products showed a 532 base pair fragment for purified SpGataE PCR product and a 686 base pair band for purified Sp20946 product (Figure 6).

Analysis of Bacterial Transformation

To analyze cell competence, 1 control plate and 3 experimental plates with increasing

masses of added pBSK plasmid vectors were analyzed. The control plate plate 1) showed no colonies, plate 2 had 6 blue colonies for a transformation efficiency of 10800 cfu/µg, plate 3 had 312 blue colonies for a transformation efficiency of 56160 cfu/µg, and plate 4 had 2048 blue colonies for a transformation efficiency of 36864 cfu/µg.

To analyze ligation efficiency, plates containing 5 µg of SpGataE transformed plasmid (plate 5) and 5 µg of Sp20946 (plate 6) transformed plasmid were analyzed. Plate 5 showed 42 blue colonies and 8 white colonies for a transformation efficiency of 3600 cfu/µg and a ligation efficiency of 576 cfu/µg. Plate 6 showed 18 blue colonies and 7 white colonies for a transformation efficiency of 1800 cfu/µg and a ligation efficiency of 504 cfu/µg (Table 3).

Analysis of Sequencing Data

After select white colonies from plate 5 and plate 6 were expanded, plasmid DNA from each expanded culture was isolated, and the inserts from each were sequenced. Using a Nanodrop spectrophotometer, the concentration of the selected SpGataE plasmid isolate suspended in 50 µL of TE was measured to be 2502.2 ng/µL with an A260/A280 of 1.65, while the concentration of the selected Sp20946 plasmid isolate suspended in 50 µL of TE was measured to be 1764.4 ng/µL with an A260/A280 of 2.18 (Table 4). Insert regions from each plasmid were both sequenced (Figure 7a, b).

Subsequent alignment between the insert region of SpGataE plasmid isolate and the known sequence for SpGataE yielded no significant alignment (Figure 8a). Contrarily, alignment between the insert region in the Sp20946 plasmid isolate and the known sequence for Sp20946 yielded alignment with 96% query cover, 94.29% identity, and an E-value of 0 (Figure 8b). NCBI nucleotide BLAST search of the insert region from SpGataE plasmid isolate returned

matches to several cloning vectors, one of which was a pGEM-T vector for 96% query cover, 99.89% identity, a max score of 1581, and an E-value of 0 (Figure 8c).

Discussion and Conclusions

Unknown Gene Identification

Prior to conducting BLAST searches of the unknown gene, it was expected that the most significant match would be for a gene found in S. purpuratus since it was known a priori that the sequence came from this organism. In both nucleotide BLAST and translated nucleotide BLAST, S. purpuratus myoneurin (LOC100893972) yielded the highest max score, highest percent identity, highest query cover among all matched candidates, indicating that alignment with S. purpuratus myoneurin was high and not due to random chance. Furthermore, S. purpuratus myoneurin exhibited an E-value of 0 in both searches which further suggested that the match between the unknown gene Sp20946 and S. purpuratus myoneurin was significant (Figure 1a, b).

Continuing from the BLAST results, I expected Pfam search to return domains known to be found within S. purpuratus myoneurin protein. Pfam search returned significant matches to Chromo domain and CBX7-C domains (Figure 2), which were both known to be found in S. purpuratus myoneurin (U.S. National Library of Medicine). This conformed with the expectation set by the BLAST results.

The phylogenetic tree constructed from alignment between translated Sp20946 and myoneurin from protostome and other deuterostome species did not appear to reflect known evolutionary relationships in that the tree suggested translated Sp20946, putatively S. purpuratus myoneurin from past analyses, was as closely related to myoneurin from protostome species as it was to myoneurin from deuterostome species (Figure 3a). However, low bootstrap values in

some clades in the tree that were responsible for this grouping pattern cast doubt on the validity of the tree itself. Due to a lack of myoneurin sequences from non-chordate species, several predicted sequences were used, which may have skewed the groupings and bootstrap values.

Additionally, BLAST search of S. purpuratus myoneurin revealed two splice variants of myoneurin that exhibited closer grouping to chordate myoneurin sequences than did the originally used S. purpuratus myoneurin amino acid sequence (Figure 3b). This may have been a contributor to the strange alignment results yielded by the original tree (Figure 3a). Despite this improved grouping pattern, the modified tree that incorporated these splice variants similarly exhibited low bootstrap values, detracting from the validity of the tree.

Multiple sequence alignment between translated Sp20946 and myoneurin from protostome and other deuterostome species qualitatively exhibited appreciable similarity. Several stretches of amino acid residues were either similar or identical between the Sp20946 translated sequence and myoneurin amino acid sequences from other species (Figure 4), which aligned with the expected myoneurin identity established by BLAST results.

Analysis of PCR Product

Gel electrophoresis of unpurified PCR product yielded a 548 base pair band for the amplified region of SpGataE and a 697 base pair band for the amplified region of Sp20946, which were both slightly below the expected 569 base pair length for SpGataE and 764 base pair length for Sp20946 corresponding to a 3.87% error for SpGataE and 8.77% error for Sp20946 (Figure 5). Both of these percent error values were below the 10% threshold, so the result was deemed acceptable. A likely source of error was the difficulty in discerning the distance between the wells and the location of the bands. The Sp20946 band was significantly less intense than

that for SpGataE, so high exposure had to be used to visualize both bands simultaneously. A consequence of this method was that the wells were difficult to discern in the resultant image. For this reason, it is likely that errors in distance measurement occurred.

After purification, the concentration of SpGataE was over two-fold higher than that of Sp20946 (Table 2), which conformed with the intensity patterns on the previously run gel. A260/A280 ratios for both samples were within the 1.6-2.0 range to suggest that purification was successful (Figure 6). Gel electrophoresis of the purified PCR products showed a 532 base pair band for SpGataE and a 686 base pair band for Sp20946 (Figure 6). As with the results of the previous gel, these bands were shorter than expected. Though a measurement error was still a possible source of error, the high clarity of the wells and locations of the bands should have reduced this type of error. Other contributors to the observed discrepancy could have been minor degradation of the amplified sequences during the purification process and stochasticity in pore sizes across the lanes of the gel.

Analysis of Bacterial Transformation

For plate 1, no DNA was added, so the presence of 0 colonies after incubation conformed with expectation. The expected transformation efficiencies of plates 2 through 4 were on the order of 10^4- 10^6 cfu/µg, so the experimental transformation efficiencies, which were on the order of 10^4 cfu/µg, were acceptable despite being on the lower side of expected competence. Ligation efficiencies were expected to be 10-100x lower than the experimental transformation efficiencies, which was observed in both plates 5 and 6 (Table 3). To increase competence in future experiments, the time bacterial cells spend off-ice could be further minimized. Additionally, ligation efficiencies could be improved by using more insert DNA during ligation.

Analysis of Sequencing Data

Nonzero concentrations for SpGataE and Sp20946 plasmid isolates coupled with A260/A280 ratios of greater than 1.6 both suggest that plasmid isolation was successful. During the isolation procedure, plasmid isolate was visible at the bottom of each 1.5 mL tube prior to addition of TE buffer, so the Nanodrop results were consistent with this crude experimental observation. In the case of Sp20946 plasmid isolate, the A260/A280 ratio was greater than 2, which suggests that the Sp20946 plasmid isolate may have had RNA contamination (Table 4).

Sequencing of plasmid isolates was also deemed successful based on the clarity of peaks corresponding to each nucleotide. The Sp20946 plasmid isolate sequenced region exhibited slightly higher noise than that of the SpGataE plasmid isolate sequenced region, though the peaks were nonetheless still well-defined (Figure 7a, b). However, failure to find significant alignment between the SpGataE plasmid isolate sequenced region and the known sequence for SpGataE suggests that ligation of the SpGataE insert was unsuccessful (Figure 8a). Unlike the SpGataE plasmid isolate sequenced region, the Sp20946 plasmid isolate sequenced region exhibited significant alignment with the known sequence of Sp20946 as evidenced by a high query cover, high percent identity, and an E-value of 0 (Figure 8b).

Entry of the sequenced region into nucleotide BLAST returned significant alignment to plasmid vectors (Figure 8c), including a pGEM-T vector, to suggest that the plasmid may have ligated onto itself in such a way that disrupted the LacZ gene. Another possibility is that the selected colony was not actually white. The white colonies on the SpGataE culture plate were often times exceptionally small, making true determination of color fairly difficult. Hence, the sampled colony whose plasmid was sequenced may have truly been a faint blue. Although BLAST search of the SpGataE plasmid insert region returned higher max scores, query

coverages, and percent identities for plasmid vectors other than pGEM-T, these better matchings were likely a product of errors during sequencing and shorter subject sequence lengths. Since the sequencing data was high quality, it is unlikely that the negative finding for the SpGataE plasmid isolate and positive finding for the Sp20946 plasmid were erroneous.

In future experiments, restriction enzyme digestion of plasmid isolate and subsequent gel electrophoresis could be used to more quickly identify which allegedly white colonies were likely to contain the correct insert so that sequencing would be more likely to return a positive finding. Due to time constraints, this experimental step was not performed.

Concluding Statement

The objectives of the above experiments were to characterize unknown gene Sp20946 and clone a region of said gene into a plasmid vector to assist in further study. Since BLAST, Pfam, and multiple sequence alignment experiments supported the putative identity of the gene as S. purpuratus myoneurin, the aforementioned computational exploration was successful in identifying Sp20946. As for gene cloning, the experiments were moderately successful. Transformation efficiency and ligation efficiency were both on the lower side of acceptable, so it is recommended for these experiments to be performed again. Sequencing of the inserted region revealed successful incorporation of the Sp20946 insert into the plasmid vector, but failure to incorporate the SpGataE insert may be suggestive of experimental shortcomings. Due to evidenced successful ligation, Sp20946-plasmid containing cells from the selected expanded colony could be used as part of subsequent in-situ hybridization experiments for studying S. purpuratus myoneurin. Regarding SpGataE, a different colony would need to be used for further study.

Tables

Plate Designations and Identities of DNA Additives	
Plate Number	DNA Additive
1	No DNA
2	1 ng PBSK plasmid
3	10 ng PBSK plasmid
4	100 ng PBSK plasmid
5	5 ul SpGataE plasmid
6	5 ul Sp20946 plasmid

Table 1: Plate Designations and Identities of DNA Additives Table summary of plate designations and corresponding DNA additives.

Nanodrop Results for Purified PCR Products		
	Concentration (ng/ul)	A260/A280
SpGataE	65.5	1.84
Sp20946	29.7	1.83

Table 2: Table Summary of Nanodrop Results for Purified PCR Products Summary of Nanodrop results for purified SpGataE and Sp20946 PCR products.

Tables (Continued)

Bacterial Transformation Results					
Plate Number	White Colony Count	Blue Colony Count	Total Colonies	Transformation efficiency (cfu/µg)	Ligation Efficiency (cfu/µg)
1	0	0	0	N/A	N/A
2	0	6	6	10800	N/A
3	0	312	312	56160	N/A
4	0	2048	2048	36863	N/A
5	8	42	50	3600	576
6	7	18	25	1800	504

Table 3: Summary of Bacterial Transformation Results Table displaying number of white and blue colonies found on each culture plate. Transformation and ligation efficiencies, where applicable, were calculated.

Nanodrop Results for Plasmid Isolates		
	Concentration (ng/ul)	A260/A280
SpGataE	2502.2	1.65
Sp20946	1764.4	2.18

Table 4: Summary of Nanodrop Results for Recombinant Plasmid Isolates Table displaying the concentrations and purities of isolated plasmid DNA from SpGataE and Sp20946 transformed bacteria.

Figures

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download New Select columns Show 10

select all *10 sequences selected* GenBank Graphics Distance tree of results New MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
PREDICTED: Strongylocentrotus purpuratus myoneurin (LOC100893972), mRNA	Strongyloce...	3508	3508	99%	0.0	100.00%	2345	XM_003727353.3
PREDICTED: Strongylocentrotus purpuratus myoneurin-like (LOC115923504), mR...	Strongyloce...	2933	2933	88%	0.0	97.36%	2021	XM_030984312.1
PREDICTED: Lytechinus variegatus polycomb group protein Pc-like (LOC1214228...	Lytechinus ...	1338	1399	91%	0.0	77.82%	2825	XM_041618114.1
PREDICTED: Limulus polyphemus chromobox protein homolog 2-like (LOC106475...	Limulus pol...	108	108	4%	5e-18	84.54%	3236	XM_013936339.2
PREDICTED: Scophthalmus maximus E3 SUMO-protein ligase CBX4-like (LOC11...	Scophthalm...	99.6	99.6	4%	2e-15	82.47%	2992	XM_035638078.1
PREDICTED: Limulus polyphemus chromobox protein homolog 2-like (LOC106460...	Limulus pol...	98.7	98.7	6%	2e-15	78.86%	5007	XM_013920380.2
PREDICTED: Acanthochromis polyacanthus chromobox 4 (cbx4), mRNA	Acanthochr...	96.9	96.9	5%	8e-15	81.63%	3242	XM_022212110.1
PREDICTED: Amphiprion ocellaris E3 SUMO-protein ligase CBX4-like (LOC11157...	Amphiprion ...	96.9	96.9	5%	8e-15	81.63%	3869	XM_023283538.2
PREDICTED: Amphiprion ocellaris E3 SUMO-protein ligase CBX4-like (LOC11157...	Amphiprion ...	96.9	96.9	5%	8e-15	81.63%	3942	XM_035954330.1
PREDICTED: Bombus bifarius polycomb group protein Pc (LOC117209186), mRNA	Bombus bif...	95.1	95.1	6%	3e-14	77.60%	3144	XM_033450884.1

Figure 1a: BLASTn Results Above lists the top 10 hits for the query nucleotide sequence. The top two matches are mRNA sequences for myoneurin and myoneurin-like proteins from Strongylocentrotus purpuratus. The other eight matches consist of sequences for polycomb group (PcG) proteins or other CBX proteins, which is homologous to PcG group proteins. The myoneurin mRNA sequence had an E-value of 0, percent identity of 100%, and a query cover of 99%. The myoneurin-like mRNA sequence had an E-value of 0, percent identity of 97.36%, and a query cover of 88%.

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download New Select columns Show 10

select all *10 sequences selected* GenPept Graphics

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
myoneurin [Strongylocentrotus purpuratus]	Strongylocentrotus purpuratus	641	641	69%	0.0	100.00%	456	XP_003727401.1
myoneurin-like [Strongylocentrotus purpuratus]	Strongylocentrotus purpuratus	634	634	69%	0.0	98.47%	457	XP_030840172.1
polycomb group protein Pc-like [Lytechinus variegatus]	Lytechinus variegatus	281	468	58%	7e-84	69.44%	445	XP_041474048.1
PREDICTED: chromobox protein homolog 7-like [Priapulus caudatus]	Priapulus caudatus	94.4	94.4	12%	1e-18	55.42%	120	XP_014671719.1
hypothetical protein BSL78_25981 [Apostichopus japonicus]	Apostichopus japonicus	98.2	98.2	11%	1e-18	61.04%	250	PIK37189.1
chromobox protein homolog 7-like isoform X2 [Acanthaster planci]	Acanthaster planci	98.6	152	16%	7e-18	59.21%	387	XP_022101857.1
chromobox protein homolog 7-like isoform X1 [Acanthaster planci]	Acanthaster planci	95.9	150	16%	6e-17	59.74%	388	XP_022101855.1
polycomb group protein Pc-like [Aedes albopictus]	Aedes albopictus	95.5	95.5	10%	7e-17	58.21%	380	XP_019533021.2
AAEL014098-PA [Aedes aegypti]	Aedes aegypti	95.5	95.5	10%	7e-17	57.35%	377	EAT33624.1
polycomb group protein Pc-like [Aedes albopictus]	Aedes albopictus	95.5	95.5	10%	8e-17	58.21%	380	XP_029734931.1

Figure 1b: BLASTx Results Above lists the top 10 BLASTx results, which translates the query nucleotide sequence into a protein. The top two results are for myoneurin and myoneurin-like proteins from Strongylocentrotus purpuratus, which agrees with the BLASTn results. However, the next eight results slightly differ from the BLASTn results, though all results belong to CBX or PcG families as well. Myoneurin exhibits an E-value of 0, percent identity of 100%, and a query cover of 69%. Myoneurin-like exhibits an E-value of 0, percent identity of 98.47%, and a query cover of 69%.

Figures (Continued)

Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
Chromo	Chromo (CHRomatin Organisation MOdifier ...	Domain	CL0049	14	63	18	63	5	54	54	52.9	2.8e-14	n/a	Show
CBX7_C	CBX family C-terminal motif	Domain	n/a	340	374	341	374	2	34	34	49.7	2.4e-13	n/a	Show

Insignificant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
DUF2457	Protein of unknown function (DUF2457)	Domain	n/a	2	121	62	86	307	331	466	-2.4	2200	n/a	Show
DUF2457	Protein of unknown function (DUF2457)	Domain	n/a	192	342	205	338	59	196	466	19.1	0.00066	n/a	Show
DUF2457	Protein of unknown function (DUF2457)	Domain	n/a	360	444	394	436	358	404	466	-0.8	730	n/a	Show

Figure 2: Pfam Search Results PFAM matches for the unknown sequence include the Chromo family, with
an E-value of 2.8e-14, and the CBX7-C family, with an E-value of 2.4e-13. Insignificant matches for the unknown sequence include three proteins of unknown function with E-values of 0.00066, 730, and 22000.

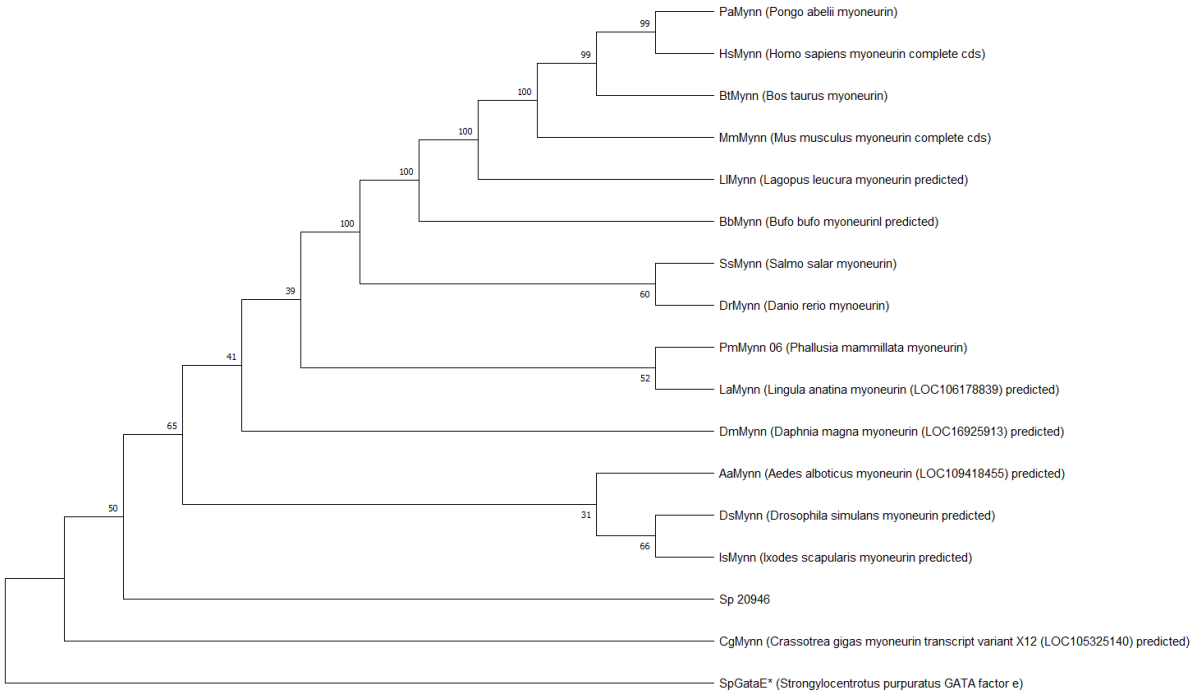

Figure 3a: Phylogenetic Tree of Myoneurin across Species Phylogenetic tree of translated Sp20946 and myoneurin amino acid sequences from various protostome and deuterostome species.

Figures (Continued)

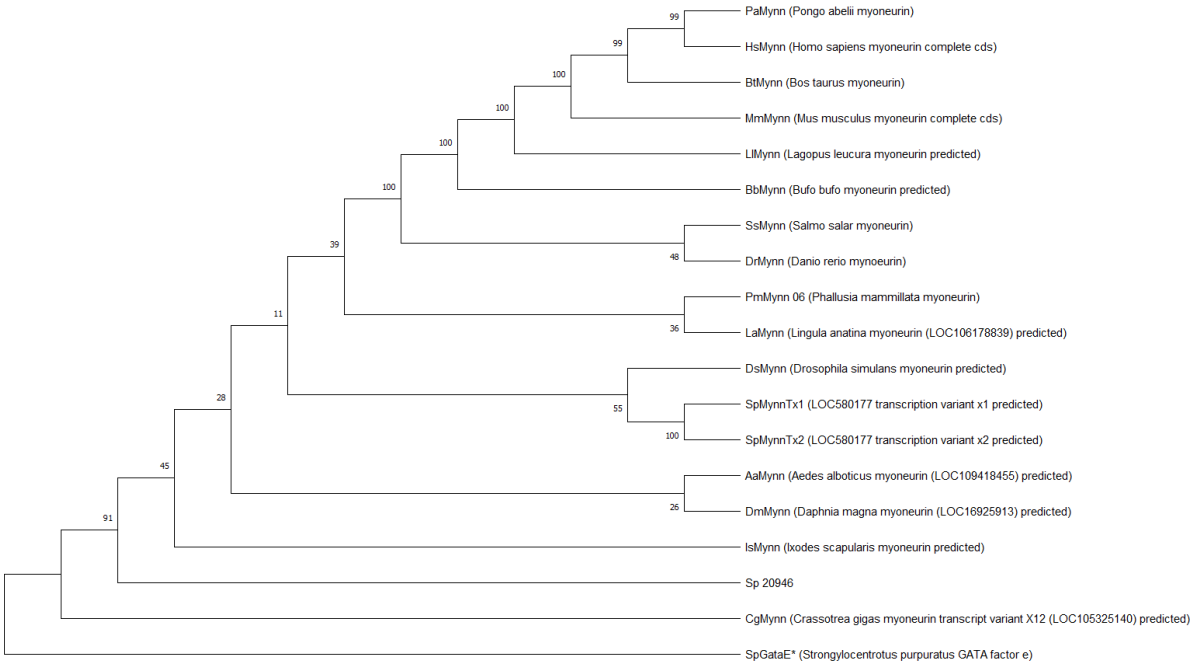

Figure 3b: Phylogenetic Tree with Added S. Purpuratus Myoneurin Splice Variants
Phylogenetic tree of myoneurin amino acid sequences from protostome and deuterostome species alongside S. Purpuratus myoneurin splice variants.

Figures (Continued)

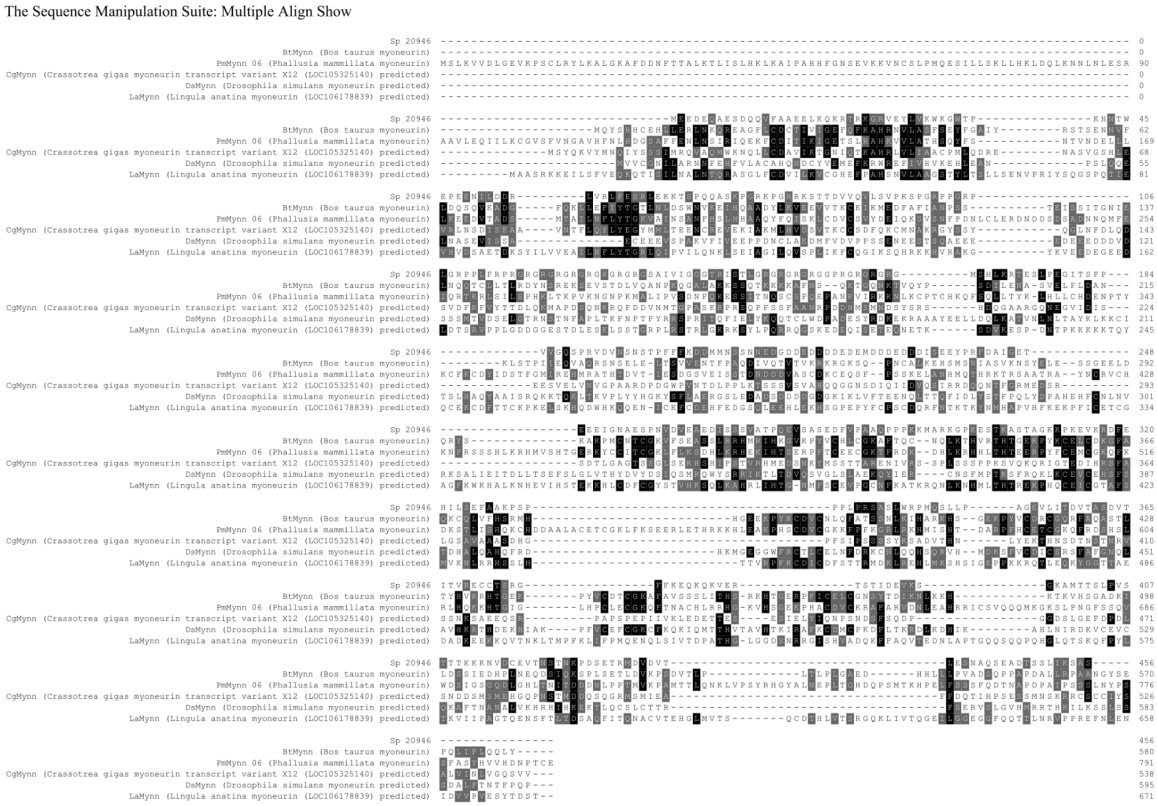

Figure 4: Multiple Sequence Alignment Results Visualized multiple alignment results among Sp20946, chordate myoneurin, and non-chordate myoneurin. Identical amino acid residues are highlighted in black. Similar amino acid residues are highlighted in dark gray.

Figures (Continued)

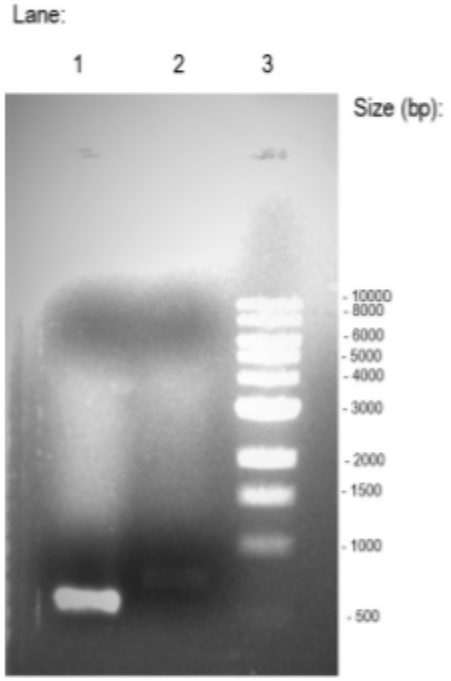

Figure 5: Gel Electrophoresis of PCR Product Gel electrophoresis results for amplified PCR product. The lane designations are as follows: Lane 1) SpGataE; Lane 2) Sp20946; Lane 3: DNA ladder. Annotations for the sizes of ladder sequences appear on the right side of the gel image.

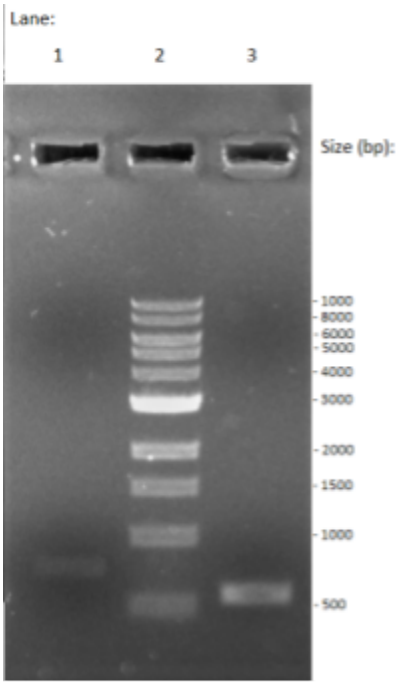

Figure 6: Gel Electrophoresis of Purified PCR Product Gel electrophoresis results for purified PCR product. Lane designations are as follows: Lane 1) Sp20946; Lane 2) DNA Ladder; Lane 3) SpGataE. Annotations for the sizes of the ladder sequences appear on the right side of the gel image.

Figures (Continued)

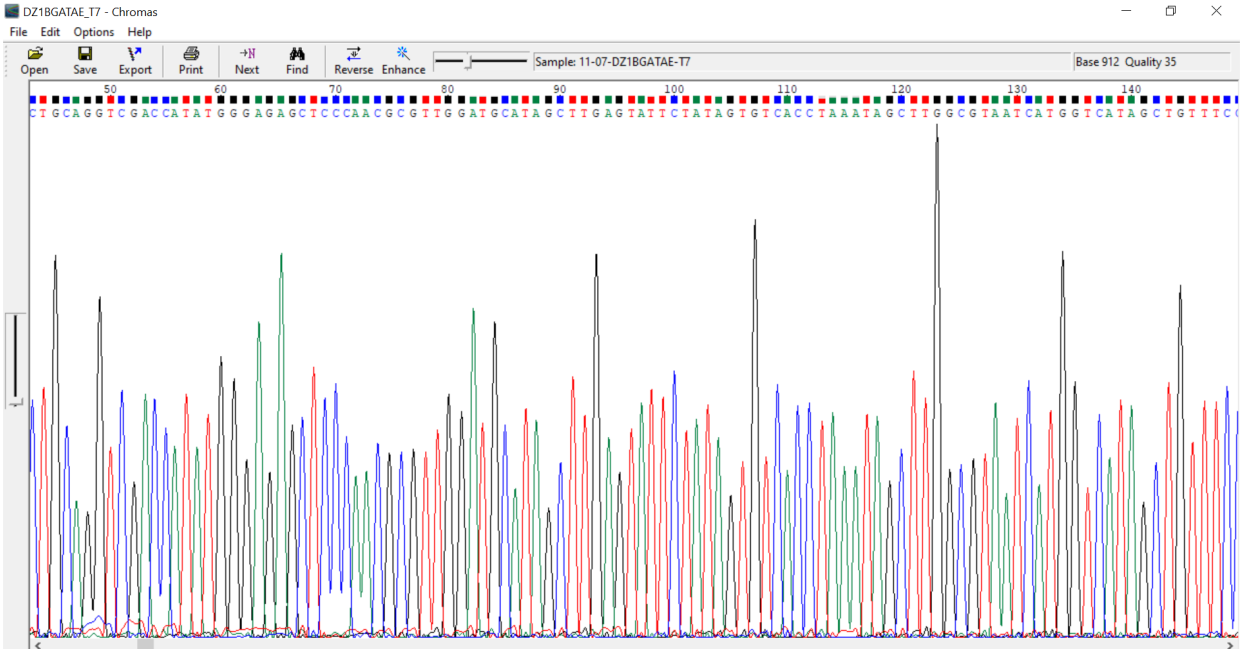

Figure 7a: SpGataE Insert Region Sequence Excerpt Sequence excerpt for insert region in isolated SpGataE recombinant plasmid

Figure 7b: Sp20946 Insert Region Sequence Excerpt Sequence excerpt for insert region in isolated Sp20946 recombinant plasmid.

Figures (Continued)

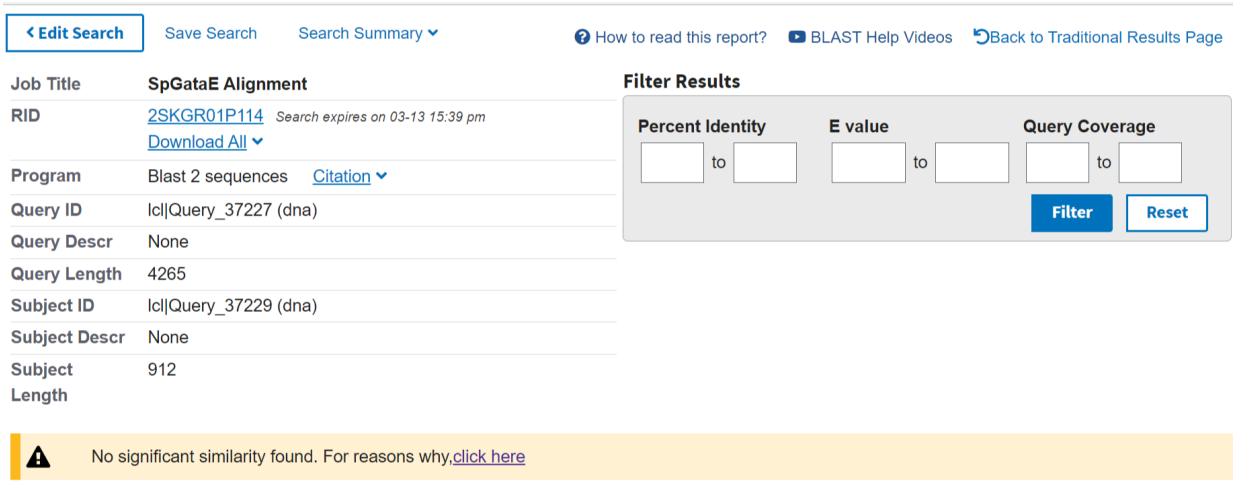

Figure 8a: SpGataE BLAST Alignment Results Alignment results between sequenced SpGataE plasmid insert region and known sequence for SpGataE. No significant alignment was observed.

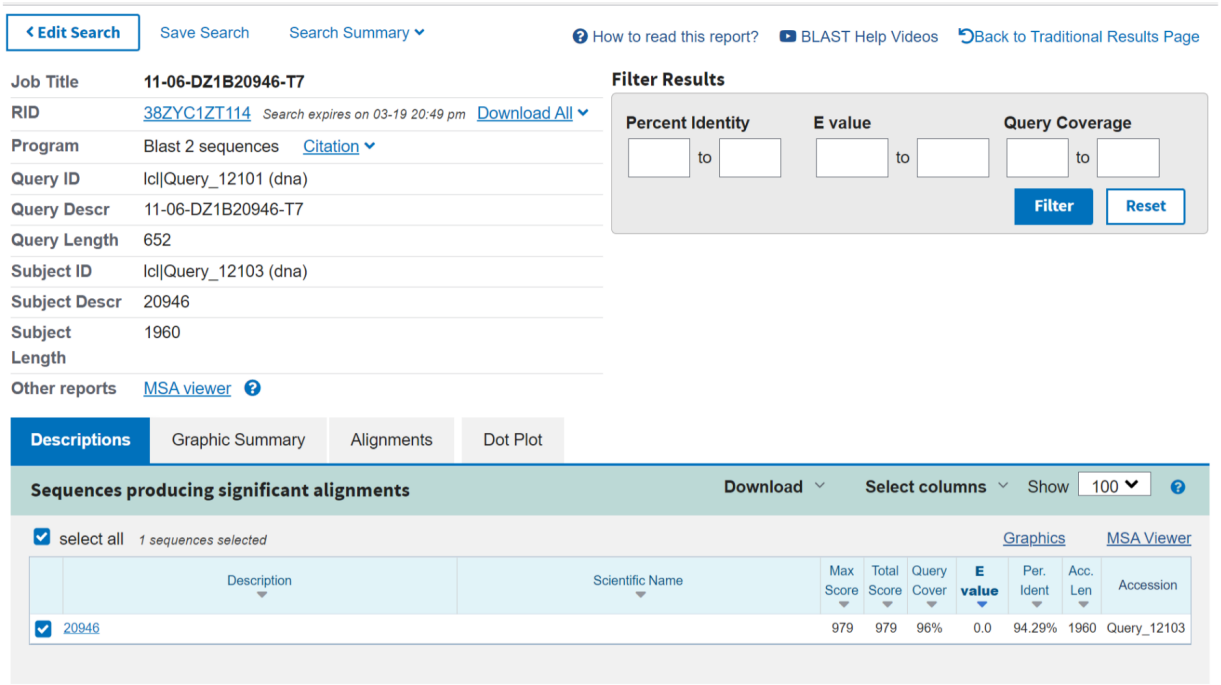

Figure 8b: Sp20946 BLAST Alignment Results Alignment results between sequenced Sp20946 plasmid insert region and known sequence for Sp20946. Alignment yielded in 96% query cover, 94.29% identity, and an E-value of 0.

Figures (Continued)

Sequences producing significant alignments Download Select columns Show 50

select all *50 sequences selected* GenBank Graphics Distance tree of results MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Cloning vector psiLentGene Basic, complete sequence	Cloning vector psiLentGene Basic	1637	1637	100%	0.0	99.78%	3015	AY508731.1
Cloning vector pG-Sable, complete sequence	Cloning vector pG-Sable	1612	1612	98%	0.0	99.89%	4785	KC537291.1
Cloning vector pMono_T-vector, complete sequence	Cloning vector pMono_T-vector	1612	1612	98%	0.0	99.89%	5010	JN681269.1
Cloning vector U6p-gRNA1, complete sequence	Cloning vector U6p-gRNA1	1611	1611	97%	0.0	100.00%	3493	KJ796490.1
Cloning vector pVEcLa4, complete sequence	Cloning vector pVEcLa4	1606	1606	97%	0.0	100.00%	6217	MK388215.1
Cloning vector pFH6, complete sequence	Cloning vector pFH6	1597	1597	97%	0.0	100.00%	3620	KY080689.1
Cloning vector pGEM-5Zf(-)	Cloning vector pGEM-5Zf(-)	1589	1589	99%	0.0	98.79%	3001	X65309.2
Cloning vector pGEM-5Zf(+)	Cloning vector pGEM-5Zf(+)	1589	1589	99%	0.0	98.79%	3000	X65308.2
Cloning vector pCWP1Fluc, complete sequence	Cloning vector pCWP1Fluc	1585	1585	97%	0.0	99.33%	8083	MF062157.1
Cloning vector pVLH/int(+), complete sequence	Cloning vector pVLH/int(+)	1585	1585	97%	0.0	99.33%	8658	AF338824.1
Expression vector pB-TUB_DEST_GFP DNA, complete sequence	Expression vector pB-TUB_DEST_GFP	1585	1585	97%	0.0	99.33%	6655	LC136879.1
Cloning vector pDuExAc6, complete sequence	Cloning vector pDuExAc6	1585	1680	97%	0.0	99.33%	7271	GU370778.1
Plant expression vector pDuExP (pDuExAn6), complete sequence	Plant expression vector pDuExP (pDuExAn6)	1585	1680	97%	0.0	99.33%	7315	EF565883.1
Plant expression vector pDuExAn1, complete sequence	Plant expression vector pDuExAn1	1585	1680	97%	0.0	99.33%	7306	EF565882.1
Cloning vector pGEM-WIZ, complete sequence	Cloning vector pGEM-WIZ	1585	1585	97%	0.0	99.33%	3130	DQ917671.1
Cloning vector pGEM-S1, complete sequence	Cloning vector pGEM-S1	1585	1585	97%	0.0	99.33%	3038	DQ917670.1
Cloning vector pGEM-URA3, complete sequence	Cloning vector pGEM-URA3	1585	1585	97%	0.0	99.33%	4350	AF173954.1
Cloning vector pCWP2Fluc, complete sequence	Cloning vector pCWP2Fluc	1584	1584	96%	0.0	100.00%	8457	MF062156.1
Cloning vector pSpark II, complete sequence	Cloning vector pSpark II	1584	1584	96%	0.0	100.00%	3001	GQ489185.1
Cloning vector pCPT8, complete sequence	Cloning vector pCPT8	1584	1584	99%	0.0	98.68%	3012	AF092546.1
Cloning vector pGEM-T_blanc_GFPco8g_tGAL_URA3locus_tTK, comple...	Cloning vector pGEM-T_blanc_GFPco8g_tGA...	1581	1581	96%	0.0	99.89%	7148	MN271957.1

Figure 8c: Nucleotide BLAST Results for SpGataE Plasmid Isolate Insert Region
Nucleotide database search results of the SpGataE plasmid isolate insert region yielded matches to cloning vectors, pGEM-T being among the list matches.

References

Adonin, L., Drozdov, A., Barlev, N. A., 2021. Sea Urchin as a Universal Model for Studies of Gene Networks. Front. Genet. 11:627259.

Jones, D. O., Cowell, I. G., Singh, P. B., 2000. Mammalian chromodomain proteins: Their role in Genome Organisation and expression. BioEssays 22(2), 124–137.

U.S. National Library of Medicine. (n.d.). Myoneurin [strongylocentrotus purpuratus] - protein - NCBI. National Center for Biotechnology Information. Retrieved January 28, 2022, from https://www.ncbi.nlm.nih.gov/protein/XP_003727401.1?report=genbank&log%24=protalign&blast_rank=1&RID=00HTF183013

Wang, R., Taylor, A. B., Leal, B. Z., Chadwell, L. V., Ilangovan, U., Robinson, A. K., Schirf, V., Hart, P. J., Lafer, E. M., Demeler, B., Hinck, A. P., McEwen, D. G., Kim, C. A., 2010. Polycomb group targeting through different binding partners of ring1b C-terminal domain. Structure 18(8), 966–975.