Data S2. Octopus Myo-VIIa genomic alignment, eelgrass NdhL intron retention, related to Figure 5.

Octopus Myo-VIIa genomic alignments

We give more detailed alignment information, including genome coordinates, for Myo-VIIa anchor and targets in the two genomes (*O. sinensis* and *O. bimaculoides*). The unannotated alternative first exon "1b" (expressed specifically in statocyst tissue) is present in both octopus genomes. Note that we do not know the starting point of exon "1b".

Sequence matching exon 2 is missing from the *O. bimaculoides* genome. We show the sense-strands (reverse-complements of the anchor and targets).

Alignment to O. sinensis genome.

Sequences matching exons 1a,1b, and 2 are present in the O. sinensis genome (though exon 1b is not annotated) and have the expected gt or ag dinucleotide splicing signals.



anchor is in blue; target 1 in green; target 2 in red.

brown = genomic sequence from *O. sinensis*, showing splice signals. Splice dinucleotides and ATGs <u>underlined</u>.

lowercase-orange = SNP differences in O. sinensis from anchor-targets (O. bimaculoides)

All upstream ATGs have downstream stop codons shortly after, and the annotated start codon has an upstream stop shortly before. Thus the alternative first exons do not introduce additional protein sequence at the N-terminus.

Numbers are genome coordinates for *O. sinensis* chromosome LG8 = NC_043004.1 "Octopus vulgaris isolate Ov201803 linkage group LG8, ASM634580v1". The *O. sinensis* myoVIIa gene model is LOC115214860.

Alignment to *O. bimaculoides* genome.

In *O. bimaculoides*, there is no good match for the relevant exon 2 portion nor for the anchor (uppercase sequence): atataatTGGATTTTAAACAAAAGCAAAAATGG. Sequences matching exons 1a and 1b are present in the *O. bimaculoides* genome, though they are unannotated. The splice donor for myoVIIa exon 1a matches a splice donor in the annotated noncoding RNA XR_008264717.1 (gene LOC128248543) that is located upstream of the annotated *O. bimaculoides* myoVIIa transcript.

9,633,631	9,633,645	9,633,690			
CATCCCGGAT1	TTCTACTCAATTCTGGCGTTTCTGTTGCC	GGCATGCCTTATAACTTGTG <u>gt</u> aagtat O.bimaculoides splice donor, XR O	08264717.1		
CATCCCGGATT	TTCTACTCAATTCTacCGTct-TGTTGCC	GGC <u>ATG</u> CCTTATAACTTGTG <u>gt</u> aggta	"1a"		
		MetProTyrAsnLeu*** MetV			
consensus 1	TACTCAATTCTGGCGTTTCTGTTGCC GGC <u>ATG</u> CCTTATAACTTGGATATAATTGGATTTTAAACAAAAAGCAAAA <mark>ATG</mark> G				
_					
consensus_2	TTTATTATCACCAAT <u>ATG</u> GACGGAAA <mark>TAGTGTATCCATTTATTAAGATATAA</mark> TTGGATTT <u>TAA</u> ACAAAAAGCAAAA <mark>ATG</mark> G				
	MetAspGlyAsnSerValSerIleTyr***				
	exon 2 no match in 0. bimaculoides				
	9,658,490	9,658,536			
A <u>ATG</u> ACTCACI	TTATTTATTATCACCAAT <u>ATG</u> GACGGAAA	TAGTGTATCCATTTATTAAG <u>gt</u> aagc	"1b"		
MetThrHisH	PheIleTyrTyrHisGlnTyrGlyArgLys	(expressed in statocyst tissue)			
	MetAspGlyAs	nSerValSerIleTyr***			

Numbers are genome coordinates for *O. bimaculoides* chromosome 8 = NC_068988.1 "Octopus bimaculoides isolate UCB-OBI-ISO-001 chromosome 8, ASM119413v2, whole genome shotgun sequence". The *O. bimaculoides* myoVIIa gene model is LOC106880717.

Possible mis-assembly of Myo-VIIa gene in O. sinensis

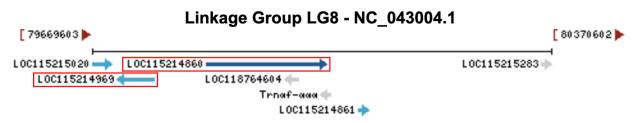
There appears to be an assembly issue for the *O. sinensis* Myo-VIIa gene on chromosome LG8 that matches our anchor-targets. There are several *O. sinensis* genes annotated with "myosin-VIIa" in the genome; they are listed below, including their protein domain content.

Name/Gene ID	Description	Location	a.a.	transcripts, protein domains
LOC115224051	unconventional	Chromosome LG24, NC_043020.1	920	one variant. Has Motor_domain and IQ
	myosin-VIIa	(1052001410643911)		domain.
LOC115215798	myosin-VIIa	Chromosome LG9, NC_043005.1	1,239	two variants, same a.a. size; for X2, longer
		(4937897149480887)		RNA. No motor domain. Has [MyTH4,
				B41/FERM1_F1_Myosin-VII,
				FERM_C1_MyoVII], [MyTH4,
				B41/Ubl1_cv_Nsp3_N-like, PH-like].
LOC115214860	unconventional	Chromosome LG8, NC_043004.1	855	one variant. Has MYSc_Myo7 motor and IQ
	myosin-VIIa	(7984399680027230)		domain.
LOC115214969	myosin-VIIa	Chromosome LG8, NC_043004.1	1,237	one variant. No motor domain. Has [MyTH4,
		(7970714879764229,		B41/FERM1_F1_Myosin-VII, PH-like,
		complement)		FERM_C1_MyoVII], SH3, [MyTH4,
				B4/FERM2_F1_Myosin-VII, PH-like,
				FERM_C2_MyoVII].
LOC115229165	unconventional	NW_021832531.1		the RNA is annotated as a pseudogene
	myosin-VIIa-like	(136771143156)		
LOC115217708	unconventional	Chromosome LG11, NC_043007.1	2,422	variant X1, the longest. Has MYSc_Myo22
	myosin-X	(7082020471380859)		motor and [MyTH4,
				B41/FERM_F1_DdMyo7_like, PH-like],
				[MyTH4, B41/FERM_F1_DdMyo7_like,
				PH-like].

Red is the myosin-VIIa gene that matches our anchor-targets. The two genes on chromosome LG8 are marked by pale-orange background.

Slash "/" means the domains overlap. Square brackets "[]" highlight repeat structure.

Neither of the two myo-VIIa genes on Ch. LG8 has a full complement of protein domains: LOC115214860 has the N-terminal myosin motor domain, but lacks the tail domains. LOC115214969 has all the tail domains but lacks the motor domain. The two genes are adjacent on the chromosome, but are in head-to-head orientation, as shown in this NCBI screenshot (the genes have been marked with red boxes post-facto):



If LOC115214860 was inverted, then all domains would be present in the correct linear order.

Zostera marina (eelgrass) NADPH quinone oxidoreductase subunit L (NdhL) intron retention

We have confirmed the intron retention event by sequence extension of target 4 from raw reads to reach the end of exon 2 (data not shown).

We show here the predicted translation of the intron retention isoform of NdhL (target 4 of Figure 5D). It causes a frameshift and termination shortly after the end of exon 3 (where the other targets are located). The anchor is in exon 4.

anchor is in blue. target4 is in red. target1 is in green, target2 and target3 are aligned underneath target1, differences in magenta. Intron sequence is shown in lowercase, splice dinucleotides underlined. Reverse-complements of anchor-targets are shown.

The intron retention frameshift and termination occur within the second transmembrane domain of the protein, as shown below. These topologies were predicted by CCTOP (<u>https://cctop.ttk.hu/</u>) The third transmembrane domain in the full-length protein is not predicted by all programs.

>full-length_NdhL MTHLLLPLPSKVTGAFNHREWSCHRVPHPVSSAQRTRPLISASISKTKKINGRLMCNIESSKA**TNSTLLHLGVLLTSIA** DEPAFAVTGS<u>NNYEODLTS**VLIOSGAFAFFYFLIMPPIIM**NWLRLRWYKRKLFET**YLOFMFVFLFFPGILLWAPFI**NFR <u>RLPRD</u>PTMKHPWSTPRDSST</u>

>intron-retention_NdhL

MTHLLLPLPSKVTGAFNHREWSCHRVPHPVSSAQRTRPLISASISKTKKINGRLMCNIESSKA**TNSTLLHLGVLLTSIA** DEPAFAVTGS<u>NNYEODLTS**VLIOSGAFAFFYFLIMPV**YNCKVILT</u>

blue = cytoplasmic; **red = transmembrane;** green = extracellular. The InterPro NdhL domain is <u>underlined</u>.