

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

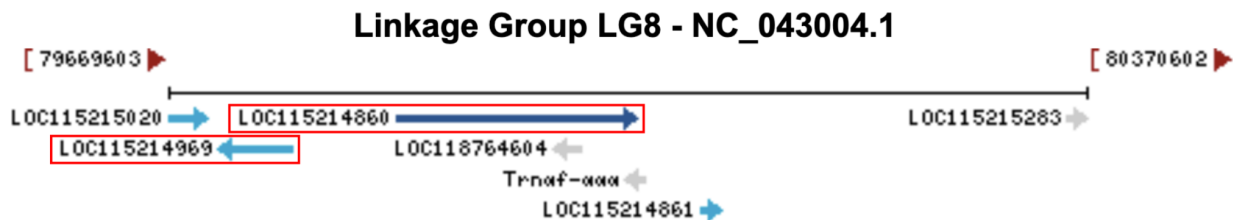
There appears to be an assembly issue for the *O. sinensis* Myo-VIIa gene on chromosome LG8 that matches our anchor-targets. There are several *O. sinensis* genes annotated with “myosin-VIIa” in the genome; they are listed below, including their protein domain content.

Name/Gene ID	Description	Location	a.a.	transcripts, protein domains
LOC115224051	unconventional myosin-VIIa	Chromosome LG24, NC_043020.1 (10520014..10643911)	920	one variant. Has Motor_domain and IQ domain.
LOC115215798	myosin-VIIa	Chromosome LG9, NC_043005.1 (49378971..49480887)	1,239	two variants, same a.a. size; for X2, longer 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 myosin-VIIa	Chromosome LG8, NC_043004.1 (79843996..80027230)	855	one variant. Has MYSc_Myo7 motor and IQ domain.
LOC115214969	myosin-VIIa	Chromosome LG8, NC_043004.1 (79707148..79764229, complement)	1,237	one variant. No motor domain . Has [MyTH4, B41/FERM1_F1_Myosin-VII, PH-like, FERM_C1_MyoVII], SH3, [MyTH4, B4/FERM2_F1_Myosin-VII, PH-like, FERM_C2_MyoVII].
LOC115229165	unconventional myosin-VIIa-like	NW_021832531.1 (136771..143156)	---	the RNA is annotated as a pseudogene
LOC115217708	unconventional myosin-X	Chromosome LG11, NC_043007.1 (70820204..71380859)	2,422	variant X1, the longest. Has MYSc_Myo22 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.

```
>LFYR01000468.1:167070-167279 Zostera marina strain Finnish scaffold_137, whole genome shotgun
sequence
GCAAGATTTGACTTCAGTGCTTATACAATCAGGAGCATTTGCTTTCTTCTACTTCCTTATCATGCCGgta
...SerGlyAlaPheAlaPhePheTyrPheLeuIleMetProVal
target2          TGTTTCTTCTACTTCCTTATCATGCC
                  ValPhePheTyrPheLeuIleMetPro
target3          TGCTTTCTTCTACTTCCTTGCATGCC
                  AlaPhePheTyrPheLeuValMetPro

tataattgcaaagtgataacttacataataattatctcattgactttacaactgcgaacagtactgattga
TyrAsnCysLysValIleLeuThr*****LeuPheHis***LeuTyrAsnCysGluGlnTyr***LeuI

tcattgattgaactgcatgatcgaagCCTATCATCATGAATTGGCTTCGATTGAGATGGTACAAGCGCAA
leIleAsp***ThrAla***SerLysProIleIleMetAsnTrpLeuArgLeuArgTrpTyr...
```

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 (<https://cctop.ttk.hu/>) The third transmembrane domain in the full-length protein is not predicted by all programs.

```
>full-length_NdhL
MTHLLLPLPSKVTGAFNHREWSCHRVPHPVSSAQRTRPLISASISKTKKINGRLMCNIESSKATNSTLLHLGVLLTSIA
DEPAFAVTGSNNYEODLTSVLIOSGAFAFYFLIMPPIIMNWLRLRWYKRKLFETYLOFMFVFLFFPGILLWAPFINFR
RLPRDPTMKHPWSTPRDSST

>intron-retention_NdhL
MTHLLLPLPSKVTGAFNHREWSCHRVPHPVSSAQRTRPLISASISKTKKINGRLMCNIESSKATNSTLLHLGVLLTSIA
DEPAFAVTGSNNYEODLTSVLIOSGAFAFYFLIMPVYNCKVILT
```

blue = cytoplasmic; red = transmembrane; green = extracellular.
The InterPro NdhL domain is underlined.