Science Advances
AAAS

Supplementary Materials for

Evolution of a chordate-specific mechanism for myoblast fusion

Haifeng Zhang *et al.*

Corresponding author: Pengpeng Bi, pbi@uga.edu; Alberto Stolfi, alberto.stolfi@biosci.gatech.edu

Sci. Adv. **8**, eadd2696 (2022)
DOI: 10.1126/sciadv.add2696

The PDF file includes:

Figs. S1 to S18

Other Supplementary Material for this manuscript includes the following:

Tables S1 and S2
Supplemental Files 1 to 3
Movies S1 to S4

Figure S1

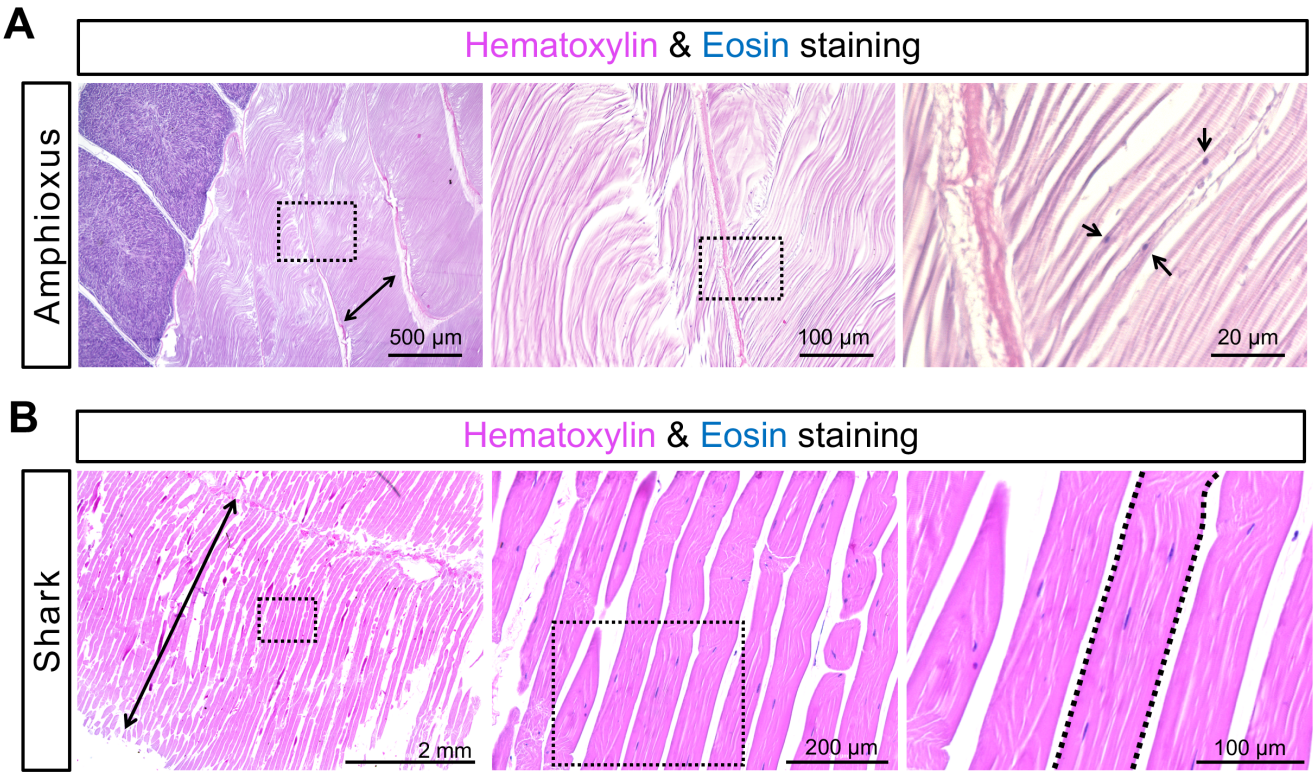


Figure S1. Muscle histological analyses of adult amphioxus and shark.

Histology of longitudinal sections of muscle tissues dissected from adult amphioxus (**A**) and shark (*Squalus acanthias*) (**B**). Arrows point to mononucleated myocytes in A. One multinucleated myofiber is outlined in B.

Figure S2

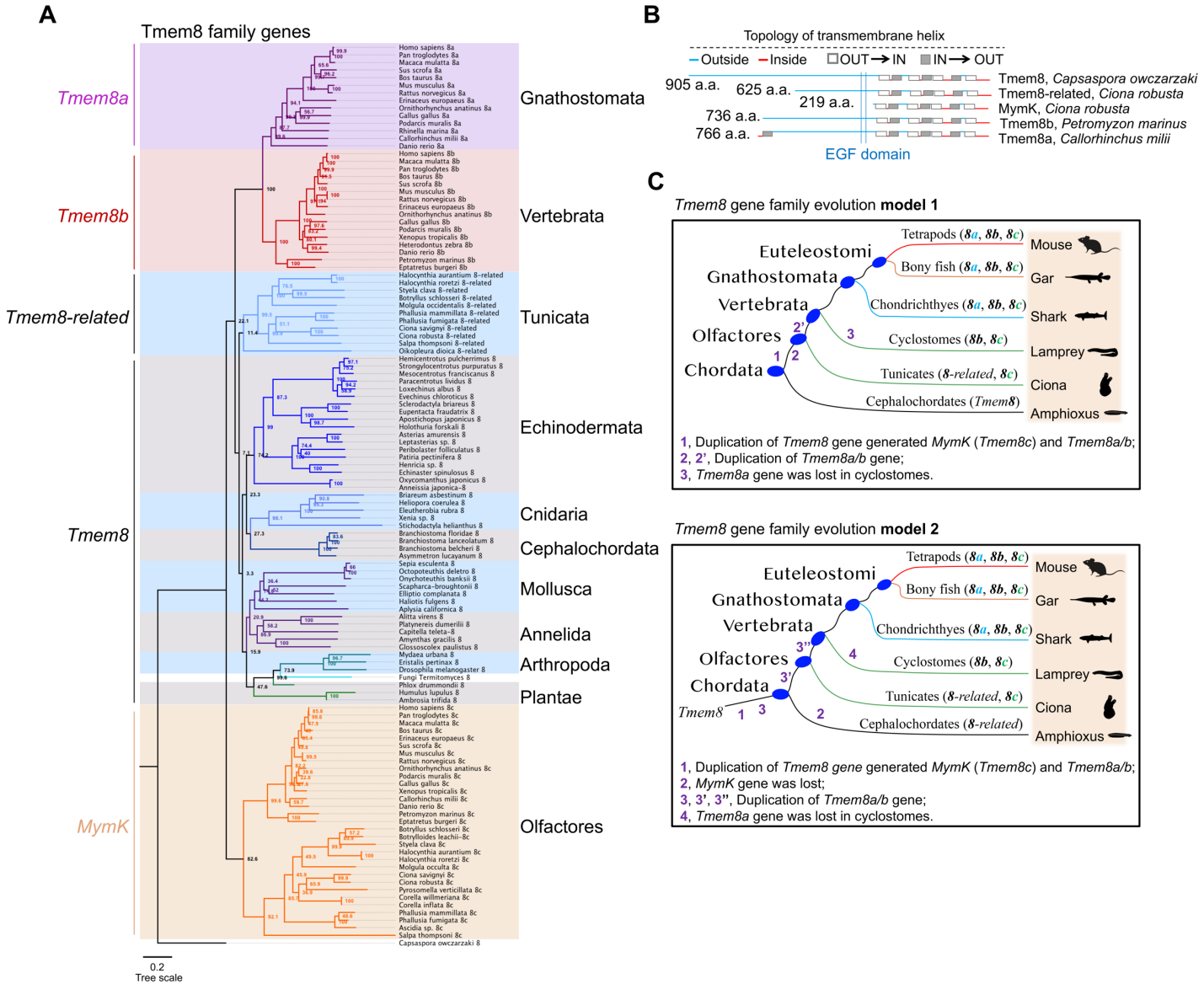


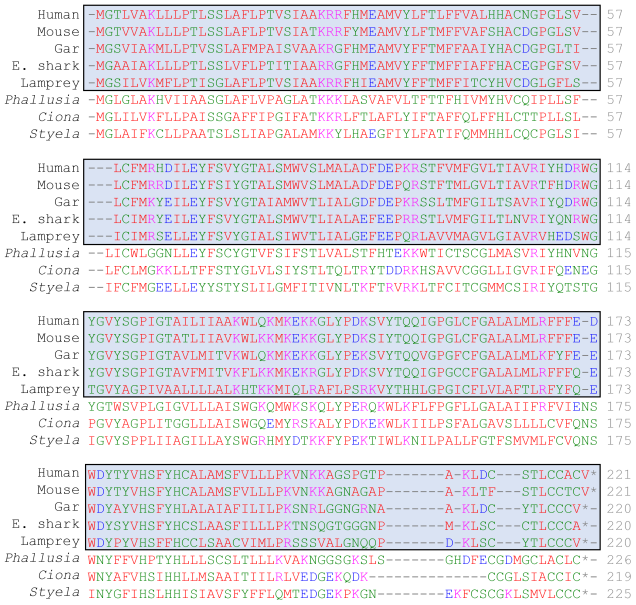
Figure S2. Phylogenetic analysis of the Tmem8 gene family.

(**A**) Phylogenetic tree of *Tmem8* family genes inferred by a distance-based method (neighbour joining). The bootstrap percentages obtained from 1,000 replicates were shown in the cladogram. (**B**) Topology and domain predictions (SCAMPI model) for Tmem8 family transmembrane proteins. EGF domain: epidermal growth factor like domain. (**C**) Scenarios of two-round gene duplications that gave rise to various Tmem8 family gene members in Chordata. Model 1 represents a more parsimonious scenario of evolution.

Figure S3

A

Alignment of non-vertebrate and vertebrate MymK proteins

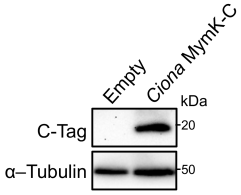


B

Clade	Species	Exon 1 (bp)	Exon 2 (bp)	Exon 3 (bp)	Exon 4 (bp)	Exon5 (bp)	ORF (a.a.)	Protein identity vs. human MymK
Jawed vertebrate	Human	135 ■	115 ▶	149 ◣	117 ■	147 ■	221	100%
	Zebrafish					144 ■	220	75.2%
	Elephant shark					144 ■	220	74.8%
Jawless vertebrate	Arctic lamprey					150 ■	222	56.1%
	Sea lamprey					144 ■	220	55.5%
Ascidian (tunicates)	*P. mammillata*	135 ■	146 ◤	121 ◀	122 ◤	151 ◀	225	38.1%
	C. robusta					133 ◀	219	36.6%
	H. aurantium					151 ◀	225	36.3%
	H. roretzi					151 ◀	225	35.8%
	B. schlosseri					151 ◀	225	35.1%
	B. leachii					151 ◀	225	34.2%
	M. occulta					160 ◀	228	34.1%
	S. clava					151 ◀	225	33.8%
	C. inflata					136 ◀	220	33.6%
Thaliacean (tunicates)	*P. verticillata*	135 ■	146 ◤	121 ◀	122 ◤	151 ◀	225	34.9%
	S. thompsoni	138 ■	158 ◤	121 ◀	122 ◤	163 ◀	234	26.5%

Nucleotide number and complementarity of exon: ■ 3N ▶ 3N+1 ◣ 2+3N ◤ 3N+2 ◀ 1+3N

C



D

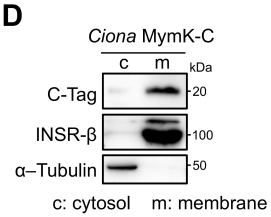


Figure S3. Sequence alignments, gene structures and validation of MymK ortholog expression.

(**A**) Alignment of MymK proteins from eight species: Human (*Homo sapiens*), Mouse (*Mus musculus*), Gar (*Lepisosteus oculatus*), Elephant shark (*Callorhinchus milii*), Lamprey (*Petromyzon marinus*), *Phallusia* (*Phallusia mammillata*), *Ciona* (*Ciona robusta*), *Styela* (*Styela clava*). Substantial degree of sequence identity is observed for vertebrate MymK sequences (boxed) but not MymK from tunicates (*Phallusia*, *Ciona*, *Styela*). (**B**) Exon size and structure for various MymK orthologs. (**C**, **D**) Western blots confirming the expression of C-tagged *Ciona* MymK protein in human myoblasts (**C**, total protein; **D**, after membrane fractionation). α-Tubulin blot was used as a positive control of cytosolic proteins. Insulin receptor β (INSR-β) blot was used as a positive control of membrane proteins.

Figure S4

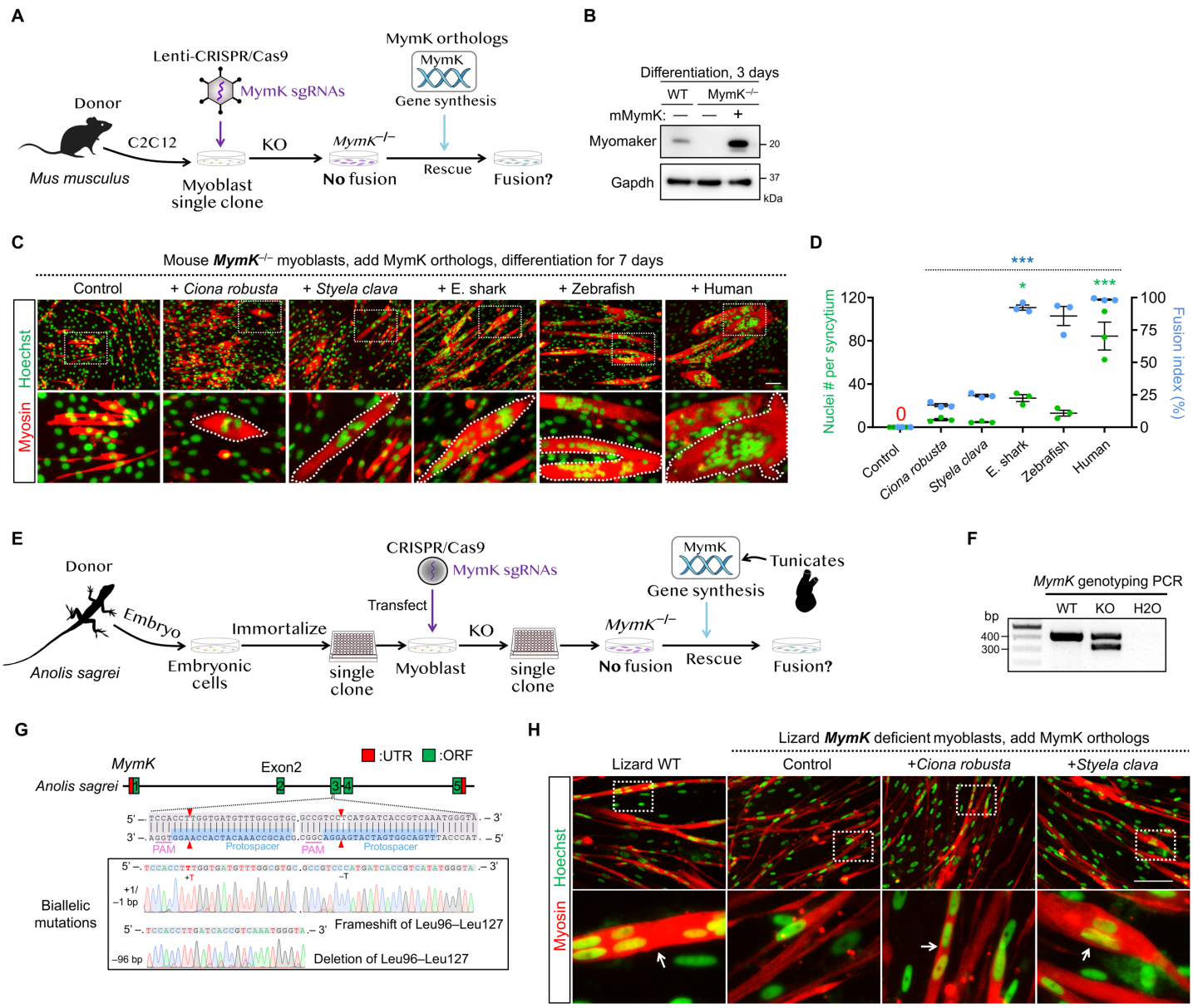


Figure S4. Tunicate MymK proteins induce fusion of mouse and lizard *MymK*$^{-/-}$ myoblasts.

(**A**) Schematic of experimental design to generate mouse *MymK*-deficient myoblasts and test the fusogenic activity of *MymK* orthologs. Expression of Cas9 and three single guide RNA (sgRNAs) that targets the first three exons are delivered by lentiviral infection. (**B**) Western blots validated the successful depletions of MymK protein in one isolated single clone of CRISPR treated mouse C2C12 myoblasts. (**C**) Myosin immunostaining of mouse *MymK*$^{-/-}$ myoblasts transfected with MymK orthologs. Tunicate (*Styela* and *Ciona*) MymK genes induced formations of muscle syncytia (outlined), which are smaller compared to syncytia induced by MymK proteins from jawed vertebrates; E. shark: elephant shark. (**D**) Measurements of myoblast fusion in **C** after seven days of differentiation. (**E**) Schematic of experimental design to isolate myogenic clone from lizard (*Anolis sagrei*) embryonic cells, inactivate *MymK* gene and finally test the fusogenic activity of tunicate MymK proteins. CRISPR/Cas9-mediated mutagenesis of lizard *MymK* gene was performed using a pair of sgRNAs targeting exon 3. (**F**, **G**) Genotyping PCR (**F**) and Sanger sequencing (**G**) analyses of lizard *MymK* locus revealed a mutant myoblast clone where two alleles of *MymK* gene are disrupted. The predicted sgRNAs cut sites are located between the codons Leu96 and Leu127. KO: knockout. (**H**) Myosin immunostaining of lizard *MymK*-deficient myoblasts transfected with various *MymK* orthologs. Consistent with the results in human and mouse myoblasts, expression of tunicate MymK proteins induce fusion of lizard myoblasts. Cells are differentiated for nine days before analysis. Arrows point to multinucleated myotubes. Scale bars, 100 μm. Data are means ± SEM. * $P < 0.05$; *** $P < 0.001$, compared to control group, one-way ANOVA.

Figure S5

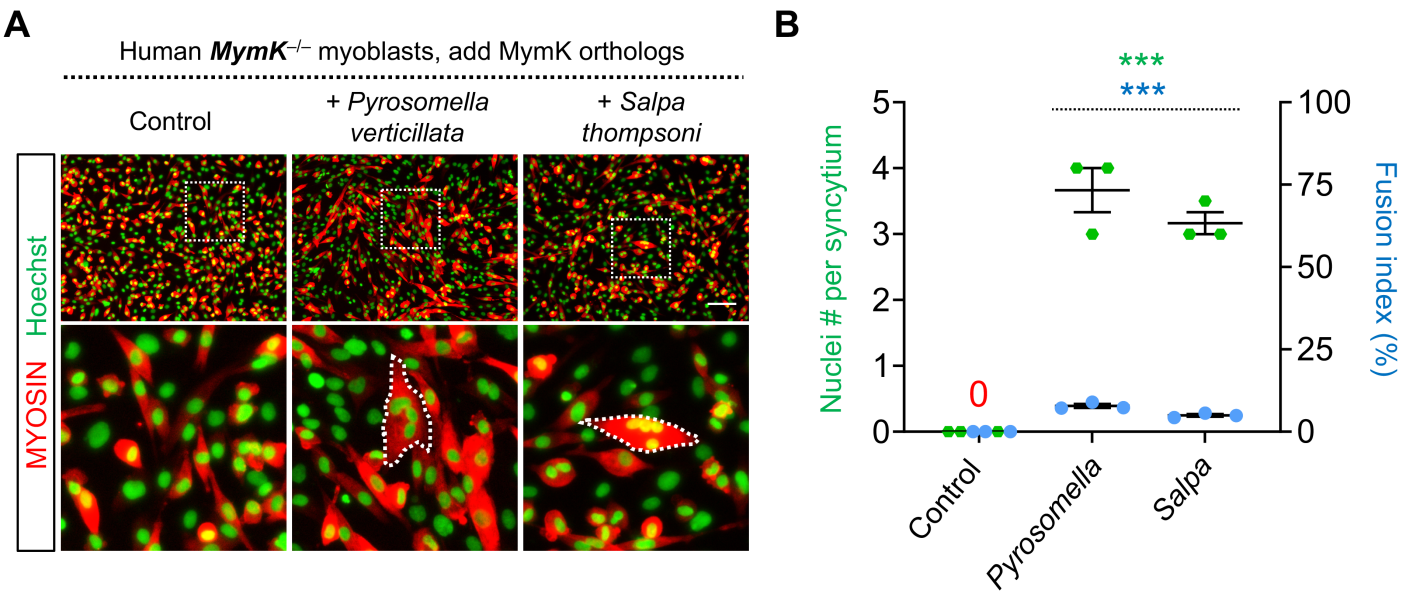


Figure S5. *MymK* proteins discovered from other tunicate species also induced myoblast fusion.

(**A**) Myosin immunostaining of human *MymK*$^{-/-}$ myoblasts transfected with control (empty vector) or expression vectors of MymK orthologs from thaliaceans. (**B**) Measurements of myoblast fusion in A after seven days of differentiation. Scale bars, 100 µm. Data are means ± SEM. *** $P < 0.001$, compared to control group, one-way ANOVA.

Figure S6

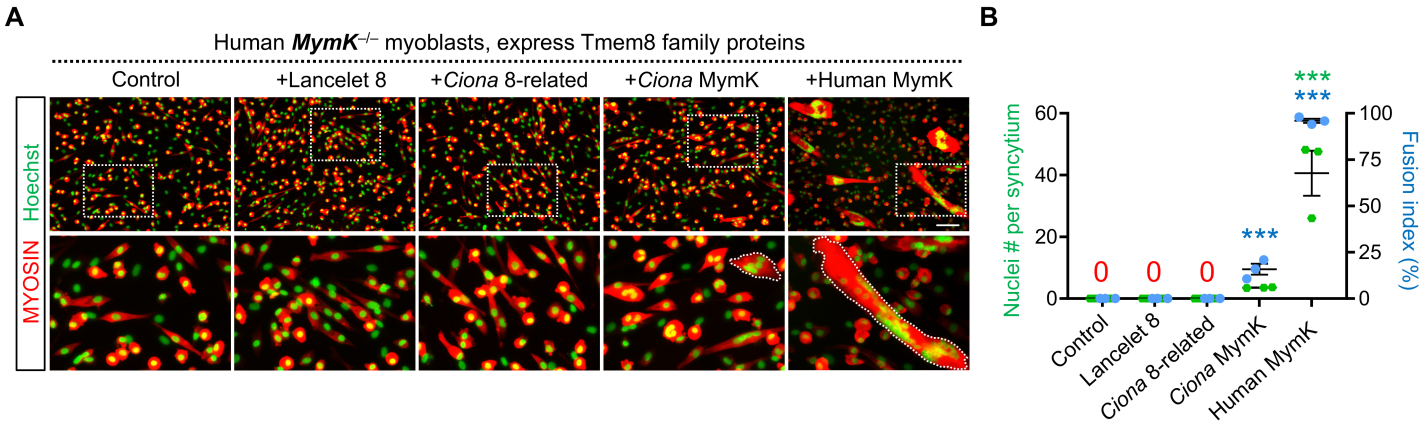


Figure S6. Expression of *Tmem8*-related genes does not induce myoblast fusion.

(**A**) Myosin immunostaining of human *MymK*$^{-/-}$ myoblasts transfected with control (empty vector) or expression vectors of Tmem8 family genes. Only *MymK* genes induce cell multinucleations (outlined). Lancelet 8: Tmem8 gene from *Branchiostoma floridae*; *Ciona* 8-related: Tmem8-related gene from *Ciona robusta*; *Ciona* MymK: Tmem8c gene from *Ciona robusta*. (**B**) Measurements of myoblast fusion in **A** after five days of differentiation. Scale bars, 100 µm. Data are means ± SEM. *** $P < 0.001$, compared to control group, one-way ANOVA.

Figure S7

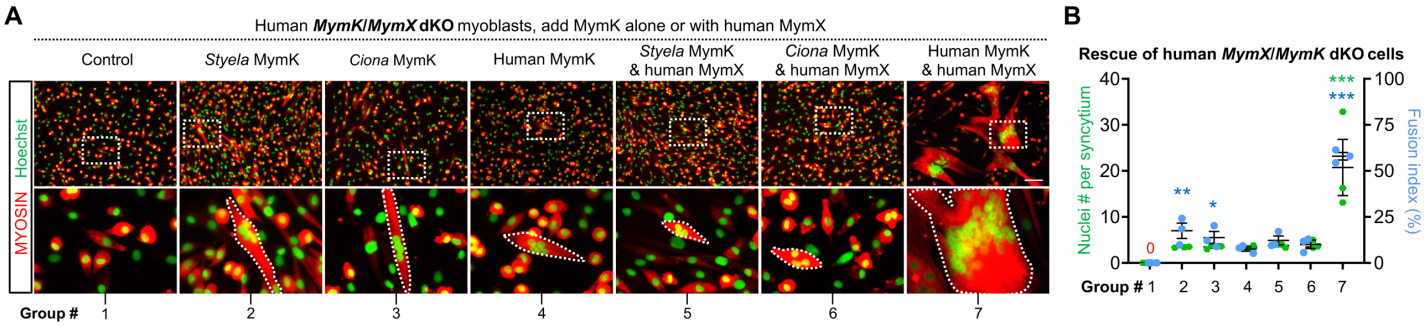


Figure S7. The fusogenic activity of tunicate MymK proteins is independent of MymX.

(**A**) Myosin immunostaining of human *MymK*/*MymX* dKO myoblasts transfected with MymK orthologs in presence or absence of human MymX re-expression. Muscle syncytia are outlined. Scale bar, 100 µm. (**B**) Measurements of myoblast fusion for the expression groups in **A** after differentiation. Data are means ± SEM. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$, compared to control group, one-way ANOVA.

Figure S8

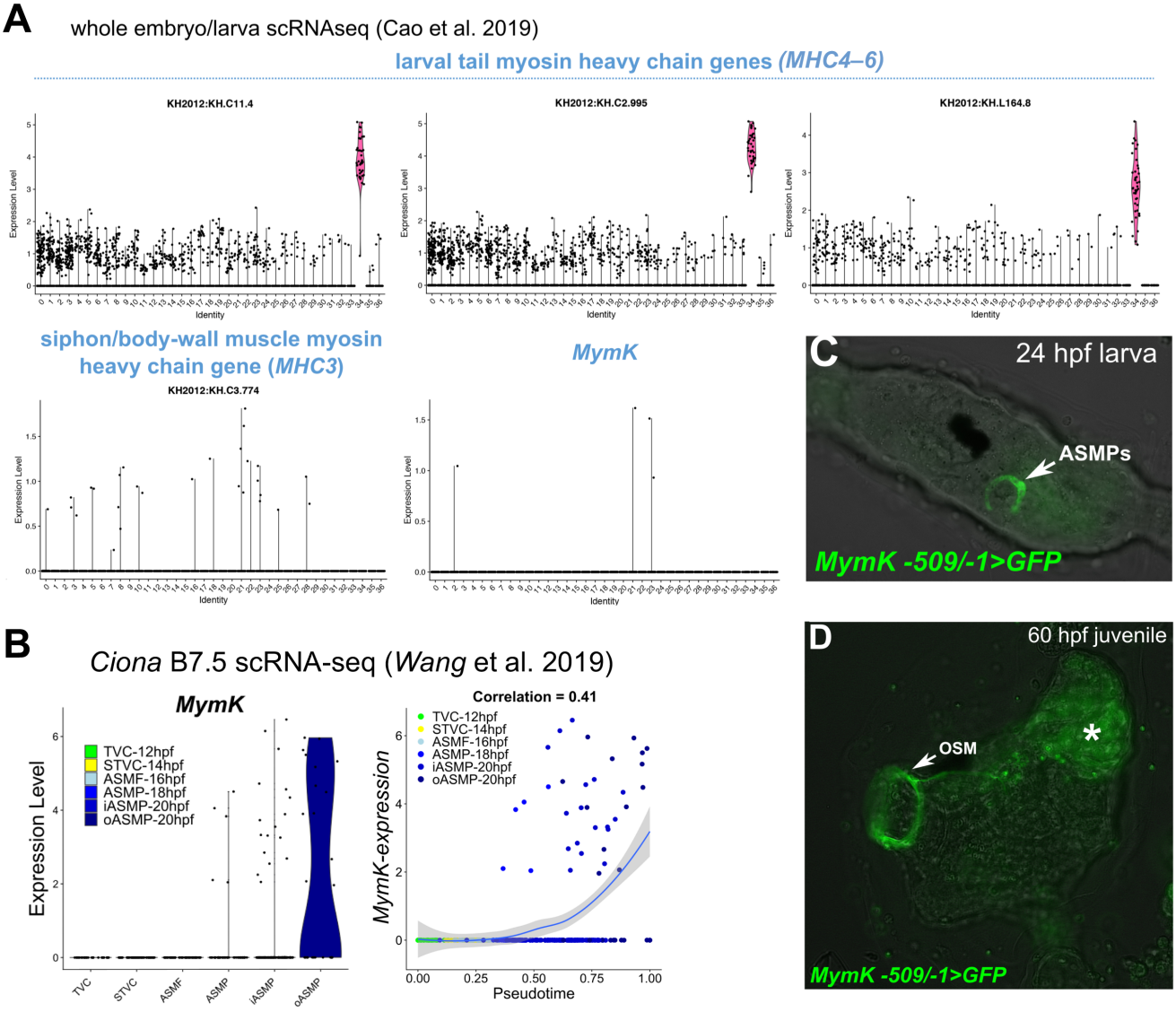


Figure S8. Mononucleated tail muscle cells from *Ciona robusta* larvae do not express *MymK*.

(**A**) Single-cell RNA sequencing results of *Ciona* larval samples (18 hpf, SRA accession: PRJNA542748) Top row: mononucleated tail muscle cells (cluster # 34) identified by the abundant expression of larval tail muscle-specific *Myosin heavy chain* paralogs (*MHC4–6*). Note that the expression of post-metamorphic siphon/body-wall muscle marker *MHC3* and *MymK* is not detected in the mononucleated tail muscle cells. (**B**) Violin plot comparing *MymK* expression levels in selected TVC derivatives (see Fig. 2E) at different developmental stages, showing initial infrequent expression starting in ASMFs, increasing in ASMPs at 18 hpf, and enriched primarily in oASMPs at 20 hpf. The pseudotemporal expression profile is shown on right. (**C**) Larva transfected with *MymK -509/-1>GFP* revealing MymK promoter activity in ASMPs right before settlement and metamorphosis. (**D**) Post-metamorphic juvenile transfected with *MymK -509/-1>GFP* revealing expression by oral siphon muscles (OSM) in addition to ASMs (see Fig. 2D). Asterisk indicates autofluorescent reabsorbed larval tail.

Figure S9

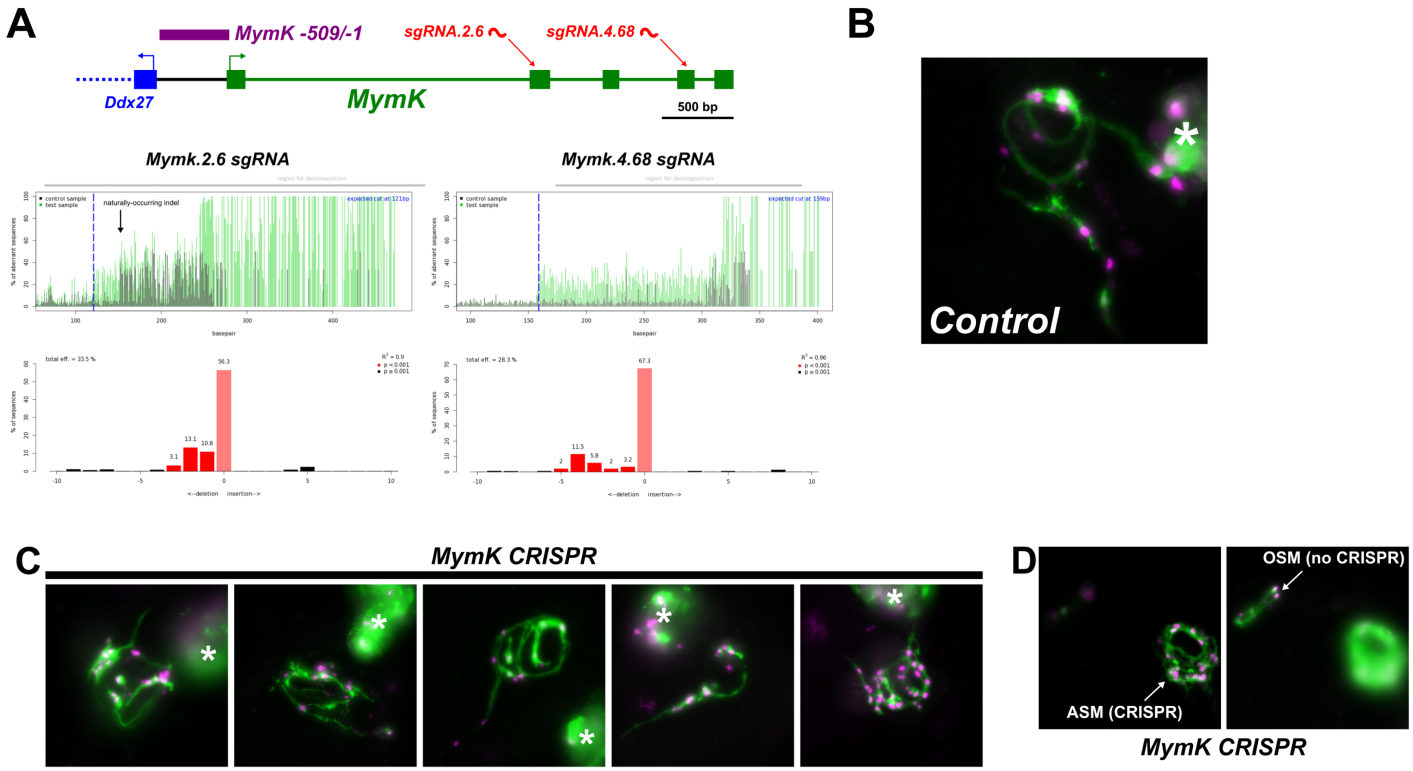


Figure S9. Knocking out *MymK* from *Ciona robusta* by CRISPR/Cas9.

(**A**) Diagram of *MymK* locus in *C. robusta,* showing location of the *MymK -509/-1* promoter fragment (in Fig. 2D) and target sites of single guide RNAs (sgRNAs) used for CRISPR mutagenesis. Lower panels show "peakshift" assay for single guide RNA (sgRNA) validation, performed by Sanger sequencing amplicons from larvae subjected to CRISPR/Cas9-mediated mutagenesis using individual sgRNA designed to target *MymK* in *Ciona robusta.* Images generated by TIDE (https://tide.nki.nl/). (**B**) Negative control juvenile (60 hpf, transfected with *Mesp>Cas9* alone) showing clear formation of circular myofibers. (**C**) Representative images of developing ASMs in juveniles (60 hpf) subjected to B7.5-specific mutagenesis of *MymK* by CRISPR/Cas9. In 10 of 19 *MymK* CRISPR juveniles, ASMs were disorganized and did not form enclosed rings of circular myofibers around the atrial siphon primordia. Asterisks indicate reabsorbed larval tail muscle cells. (**D**) Different focal plane views of a single *MymK* CRISPR juvenile (60 hpf) with both OSMs and ASMs labeled. Defects are only seen in the ASMs, which was expected due to lack of expression of Cas9 in the OSM lineage (*Mesp–*), which in turn serves as an internal negative control. Note that the oral siphon opening is perpendicular to the focal plane (not parallel to the focal plane, like the atrial siphon openings), which means the OSMs are imaged in profile. Muscle plasma membranes and nuclei in panels **B–D** are labeled by *MRF>CD4::GFP* and *MRF>H2B::mCherry,* respectively. All microscopy images were captured on a compound epifluorescence microscope. Images in **B**, **C** are projections of Z stacks, while images in **D** are single focal plane images.

Figure S10

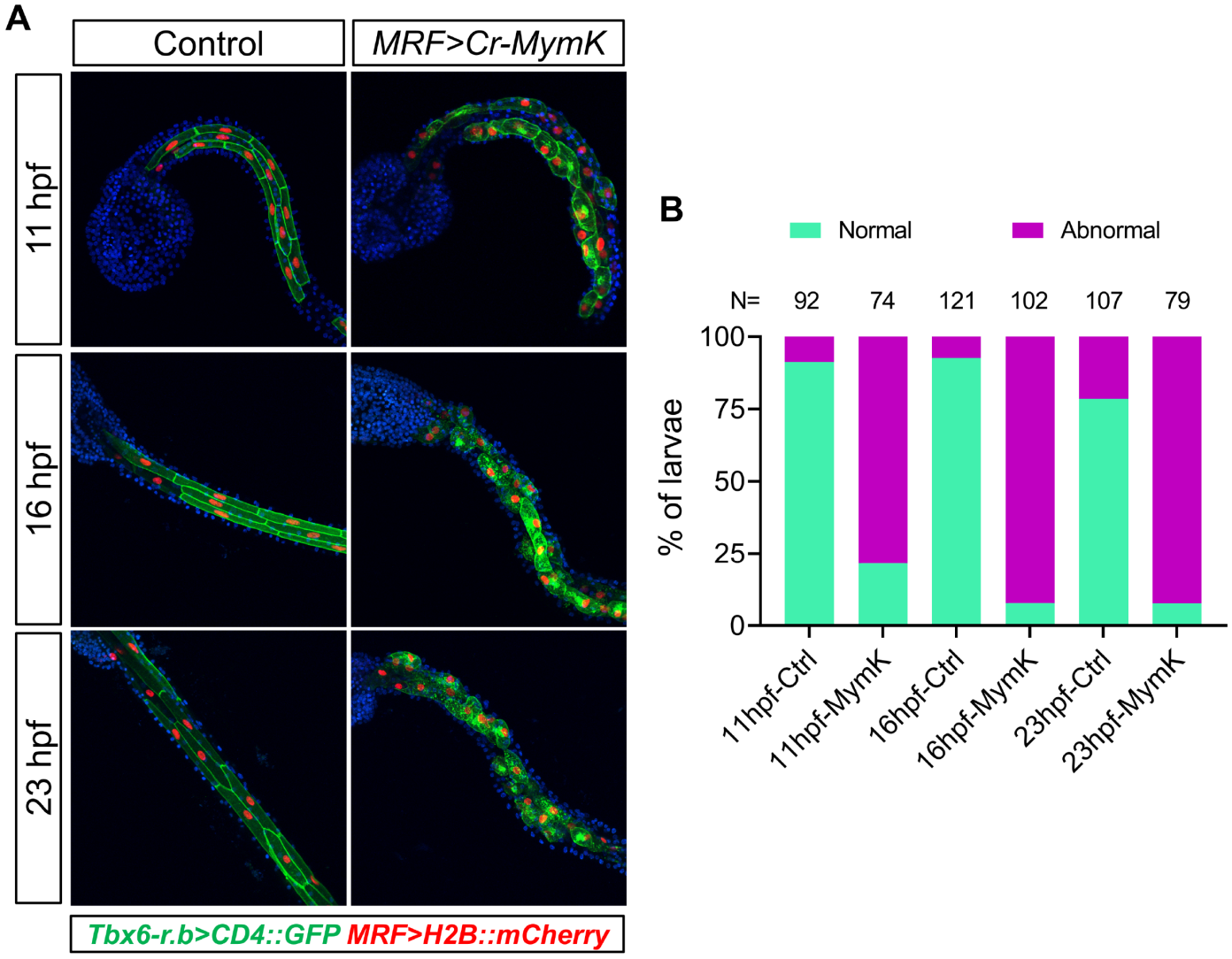


Figure S10. Overexpression of MymK in mononucleated tail cells perturbs morphology but does not induce cell fusion.

(**A**) Comparison of mononucleated tail muscle cell morphology between negative control embryos/larvae and those overexpressing *Ciona robusta* MymK driven by the MRF promoter (*MRF>Cr-MymK*). Cells lost their characteristic shape and cell-cell junctions, but did not show ectopic bi- or multi-nucleation. Cells were labeled with a combination of *Tbx6-r.b>CD4::GFP* (green) and *MRF>H2B::mCherry* (red) reporter plasmids and counterstained with DAPI (blue). All images are projections of Z stacks acquired on a scanning confocal microscope. (**B**) Scoring of "normal" vs. "abnormal" tail muscle morphology in embryos/larvae represented by images in panel **A**. Individuals were scored under confocal microscope as showing "abnormal" morphology if over 50% of labelled tail muscle cells show rounded or other irregular shape other than the regular polygonal shape observed in majority of negative control cells. N, numbers of larvae assayed for each condition.

Figure S11

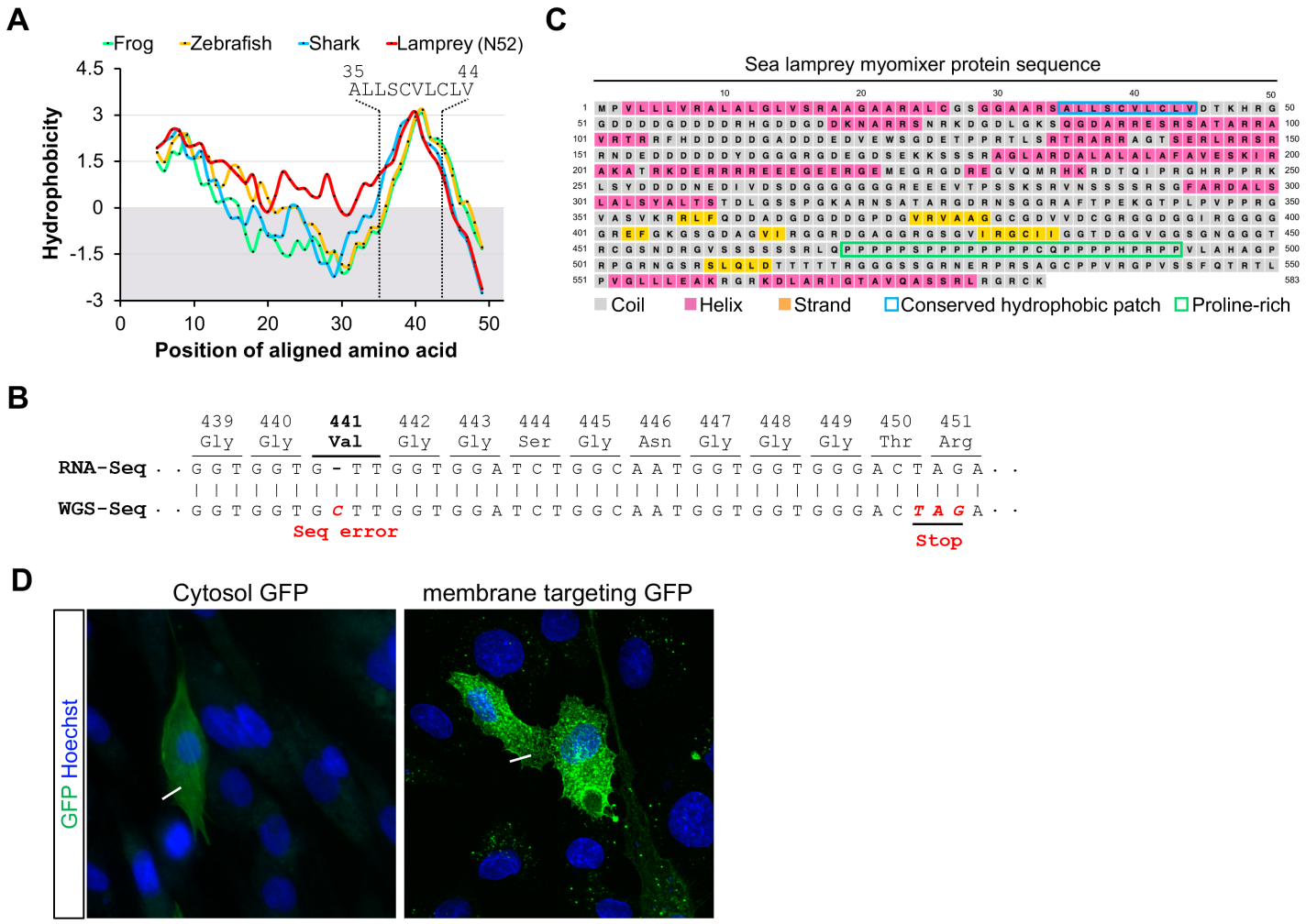
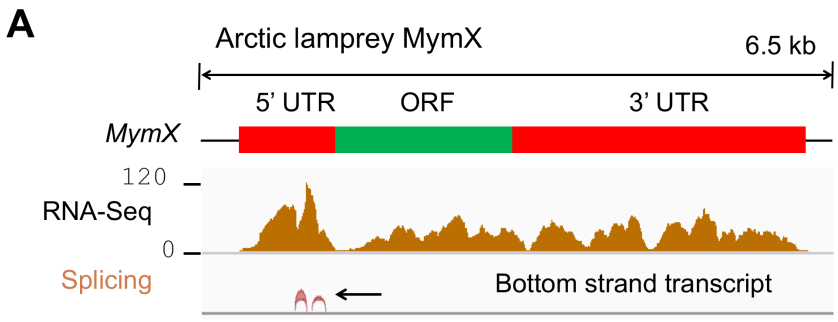


Figure S11. Analysis of lamprey MymX protein and coding sequence.

(**A**) Hydrophobicity signatures of MymX proteins for the aligned regions in Fig. 3B. The conserved hydrophobic patch (Ala35–Val44) from lamprey MymX is indicated. (**B**) Alignment of sequencing (seq) reads of RNA (SRA accession: PRJNA50489) with genome data (GenBank: AEFG01021847). A single-nucleotide gap of a cytosine insertion is highlighted (red). Cloning and sequencing of this region validated the correct reading frame shown in RNA-seq reads. WGS: whole genome shotgun. (**C**) Prediction of the secondary structure for full-length sea lamprey MymX (583 amino acids). (**D**) The fluorescence images of human myoblasts transfected with GFP (cytosol) and membrane targeting GFP (fused to N-terminus of myristoylated alanine rich c-kinase substrate, MARCKS) introduced as protein localization controls (see also Fig. 3E).

Figure S12



B

```
    Sea lamprey MymX MPVLLLVRALALGLVSRAAGAARALCGSGGAARSALLSCVLCLVDTKHRGGDDDD---GD 57
Arctic lamprey MymX ************************************************G***DGG** 60

    Sea lamprey MymX DDDRHGDDGDDKNA--RRSNRKDGDLGKSQGDARRESRSATARRAVRTRRFHDDDDDG-- 113
Arctic lamprey MymX ******V*******SR*******V********D**********************GDGA 120

    Sea lamprey MymX ADDDEDVEWSGDETPPRTLSRTRARRAGTSERLRRSRRNDEDD-DDDDYDGGGRGDEGDS 172
Arctic lamprey MymX D*****A***********************************D*******R*******C 180

    Sea lamprey MymX EKKSSSRAGLARDALALALAFAVESKIRAKATRKDERRRREEEGEER-GEMEGRGDREGV 231
Arctic lamprey MymX ************************************************G*********** 240

    Sea lamprey MymX QMRHKRDTQIPRGHRPPRKLSYDDD-DNEDIVDSDGGG-GGGGREEEVTPSSKSRVNSSS 289
Arctic lamprey MymX ******E*****************G***********G********************** 300

    Sea lamprey MymX SRSGFARDALSLALSYALTSTDLGSSPGKARNSATARGDRNSGGRAFTPEKGTPLPVPPR 349
Arctic lamprey MymX **********************E************************************* 360

    Sea lamprey MymX GVASVKRRLFQDDADGDGDDGPDGVRVAAGGCGDVVDCGRGGDGGIRGGGGGREFGKGSG 409
Arctic lamprey MymX ************G*******R*****************************S****R*** 420

    Sea lamprey MymX DAGVIRGGRDGAGGRGSGVIRGCIIGGTDGGVGGSGNGGGTRCGSNDRGVSSSSSSRLQP 469
Arctic lamprey MymX *V*****************T************A*******************-****** 479

    Sea lamprey MymX PPPPSPPPPPPPPCQPP-PPHPRPPVLAHAGPRPGRNGSRSLQLDTTTT-TRGGGSSGRN 527
Arctic lamprey MymX *****************P**********R**************F***I*T*TTK****** 539

    Sea lamprey MymX ERPRSAGCPPVRGPVSSFQTRTLPVGLLLEAKRGRKDLARIGTAVQASSRLRGRCK*     583
Arctic lamprey MymX ****************************T*************************     595
```

Figure S12. Discovery of MymX protein in arctic lamprey.

(**A**) RNA sequencing tracks that confirmed transcription of *MymX* genes in arctic lamprey (*Lethenteron camtschaticum*) (SRA accession: PRJNA371391). No splicing junction was detected in the hypothetical ORF. (**B**) Alignment of protein sequences encoded by *MymX* gene identified in sea lamprey and arctic lamprey. Identical amino acids are shown as asterisks. AxLyCxL motif is highlighted.

Figure S13

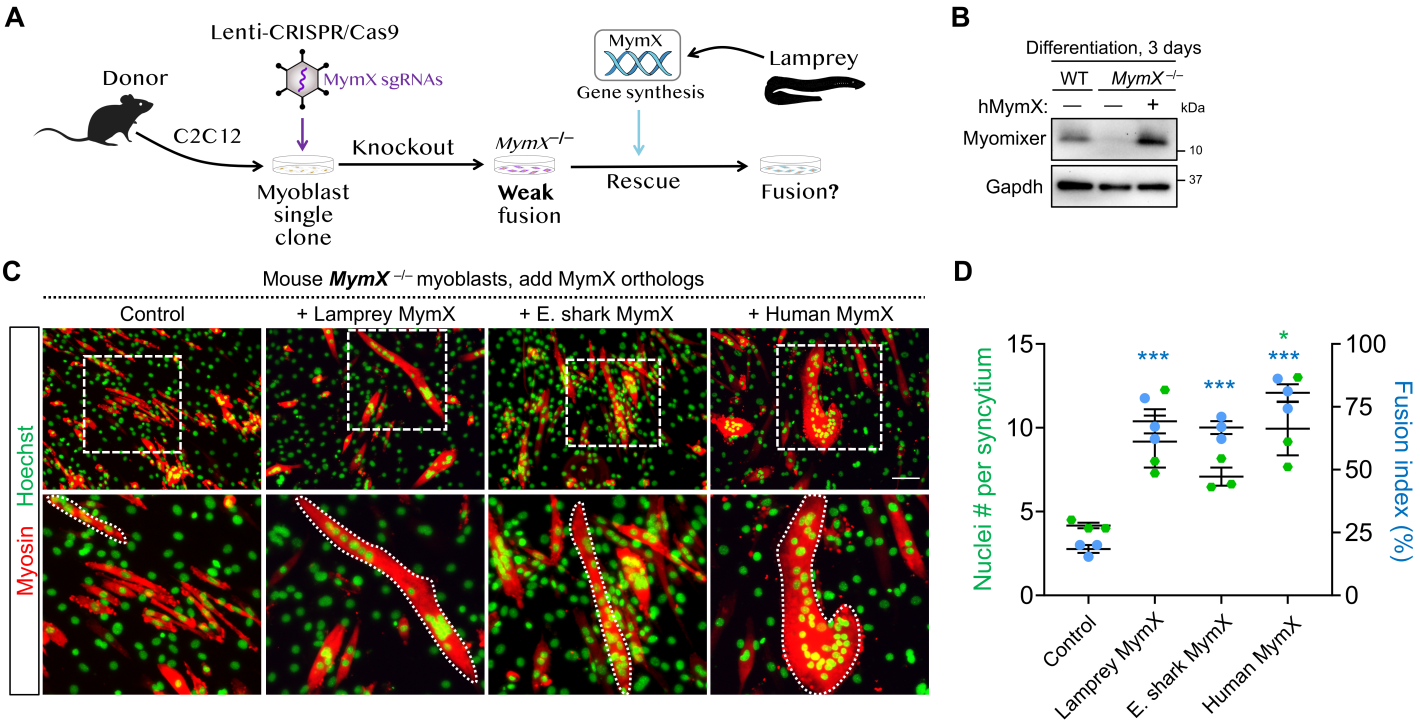


Figure S13. Lamprey MymX protein can rescue the fusion of mouse *MymX*$^{-/-}$ myoblasts.

(**A**) Schematic of experimental design to generate mouse *MymX*-deficient myoblasts by CRISPR/Cas9 and test the fusogenic function of lamprey MymX proteins. A pair of sgRNAs that target the single coding exon of MymX is applied. (**B**) Western blots validating the successful depletions of MymX protein in an isolated single clone of CRISPR treated mouse C2C12 myoblasts. (**C**) Myosin immunostaining of mouse *MymX*$^{-/-}$ myoblasts transfected with control (empty vector) or MymX ortholog expression vectors. Sea lamprey MymX protein can induce myotube formations (outlined). E. shark: elephant shark. Scale bar, 100 μm. (**D**) Measurements of myoblast fusion in **C** after five days of differentiation. Data are means ± SEM. * $P < 0.05$; *** $P < 0.001$, compared to control group, one-way ANOVA.

Figure S14

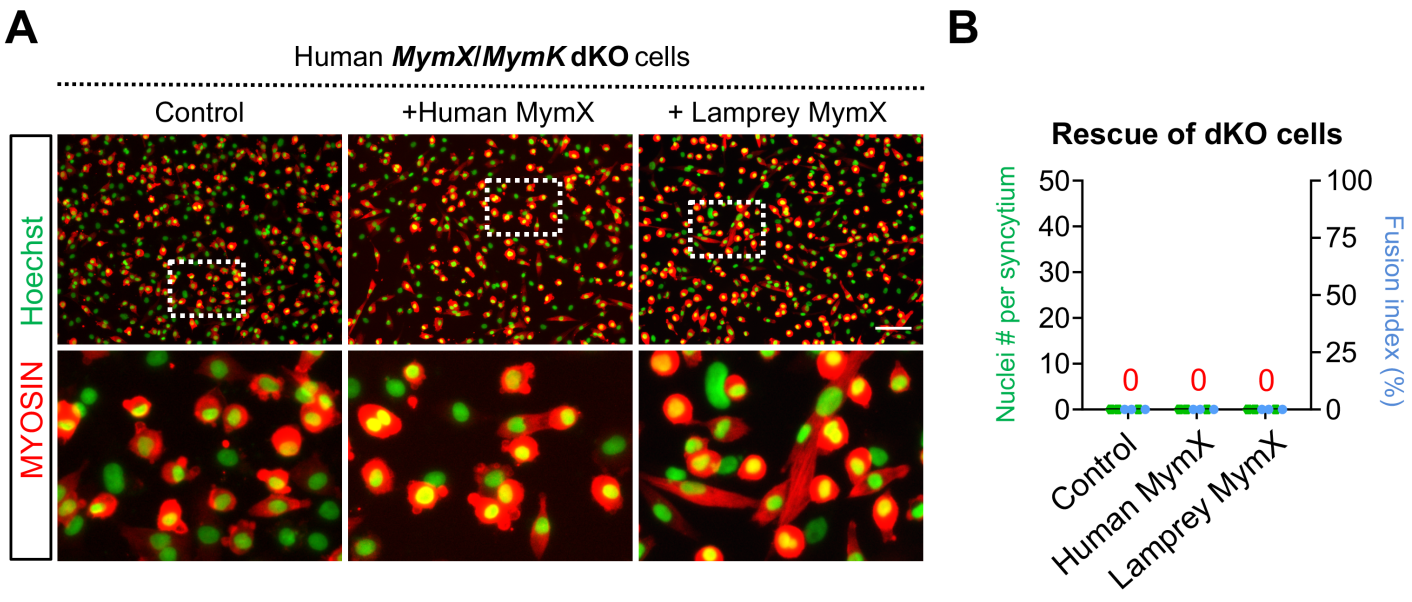


Figure S14. Lamprey MymX protein requires MymK for function.

(**A**) Myosin immunostaining of human *MymK*/*MymX* dKO myoblasts transfected with human or lamprey MymX expression vector. Scale bar, 100 µm. (**B**) Measurements of myoblast fusion in **A** after differentiation. Data are means ± SEM.

Figure. S15

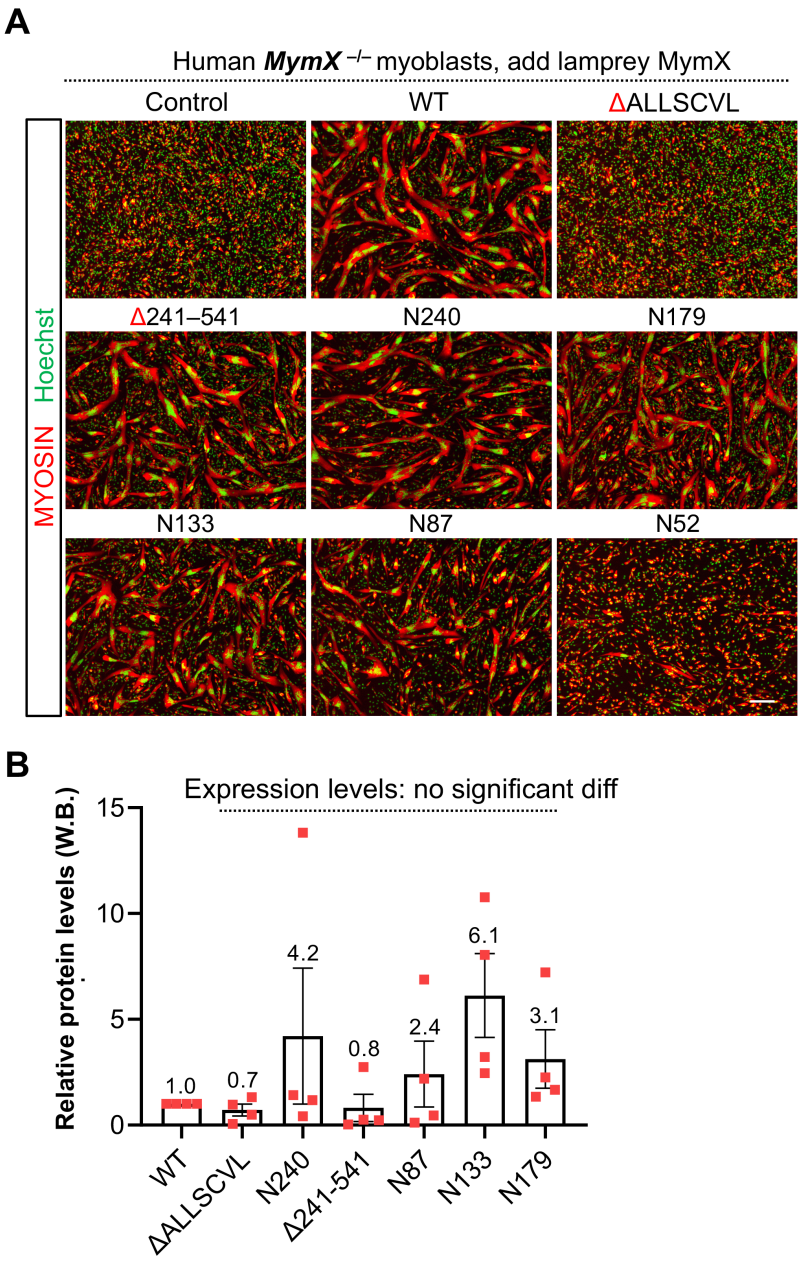


Figure S15. Structure and function analyses of lamprey MymX mutants.

(**A**) Macroscopic views of myosin immunostaining of human *MymX*$^{-/-}$ myoblasts transfected with MymX expression or control vector. Cells are differentiated for four days. Scale bars, 200 μm. (**B**) Measurement of lamprey MymX protein expression levels in human myoblasts by analyzing band intensity from Western blots. Numbers are the average fold-changes normalized to the expression level of wild-type (WT) lamprey MymX protein. Data are means ± SEM. n = 4, compared to full length lamprey MymX group (WT), one-way ANOVA.

Figure. S16

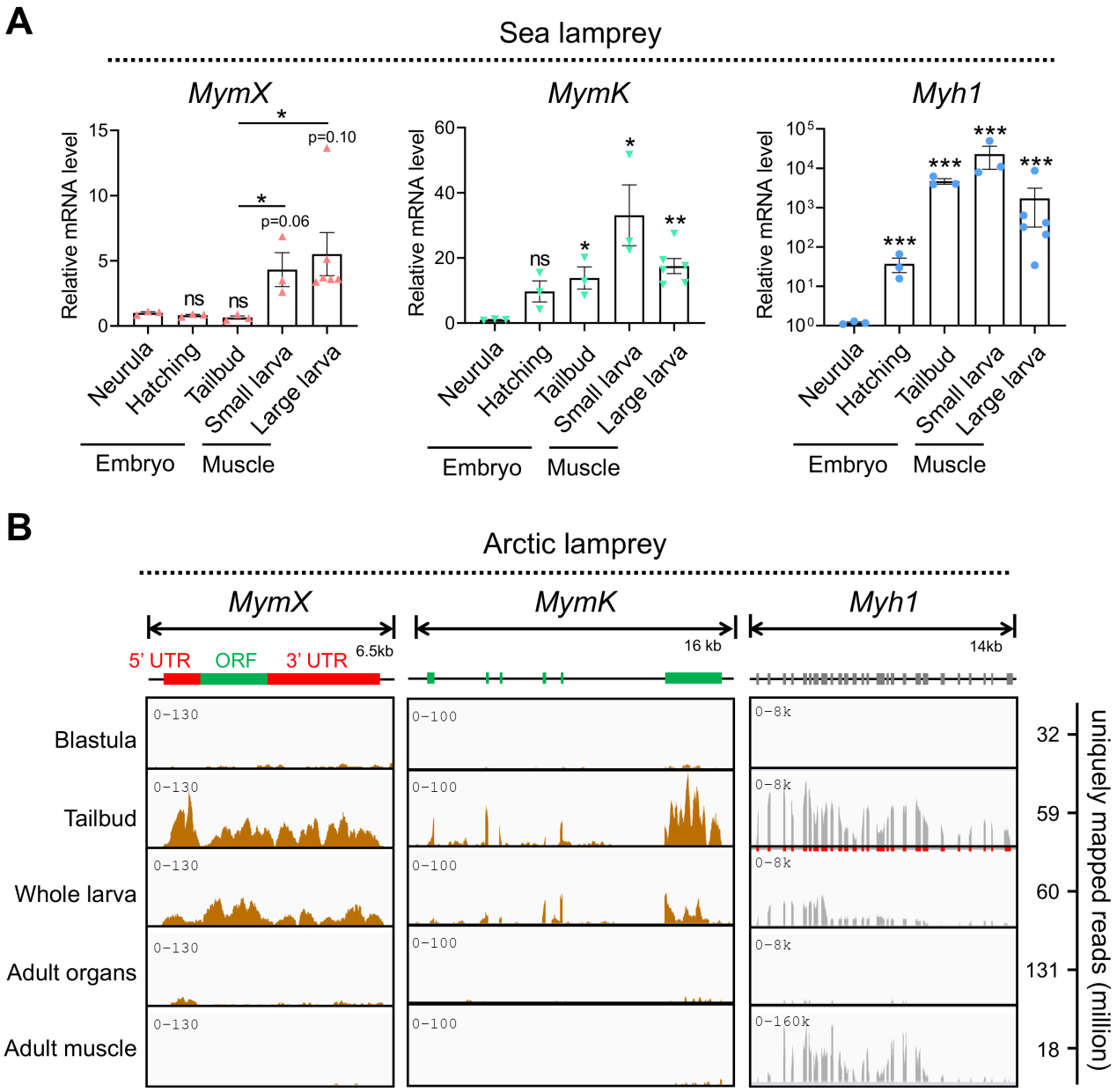


Figure S16. Expression pattern of *MymX* and *MymK* genes during lamprey development.

(**A**) qPCR results that measured the expression levels of sea lamprey *MymX*, MymK and myosin heavy chain 1 (*Myh1*) genes at various embryonic development and larval growth stages. Larva sizes refer to the grouping in Fig. 4D. (**B**) RNA-seq results of *MymX*, *MymK* and *Myh1* genes from arctic lamprey (*Lethenteron camtschaticum*) embryos (SRA accession: PRJNA371391), whole larva (17-day old, SRA accession: PRJNA553689), adult organs (notochord, ovary, testis, kidney and heart, SRA accession: PRJNA354821) and muscle tissues (SRA accession: PRJNA354821). Note that *MymX* gene in arctic lamprey is specifically expressed during muscle development and growing stages. Data are means ± SEM. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Figure S17

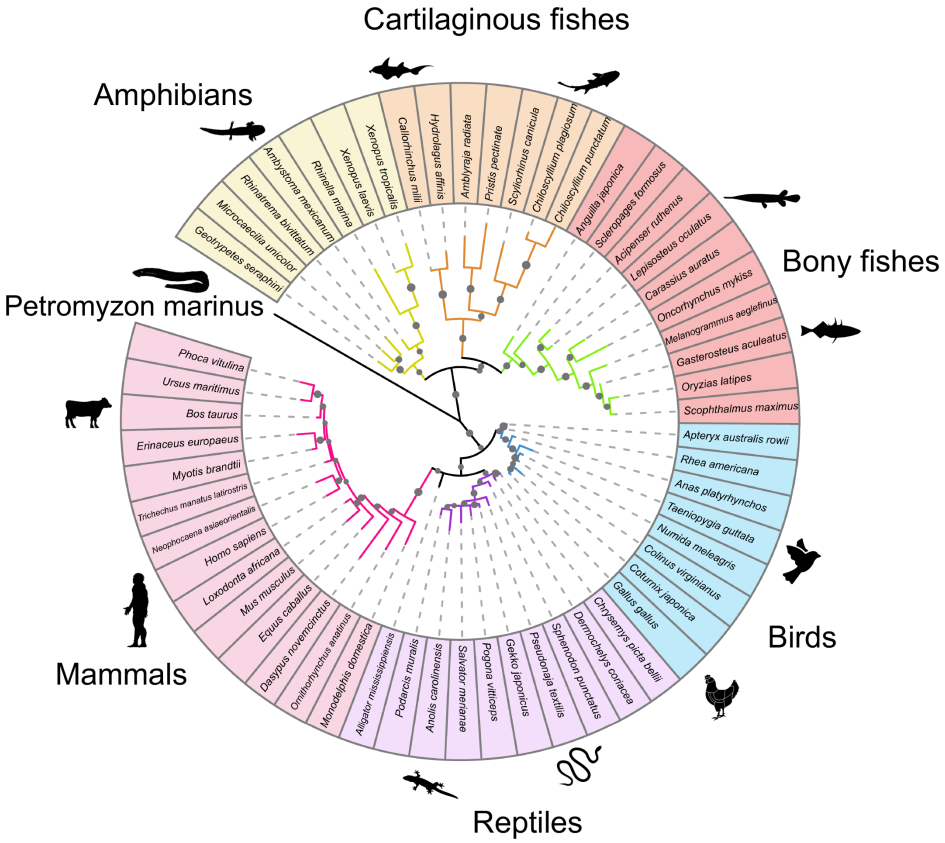


Figure S17. Phylogenetic tree of *MymX* orthologs.

Figure S18

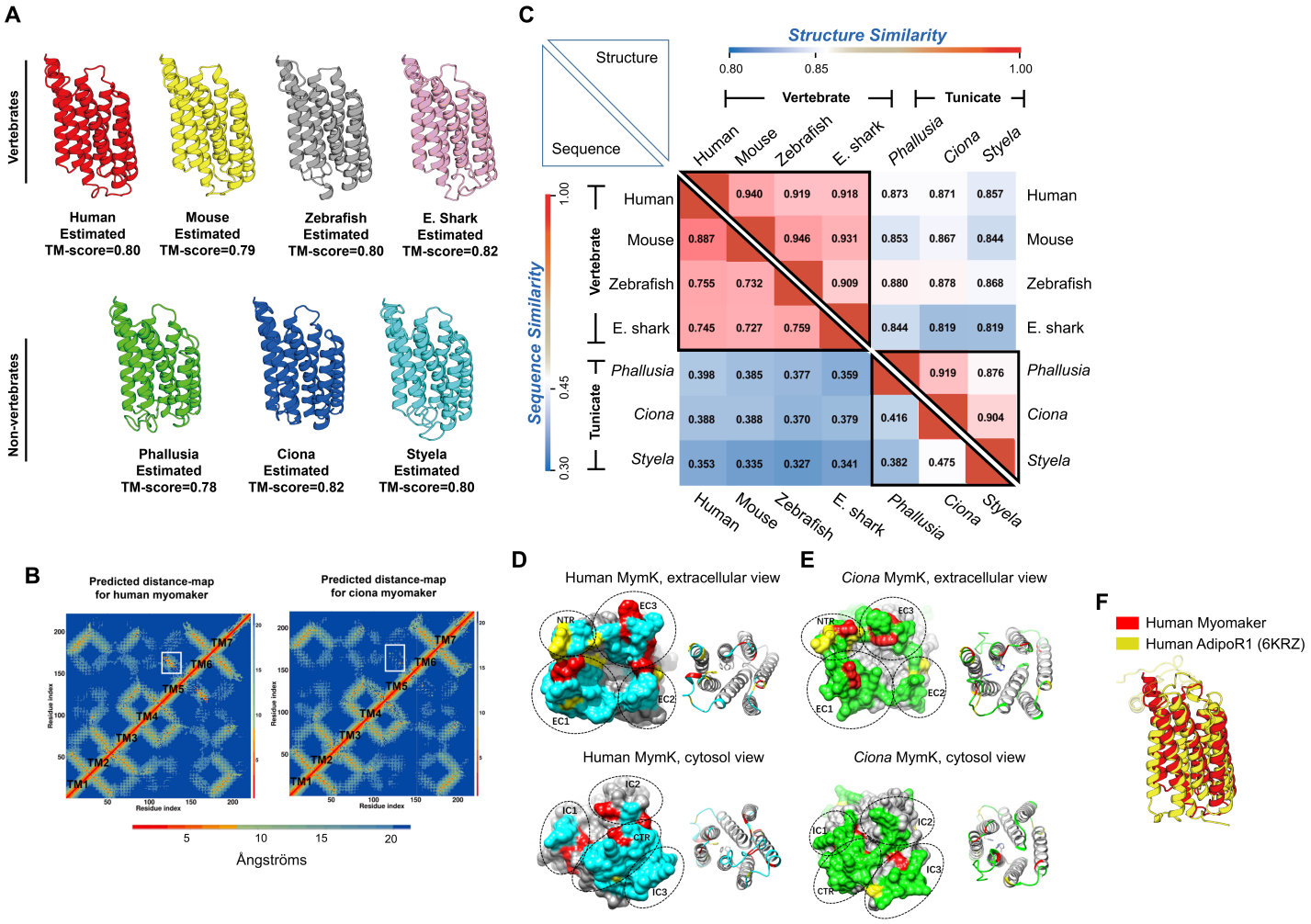


Figure S18. Structural modeling and comparison of vertebrate vs non-vertebrate MymK proteins.

(**A**) Protein structural models for MymK from vertebrates (top) and tunicates (bottom). TM-score: template modelling score. (**B**) The distance maps of human MymK protein (left) and Ciona MymK protein (right) that were derived from deep learning. (**C**) Similarities of amino-acid sequence (lower left triangle) and predicted structure (upper right triangle) between MymK proteins. (**D**) The extracellular (top) and intracellular (bottom) faces of human MymK protein model shown by surface (left) and cartoon (right) representations. Cyan highlights the surface residues; red highlights the conserved residues among vertebrates; yellow highlights the conserved residues among all species. (**E**) The extracellular (top) and intracellular (bottom) faces of *Ciona* MymK protein model shown by surface (left) and cartoon (right) representations. Green highlights the surface residues; red highlights the conserved residues among tunicates; yellow highlights the conserved residues among all species. EC: extracellular; IC: intracellular; NTR: N-terminus region; CTR:C-terminus region. (**F**) Superimposition of human MymK structure model and human adiponectin receptor 1 structure (AdipoR1, PDB ID: 6KRZ).

Supplementary Table 1. sgRNA and oligonucleotide sequences.

Purpose	Species	Gene	Sequence 5' → 3'	Amplicon (bp)	Application
CRISPR sgRNA	Human	MymK	sgRNA1: CTCACAGCTACAGAAGATGA	N/A	Fig. 1D
			sgRNA2: AAAGAAGAAGCGTAGCATCA		
	Human	MymX	sgRNA1: GGCTCCCAGGACATGCGAG		Fig. 3F
			sgRNA2: ACCTCTCCCTCCTCTCCAGG		
	Mouse	MymK	sgRNA1: TAGCGATGCTCACTGTCGGG		Fig. S4A
			sgRNA2: GCGTCCTTACCATCGCTGTG		
			sgRNA3: AGACAAACCAGGCCCATCAC		
	Mouse	MymX	sgRNA1: GCTGCTGCCTGTTGCCCGCC		Fig. S13A
			sgRNA2: GAGGCCTCTCCAGAATCCGG		
			sgRNA3: CCTCTGGGAGTGGTCCACTC		
	Lizard	MymK	sgRNA1: CACGCCAAACATCACCAAGG		Fig. S4A
			sgRNA2: TTTGACGGTGATCATGAGGA		
	Ciona	MymK	sgRNA2.6: gCAATGGAGTTGTGCAGAGG		Fig. 2H
			sgRNA4.68: gCAGCTCCAAGTGCAAACGA		
Genotype Analysis	Human	MymK	F: CTTCCTTCCCAGCCATCCAG R: GGGCTAGTGAGCAGGGACTA	492	Fig. 1D
	Human	MymX	F: AACTGAAGGGAGGGGGAACT R: TGGAGGACAGAGGGGCAATA	599	Fig. 3F
	Mouse	MymK	F: GCCTTTACCACCTTCTCCCC R: CCCACCTCACACCTTCCTTC	6,278	Fig. S4A
	Mouse	MymX	F: AGTTCAGGCTTCAGGTCAGAG R: GCTAGGGGAGTGGGAACTGT	743	Fig. S13A
	Lizard	MymK	F: CTTCATGTATGTTGGAAGGAGGCT R: GTTGAGTGTAGACGCTTTTCTCTG	395	Fig. S4A
	Lamprey	MymX	F: ATTGGAGGCACAGATGGTGG R: GGGGTTTCACGACAAGCAAC	498	Fig. 3e
	Ciona	MymK	F: CGCGATCACAAATGACGAAAC R: CCCGCAATTACAACATGCTAG	1400	Fig. S8
RT-PCR	Lamprey	18s	F:CGCTTTGGTGACTCTGGATAA R: GCCTGCCATCGTAAGTTGATA	100	Fig. 4H
	Lamprey	MymX	F: TTCTCTCCTGCGTGCTGTG R: GCGTTCTTGTCATCGCCATC	106	Fig. 4H
	Lamprey	MymK	F: AGAATTCGAGGAGCCGCAG R: AGGAACGCCCTCAGTTGGAT	174	Fig. 4H