A sex-specific transcription factor controls male identity in a simultaneous hermaphrodite

Tracy Chong[1], James J. Collins III[1], John L. Brubacher[2], David Zarkower[3], & Phillip A. Newmark[1, 4*]

[1]Department of Cell and Developmental Biology, University of Illinois at Urbana-Champaign, 601 South Goodwin Avenue, Urbana, Illinois 61801, USA.

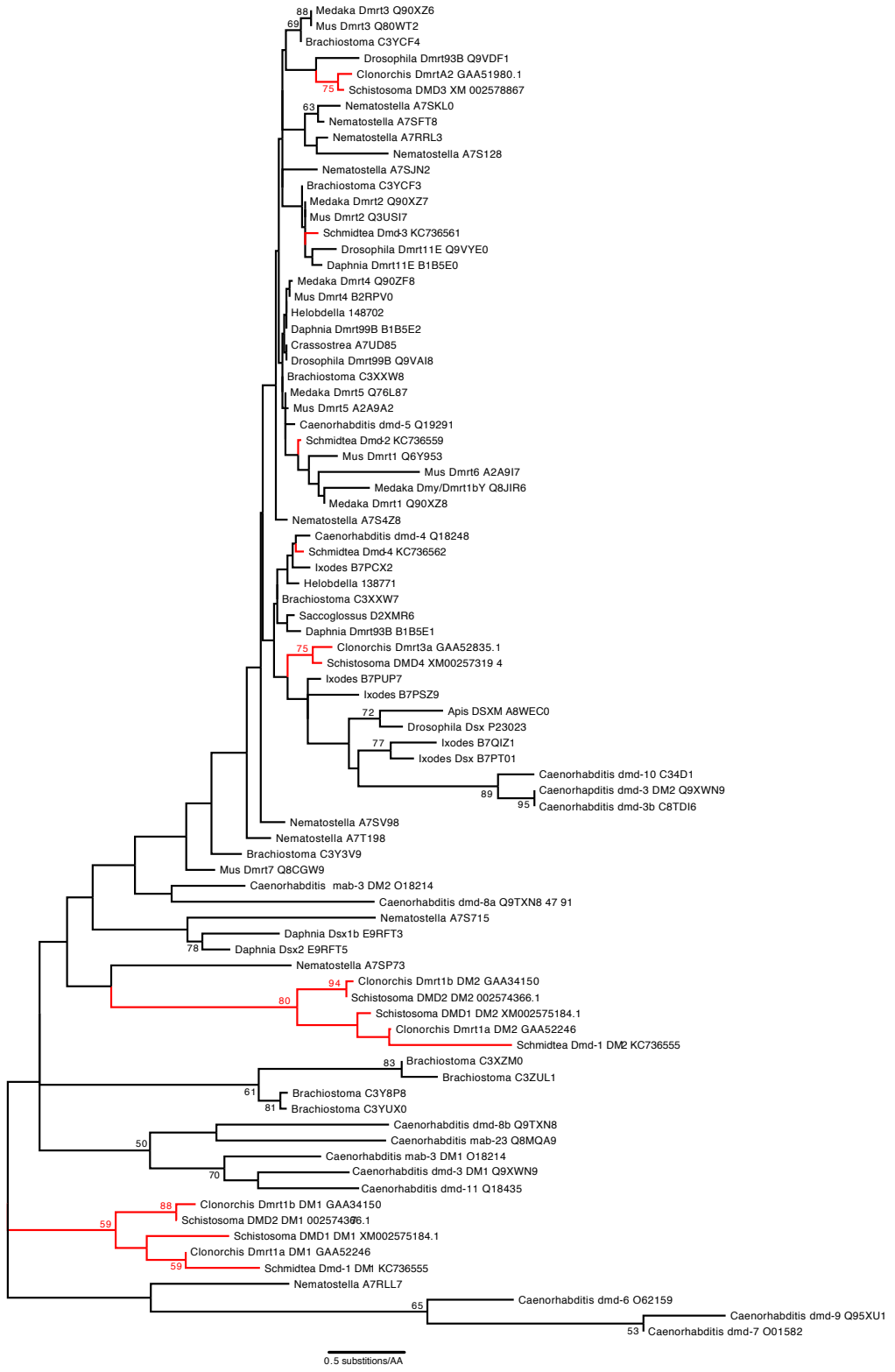
[2]Faculty of Humanities and Sciences, Canadian Mennonite University, 500 Shaftesbury Boulevard, Winnipeg, MB R3P 2N2, Canada.

[3]Department of Genetics, Cell Biology and Development, University of Minnesota Medical School, 6-160 Jackson Hall, 321 Church St. SE, Minneapolis, MN 55455, USA.

[4]Howard Hughes Medical Institute, University of Illinois at Urbana-Champaign, 601 South Goodwin Avenue, Urbana, Illinois 61801, USA.

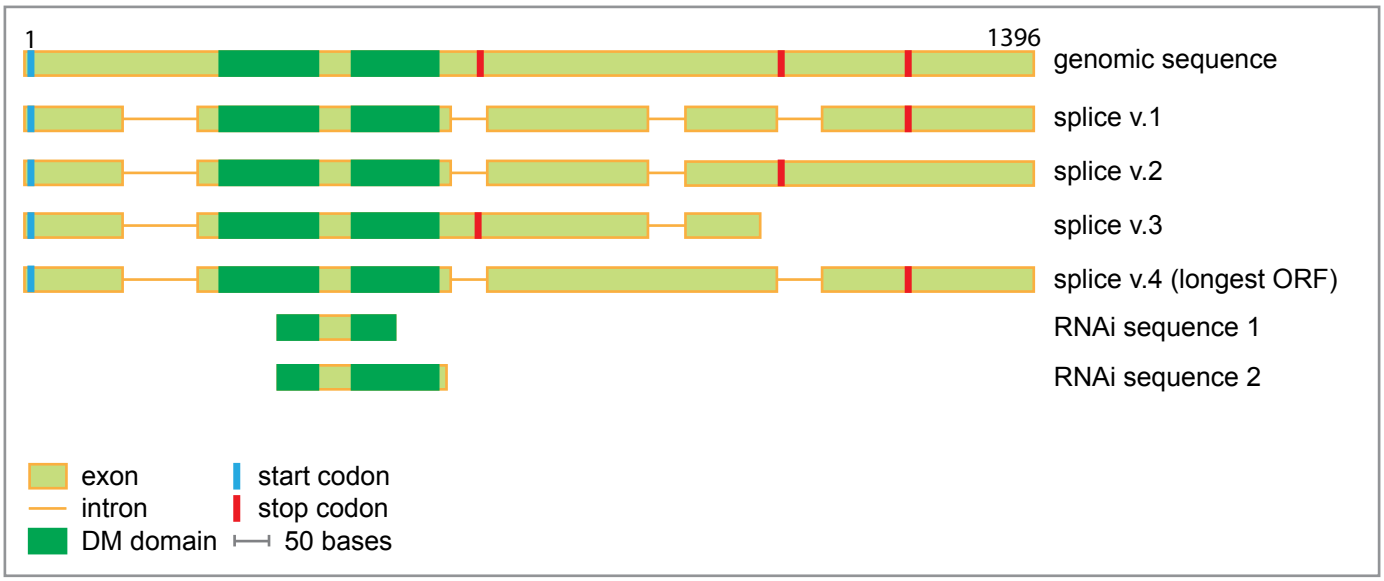
[*] P.A.N: pnewmark@life.illinois.edu

Supplementary Figure S1- Molecular phylogenetic analysis of DM domains by maximum likelihood method.



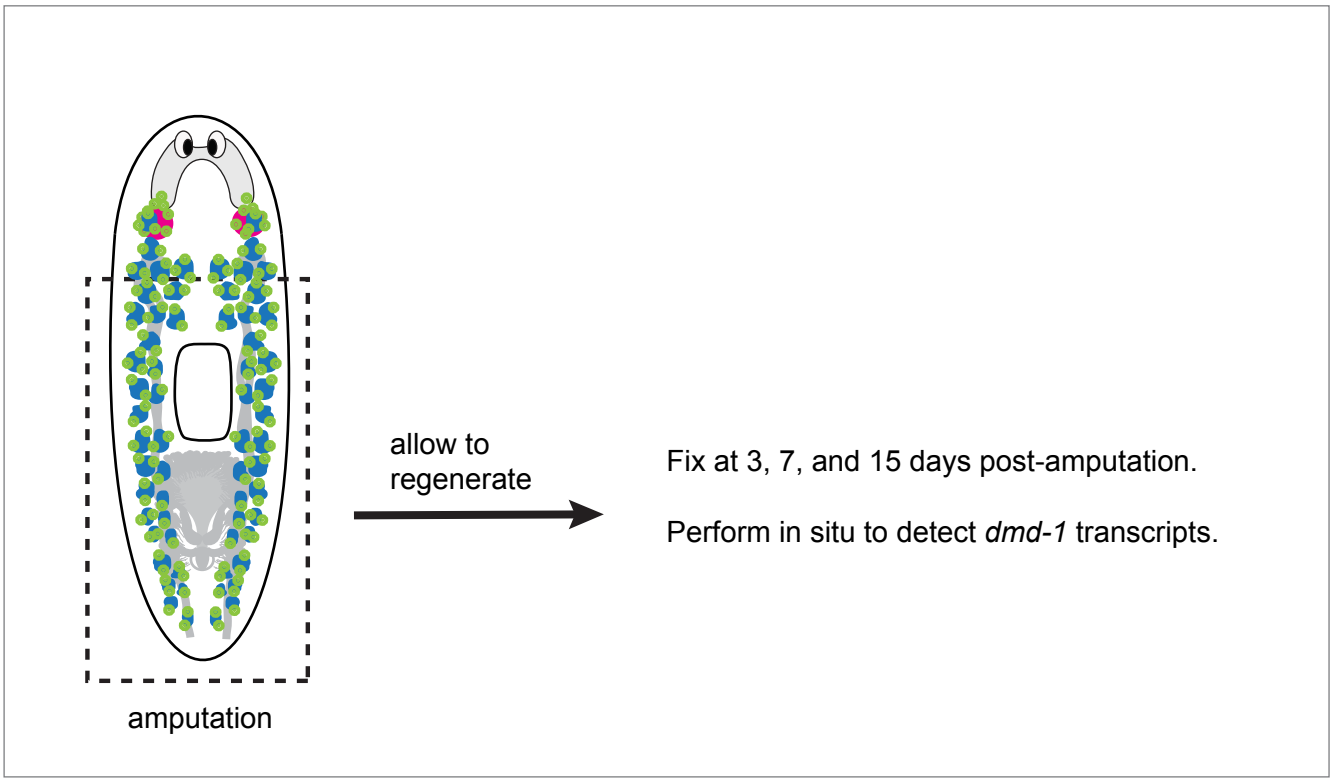
The phylogenetic tree with the highest log likelihood after 500 bootstrap replicates is shown. The tree is drawn to scale with branch lengths measured as the number of substitutions per amino acid (scale bar). The percentage of trees in which associated taxa clustered together in more than 50% of bootstrap replicates is shown next to branches. Branches involving genes from Platyhelminthes species are colored red. The following species were used for phylogenetic analysis and are designated in the figure by genus name. Lophotrochozoans: the planarian *Schmidtea mediterranea*; the flukes *Schistosoma mansoni* and *Clonorchis sinensis*; the leech *Helobdella robusta,* and the mollusk *Crassotrea gigas*. Ecdysozoans: the branchiopod crustacean *Daphnia magna*; the tick *Ixodes scapularis*; the bee *Apis mellifera*; the fly *Drosophila melanogaster*; and the nematode worm *Caenhorabditis elegans*. Other-metazoan taxa were the anemone *Nematostella vectensis* and four chordates: the cephalochordate *Branchiostoma floridae* (amphioxus); the hemichordate *Saccoglossus kowalevskii* (acorn worm); and the vertebrates *Mus musculus* (mouse) and *Oryzias latipes* (medaka fish). Pfam or GenBank accession numbers are indicated. *Helobdella* sequences are available at http://genome.jgi-psf.org/Helro1/Helro1.home.html. For proteins with two DM domains, the two domains are indicated as DM1 or DM2.

Supplementary Figure S2 - Splice variants of *dmd-1* and regions of the gene that were used for RNAi experiments.



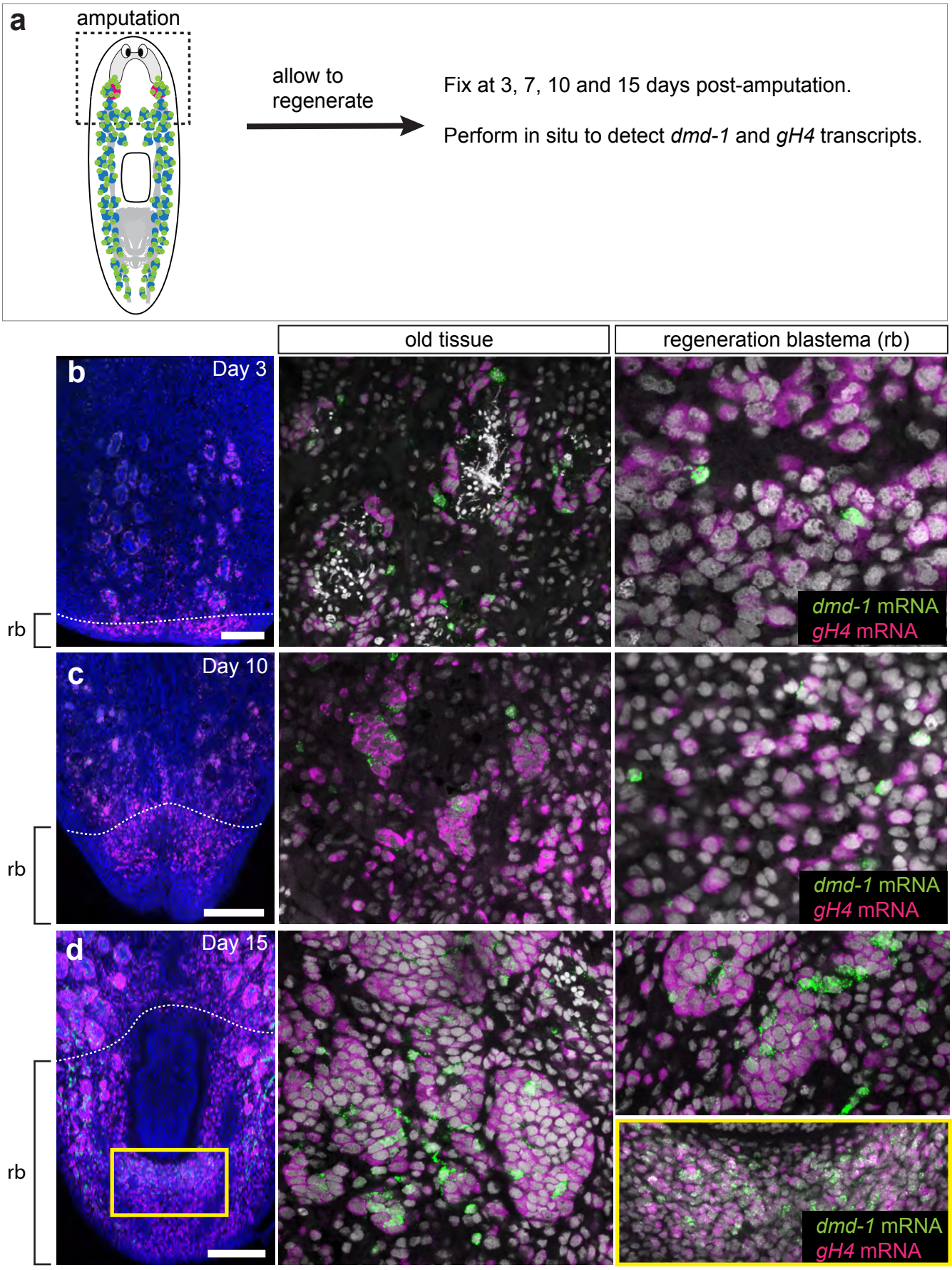
We have identified four splice variants of *dmd-1*, with the longest open reading frame encoding a predicted protein of 341 amino acids.

Supplementary Figure S3 - Experimental scheme for head regeneration.



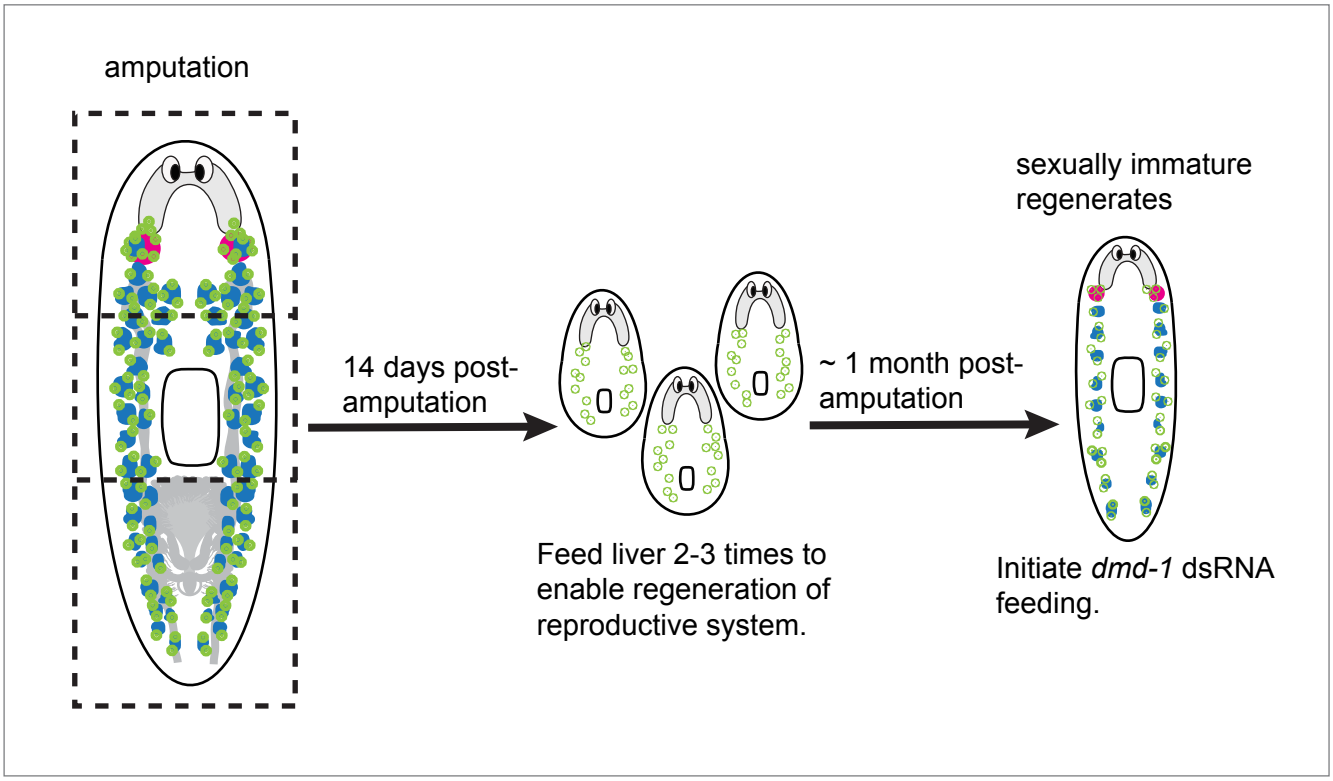
Animals were amputated between the ovaries and pharynx and allowed to regenerate new heads.

Supplementary Figure S4 - Expression of *dmd-1* in regenerating planarians.



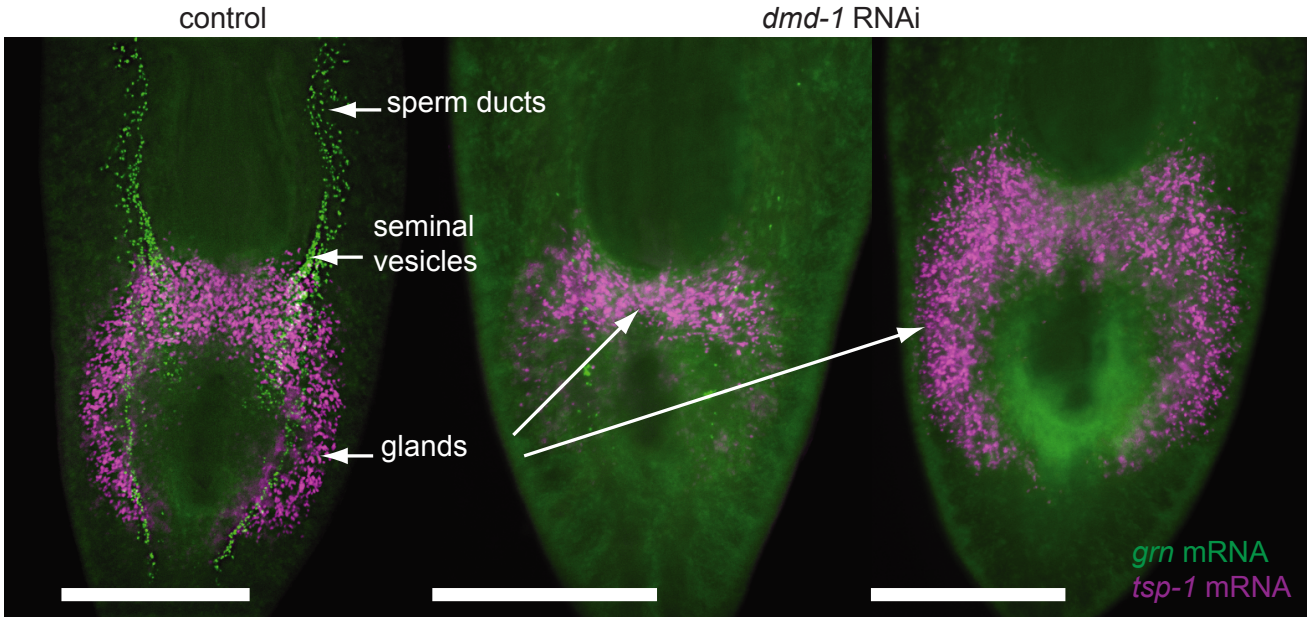
a) Animals were amputated between the ovaries and pharynx and allowed to regenerate missing posterior structures. **b-d)** Two-color FISH showing *dmd-1* (green) and *gH4* (magenta) mRNAs in the old tissue and regenerating blastema (rb) of anterior fragments at day 3 **(b)**, day 10 **(c)**, and day 15 **(d)** post-amputation. Yellow boxes in **(d)** show location of the developing copulatory apparatus. Dotted lines in the left panels show amputation plane. Scale bars, 0.2 mm.

Supplementary Figure S5 - Experimental scheme for generating sexually immature regenerates.



Sexually mature planarians were amputated. After 14 days, the animals were fed liver to enable regeneration of their reproductive system. About one month post-amputation, the sexually immature regenerating animals were used for RNAi experiments.

Supplementary Figure S6 - Female organs are present in *dmd-1*(RNAi) animals.



The sperm ducts visualized by FISH to detect *grn* [GB:DN304193.1] transcripts (green) are absent in *dmd-1*(RNAi) animals compared to controls. Meanwhile, the glands associated with egg laying (*tsp-1* RNA [DN305069.1], magenta) are present in *dmd-1*(RNAi) and control animals. Scale bars, 1 mm.

Supplementary Table S1 - Sequences of DM domain genes and dsRNA constructs.

dmd-1, splice form 1 [GenBank: KC736555], used to generate riboprobes for in situ hybridization

AGCATTAGAACATGAATTTAACTCAATGCCATTCGAAAATTAACGATTCTACCGGAAATTATTTGGAATTCAACCATTGCATTGTTGAAGATGATAAAAGTGAAAA
TAATGAAGACGTCAAACCGAATTTACTGAAAAAAGAATATTCTAGCAAATCGAGGAAAAAATCTCGTTCCTCATACATGTGTAGAAAATGTCGAGCCCATGGCCA
GTTGTTGCCGGTGAAAAACCATCGTCGAAATTGTCCGTTTAAATTGTGTCCCTGTCCATTGTGTACAGTCGTTAGTCATGGTCGAGAAATTGTCGCTCGTCAAAT
TAGACTTTATAGAAATCAGAAAGACATCAGCAAAAGGTCTCCGCAATCACCACATTGTAGACGATGCAGAAATCATGGAGAACACAATGCTTGGAAAGGCCACA
AAAAGAAATGTCGACATTCAAATTGTCCCTGTAAATTGTGCCTGTTAATTACCATGAGAAAAACCAATGAAAAAAGTTACAGAGAAATAGTTCAAGAAACTCACG
AAGAGAAATCAATTTTAGTAAACAGCAACAATAATAGCAATAATAATAATAACAATAGTTCCGCATTCTGTCTGTCAAACCGATCAAAAGACAGATCACTGGCAG
AAGATCAGTCAGTTTTTCTACGTAACATCAACCGAACTCCAATATTGCCAACGACAGGTTCGAATATCTACGACACCAGTCACAGCAGTTCACTTTACGCTCAAA
ACTCAATGCTAAACTTTTCGAATTCCGAAAACTGCAACATTCGGCCCTTGAATTGGTCAACAGCATCAGCCTATAAAGTTGGTTTACAGCCGTGGGACACTAACG
GATCGTTTGATATTCATAACAACGCCTCATATTGCGGCAGCTCACACCTGAACAACATGATGATTGGAGGTCATAACGTTTCCCAGCATAATTCCGTGGAATATC
TACCAACGGCGACTCCATCTTGGTATTCAAATTTTTATTGAGATTGACTGGAAATCTAACACGGATATAAATATAAATATATATTAATATTTACATTGAAAAGACCT
GTATTTATTCAATTGACAAATATCTACTGGTTATATCTANATATGTAACTGTTGTATGAATGAATATTTTCTTTGGG

dmd-1, splice form 2 [GenBank: KC736556]

AGCATTAGAACATGAATTTAACTCAATGCCATTCGAAAATTAACGATTCTACCGGAAATTATTTGGAATTCAACCATTGCATTGTTGAAGATGATAAAAGTGAAAA
TAATGAAGACGTCAAACCGAATTTACTGAAAAAAGAATATTCTAGCAAATCGAGGAAAAAATCTCGTTCCTCATACATGTGTAGAAAATGTCGAGCCCATGGCCA
GTTGTTGCCGGTGAAAAACCATCGTCGAAATTGTCCGTTTAAATTGTGTCCCTGTCCATTGTGTACAGTCGTTAGTCATGGTCGAGAAATTGTCGCTCGTCAAAT
TAGACTTTATAGAAATCAGAAAGACATCAGCAAAAGGTCTCCGCAATCACCACATTGTAGACGATGCAGAAATCATGGAGAACACAATGCTTGGAAAGGCCACA
AAAAGAAATGTCGACATTCAAATTGTCCCTGTAAATTGTGCCTGTTAATTACCATGAGAAAAACCAATGAAAAAAGTTACAGAGAAATAGTTCAAGAAACTCACG
AAGAGAAATCAATTTTAGTAAACAGCAACAATAATAGCAATAATAATAATAACAATAGTTCCGCATTCTGTCTGTCAAACCGATCAAAAGACAGATCACTGGCAG
AAGATCAGTCAGTTTTTCTACGTAACATCAACCGAACTCCAATATTGCCAACGACAGGTTCGAATATCTACGACACCAGTCACAGCAGTTCACTTTACGCTCAAA
ACTCAATGCTAAACTTTTCGAATTCCGAAAACTGCAACATTCGGCCCTTGAATTGGTCAACAGCATCAGCCTATAAAGTTGGTTTACAGCCGTGGGACACTAACG
GATCGTTTGATATTCATAACAGTTAATTAGTATTATATTAACCATTAGATTGAATTTAGTGTTTTCTATTTTGCCTCAATTAGACGCCTCATATTGCGGCAGCTCAC
ACCTGAACAACATGATGATTGGAGGTCATAACGTTTTCCAGCATAATTCCGTGGAATATCTACCAACGGCGACTCCATCTTGGTATTCAAATTTTTATTGAGATT
GACTGGAAATCTAACACGGATATAAATATAAATATATATTAATATTTACATTGAAAAGACCTGTATTTATTCAATTGACAAATATCTACTGGTTATATCTANATATGT
AACTGTTGTATGAATGAATATTTTCTTTGGG

dmd-1, splice form 3 [GenBank: KC736557]

AGCATTAGAACATGAATTTAACTCAATGCCATTCGAAAATTAACGATTCTACCGGAAATTATTTGGAATTCAACCATTGCATTGTTGAAGATGATAAAAGTGAAAA
TAATGAAGACGTCAAACCGAATTTACTGAAAAAAGAATATTCTAGCAAATCGAGGAAAAAATCTCGTTCCTCATACATGTGTAGAAAATGTCGAGCCCATGGCCA
GTTGTTGCCGGTGAAAAACCATCGTCGAAATTGTCCGTTTAAATTGTGTCCCTGTCCATTGTGTACAGTCGTTAGTCATGGTCGAGAAATTGTCGCTCGTCAAAT
TAGACTTTATAGAAATCAGAAAGACATCAGCAAAAGGTCTCCGCAATCACCACATTGTAGACGATGCAGAAATCATGGAGAACACAATGCTTGGAAAGGCCACA
AAAAGAAATGTCGACATTCAAATTGTCCCTGTAAATTGTGCCTGTTAATTACCATGAGAAAAACCAATGAAAAAAGTTACAGTAACGAGAACCATTATTTTTTTTA
TTGTATAGTTTTACTGTAAGATCCAGGAGAAATAGTTCAAGAAACTCACGAAGAGAAATCAATTTTAGGAAACAGCAACAATAATAGCAATAATAATAATAACAAT
AGTTCCGCATTCTGTCTGTCAAACCGATCAAAAGACAGATCACTGGCAGAAGATCAGTCAGTTTTTCTACGTAACATCAACCGAACTCCAATATTGCCAACGACA
GGTTCGAATATCTACGACACCAGTCACAGCAGTTCACTTTACGCTCAAAACTCAATGNTAAACTTTTCGAATTCCGAAAACTGCAACATTCGGCCCTTGAATTGG
TCAACAGCATCAGCCTATAAAG

dmd-1, splice form 4 [GenBank: KC736558]

AGCATTAGAACATGAATTTAACTCAATGCCATTCGAAAATTAACGATTCTACCGGAAATTATTTGGAATTCAACCATTGCATTGTTGAAGATGATAAAAGTGAAAA
TAATGAAGACGTCAAACCGAATTTACTGAAAAAAGAATATTCTAGCAAATCGAGGAAAAAATCTCGTTCCTCATACATGTGTAGAAAATGTCGAGCCCATGGCCA
GTTGTTGCCGGTGAAAAACCATCGTCGAAATTGTCCGTTTAAATTGTGTCCCTGTCCATTGTGTACAGTCGTTAGTCATGGTCGAGAAATTGTCGCTCGTCAAAT
TAGACTTTATAGAAATCAGAAAGACATCAGCAAAAGGTCTCCGCAATCACCACATTGTAGACGATGCAGAAATCATGGAGAACACAATGCTTGGAAAGGCCACA
AAAAGAAATGTCGACATTCAAATTGTCCCTGTAAATTGTGCCTGTTAATTACCATGAGAAAAACCAATGAAAAAAGTTACAGAGAAATAGTTCAAGAAACTCACG
AAGAGAAATCAATTTTAGTAAACAGCAACAATAATAGCAATAATAATAATAACAATAGTTCCGCATTCTGTCTGTCAAACCGATCAAAAGACAGATCACTGGCAG
AAGATCAGTCAGTTTTTCTACGTAACATCAACCGAACTCCAATATTGCCAACGACAGGTTCGAATATCTACGACACCAGTCACAGCAGTTCACTTTACGGTTATT
TCCAGTTGATTGTTTGGGCCTTGGTTTTAATTGCCTTTCTTTTAGCTCAAAACTCAATGCTAAACTTTTCGAATTCCGAAAACTGCAACATTCGGCCCTTGAATTG
GTCAACAGCATCAGCCTATAAAGTTGGTTTACAGCCGTGGGACACTAACGGATCGTTTGATATTCATAACAACGCCTCATATTGCGGCAGCTCACACCTGAACA
ACATGATGATTGGAGGTCATAACGTTTTCCAGCATAATTCCGTGGAATATCTACCAACGGCGACTCCATCTTGGTATTCAAATTTTTATTGAGATTGACTGGAAAT
CTAACACGGATATAAATATAAATATATATTAATATTTACATTGAAAAGACCTGTATTTATTCAATTGACAAATATCTACTGGTTATATCTATATATGTAACTGTTGTA
TGAATGAATATTTTCTTTGGG

dmd-2, splice form 1 [GenBank: KC736559]

ATGGAAAATTCTAATAGCTTTAATAATACAAAAATTAATTTAAGTAGTGTGGAATCTATTATAGATAATGAACCAAAAGATTTTAAATCAAATCTTAAATGTTTG
CAGCCTAGGCAAATGTTTTGTCGAAATTTTGAAAAATACCAACGAACTCCTAAATGTGCACGTTGCCGCAATCATGGATATGTTTCACCTTTAAAAGGTCACAAG
CGATACTGTAGATGGAAAGATTGTGTCTGTGCAAAATGTGCTCTAATAGCGGAACGACAAAGAGTGATGGCAGCTCAGGTTGCATTGCGGAGACAACAAGCGCA
AGAGGAGCAGGAAATTAATGATTTTGTCATGTATCAAAATGAAGAAATTGTTTACCCGCCAGATTCCAAAAAAAAAA

dmd-2, splice form 2 [GenBank: KC736560]

ATGGAAAATTCTAATAGCTTTAATAATACAAAAATTAATTTAAGTAGTGTGGAATCTATTATAGATAATGAACCAAAAGATTTTAAATCAAATCTTAAATGTTTG
CAGCCTAGGCAAATGTTTTGTCGAAATTTTGAAAAATACCAACGAACTCCTAAATGTGCACGTTGCCGCAATCATGGATATGTTTCACCTTTAAAAGGTCACAAG
CGATACTGTAGATGGAAAGTATAAATCATTTTGCATTTTCTTGACTAATTATAAATTTTAGGATTGTGTCTGTGCAAAATGTGCTCTAATAGCGGAACGACAAAG
AGTGATGGCAGCTCAGGTTGCATTGCGGAGACAACAAGCGCAAGAGGAGCAGGAAATTAATGATTTTGTCATGTATCAAAATGAAGAAATTGTTTACCCGCCAG
ATTCCAAAAAAAA

dmd-3 [GenBank: KC736561]

ATGAAATTATACAATGTCTCGAGAACACCAAAATGCGCCAGATGTCGAAACCACGGTGTTGTATCTTGTTTAAAAGGGCATAAGCGTTTTTGTCACTGGAGAGAT
TGTCACTGTAGTAACTGTTTACTGGTAATCGAAAGACAAAGAGTTATGGCTGCCCAAGTTGCTCTGAGGCGTCAGCAGACTACAAGCGCTTGTAAACTGCCAACG
AACACCAAGACGGAAAAGAGTTCGTATAAACGGCTTATCAGACAAAGCATATCGACAGTGAAATCCAAAAAATCCCATACCATATTTAGGAAGCAATCTGAAA
AAAAAGAATCGAATATGAAGAACAATTCATCTGGGTCGGTCGAAAGCATTTTAGAACGGTTACCCGTTGATCTTGGAACTCAATTTACTGAACGTCTTCGGAAAC
GAAAAGCATTTGCTGATGAAAATCTTGATTCAATTTCCAGTTTAAACAATCAACTAACTTCGTTAATATATTCAACGAAAGTTGAAAATTCCAATTATTTAATATA
TAATTTAAACAAAAATGACAATTTCAAAGAGGTGTATCACAATATGGTCAGTGAACAATTAAAGTTCTGGATGGAACAGCTCAAAAAGAATTTAGGAGTATTTG
GAAGTGGAAAATCTGAAGAAATTTCCTTGTATCAAAATTATTTTAGTAGACTCTATCCAGAAATCACAAATTACAATTCAAAAAACCAAAATAAAGTTATAAACA
GTGAAAAATCGTTCTTTATTGAGCGTCTTTTGAATTGAAACTTTAAAAATAACAAATTCAAATATCCAAATTTATTTATTTAATAGAGTTCTGCATATGTTAAAAA
AAAAAAA

dmd-4 [GenBank: KC736562]

ATGGACAAAGAATCCAAAAAGTGTCAAAAATCTTTGTCTGATCAAACTCCAGAATTTTCCAGAAATTTCTTAAAATTAATGCAGCCCCCTCTTTATCCGATTACA
GAAAAGGGTGCTAGAAAACCGAAATGTGCAAGATGTAGAAATCACGGAATGGTATCGTGGTTAAAAGGCCACAAAAGACATTGTCATTACAAGGATTGCCGTT
GTGCGAAATGCAATCTTATAGCTGAAAGACAACGAGTTATGGCAGCTCAAGTAGCATTGAAAAGACAACAAGCTGCCGAAGATGCAATCGCAATCGGTTTAAGA
ACAGCTTTAAATGAATCTCAACTTCCACTTTTAACCGCTGGACCTTTATGGGGTCCAGGTTCGGTTAGTGCTCCAATTCTCAGCTCTCCAAACACATCAAATTACA
ATGAATTAAACAACAACAATAATGATGAATCTGGCAAACATTCAAACTGCACAAATGAAATCAATGATGATGGAAAAGTAAACTCAAATGATTTTACTGATTCT
GACTCAAATTCCGAGAAGACTGAAGACCATGATAAGTCATATCAAGAAACGACAAGTGAAAATATTAGAACGGATCAAATGATAAAATCAAAACAATTTGCAA
TGCACAAGAATTTCAATTTTTCCTCAGATGAAAGATACTTTCGAGCTAATAACGAAATCGGTAGTAAATCAGATCTTTTGGACTTATTAAAGCATTTTTTTCCACA
TAAAATGGGATTTGATCCATTTCTTATTGATGCATTAATGGAAAAACACAATAACAATGTGTCCAAATGTTTGGAAGAAATAATCGGGAACTCATCAAATCTATC
TTCATTCAACGAACCCTCGAAGACTAATATTCATGATCAGATTCATGAAAACAATCTGTTCGATTTAAAACGATTTATTATTCCATATTCAAATGCTAATATTTCT
CCTATGAGCATGTGCCAAAATCCGATTTTGCACCGAATTCCTTCAAATAATGATTTCAATAATATTTCCCTTGAATTCAAATTTTTAAACCTACAAATTCACAATC
CATGCCCCCATCAAATTTCCAATTGAAGCCTAATGCTGCCCCTGCTTATAATTATCCCTGTAACCCATATTCCACTTTCCCTTACTTTTATTTTAATCAGAGAATAC
AACCTAAACAATGCTCCGAAGAATTTTTTAGCAAACCACAGACAGATATTTCTGAGAAACCGATCAATTTCTCGGAAAAAGATAATATAAAAAATACAGTTTTA
ATCAAATCAGAAAATAATAAATTAATTACAAATTCAAATTCTAGAAGATCCACCTTCTGGCCATTAGTAGATTTGGTTAATGAGGAATCAACATTAAGTTCAACC
TTATAACCTTCAGTCATAATTTT

dmd-1 RNAi sequence 1

TGTGTCCCTGTCCATTGTGTACAGTCGTTAGTCATGGTCGAGAAATTGTCGCTCGTCAAATTAGACTTTATAGAAATCAGAAAGACATCAGCAAAAGGTCTCCGC
AATCACCACATTGTAGACGATGCAGAAATCATGGAGAACACAATGCTTGGAAAGGCCAC

dmd-1 RNAi sequence 2

TGTGTCCCTGTCCATTGTGTGTCGTCGTTAGTCATGGTCGAGAAATTGTCGCTCGTCAAATTAGACTTTATAGAAATCAGAAAGACATCAGCAAAAGGTCTCCGC
AATCACCACATTGTAGACGATGCAGAAATCATGGAGAACACAATGCTTGGAAAGGCCACAAAAAGAAATGTCGACATTCAAATTGTCCCTGTAAATTGTGCCTG
TTAATTACCATGAGAAAAACCAAAAAAAAA

Supplementary Table S2 – Primers used for cloning and qPCR.

Primers for cloning
dmd-1 3’ RACE outer: AACCATCGTCGAAATTGTCCG 3’ RACE inner: TGTGTCCCTGTCCATTGTGT 5’ RACE outer: AACTGCTGTGACTGGTGTCG 5’ RACE inner: TGTCGTTGGCAATATTGGAG Full-length F: AACATGAATTTAACTCAATGCCATTC Full-length R: CAATTGAATAAATACAGGTCTTTTCAATG
dmd-2 3’ RACE outer: TTTAAATCAAATCTTAAATGTTTGCAG 3’ RACE inner: GCAGCCTAGGCAAATGTTTT
dmd-3 3’ RACE outer: TCTCGAGAACACCAAAATGC 3’ RACE inner: GTCGAAACCACGGTGTTGTA
dmd-4 3’ RACE outer: TGGACAAAGAATCCAAAAAGTG 3’ RACE inner: ATTAATGCAGCCCCCTCTTT
Primers for qPCR
Smed_dmd-1 F: GACGTCAAACCGAATTTACTGA *Smed_dmd-1* R: TTTCACCGGCAACAACTG
smedwi-1 F: GATTACGATTCGTGGTCGTG *smedwi-1* R: CTGTGCCTTTAGGAACGTCA
nanos F: CAAGGACAAATGTTGCCTGTA *nanos* R: CAACCCATCGATCCAACTCT
β-tubulin F: TGGCTGCTTGTGATCCAAGA *β-tubulin* R: AAATTGCCGCAACAGTCAAATA
Sm_dmd-1 F: GGACATGGTATGTCTGAACCAG *Sm_dmd-1* R: TGTTGGGCTACTATACGACGAC
Sm_cytochrome c oxidase F: GTGGGAGTCTTTGGTTGTTG *Sm_cytochrome c oxidase* R: CCACCCACAACTTAGAAGCA

Supplementary Table S3 – RNAi feeding schedule.

Asexuals (*dmd-1* dsRNA)	7 feedings, ~6 weeks
Asexuals (*nanos* dsRNA)	6 feedings, ~5 weeks
Animals devoid of germ cells (*dmd-1* dsRNA)	2 feedings over 8 days before amputation
Recently hatched planarians (*dmd-1* dsRNA)	9 feedings, ~2 months
Sexually mature planarians (*dmd-1* dsRNA)	12-14 feedings, ~ 2 months
Sexually immature regenerates, examination of testes and ovaries (*dmd-1* dsRNA)	5 feedings, ~ 1 month, 12-13 feedings, ~ 3 months
Sexually immature regenerates, examination of sperm ducts/seminal vesicles/oviducts (*dmd-1* dsRNA)	20 feedings, ~ 5 month