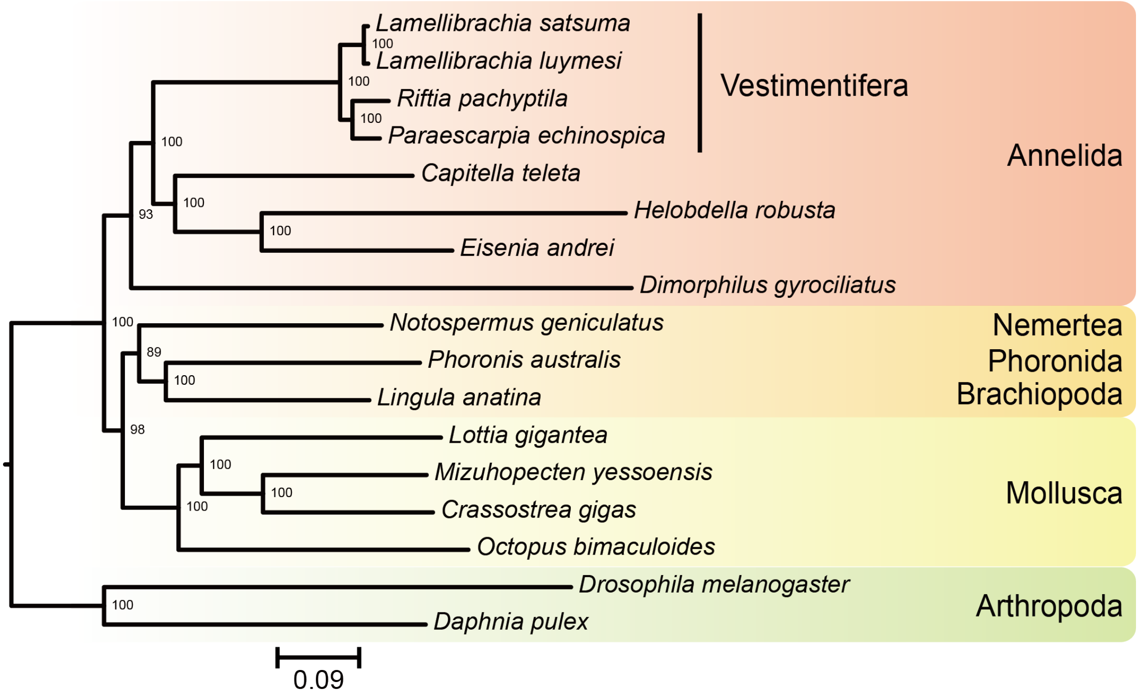
Supplementary material for

Genomic and transcriptomic analyses illuminate the molecular basis of the unique lifestyle of a tubeworm, *Lamellibrachia satsuma*

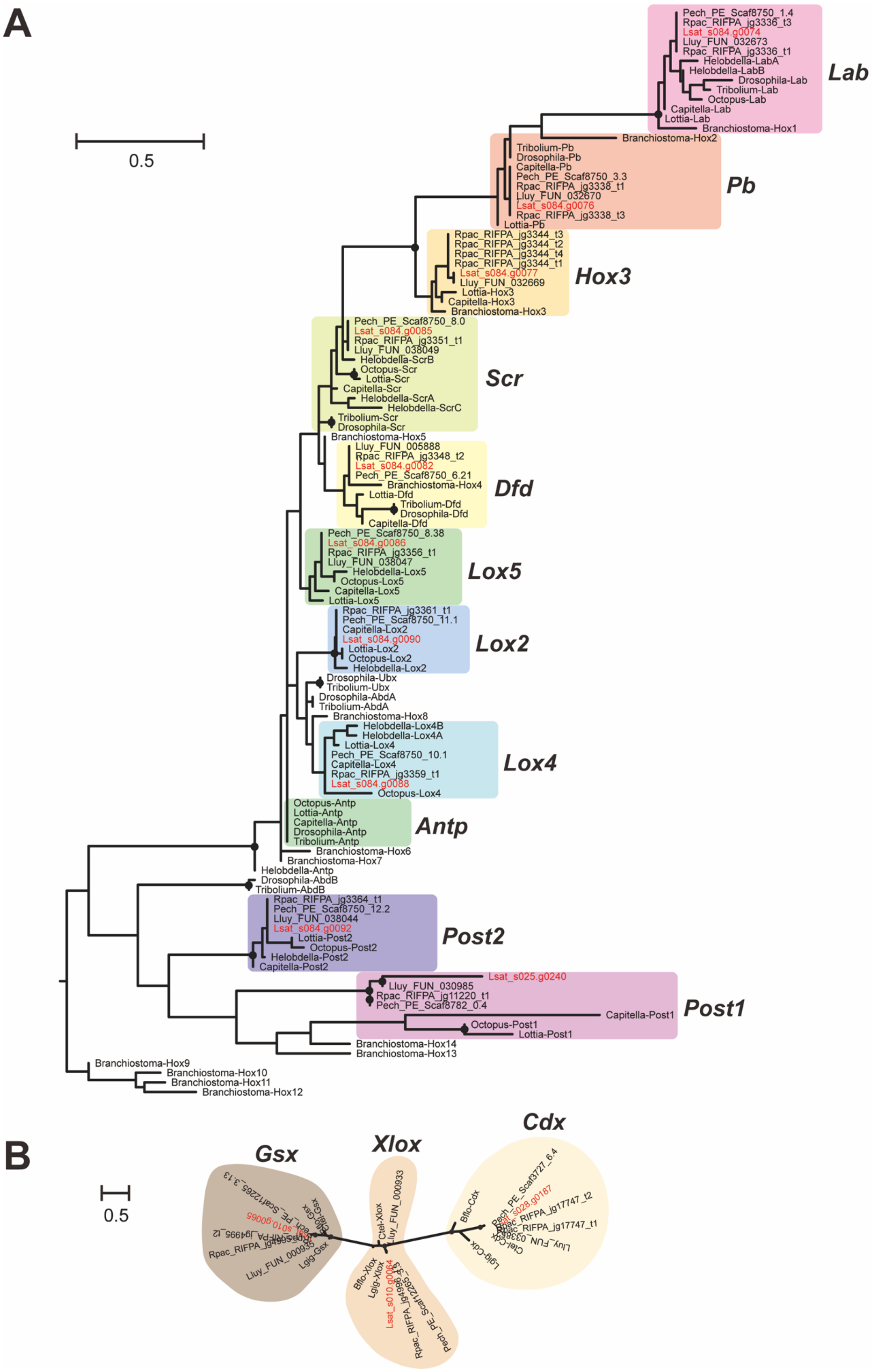
Contents

Supplementary Figures



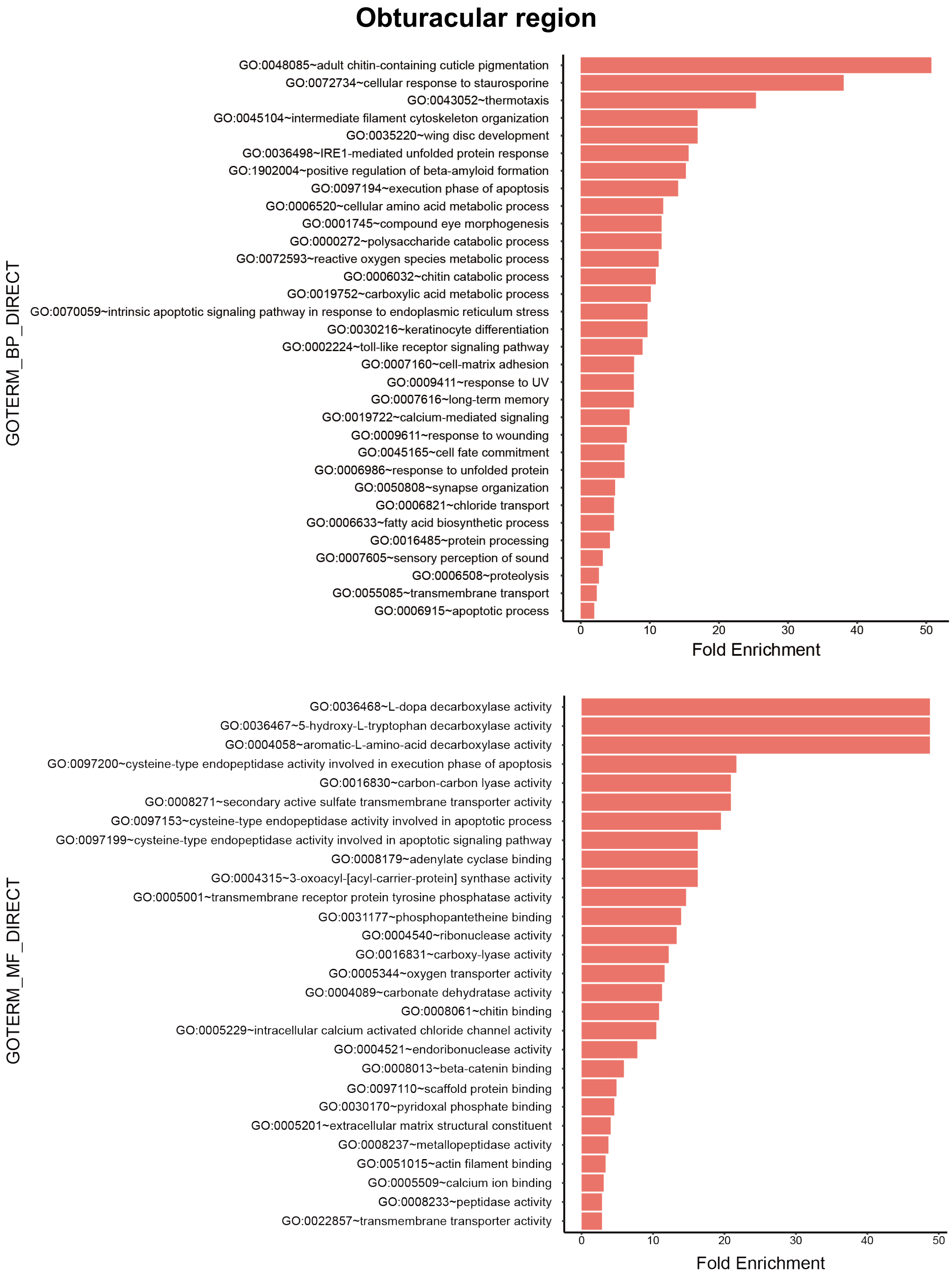
Supplementary Figure S1.

Phylogenetic relationships among four vestimentiferans and 13 metazoans. Concatenated amino acid sequences of 163 single-copy OGs were aligned, gap-trimmed, and used for construction of a maximum likelihood phylogenetic tree. Numbers on branches indicate bootstrap values.


A
0.5
Lab
Pech_PE_Scaf8750_1.4
Rpac_RIFPA_jg3336_t3
Lsat_s084.g0074
Lluy_FUN_032673
Rpac_RIFPA_jg3336_t1
Helobdella-LabA
Helobdella-LabB
Drosophila-Lab
Tribolium-Lab
Octopus-Lab
Capitella-Lab
Lottia-Lab
Branchiostoma-Hox1
Branchiostoma-Hox2
Pb
Tribolium-Pb
Drosophila-Pb
Capitella-Pb
Pech_PE_Scaf8750_3.3
Rpac_RIFPA_jg3338_t1
Lluy_FUN_032670
Lsat_s084.g0076
Rpac_RIFPA_jg3338_t3
Lottia-Pb
Hox3
Rpac_RIFPA_jg3344_t3
Rpac_RIFPA_jg3344_t2
Rpac_RIFPA_jg3344_t4
Rpac_RIFPA_jg3344_t1
Lsat_s084.g0077
Lluy_FUN_032669
Lottia-Hox3
Capitella-Hox3
Branchiostoma-Hox3
Scr
Pech_PE_Scaf8750_8.0
Lsat_s084.g0085
Rpac_RIFPA_jg3351_t1
Lluy_FUN_038049
Helobdella-ScrB
Octopus-Scr
Lottia-Scr
Capitella-Scr
Helobdella-ScrA
Helobdella-ScrC
Tribolium-Scr
Drosophila-Scr
Branchiostoma-Hox5
Dfd
Lluy_FUN_005888
Rpac_RIFPA_jg3348_t2
Lsat_s084.g0082
Pech_PE_Scaf8750_6.21
Branchiostoma-Hox4
Lottia-Dfd
Tribolium-Dfd
Drosophila-Dfd
Capitella-Dfd
Lox5
Pech_PE_Scaf8750_8.38
Lsat_s084.g0086
Rpac_RIFPA_jg3356_t1
Lluy_FUN_038047
Helobdella-Lox5
Octopus-Lox5
Capitella-Lox5
Lottia-Lox5
Lox2
Rpac_RIFPA_jg3361_t1
Pech_PE_Scaf8750_11.1
Capitella-Lox2
Lsat_s084.g0090
Lottia-Lox2
Octopus-Lox2
Helobdella-Lox2
Drosophila-Ubx
Tribolium-Ubx
Drosophila-AbdA
Tribolium-AbdA
Branchiostoma-Hox8
Lox4
Helobdella-Lox4B
Helobdella-Lox4A
Lottia-Lox4
Pech_PE_Scaf8750_10.1
Capitella-Lox4
Rpac_RIFPA_jg3359_t1
Lsat_s084.g0088
Octopus-Lox4
Antp
Octopus-Antp
Lottia-Antp
Capitella-Antp
Drosophila-Antp
Tribolium-Antp
Branchiostoma-Hox6
Branchiostoma-Hox7
Helobdella-Antp
Drosophila-AbdB
Tribolium-AbdB
Post2
Rpac_RIFPA_jg3364_t1
Pech_PE_Scaf8750_12.2
Lluy_FUN_038044
Lsat_s084.g0092
Lottia-Post2
Octopus-Post2
Helobdella-Post2
Capitella-Post2
Post1
Lsat_s025.g0240
Lluy_FUN_030985
Rpac_RIFPA_jg11220_t1
Pech_PE_Scaf8782_0.4
Capitella-Post1
Octopus-Post1
Lottia-Post1
Branchiostoma-Hox14
Branchiostoma-Hox13
Branchiostoma-Hox9
Branchiostoma-Hox10
Branchiostoma-Hox11
Branchiostoma-Hox12
B
0.5
Gsx
Pech_PE_Scaf12265_3.13
Lsat_s010.g0065
Rpac_RIFPA_jg4995_t2
Rpac_RIFPA_jg4995
Rpac_RIFPA_jg4995_t1
Lluy_FUN_000935
Lgig-Gsx
Bflo-Gsx
Ctel-Gsx
Xlox
Ctel-Xlox
Lluy_FUN_000933
Bflo-Xlox
Lgig-Xlox
Lsat_s010.g0064
Rpac_RIFPA_jg4996_t3
Pech_PE_Scaf12265_3.5
Cdx
Bflo-Cdx
Pech_PE_Scaf3727_6.4
Lsat_s028.g0187
Rpac_RIFPA_jg17747_t2
Rpac_RIFPA_jg17747_t1
Lluy_FUN_033856
Ctel-Cdx
Lgig-Cdx

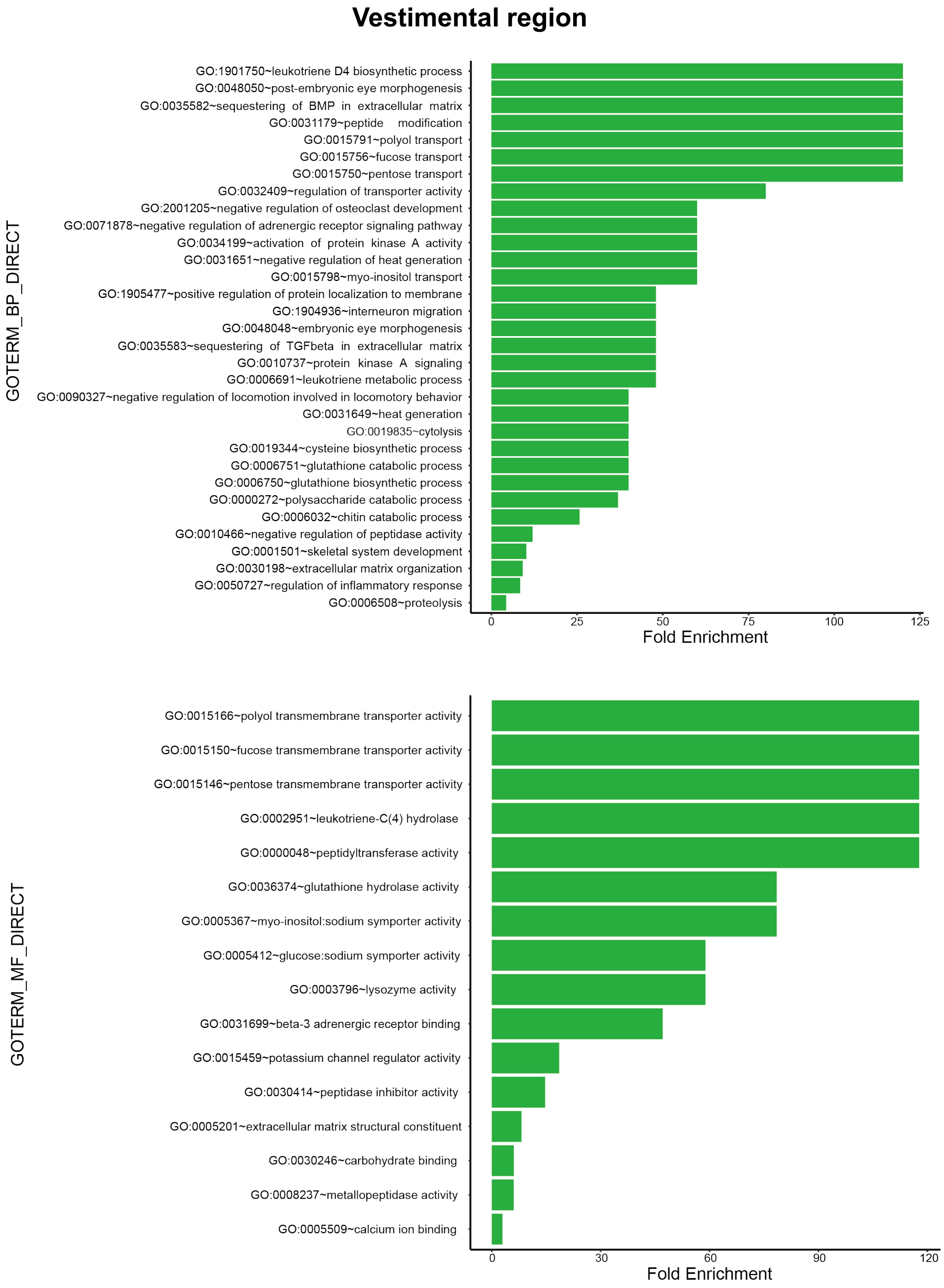
Supplementary Figure S2.

Molecular phylogenetic tree of *Hox* and *ParaHox* genes. (A) Molecular phylogenetic tree of *Hox* genes of four vestimentiferans and seven non-vestimentiferan metazoans. In order to extract *Hox* candidate genes, gene models of four vestimentiferans were searched using BLAST against *Hox* genes of *Capitella teleta*, *Helobdella robusta*, *Lottia gigantea*, *Octopus bimaculoides*, *Drosophila melanogaster*, *Tribolium castaneum*, and *Branchiostoma floridae*[1,2]. Amino acid sequences of candidate genes were aligned, gap-trimmed, and used for construction of a maximum likelihood phylogenetic tree. Although previous studies identified *Hox3* in the *Paraescarpia* genome and *Lox4* and *Lox2* in the *L. luymesi* gene models[3], our study and another study were not able to identify clear orthologs of these genes[4]. (B) Molecular phylogenetic tree of *ParaHox* genes of four vestimentiferans and three non-vestimentiferan metazoans. In order to extract *ParaHox* candidate genes, gene models of four vestimentiferans were searched using BLAST against *ParaHox* genes of *C. teleta*, *L. gigantea*, and *B. floridae*[5,6]. Amino acid sequences of candidate genes were aligned, gap-trimmed, and used for construction of a maximum likelihood phylogenetic tree. For genes of non-vestimentiferan metazoans, taxonomy and definitions are shown. Circles on branches indicate bootstrap values higher than 80%. *L. satsuma* genes are highlighted in red.



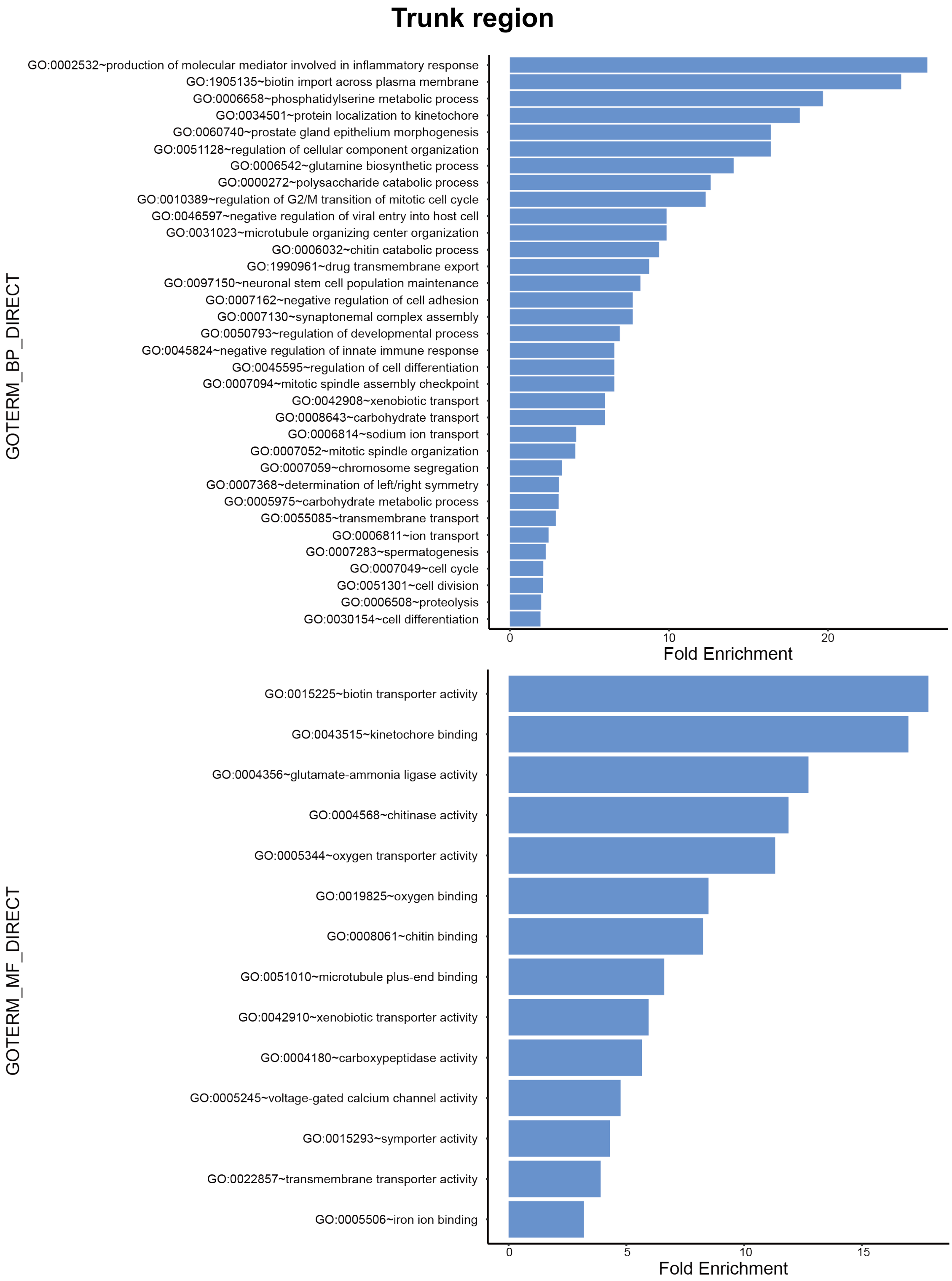
Supplementary Figure S3.

Functional enrichment analysis of obturacular region-specific DEGs based on Gene Ontology (GO) terms. GO terms with p-values < 0.05 are shown.



Supplementary Figure S4.

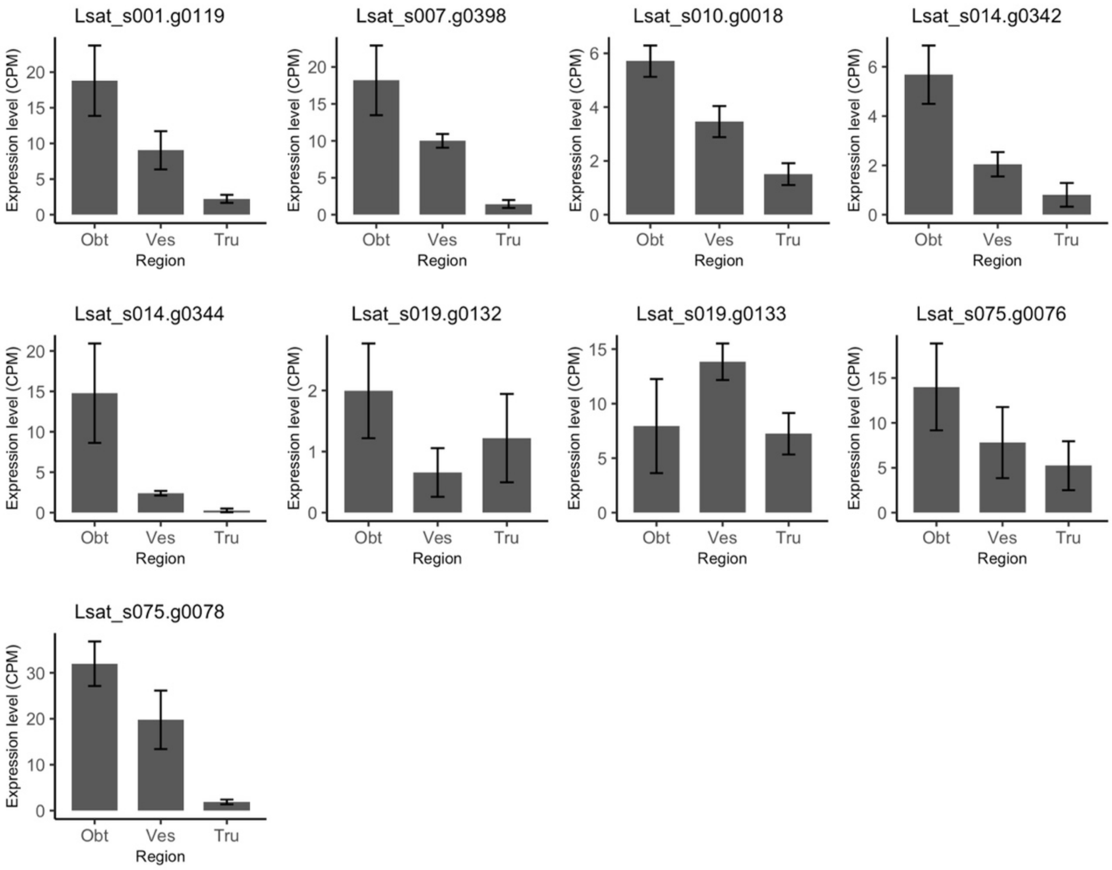
Functional enrichment analysis of vestimental region-specific DEGs based on GO terms. GO terms with *p*-values < 0.05 are shown.



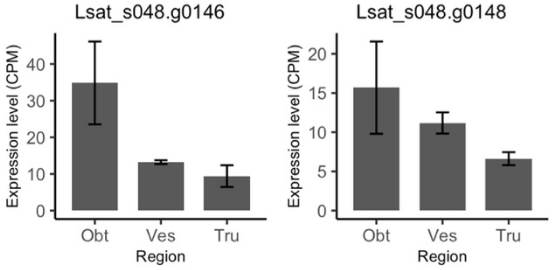
Supplementary Figure S5.

Functional enrichment analysis of trunk region-specific DEGs based on GO terms. GO terms with *p*-values < 0.05 are shown.

TLR

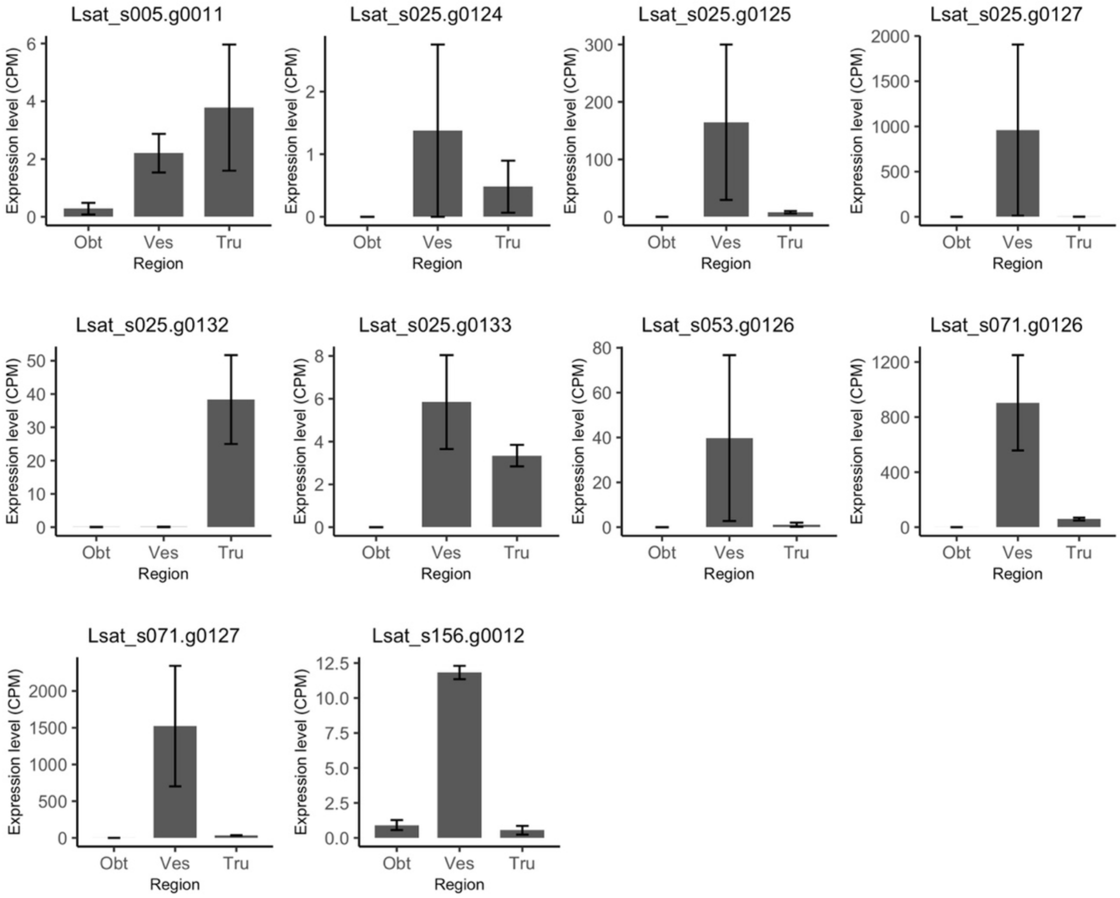


MyD88



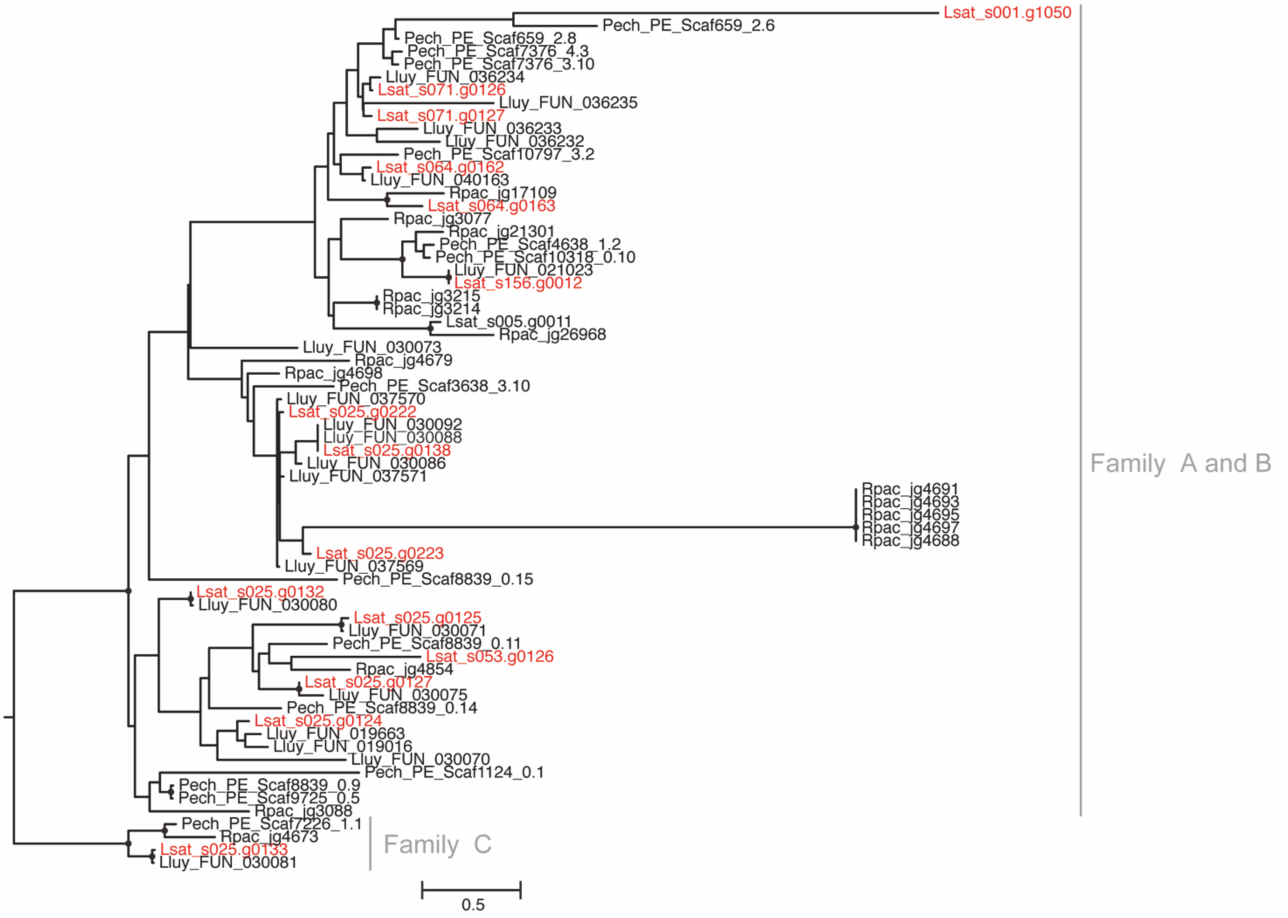
Supplementary Figure S6.

Gene expression levels (TMM-normalized CPMs) of TLR and MyD88. Data are shown as means ± SEs (n = 3). Obt, obturacular region; Ves, vestimental region; Tru, trunk region.



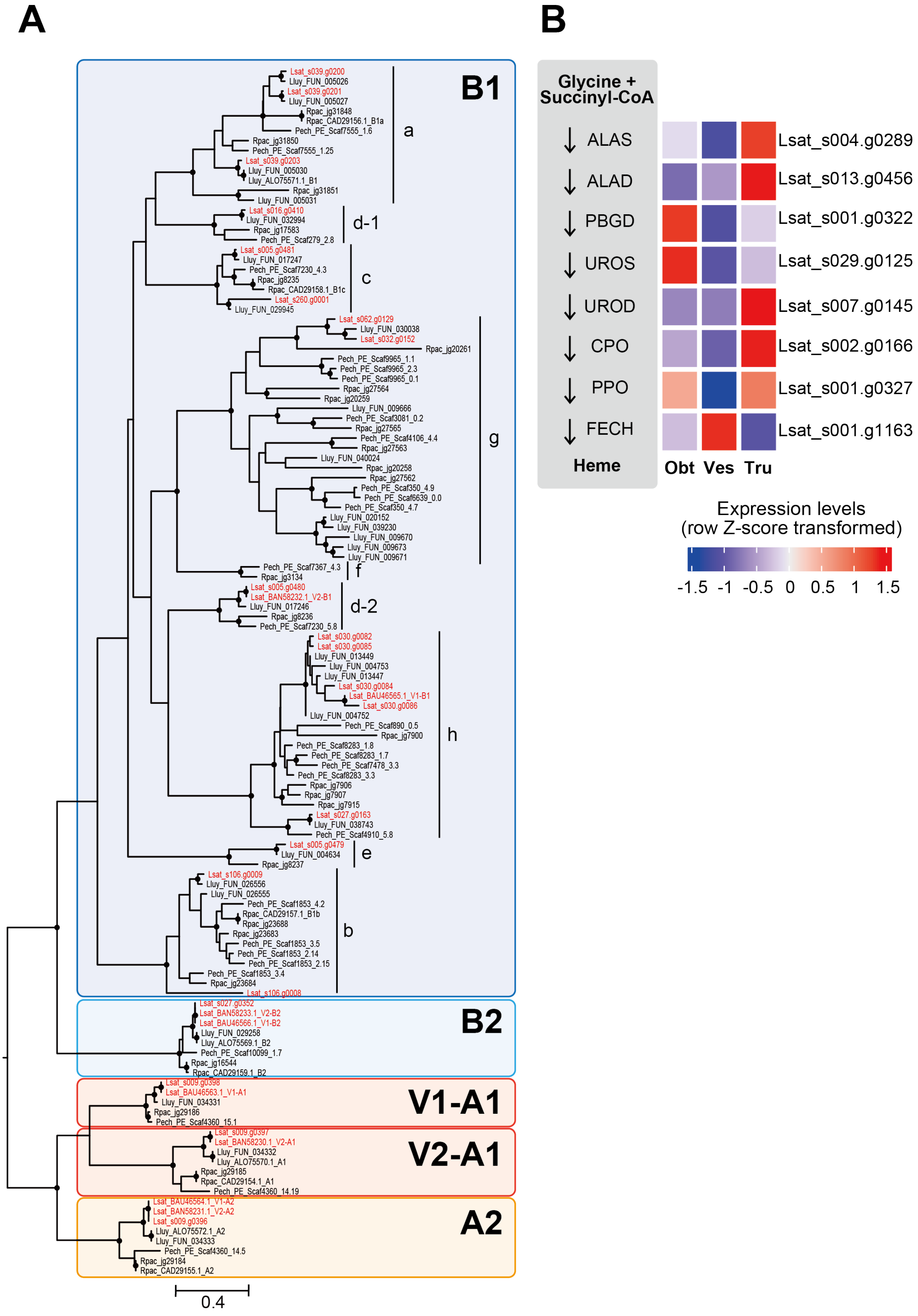
Supplementary Figure S7.

Gene expression levels (TMM-normalized CPMs) of lysozymes. Data are shown as means ± SEs (n = 3). Obt, obturacular region; Ves, vestimental region; Tru, trunk region.



Supplementary Figure S8.

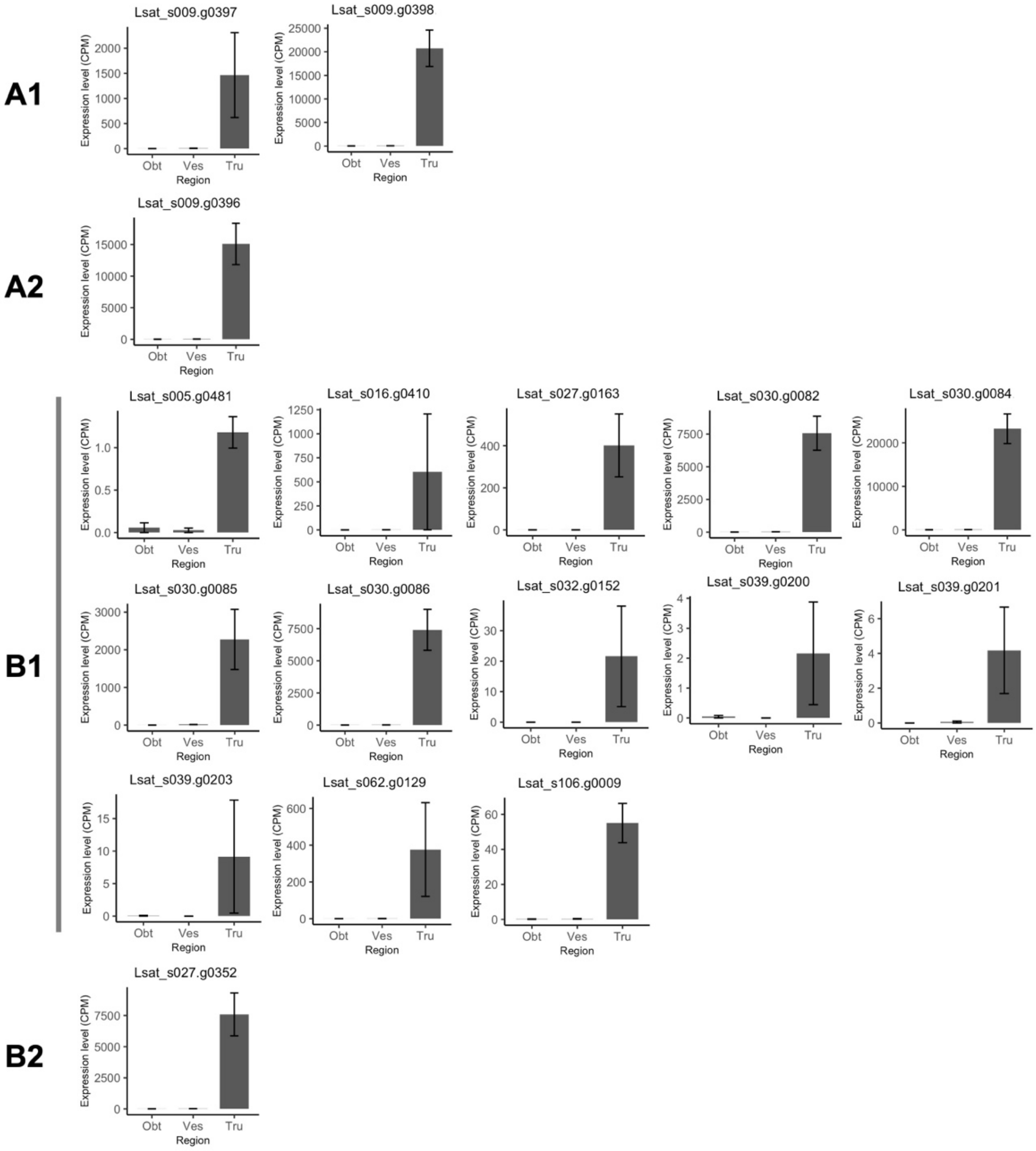
Molecular phylogenetic analysis of lysozyme genes in genomes of vestimentiferans. The maximum likelihood phylogenetic tree was constructed using aligned and gap-trimmed amino acid sequences. *L. satsuma* genes are highlighted in red. Circles on branches indicate bootstrap values higher than 80%.


A
B1
Lsat_s039.g0200
Lluy_FUN_005026
Lsat_s039.g0201
Lluy_FUN_005027
Rpac_jg31848
Rpac_CAD29156.1_B1a
Pech_PE_Scaf7555_1.6
Rpac_jg31850
Pech_PE_Scaf7555_1.25
Lsat_s039.g0203
Lluy_FUN_005030
Lluy_ALO75571.1_B1
Rpac_jg31851
Lluy_FUN_005031
a
Lsat_s016.g0410
Lluy_FUN_032994
Rpac_jg17583
Pech_PE_Scaf279_2.8
d-1
Lsat_s005.g0481
Lluy_FUN_017247
Pech_PE_Scaf7230_4.3
Rpac_jg8235
Rpac_CAD29158.1_B1c
Lsat_s260.g0001
Lluy_FUN_029945
c
Lsat_s062.g0129
Lluy_FUN_030038
Lsat_s032.g0152
Rpac_jg20261
Pech_PE_Scaf9965_1.1
Pech_PE_Scaf9965_2.3
Pech_PE_Scaf9965_0.1
Rpac_jg27564
Rpac_jg20259
Lluy_FUN_009666
Pech_PE_Scaf3081_0.2
Rpac_jg27565
Pech_PE_Scaf4106_4.4
Rpac_jg27563
Lluy_FUN_040024
Rpac_jg20258
Rpac_jg27562
Pech_PE_Scaf350_4.9
Pech_PE_Scaf6639_0.0
Pech_PE_Scaf350_4.7
Lluy_FUN_020152
Lluy_FUN_039230
Lluy_FUN_009670
Lluy_FUN_009673
Lluy_FUN_009671
g
Pech_PE_Scaf7367_4.3
Rpac_jg3134
f
Lsat_s005.g0480
Lsat_BAN58232.1_V2-B1
Lluy_FUN_017246
Rpac_jg8236
Pech_PE_Scaf7230_5.8
d-2
Lsat_s030.g0082
Lsat_s030.g0085
Lluy_FUN_013449
Lluy_FUN_004753
Lluy_FUN_013447
Lsat_s030.g0084
Lsat_BAU46565.1_V1-B1
Lsat_s030.g0086
Lluy_FUN_004752
Pech_PE_Scaf890_0.5
Rpac_jg7900
Pech_PE_Scaf8283_1.8
Pech_PE_Scaf8283_1.7
Pech_PE_Scaf7478_3.3
Pech_PE_Scaf8283_3.3
Rpac_jg7906
Rpac_jg7907
Rpac_jg7915
Lsat_s027.g0163
Lluy_FUN_038743
Pech_PE_Scaf4910_5.8
h
Lsat_s005.g0479
Lluy_FUN_004634
Rpac_jg8237
e
Lsat_s106.g0009
Lluy_FUN_026556
Lluy_FUN_026555
Pech_PE_Scaf1853_4.2
Rpac_CAD29157.1_B1b
Rpac_jg23688
Rpac_jg23683
Pech_PE_Scaf1853_3.5
Pech_PE_Scaf1853_2.14
Pech_PE_Scaf1853_2.15
Pech_PE_Scaf1853_3.4
Rpac_jg23684
Lsat_s106.g0008
b
B2
Lsat_s027.g0352
Lsat_BAN58233.1_V2-B2
Lsat_BAU46566.1_V1-B2
Lluy_FUN_029258
Lluy_ALO75569.1_B2
Pech_PE_Scaf10099_1.7
Rpac_jg16544
Rpac_CAD29159.1_B2
V1-A1
Lsat_s009.g0398
Lsat_BAU46563.1_V1-A1
Lluy_FUN_034331
Rpac_jg29186
Pech_PE_Scaf4360_15.1
V2-A1
Lsat_s009.g0397
Lsat_BAN58230.1_V2-A1
Lluy_FUN_034332
Lluy_ALO75570.1_A1
Rpac_jg29185
Rpac_CAD29154.1_A1
Pech_PE_Scaf4360_14.19
A2
Lsat_BAU46564.1_V1-A2
Lsat_BAN58231.1_V2-A2
Lsat_s009.g0396
Lluy_ALO75572.1_A2
Lluy_FUN_034333
Pech_PE_Scaf4360_14.5
Rpac_jg29184
Rpac_CAD29155.1_A2
0.4
B
Glycine +
Succinyl-CoA
↓ ALAS
Lsat_s004.g0289
↓ ALAD
Lsat_s013.g0456
↓ PBGD
Lsat_s001.g0322
↓ UROS
Lsat_s029.g0125
↓ UROD
Lsat_s007.g0145
↓ CPO
Lsat_s002.g0166
↓ PPO
Lsat_s001.g0327
↓ FECH
Lsat_s001.g1163
Heme
Obt
Ves
Tru
Expression levels
(row Z-score transformed)
-1.5
-1
-0.5
0
0.5
1
1.5

Supplementary Figure S9.

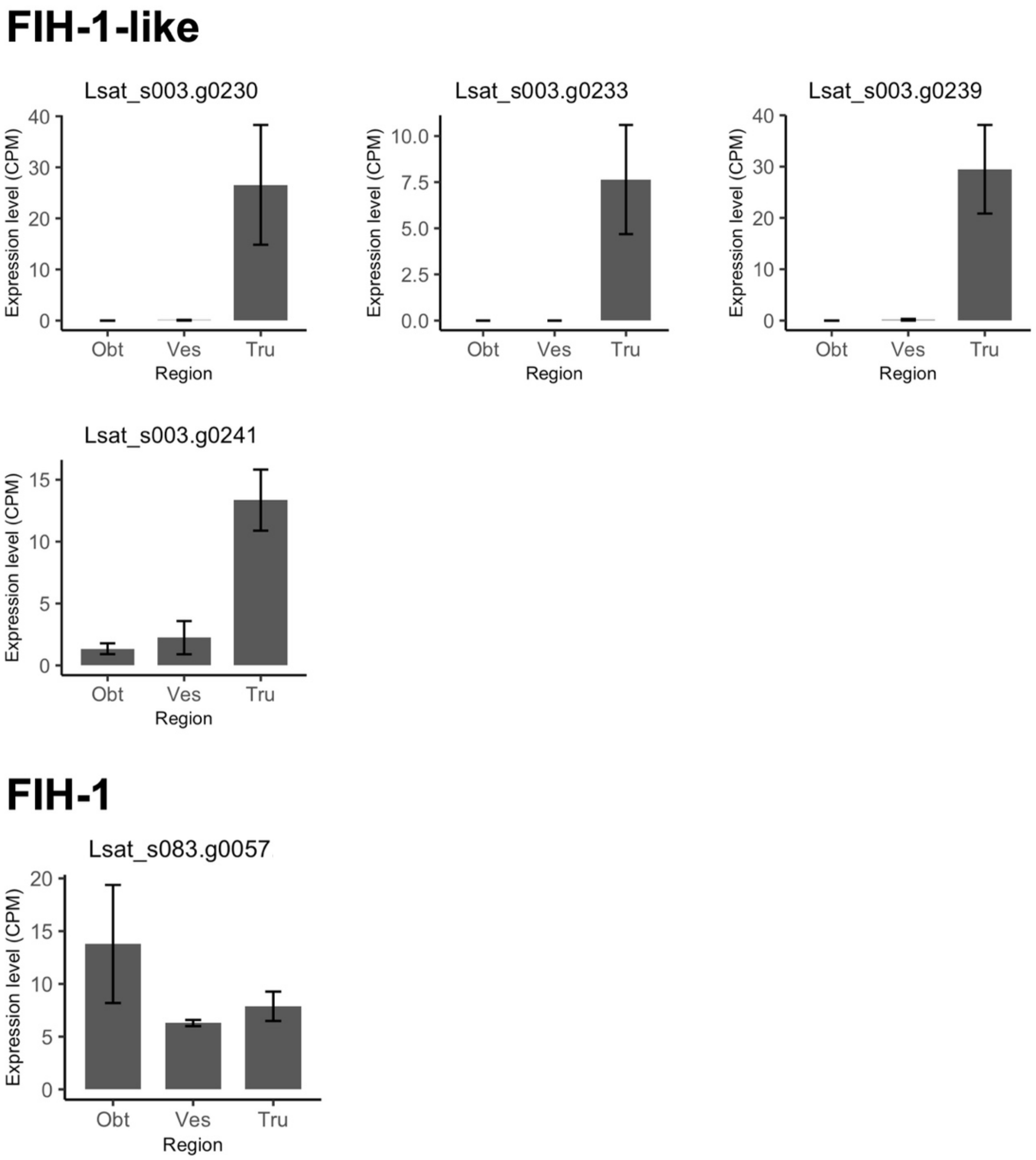
Molecular phylogenetic tree of globin subunit genes and expression levels of genes involved in heme biosynthesis. (A) Molecular phylogenetic tree of globin subunit genes of four vestimentiferans. The maximum likelihood phylogenetic tree was constructed using aligned and gap-trimmed amino acid sequences. For genes retrieved from the NCBI GenBank and UniProt/Swiss-Prot databases, taxonomy, accession number, and definitions are shown, separated by underscores. *L. satsuma* genes are highlighted in red. Circles on branches indicate bootstrap values higher than 80%. “a” – “h” correspond to eight paralogous groups proposed in a previous study[4]. Note that paralogous group d[4] was divided into two independent clades (d-1 and d-2); therefore, the total number of groups is nine. (B) Expression levels of putative genes involved in heme biosynthesis. The heme biosynthetic pathway and involved enzymes are shown on the left. A heat map of expression levels of the genes corresponding to the enzymes (row Z-score transformed) is shown on the right.

Lsat, *L. satsuma*; Lluy, *L. luymesi*; Rpac, *R. pachyptila*; Pech, *P. echinospica*; A1-B2, globin A1-B2 chain (subunit); V1, V2, V1 and V2 hemoglobin; Obt, obturacular region; Ves, vestimental region; Tru, trunk region.



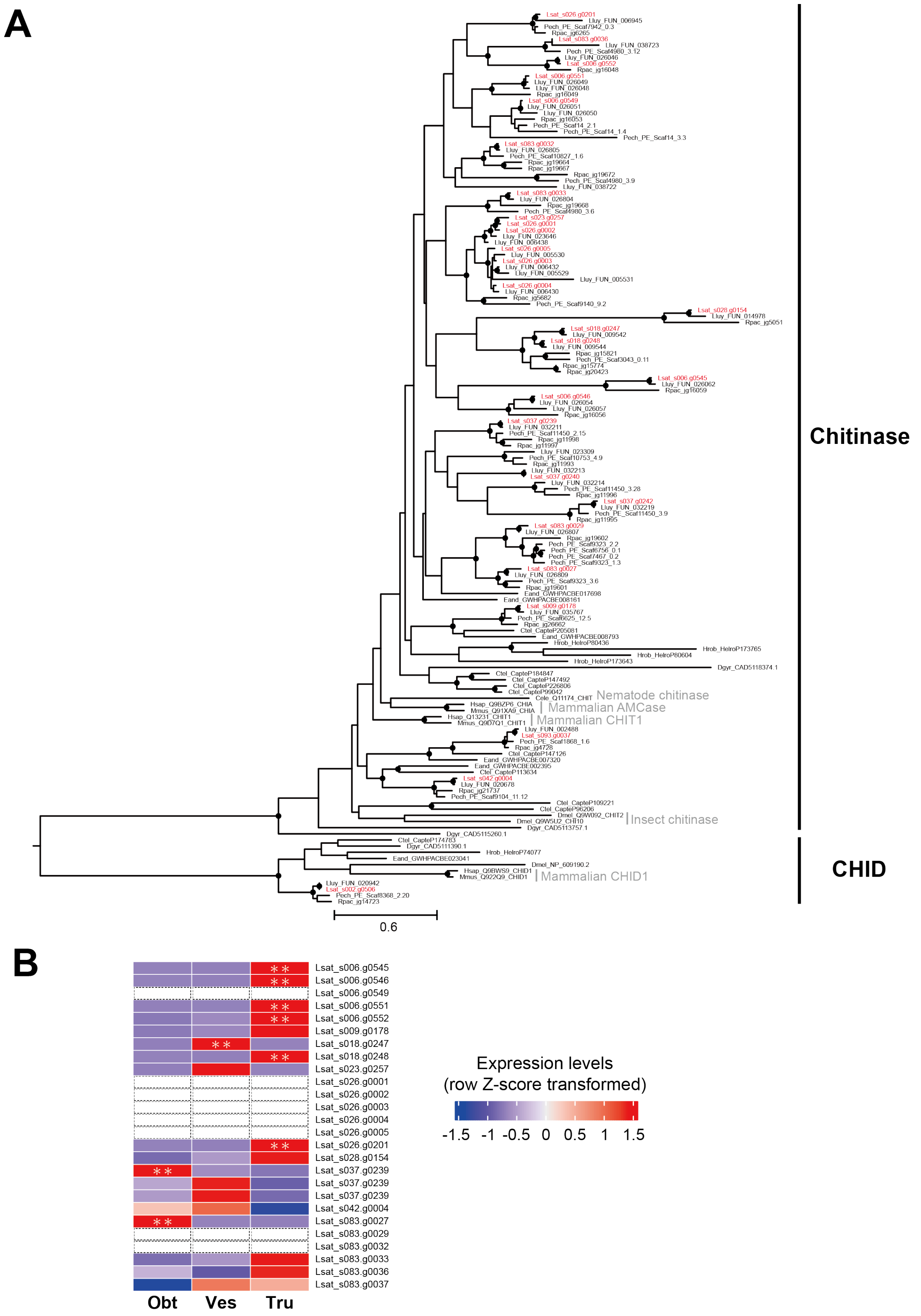
Supplementary Figure S10.

Gene expression levels (TMM-normalized CPMs) of globin subunits. Data are shown as means ± SEs (n = 3). A1-B2, globin A1-B2 chain (subunit); Obt, obturacular region; Ves, vestimental region; Tru, trunk region.



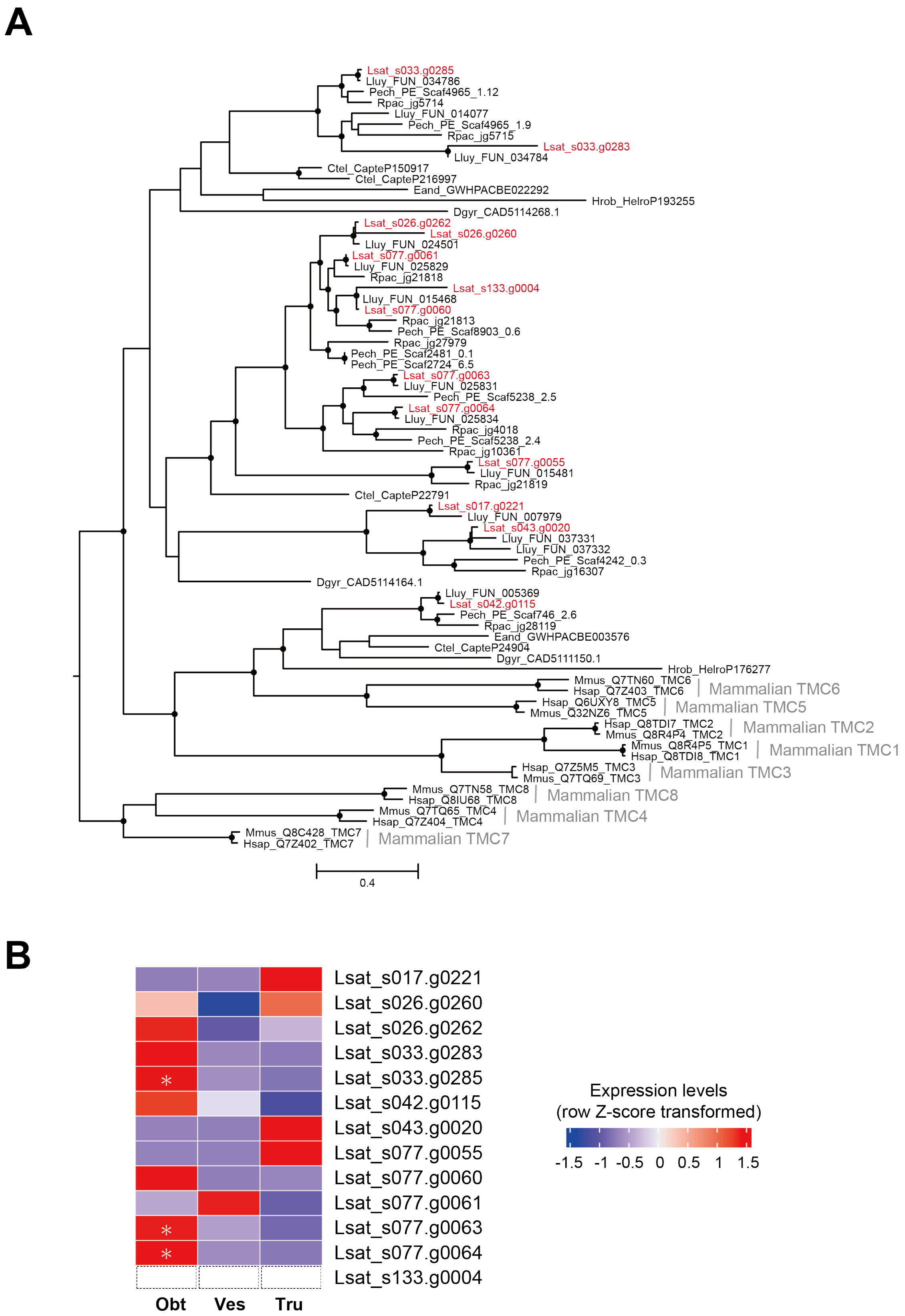
Supplementary Figure S11.

Gene expression levels (TMM-normalized CPMs) of FIH-1 and FIH-1-like proteins. Data are shown as means ± SEs (n = 3). Obt, obturacular region; Ves, vestimental region; Tru, trunk region.


A
Lsat_s026.g0201
Lluy_FUN_006945
Pech_PE_Scaf7942_0.3
Rpac_jg6265
Lsat_s083.g0036
Lluy_FUN_038723
Pech_PE_Scaf4980_3.12
Lluy_FUN_026046
Lsat_s006.g0552
Rpac_jg16048
Lsat_s006.g0551
Lluy_FUN_026049
Lluy_FUN_026048
Rpac_jg16049
Lsat_s006.g0549
Lluy_FUN_026051
Lluy_FUN_026050
Rpac_jg16053
Pech_PE_Scaf14_2.1
Pech_PE_Scaf14_1.4
Pech_PE_Scaf14_3.3
Lsat_s083.g0032
Lluy_FUN_026805
Pech_PE_Scaf10827_1.6
Rpac_jg19664
Rpac_jg19667
Rpac_jg19672
Pech_PE_Scaf4980_3.9
Lluy_FUN_038722
Lsat_s083.g0033
Lluy_FUN_026804
Rpac_jg19668
Pech_PE_Scaf4980_3.6
Lsat_s023.g0257
Lsat_s026.g0001
Lsat_s026.g0002
Lluy_FUN_023646
Lluy_FUN_006438
Lsat_s026.g0005
Lluy_FUN_005530
Lsat_s026.g0003
Lluy_FUN_006432
Lluy_FUN_005529
Lluy_FUN_005531
Lsat_s026.g0004
Lluy_FUN_006430
Rpac_jg5682
Pech_PE_Scaf9140_9.2
Lsat_s028.g0154
Lluy_FUN_014978
Rpac_jg5051
Lsat_s018.g0247
Lluy_FUN_009542
Lsat_s018.g0248
Lluy_FUN_009544
Rpac_jg15821
Pech_PE_Scaf3043_0.11
Rpac_jg15774
Rpac_jg20423
Lsat_s006.g0545
Lluy_FUN_026062
Rpac_jg16059
Lsat_s006.g0546
Lluy_FUN_026054
Lluy_FUN_026057
Rpac_jg16056
Lsat_s037.g0239
Lluy_FUN_032211
Pech_PE_Scaf11450_2.15
Rpac_jg11998
Rpac_jg11997
Lluy_FUN_023309
Pech_PE_Scaf10753_4.9
Rpac_jg11993
Lluy_FUN_032213
Lsat_s037.g0240
Lluy_FUN_032214
Pech_PE_Scaf11450_3.28
Rpac_jg11996
Lsat_s037.g0242
Lluy_FUN_032219
Pech_PE_Scaf11450_3.9
Rpac_jg11995
Lsat_s083.g0029
Lluy_FUN_026807
Rpac_jg19602
Pech_PE_Scaf9323_2.2
Pech_PE_Scaf6756_0.1
Pech_PE_Scaf7467_0.2
Pech_PE_Scaf9323_1.3
Lsat_s083.g0027
Lluy_FUN_026809
Pech_PE_Scaf9323_3.6
Rpac_jg19601
Eand_GWHPACBE017698
Eand_GWHPACBE008161
Lsat_s009.g0178
Lluy_FUN_035767
Pech_PE_Scaf6625_12.5
Rpac_jg26662
Ctel_CapteP205081
Eand_GWHPACBE008793
Hrob_HelroP80436
Hrob_HelroP173765
Hrob_HelroP80604
Hrob_HelroP173643
Dgyr_CAD5118374.1
Ctel_CapteP184847
Ctel_CapteP147492
Ctel_CapteP226806
Ctel_CapteP99042
Cele_Q11174_CHIT
Nematode chitinase
Hsap_Q9BZP6_CHIA
Mmus_Q91XA9_CHIA
Mammalian AMCase
Hsap_Q13231_CHIT1
Mmus_Q9D7Q1_CHIT1
Mammalian CHIT1
Lluy_FUN_002488
Lsat_s093.g0037
Pech_PE_Scaf1868_1.6
Rpac_jg4728
Ctel_CapteP147126
Eand_GWHPACBE007320
Eand_GWHPACBE002395
Ctel_CapteP113634
Lsat_s042.g0004
Lluy_FUN_020678
Rpac_jg21737
Pech_PE_Scaf9104_11.12
Ctel_CapteP109221
Ctel_CapteP96206
Dmel_Q9W092_CHIT2
Dmel_Q9W5U2_CHI10
Insect chitinase
Dgyr_CAD5113757.1
Dgyr_CAD5115260.1
Ctel_CapteP174783
Dgyr_CAD5111390.1
Hrob_HelroP74077
Eand_GWHPACBE023041
Dmel_NP_609190.2
Hsap_Q9BWS9_CHID1
Mmus_Q922Q9_CHID1
Mammalian CHID1
Lluy_FUN_020942
Lsat_s002.g0506
Pech_PE_Scaf8368_2.20
Rpac_jg14723
0.6
Chitinase
CHID
B
Lsat_s006.g0545
Lsat_s006.g0546
Lsat_s006.g0549
Lsat_s006.g0551
Lsat_s006.g0552
Lsat_s009.g0178
Lsat_s018.g0247
Lsat_s018.g0248
Lsat_s023.g0257
Lsat_s026.g0001
Lsat_s026.g0002
Lsat_s026.g0003
Lsat_s026.g0004
Lsat_s026.g0005
Lsat_s026.g0201
Lsat_s028.g0154
Lsat_s037.g0239
Lsat_s037.g0239
Lsat_s037.g0239
Lsat_s042.g0004
Lsat_s083.g0027
Lsat_s083.g0029
Lsat_s083.g0032
Lsat_s083.g0033
Lsat_s083.g0036
Lsat_s083.g0037
* *
Obt
Ves
Tru
Expression levels
(row Z-score transformed)
-1.5
-1
-0.5
0
0.5
1
1.5

Supplementary Figure S12.

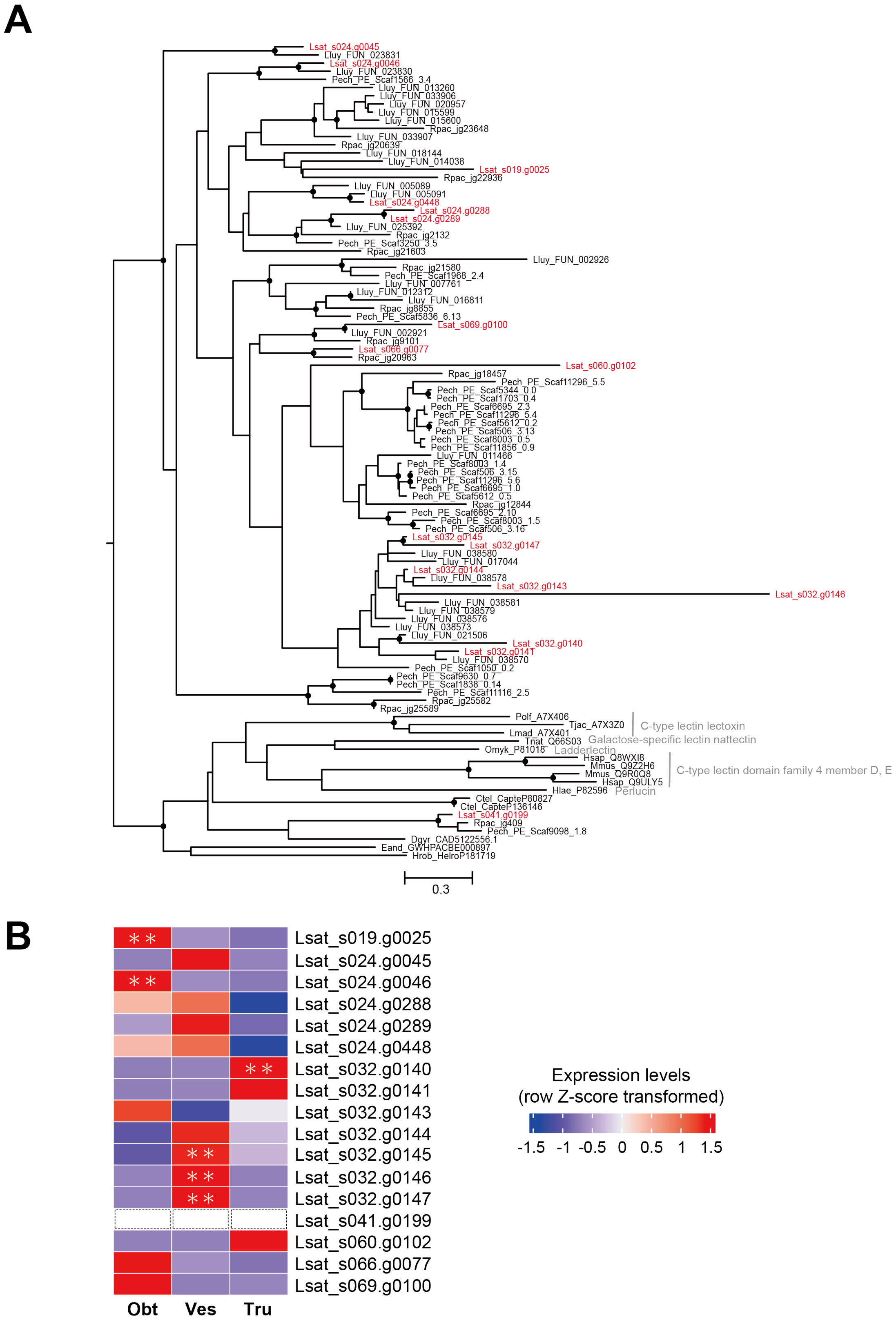
Molecular phylogenetic tree and expression levels of chitinase genes. (A) Molecular phylogenetic tree of the chitinase family in genomes of four vestimentiferans. The maximum likelihood phylogenetic tree was constructed using aligned and gap-trimmed amino acid sequences. For genes retrieved from the UniProt database, taxonomy, accession number, and definitions are shown. *L. satsuma* genes are highlighted in red. Circles on branches indicate bootstrap values higher than 80%. (B) Heat map of expression levels (row Z-score transformed) of *L. satsuma* genes belonging to the chitinase family. AMCase, acidic mammalian chitinase; CHIT, chitotriosidase; CHID, chitinase domain-containing protein; Lsat, *L. satsuma*; Lluy, *L. luymesi*; Rpac, *R. pachyptila*; Pech, *P. echinospica*; Ctel, *C. teleta*; Hrob, *Helobdella robusta*; Eand, *Eisenia andrei*, Dgyr; *Dimorphilus gyrociliatus*; Cele, *Caenorhabditis elegans*; Hsap, *Homo sapiens*; Mmus, *Mus musculus*; Dmel, *Drosophila melanogaster*; Obt, obturacular region; Ves, vestimental region; Tru, trunk region; **, region-specific DEGs.


A
Lsat_s033.g0285
Lluy_FUN_034786
Pech_PE_Scaf4965_1.12
Rpac_jg5714
Lluy_FUN_014077
Pech_PE_Scaf4965_1.9
Rpac_jg5715
Lsat_s033.g0283
Lluy_FUN_034784
Ctel_CapteP150917
Ctel_CapteP216997
Eand_GWHPACBE022292
Hrob_HelroP193255
Dgyr_CAD5114268.1
Lsat_s026.g0262
Lsat_s026.g0260
Lluy_FUN_024501
Lsat_s077.g0061
Lluy_FUN_025829
Rpac_jg21818
Lsat_s133.g0004
Lluy_FUN_015468
Lsat_s077.g0060
Rpac_jg21813
Pech_PE_Scaf8903_0.6
Rpac_jg27979
Pech_PE_Scaf2481_0.1
Pech_PE_Scaf2724_6.5
Lsat_s077.g0063
Lluy_FUN_025831
Pech_PE_Scaf5238_2.5
Lsat_s077.g0064
Lluy_FUN_025834
Rpac_jg4018
Pech_PE_Scaf5238_2.4
Rpac_jg10361
Lsat_s077.g0055
Lluy_FUN_015481
Rpac_jg21819
Ctel_CapteP22791
Lsat_s017.g0221
Lluy_FUN_007979
Lsat_s043.g0020
Lluy_FUN_037331
Lluy_FUN_037332
Pech_PE_Scaf4242_0.3
Rpac_jg16307
Dgyr_CAD5114164.1
Lluy_FUN_005369
Lsat_s042.g0115
Pech_PE_Scaf746_2.6
Rpac_jg28119
Eand_GWHPACBE003576
Ctel_CapteP24904
Dgyr_CAD5111150.1
Hrob_HelroP176277
Mmus_Q7TN60_TMC6
Hsap_Q7Z403_TMC6
Mammalian TMC6
Hsap_Q6UXY8_TMC5
Mmus_Q32NZ6_TMC5
Mammalian TMC5
Hsap_Q8TDI7_TMC2
Mmus_Q8R4P4_TMC2
Mammalian TMC2
Mmus_Q8R4P5_TMC1
Hsap_Q8TDI8_TMC1
Mammalian TMC1
Hsap_Q7Z5M5_TMC3
Mmus_Q7TQ69_TMC3
Mammalian TMC3
Mmus_Q7TN58_TMC8
Hsap_Q8IU68_TMC8
Mammalian TMC8
Mmus_Q7TQ65_TMC4
Hsap_Q7Z404_TMC4
Mammalian TMC4
Mmus_Q8C428_TMC7
Hsap_Q7Z402_TMC7
Mammalian TMC7
0.4
B
Lsat_s017.g0221
Lsat_s026.g0260
Lsat_s026.g0262
Lsat_s033.g0283
Lsat_s033.g0285
Lsat_s042.g0115
Lsat_s043.g0020
Lsat_s077.g0055
Lsat_s077.g0060
Lsat_s077.g0061
Lsat_s077.g0063
Lsat_s077.g0064
Lsat_s133.g0004
*
*
*
Obt
Ves
Tru
Expression levels
(row Z-score transformed)
-1.5
-1
-0.5
0
0.5
1
1.5

Supplementary Figure S13.

Molecular phylogenetic tree and expression levels of TMC genes. (A) Molecular phylogenetic tree of the TMC family in the genomes of four vestimentiferans. The maximum likelihood phylogenetic tree was constructed using aligned and gap-trimmed amino acid sequences. For genes retrieved from the UniProt database, taxonomy and accession number, and definitions are shown. *L. satsuma* genes are highlighted in red. Circles on branches indicate bootstrap values higher than 80%. (B) Heat map of expression levels (row Z-score transformed) of *L. satsuma* genes belonging to the TMC family.

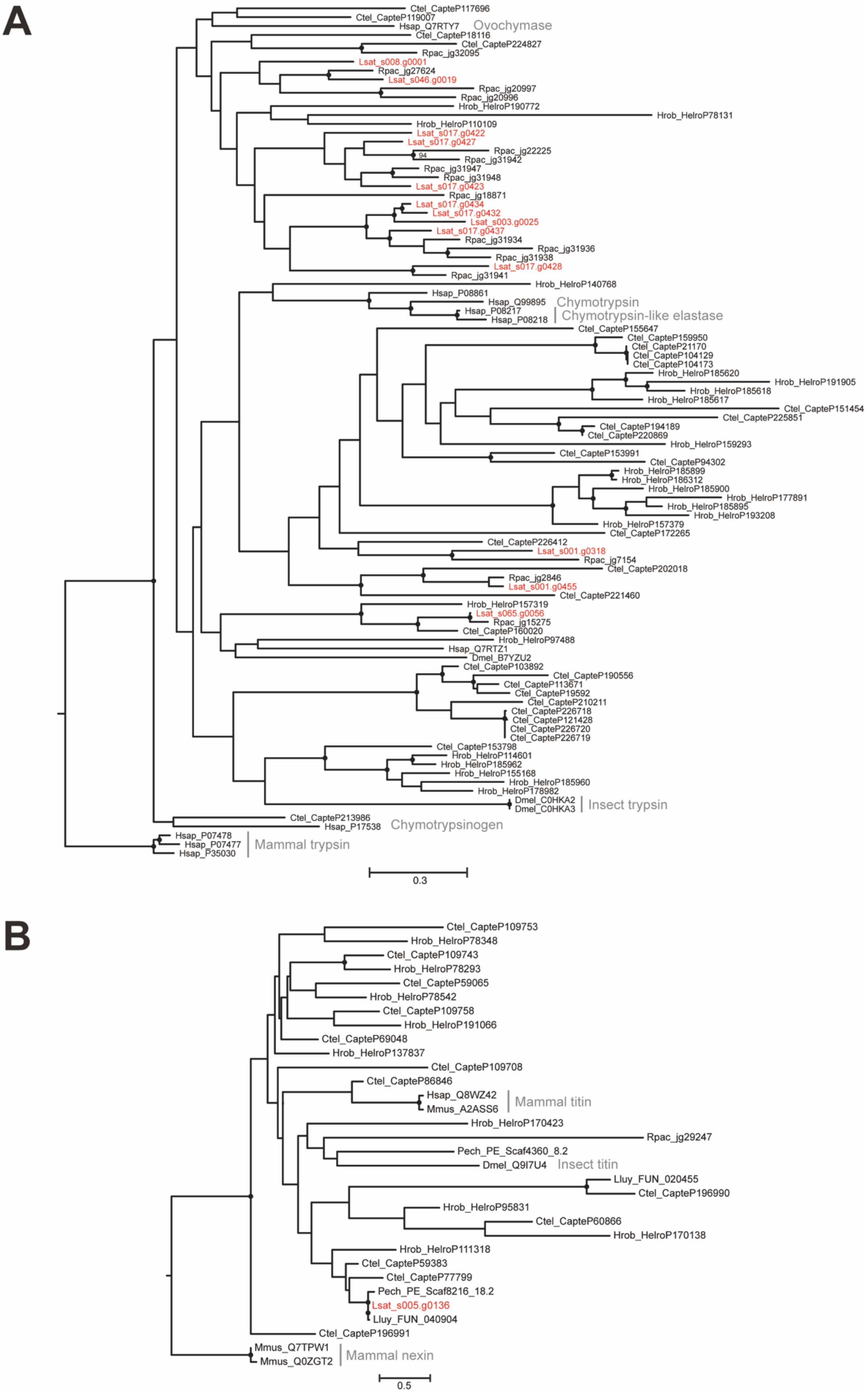
Lsat, *L. satsuma*; Lluy, *L. luymesi*; Rpac, *R. pachyptila*; Pech, *P. echinospica*; Ctel, *C. teleta*; Hrob, *Helobdella robusta*; Eand, *Eisenia andrei*, Dgyr; *Dimorphilus gyrociliatus*; Hsap, *Homo sapiens*; Mmus, *Mus musculus*; Obt, obturacular region; Ves, vestimental region; Tru, trunk region; *, genes with expression levels significantly higher in a specific region than other regions (FDR < 0.05).


A
Lsat_s024.g0045
Lluy_FUN_023831
Lsat_s024.g0046
Lluy_FUN_023830
Pech_PE_Scaf1566_3.4
Lluy_FUN_013260
Lluy_FUN_033906
Lluy_FUN_020957
Lluy_FUN_015599
Lluy_FUN_015600
Rpac_jg23648
Lluy_FUN_033907
Rpac_jg20639
Lluy_FUN_018144
Lluy_FUN_014038
Lsat_s019.g0025
Rpac_jg22936
Lluy_FUN_005089
Lluy_FUN_005091
Lsat_s024.g0448
Lsat_s024.g0288
Lsat_s024.g0289
Lluy_FUN_025392
Rpac_jg2132
Pech_PE_Scaf3250_3.5
Rpac_jg21603
Lluy_FUN_002926
Rpac_jg21580
Pech_PE_Scaf1968_2.4
Lluy_FUN_007761
Lluy_FUN_012312
Lluy_FUN_016811
Rpac_jg8855
Pech_PE_Scaf5836_6.13
Lsat_s069.g0100
Lluy_FUN_002921
Rpac_jg9101
Lsat_s066.g0077
Rpac_jg20963
Lsat_s060.g0102
Rpac_jg18457
Pech_PE_Scaf11296_5.5
Pech_PE_Scaf5344_0.0
Pech_PE_Scaf1703_0.4
Pech_PE_Scaf6695_2.3
Pech_PE_Scaf11296_5.4
Pech_PE_Scaf5612_0.2
Pech_PE_Scaf506_3.13
Pech_PE_Scaf8003_0.5
Pech_PE_Scaf11856_0.9
Lluy_FUN_011466
Pech_PE_Scaf8003_1.4
Pech_PE_Scaf506_3.15
Pech_PE_Scaf11296_5.6
Pech_PE_Scaf6695_1.0
Pech_PE_Scaf5612_0.5
Rpac_jg12844
Pech_PE_Scaf6695_2.10
Pech_PE_Scaf8003_1.5
Pech_PE_Scaf506_3.16
Lsat_s032.g0145
Lsat_s032.g0147
Lluy_FUN_038580
Lluy_FUN_017044
Lsat_s032.g0144
Lluy_FUN_038578
Lsat_s032.g0143
Lsat_s032.g0146
Lluy_FUN_038581
Lluy_FUN_038579
Lluy_FUN_038576
Lluy_FUN_038573
Lluy_FUN_021506
Lsat_s032.g0140
Lsat_s032.g0141
Lluy_FUN_038570
Pech_PE_Scaf1050_0.2
Pech_PE_Scaf9630_0.7
Pech_PE_Scaf1838_0.14
Pech_PE_Scaf11116_2.5
Rpac_jg25582
Rpac_jg25589
Polf_A7X406
Tjac_A7X3Z0
Lmad_A7X401
C-type lectin lectoxin
Tnat_Q66S03
Galactose-specific lectin nattectin
Omyk_P81018
Ladderlectin
Hsap_Q8WXI8
Mmus_Q9Z2H6
Mmus_Q9R0Q8
Hsap_Q9ULY5
C-type lectin domain family 4 member D, E
Hlae_P82596
Perlucin
Ctel_CapteP80827
Ctel_CapteP136146
Lsat_s041.g0199
Rpac_jg409
Pech_PE_Scaf9098_1.8
Dgyr_CAD5122556.1
Eand_GWHPACBE000897
Hrob_HelroP181719
0.3
B
Lsat_s019.g0025
Lsat_s024.g0045
Lsat_s024.g0046
Lsat_s024.g0288
Lsat_s024.g0289
Lsat_s024.g0448
Lsat_s032.g0140
Lsat_s032.g0141
Lsat_s032.g0143
Lsat_s032.g0144
Lsat_s032.g0145
Lsat_s032.g0146
Lsat_s032.g0147
Lsat_s041.g0199
Lsat_s060.g0102
Lsat_s066.g0077
Lsat_s069.g0100
* *
Obt
Ves
Tru
Expression levels
(row Z-score transformed)
-1.5
-1
-0.5
0
0.5
1
1.5

Supplementary Figure S14.

Molecular phylogenetic tree and expression levels of C-type lectin genes. (A) Molecular phylogenetic tree of the C-type lectin family in the genomes of four vestimentiferans. The maximum likelihood phylogenetic tree was constructed using aligned and gap-trimmed amino acid sequences. For genes retrieved from the UniProt database, taxonomy and accession numbers are shown. *L. satsuma* genes are highlighted in red. Circles on branches indicate bootstrap values higher than 80%. (B) Heat map of expression levels (row Z-score transformed) of *L. satsuma* genes belonging to the C-type lectin family.

Lsat, *L. satsuma*; Lluy, *L. luymesi*; Rpac, *R. pachyptila*; Pech, *P. echinospica*; Ctel, *C. teleta*; Hrob, *Helobdella robusta*; Eand, *Eisenia andrei*, Dgyr; *Dimorphilus gyrociliatus*; Polf, *Philodryas olfersii* (Green snake); Tjac, *Thrasops jacksonii* (Jackson's black tree snake); Lmad, *Leioheterodo*n *madagascariensis* (Malagasy giant hognose snake); Tnat, *Thalassophryne nattereri* (Copper Joe toadfish); Omyk, *Oncorhynchus mykiss* (Rainbow trout); Hlae, *Haliotis laevigata* (Smooth Australian abalone); Hsap, *Homo sapiens*; Mmus, *Mus musculus*; Obt, obturacular region; Ves, vestimental region; Tru, trunk region; **, region-specific DEGs.


A
Ctel_CapteP117696
Ctel_CapteP119007
Hsap_Q7RTY7
Ovochymase
Ctel_CapteP18116
Ctel_CapteP224827
Rpac_jg32095
Lsat_s008.g0001
Rpac_jg27624
Lsat_s046.g0019
Rpac_jg20997
Rpac_jg20996
Hrob_HelroP190772
Hrob_HelroP78131
Hrob_HelroP110109
Lsat_s017.g0422
Lsat_s017.g0427
Rpac_jg22225
94
Rpac_jg31942
Rpac_jg31947
Rpac_jg31948
Lsat_s017.g0423
Rpac_jg18871
Lsat_s017.g0434
Lsat_s017.g0432
Lsat_s003.g0025
Lsat_s017.g0437
Rpac_jg31934
Rpac_jg31936
Rpac_jg31938
Lsat_s017.g0428
Rpac_jg31941
Hrob_HelroP140768
Hsap_P08861
Hsap_Q99895
Chymotrypsin
Hsap_P08217
Hsap_P08218
Chymotrypsin-like elastase
Ctel_CapteP155647
Ctel_CapteP159950
Ctel_CapteP21170
Ctel_CapteP104129
Ctel_CapteP104173
Hrob_HelroP185620
Hrob_HelroP191905
Hrob_HelroP185618
Hrob_HelroP185617
Ctel_CapteP151454
Ctel_CapteP225851
Ctel_CapteP194189
Ctel_CapteP220869
Hrob_HelroP159293
Ctel_CapteP153991
Ctel_CapteP94302
Hrob_HelroP185899
Hrob_HelroP186312
Hrob_HelroP185900
Hrob_HelroP177891
Hrob_HelroP185895
Hrob_HelroP193208
Hrob_HelroP157379
Ctel_CapteP172265
Ctel_CapteP226412
Lsat_s001.g0318
Rpac_jg7154
Ctel_CapteP202018
Rpac_jg2846
Lsat_s001.g0455
Ctel_CapteP221460
Hrob_HelroP157319
Lsat_s065.g0056
Rpac_jg15275
Ctel_CapteP160020
Hrob_HelroP97488
Hsap_Q7RTZ1
Dmel_B7YZU2
Ctel_CapteP103892
Ctel_CapteP190556
Ctel_CapteP113671
Ctel_CapteP19592
Ctel_CapteP210211
Ctel_CapteP226718
Ctel_CapteP121428
Ctel_CapteP226720
Ctel_CapteP226719
Ctel_CapteP153798
Hrob_HelroP114601
Hrob_HelroP185962
Hrob_HelroP155168
Hrob_HelroP185960
Hrob_HelroP178982
Dmel_C0HKA2
Dmel_C0HKA3
Insect trypsin
Ctel_CapteP213986
Hsap_P17538
Chymotrypsinogen
Hsap_P07478
Hsap_P07477
Hsap_P35030
Mammal trypsin
0.3
B
Ctel_CapteP109753
Hrob_HelroP78348
Ctel_CapteP109743
Hrob_HelroP78293
Ctel_CapteP59065
Hrob_HelroP78542
Ctel_CapteP109758
Hrob_HelroP191066
Ctel_CapteP69048
Hrob_HelroP137837
Ctel_CapteP109708
Ctel_CapteP86846
Hsap_Q8WZ42
Mmus_A2ASS6
Mammal titin
Hrob_HelroP170423
Rpac_jg29247
Pech_PE_Scaf4360_8.2
Dmel_Q9I7U4
Insect titin
Lluy_FUN_020455
Ctel_CapteP196990
Hrob_HelroP95831
Ctel_CapteP60866
Hrob_HelroP170138
Hrob_HelroP111318
Ctel_CapteP59383
Ctel_CapteP77799
Pech_PE_Scaf8216_18.2
Lsat_s005.g0136
Lluy_FUN_040904
Ctel_CapteP196991
Mmus_Q7TPW1
Mmus_Q0ZGT2
Mammal nexin
0.5

Supplementary Figure S15.

(A) Molecular phylogenetic tree of chymotrypsin-like serine protease genes.

(B) Molecular phylogenetic tree of titin-like protein genes.

The maximum likelihood phylogenetic tree was constructed using aligned and gap-trimmed amino acid sequences. For genes retrieved from the UniProt database, taxonomy and accession numbers are shown. *L. satsuma* genes are highlighted in red. Circles on branches indicate bootstrap values higher than 80%. Lsat, *L. satsuma*; Rpac, *R. pachyptila*; Ctel, *C. teleta*; Hrob, *Helobdella robusta*; Hsap, *Homo sapiens*; Mmus, *Mus musculus*; Dmel, *Drosophila melanogaster.*

Supplementary table footnotes

Supplementary Table S1.

Statistics of gene models used in this study. For gene models downloaded from the NCBI RefSeq or GenBank database, accession numbers are shown. Only the longest transcript variants of each gene were counted. The longest transcript variants of each gene were translated into amino acid sequences, and then used for the analysis with BUSCO v.5.3.2.

S, complete and single-copy BUSCOs; D, complete and duplicated BUSCOs.

Supplementary Table S2.

Statistics of RNA-seq, quality trimming, and mapping data.

Supplementary Table S3.

Read counts for RNA-seq quantified with Salmon v.1.8.0

Supplementary Table S4.

TPMs of RNA-seq samples calculated with EdgeR v.3.36.0.

Supplementary Table S5.

CPMs of RNA-seq samples calculated with EdgeR v.3.36.0.

Supplementary Table S6-8

Pairwise comparison of CPMs between obturacular and vestimental region, obturacular and trunk region, and trunk and vestimental region, respectively.

Supplementary Table S9.

Statistics of the region-specific DEGs.

References

1. Simakov, O., Marletaz, F., Cho, S.-J., et al. 2013, Insights into bilaterian evolution from three spiralian genomes. *Nature*, **493**, 526–31.
2. Albertin, C. B., Simakov, O., Mitros, T., et al. 2015, The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature*, **524**, 220–4.
3. Sun, Y., Sun, J., Yang, Y., et al. 2021, Genomic Signatures Supporting the Symbiosis and Formation of Chitinous Tube in the Deep-Sea Tubeworm *Paraescarpia echinospica*. *Mol. Biol. Evol.*, **38**, 4116–34.
4. de Oliveira, A. L., Mitchell, J., Girguis, P., and Bright, M. 2022, Novel Insights on Obligate Symbiont Lifestyle and Adaptation to Chemosynthetic Environment as Revealed by the Giant Tubeworm Genome. *Mol. Biol. Evol.*, **39**, msab347.
5. Fröbius, A. C., and Seaver, E. C. 2006, *ParaHox* gene expression in the polychaete annelid *Capitella* sp. I. *Dev. Genes Evol.*, **216**, 81–8.
6. Brooke, N. M., Garcia-Fernàndez, J., and Holland, P. W. H. 1998, The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature*, **392**, 920–2.