Identification of a unique lamprey gene with tandemly repeated sequences and pharyngeal chondrocyte-specific expression

Identification of the unique gene with a tandemly-repeated sequences from lamprey

which shows specific expression in the pharyngeal chondrocytes

Hiromasa Yokoyama, Yoshiaki Morino and Hiroshi Wada*

Graduate School of Life and Environmental Sciences, University of Tsukuba

* Author for correspondence: H. Wada

Graduate School of Life and Environmental Sciences, University of Tsukuba

1-1-1 Tennodai, Tsukuba 305-8572 JAPAN

e-mail: hwada@biol.tsukuba.ac.jp

tel and fax: +81-29-853-4671

Abstract

Recent studies on invertebrate cartilage proposed that the origin of the vertebrate cartilage is in the chondroid connective tissues of invertebrate ancestors. A further issue was raised regarding vertebrate cartilage evolution—whether the evolution of cartilage from the chondroid connective tissues occurred in the common ancestors of vertebrates. Alternatively, the evolution of cartilage may have occurred independently in agnathans and in gnathostomes, because extant agnathans (cyclostomes) are known to possess a matrix composition distinct from that of gnathostomes. Here, we identified the gene which is likely to encode the second type of matrix protein unique to lamprey cartilage, pharymprin. *Pharymprin* shows specific expression in larval pharyngeal chondrocytes. Similar to those of lamprins, the known matrix proteins of the lamprey trabecular cartilage, pharymprin is also composed of repeat sequences. However, the repeat sequence is quite distinct from that of lamprins. The presence of two distinct matrix proteins in lamprey cartilage may be more consistent with the hypothesis that the true cartilage evolved independently in cyclostomes and gnathostomes.

Key words: lamprey, cartilage, pharyngeal arch, matrix, pharymprin

Introduction

Vertebrates have four skeletal tissue types: cartilage, bone, enamel, and dentine (Hall, 2005). Evolution of these skeletons contributed to the elaboration of the unique body plan and life strategies of the vertebrates. However, the evolutionary sequence of the emergence of these skeletal tissues is debatable (Donoghue et al., 2006). Although extant species of Agnatha, the basal group of vertebrates, lack mineralized skeletal tissues, such as bone, enamel, or dentine, fossil species of agnathans did possess mineralized skeletal tissues as their dermal skeleton (Hardisty, 1981; Janvier, 1996; Donoghue et al., 2006).

Homology between the cartilage of extant Agnatha, such as lamprey and hagfish, with that of gnathostomes, is supported by their shared genetic machinery. For both types of skeleton, differentiation is regulated by the group E Sox genes (*Sox8*, *9*, and *10* of amniotes; McCauley and Bronner-Fraser, 2006; Zhang et al., 2006; Ohtani et al., 2008). However, *SoxE* regulation of cartilage differentiation may be traced further back, to the ancestral chordates (Kaneto and Wada, 2011; Jandzik et al., 2015), or possibly to the common ancestors of bilaterians (Tarazona et al., 2016). These studies highlight the alternative hypothesis that cartilage is not a novel characteristic of

vertebrates, but the origin of the cartilage date back to the common ancestors of

bilaterians. This reminded us the hypothesis of Cole and Hall (2004a), proposing that

the cartilage evolved from chondroid connective tissue of the metazoan ancestors. Cole

and Hall (2004a) defined the cartilage as rigid connective tissue composed of large cells

(chondrocytes) distinct from other connective tissue cells. The chondrocytes are

embedded within an extracellular matrix of varying abundance that has high amounts

of fibrous protein. According to this criteria, they proposed that cartilage evolved

convergently in vertebrates and cephalopods (Cole and Hall, 2004a; Cole and Hall,

2004b; Tarazona et al., 2016). This hypothesis raises another issue—whether the

evolution from the ancestral chondroid connective tissue to the true cartilage occurred

in the common ancestors of agnathans and gnathostomes, or whether it occurred

independently in each of these lineages. The latter issue stems from the differences in

the biochemical nature of the cartilage of gnathostomes and that of extant agnathans,

the cyclostomes (Wright et al., 1983).

Whereas fibrillar collagen (type II collagen) is the most abundant matrix in the

gnathostome cartilage, it is a rather minor component of cyclostome cartilage, and

present only in the peripheral catilage (Wright et al., 1983; Hall, 2005). Additionally,

the next major matrix component of gnathostome cartilage, aggrecan, is not encoded in

the lamprey genome (Smith et al., 2018). Instead, cyclostome cartilage is known to have

a unique matrix composition, where the abundant matrix protein is more similar to

elastin and insoluble to CNBr (Wright et al., 1983).

In the case of lamprey, the elastin-like molecule lamprin was identified as the

major matrix component (44-51% of dry weight) of annular cartilage (Wright et al.,

1983; Robson et al., 1993; Robson et al., 1997; Robson et al., 2000). Lamprin has several

repeats of GGLGY, which are flanked by short unique sequences (Robson et al., 1993;

Robson et al., 2000). Although the repeat sequence is similar to that of avian egg shell

protein or insect amniote proteins the repeat sequence evolved convergently in each

lineage (Robson et al., 2000). Lamprin function as the major matrix of other cartilage,

such as trabecular and piston cartilage, whereas the major matrix component of

branchial, pericardial, and nasal cartilages remains to be characterized (Robson et al.,

1997). Hagfish are also known to possess a unique cartilage matrix, myxinine; however,

their amino acid sequence has not yet been characterized (Wright et al., 1984).

To understand the early evolution of the vertebrate skeleton, molecular

characterizations of cyclostome cartilages are of critical importance. Here, we report the

molecular identification of the matrix component of the lamprey pharyngeal arch. Our

study is based on the pioneering study by Wright et al. (2001), who reported the

N-terminal sequence of the matrix of pharyngeal cartilage. By referring to this sequence,

we identified the novel gene which is likely to encode the matrix protein of the lamprey

pharyngeal cartilage.

Materials and Methods

Animals

Adult Arctic lampreys (*Lethenteron camtschaticum*) were collected from the

Shiribeshi-Toshibetsu River, Hokkaido, Japan. Mature eggs were retrieved from

females that were anesthetized with ethyl 3-aminobenzoate methanesulfonate (MS-222)

and fertilized *in vitro*. Embryos were cultured at 16°C. Developmental stages were

designated based on a previous report (Tahara, 1988). Adult specimens of *L.* sp. N were

collected from the Kamo River, Toyama, Japan (Yamazaki and Goto, 1998; Yamazaki et

al., 2006).

Molecular characterization of pharymprin

We performed transcriptome analysis on samples of stage (st.) 25-26 larvae. Total RNA

was extracted using Trizol reagent (Life Technologies, Carlsbad, CA, USA), treated with

DNase, and cleaned using the RNeasy kit (Qiagen, Hilden, Germany). Construction of

paired-end (200 bp), duplex-specific nuclease normalized library and sequencing on the

Hiseq4000 platform (Illumina) were performed by the Beijing Genomics Institute. The

raw reads were deposited in the DDBJ Sequence Reads Archives (XXXXXXXXXXX).

The low-quality reads were filtered and trimmed with Trimmomatic (v0.33; Bolger et al., 2014). *De novo* assembly was conducted using Trinity (trinity_r20140707; Grabherr et al., 2011).

From the assembled gene models, we searched for the N-terminal sequence of the matrix protein in the lamprey pharyngeal cartilage (Wright et al., 2001). Among the three amino acid sequence, GAGADVCGAP and YAPGPG did not retrieve sequence by BLAST search. By using the longest sequence (AGPGGYGQGGGGPGGYGPG), we retrieved a single gene model. By referring to the assembled sequence, we designed a pair of primers with which we amplified cDNAs of pharymprins from *L. camtschaticum* and *L.* sp. N (Genbank Acc.Nos. LC413899, and LC413900).

In situ hybridization

RNA probes for *in situ* hybridization were made using PCR fragments. We prepared template DNA for probe synthesis by using a reverse primer with a 20-bp T3 promoter sequence. *In situ* hybridization was performed as described by Ogasawara et al. (2000) with some minor modifications.

Results and Discussion

Molecular characterization

Wright et al. (2001) partially characterized the CNBr insoluble matrix protein of the

Petromyzon marinus pharyngeal cartilage. They determined partial amino acid

sequences of peptides generated by thermolysin digestion of CNBr-insoluble protein. By

referring to one of the amino acid sequences of the CNBr-insoluble protein from

pharyngeal cartilage (AGPGGYGQGGGGGPGGYGPG; Wright et al., 2001), we retrieved

a gene model sequence by BLAST search from our transcriptome data of st. 26 embryo

of *Lethenteron camtschaticum*. Because the amino acid sequence contains several

repeat sequences, we thought we should avoid the risk of using the wrong gene model.

Thus, we performed RT-PCR from the RNA of st. 25-26 larvae using the primers

designed from the gene model (Fig. 1). The resultant sequence is two repeats shorter

than the predicted gene model of the transcriptome (Fig. 1). Next, we examined the

amino acid composition of pharymprin after a conceptual treatment with CNBr that

cleaves proteins at methionine residue. The Amino acid composition of the pharymprin

of *L. camtschaticum* is reasonably fit with that reported for the CNBr-insoluble matrix

of *P. marinus* (Table 1). These evidences support that the isolated gene is likely to

encode the CNBr-insoluble matrix protein of pharyngeal cartilage. Thus, we designated

this gene as *pharymprin* (pharyngeal cartilage matrix distinct from lamprin).

Pharymprin of *L. camtschaticum* possessed 28 repeats of

GG(F/Y)(G/A)P(G/A)GGP with some variations (Fig. 1); these are flanked by unique

sequences in the N-terminus and C-terminus. The unit sequence GGY(G/A)P(G/A)GGP

is distinct from that of lamprin: GGLGY(Robson et al., 1993; Robson et al., 2000). By

comparing it to the genome sequence of *L. camtschaticum*

(http://jlampreygenome.imcb.a-star.edu.sg/), we extrapolated the genome structure of

pharymprin and found that the coding sequence did not contain any introns.

Expression of pharymprin during development

To determine whether *pharymprin* is expressed as part of the pharyngeal cartilage

matrix, we performed *in situ* hybridization using larvae. No expression was detected in

specimens from st. 24 to 27. The earliest signal was detected from larvae of st. 28 in the

chondrocytes of pharyngeal arch 3 and 4 (Fig. 2A). The signal was detected in the

middle of the cartilage rod along the dorsal–ventral axis. (Yao et al., 2008) described the

development of the lamprey pharyngeal cartilage using histological Weigert staining.

The patterns of *pharymprin* expression and Weigert staining are similar with respect to

the first signal that arises from the middle of the cartilage rod along the dorsal–ventral

axis.

During development, the pharymprin expression emerges in chondrocytes of

more posterior pharyngeal arches (Fig. 3A-E). By the end of st. 28, the *pharymprin*

expression is also detected in the epitrematic process (EP), hypotrematic bars(HTB), but

not in the subchordal rods (Fig. 3E, F). Note that matrix secretion for the subchordal rod

was visualized by weigert staining in 50-day-old larvae, but not in st. 30 larvae; namely,

the matrix secretion is later in subchordal rods than the other pharyngeal skeletal

elements (Yao et al., 2008). In stage 29, the signal becomes weaker in the middle part

(along the dorso–ventral axis) of the rod from the anterior pharyngeal arch (Fig. 3F). In

stage 30, the expression was also detected in hypobranchial bars (HBB). Although

trabecular and parachordal cartilage was visualized in st. 30 larvae using Weigert

staining (Yao et al., 2008), the pharymprin expression was not observed in trabecular

and parachordal cartilage of younger than st. 30 (Fig. 3E, F). This is consistent with

lamprin expression in trabecular cartilage (McBurney et al., 1996).

Molecular evolution of the *pharymprin*

To trace the molecular evolutionary process of the *pharymprin*, we then searched for the *pharymprin* sequence in *L.* sp. N and *P. marinus* (Fig. 3). *Pharymprin* cDNA from *L.* sp. N, amplified from RNA of ammocoete larvae by using the primers for *L. camtschaticum,* showed an almost identical sequence to that of *L. camtschaticum* (Fig. 3). Although some amino acid substitutions were observed in repeat sequences, the number of repeats was the same. In the current gene models of *P. marinus*, we did not retrieve the full cDNA sequence of pharymprin (Smith et al., 2018). However, absence of introns in the *L. camtschaticum* pharymprin encouraged us to examine the genome sequence assembly of *P. marinus*. We retrieved almost a full sequence of the pharymprin encoding region. Because we did not find the first methionine (Fig. S1), the *pharymprin* gene of *P. marinus* is likely to possess introns in the N-terminus. We found that the overall structure of the pharymprin molecule of *P. marinus* was similar to that of *Lethenteron*. In the genome sequence of *P. marinus pharymprin*, a unique insert was found in the C-terminal region (shaded in Fig. 3). We cannot, however, exclude the possibility that this region may be spliced out in mRNA, because the insert started from GT and ended with AG (Fig. S1).

We searched for similar sequences from the genomes (unmasked assembly

sequences) of human, mouse, frog, zebrafish, amphioxus, fly, nematode, limpet,

polycheate and jelly fish, but we did not find any.

Implication for the evolution of vertebrate cartilage

Here, we reported the molecular characterization of the gene: *pharymprin*, which is

likely to encode the second matrix protein unique to lamprey cartilage. Although both

lamprin and pharymprin consist of repeat sequences, their repeat unit is quite different:

GGLGY for lamprin and GG(Y/F)(G/A)AAGGP for pharymprin. Although we cannot

exclude the possibility that *lamprin* and *pharymprin* evolved from a single ancestral

gene, we prefer the idea that lampreys evolved two distinct molecules for the evolution

of their cartilage, because of the distinct sequence of the repeat unit. The gnathostomes

kept type II collagen as their main matrix molecule since the divergence of extant

cartilaginous fish and the rest of the gnathostomes (Hall, 2005). In addition, aggrecan

was recruited as the second major matrix component in gnathostomes. On the other

hand, we present an evidence that the lamprey possess two matrix protein for their

cartilage, and they are likely to have evolved independently. In addition, Robson et al.

(1997) suggested a third type of matrix molecule in the lamprey notochord sheath and fins. These observations may favor the idea that the evolution from the ancestral chondroid connective tissue to the cartilage occurred independently in cyclostomes and gnathostomes. We note that the hypothesis of independent evolution of cartilage is not unexpected considering that the acquired immune system of lamprey and gnathostomes developed using distinct molecular mechanisms (Pancer et al., 2004). This hypothesis should be examined using a more comprehensive comparison of the matrix molecules of cyclostomes and gnathostomes.

References

Bolger, A.M., Lohse, M. and Usadel, B., 2014. Trimmomatic: a flexible trimmer for
Illumina sequence data. Bioinformatics 30, 2114-2120.

Cole, A.G. and Hall, B.K., 2004a. Cartilage is a metazoan tissue; integrating data from
nonvertebrate sources. Acta Zool. 85, 69-80.

Cole, A.G. and Hall, B.K., 2004b. The nature and significance of invertebrate cartilages
revisited: distribution and histology of cartilage and cartilage-like tissues within
the Metazona. Zoology 107, 261-273.

Donoghue, P.C.J., Sansom, I.J. and Downs, J.P., 2006. Early evolution of vertebrate

skeletal tissues and cellular interactions, and the cannalization of skeletal

development. J. Exp. Zool. 306B, 278-294.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Fan, L.,

Raychowdhury, R. and Zeng, Q., 2011. Full-length transcriptome assembly from

RNA-Seq data without a reference genome. Nat. Biotech. 29, 1546-1696.

Hall, B.K., 2005. Bone and Cartilage, Elsevier, San Diego.

Hardisty, M.W., 1981. The skeleton, in: Hardisty, M.W. and Potter, I.C. (Eds.), The

Biology of Lampreys, vol. 3. Academic Press, New York, pp. 333-376.

Jandzik, D., Garnett, A.T., Square, T.A., Cattell, M.V., Yu, J.K. and Medeiros, D.M.,

2015. Evolution of the new vertebrate head by co-option of an ancient chordate

skeletal tissue. Nature 518, 534-7.

Janvier, P., 1996. Early vertebrates. Oxford University Press, Oxford.

Kaneto, S. and Wada, H., 2011. Regeneration of amphioxusoral cirri and its skeletal

rods: implications for the origin of the vertebrate skeleton. J. Exp. Zool. 316B,

409-417.

McBurney, K.M., Keely, F.W., Kibenge, F.S.B. and Wright, G.M., 1996. Spatial abd

temporal distribution of lamprin mRNA during chondrogenesis of trabecular

cartilage in the sea lamprey. Anat. Embryol. 193, 419-426.

McCauley, D.W. and Bronner-Fraser, M., 2006. Importance of SoxE in neural crest

development and the evolution of the pharynx. Nature 441, 750-752.

Ogasawara, M., Shigetani, Y., Hirano, S., Satoh, N. and Kuratani, S., 2000.

Pax1/Pax9-related genes in an Agnathan vertebrate, Lampetra japonica:

expression pattern of LjPax9 implies sequential evolutionary events toward the

gnathostome body plan. Dev. Biol. 223, 399-410.

Ohtani, K., Yao, T., Kobayashi, M., Kusakabe, R., Kuratani, S. and Wada, H., 2008.

Expression of Sox and fibrillar collagen genes in lamprey larval chondrogenesis

with implications for the evolution of vertebrate cartilage. J. Exp. Zool. 310B,

596-607.

Pancer, Z., Amemiya, C.T., Ehrhardt, G.R.A., Ceitlin, J., Gartland, G.L. and Cooper,

M.D., 2004. Somatic diversification of variable lymphocyte receptors in the

agnathan sea lamprey. Nature 430, 174-180.

Robson, P., Wright, G., Sitarz, E., Maiti, A., Rawat, M., Youson, J. and Keeley, F., 1993.

Characterization of lamprin, an unusual matrix protein from lamprey cartilage.

Implications for evolution, structure, and assembly of elastin and other fibrillar

proteins. J Biol Chem. 268, 1440-1447.

Robson, P., Wright, G.M., Youson, J.H. and Keely, F.W., 1997. A family of

non-collagen-based cartilages in the skeleton of the sea lamprey, *Petromyzon*

marinus. Comp.Biochem. Physiol. 118B, 71-78.

Robson, P., Wright, G.M., Youson, J.H. and Keely, F.W., 2000. The structure and

organization of lamprin genes: multiple-copy genes with alternative splicing and

convergent evolution with insect structural proteins. Mol. Biol. Evol. 17,

1739-1752.

Smith, J.J., Timoshevskaya, N., Ye, C., Holt, C., Keinath, M.C., Parker, H.J., Cook, M.E.,

Hess, J.E., Narum, S.R., Lamanna, F., Kaessmann, H., Timoshevskiy, V.A.,

Waterbury, C.K.M., Saraceno, C., Wiedemann, L.M., Robb, S.M.C., Baker, C.,

Eichler, E.E., Hockman, D., Sauka-Spengler, T., Yandell, M., Krumlauf, R., Elgar,

G. and Amemiya, C.T., 2018. The sea lamprey germline genome provides insights

into programmed genome rearrangement and vertebrate evolution. Nat. Genet. 50, 270-277.

Tahara, Y., 1988. Normal stages of development in the lamprey, *Lampetra reissneri* (Dybowski). Zool. Sci. 5, 109-118.

Tarazona, O.A., Slota, L.A., Lopez, D.H., Zhang, G. and Cohn, M.J., 2016. The genetic program for cartilage development has deep homology within Bilateria. Nature 533, 86-9.

Wright, G.M., Keeley, F.W. and Robson, P., 2001. The unusual cartilagenous tissues in jawless craniates, cephalochordates and invertebrates. Cell Tissue Res. 304, 165-174.

Wright, G.M., Keeley, F.W. and Youson, J.H., 1983. Lamprin: a new vertebrate protein comprising the major structural protein of adult lamprey cartilage. Experientia 39, 495-497.

Wright, G.M., Keely, F.W., Youson, J.H. and Babineau, D.L., 1984. Cartilage in the Atlantic Hagfish, *Myxine glutinosa*. Am. J. Anat. 169, 407-424.

Yamazaki, Y. and Goto, A., 1998. Genetic structure and differentiation of four

Lethenteron taxa from the Far East, deduced from allozyme analysis. Env. Biol.

Fish. 52, 149-161.

Yamazaki, Y., Yokoyama, R., Nishida, M. and Goto, A., 2006. Taxnomy and molecular

phylogeny of *Lethenteron* lampreys in eastren Eurasia. J. Fish Biol. 68, 251-269.

Yao, T., Ohtani, K. and Wada, H., 2008. Whole-mount observation of pharyngeal and

trabecular cartilage development in lampreys. Zool. Sci. 25, 976-981.

Zhang, G., Miyamoto, M.M. and Cohn, M.J., 2006. Lamprey type II collagen and *Sox9*

reveal an ancient origin of the vertebrate collagenous skeleton. Proc. Natl. Acad.

Sci. USA 103, 3180-3185.

Legends for Figures

Figure 1.   Nucleotide and predicted amino acid sequence of pharymprin from *L.*

camtschaticum. Repeat sequences are highlighted by boxes. Nucleotide sequences

which is retrieved from the assembled gene model of transcriptome, but not confirmed

by the RT-PCR products were double-underlined.

Figure 2. Expression of pharymprin in larvae of *L. camtschaticum.* (A) Expression was

detected in the skeletal rods of third and fourth pharyngeal arches (PA3 and PA4) from

st. 28. (B-D) During st. 28, the expression emerges in more posterior arches. (E) Before

the end of st. 28, expression comes out in the epitrematic process (EP), hypotrematic

bars (HTB). (F) At stage 29, the signal becomes weaker in the middle part (along the

dorso–ventral axis) of the rod from the anterior pharyngeal arch. (G) Ventral view of

yiew of st. 30 larva, showing the expression in hypobranchial bars (HBB). (H, I) Sagital

(H) and transverse (I) section of st. 28 larva shown as (h, i) in (C). Scale bars: 100 μm in

A-H, and 50 μm in I.

Figure 3. Alignment of the amino acid sequences of pharymprin from *L. camtschaticum,*

L. sp. N and *P. marinus.*   Repeat sequences are boxed. Sequences possibly spliced out

from the *P. marinus* sequence was shaded.

Table 1. Amino acid composition of the CNBr-insoluble protein from pharyngeal cartilage and phary

	Branchial cartilage (Robson et al. 1997)	%	Lc_pharymprin	%	Pm_pharymprin	%
ASX	66	6.6	9	2.0	10	2.4
THR	14	1.4	3	0.7	2	0.5
SER	43	4.3	18	4.0	16	3.8
GLX	60	6.0	9	2.0	9	2.1
PRO	131	13.1	69	15.3	64	15.1
GLY	323	32.3	168	37.3	156	36.9
ALA	169	16.9	96	21.3	88	20.8
VAL	30	3.0	7	1.6	9	2.1
MET	<2.0	<0.2	1	0.2	0	0.0
ILU	15	1.5	1	0.2	2	0.5
LEU	31	3.1	5	1.1	6	1.4
TYR	47	4.7	32	7.1	32	7.6
PHE	16	1.6	7	1.6	6	1.4
LYS	13	1.3	9	2.0	8	1.9
HIS	14	1.4	3	0.7	2	0.5
ARG	29	2.9	13	2.9	13	3.1
Total	1000	100	450	100	423	100

mprin

```
ATGGTGATTGCACACCTGCTAACCTTCTCTCTCGTTTCATCAGGATATGGTCAAGGGGCA          60
 M  V  I  A  H  L  L  T  F  S  L  V  S  S  G  Y  G  Q  G  A           20

GGAGCTGGGCTTCCTTCAAGAACAGCTGGCCCTGGGGGGTACGGTCAAGGGGGTGCTGGA         120
 G  A  G  L  P  S  R  T  A  G  P  G  G  Y  G  Q  G  G  A  G           40

CCTGGGGGGTACGGCCCAGGGGGCAGGAGGTTACGGTCCAGGGGGGAGCAAAGGCCGGTAAC       180
 P  G  G  Y  G  P  G  A  G  G  Y  G  P  G  G  A  K  A  G  N           60

GGTGCAGGCGCAGGCAGGCCCAATGGATACGGCCCAGCGCTAGGAGGGCCAGGGGCATTT        240
 G  A  G  A  G  R  P  N  G  Y  G  P  A  L  G  G  P  G  A  F           80

GTGCCCAGGGCAAGGCCAAGTGTGTACGTTGCTGCGCCAAGTAGAGCTAATGGATATGGT        300
 V  P  R  A  R  P  S  V  Y  V  A  A  P  S  R  A  N  G  Y  G          100

CCAGGAGGAAGCAAAGCTGGTGGTCCAGGAGCTGGGGGACCTGGGGGTTATGGTCCAGGA        360
 P  G  G  S  K  A  G  G  P  G  A  G  G  P  G  G  Y  G  P  G          120

GCAGGAAGGCCTGGTGGCTATGCACCAGCAGCAGGCGGGCCTGGTGGCTATGCACCAGCA        420
 A  G  R  P  G  G  Y  A  P  A  A  G  G  P  G  G  Y  A  P  A          140

GCAGGAGGGCCTGGTGGCTATGCACCAGCAGCAGGAGGACCTGGTGGCTATGCACCAGCA        480
 A  G  G  P  G  G  Y  A  P  A  A  G  G  P  G  G  Y  A  P  A          160

GCAGGAGGACCTGGTGGTTTTGGCCCAGCAGCAAGTAAAGCCGGTGGTTATGGTCCAGGG        540
 A  G  G  P  G  G  F  G  P  A  A  S  K  A  G  G  Y  G  P  G          180

GCTGGGGGGACCTGGGGGGTTATGGTCCAGCAGCAGGAGGACCTGGTGGCTATGCACCAGCA       600
 A  G  G  P  G  G  Y  G  P  A  A  G  G  P  G  G  Y  A  P  A          200

GCAGGAGGGCCTGGTGGCTATGCACCAGCAGCAGGCGGGCCTGGTGGCTATGCACCAGCA        660
 A  G  G  P  G  G  Y  A  P  A  A  G  G  P  G  G  Y  A  P  A          220

GCAGGAGGACCTGGTGGTTTTGGCCCAGCAGCAAGTAAAGCCGGTGGTTATGGTCCAGGG        720
 A  G  G  P  G  G  F  G  P  A  A  S  K  A  G  G  Y  G  P  G          240

GCTGGAGGACCTGGTGGCTATGCACCAGCAGCAGGCGGGCCTGCTGGTTATGCACCGGCA        780
 A  G  G  P  G  G  Y  A  P  A  A  G  G  P  A  G  Y  A  P  A          260

GCAGGAGGGCCTGGCGGTTATGGTCCAGCGGCAGGCGGGGCTAATGGATATGGCCCTGGG        840
 A  G  G  P  G  G  Y  G  P  A  A  G  G  A  N  G  Y  G  P  G          280

GCTAGTAAAGCGGCAGGAGGACCCGGCGGTTATGCACCGGCAGCAGGAGGACCTGGACAG        900
 A  S  K  A  A  G  G  P  G  G  Y  A  P  A  A  G  G  P  G  Q          300

TCTGGTCCAGGGGGCAGGAGGAGCAGGCGGGAATGGGGCTGGAGGTAAGCCTAACGGACAT        960
 S  G  P  G  A  G  G  A  G  G  N  G  A  G  G  K  P  N  G  H          320

GCCCCAGCAGCCGGTGGCCAACCTGGCTTTGCACCAAGGCCTAGCGCATTTGGTCCAGGG       1020
 A  P  A  A  G  G  Q  P  G  F  A  P  R  P  S  A  F  G  P  G          340

TCAGGGGGGACCTGGATCTCATCCAGGAGCAGGAGCTGGTGGGTATGGACAAAAGGCATCT      1080
 S  G  G  P  G  S  H  P  G  A  G  A  G  G  Y  G  Q  K  A  S          360

GGCTATGCCCCAAGGGCAGGGGGACCTGGTGGTTATGCTCCAGCTGCGAGGAGGCCGAAT      1140
 G  Y  A  P  R  A  G  G  P  G  G  Y  A  P  A  A  R  R  P  N          380

GGCTATGCCCCAGCGGTAGGGAGGCAAAACGGATATGGACCGGCGGCAGGGAGACCTGGA       1200
 G  Y  A  P  A  V  G  R  Q  N  G  Y  G  P  A  A  G  R  P  G          400

TATGGTCCAAGTTCCGGAGGAGCTGGAGGACCTGCCCCAGGTGACCAGAAGCCAGCAGGT      1260
 Y  G  P  S  S  G  A  G  G  P  A  P  G  D  Q  K  P  A  G          420

GGCCAGAAGGCAGGTGGTGCTGGTGGGTACGGACCTGGGGCCGGAGGAGCAGGATACGCG       1320
 G  Q  K  A  G  G  A  G  G  Y  G  P  G  A  G  G  A  G  Y  A          440

CCCGCAGTCGCAACTCGCTCACAGTTCTACATGAAGCTCTCCTCTAAGTGTTCCCAAGG       1380
 P  A  V  A  T  R  S  Q  F  Y  M  K  L  S  S  K  C  S  Q  R          460

AGACAACTCGGTAAGAGTGCGTGA                                           1404
 R  Q  L  G  K  S  A  *                                              467
```
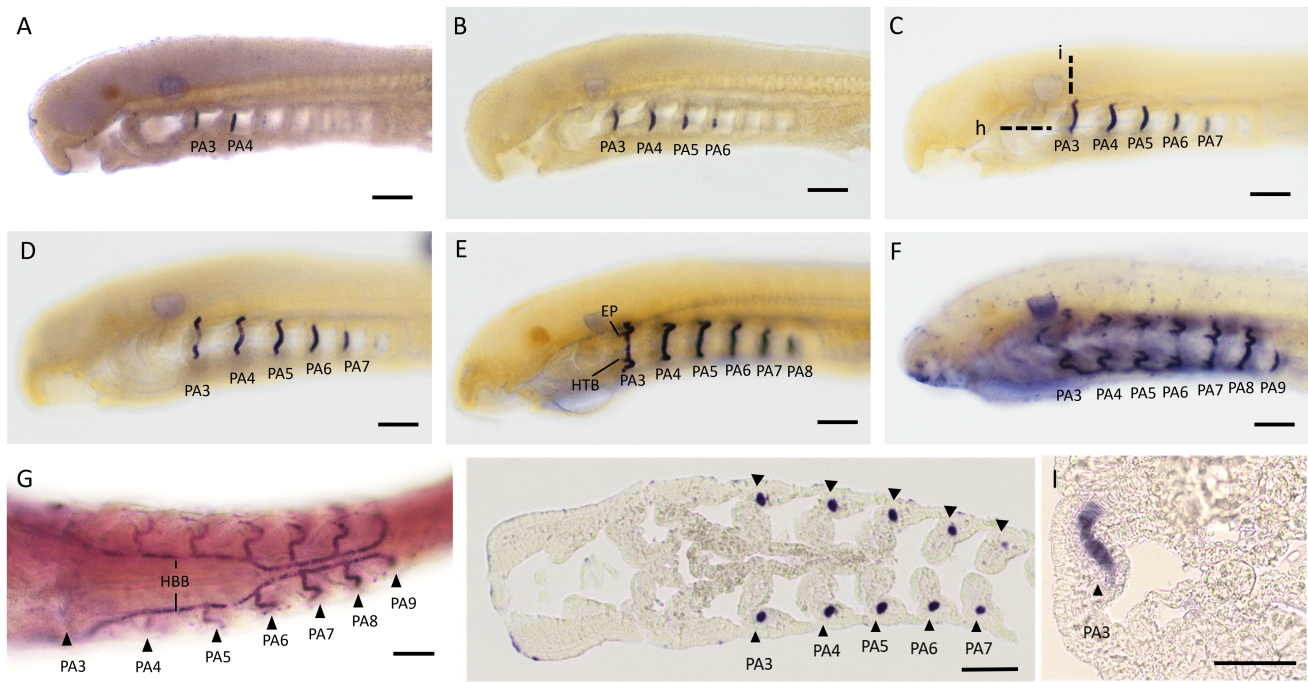
```
L. camtsyaticum   M V I A H L L T F S L V S S G Y G Q G A G A G L P S R T A G   30
L. sp. N          - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -    0
P. marinus        - V I A H L L T F Y L V S S G Y G Q G A G A A L P S R T A G   29

L. camtsyaticum   P G G Y G Q G G A G P G G Y G P G A G G Y G P G G A K A G N   60
L. sp. N          - - - - - - - - - - - - - - - - - - - - - - - - - - - - -    0
P. marinus        P G G Y G Q G G G G P G G Y G P G A G G Y G P G G A K V G N   59

L. camtsyaticum   G A G A G R P N G Y G P A L G G P G A F V P R A R P S V Y V   90
L. sp. N          - - - - - - - - - - - - - - - - - - G A F V P R A R P S V Y V   13
P. marinus        G A G A G R P N G Y G P A I G G P G A F V P R A R P S V Y V   89

L. camtsyaticum   A A P S R A N G Y G P G G S K A G G P G A G G P G G Y G P G   120
L. sp. N          A A P S R A N G Y G P G G S K A G G P G A G G P G G Y G P G   43
P. marinus        A A P S R A N G Y G P G G S K A G G P G A G G P G G Y G P G   119

L. camtsyaticum   A G R P G G Y A P A A G G P G G Y A P A A G G P - - - - - -   147
L. sp. N          A G G P G G Y G P G A G G P G G Y G P G A G G P - - - - - -   70
P. marinus        A G R P G G Y A P A A G G P G G Y A P A A G G P G G Y A P A   149

L. camtsyaticum   - - - - - - G G Y A P A A G G P G G Y A P A A G G P G G F G   168
L. sp. N          - - - - - - G G Y A P A A G G P G G Y G P A A G G P G G F G   91
P. marinus        A G G G G P G G Y A P A A G G P G G Y A P A A G G P G G F G   179

L. camtsyaticum   P A A S K A G G Y G P G A G G P G G Y G P A A G G P G G Y A   198
L. sp. N          P A A S K A G G Y G P G A G G P G G Y G P G A G G P G G Y A   121
P. marinus        P A A S K A G G Y G P G A G G P G G Y A P A A G G P G G Y A   209

L. camtsyaticum   P A A G G P G G Y A P A A G G P G G Y A P A A G G P G G F G   228
L. sp. N          P A A G G P G G Y A P A A G G P G G Y A P A A G G P G G F G   151
P. marinus        P A A G G P G G Y A P A A G G P G G Y A P A A G G - - - - -   234

L. camtsyaticum   P A A S K A G G Y G P G A G G P G G Y A P A A G G P A G Y A   258
L. sp. N          P A A S K A G G Y G P G A G G P G G Y A P A A G G P A G Y A   181
P. marinus        - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   234

L. camtsyaticum   P A A G G P - - - - - - - - - - - - - - - - - - - - - - - -   264
L. sp. N          P A A G G P - - - - - - - - - - - - - - - - - - - - - - - -   187
P. marinus        - - - - - - L V A M H Q Q Q E E L V V L A Q Q Q A K P V V M   258

L. camtsyaticum   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   264
L. sp. N          - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   187
P. marinus        V Q G L E D L V A M H R Q Q E D L V A M H R Q Q E D L V A M   288

L. camtsyaticum   - - - - - - - - - - G G Y G P A A G G A N G Y G P G A S K A A   285
L. sp. N          - - - - - - - - - - G G Y G P A A G G A N G Y G P G A S K A A   208
P. marinus        H R Q Q E D L A V - - Y G P A A G G A N V Y G P G A S K A A   316

L. camtsyaticum   G G P G G Y A P A A G G P G Q S G P G A G G A G G N G A G G   315
L. sp. N          G G P G G Y A P A A G G P G Q S G P G A G G P G A N G A G G   238
P. marinus        G G P G G Y A P A A G G P G Q S G P G A G G P G G N G A G G   346

L. camtsyaticum   K P N G H A P A A G G Q P G F A P R P S A F G P G S G G P G   345
L. sp. N          K P N G H A P A A G G Q P G F A P R P S A F G P G S G G P G   268
P. marinus        K P N G H A P A A G G Q P G F A P R P S A F G P G S G G P G   376

L. camtsyaticum   S H P G A G A G G Y G Q K A S G Y A P R A G G P G G Y A P A   375
L. sp. N          S H P G A G A G G Y G Q K A S G Y A P R A G G P G G Y A P A   298
P. marinus        S Y P G A G A G G Y G Q K A S G Y A P R A G G P G G Y A P A   406

L. camtsyaticum   A R R P N G Y A P A V G R Q N G Y G P A A G R P G Y G P S S   405
L. sp. N          A R R P N G Y A P A V G R Q N G Y G P A A G R P G Y G P S S   328
P. marinus        A R R P N G Y A P A V G R Q N G Y G P A A G R P G Y G P S S   436

L. camtsyaticum   G G A G G P A P G D Q K P A G G Q K A G G A G G Y G P G A G   435
L. sp. N          G G A G G P A P G D Q K P A G G Q K A G G A G G Y G P G A G   358
P. marinus        G G A G G P A P G D Q K P A G G Q K A G G A G G Y G P G A G   466

L. camtsyaticum   G A G Y A P A V A T R S Q F Y M K L S S K C S Q R R Q L G K   465
L. sp. N          G G G Y A P A V A T R S Q F Y M K L S S K C S Q R R Q L V T   388
P. marinus        G G G Y A P A A G A R S Q F Y M K L S S K C S Q R R Q L G K   496

L. camtsyaticum   S A *                                                       467
L. sp. N          N G N                                                       391
P. marinus        S A *                                                       498
```

```
GTGATTGCACACCTGCTAACCTTCTATCTCGTTTCATCAGGATATGGCCAAGGGGCAGGA      60
 V  I  A  H  L  L  T  F  Y  L  V  S  S  G  Y  G  G  Q  G  A  G      20

GCTGCGCTTCCTTCAAGAACAGCTGGCCCTGGGGGGTACGGTCAGGGGGGGCGGTGGACCT     120
 A  A  L  P  S  R  T  A  G  P [G  G  Y  G  Q  G  G  G  G  P       40

GGGGGGTACGGCCCAGGGGGCAGGAGGTTACGGTCCAGGGGGGAGCAAAGGTCGGTAATGGT     180
 G  G  Y  G  P  G  A [G  G  Y  G  P  G  G  A  K  V] G  N  G      60

GCAGGCGCAGGCAGGCCCAATGGATACGGTCCAGCGATCGGAGGGCCAGGCGCATTTGTA     240
 A  G  A  G  R  P [N  G  Y  G  P  A  I  G  G  P] G  A  F  V      80

TCCAGGGCAAGGCCAAGTGTATACGTTGCTGCGCCAAGTAGAGCGAATGGATATGGTCCA     300
 P  R  A  R  P  S  V  Y  V  A  A  P  S  R  A [N  G  Y  G  P      100

GGCGGAAGCAAAGCTGGTGGTCCAGGAGCTGGGGGACCTGGGGGGTTATGGTCCAGGAGCT     360
[G  G  S  K  A] G  G  P  G  A  G  G  P [G  G  Y  G  P  G  A      120

GGAAGACCTGGTGGCTATGCACCAGCAGCAGGAGGGCCTGGTGGCTATGCACCAGCAGCA     420
[G  R  P] G  G  Y  A  P  A  A  G  G  P  G  G  Y  A  P  A  A      140

GGAGGGCCTGGTGGCTATGCACCAGCAGCAGGAGGAGGAGGACCTGGTGGCTATGCCCCA     480
[G  G  P] G  G  Y  A  P  A  A  G  G  G  G  P [G  G  Y  A  P      160

GCAGCAGGAGGACCTGGTGGCTATGCACCAGCAGCAGGAGGACCTGGTGGTTTTGGCCCG     540
 A  A  G  G  P] G  G  Y  A  P  A  A  G  G  P  G  G  F  G  P      180

GCCGCAAGTAAAGCTGGTGGTTATGGTCCAGGAGCTGGGGGACCTGGTGGTTATGCACCA     600
 A  A  S  K  A] G  G  Y  G  P  G  A  G  G  P  G  G  Y  A  P      200

GCAGCAGGAGGGCCTGGTGGCTATGCACCAGCAGCAGGAGGGCCTGGTGGCTATGCACCA     660
 A  A  G  G  P] G  G  Y  A  P  A  A  G  G  P  G  G  Y  A  P      220

GCAGCAGGAGGACCTGGAGGCTATGCACCAGCAGCAGGAGGCCTGGTGGCTATGCACCAG     720
 A  A  G  G  P] G  G  Y  A  P  A  A  G  G  L  V  A  M  H  Q      240

CAGCAGGAGGAGCTGGTGGTTTTGGCCCAGCAGCAAGCAAAGCCGGTGGTTATGGTCCAG     780
 Q  Q  E  E  L  V  V  L  A  Q  Q  Q  A  K  P  V  V  M  V  Q      260

GGGCTGGAGGACCTGGTGGCTATGCACCGGCAGCAGGAGGACCTGGTGGCTATGCACCGG     840
 G  L  E  D  L  V  A  M  H  R  Q  Q  E  D  L  V  A  M  H  R      280

CAGCAGGAGGACCTGGTGGCTATGCACCGGCAGCAGGAGGACCTGGCGGTTTATGGTCCA     900
 Q  Q  E  D  L  V  A  M  H  R  Q  Q  E  D  L  A  V  Y  G  P      300

GCAGCAGGAGGGGCTAATGTATATGGCCCCTGGGGCTAGTAAAGCGGCAGGAGGACCTGGC     960
 A  A  G  G  A [N  V  Y  G  P  G  A  S  K  A] A  G  G  P  G      320

GGTTATGCACCAGCAGCAGGAGGACCTGGACAGTCTGGTCCAGGGGCAGGAGGACCAGGC    1020
 G  Y  A  P  A  A  G  G  P  G  Q  S  G  P  G  A  G  G  P  G      340

GGGAATGGGGCTGGAGGTAAGCCTAACGGACATGCCCCAGCAGCCGGTGGCCAACCTGGC    1080
 G  N  G  A  G  G  K  P  N  G  H  A  P  A  A  G  G  Q  P  G      360

TTTGCACCAAGGCCTAGCGCATTTGGTCCAGGGTCAGGGGGACCTGGATCTTATCCAGGA    1140
 F  A  P  R  P  S  A  F  G  P  G  S  G  G  P  G  S  Y  P  G      380

GCAGGAGCTGGTGGATATGGACAAAAGGCATCTGGCTATGCCCCAAGGGCAGGGGGACCT    1200
 A  G  A  G  Y  G  Q  K  A [S  G  Y  A  P  R  A  G  G  P      400

GGTGGTTATGCTCCAGCTGCGAGGAGGCCAAATGGCTATGCCCCAGCGGTAGGGAGGCAA    1260
 G  G  Y  A  P  A  A  R  R  P  N  G  Y  A  P  A  V  G  R  Q]     420

AACGGATATGGACCGGCGGCAGGGAGACCTGGATATGGTCCAAGTTCAGGAGGAGCTGGA    1320
[N  G  Y  G  P  A  A  G  R  P] G  Y  G  P  S  S  G  G  A  G     440

GGACCTGCCCCAGGTGACCAGAAACCAGCAGGTGGCCAGAAGGCAGGTGGTGCTGGTGGG    1380
 G  P  A  P  G  D  Q  K  P  A  G  G  Q  K  A  G  G  A [G  G     460

TACGGACCTGGGGCTGGAGGAGGAGGATACGCGCCCGCAGCCGGAGCTCGCTCACAGTTC    1440
 Y  G  P  G  A  G  G  G] G  Y  A  P  A  A  G  A  R  S  Q  F     480

TACATGAAGCTCTCCTCTAAGTGTTCCCAAAGGAGACAGCTCGGTAAGAGTGCGTGA       1497
 Y  M  K  L  S  S  K  C  S  Q  R  R  Q  L  G  K  S  A  *        498
```