

Title:

**Diversification of metazoan Kexin-like proprotein convertases: insights from the leech *Helobdella***

Authors:

Wei-Yu Tao<sup>1</sup>, Ya-Chih Cheng<sup>1,4</sup>, Mi Hye Song<sup>2</sup>, David A. Weisblat<sup>3</sup>, and Dian-Han Kuo<sup>1\*</sup>

Author Affiliation:

1: Department of Life Science, National Taiwan University, Taipei, Taiwan

2: Department of Biological Sciences, Oakland University, Rochester, MI, U.S.A.

3: Department of Molecular and Cell Biology, University of California, Berkeley, U.S.A.

4: Present address: Institute of Environmental and Occupational Health Sciences, National Taiwan University, Taipei, Taiwan

\*: Correspondence: [dhkuo@ntu.edu.tw](mailto:dhkuo@ntu.edu.tw); [dianhankuo@gmail.com](mailto:dianhankuo@gmail.com)

Keywords:

Protein convertase, PCSK, gene family, cell signaling, neuropeptides, hormones, BMP

## Abstract

Intercellular communication is quintessential for multicellularity and often mediated by secreted peptide ligands. In Metazoa, proprotein convertases are a major class of endoproteases partaking in the proteolytic processing of these ligands, which is in turn required for their signaling activities. In vertebrates, the best-studied convertase substrates are neuropeptides, peptide hormones, and members of the TGF $\beta$ /BMP-family. Each ligand is processed by a particular subset of convertases. Therefore, the diversification of convertases may have contributed to the growing complexity of cellular communication in metazoan evolution. However, proprotein convertases have not been systematically explored in Metazoa. Here, we sampled the representative metazoan genomes and established that six Kexin-like proprotein convertases were present in the last common ancestor of protostomes and deuterostomes. Among these, we identified a novel PCSKX orthologous group (OG) that was lost in vertebrates. Spiralian protosomes have, in general, maintained all six OGs. Therefore, we characterized the functional divergence of the Kexin-like OGs in the leech *Helobdella*, an experimentally tractable spiralian. Gene expression patterns suggested that PCSK1 and PCSK2 are specialized for the processing of neuropeptides and peptide hormones in bilaterians and that the newly identified PCSKX is probably functionally similar to furin and PCSK7. Finally, we showed that, distinct from the BMP morphogen in vertebrate embryos, the convertase-mediated proteolytic cleavage is not required for the short-range BMP signaling in the dorsoventral patterning of leech ectoderm. Together, our data revealed the complexity of the Kexin-like proprotein convertase gene family and their roles in generating diverse patterns of cellular communication in Metazoa.

## **Introduction**

Endoproteolytic processing of secreted peptides is one of the common mechanisms for the post-translational modifications that are used to activate and inactivate biomolecules in the eukaryotic cells. In endoproteolytic processing, proprotein convertases cleave the inhibitory prodomain from the precursor peptides to produce bioactive signaling ligands, receptors, and enzymes in the secretory pathway. The metazoan proprotein convertases are structurally similar to the prokaryotic Subtilisin and the yeast Kexin and are therefore named Proprotein Convertase Subtilisin/Kexin (PCSK) (Seidah 2011). Metazoan PCSKs are involved in a great variety of physiological and developmental processes. For example, PCSK1 (or PC1/3) and PCSK2 (or PC2) cleave inactive precursors in secretory vesicles to produce the active forms of peptide hormones and neuropeptides (Furuta, et al. 1997; Siekhaus and Fuller 1999; Zhu, et al. 2002; Husson, et al. 2006; Rhea, et al. 2010; Zhang, et al. 2010). In contrast, furin (or PCSK3), PCSK5 (or PC5/6), PCSK6 (or PACE4) and PCSK7 (or PC7) are best known for their role in the processing of TGF $\beta$ /BMP ligands, a family of signaling molecules that have multiple important developmental roles (Dubois, et al. 1995; Cui, et al. 1998; Essalmani, et al. 2008; K nnapuu, et al. 2009; Nelsen and Christian 2009), as well as a number of other secreted or transmembrane proteins (Seidah and Chr tien 1997, 1999).

Although proprotein convertases such as furin and PCSK5-7 are often viewed as “housekeeping proteins” due to their ubiquitous expression in adult tissue, they may also make significant contributions to the regulation of morphogen signals during embryonic development, since a subtle difference in the proprotein convertase-mediated cleavage of the ligand peptides can introduce a profound effect on the range and strength of TGF $\beta$ /BMP signaling and influence the shape of a morphogen gradient (Constam 2014). Intriguingly, the order in which the multiple convertase sites on the vertebrate BMP4 proprotein are cleaved can determine its signaling range, by specifying different modes of interactions between the inhibitory prodomain and the ligand peptide (Cui, et al. 2001; Degnin, et al. 2004; Tilak, et al. 2014). Moreover, proprotein convertases that are broadly expressed in adult tissue may exhibit developmentally regulated expression during embryogenesis, being specifically upregulated in cells that produce TGF $\beta$ /BMP ligands (Constam, et al. 1996; Nelsen, et al. 2005). Functional data showed that there seems to be a tissue-specific requirement for individual convertase sites on a BMP ligand (Goldman, et al. 2006; Sopory, et al. 2010; Akiyama, et al. 2012). Given that

each individual convertase has distinct, preferred cleavage sites (Nelsen and Christian 2009), tissue-specific expression of convertase activity may be a critical factor in determining the differing strength and range of TGF $\beta$ /BMP signaling between tissue types.

While it has been shown that the requirement for and consequence of convertase-mediated cleavage differ between BMP ligands and between species (Fritsch, et al. 2012), the diversity in the roles of convertase-mediated cleavage in the regulation of cellular signaling has been largely unexplored. In the leech *Helobdella*, a novel form of short-range BMP signal patterns the dorsoventral axis of ectoderm (Kuo and Weisblat 2011b). Thus, to investigate the hypothesis that differential expression of selected Kexin-like proprotein convertases contributes to such a non-canonical deployment of this BMP signal, we set out to characterize the genes encoding proprotein convertases in the leech genome.

Proprotein convertases have only been studied in a small number of mostly vertebrate model species, and the pre-vertebrate evolutionary history of this gene family is largely unknown. To provide a phylogenetic framework for our studies on the leech proprotein convertases and to gain a bird's-eye view on the evolution of this gene family in the animal kingdom, we first carried out a survey of convertase-like genes across metazoan species. Genes that are currently described as PCSKs were in fact divided into three distantly-related groups based on their distinct structural profiles. We focused on the classical Kexin-like proprotein convertases, which are responsible for processing BMP ligands. To assign individual leech convertases to specific orthologous groups (OGs), we performed a phylogenetic analysis of proprotein convertases from representative metazoan genomes. Six Kexin-like convertases were inferred to be present in the genome of the last common ancestor of protostomes and deuterostomes. Among these, we identified a novel OG named PCSKX, which was lost in the vertebrates. On the other hand, PCSK7 has a premetazoan origin, and the five other Kexin-like convertases arose from gene duplications that occurred before the protostome-deuterostome split. Additional lineage-specific gene duplications and losses further diversified the proprotein convertase gene complement among extant bilaterians.

Spiralians, including the leech *Helobdella*, have retained all six ancestral OGs of Kexin-like convertases. We analyzed the expression pattern of individual proprotein convertase genes in the adult leech. Our results suggest that, similar to other animal species, PCSK1 and PCSK2 were likely involved in the production of neurohormones and synaptic neuropeptides in

the leech. Finally, we showed that none of the leech Kexin-like convertases is specifically upregulated in the BMP-producing cells and that convertase-mediated cleavage is not required for the normal signaling activity of BMP in the dorsoventral patterning of ectoderm. Therefore, the expression and functional requirement of proprotein convertases in this patterning system are very different from that in the standard BMP morphogen system as observed in vertebrate embryos. Together, our data revealed a dynamic and complex picture for the evolution of proprotein convertases in Metazoa. In line with the diversity in number and composition of proprotein convertase genes, their roles in the regulation of the peptide ligands have also undergone lineage-specific modifications. Diversification in gene repertoire and gene function of proprotein convertases is expected to play a role in the evolution of complex cellular communication systems in Metazoa.

## **Results and Discussion**

### **Proprotein convertase-like genes in metazoan genomes**

In mammals, the PCSK designation was applied to nine different genes, whose gene products are involved in the proteolytic processing of secreted proteins (Seidah 2011). Among these, PCSK1-7 are characterized by a P domain (Pfam PF01483) on the C-terminal side of the catalytic peptidase S8 domain (Pfam PF00082) and are structurally similar to the yeast Kexin. Homologs of Kexin-like convertases, such as PCSK2/PC2 and PCSK3/furin, have been identified and characterized in a number of invertebrate species (Fan and Nagle 1996; Agata, et al. 1998; Oliva, et al. 2000; Thacker and Rose 2000; Rayburn, et al. 2003; Bertrand, et al. 2006; Cummins, et al. 2009). Published records for homologs of PCSK8 and PCSK9 outside of vertebrates are sparse, but their phylogenetic distribution suggests an ancient, pan-eukaryotic origin for each of these genes. PCSK8 is also known as Site-1 Protease (S1P), or Subtilisin/Kexin-Isozyme 1 (SKI-1); its homologs were reported in *Drosophila* (Seegmiller, et al. 2002) and in *Arabidopsis* (Liu, et al. 2007) but not in yeast. PCSK9, or Neural Apoptosis-Regulated Convertase 1 (NARC-1), is structurally similar to the fungi proteinase Ks (Naureckiene, et al. 2003; Seidah, et al. 2003); its homologs were reported in a few metazoan species, including amphioxus, sea cucumber, and gastropods (Cameron, et al. 2008; Ren, et al. 2016). Here, we will use “Kexin-like,” “S1P-like,” and “PK-like” to refer to the convertases

that are structurally similar to vertebrate PCSK1-7, PCSK8, and PCSK9 respectively, for reasons entailed later in this section.

To further explore the evolution of the proprotein convertase genes in Metazoa, we searched for the Kexin-like, S1P-like, and PK-like convertase genes in selected genomes represented in the ENSEMBL databases. A Kexin-like convertase gene was defined by having an amino acid sequence that matches both Pfam PF00082 and Pfam PF01483 profiles. S1P-like convertase genes were defined by matching both Pfam PF00082 and InterPro IPR034185 (Site-1 peptidase catalytic domain). PK-like convertase genes were defined by matching all of the following three profiles, namely Pfam PF00082, Pfam PF05922 (Peptidase inhibitor I9) and InterPro IPR034193 (Proteinase K-like catalytic domain). Gene models that match any one, but not all, of the defining profiles were manually examined and analyzed by reciprocal BLAST search to identify incomplete models and to remove the false positives. The complete set of convertase genes identified in this study is listed in Table S1.

Among the three non-metazoan opisthokont genomes searched, one Kexin-like gene (*KEX2*) and three S1P-like genes (*RRT12*, *PRB1*, and *YSP3*) were found in the yeast *Saccharomyces cerevisiae*; one Kexin-like, one S1P-like and three PK-like genes were found in the filopodia-bearing, amoeboid protist *Capsaspora owczarzaki*; and two Kexin-like genes and one S1P-like were found in the choanoflagellate *Monosiga brevicollis*. We also searched the genome of *Acanthamoeba castellanii*, and found three Kexin-like, one S1P-like, and nine PK-like genes. *A. castellanii* belongs to Amoebozoa, a distant outgroup for Opisthokonta (Cavalier-Smith, et al. 2014).

Within Metazoa, we searched 23 genomes, including five vertebrates (the human *Homo sapiens*, the chicken *Gallus gallus*, the western clawed frog *Xenopus tropicalis*, the zebrafish *Danio rerio*, and the lamprey *Petromyzon marinus*), three invertebrate deuterostomes (the ascidian *Ciona intestinalis*, the amphioxus *Branchiostoma floridae*, and the sea urchin *Strongylocentrotus purpuratus*), four ecdysozoans (the fruit fly *Drosophila melanogaster*, the red flour beetle *Tribolium castaneum*, the water flea *Daphnia pulex*, and the nematode *Caenorhabditis elegans*), seven spiralian (the limpet *Lottia gigantea*, the oyster *Crassostrea gigas*, the octopus *Octopus bimaculoides*, the polychaete worm *Capitella teleta*, the leech *Helobdella robusta*, the brachiopod *Lingula anatina*, and the rotifer *Adineta vaga*), and four non-bilaterians (the sea anemone *Nematostella vectensis*, the placozoan *Trichoplax adherens*,

the ctenophore *Mnemiopsis leidyi*, and the sponge *Amphimedon queenslandica*). The copy number of Kexin-like genes in these metazoan genomes ranged from three (*Trichoplax*) to nineteen (*Adineta*) and was never zero in any of these species (Figure 1A). The copy number of S1P-like convertases ranged from zero (*Caenorhabditis*) to three (*Lingula*); a single copy was found in 19 out of the 23 metazoan genomes (Figure 1A). The PK-like convertase genes were absent from 15 out of the 23 genomes and were highly duplicated in others, with the most extreme case being the sea urchin genome which contained as many as unique 62 PK-like loci (Figure 1A).

A fourth, and the final, group of genes encoding S8 domain-containing peptidases, namely the tripeptidyl peptidase II (TPPII) family, was also broadly represented in the metazoan genomes. TPPII functions in cytosolic proteolysis (Rockel, et al. 2012), but not in the secretory pathway. By definition, TPPII is not a proprotein convertase. For the sake of completeness, we nevertheless went on to document TPPII-like genes in the metazoan genomes. Genes of this family can be structurally defined by matching Pfam PF00082 and Pfam PF12580 (TPPII). The copy number of TPPII-like genes in the 23 metazoan genomes examined ranged between zero and two. TPPII-like was absent from the genomes of *Mnemiopsis* and *Petromyzon*. Two copies of TPPII-like genes were found in *Amphimedon*, *Lingula*, *Adineta*, and *Branchiostoma*. In the remaining 17 genomes, each contained only one TPPII-like gene (Figure 1A). Outside of Metazoa, single TPPII-like genes were found in the genomes of *Acanthamoeba*, *Capsaspora*, and *Monosiga*, but not in the yeast *Saccharomyces*, suggesting a gene loss in the lineage leading to this species. Overall, the copy number of S1P-like and TPPII-like genes showed little variation between metazoan genomes, whereas the copy number variation was greater for the Kexin-like and the PK-like genes (Figure 1B)

In addition to the four structurally defined S8-like peptidase gene families we found in metazoan genomes, Kexin-like, S1P-like, PK-like, and TPPII-like, a more diverse repertoire of peptidase S8-like genes was uncovered outside of Metazoa (Bryant, et al. 2009; Li, et al. 2017). Based on the following observations, we postulated that the four metazoan S8-like families are probably distantly related and not evolutionary sisters. First, from their distribution across species, each of the four families has a very deep origin, probably pre-Eukaryota. Second, each family has a distinct structural profile from each other and recognizes a distinct set of target-site sequences (Seidah, et al. 1999; Seidah, et al. 2003; Seidah 2011). Third,

sequences of members from different families cannot be aligned without catastrophic loss of phylogenetic information (data not shown). Given that TPPII-like does not have a proprotein convertase function and that the four families of S8-like genes in the metazoan genomes are distantly related, we speculate that the propeptidase function of the three families of S8-like peptidases arose by functional convergence. Thus, although PCSK has been used to refer to the metazoan S8 peptidase genes possessing a proprotein convertase activity, one should refrain from using PCSK for the larger gene families on a phylogenetic basis; we therefore use specific names such as Kexin-like, S1P-like, and PK-like to refer to the gene families here.

### **Kexin-like proprotein convertase genes in the *Helobdella* genome**

Pfam query and BLAST search of the *Helobdella robusta* database yielded 17 candidate gene models encoding Kexin-like genes. Among these, nine models were determined as being complete or nearly complete, based on the criterion that they have all the structure domains that are typical of known Kexin-like convertases, namely a prodomain, a catalytic domain, and a P domain. The remaining eight are apparently incomplete models. We next examined the details of these incomplete models to revise these models. By manually examining the genome location of these models, we found that three models (JGI Protein IDs: 74170, 137839, and 149650) are adjacent to each other and that each corresponds to a unique region of a typical Kexin-like convertase. Therefore, we propose that these three models should be merged to represent a single gene in the *Helobdella* genome. Another two models (JGI Protein IDs: 81814 and 184031) each represent a different portion of a typical Kexin-like convertase, and each is located at the end of a large scaffold. Therefore, we proposed that these two models may represent a single gene and that they were separated due to incomplete genome assembly. Third, one gene model (JGI Protein ID: 190143) contains a gap in the sequence between two predicted exons, and two models (JGI Protein IDs: 93024 and 93071) are located on small scaffolds, 1,371 and 1,142 base-pairs in length respectively. Of these two small scaffolds, gene model 93024 fits into the gap in gene model 190143; thus, we proposed that these two models also represent a single gene. Finally, the sequence of the genome scaffold containing the remaining orphaned model (JGI Protein ID: 93071) was derived from two relatively short Sanger reads with poor quality and is highly similar to a small region on Scaffold 34 that contains model 82977. Therefore, we determined that model 93071



and the genome scaffold containing it (scaffold 3727) are probably artifacts resulting from poor sequencing reads. Together, our *in silico* approach identified 12 candidate Kexin-like convertase genes in the *Helobdella* genome (Table S1). These predictions were confirmed by cloning and sequencing of cDNA for each of the 12 predicted genes. A single S1P-like (JGI protein ID: 194246) and a single TPPII-like (JGI protein ID: 107485) were identified in the *Helobdella* genome. No PK-like gene was found.

### Phylogeny of metazoan Kexin-like genes

To assign orthologies to the 12 *Helobdella* Kexin-like convertases, we next performed a phylogenetic analysis. From the gene list uncovered by Pfam profile search, gene models that contain only partial sequence, as determined by whether all the Pfam-defined domains are completely covered by the model sequence, were excluded from our analysis. As a result, all four *Petromyzon* models, and three *Amphimedon*, one *Nematostella*, one *Crassostrea*, one *Octopus*, five *Lingula*, three *Adineta*, and two *Strongylocentrotus* models were removed from the dataset. Furthermore, given the high degree of conservation among gnathostome Kexin-like gene complements, only three vertebrate species (*H. sapiens*, *G. gallus*, and *X. tropicalis*) were included in our analysis to reduce computation load. The 12 *Helobdella* Kexin-like protein sequences were revised based on the cloned cDNA sequences.

The amino acid sequence alignment of the 156 genes from the genomes of three non-metazoan opisthokonts and 21 metazoans was produced, and a Maximum-Likelihood (ML) tree was then generated by applying LG+G model to the dataset, with *Saccharomyces* Kexin as an outgroup. In this tree, four well supported OGs were recognized: PCSK7, PCSK1, PCSK2, and a novel OG consisting entirely of previously uncharacterized invertebrate proprotein convertase genes (Figure 2A). We named this novel OG “PCSKX.” One *Helobdella* ortholog was found in each of the four OGs. We therefore named these *Helobdella* genes *Hro-pcsk7*, *Hro-pcsk1*, *Hro-pcsk2*, and *Hro-pcskx*. Among the remaining eight *Helobdella* Kexin-like genes, six were in two poorly supported groups clustered with the vertebrate furin/PCSK4 genes and protostome furin1 genes, and two were in groups consisting of apparently long-branch genes.

We hypothesized that such a tree topology might result from a long-branch attraction (LBA) artifact. We next sought to eliminate the potential LBA by using a 77-gene/15-species

dataset, in which species other than *Helobdella* containing duplicated, long-branch genes were removed. Besides *Helobdella*, each of the remaining ten bilaterian species in this dataset has at most seven Kexin-like genes in its genome. In the tree constructed under LG+G model, seven of the *Helobdella* genes were in a clade that also includes protostome furin1 genes and vertebrate furin/PCSK4 genes. We therefore named these seven *Helobdella* furin1 paralogs *Hro-furin1.1*, *Hro-furin1.2*, *Hro-furin1.3*, *Hro-furin1.4*, *Hro-furin1.5*, *Hro-furin1.6*, and *Hro-furin1.7* respectively (Figure 3A). The remaining *Helobdella* gene was in a poorly supported group that includes the protostome furin2 genes and a sea urchin gene (Figure 3A). This *Helobdella* gene is named *Hro-furin2-like*. Hence, the removal of long-branch genes had apparently improved the phylogenetic resolution to allow us to assign each of the 12 *Helobdella* Kexin-like genes to a specific OG. By analyzing the complete 156-gene/24-species dataset using the CAT-GTR model, which generally outperforms other models in resolving long branches (Stamatakis 2006; Lartillot, et al. 2007; Price, et al. 2010; Izquierdo-Carrasco, et al. 2011), we obtained a similar grouping of the *Helobdella* furin-like genes (Figure 2B). The fact that similar results were obtained using two different approaches gives us confidence that the orthology assignments for the *Helobdella* furin-like genes are likely to reflect their true phylogenetic affiliations.

The placement of one of the choanoflagellate Kexin genes in the 77-gene/15-species LG+G tree is inconsistent with the currently accepted phylogenetic hypothesis. We next used the CAT-GTR model to analyze the 77-gene/15-species dataset. In the resulting tree (Figure 3B), the first node from the root separates PCSK7 from the other Kexin-like genes. On the other side, the *Capsaspora* gene branches off first, followed by a group consisting of the two *Monosiga* paralogs. Although these nodes are not well supported, they are consistent with the currently accepted species tree. The grouping of PCSK1, PCSK2, PCSKX, furin1, furin2 (containing non-chordate species only), and PCSK5/6 (containing chordate species only) was identical between LG+G and CAT-GTR trees, but the branching order differs. In both cases, PCSK1 and the large furin-like group (furin1 + furin2 + PCSK5/6) are sister clades. In the CAT-GTR tree, PCSK2 and PCSKX are sister clades, and the (PCSK2 + PCSKX) clade is, in turn, the sister of the (PCSK1 + furin-like) clade. In the LG+G tree, PCSK2 is the sister of the (PCSK1 + furin-like) clade, and PCSKX is the sister of the (PCSK2 + (PCSK1 + furin-like)) clade. However, none of these deep nodes is well-supported.

To further resolve the interrelationships between the identified OGs, we next removed the *Helobdella* genes, which also represent long branches themselves, and performed the ML tree search using either LG+G or CAT-GTR model. In LG+G trees as well as CAT-GTR trees, the interrelationships between PCSKX, PCSK1, PCSK2, and furin-like groups were the same regardless of whether the *Helobdella* genes were included or not (Figure 3A versus Figure 4A; Figure 3B versus Figure 4B). When polytomy was allowed for conditions in which the bootstrap support was < 50%, the topologies of LG+G tree and CAT-GTR tree were very similar to each other, with only one difference detected: the separation of the two sea urchin furin-like genes into the two furin-like subgroups in the CAT-GTR tree (*c.f.* Figure 4A' and Figure 4B'). The support for the sisterhood between PCSK1 and furin-like genes was nevertheless improved in the CAT-GTR tree. Notably, this CAT-GTR tree (Figure 4B) provided a plausible parsimonious solution to the phylogeny of the furin-like genes. In the moderately supported furin-like group, a single *Nematostella* gene branched off first; next, a well-supported furin1 subgroup and a poorly supported (furin2 + PCSK5/6) subgroup separated. This branching pattern provided the best solution to reconcile the species tree. Therefore, we proposed the (furin2 + PCSK5/6) grouping as our working phylogenetic hypothesis.

We next attempted to infer the evolutionary history for each of the six OGs by reconciling our gene trees to the species tree. The species tree used here was synthesized from recent phylogenomic studies (Laumer, et al. 2015; Telford, et al. 2015; Feuda, et al. 2017; Kocot, et al. 2017). PCSK7 was identified in many sampled metazoan genomes, including the basally branching species such as *Amphimedon*, *Trichoplax*, and *Nematostella* (Figure 2, Figure 4C). Although the sea urchin PCSK7 gene was missing from our phylogenetic analysis, a reciprocal BLAST search identified one of the two incomplete *Strongylocentrotus* gene models that were dropped from the dataset as being most similar to PCSK7, suggesting that PCSK7 is likely present in this species. Given that all the *Capsaspora* and *Monosiga* Kexin-like genes branch off after the PCSK7 group, PCSK7 might have arisen by a gene duplication event in the ancestral filozoan or holozoan, followed by gene losses in lineages such as *Capsaspora* and *Monosiga*. Within Metazoa, PCSK7 was apparently lost in the ecdysozoan lineage.

The PCSKX is a novel Kexin-like OG identified in this study, with members present in *Nematostella*, several spiralian species, sea urchin, and amphioxus, but not in ecdysozoans

and vertebrates (Figure 2, Figure 4C). This suggests two independent gene losses in these respective lineages. This OG was likely to have arisen before the common ancestor of planulozoans, the group consisting of cnidarians and bilaterians. Similar to PCSKX, the PCSK1 orthology group contains genes from *Nematostella* and bilaterians, and therefore, the PCSK1 OG likely predated the planulozoans, and was lost independently in some bilaterian species, e.g. *Caenorhabditis*, *Drosophila*, *Strongylocentrotus*, and *Ciona*.

We could not robustly resolve the phylogenetic relationships among the furin-like subgroups. Nevertheless, three better-supported subgroups can be identified: furin1 (present across all bilaterians; including the vertebrate furin/PCSK4), furin2 (in protostomes only), PCSK5/6 (in vertebrate only). If we accept the (furin2 + PCSK5/6) grouping, a simple scenario for the evolution of furin-like genes emerges: first, a gene duplication took place in the ancestral planulozoan and gave rise to a PCSK1 gene and a furin-like gene. Next, in the ancestral bilaterian, the furin-like gene duplicated into furin1 and furin2. In vertebrates, furin1 further duplicated into furin (a.k.a. PCSK3) and PCSK4, while furin2 duplicated into PCSK5 and PCSK6. In a more complicated alternative scenario, a series of duplications from the planulozoan furin-like gene led to three genes, furin1, furin2, and PCSK5/6, in the ancestral bilaterian. Furin1 was preserved in all bilaterian lineages. In contrast, furin2 was lost in the chordates, and PCSK5/6 was lost in protostomes.

PCSK2 was found across bilaterians but not in any non-bilaterian metazoan. This may lead one to conclude that PCSK2 originated in Urbilateria. However, this is inconsistent with the species represented in its candidate sister groups, i.e. the PCSKX, PCSK1, and furin-like OGs. Since these all contain genes from non-bilaterian metazoan species, even by considering the phylogenetic uncertainty between these OGs, none of the possible tree topologies is consistent with an urbilaterian origin of the PCSK2 group. To reconcile this discrepancy, we hypothesize that PCSK2 originated before the ancestral planulozoan and was lost in the lineage leading to *Nematostella*. We have not found PCSK2 in the genomes of the hydra and the coral *Acropora*, suggesting that the loss of PCSK2 probably took place in the common ancestor of modern cnidarians. However, an orphaned *Nematostella* gene, Nve-234208, tended to lie near PCSK2 in our phylogenetic analyses and may be in fact a highly derived PCSK2. Orthologs of Nve-234208 were also identified by reciprocal BLAST in the genomes of hydra and coral, suggesting that this is not an anomaly unique to *Nematostella*. In

that case, the cnidarian PCSK2 may have undergone a rapid sequence divergence instead of a gene loss.

We could not reach a single consensus conclusion on the interrelationships between the bilaterian Kexin-like OGs. Our data provide soft support for the sisterhood between PCSK1 and furin-like OGs. By constraining them as sister groups, there were three possible scenarios (Figures 4D), all of which were equal in the number of gene losses required to reconcile the species tree. In our phylogenetic analyses, however, PCSK1 was most likely the sister of the furin-like genes, and therefore, the phylogenetic uncertainty mostly lies in the branching order of PCSK2, PCSKX, and the (PCSK1 + furin-like) clade.

In summary, we propose that the last common ancestor of modern bilaterians had six Kexin-like genes, namely PCSK7, PCSKX, PCSK2, PCSK1, furin1, and (furin2+PCSK5/6). We have identified an early origin of PCSK7 but not been able to fully resolve the details in the early history of Kexin-like genes. Trees that include genes from the most basally branching metazoans, *Amphimedon* and *Mnemiopsis*, could not be reasonably reconciled, and this indicates a skewed tree topology, presumably resulting from LBA. We also find that taxon-specific duplication events produced further local diversification of selected Kexin-like genes. For example, both the leech *Helobdella* and the brachiopod *Lingula* have multiple furin-like genes. It is interesting to note the bdelloid rotifer *Adineta* has four paralogs for each of PCSKX, PCSK1, PCSK2, and furin2, consistent with the degenerate tetraploid genome that is characteristic of this asexual group (Hur, et al. 2009; Flot, et al. 2013; Nowell, et al. 2018). Furthermore, although amphioxus has been inferred as having a genome organization representative of the ancestral chordate (Putnam, et al. 2008), the amphioxus Kexin-like genes are unexpectedly highly duplicated, as multiple paralogs of PCSK7, PCSKX, PCSK2, and furin-like were found in this species (Figure 2). On the other hand, selected OGs have been lost in specific taxa. For example, ecdysozoans have lost PCSK7 and PCSKX, and the vertebrates have lost PCSKX. From our dataset, it seems obvious that the spiralian species have maintained a complete set of the six Kexin-like genes that were likely present in the last common ancestor of bilaterians (Figure 4C). Although we did not succeed in identifying the amphioxus furin1 ortholog(s), we could not exclude that they are in fact represented by the highly duplicated long-branch furin-like genes (Figure 2). In that case, the amphioxus genome would also contain all six OGs since clear-cut orthologs of PCSK1, PCSK2, (furin2+PCSK5/6),

PCSK7, and PCSKX were identified in amphioxus, and we expect that the presence of all six OGs should be prevalent in the basally branching deuterostome taxa, such as echinoderms and hemichordates.

### **Functional divergence of proprotein convertase genes**

The fact that the six Kexin-like orthology groups were already in place in Urbilateria, and have been conserved to varying degrees since that time is indicative of a certain degree of functional divergence, which is expected to facilitate the preservation of duplicated genes (Ohno 1970). Consistent with this, the mammalian Kexin-like convertases are not functionally redundant. PCSK1 and PCSK2 are involved in the processing of neuropeptides and hormones as they are expressed in brain, neuroendocrine and endocrine tissues, but their null mutant phenotypes differ (Smeekens, et al. 1991; Furuta, et al. 1997; Zhu, et al. 2002; Refaie, et al. 2012); PCSK4 is expressed in the reproductive system and is required for fertility (Seidah, et al. 1992; Mbikay, et al. 1997; Tadros, et al. 2001; Qiu, et al. 2005); furin and PCSK5-7 are broadly expressed, but these genes also exhibit distinct mutant phenotypes (Roebroek, et al. 1998; Constam and Robertson 2000; Thomas 2002; Essalmani, et al. 2006; Wetsel, et al. 2013).

We hypothesize that the origin of functional specialization observed among the mammalian proprotein convertases can be traced back to the bilaterian ancestor and will thus show parallels to modern representatives of distantly related taxa. This hypothesis would predict (1) that the expression of PCSK1 and PCSK2 should be restricted to the nervous system and/or endocrine system across Bilateria, and (2) that furin1, furin2 and PCSK7 should be broadly expressed in most bilaterian species. The newly discovered PCSKX group has obviously not been characterized, and therefore, new expression data will help to understand its biological function. Having confirmed that the spiralian, in general, have all six Kexin-like OGs, gene expression data from a spiralian species is well suited for testing our hypothesis. Therefore, we characterized expression patterns of the 12 Kexin-like convertases in differentiated tissues by whole-mount *in situ* hybridization (WMISH) to juvenile specimens of the leech *Helobdella austinensis*, an experimentally tractable spiralian species (Weisblat and Kuo 2009, 2014).

#### *Tissue-specific expression of PCSK1 and PCSK2*

As predicted, both *Hau-pcsk1* and *Hau-pcsk2* exhibited tissue-restricted expression in the juvenile leech. *Hau-pcsk1* transcript was prominently expressed in a few pairs of cells in the supraoesophageal ganglion (Figure 5A). In ganglia of the ventral nerve cord, however, lower levels of *Hau-pcsk1* expression were detected in a specific set of cells (Figure 5B). In contrast, *Hau-pcsk2* was broadly expressed in the central nervous system (CNS), including the ventral nerve cord and supraoesophageal ganglion (Figure 5C, 5D). Therefore, *Hau-pcsk1* has a more restricted expression than *Hau-pcsk2* in the CNS. Outside of the CNS, *Hau-pcsk2* is also expressed in clusters of large cells flanking the gut on the dorsal side (Figure 4C-4E). Based on their sizes and positions, we speculated that these cells are likely not neurons but exocrine cells of the salivary glands (Moser and Desser 1995; Kuo and Weisblat 2011a).

Invertebrate PCSK1 orthologs have not been characterized in species other than *Aplysia* (Gorham, et al. 1996; Ouimet and Castellucci 1997) and *Helobdella* (this study). In *Aplysia*, the PCSK1 homolog is expressed in the ganglionic neuroendocrine cells that produce the reproductive hormone (Ouimet and Castellucci 1997) as well as the exocrine atrial gland that produces the reproductive pheromone (Gorham, et al. 1996). Our data showed that the leech PCSK1 is prominently expressed in the supraoesophageal ganglion, representing the brain of the leech. Given that many invertebrate peptide hormones, including those of the leech, are produced by neurosecretory cells of the brain (Grimmelikhuijzen, et al. 1980; Engelhardt, et al. 1982; Vigna, et al. 1984; Pestarino 1990; van Heumen and Roubos 1990; Gomot, et al. 1992; Meester, et al. 1992; Sonetti, et al. 1992; Kellner-Cousin, et al. 1994; Kellner, et al. 2002; Wang, et al. 2007; Miller and Newmark 2012), these specialized *Hau-pcsk1*-expressing cells may well be some of the neurosecretory cells of the leech.

On the other hand, *Hau-pcsk2* is broadly expressed in the CNS, and this is similar to the expression pattern of planarian PCSK2 (Shimoyama, et al. 2016). Not all invertebrate PCSK2 genes exhibit broad CNS expression, however. For example, the expression of *Drosophila* PCSK2 seems to be restricted to a small subset of neurons and endocrine cells (Siekhaus and Fuller 1999). Such a difference may be indicative of the varying abundance of peptidergic neurons among these species. Alternatively, it may signify the evolution of other neural functions than neuropeptide processing for PCSK2s in leech and planarian (Lophotrochozoa) compared to *Drosophila* (Ecdysozoa). It is noteworthy that, in addition to the neural tissue, PCSK2 was shown to be expressed in the atrial gland, an exocrine gland, in *Aplysia* (Nagle, et

al. 1995). This seems to parallel the expression of *Hau-pcsk2* in the salivary gland. We interpreted these as independent recruitment of PCSK2 expression to the leech salivary and *Aplysia* atrial gland, however, since the two glands are not homologous. In any case, our data were consistent with the notion that the ancestral function of PCSK1 and PCSK2 in Bilateria is in neurosecretion. Further, our results are consistent with other data that, although PCSK1 and PCSK2 are generally expressed in the neural tissue, they can be occasionally recruited to other tissue types, such as exocrine glands, in some evolutionary lineages.

Our molecular phylogeny data suggested that PCSK1 and PCSK2 are not sister groups (Figure 4D), and therefore, this phylogenetic pattern would suggest independent recruitment of PCSK1 and PCSK2 to the neural and neuroendocrine functions or loss of these functions from other OGs. Unfortunately, the currently available data are not sufficient to resolve the detailed early evolutionary history of PCSK1 and PCSK2 OGs, or of the cell types that specifically express these genes. Further expansion of taxon sampling may help to explore the origins of the peptidergic nervous system and the peptide hormone system.

#### *Widespread expression of PCSK7, PCSKX, Furin1, and Furin2*

Consistent with our prediction, we found that furin-like PCSKs, namely *Hau-furin1.1*, *Hau-furin1.2*, *Hau-furin1.3*, *Hau-furin1.4*, *Hau-furin1.5*, *Hau-furin1.6*, and *Hau-furin2-like*, were broadly expressed (Figure 5F-5K and 5M), although our data suggest the possibility that some these genes are differentially expressed as well (*c.f.*, Figure 5G and 5H). Likewise, *Hau-pcsk7* and *Hau-pcskx* were also broadly expressed (Figure 5N, 5O). In contrast, the expression of *Hau-furin1.7* is restricted to the rear sucker (Figure 5L). These data suggested that genes encoding leech PCSK7, PCSKX, and all but one furin-like convertase exhibit far less tissue-type specificity in their expression than do PCSK1 and PCSK2. Our data also revealed that PCSKX is similar to PCSK7 and furin-like in exhibiting broad expression. We speculate that functional differences among these broadly expressed proprotein convertases may lie in their substrate preferences and/or in their subcellular localization, but testing this idea is beyond the scope of this study. If the absence of adult tissue-type specialization for these genes is conserved among bilaterians, it would suggest that tissue-specific expression patterns of mammalian PCSK4 and *Hau-pcsk1.7* may reflect neo-functionalization after gene duplication.



## Expression patterns of proprotein convertase genes in DV patterning of the germinal band

A key function of furin and PCSK5-7 in vertebrates is to process the developmentally important BMP ligands by removing the prodomain (Dubois, et al. 1995; Cui, et al. 1998; Constam and Robertson 2000; Essalmani, et al. 2008; Nelsen and Christian 2009). Although they are broadly expressed in adult tissue, selected proprotein convertase genes are upregulated in cells of the embryonic signaling centers that produce active BMP and other TGF $\beta$  ligands in vertebrates (Constam, et al. 1996; Nelsen, et al. 2005). To see if this phenomenon is conserved between the leech and vertebrates, we next examined the expression patterns of the eight furin-like PCSKs, PCSK7 and PCSKX in stage-8 embryos of *Helobdella*. In this stage, BMP genes are expressed and are involved in dorsoventral patterning of the trunk ectoderm (Kuo and Weisblat 2011b), and we focused on testing if any of these proprotein convertases is upregulated in the critical BMP-expressing cells in the dorsoventral patterning process.

In the leech, ectoderm and mesoderm of the segmented trunk arise from five bilateral pairs of teloblast stem cells (Weisblat and Shankland 1985). These five pairs of teloblasts arise from the D quadrant of the four-cell embryo through a stereotypic series of asymmetric cell divisions (Figure 6A). The M teloblasts give rise to the mesoderm, and the N, O/P, O/P and Q teloblasts to the ectoderm. Each teloblast undergoes a repetitive series of asymmetric cell division to produce a bandlet of blast cells, each of which develops in a lineage-specific manner and gives rise to a definitive set of differentiated cells. In stage 8, the five bandlets from the five ipsilateral teloblasts converge at the posterior end of the embryo to form left and right germinal bands, respectively (Figure 6B). Within each germinal band, the four ectodermal bandlets form a parallel array superficial to the mesodermal bandlet. The dorsalmost ectodermal bandlet, the q bandlet, expresses BMP5-8 (Figure 6C), which induces the dorsolateral, P fate in cells of the adjacent bandlet produced by one of the O/P teloblasts (Kuo and Weisblat 2011b). Cells adopt the ventrolateral, O fate in the other O/P-derived bandlet that is not in direct contact with the q bandlet (Huang and Weisblat 1996; Kuo and Weisblat 2011b). Fates of the ventralmost ectodermal bandlet, the n bandlet, are not regulated by BMP5-8 signaling (Kuo and Weisblat 2011b). In addition to the q bandlet-specific BMP5-8 expression (Figure 6C), three other BMP genes (BMP2/4a, BMP2/4b, and ADMP) are broadly expressed

in the germinal band (Kuo and Weisblat 2011b). The BMP2/4s also contribute to generating a dorsoventral gradient of BMP activity in the germinal band, but the function of ADMP is largely unknown (Kuo and Weisblat 2011b).

Given the importance of the q bandlet in the BMP-mediated dorsoventral patterning in the leech, we predicted that one or more proprotein convertases would be upregulated in this region. In contrast to this prediction, however, the proprotein convertases, if expressed in the germinal band, showed a broad but weak expression throughout the germinal bands, and none exhibited an elevated expression level in the q bandlet (Figure 6D-M). For *Hau-furin1.1*, *Hau-furin1.2*, *Hau-furin1.3*, and *Hau-furin1.4*, the staining appeared to be somewhat stronger in the mesodermal bandlets (Figure 6D-6G). Staining for *Hau-furin1.5* was barely above the background level and was also stronger in the mesodermal bandlets (Figure 6H). *Hau-furin1.6* was expressed in all but the four O/P teloblasts (Figure 6I, 6I'). Expression of *Hau-furin1.7*, *Hau-furin2-like*, and *Hau-pcsk7* was not detected in stage 8 embryos (Figure 5J-L); this latter result was unlikely to arise from any technical issue since these probes did detect transcript expression, at even lower probe concentrations (1/5 of that used on the stage-8 embryo) when applied to juvenile specimens. *Hau-pcskx* transcript was somewhat elevated in the cytoplasm of macromeres as well as cells of the isolated o/p bandlets before they converge to form the germinal band (Figure 6M). Given that a proprotein convertase generally cleaves multiple substrates, these PCSKs may act on other molecules than BMPs. In any case, our observation implies that proteolytic processing of the BMP ligands may not have the same role in generating their axial patterning activity in the leech as in the vertebrates.

### **Proteolytic processing by proprotein convertase is not required for the dorsalizing activity of Hau-BMP5-8**

The low levels of proprotein convertase transcripts in the q bandlet suggest that levels of convertase activity are low in the germinal band tissue. If so, it could be (1) that cleavage by PCSK is irrelevant to the P fate-inducing activity of Hau-BMP5-8, or (2) that convertase-mediated cleavage is a rate-limiting step in P fate induction. To distinguish between these two alternatives, we sought to determine whether convertase-dependent proteolytic processing of Hau-BMP5-8 is required for its P fate-inducing activity, as assayed by the induction of *Hau-six1/2a* expression in the O-P equivalence group (Figure 7C). In BMPs and other members of

the TGF $\beta$  family, the linker sequence between the prodomain and the bioactive TGF $\beta$  cysteine-knot domain contains at least one, often more than one, convertase cleavage site(s) (Derynck and Miyazono 2008). In Hau-BMP5-8, three such -RXXR- sites were identified, and we mutated all three sites into -GXXG- to prevent recognition and cleavage by convertases (Figure 7A). To see if this convertase-resistant Hau-BMP5-8 can induce P fate, we injected *in vitro* transcribed mRNA encoding the mutant Hau-BMP5-8 into the N teloblast and found that the o bandlet, which neighbors the n bandlet, was invariably induced to express the P fate marker, *Hau-six1/2a* ectopically (Figure 7E). This result was identical to that obtained from embryos in which wild-type Hau-BMP5-8 was expressed in the n bandlet (Figure 7D), indicating that proteolytic processing by proprotein convertase at the linker sequence is not required for the P fate-inducing activity of Hau-BMP5-8.

To further test the rate-limiting hypothesis, we next over-expressed the full-length wild-type Hau-furin1.1 in the germinal band by injecting its mRNA into a NOPQ proteloblast. In such embryos, no change in the expression of the P fate marker *Hau-six1/2a* was observed (Figure 7F). This result negated the hypothesis that convertase activity was rate-limiting for P-fate induction. Together, these results showed that proteolytic processing by convertase was neither necessary for P fate-inducing activity of Hau-BMP5-8, nor could it enhance BMP signaling in the leech germinal band.

On the other hand, the presence of proprotein convertase, even at a background level, should cleave at least a portion of the Hau-BMP5-8 protein. In an attempt to verify the cleavage of Hau-BMP5-8 in the germinal band, we generated a tagged Hau-BMP5-8 construct, in which an HA tag was inserted to the N-terminal region of the prodomain, and a FLAG tag was placed downstream to the convertase sites at the N-terminus of the TGF $\beta$  cysteine-knot domain. Cleavage of Hau-BMP5-8 should result in a separation of the HA-tagged prodomain fragment from the FLAG-tagged TGF $\beta$  cysteine-knot domain fragment. This tagged Hau-BMP5-8 is biologically active since it is capable of inducing *Hau-six1/2a* expression in the o bandlet when ectopically expressed in the n bandlet (Figure 7G). We then performed Western blot using lysates of embryos expressing the tagged Hau-BMP5-8 in the q bandlet or in the entire ectoderm. Despite multiple trials under various conditions and using different antibodies, such Western blot experiments were not successful due to commercial antibodies' prevalent

cross-reactivities to endogenous leech antigens. We also replaced both HA and FLAG tags in the Hau-BMP5-8 construct with GFP, but unfortunately, the same technical issue persisted.

Convertase-mediated cleavage separates the bioactive TGF $\beta$  cysteine-knot domain from the prodomain (Panganiban, et al. 1990; Dubois, et al. 1995; Cui, et al. 1998; Birsoy, et al. 2005), and it has been shown that association with prodomain may bring about different consequences for different TGF $\beta$ -family ligands, in different species, and in different cellular environments: in some cases, the prodomain anchors the TGF $\beta$ -family ligand to the extracellular matrix; in others, the prodomain prevents the ligand from binding its receptor; in yet other cases, the prodomain has no effect on signaling activity of the ligand (Gregory, et al. 2005; Sengle, et al. 2011; Jiang, et al. 2016). Although we had not succeeded in verifying the cleavage of Hau-BMP5-8 in the germinal band, we nevertheless demonstrated that such cleavage is irrelevant to Hau-BMP5-8's ability to induce the P fate. This finding is in stark contrast to the situation in vertebrate embryos, in which overexpression of BMP4 or BMP7 with mutated cleavage sites has a dominant-negative effect (Hawley, et al. 1995); on the other hand the apparent irrelevance of convertase-mediated cleavage in the leech germinal band resembles the convertase-resistant mutant of the zebrafish nodal homolog Southpaw, which exhibits a similarly short-range signal activity, rather than a complete loss of signal (Tessadori, et al. 2015). The endogenous Hau-BMP5-8 may likewise be in a prodomain-bound form, regardless of whether it is proteolytically processed or not, and the presence of the prodomain may restrict the signaling range of BMP5-8, without interfering with its receptor binding.

Although technical difficulties prevented us from demonstrating the cleavage of Hau-BMP5-8 directly, finding that over-expression of proprotein convertases did not expand the signal range of Hau-BMP5-8 suggests that its prodomain and TGF $\beta$  cysteine-knot domain may have remained in association even after being cleaved. This would explain the convertase-independent, short-range signaling activity in the q bandlet and for the absence of a lineage-specific expression of proprotein convertase. We note that the *Drosophila* Gbb ligand, also a BMP5-8 homolog, has a stronger signaling activity and a longer signaling range when covalently linked to the C-terminal portion of its prodomain than as the fully processed short ligand (Akiyama, et al. 2012). In contrast to Gbb, however, the unprocessed Hau-BMP5-8 would contain the complete prodomain because it lacks the homologous cleavage site within its prodomain; this may account for the different behaviors of these two BMP5-8 ligands. In any

case, together, these cases have revealed diversity in the regulation of how the BMP ligands are distributed in extracellular space and the differing role of proprotein convertases in regulating the signaling range of BMP ligands in different biological contexts.

## Conclusions

Proprotein convertases are a well-studied group of endoproteases that have multiple roles in animal development and physiology. Here, we presented a large-scale phylogenetic survey of proprotein convertases across metazoan species. Our general findings were: (1) that there are six ancestral bilaterian OGs of Kexin-like convertases, including the previously unknown PCSKX; (2) spiralian, in general, have retained all six OGs while specific OGs were apparently lost in other lineages.

We have also used the leech *Helobdella* as a tractable spiralian model in which to carry out to begin a functional characterization of convertase OGs. Our results show that functional divergence among the PCSK OGs was probably already in place in the bilaterian ancestor. In contrast to the absolute requirement for convertase processing of BMP ligands in vertebrate embryos, we demonstrate an instance of short-range, convertase-independent BMP signaling in dorsoventral patterning of the leech embryo. Overall, our findings highlight the somewhat overlooked significance of such conserved “housekeeping” enzymes in the diversification of complex biological signaling systems and illustrate the importance of studying non-model organisms for illuminating the breadth of developmental mechanisms.

## Methods

### Gene Identification and Phylogenetic Analysis.

The *Helobdella* proprotein convertase genes were identified by querying the filtered model protein dataset in the JGI *Helobdella robusta* database (<https://genome.jgi.doe.gov/Helro1/Helro1.home.html>) for proteins containing the Peptidase\_S8 domain (Pfam: PF00082) and by a BLASTP search against using human furin as a query. The results from these two searches were compared, and the retrieved sequences were analyzed using HMMER to evaluate the completeness of the gene models. The incomplete models were manually curated by referencing the sequences of the isolated cDNA clones.

For phylogenetic reconstruction, gene models from other species were collected by querying the Ensembl genome databases. The amphioxus genome is not included in the Ensembl collection, and we queried the JGI *Branchiostoma floridae* database (<https://genome.jgi.doe.gov/Braf11/Braf11.home.html>) instead. Models missing any of the defining criteria were further analyzed by BLASTP search against the GenBank to identify the false positives and the incomplete models. Incomplete models were included in the gene copy analysis but were excluded from the phylogenetic analysis. Descriptive statistical data were processed and plotted using R packages. Conceptually translated amino acid sequences were aligned using MUSCLE (Edgar 2004b, a). RAXML (Stamatakis 2014) was used to produce Maximum Likelihood trees. Phylogenetic trees and statistical plots were edited to improve visual presentation by using Adobe Illustrator CS6.

### **Molecular Cloning of *Helobdella* proprotein convertases genes**

DNA fragments corresponding to the coding region of the *Helobdella* proprotein convertase genes were PCR amplified from a pool of mix-stage *H. austinensis* cDNA. *H. austinensis* is a sibling species of *H. robusta* and is more ready to breed and easier to maintain in the laboratory environment (Kutschera, et al. 2013). All experiments described here were performed with *H. austinensis* instead of *H. robusta*. PCR primers (Table S2) were designed using sequence information uncovered from the *H. robusta* genome database, and PCR reaction was performed using the proof-reading thermostable DNA polymerase Phusion (ThermoFisher). The obtained cDNA fragments were inserted into pGEM-T vector (Promega) by following standard cloning procedures. The cDNA clones were isolated and sequenced, and the sequence information was used to correct and supplement the gene models in the JGI *H. robusta* database, in case of mistaken annotation.

### **Whole-Mount *In Situ* Hybridization**

Linear DNA templates for probe synthesis were PCR amplified from the cDNA clones in the pGEM-T vector using SP6 and T7 primers and purified by gel extraction. Digoxigenin labeled antisense probes were *in vitro* transcribed with either SP6 or T7 RNA polymerase (ThermoFisher), depending on the orientation of cDNA insertion in the vector. Probe synthesis and whole-mount *in situ* hybridization were performed as previously described (Weisblat and

Kuo 2009). The color reaction was carried out by using BM Purple alkaline phosphatase substrate (Roche). Following the color reaction, the specimens were stored in ethanol and were mounted in 70% glycerol solution for imaging. Digital images were obtained by using an Olympus DP80 CCD camera mounted on an Olympus BX63 microscope. The acquired images were adjusted for size, contrast, and brightness by using Adobe Photoshop CS6, and the adjusted images were assembled into figure panels and annotated using Adobe Illustrator CS6.

## Expression of Molecular Constructs

Two different approaches were used to express Hau-BMP5-8 variants in the specified cell lineage. In one, synthesized capped RNA was injected into a teloblast or proteloblast (Zhang and Weisblat 2005); alternatively, plasmid DNA carrying the leech EF1 $\alpha$  upstream sequence that is capable of driving transgene expression (Pilon and Weisblat 1997) was injected instead. These two methods yielded similar results. All experiments were performed with a >30 sample size; similar to the previously published Hau-BMP5-8 mis-expressing experiments (Kuo and Weisblat 2011b), invariable results were obtained in each set of experiments

Capped RNA was *in vitro* transcribed from linearized plasmid template using mMESSAGE mMACHINE SP6 Kit (ThermoFisher) and was further polyadenylated by using Poly(A) Tailing Kit (ThermoFisher). The plasmid template for the wild-type Hau-BMP5-8, pCS107-Hau-bmp5-8, and the dosage of mRNA injection was also the same as previously described (Kuo and Weisblat 2011b). The Hau-BMP5-8 variants with epitope tag insertion and/or site-directed mutagenesis were generated by inverse PCR from the wild-type pCS107-Hau-bmp5-8 plasmid template, using a previously described strategy (Fisher and Pei 1997). The proofreading thermophilic DNA polymerase Phusion (ThermoFisher) was used in these PCR reactions. Constructs with multiple tags and/or mutations were produced by stepwise introduction of individual insertions and mutations. The primer pairs that were used to generate a specific construct were listed in Table S3.

The coding region of each Hau-BMP5-8 variant was then subcloned from the pCS107 plasmid into a site immediately downstream to the leech EF1 $\alpha$  upstream sequence in pMiEF1P, which was modified from a Minos transposon vector (Pavlopoulos and Averof 2005). Plasmid

DNA was injected at a concentration of 100 ng/μl in micropipette as previously described (Gline, et al. 2009).

Overexpression of Hau-furin1.1 was performed by plasmid injection. The full-length coding region cDNA of Hau-furin1.1 was PCR amplified by using a primer pair listed in Table S3 and then inserted into the pMIEF1P to produce the expression plasmid pMIEF1P-Hau-furin1.1.

## Acknowledgments

This work was supported by a grant from MOST, Taiwan (102-2311-B-002-024-MY2) to DHK and NIH grant 2R01GM074619 to DAW.

## References

- Agata K, Soejima Y, Kato K, Kobayashi C, Umesono Y, Watanabe K. 1998. Structure of the planarian central nervous system (CNS) revealed by neuronal cell markers. *Zool Sci* 15:433-440.
- Akiyama T, Marqués G, Wharton KA. 2012. A large bioactive BMP ligand with distinct signaling properties is produced by alternative proconvertase processing. *Sci Signal* 5:ra28.
- Bertrand S, Camasses A, Paris M, Holland ND, Escriva H. 2006. Phylogenetic analysis of amphioxus genes of the proprotein convertase family, including aPC6C, a marker of epithelial fusions during embryology. *Int J Biol Sci* 2:125-132.
- Birsoy B, Berg L, Williams PH, Smith JC, Wylie CC, Christian JL, Heasman J. 2005. XPACE4 is a localized pro-protein convertase required for mesoderm induction and the cleavage of specific TGFβ proteins in *Xenopus* development. *Development* 132:591-602.
- Bryant MK, Schardl CL, Hesse U, Scott B. 2009. Evolution of a subtilisin-like protease gene family in the grass endophytic fungus *Epichloë festucae*. *BMC Evol Biol* 9:168.
- Cameron J, Holla ØL, Berge KE, Kulseth MA, Ranheim T, Leren TP, Laerdahl JK. 2008. Investigations on the evolutionary conservation of PCSK9 reveal a functionally important protrusion. *FEBS J* 275:4121-4133.
- Cavalier-Smith T, Chao EE, Snell EA, Berney C, Fiore-Donno AM, Lewis R. 2014. Multigene eukaryote phylogeny reveals the likely protozoan ancestors of opisthokonts (animals, fungi, choanozoans) and Amoebozoa. *Mol Phylogenet Evol* 81:71-85.



- Constam DB. 2014. Regulation of TGF $\beta$  and related signals by precursor processing. *Sem Cell Dev Biol* 32:85-97.
- Constam DB, Calfon M, Robertson EJ. 1996. SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone morphogenetic proteins at distinct sites during embryogenesis. *J Cell Biol* 134:181-191.
- Constam DB, Robertson EJ. 2000. SPC4/PACE4 regulates a TGF $\beta$  signaling network during axis formation. *Genes Dev* 14:1146-1155.
- Cui Y, Hackenmiller R, Berg L, Jean F, Nakayama T, Thomas G, Christian JL. 2001. The activity and signaling range of mature BMP-4 is regulated by sequential cleavage at two sites within the prodomain of the precursor. *Genes Dev* 15:2797-2802.
- Cui Y, Jean F, Thomas G, Christian JL. 1998. BMP-4 is proteolytically activated by furin and/or PC6 during vertebrate embryonic development. *EMBO J* 17:4735-4743.
- Cummins SF, York PS, Hanna PH, Degnan BM, Croll RP. 2009. Expression of prohormone convertase 2 and the generation of neuropeptides in the developing nervous system of the gastropod *Haliothis*. *Int J Dev Biol* 53:1081-1088.
- Degnin C, Jean F, Thomas G, Christian JL. 2004. Cleavages within the prodomain direct intracellular trafficking and degradation of mature bone morphogenetic protein-4. *Mol Biol Cell* 15:5012-5020.
- Derynck R, Miyazono K. 2008. The TGF- $\beta$  Family. In. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. p. 1114.
- Dubois CM, Laprise M-H, Blanchette F, Gentry LE, Leduc R. 1995. Processing of transforming growth factor  $\beta$ 1 precursor by human furin convertase. *J Biol Chem* 270:10618-10624.
- Edgar RC. 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform* 5:113.
- Edgar RC. 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.
- Engelhardt R-P, Dhainaut-Courtois N, Tramu G. 1982. Immunohistochemical demonstration of a CCK-like peptide in the nervous system of a marine annelid worm, *Nereis diversicolor* O.F. Müller. *Cell Tissue Res* 227:401-411.

- Essalmani R, Hamelin J, Marcinkiewicz J, Chamberland A, Mbikay M, Chrétien M, Seidah NG, Prat A. 2006. Deletion of the gene encoding proprotein convertase 5/6 causes early embryonic lethality in the mouse. *Mol Cell Biol* 26:354-361.
- Essalmani R, Zaid A, Marcinkiewicz J, Chamberland A, Pasquato A, Seidah NG, Prat A. 2008. *In vivo* functions of the proprotein convertase PC5/6 during mouse development: Gdf11 is a likely substrate. *Proc Nat Acad Sci U S A* 105:5750-5755.
- Fan X, Nagle GT. 1996. Molecular cloning of *Aplysia* neuronal cDNAs that encode carboxypeptidases related to mammalian prohormone processing enzymes. *DNA Cell Biol* 15:937-945.
- Feuda R, Dohrmann M, Pett W, Philippe H, Rota-Stabelli O, Lartillot N, Wörheide G, Pisani D. 2017. Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. *Curr Biol* 27:3864-3870.
- Fisher CL, Pei GK. 1997. Modification of a PCR-based site-directed mutagenesis method. *BioTechniques* 23:570-574.
- Flot J-F, Hespels B, Li X, Noel B, Arkhipova I, Danchin EG, Hejnal A, Henrissat B, Koszul R, Aury JM, et al. 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature* 500:453-457.
- Fritsch C, Sawala A, Harris R, Maartens A, Sutcliffe C, Ashe HL, Ray RP. 2012. Different requirements for proteolytic processing of bone morphogenetic protein 5/6/7/8 ligands in *Drosophila melanogaster*. *J Biol Chem* 287:5942-5953.
- Furuta M, Yano H, Zhou A, Rouillé Y, Holst JJ, Carroll R, Ravazzola M, Orci L, Furuta H, Steiner DF. 1997. Defective prohormone processing and altered pancreatic islet morphology in mice lacking active SPC2. *Proc Nat Acad Sci U S A* 94:6646-6651.
- Gline SE, Kuo D-H, Stolfi A, Weisblat DA. 2009. High resolution cell lineage tracing reveals developmental variability in leech. *Dev Dyn* 238:3139-3151.
- Goldman DC, Hackenmiller R, Nakayama T, Sopory S, Wong C, Kulesa H, Christian JL. 2006. Mutation of an upstream cleavage site in the BMP4 prodomain leads to tissue-specific loss of activity. *Development* 133:1933-1942.
- Gomot A, Gomot L, Marchand CR, Colard C, Bride J. 1992. Immunocytochemical localization of insulin-related peptide(s) in the central nervous system of the snail *Helix aspersa* Müller: involvement in growth control. *Cell Mol Neurobiol* 12:21-32.

- Gorham EL, Nagle GT, Smith JS, Shen H, Kurosky A. 1996. Molecular cloning of prohormone convertase 1 from the atrial gland of *Aplysia*. *DNA Cell Biol* 15:339-345.
- Gregory KE, Ono RN, Charbonneau NL, Kuo CL, Keene DR, Bächinger HP, Sakai LY. 2005. The prodomain of BMP-7 targets the BMP-7 complex to the extracellular matrix. *J Biol Chem* 280:27970-27980.
- Grimmelikhuijzen CJ, Sundler F, Rehfeld JF. 1980. Gastrin/CCK-like immunoreactivity in the nervous system of coelenterates. *Histochemistry* 69:61-68.
- Hawley SHB, Wünnenberg-Stapleton K, Hashimoto C, Laurent MN, Watabe T, Blumberg BW, Cho KWY. 1995. Disruption of BMP signals in embryonic *Xenopus* ectoderm leads to direct neural induction. *Genes Dev* 9:2923-2935.
- Huang FZ, Weisblat DA. 1996. Cell fate determination in an annelid equivalence group. *Development* 122:1839-1847.
- Hur JH, Van Doninck K, Mandigo ML, Meselson M. 2009. Degenerate tetraploidy was established before bdelloid rotifer families diverged. *Mol Biol Evol* 26:375-383.
- Husson SJ, Clynen E, Baggerman G, Janssen T, Schoofs L. 2006. Defective processing of neuropeptide precursors in *Caenorhabditis elegans* lacking proprotein convertase 2 (KPC-2/EGL-3): mutant analysis by mass spectrometry. *J Neurochem* 98:1999-2012.
- Izquierdo-Carrasco F, Smith SA, Stamatakis A. 2011. Algorithms, data structures, and numerics for likelihood-based phylogenetic inference of huge trees. *BMC Bioinform* 12:470.
- Jiang H, Salmon RM, Upton PD, Wei Z, Lawera A, Davenport AP, Morrell NW, Li W. 2016. The prodomain-bound form of bone morphogenetic protein 10 is biologically active on endothelial cells. *J Biol Chem* 291:2954-2966.
- Künnapu J, Björkgren I, Shimmi O. 2009. The *Drosophila* DPP signal is produced by cleavage of its proprotein at evolutionary diversified furin-recognition sites. *Proc Nat Acad Sci U S A* 106:8501-8506.
- Kellner-Cousin K, Mialhe E, Mathieu M. 1994. Identification of insulin-like peptides in cerebral ganglia neurosecretory cells of the mussel *Mytilus edulis*. *Tissue Cell* 26:891-899.
- Kellner K, Heude-Berthelin C, Mathieu M. 2002. Immunocytochemical demonstration of glucagon-like peptides in *Mytilus edulis* cerebral ganglia and an *in vitro* effect of vertebrate glucagon on glycogen metabolism. *Tissue Cell* 34:109-116.

- Kocot KM, Struck TH, Merkel J, Waits DS, Todt C, Brannock PM, Weese DA, Cannon JT, Moroz LL, Lieb B, et al. 2017. Phylogenomics of Lophotrochozoa with consideration of systematic error. *Syst Biol* 66:256-282.
- Kuo D-H, Weisblat DA. 2011a. Intermediate filament genes as differentiation markers in the leech *Helobdella*. *Dev Genes Evol* 221:225-240.
- Kuo D-H, Weisblat DA. 2011b. A new molecular logic for BMP-mediated dorsoventral patterning in the leech *Helobdella*. *Curr Biol* 21:1282-1288.
- Kutschera U, Langguth H, Kuo D-H, Weisblat DA, Shankland M. 2013. Description of a new leech species from North America, *Helobdella austinensis* n. sp. (Hirudinea: Glossiphoniidae), with observations on its feeding behaviour. *Zoosyst Evol* 89:239-246.
- Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol* 7:S4.
- Laumer CE, Bekkouche N, Kerbl A, Goetz F, Neves RC, Sorensen MV, Kristensen RM, Hejnol A, Dunn CW, Giribet G, et al. 2015. Spiralian phylogeny informs the evolution of microscopic lineages. *Curr Biol* 25:2000-2006.
- Li J, Gu F, Wu R, Yang J, Zhang K-Q. 2017. Phylogenomic evolutionary surveys of subtilase superfamily genes in fungi. *Sci Rep* 7:45456.
- Liu J-X, Srivastava R, Che P, Howell SH. 2007. Salt stress responses in *Arabidopsis* utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. *Plant J* 51:897-909.
- Mbikay M, Tadros H, Ishida N, Lerner CP, De Lamirande E, Chen A, El-Alfy M, Clermont Y, Seidah NG, Chrétien M, et al. 1997. Impaired fertility in mice deficient for the testicular germ-cell protease PC4. *Proc Nat Acad Sci U S A* 94:6842-6846.
- Meester I, Ramkema MD, van Minnen J, Boer HH. 1992. Differential expression of four genes encoding molluscan insulin-related peptides in the central nervous system of the pond snail *Lymnaea stagnalis*. *Cell Tissue Res* 269:183-188.
- Miller CM, Newmark PA. 2012. An insulin-like peptide regulates size and adult stem cells in planarians. *Int J Dev Biol* 56:75-82.
- Moser WE, Desser SS. 1995. Morphological, histochemical, and ultrastructural characterization of the salivary glands and proboscises of three species of glossiphoniid leeches (Hirudinea: Rhynchobdellida). *J Morphol* 225:1-18.

- Nagle GT, Garcia AT, Knock SL, Gorham EL, Van Heumen WRA, Kurosky A. 1995. Molecular cloning, cDNA sequence, and localization of a prohormone convertase (PC2) from the *Aplysia* atrial gland. *DNA Cell Biol* 14:145-154.
- Naureckiene S, Ma L, Sreekumar K, Purandare U, Lo CF, Huang Y, Chiang LW, Grenier JM, Ozenberger BA, Jacobsen JS, et al. 2003. Functional characterization of Narc 1, a novel proteinase related to proteinase K. *Arch Biochem Biophys* 420:55-67.
- Nelsen S, Berg L, Wong C, Christian JL. 2005. Proprotein convertase genes in *Xenopus* development. *Dev Dyn* 233:1038-1044.
- Nelsen SM, Christian JL. 2009. Site-specific cleavage of BMP4 by furin, PC6, and PC7. *J Biol Chem* 284:27157-27166.
- Nowell RW, Almeida P, Wilson CG, Smith TP, Fontaneto D, Crisp A, Micklem G, Tunnacliffe A, Boschetti C, Barraclough TG. 2018. Comparative genomics of bdelloid rotifers: insights from desiccating and nondesiccating species. *PLoS Biol* 16:e2004830.
- Ohno S. 1970. *Evolution by Gene Duplication*. New York, NY: Springer-Verlag.
- Oliva AA, Jr., Chan SJ, Steiner DF. 2000. Evolution of the prohormone convertases: identification of a homologue of PC6 in the protochordate amphioxus. *Biochim Biophys Acta* 1477:338-348.
- Ouimet T, Castellucci VF. 1997. Gene expression and regulation of *Aplysia californica* prohormone convertases aPC1B and aPC2. *J Neurochem* 68:1031-1040.
- Panganiban GEF, Rashka KE, Neitzel MD, Hoffmann FM. 1990. Biochemical characterization of the *Drosophila dpp* protein, a member of the transforming growth factor  $\beta$  family of growth factors. *Mol Cell Biol* 10:2669-2677.
- Pavlopoulos A, Averof M. 2005. Establishing genetic transformation for comparative developmental studies in the crustacean *Parhyale hawaiiensis*. *Proc Nat Acad Sci U S A* 102:7888-7893.
- Pestarino M. 1990. Immunocytochemical localization of glucagon-related peptides in the cerebral ganglion of the protochordate ascidian, *Styela plicata*. *J Neuroendocrinol* 2:641-645.
- Pilon M, Weisblat DA. 1997. A *nanos* homolog in leech. *Development* 124:1771-1780.
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490.

- Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064-1071.
- Qiu Q, Basak A, Mbikay M, Tsang BK, Gruslin A. 2005. Role of pro-IGF-II processing by proprotein convertase 4 in human placental development. *Proc Nat Acad Sci U S A* 102:11047-11052.
- Rayburn LYM, Gooding HC, Choksi SP, Maloney D, Kidd AR, III, Siekhaus DE, Bender M. 2003. *amontillado*, the *Drosophila* homolog of the prohormone processing protease PC2, is required during embryogenesis and early larval development. *Genetics* 163:227-237.
- Refaie S, Gagnon S, Gagnon H, Desjardins R, D'Anjou F, D'Orléans-Juste P, Zhu X, Steiner DF, Seidah NG, Lazure C, et al. 2012. Disruption of proprotein convertase 1/3 (PC1/3) expression in mice causes innate immune defects and uncontrolled cytokine secretion. *J Biol Chem* 287:14703-14717.
- Ren C, Chen T, Jiang X, Sun H, Qian J, Hu C, Wang Y. 2016. Two proprotein convertase subtilisin/kexin type 9 (PCSK9) paralogs from the tropical sea cucumber (*Stichopus monotuberculatus*): molecular characterization and inducible expression with immune challenge. *Fish Shellfish Immunol* 56:255-262.
- Rhea JM, Wegener C, Bender M. 2010. The proprotein convertase encoded by *amontillado* (*amon*) is required in *Drosophila corpora cardiaca* endocrine cells producing the glucose regulatory hormone AKH. *PLoS Genet* 6:e1000967.
- Rockel B, Kopec KO, Lupas AN, Baumeister W. 2012. Structure and function of tripeptidyl peptidase II, a giant cytosolic protease. *Biochim Biophys Acta* 1824:237-245.
- Roebroek AJM, Umans L, Pauli IGL, Robertson EJ, van Leuven F, Van de Ven WJM, Constam DB. 1998. Failure of ventral closure and axial rotation in embryos lacking the proprotein convertase Furin. *Development* 125:4863-4876.
- Seegmiller AC, Dobrosotskaya I, Goldstein JL, Ho YK, Brown MS, Rawson RB. 2002. The SREBP pathway in *Drosophila*: regulation by palmitate, not sterols. *Dev Cell* 2:229-238.
- Seidah NG. 2011. The proprotein convertases, 20 years later. *Meth Mol Biol* 768:23-57.
- Seidah NG, Benjannet S, Wickham L, Marcinkiewicz J, Jasmin SB, Stifani S, Basak A, Prat A, Chrétien M. 2003. The secretory proprotein convertase neural apoptosis-regulated

convertase 1 (NARC-1): liver regeneration and neuronal differentiation. Proc Nat Acad Sci U S A 100:928-933.

Seidah NG, Chrétien M. 1997. Eukaryotic protein processing: endoproteolysis of precursor proteins. Curr Opin Biotech 8:602-607.

Seidah NG, Chrétien M. 1999. Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. Brain Res 848:45-62.

Seidah NG, Day R, Hamelin J, Gaspar A, Collard MW, Chrétien M. 1992. Testicular expression of PC4 in the rat: molecular diversity of a novel germ cell-specific Kex2/subtilisin-like proprotein convertase. Mol Endocrinol 6:1559-1570.

Seidah NG, Mowla SJ, Hamelin J, Mamarbachi AM, Benjannet S, Touré BB, Basak A, Munzer JS, Marcinkiewicz J, Zhong M, et al. 1999. Mammalian subtilisin/kexin isozyme SKI-1: a widely expressed proprotein convertase with a unique cleavage specificity and cellular localization. Proc Nat Acad Sci U S A 96:1321-1326.

Sengle G, Ono RN, Sasaki T, Sakai LY. 2011. Prodomains of transforming growth factor  $\beta$  (TGF $\beta$ ) superfamily members specify different functions: extracellular matrix interactions and growth factor bioavailability. J Biol Chem 286:5087-5099.

Shimoyama S, Inoue T, Kashima M, Agata K. 2016. Multiple neuropeptide-coding genes involved in planarian pharynx extension. Zool Sci 33:311-319.

Siekhaus DE, Fuller RS. 1999. A role for *amontillado*, the *Drosophila* homolog of the neuropeptide precursor processing protease PC2, in triggering hatching behavior. J Neurosci 19:6942-6954.

Smeekens SP, Avruch AS, LaMendola J, Chan SJ, Steiner DF. 1991. Identification of a cDNA encoding a second putative prohormone convertase related to PC2 in AtT20 cells and islets of Langerhans. Proc Nat Acad Sci U S A 88:340-344.

Sonetti D, van Heumen WR, Roubos EW. 1992. Light- and electron-microscopic immunocytochemistry of a molluscan insulin-related peptide in the central nervous system of *Planorbarius corneus*. Cell Tissue Res 267:473-481.

Sopory S, Kwon S, Wehrli M, Christian JL. 2010. Regulation of Dpp activity by tissue-specific cleavage of an upstream site within the prodomain. Dev Biol 346:102-112.

- Stamatakis A. 2006. Phylogenetic models of rate heterogeneity: a high performance computing perspective. Proceedings 20th IEEE International Parallel and Distributed Processing Symposium; Rhodes Island, Greece. p. 278.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313.
- Tadros H, Chrétien M, Mbikay M. 2001. The testicular germ-cell protease PC4 is also expressed in macrophage-like cells of the ovary. *J Reprod Immunol* 49:133-152.
- Telford MJ, Budd GE, Philippe H. 2015. Phylogenomic insights into animal evolution. *Curr Biol* 25:R876-R887.
- Tessadori F, Noël ES, Rens EG, Magliozzi R, Evers-van Gogh IJA, Guardavaccaro D, Merks RMH, Bakkers J. 2015. Nodal signaling range is regulated by proprotein convertase-mediated maturation. *Dev Cell* 32:631-639.
- Thacker C, Rose AM. 2000. A look at the *Caenorhabditis elegans* Kex2/Subtilisin-like proprotein convertase family. *BioEssays* 22:545-553.
- Thomas G. 2002. Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Nat Rev Mol Cell Biol* 3:753-766.
- Tilak A, Nelsen SM, Kim H-S, Donley N, McKnite A, Lee H, Christian JL. 2014. Simultaneous rather than ordered cleavage of two sites within the BMP4 prodomain leads to loss of ligand in mice. *Development* 141:3062-3071.
- van Heumen WR, Roubos EW. 1990. Ultrastructural evidence for synthesis, storage and release of insulin-related peptides in the central nervous system of *Lymnaea stagnalis*. *Neuroscience* 39:493-500.
- Vigna SR, Morgan JL, Thomas TM. 1984. Localization and characterization of gastrin/cholecystokinin-like immunoreactivity in the central nervous system of *Aplysia californica*. *J Neurosci* 4:1370-1377.
- Wang S, Tulina N, Carlin DL, Rulifson EJ. 2007. The origin of islet-like cells in *Drosophila* identifies parallels to the vertebrate endocrine axis. *Proc Nat Acad Sci U S A* 104:19873-19878.
- Weisblat DA, Kuo D-H. 2014. Developmental biology of the leech *Helobdella*. *Int J Dev Biol* 58:429-443.



- Weisblat DA, Kuo D-H. 2009. *Helobdella* (Leech): a model for developmental studies. In. Emerging Model Organisms: a Laboratory Manual. Cold Spring Harbor, NY: CSHL Press. p. 245-267.
- Weisblat DA, Shankland M. 1985. Cell lineage and segmentation in the leech. *Philos Trans R Soc Lond B* 312:39-56.
- Wetsel WC, Rodriguez RM, Guillemot J, Rousselet E, Essalmani R, Kim IH, Bryant JC, Marcinkiewicz J, Desjardins R, Day R, et al. 2013. Disruption of the expression of the proprotein convertase PC7 reduces BDNF production and affects learning and memory in mice. *Proc Nat Acad Sci U S A* 110:17362-17367.
- Zhang SO, Weisblat DA. 2005. Applications of mRNA injections for analyzing cell lineage and asymmetric cell divisions during segmentation in the leech *Helobdella robusta*. *Development* 132:2103-2113.
- Zhang X, Pan H, Peng B, Steiner DF, Pintar JE, Fricker LD. 2010. Neuropeptidomic analysis establishes a major role for prohormone convertase-2 in neuropeptide biosynthesis. *J Neurochem* 112:1168-1179.
- Zhu X, Zhou A, Dey A, Norrbom C, Carroll R, Zhang C, Laurent V, Lindberg I, Ugleholdt R, Holst JJ, et al. 2002. Disruption of PC1/3 expression in mice causes dwarfism and multiple neuroendocrine peptide processing defects. *Proc Nat Acad Sci U S A* 99:10293-10298.

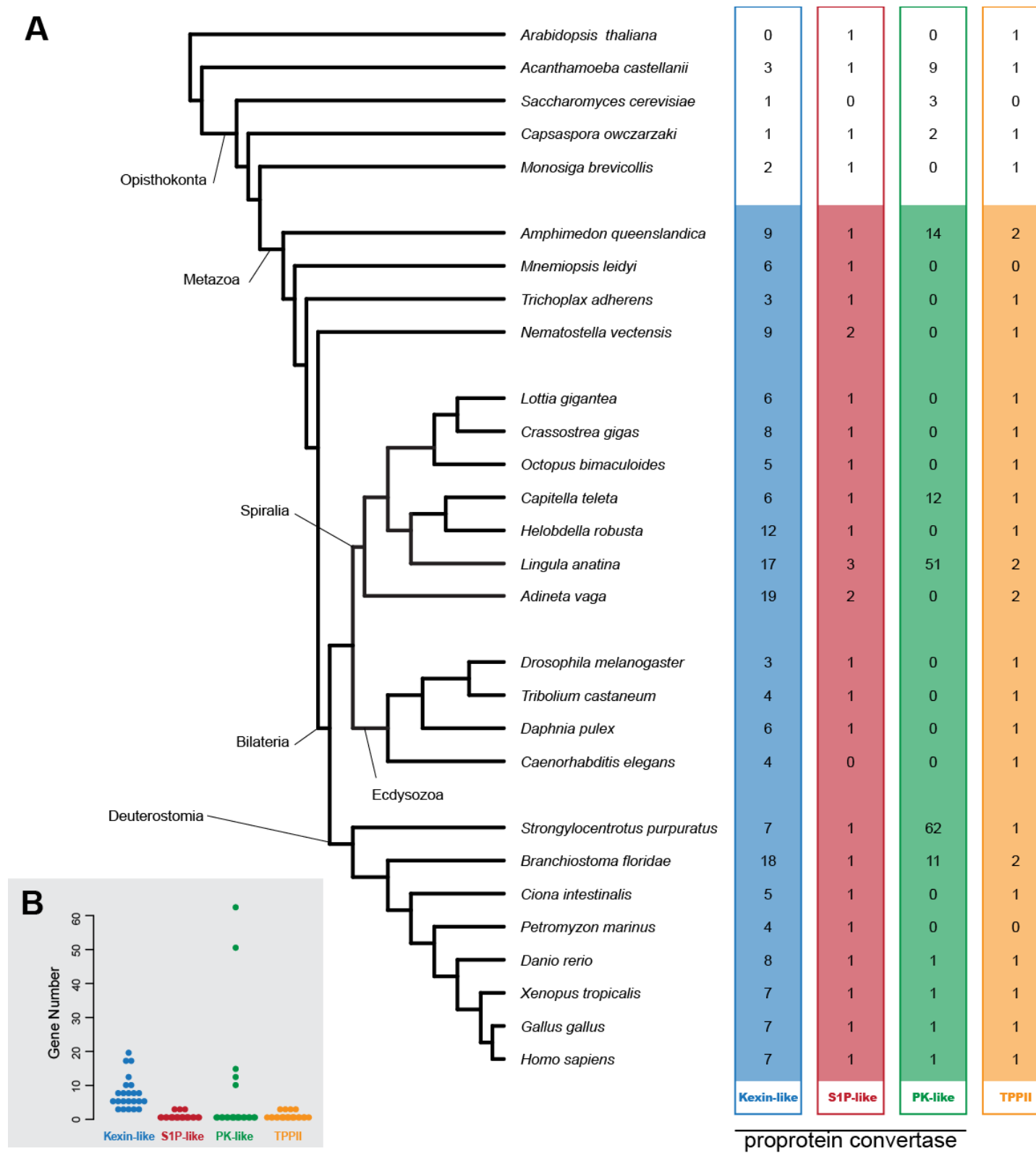


Figure 1

**Figure 1.** Phylogenetic distribution of metazoan S8 domain-containing peptidases. **A.** The number of Kexin-like, S1P-like, PK-like, and TPPII genes in 23 metazoan and five non-metazoan genomes analyzed in this study. **B.** Beeswarm plot showing the variability in the gene family size of Kexin-like (blue dots), S1P-like (red dots), PK-like (green dots), and TPP (orange dots) in individual metazoan genomes. Each dot represents the number of genes in the specified gene family in a metazoan genome.

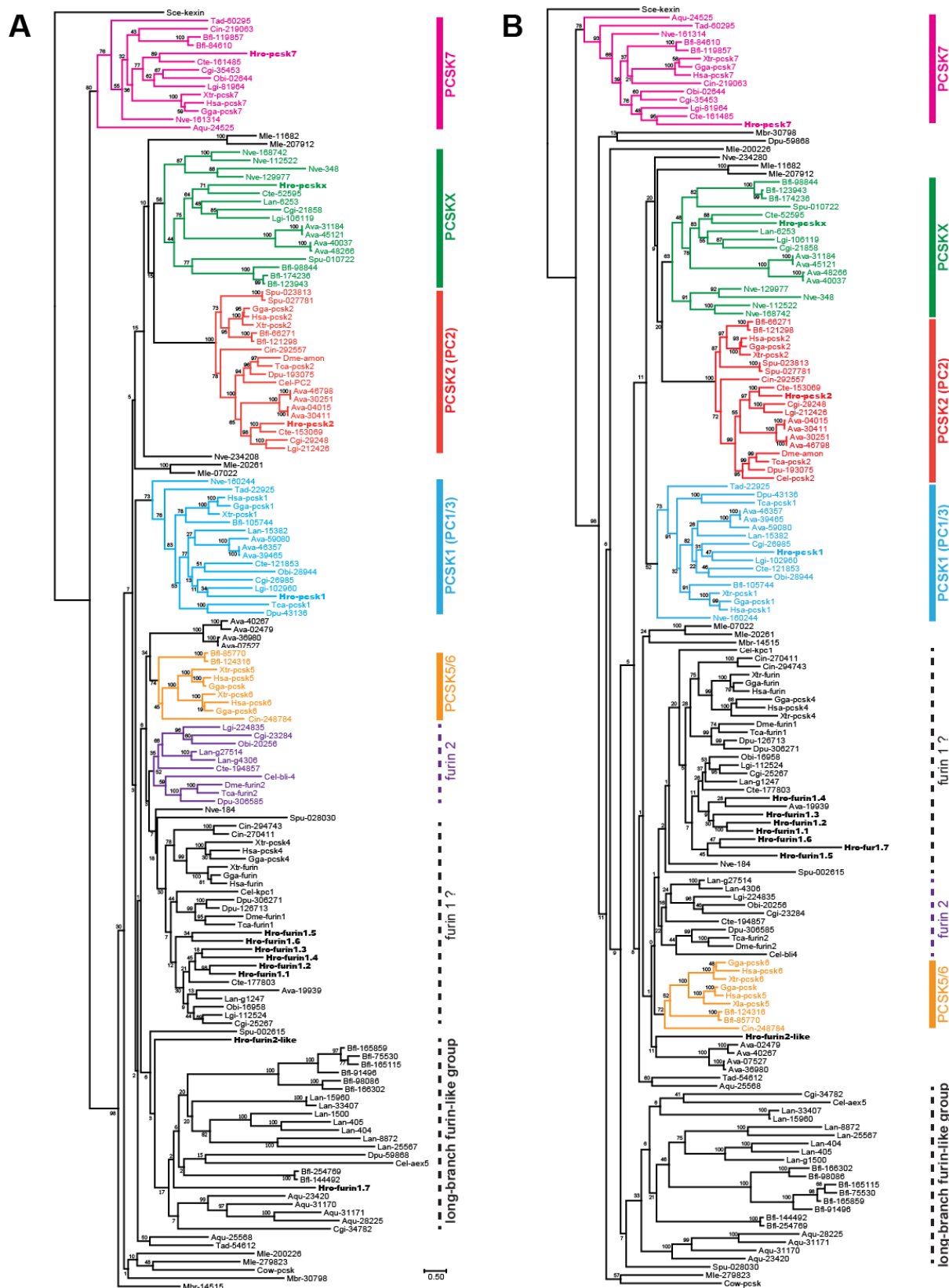


Figure 2.

**Figure 2.** Phylogenetic analysis of the 156 Kexin-like proteins from 21 metazoan and three opisthokont genomes. **A.** The best maximum-likelihood tree produced by using the LG model. **B.** The best maximum-likelihood tree produced by using the CAT-GTR model. 500X bootstraps were performed for each tree. The three-letter prefix in the taxon name indicates the species from which the protein was identified. The amino acid sequences retrieved from the genome of *Helobdella robusta* were labeled in bold types. See Table S1 for the details of the taxon label. The thick vertical bars on the right side of the tree mark better supported OG, whereas the dashed line denotes a poorly supported OG. The branching order of genes in the better supported OGs is generally in agreement with the species tree.

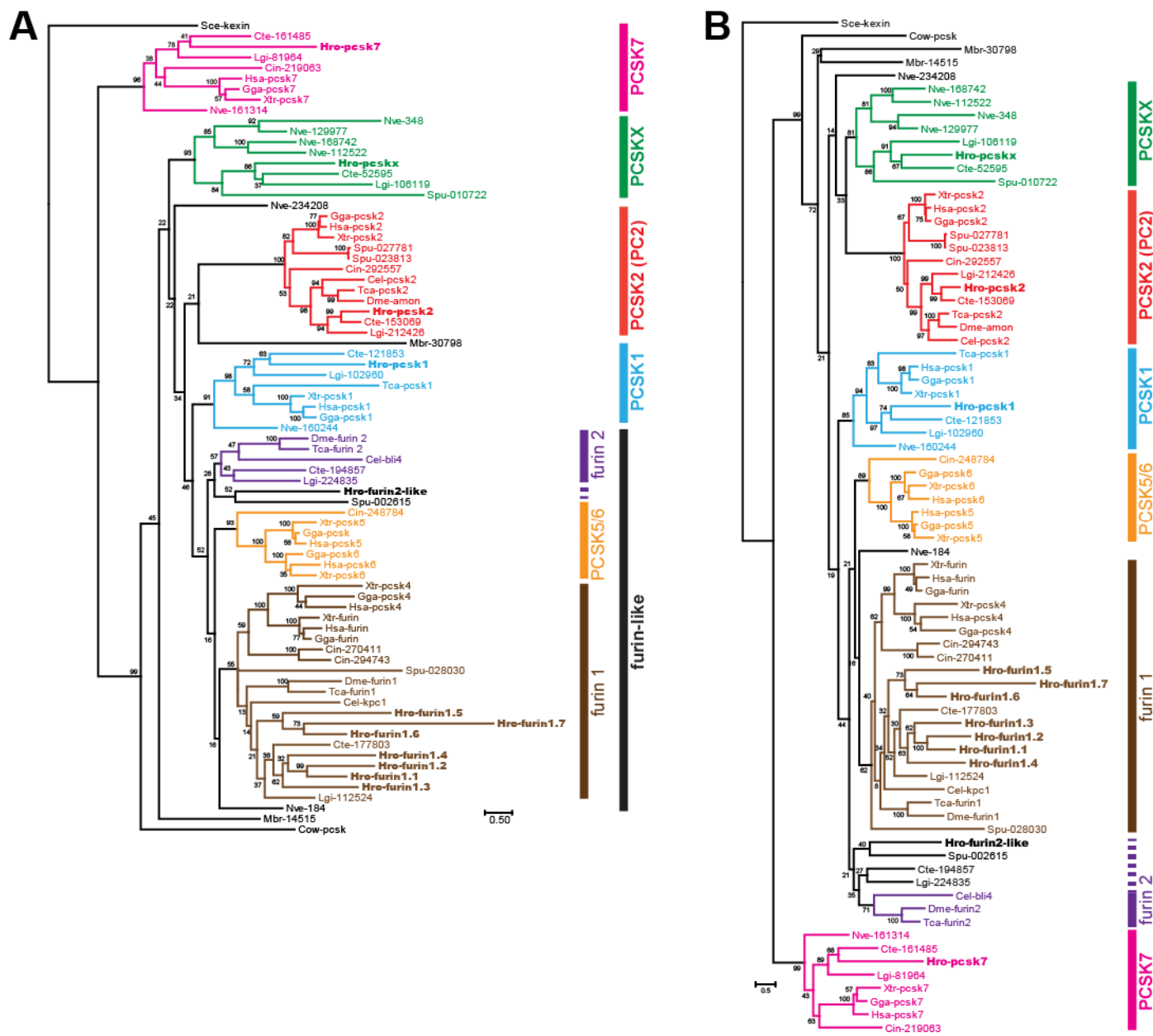
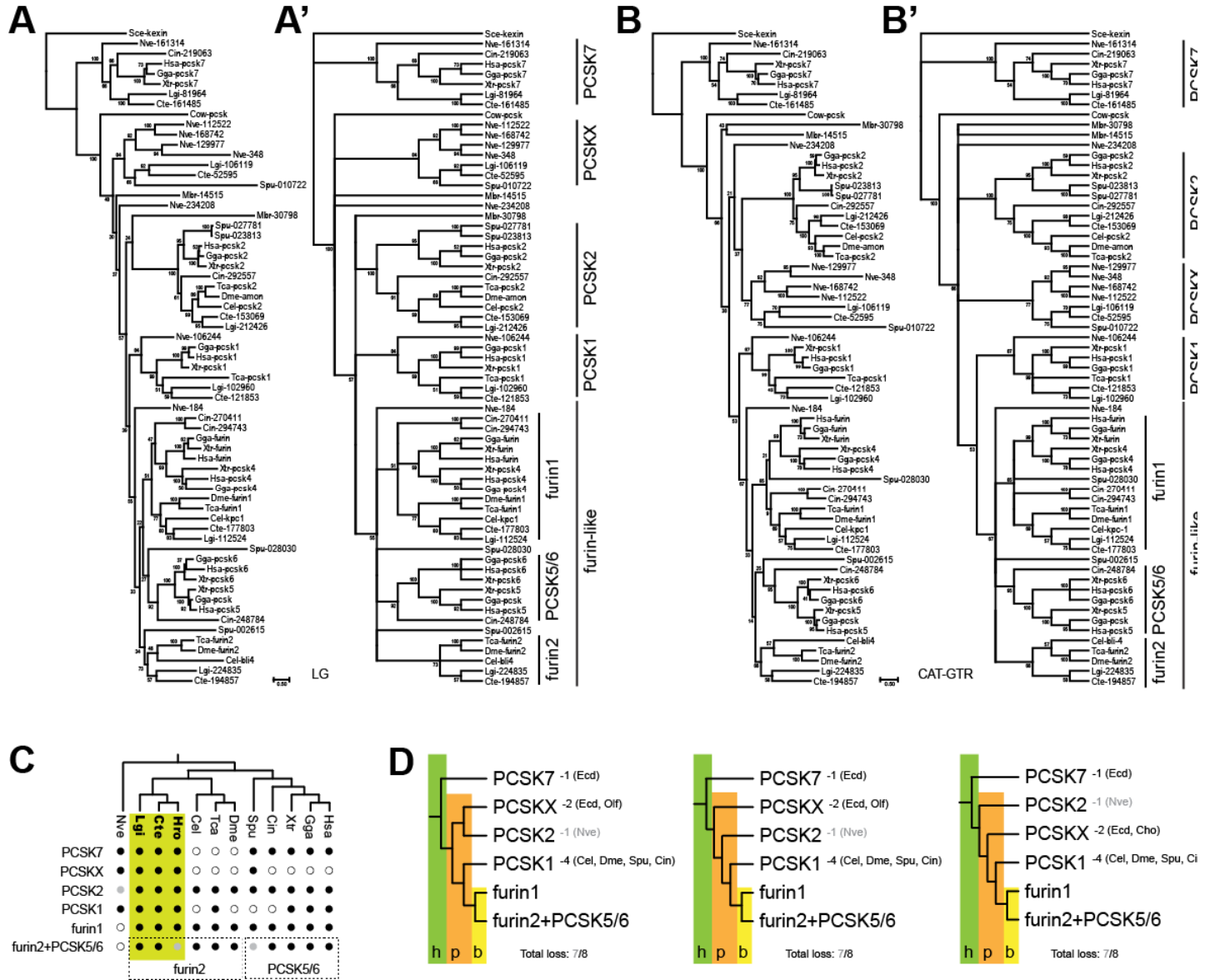


Figure 3.

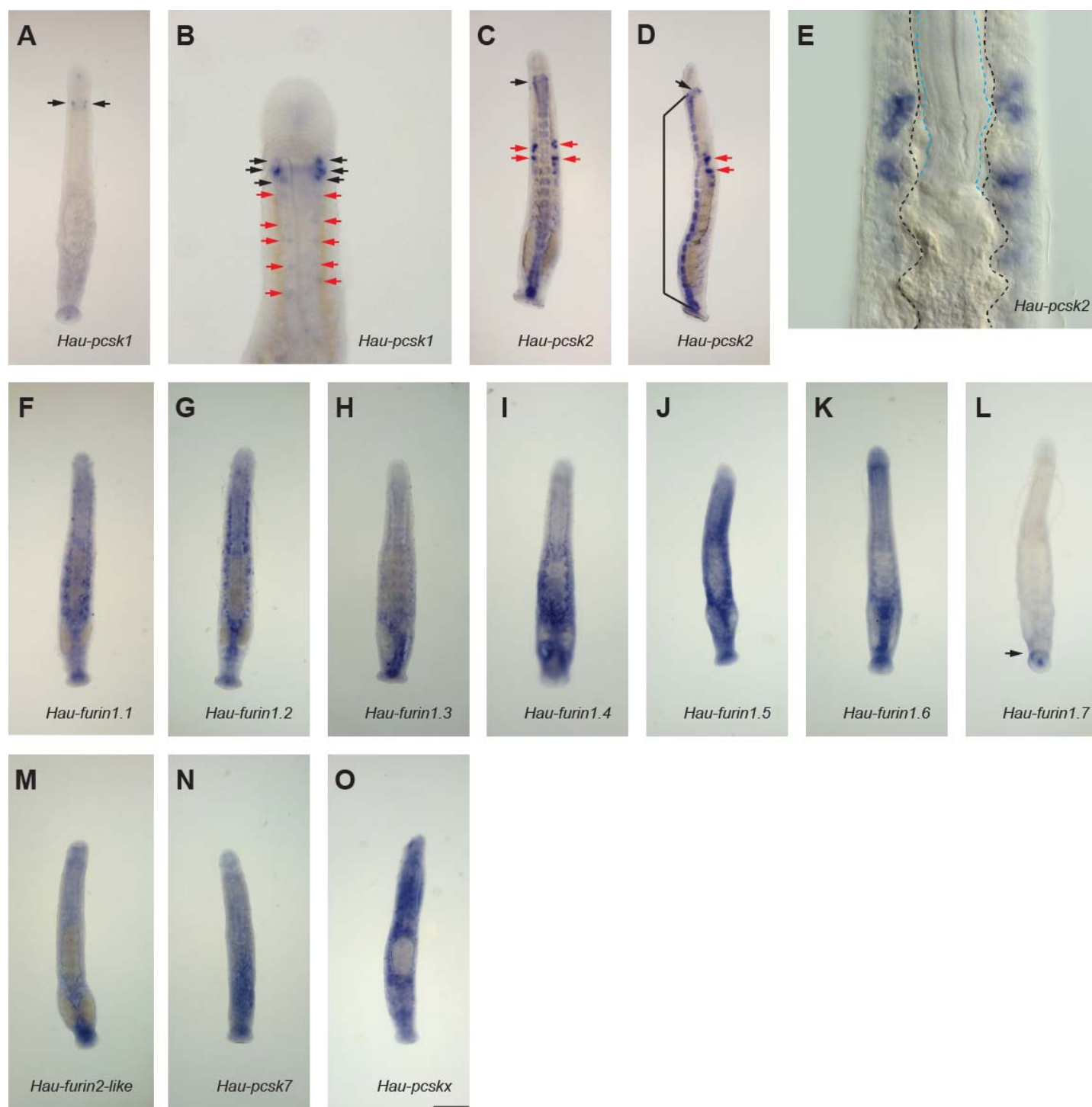
**Figure 3.** Assignment of the leech Kexin-like genes to specific orthologous groups (OGs) based on phylogenetic trees of the 77 proteins from 12 metazoan and three opisthokont genomes. **A.** The best maximum-likelihood tree produced by using the LG+G model. **B.** The best maximum-likelihood tree produced by using the CAT-GTR model. 500X bootstraps were performed for each tree. The three-letter prefix in the taxon name indicates the species from which the protein was identified (refer to Figure 1). The amino acid sequences retrieved from the genome of *Helobdella robusta* (Hro) are labeled in bold type. See Table S1 for the details of the taxon label. The thick vertical bar on the right side of the tree marks a well supported OG.



**Figure 4.**

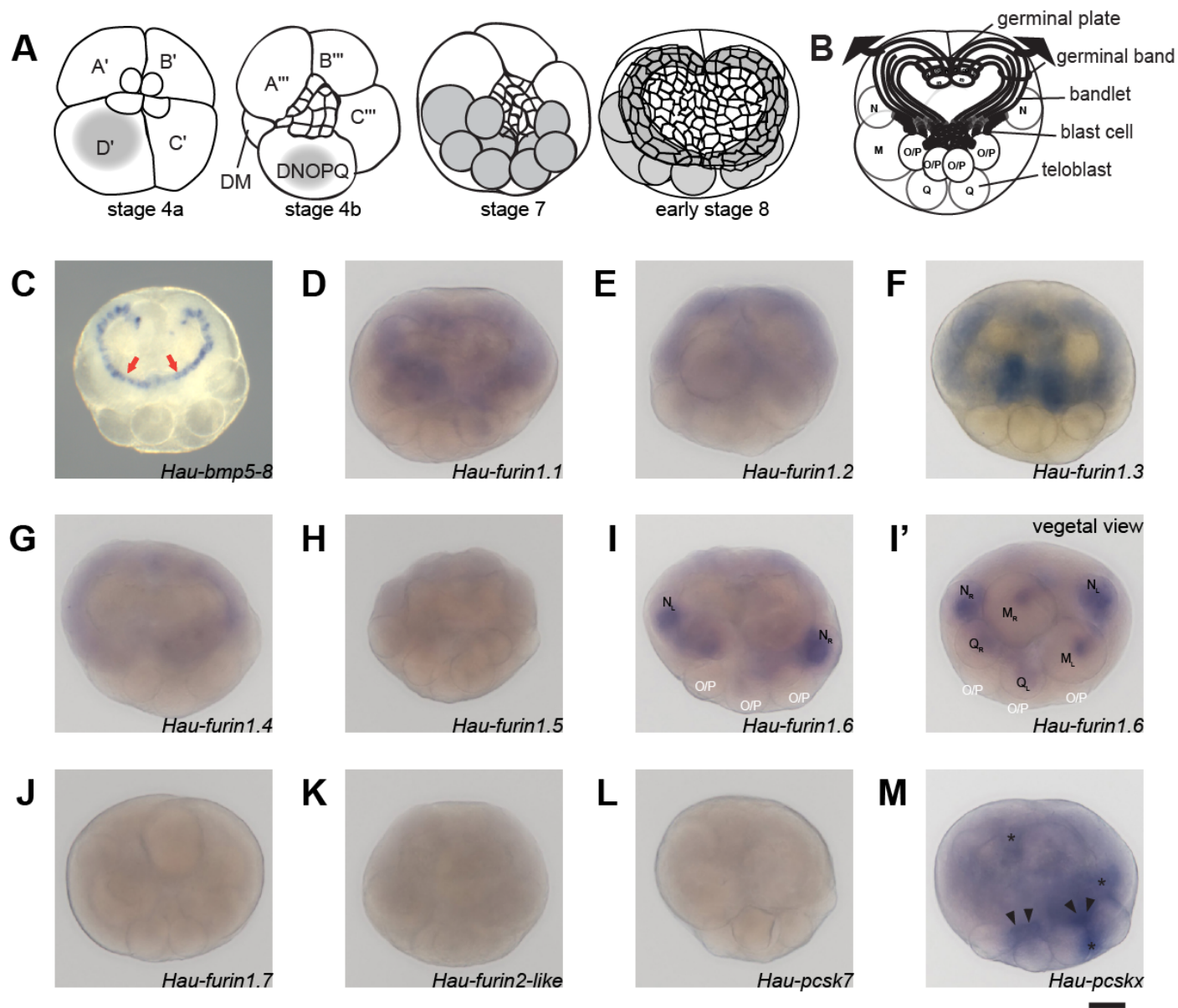


**Figure 4.** Phylogenetic analysis of the 65 Kexin-like protein from the 11 metazoan species that have less divergent Kexin-like gene complements and the three opisthokont outgroups revealed the evolutionary history of metazoan Kexin-like genes. **A.** The best maximum-likelihood tree produced by using the LG model. **B.** The best maximum-likelihood tree produced by using the CAT-GTR model. 500X bootstraps were performed for each tree. Panels A' and B' show cladograms of the trees in panels A and B respectively, but with nodes having less than .50 bootstrap support collapsed and shown as polytomies. **C.** The pattern of gene loss in the metazoan Kexin-like gene family. A closed black circle indicates the presence of identified ortholog(s) in a genome. An open circle denotes the absence of specific ortholog in a genome. A grey closed circle denotes the potential existence of derived homolog(s) in a genome (see text for details). Note that spiralian species (yellow) contains all six OGs. **D.** Three alternative phylogenetic scenarios which all entail equal numbers of gene losses. Ecd: gene loss in the common ancestor of ecdysozoans; Ver: gene loss in the common ancestor of Olfactores (Vertebrata + Urochordata); Nve: gene loss in *Nematosella vectensis*; Cel: gene loss in *Caenorhabditis elegans*; Dme: gene loss in *Drosophila melanogaster*; Spu: gene loss in *Strongylocentrotus purpuratus*; Cin: gene loss in *Ciona intestinalis*. The green shade marked with "h" indicates that the gene duplication event giving rise to PCSK7 and the ancestor of the remaining five OGs was premetazoan, probably in a holozoan ancestor. The orange shade marked with "p" indicates gene duplication events in the planulozoan ancestor, from which the modern-day PCSKX, PCSK2, and PCSK1 orthologous group arose. The yellow shade marked with "b" indicates the gene duplication event that gave rise to furin1 and furin2+PCSK5/6 OGs in the bilaterian ancestor.



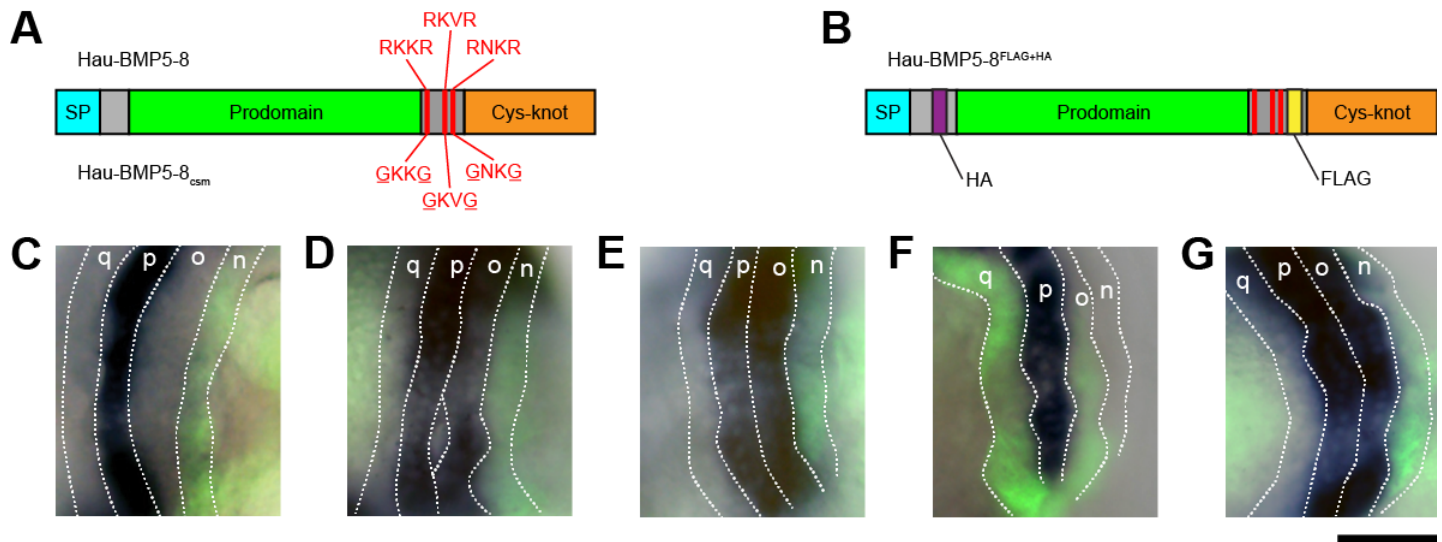
**Figure 5.**

**Figure 5.** Expression patterns of the Kexin-like genes in juvenile *Helobdella austinensis* revealed by whole-mount *in situ* hybridization. *Hau-pcsk1* (A and B) is prominently expressed in paired cell clusters in the supraoesophageal ganglion (black arrows). However, lower levels of *Hau-pcsk1* transcript were also detected in cells of the ventral segmental ganglia (red arrows in B). *Hau-pcsk2* (C-E) is broadly expressed in the central nervous system, consisting of the supraoesophageal ganglion (black arrows in C and D) and the 32 ventral segmental ganglia (bracket in D). In addition to the neural tissue, *Hau-pcsk2* was also expressed in cells of the paired salivary glands (red arrows in C and D). High-power view (E) shows that the large cells expressing *Hau-pcsk2* are adjacent to the wall of the digestive tract (black dashed lines). The retracted proboscis is marked with blue dashed lines. *Hau-furin1.1* (F), *Hau-furin1.2* (G), *Hau-furin1.3* (H), *Hau-furin1.4* (I), *Hau-furin1.5* (J), *Hau-furin1.6* (K), *Hau-furin2-like* (M), *Hau-pcsk7* (N) and *Hau-pcskx* (O) were broadly expressed in juvenile leech, although sites of elevated gene expression differ somewhat among these genes. In contrast, *Hau-furin1.7* (L) was specifically expressed in the rear sucker (black arrow). Panels A-B and E-O are the dorsal view; panel C shows the ventral view; panel D shows the lateral view. Scale bar = 150  $\mu$ m in panels A, C, D, and F-O; 75  $\mu$ m in panel B; 30  $\mu$ m in E.



**Figure 6.**

**Figure 6.** Expression of the Kexin-like genes in stage 8 *Helobdella austinensis* embryos revealed by whole-mount *in situ* hybridization. **A.** Developmental origin of teloblasts. Grey shade marks teloplasm, a special pool of cytoplasm that contains the teloblast determinants and is segregated into teloblasts during cleavage. **B.** Anatomical organization of the stage-8 leech embryo (see text for details). **C.** *Hau-bmp5-8* is specifically expressed in the dorsal, q bandlet. The image was taken from Kuo and Weisblat, 2011. **D-M.** Expression patterns of Kexin-like genes in stage-8 leech embryos. See the text for a detailed description. The asterisks in panel M mark the cytoplasm associated with macromere nuclei; the black arrowheads mark the isolated o/p bandlets entering the germinal bands. All panels except I' are in the animal-pole view; panel I' is the vegetal-pole view. Scale bar = 100  $\mu$ m.



**Figure 7.** Cleavage by convertase is not required for the P fate-inducing activity of Hau-BMP5-8. **A.** Schematic representation of the wild-type Hau-BMP5-8 and Hau-BMP5-8<sub>csm</sub>, in which all three putative convertase sites were mutated to make it resistant to convertase-mediated proteolytic cleavage. **B.** Schematic representation of HauBMP5-8<sup>FLAG+HA</sup>, in which a FLAG epitope tag was inserted between the TGFβ cysteine knot domain and the convertase cleavage site and an HA epitope tag was inserted between the signal peptide and the prodomain. In the mature, convertase-processed form, the HA tag should label the prodomain and the FLAG tag the active signaling ligand. **C.** Normal embryo in which the n bandlet was labeled with lineage tracer; the P fate marker *Hau-six1/2a* was only expressed in the p bandlet, as revealed by *in situ* hybridization. **D.** Embryo whose n bandlet mis-expressing the wild-type Hau-BMP5-8, *Hau-six1/2a* was ectopically expressed in the o bandlet, as well as the p bandlet. **E.** Similarly, *Hau-six1/2a* was also ectopically expressed in the o bandlet when the convertase-resistant Hau-BMP5-8<sub>csm</sub> was mis-expressed in the n bandlet. **F.** When Hau-furin1.1 was mis-expressed in the o, p and q bandlets, the expression pattern of *Hau-six1/2a* remained normal, restricted to the p bandlet. No expansion of *Hau-six1/2a* expression to the o bandlet was observed. **G.** Hau-BMP5-8<sup>FLAG+HA</sup> mis-expressed in the n bandlet induced ectopic *Hau-six1/2a* expression in the o bandlet as did wild-type Hau-BMP5-8 (D). Scale bar: 50 μm.