Globins and hemogenic cells in the marine Annelid *Platynereis dumerilii* shed new light on blood evolution in Bilaterians.

Solène Song [1, 2], Viktor Starunov [3], Xavier Bailly [4], Christine Ruta [5], Pierre Kerner [1], Annemiek J.M. Cornelissen [2] and Guillaume Balavoine [1]*

1. Institut Jacques Monod, CNRS, UMR 7592, Université Paris Diderot, Sorbonne Paris Cité, Paris, France
2. Laboratoire Matière et Systèmes Complexes, CNRS UMR 7057, Université Paris Diderot, Sorbonne Paris Cité, Paris, France
3. Laboratory of Evolutionary Morphology, Zoological Institute, Russian Academy of Sciences, Universitetskaja nab. 1, 199034 Saint Petersburg, Russia
4. Université Pierre et Marie Curie - CNRS, FR2424, Centre de Ressources Biologiques Marines, Station Biologique de Roscoff, Roscoff, France
5. Laboratório de Invertebrados, Núcleo em Ecologia e Desenvolvimento Sócio-Ambiental de Macaé. Campus Macaé, Universidade Federal do Rio de Janeiro, Macaé, RJ, Brazil.

* Corresponding author

Abstract

How vascular systems and their respiratory pigments evolved is still debated. To unravel blood evolution in Bilaterians, we studied the marine Annelid *Platynereis dumerilii*. *Platynereis* exhibits a closed vascular system filled with extracellular hemoglobin. An exhaustive screen in *Platynereis* genome reveals a family of 17 globins. Seven extracellular globins are produced by specialized hemogenic cells lining the vessels of the segmental appendages of the worm, serving as gills. Extracellular globins are absent in juveniles then accumulate considerably as the worm size and activity increase, culminating in swarming adults. Phylogenetic analyses with deep screenings in complete genomes establish that five globin genes (stem globins) were present in the last common ancestor of Bilaterians. All known Bilaterian blood globins are derived convergently from a single presumably ubiquitously expressed, intracellular stem globin gene. All five globin types are retained in *Platynereis,* reinforcing this species status as a key slow evolving genome within Bilaterians.

Introduction

Most animals develop at least one type of circulatory system to circumvent the impairment of gas, nutrient and waste exchanges when body size and tissue thickness increase. Vascular systems are diverse, representing different solutions to the same purpose in animals with varied body plans. Sponges have an aquiferous system of ciliated channels that transport the external medium throughout the body. Some cnidarians, ctenophores and flatworms possess gastrovascular systems that are ramifications of the digestive cavity. In many Bilaterian animals, two internal liquid compartments can take up circulatory functions: the coelomic cavity and the blood vascular system. Both systems are empty spaces forming between epithelial layers and they are often anatomically closely associated. The coelomic fluid is a liquid cavity entirely delimitated by the apical side of a mesodermal epithelium. The coelomic fluid usually contains motile cells that are called coelomocytes or hemocytes, that are mostly cells specialized for the immune response and debris cleaning (Grigorian and Hartenstein, 2013). By contrast, the blood vascular system forms at the basal side of epithelia. Blood fills spaces appearing between the lamina of two different epithelia abutting on their basal sides. The two epithelia are mesodermal in origin for the main vessels but in some animals, one can be endodermal (gut sinus) or ectodermal (skin capillary-like vessels).

In most Bilaterians, the lumen of vessels is delimitated only by extracellular matrix. In Vertebrates however, a continuous internal cell layer, the endothelium, completely lines the vessels. The blood may contain hemocytes or not. Similar to coelomocytes, these cells are mostly specialized in immune defences but in Vertebrates, a unique population of red cells has evolved that carries large quantities of oxyphoric hemoglobins. In Arthropods and many mollusks, the coelomic compartment is secondarily reduced and the blood vascular system is open in the space directly surrounding tissues, thus creating a “hemocoel” filled with hemolymph.

The evolution of the blood vascular system (as well as of the coelom) is still debated. The general ideas on the evolution of fluid-filled body cavities are linked with opinions on the general body plan evolution in metazoans. Simplifying the debate, we can classify two opposed extreme hypotheses:

- On one side, many authors believe that the last common ancestor of Bilaterians (a species often called *Urbilateria*) was a simply organized animal with no body cavity and limited diversification of organ systems (Erwin and Davidson). For some authors (Hejnol and Martindale, 2008), the modern acoelomorph worms (a divergent group of minute flatworm-like animals) may be close to the level of organization of the Urbilaterian ancestor. In both cases, the small size or level of activity of this hypothetical ancestor did not require a

circulatory system of any kind (coelomic or vascular). Gases can be exchanged with the environment through thin tissue layers. Nutrients, likewise, can be distributed on small distances by intercellular contacts. In this type of scenario, circulatory systems have evolved multiple times in different Bilaterian lineages, in correlation with the evolution of larger sizes and more active life styles.

- On the other side, some authors hypothesize a more complex type of Bilaterian ancestor (Balavoine and Adoutte 2003; De Robertis 2008; Couso 2009; Denes et al. 2007; Lauri et al. 2014). This type of hypothesis is sometimes referred to as the "complex *Urbilateria*" idea. This proposed ancestor has a body size above the centimetre threshold. It has a pronounced anterior-posterior differentiation, a condensed nervous system with a brain and one or several ventral nerve cords, a through gut, a segmented trunk, a coelom and a vascular circulatory system.

The "simple *Urbilateria*" hypotheses is strongly based on phylogenetic arguments, i.e. the general topology of the metazoan phylogenetic tree and the variety of the body organisations. The "complex *Urbilateria*" hypothesis takes more into account the stunning similarities that have been discovered in developmental patterning genes between distant phyla in the last decades, such as the Hox genes or the nervous system patterning genes (Arendt et al., 2008; Gaunt, 2018).

What are the arguments that can be brought forward to discuss the origin(s) of blood vascular systems (BVS)?

Metazoan phylogeny (Figure 1) suggests a complex picture:

- Vascular systems of any kind are found only in Bilaterians.
- A closed BVS is found in a small minority of phyla, including Chordates (but not Urochordates in which the system is open) and Annelids.
- A large proportion of phyla have a so-called open BVS. This means that while proper vessels delineated by epithelia may be present, they are open to their extremities into the general cavity of the animal or the fluid that directly baths the tissues. This is the condition found with the hemocoel of Arthropods or the Mollusks, but also in a number of smaller phyla
- Some phyla do not have a BVS at all. In Nematodes, the body cavity directly bathes tissues and fluid movements are produced by body wall muscular contractions. In flatworms (Platyhelminths), there is no body cavity and nutrients are distributed through the tissues by a diverticulate gastro-vascular system.

It is thus entirely conceivable from this picture that *Urbilateria* did not have a BVS of any kind and that the various forms of BVS evolved independently, maybe as a requirement for body size increase. Yet developmental gene studies have uncovered a remarkably similar role of conserved transcription factors (the Nkx2.5 family of genes) in the formation of a

cardiac pump-like organ in various Bilaterian taxa belonging in all three Bilaterian superphyla (notably Vertebrates, Arthropods and Annelids) (Hartenstein and Mandal, 2006; Saudemont et al., 2008).

One major function of the BVS is performing gas exchanges, bringing dioxygen to the tissues and taking back waste products (eg. CO2). To perform this respiratory function efficiently, many species use respiratory pigments that bind dissolved gases in a cooperative way and considerably increase their solubility in the blood or hemolymph. A general picture of the nature of circulating respiratory pigments used in Bilaterians (Figure 1) suggests again a very complex evolutionary history:

- Some species have no known respiratory pigments despite having a circulatory system or at least a fluid filled cavity (Nematodes, Echinoderms, Urochordates, Cephalochordates). It is assumed that in these species, either gases diffuse freely in thin layers of tissues or freely dissolved in the hemolymph.
- Some groups have circulating hemoglobins (“red blood”). The status of these hemoglobins is very diverse. They can be extracellular hemoglobins dissolved in the blood, as in many Annelids. They can be contained in red cells as in the Vertebrates or in some Annelids such as Capitellids (Mangum et al., 1992)
- Some groups have circulating dissolved hemerythrins (“pink blood”), like Priapulids, Brachiopods and some Annelids (Sipunculidae, Magelonidae) (Mangum et Kondon 1975; Klippenstein 2015; Zhang et Kurtz 1991)
- Mollusks and many arthropods have circulating dissolved hemocyanins (“blue blood”).

Circulating respiratory pigments of different types are generally not present together in the blood, suggesting that the recruitment of each type of pigment for the respiratory function occurred multiple times independently in the evolution of Bilaterians (Martín-Durán et al., 2013).

The situation with globins is particularly complex and the evolution of globins is still not understood in depth. Globins are ancient proteins present in all groups of living organisms (Vinogradov et al., 2007). They are characterized by a unique molecular structure, the “globin fold” made of eight alpha helices which shelters a heme prosthetic group, itself responsible for binding diatomic gases, such as O_2 or NO. The oxygen binding function has been particularly well described in specific groups (Vertebrates, Annelids) but the globin superfamily is prevalent across the animal kingdom and their possible functions beyond that scope are diverse and still poorly known (Vinogradov and Moens, 2008). The globin motif is sometimes found in composite proteins that have evolved by fusion of pre-existing protein domains but all respiratory “hemoglobins” derive from proteins with stand-alone globin motif. Many phylogenetic studies have focused on the remarkable hemoglobins of Mammals and

Vertebrates in generals, starting with the seminal work of Zuckerkandl and Pauling (Zuckerkandl and Pauling, 1962). A number of studies have focused on the evolution of Metazoan globins (Blank and Burmester, 2012; Dröge et al., 2012; Hoffmann et al., 2012; Lechauve et al., 2013). These studies have revealed that:

- Vertebrate hemoglobins, the related myoglobins and cytoglobins are only a small branch in a vast tree of animal globins, some of which are used for respiratory functions but many others have still unknown functions.
- Some studies have highlighted the ancientness of neuroglobins, the first globins discovered in the Vertebrate neurons (Burmester et al., 2000). These globins have homologues in groups as distantly related as Cnidarians and Acoelomorphs. Crucially, expressions in these two groups also occur in neural or sensory cells. This led to the suggestion that neuroglobins would be the most ancient type of globins occuring in animals and that hemoglobins would be derived from this ancestral neuroglobins by gene duplication and functional divergence (Lechauve et al., 2013).
- Another study (Roesner et al., 2005) has identified globin X, a molecule present in Teleost fishes and Amphibians but not in Amniotes, expressed in neural cells as well in these Vertebrates (Blank et al., 2011), as an ancient globin type. Another Panmetazoan clade of globins but missing in all Vertebrates, called X-like, is also identified (Blank and Burmester, 2012).
- The same study (Blank and Burmester, 2012) showed that many of these Globins X or X-like carry a prototypical N-terminal acylation site (myristoylation or palmitoylation) suggesting that these proteins are linked to membranes in the cell. The authors propose that Vertebrate hemoglobins could derive from molecules that were initially membrane bound.

These phylogenetic studies however have relatively incomplete samplings of species and molecules. Gene sequences are identified from animals whose complete genome was not available at the time of the study and therefore represent only a subset of the diversity that must exist in Metazoans. A larger sampling of Metazoan genomes is now available and there is a need to update these studies by using exhaustive screens for all globin-like sequences in fully sequenced genomes.

Many Annelids are remarkable for the similarities of their BVS with the Vertebrate BVS:

- The Annelid BVS is usually closed. It is one of two groups (with Nemertines) in the Protostomes where a closed circulatory system exists. *Platynereis* blood vessels are devoid of endothelium, similar to the vast majority of BVS in protostomes. They are located in between spacious pairs of coelomic cavities in each segment and have a strictly metameric organisation along the trunk of the animal.
- The blood of many Annelids is red containing a high concentration of respiratory hemoglobin. In contrast to the Vertebrates, this circulating hemoglobin is extracellular

instead of enclosed in red blood cells. Its structure is very different from the hemoglobin of Vertebrates. Instead of a heterotetramer of globin sub-units, many Annelids possess a giant hemoglobin molecule, originally referred to as “erythrocruorin”, organised in a hexagonal bilayer of no less than 144 individual globin peptides (Royer et al., 2000; Royer et al., 2006; Royer et al., 2007), assembled with the help of linker proteins.

Before this study, there was little known on the nature of the cells that produce and secrete respiratory pigments outside of Vertebrates. In Annelids, earlier studies in a limited number of species suggest that extracellular circulating hemoglobins are secreted by the cardiac body, a convoluted organ present within the anterior pulsatile vessel of some but not all Annelids (Kennedy and Dales, 1958). In some other cases (Jhiang et al., 1986), mesodermal cells bordering the gut blood sinus (also called “chloragogen” cells) are responsible for blood production. These studies reveal an unsuspected variability in the localisation of these cells in Annelids.

In this work, we explored the globin content of the genome of the marine Annelid *Platynereis dumerilii*. We determined the nature and locations of the cells responsible for producing the hemoglobin. In the following, we will refer to these cells as “hemogenic cells”. *Platynereis dumerilii* is a medium-sized Annelid that can be easily cultured in the laboratory, giving a large number of offspring all year round. The life cycle of *Platynereis* (Figure 2), fairly typical of marine Annelids includes a microscopic (160 μm diameter) lecithotrophic trochophore larva that elongates after 2,5 days in a minute (400 μm) three-segment worm. Once settled in the benthos in a silk tube, the larva will grow by posterior addition of segments and in cross-section to reach a considerably larger size (5-6 cm). The worm lives a relatively sedentary life, feeding on a variety of fresh or decaying food in the benthos for the longest part of its lifespan. Nearing the end of its life, the worm undergoes a rather dramatic sexual metamorphosis: the coelom, from head to tail, entirely fills up with gametes; the worm segmental appendages, called parapodia, change shape and acquire a swimming locomotory function; the gut degenerates as adults do not eat. Both males and females exit their tubes to swarm at the surface of the sea. At this stage they acquire a very fast swimming behaviour and have only a few hours for mating before they die of exhaustion. Despite its status of emergent model animal for evolution and development studies (Fischer and Dorresteijn, 2004; Fischer et al., 2010; Zantke et al., 2014), the composition of its red blood has not been yet characterized, although it can be inferred that they possess the Annelid extracellular giant hemoglobin (“erythrocruorin”) mentioned earlier.

In the present work, we identified the multigenic family encoding for the globins in *Platynereis* publicly available sequenced genomes and transcriptomes. In particular, we found all extracellular globin genes, presumably coding for the subunits that take part in the

making of the worm giant extracellular respiratory pigment. Based on these sequences, we combined *in situ* hybridization and electron microscopy 3D reconstructions to identify the cells that produce these extracellular globins. We show that the morphology of these cells is indeed consistent with a hemogenic activity. We described the evolution of this hemogenic activity along the life cycle of the animal in terms of location and intensity. The hemogenic cells are first found in lateral vessels appearing in between the segments. Then, in growing worms, they appear around the vessels inside parapodia, the gill-like appendages on which the worm increasingly relies for respiration. The extracellular globin genes expression intensities, measured using quantitative PCR, culminate at the onset of maturation and then collapses at the end of maturation process as the worm is approaching its reproduction and end of its life.

We analyse the evolution of globin genes across metazoans using completely sequenced genomes. From this phylogenetic analysis arises a new scenario for the evolution of globins suggesting a set of five globins in the Bilaterian ancestor, which we propose to call “stem-globins” and consequently a new classification based on gene evolution.

Materials and Methods

Animal culture and collection

Platynereis embryos, juveniles and adults were bred in the Institut Jacques Monod according to the protocols compiled by A. Dorresteijn (www.platynereis.de).

Survey of *Platynereis dumerilii* globin genes

Platynereis dumerilii globin genes were identified with consecutive steps of sequence similarity searches. We first used human globins as queries against expressed sequence tags (ESTs) from *Platynereis* Resources (4dx.embl.de/platy/), and Jékely lab transcriptome (http://jekely-lab.tuebingen.mpg.de/blast/#). Complete coding sequences were assembled from EST fragments using CodonCode Aligner (CodonCode Corporation, USA). The sequences were reciprocally blasted against Genbank non redundant protein database. This allowed constituting a first list of sequences which were recognized as *bona fide* globins. In order to get an exhaustive repertoire of *Platynereis* globin genes, including highly divergent sequences and sequences that might be missing from the transcriptomes, we performed a tblastn search in the *Platynereis* genomic database (4dx.embl.de/platy/) using as query a concatenation of three highly divergent globin peptide sequences in the *Platynereis* transcriptome and a human alpha chain hemoglobin. Exons recovered with this method were systematically reblasted on the transcriptome sequences. For each *Platynereis* globin,

intron positions were mapped on genomic DNA. Identification of extracellular globin genes was done using Phobius signal peptide predictor (http://phobius.sbc.su.se/, Stockholm Bioinformatics Center).

Survey of globin genes in animal genomes

Gene searches were carried out using the tblastn or blastp algorithms implemented in ngKlast (Korilog V 4.0, Questembert, France). We used a concatenation of the *Platynereis* globins as a query. This complex query sequence covers as much as the molecular diversity of globins as possible and even very derived sequences were recovered in the hit lists. Also, a concatemer query gives a clear drop in the list of hits with decreasing E-values, when sequences with no significant similarity with globins are reached, which helps setting a sensible E-value cut-off. We used peptide predictions from 22 genome datasets widely distributed among the metazoans phyla: Porifera, Cnidaria, Placozoa, Ecdysozoa, Lophotrochozoa and Deuterostomia. In addition, the globin genes search was performed on the transcriptomes of *Lumbricus terrestris, Arenicola marina* and *Alvinella pompejana*. Predicted peptides from the lists of BLAST hits were reciprocally BLASTed against the human proteins.

The putative extracellular globin genes were identified using Phobius signal peptide predictor (http://phobius.sbc.su.se/, Stockholm Bioinformatics Center).

The putative N-terminal myristoylation predictions were carried out using NMT predictor (http://mendel.imp.ac.at/myristate/SUPLpredictor.htm) (Maurer-Stroh et al, 2002). We retained only those classified as “reliable zone”.

Phylogenetic analyses

The amino-acid sequences of the identified globin genes in 25 species were aligned with MUSCLE 3.7 (Edgar, 2004) as implemented on the LIRMM web (http://phylogeny.lirmm.fr/phylo_cgi/one_task.cgi?task_type=muscle) under default parameters and adjusted manually in Bioedit. A selection of aligned positions was produced to eliminate unaligned or ambiguously aligned regions (suppl. file 1). This resulted in a alignment of 275 globin sequences displaying 231 phylogeny informative positions. The resulting file contains The phylogenetic trees were constructed using two different approaches: the maximum likelihood (ML) and the Bayesian analyses. Maximum likelihood trees were generated using PhyML3.0 (http://phylogeny.lirmm.fr/phylo_cgi/one_task.cgi?task_type=phyml) (Guindon and Gascuel, 2003), using the LG model of amino-acid substitutions (Le and Gascuel, 2008). This model has been shown to perform better than the more widely used WAG model. The classical Bootstrap test for short sequences such as globins is inappropriate. Therefore statistical

support for nodes was assessed using SH-like test (aLRT score) (Anisimova and Gascuel, 2006). Bayesian analysis was performed with MrBayes 3.2.6 (Ronquist and Huelsenbeck, 2003) using either LG or WAG fixed model, run for respectively 32 174 500 generations and 24 972 000 generations , using all compatible consensus and a burn'in value of 0.25. The calculation was performed with 4 chains including one heated chain at temperature 0.5. The average standard deviations of split frequencies were respectively 0.031 and 0.017.
Bayesian posterior probabilities were used for assessing the confidence value of each node. Phylogenetic trees were visualized and rooted using FigTree V.1.5.0 (http://tree.bio.ed.ac.uk/software/figtree/).

Cloning of extracellular globin genes and in situ probes design

Almost complete cDNA fragments were cloned by nested PCR using sequence-specific primers on cDNA from mixed larval stages and posterior regenerating segments. PCR products were cloned into the PCR2.1 vector following the manufacturer's instructions (Invitrogen, France) and sequenced. These plasmids were used as template to produce Dig RNA antisense probes for whole-mount *in situ* hybridization (WMISH) using Roche reagents.

Visualization of *Platynereis* extracellular globin genes expression patterns by whole mount *in situ* hybridization

The animals were fixed in 4% paraformaldehyde (PFA), 1 x PBS, 0.1% Tween20 and stored at -20°C in methanol 100%.
NBT/BCIP whole-mount in situ hybridization was performed as previously described (https://www.ijm.fr/fileadmin/www.ijm.fr/MEDIA/equipes/Balavoine/IJM_HybridationInSituNBT_BCIP.doc) on larval stages (24h, 48h, 72hpf), juvenile larvae, different stage of maturing worms and mature epitoke worms, as well as on regenerated posterior "tails", 9 days after posterior amputation. Bright-field images were taken on a Leica microscope.

Quantitative PCR

Total RNA was extracted using RNAeasy kit (Qiagen). The extraction was done from a pool of 48hpf larvae, a pool of 6 weeks worms, whole 50 segments juvenile worms, a stretch of 20 segments starting 24 segments after the head for maturing stages. The tissue was disrupted and homogenized inside RLT buffer. First-strand cDNAs were synthesized using 100ng of total RNA, random primers and the superscript II Reverse Transcriptase (Invitrogen, Life Technology). Specific primer pairs were designed with Applied Biosystems Primer Express software to amplify specific fragments between 50 and 60bp. The composition of the PCR mix was: 4uL of diluted cDNA, 5uL of SYBR GREEN master mix, 1 uL of primers fwd, 1uL of primers reverse, in a 10uL final volume. qPCR reactions were run in 96-well

plates, in real-time Applied Biosystem StepOne thermocycler. The PCR FAST thermal cycling program begins with polymerase activation and DNA denaturation at 95 °C for 20 sec, followed by 40 cycles with denaturation for 3s at 95°C and annealing/extension for 30s at 60°C. After amplification, melting curve analyses were performed between 95°C and 60°C with increase steps of 0.3 °C to determine amplification product specificity. The slopes of he standard curves was calculated and the amplification efficiencies (E) were estimated as E=10^(-1/slope). qPCR were run with biological triplicates for juvenile worms of 50 segments and mature worms, and each sample was run with technical duplicates and the ribosomal protein small subunit 9 (rps9) and *D-adenosylmethionine synthetase* (sams) housekeeping gene. Relative expression level of each target gene was obtained by 2^(-deltaCtsample) where deltaCtsample=Ct sample-Ct gene of reference.

Transmission electron microscopy

For electron microscopy animals were relaxed in 7,5% MgCl2, fixed in 2,5% glutaraldehyde buffered in 0,1M phosphate buffer and 0,3M NaCl, rinsed 3 times in the same buffer, and postfixed in 1%OsO4 in the same buffer for 1h. The specimens were dehydrated in ascending acetone series, transferred in propylene oxide and embedded in Epon-Araldite resin (EMS). Ultrathin sections (60-80nm) were cut with Reichert Ultracut E or Leica Ultracut UCT and counterstained with 2% uranyl acetate and Reynolds lead citrate. Images were acquired using Zeiss Libra 120, FEI Technai or Jeol JEM-1400 transmission Electron microscopes and processed with Fiji and Adobe Photoshop software. For 3D-reconstructions the series of semithin (700 nm) sections were cut with Diatome HistoJumbo diamond knife (Blumer et al., 2002), stained with methylene blue/basic fuchsine (D'amico, 2009), and digitalized at 40x magnification using Leica DM2500 microscope with camera. The images were aligned with IMOD and ImodAlign tool. The reconstructions were made with Fiji TrackEM2 plugin.

Results

Hemogenic cells line specific lateral vessels and then parapodial vessels in the growing *Platynereis* juveniles.

Analysis of the developing blood vessels by transmission electron microscopy

The *Platynereis* BVS is built progressively during development. The acoelomate larvae have no circulatory system and rely on yolk stored in their four big macromeres for nutrients (lecithotrophy). The first signs of a BVS appear when a feeding juvenile with three segments has settled on the substrate around 10 days after fertilization, in the form of a pulsatile dorsal

vessel (Suppl. Figure 1). Then the juvenile starts to add segments at the posterior tip of the body and this sequential addition last during most of the benthic life of the animal. The BVS undergoes two types of addition during this juvenile growth:

- New vessels appear in the growing trunk, especially at the level of the segmental appendages, which serve both as legs and branchiae (gills).
- New metameric BVS unit are put in place in each new segment added. Each metameric unit communicate with the contiguous segments by the pulsatile dorsal vessel, the quiescent ventral vessels and derivations of the lateral vessels.

The blood becomes visibly red only when the worm reaches a certain size, around 10-15 segments. This suggests that blood hemoglobin producing cells become active progressively. To localize these cells and analyse their cytological properties, we observed ultra-thin sections of worms by electron transmission microscopy TM at two juvenile stages where blood is visibly red (Figure 3). We reconstructed the 3D pattern of blood vessels and hemogenic cells also by using large series of semi-thin sections (Figure 3). One stage corresponds to juvenile worms with around 25-30 segments (Figure 3A-B). In these worms, the BVS metameric unit is still fairly simple with dorsal and ventral vessels and a couple of lateral vessels (Figure 3A-B, red). The lateral appendage vasculature is still little developed. At this stage, sheaths of very characteristic putative hemogenic cells (hc, in blue) are already present around a pair of lateral vessels (tbv, Figure 3A-B). The other stage corresponds to pre-mature worms with around fifty segments. These worms have much more developed lateral appendage network with lateral vessels colonizing the gill-like parapodia (Figure 3C-D, red). In the mid body segments of these worms, the putative hemogenic cells sheaths are visible around the lateral vessels (tbv) inside the trunk but also new hemogenic cells (hc) appear around parapodial vessels, more densely on the dorsal side.

Putative hemogenic cells semi-thin and TM images are given in Figure 4. On semi-thin sections, they appear as thick sheaths of cells surrounding the lateral vessel lumen (Figure 4A-B). They show all the characteristics of an intense protein synthesis and secretion activity (Figure 4C-G). The cytoplasm is mainly filled with a dense rough endoplasmic reticulum (rer, Figure 4D-E). The hemogenic cells are organized in a simple epithelium with basal sides facing the vessel lumen and apical sides facing the coelomic cavity (Figure 4C). A clear basal lamina delineates the vessel lumen on the basal side (Figure 4E). Electron-dense bodies are visible in the cytoplasm of hemogenic cells and are interpreted as ferritin granules (Figure 4F). Last, secretion vesicles are clearly identified on the basal side (Figure 4G), indicating intense exocytosis process. The vesicles contain a uniformly granulated matrix identical to the aspect of the blood. These small granules are thought to represent the particles of the giant hexagonal bilayer of the erythrocruorin. These hemogenic cells are

remarkably similar to the previous descriptions of hemoglobin secreting cells in the heart-bodies (an organ inside the dorsal vessel) of several sedentarian Annelids (Dales and Pell, 1970).

We analysed also semi-thin sections in very small juvenile worms. Several examples are shown in suppl. Figure 2 with worms with 4, 5 and 6 segments. These worms have already a distinct BVS with dorsal and ventral vessel but no visible lateral ramifications (red). The first hemogenic cells are visible in the posterior most segments as very small groups of cells (blue). In the forlast posterior segment of the worm with 6 segments putative degenerate hemogeni cells are also visible (green).

Identification of globin genes in the Platynereis genome and expression in the hemogenic cells

Do the secretory cells identified around lateral vessels actually produce blood globins? We identified in the genome of *Platynereis* no less than 17 globin genes, including 7 coding for extracellular globins, as predicted by the presence of a secretion signal peptide at their N-terminus (Figure 5). We conducted systematic expression analysis by *in situ* hybridization *in toto* (WMISH) on a series of *Platynereis* developmental stages. None of the globin probes gave a significant signal on early larval stages (24 hpf trochophore, 48 hpf metatrochophore, 72 hpf nectochaete). Expression patterns for growing juveniles are shown in Figure 6. Only the 7 extracellular globins have shown specific patterns. The 10 intracellular globins gave no signal or non-specific signal in silk glands after long staining periods (not shown). All 7 extracellular globin genes display the same type of expression patterns, but not necessarily at the same time. In particular, we saw no expression of Egb-A1c in juvenile worms. The expression for this gene starts only in worms that are approaching sexual maturity and are about to enter epitoque metamorphosis. For all the most precocious extracellular globins, the expression starts when the worm is about 5 segments (not shown), in the posterior most segment. In older juveniles (Figure 6A), the expression is located along lateral vessels in the trunk. The expression shows a graded pattern along the anterior-posterior axis. Expression is stronger in the mid-body segments and absent in the rostral most and caudal most segments. This graded expression continues as the worms continue to develop. In bigger worms, the expression starts massively in the parapodia of the mid-body along vessels that irrigate the gill part again in perfect correspondence with the pattern of secretory cells revealed by TEM (Figure 6B-C). The central most segments display expression in the parapodia when the more rostrally still display expression around lateral vessels in the trunk. This seems to show that there is a progressive migration of the hemogenic activity from the trunk to the parapodia and a more intense expression in the mid-body segments (suppl. Figure 2). The fading expression inside the trunk is in accordance with some of the TEM

pictures we obtained in older pre-mature worms. These pictures show hemogenic cells with clear signs of degeneration (suppl. Figure 1 & 3). In worms engaged in sexual metamorphosis, we find a decreasing intensity of expression of all extracellular globins in parapodia (suppl. Figure 4). The expression shows a dashed lining of the vessels as if expression is fading. In fully mature swarming worms, hemoglobin content is peaking but there is no mRNA expression of any blood globin left (not shown).

Quantitative PCR suggests globin switching during the life cycle

The stages for qPCR were divided as follows: the larvae 48hpf, 6 weeks larvae, 50 segments larvae, metamorphing stage I, metamorphing stage II, mature swimming worms. Metamorphing stage I is defined according to the following criteria: eyes start to bulge and parapodia are transforming. Metamorphing stage II: eyes have grown to maximal volume, and parapodia are fully transformed, but the animal is not displaying active swimming behaviour yet (Figure 7). The relative expression levels of the extracellular genes with respect to reference genes show important variations from one biological triplicate to the other. Different individuals of the same apparent stage can display quite different compositions of extracellular globin genes mRNA. We have no clear explanation for this important variability but globin mRNA expression is known to be highly responsive to physiological conditions, in particular the amount of O_2 available in the environment. However, when combining the expression levels of all extracellular globin genes, as an indirect measure of general blood production (Figure 7), the general trend is that blood production picks up at 6 weeks, culminates between 50 segments and the beginning of the sexual metamorphosis (75-80 segments) and collapses rapidly after, ending by the time the worm starts swimming, which will be followed by reproduction and death. The data for the individual genes, despite their variability show nevertheless trends that confirm our WMISH observations for some genes (supplementary Figure 6 and 7). The globins A2, B1 and B2 (Figure 7) are expressed at levels that follow the general tendency described above and at higher levels than the four A1 globins. The four paralogous A1 globins (Figure 6), meanwhile, are expressed at different levels. While A1a and A1d follow the general tendency, A1c is expressed significantly only in sexually maturing animals. In contradiction with at least some of our WMISH results, A1b was detected only at very low levels in all biological replicates. While A1c may be an adult-specific erythrocruorin sub-unit, A1d may be a paralog particularly responsive to physiological conditions.

Metazoan-wide phylogenetic analyses reveals that five stem globins existed in Urbilateria

Our complete genome surveys uncovered a highly variable number of globin sequences in the different Metazoan species, ranging from 3 in the Arthropods *Strigamia* and *Drosophila*, to 32 in the Annelid *Capitella teleta*. It is important to mention that in this survey, we systematically withdrew the globin domains that are part of multidomain proteins, such as the androglobin of Vertebrates. We also did not take into account the globins with multiple serially arranged globin folds as occur for example in Echinoderms (Bailly and Vinogradov, 2008; Christensen et al, 2015). One exception was made nevertheless for the Crustacean *Daphnia pulex*. In this species, 11 globins are didomain-organized. As it suggested multiple tandem duplications of an ancestral didomain globin, we chose to separate the N- and C-terminal globin domains in each of these peptides and to treat them as individual OTUs in our trees.

The complete maximum-likelihood is shown in supplementary Figure 8. Figure 8 shows a simplified tree version (Figure 8A) where gene names do not appear and branches have been color-coded according to the clade of the metazoan tree in which they occur (Figure 8B). Only the *Platynereis* gene locations are shown in this simplified tree. The tree in unrooted. The red diamonds indicate a number of nodes that are supported by aLRT values superior to 0.95 (Figure 8A). In this tree, we searched for groupings that would indicate pan-metazoan or pan-Bilaterian genes. Concretely, we looked for sub-trees that would contain genes coming from a broad selection of species. These clades are likely to derive from a single ancestral gene. We found four well supported clades (aLRT>0.95), that we named type I to IV, that contain a broad sampling of Bilaterian species but no non-Bilaterian species. These are likely descendants of single pan-Bilaterian globin genes. In addition, one species rich clade containing both Bilaterian and non-Bilaterian genes (Porifera, Placozoa, Cnidaria) was found but with only a marginal aLRT support. We named this clade type V as we consider that it contains very likely descendants of a single pan-metazoan gene. Last, we found a well-supported clade containing only Spiralian genes (Mollusca, Annelida, Brachiopoda, Platyhelminthes), which we termed Sp-globins (Sp-Gb). We conducted bayesian phylogenetic analyses of the same sampling of genes and found good support for all six clades, except type V.

These results suggest that all Bilaterian globins derive from five ancestral genes that were present in *Urbilateria*, thus defining five types of Bilaterian globin. In the metazoan ast common ancestor, the situation is much less clear but at least a type V globin must have been present. All blood respiratory globins, either intracellular (human) or extracellular (Annelids, *Daphnia*) are found in clade/type I in this tree. This means that they all derive from a single Urbilaterian ancestral gene (supplementary Figure 9).

Last, we mapped on our ML tree the globins that are thought to be membrane bound because they harbor a predicted myristoylation signal (supplementary Figure 10, blue

branches). None of the type I globins is predicted to be membrane-bound. This is in contrast with the type II, III, IV and Sp-Gb, that all contain numerous myristoylated sequences. As our methods of prediction very likely underestimate the proportion of myristoylated sequences, it is likely that more globins of these clades are actually membrane-bound. There is a strong probability that the ancestral globins of these clades were themselves linked to the membranes.

Figure 9 proposes a general interpretation of the evolution of metazoan globins based on our trees.

Discussion

Here we report the first exhaustive screening for globin genes in an Annelid and more generally in a Protostome animal. We performed a broad phylogenetic study of the metazoan globins based on complete genome sequence and therefore exhaustive sampling of globin genes in most species. This study helps in understanding the evolution of the globin superfamily in metazoans. It gives some hints in the evolution of the blood itself.

We can sum up the principal results of our study in the following points that will be discussed extensively in the following parts:

- The globin superfamily already had an extensive history in metazoans before some of them were recruited in the blood for gas exchange functions. At least five genes were already present in the Bilaterian ancestor, of which at least three were probably membrane-bound proteins. We propose a new nomenclature for these globins (type I to V) based on the highly supported clades found in all trees.
- All Bilaterian family of circulating blood globins derive from a likely single type I globin gene present in the Urbilaterian ancestor. This ancestral type I globin was intracellular, not bound to the membrane.
- Hemoglobin based blood has been “invented” independently several times: in other words, the descendants of this ancestral type I globin were recruited several times independently in different clades to ensure the circulatory function. We contradict a previous study (Belato et al., 2019) that suggested the existence of Annelid-like extracellular globins in the Urbilaterian ancestor.
- Each of these functional recruitments has been correlated with a series of gene duplications. These duplications gave in each « red blood » clade a family of closely related proteins, capable of forming hetero-multimers.
- *Platynereis* possesses representants of all ancestral Bilaterian globins types, reinforcing its status as a « genetic living fossil » and making it an ideal model for exploring further the functions of this complex globin superfamily.

- The extracellular blood globins of *Platynereis* are produced in hemogenic cells that are first part of the lateral vessel walls in smaller worms, then small vessels in the appendages (parapodia) in the growing worms. These appendages serve not only as locomotory organs but also as gills.
- Blood globin production culminates in worms that are getting close to sexual metamorphosis. At least two paralogous globins are expressed at higher levels in maturing worms suggesting a « globin switch » during development.

Our phylogeny suggests a new nomenclature for Metazoan globins

Our study of the phylogenetic relationships of metazoan globins fills an important gap in the literature. Many studies have actually been published on the phylogeny of globins starting with the first molecular tree (Zuckerkandl and Pauling, 1962). To cite only some of the recent works, there has been studies on the evolution of globins at the level of the tree of life (Vinogradov et al., 2006; Vinogradov et al., 2007), and at the level of the eukaryotic tree (Vinogradov et al., 2013). These studies emphasize the universality of the globin fold motif present since the last common ancestor of all living cells (LUCA, last universal common ancestor) and the likely functions of these proteins as enzymes or sensors. Many other studies have analysed globin evolution in specific metazoan groups such as Annelids (Bailly et al., 2007), Deuterostomes (Hoffmann et al., 2012), Echinoderms (Christensen et al., 2015), Pancrustaceans (Burmester, 2015), Cephalochordates (Ebner et al., 2010), Chordates (Storz et al., 2011), Agnathans (Fago et al., 2018; Rohlfing et al., 2016). This proliferation of studies illustrates the extraordinary diversification of globin genes in many of the metazoan groups. Other studies have targeted the origin of specific globin genes such as cytoglobins (Burmester et al., 2002a).

The previous most comprehensive studies (Blank and Burmester, 2012; Dröge et al., 2012; Lechauve et al., 2013) have exemplified the existence of large conserved sub-families of non-blood globins in Metazoans and have started to show the complexity of animal globin evolution. One of these sub-families, the neuroglobins (Ngb) are hexacoordinate intracellular globins first discovered in the mammalian nervous system and whose function remain poorly understood (Ascenzi et al., 2014; Burmester and Hankeln, 2009; Burmester et al., 2000). Neuroglobin orthologues have been discovered in a number of animal groups other than Vertebrates as well as in Choanoflagellates and has been proposed as an ancestral Metazoan globin from which all the other animal globins could be derived by gene duplication (Lechauve et al., 2013). The other proposed sub-family has been called “Globin-X” (Roesner et al., 2005). Globin-X (and Globin-X-like) proteins are widespread among Metazoans. They present very often one or two acylation sites at the N-terminus, either

myristoylation or palmitoylation, making them membrane-bound proteins (Blank and Burmester, 2012). The functions of globin-X proteins are not known, although a role in protecting the cell from reactive oxygen species (ROS) is suggested. Blank and Burmester's work (2012) suggest that the blood globins of Vertebrates may derive from ancient membrane-bound globins, via the loss of acylation sites, opening, among other things, the possibility to form multimers.

Despite bringing forward interesting conceptions, previous works on Metazoan globins all suffer from biased sampling. The sampling of globins of Blank and Burmester (2012) was made by searching NCBI Genbank with terms related to globins, while the same database was screened with a Ngb probe in Lechauve et al (2013). In this work, we have screened complete genome sequences of a representative sampling of metazoans, using a concatemer sequence probe representing the diversity of human and *Platynereis* sequences. We are thus confident that our samplings of globins in each genome are very close to exhaustivity.

In trees based on maximum-likelihood and bayesian algorithms, a number of robust nodes emerged that correspond to globins present in a large sampling of Bilaterian species, thus likely to represent ancestral Bilaterian genes. Based on these results, we propose a new nomenclature of globin-like molecule in Metazoans. In our opinion, this new classification reflects better the evolution of Metazoan globins than those proposed by previous work in which a part of the diversity of non-blood globin sequences was missing.

The five groups of globin molecules we identified are the following:

- Type I: this corresponds to the totality of circulating blood globins, either extracellular or carried by red cells, found in metazoans. In addition, this group also includes many non-circulating globins. These additional non-blood type I globins are present in animals with hemoglobin blood, with a non-hemoglobin blood or with no blood at all. The large majority of Bilaterian taxa are represented in this group but no globin of non-Bilaterian taxa (Cnidarians or Sponges) is present, indicating that this clade originated in an ancestor of Bilaterians. Globin proteins can display two different chemistries, depending on if the heme is attached to the globin through by one or two histine side chains, respectively referred to as "pentacoordinate" or "hexacoordinate" (Kakar et al., 2010). The Vertebrate hemoglobins and myoglobins, responsible for conveying and storing oxygen are pentacoordinate. However most globins of type I are hexacoordinate suggesting that the Vertebrate blood globins have made the transition from hexacoordination to pentacoordination.

- Type II: this clade contains multiple globins from all Bilaterian superphyla (Deuterostomes, Ecdysozoans, Spiralians). It corresponds to « globinX-like » (Blank and Burmester, 2012). No globin of non-Bilaterian taxa is present, indicating that this clade originated in an ancestor of Bilaterians. No type II globin is present in Vertebrates (although

it is present in Cephalochordates), indicating that a type II globin was present in the common ancestor of Chordates and lost secondarily in Vertebrates.

- Type III: this clade contains globins from all Bilaterian superphyla. It corresponds to « GlobinX » (Blank and Burmester, 2012). No globin of non-Bilaterian taxa is present, indicating that this clade originated in an ancestor of Bilaterians. Our sampling contains only one Vertebrate species, human, in which type III globins are not present. It is known from other studies that type III globins are present in non-Amniote Vertebrates, but absent from Amniotes, indicating secondary loss.
- Type IV: this clade contains globins from all Bilaterian superphyla, although it is absent from Chordates and also from all Ecdysozoans except Priapulida. This clade has apparently gone undetected in previous studies. No globin of non-Bilaterian taxa is present, indicating that this clade originated in an ancestor of Bilaterians.
- Type V: this clade is the least statistically supported of all. It contains however globins from a broad variety of Metazoans, including Sponges and Cnidarians. It is missing from Ecdysozoans. It corresponds to some of the « neuroglobins » identified in previous studies. This is the only globin type whose existence is clearly suggested in the last common ancestor of Metazoans. This is however a much narrower group than the « neuroglobins » previously described (Lechauve et al., 2013), that includes many globins from our type II and III in a broad « neuroglobin-like » class.
- In addition, a well supported clade contains only globins from Spiralian species (Sp-Gb). It may have appeared by a gene duplication of any of the other globin types.

The animal globin superfamily has an ancient history independent from the history of blood

The existence of these clades clearly indicates that at least five globin genes existed in *Urbilateria*. We call these five hypothetical molecules the “Bilaterian stem-globins”. Few globin sequences fall outside of these five well-defined clades (Figure 9). These additional globins fall into two cases. In one case, these sequences may derive from the stem globins but could have evolved rapidly so as to blur sequence similarities. This is what we would suppose for the large clade of *C. elegans* globins found alongside a single type I globin. The second case concerns the clade of sea anemone globins, which includes almost all the sea anemone globins in our dataset (except from a singe type V globin). Most of the members of this group are myristoylated, suggesting that it could be related to any of the Bilaterian stem globins that were potentially membrane-bound (type II, III and IV). Alternatively, the sea anemone membrane-bound globins may derive from an ancestral Eumetazoan membrane-bound globin that would also have given birth to Bilaterian type II, III and IV. Reconstituting

metazoan globin history outside Bilaterians is thus at this stage still tentative because of the small sampling of species.

Besides, some globin groups from previous globin phylogenies are not included in our dataset. Notably, “neuroglobin-like” sequences are identified outside metazoans, in the closely related Choanoflagellates (Lechauve et al., 2013). Further investigations are needed to verify whether these Choanoflagellate globins can be classified in our type V.

The deep relationships between Bilaterian stem globins are not resolved by our trees (that are not rooted in any case). The idea that Bilaterian blood globins (and all other type I globins as defined by our trees) may derive from initially membrane-bound molecules (Blank and Burmester, 2012) is thus not corroborated by our study. It is in our view equally possible that type I derives from an ancestral duplication of the cytoplamic type V, while type II, III and IV would have acquired membrane tethering independently.

What were the functions of these numerous non-blood globins derived from the Urbilaterian stem-globins? Here again, very few functional studies have been performed and we can only speculate. Burmester and Hankeln (2009) have reviewed the possibilities for the special case of neuroglobins (type V), but the same possibilities seem to exist for the other type of non-blood globins (type I to IV). Briefly, these globins, that are all intracellular, cytoplasmic or membrane-bound or in some cases located in the nucleus, can display a “classic” function, similar to the Vertebrate myoglobin (Mb) in storing O_2 in hypoxic conditions or facilitating the delivery of O_2 to the mitochondria. New functions have also emerged in the literature, such as the regulation of reactive oxygen species (ROS) or reactive nitrogen species (RNS) that can be deleterious to the cell. It has been also been speculated that they could function as O_2 or redox sensors by interacting with other proteins such as G proteins (Wakasugi et al., 2003) or cytochromes to transmit a signal.

The five Urbilaterian stem globins have undergone remarkably dissimilar fates in different Bilaterian groups (Figure 9). They have all been conserved in the Annelid *Platynereis dumerilii* and the Mollusks *Pinctada* (Bivalvia) and *Lottia* (Snail). In other Bilaterian groups, one or more of the ancestral types have been lost. Amniotes have kept only type I and type V globins. In an extreme situation, the Urochordate *Ciona* has only type I globins left but six type I paralogues are present, maybe compensating for the loss of the other types.

A novel scenario for the evolution of blood globins

Intra-species globin gene radiations occurred independently multiple times

Oxyphoric circulatory globins have evolved several times during the course of metazoan evolution. All circulatory multigenic families in our dataset derive from a likely

unique type I globin, possibly after recruitment for the circulatory function and several rounds of gene duplications. These multigenic families encompass for example the human hemoglobin family as well as the extracellular globin families of Annelids and *Daphnia*.

In the case of the hemoglobins of Vertebrate erythrocytes (RBC), alternate globins circulate depending on the development stage. Adult human hemoglobins form heterotetramers comprising two types of sub-units, α and β, coded by paralogous genes. During the course of human embryogenesis and foetal development, different types of hemoglobins are produced (δ, γ, ε, ζ), coded by other paralogous genes located in two clusters of α and β-related paralogues. The crucial emerging property of the hemoglobin tetramer is cooperativity: the binding of an O_2 molecule to one of the sub-unit induces a conformational change in the tetramer that makes the binding of O_2 easier to the three remaining sub-units.

Elsewhere among metazoans, other spectacular gene radiations correlated with hemoglobin-based blood evolution have occurred. Three of the concerned groups are sampled in our tree:

- Our phylogenetic analysis is in accordance with the scenario that early gene duplications, probably in the last common ancestor of modern Annelids, produced the initial two families (A, B) encompassing the four extracellular globin sub-families A1, A2, B1, B2 that form the hexagonal bilayer hemoglobin assemblage. A recent work based on a massive sampling of Annelid species transcriptomes recovered a variable number of extracellular globins, ranging from 1 to 12 depending on the species (Belato et al., 2019). Gene trees made in this later work suggest that, while the early gene duplication A/B indeed occurred prior to Annelid diversification, the sub-families A1, A2, B1 and B2 are not recognizable at the scale of the Annelid tree. This is in accordance with what we obtain at the much smaller scale of our Annelid sampling. This fact might be interpreted as evidence that the Annelid last common ancestor has only two type I paralogous globin genes, A and B, and that additional gene duplications occurred independently in many Annelid lineages. Alternatively, four paralogous genes (A1, A2, B1, B2) may have existed early on in Annelid history, accounting for the hexagonal bilayer hemoglobins found in divergent Annelid groups, but in this case the gene sequences are not informative enough to reconstitute early duplication steps. Our work differs however in a striking way from the results of Belato et al (2019). These authors identify a number of extracellular globins related to the Annelid A and B extracellular globins from a large and diversified subset of Bilaterian species including Echinoderms, Hemichordates, Brachiopods, Nemertines, Mollusks and Priapulids. We find no evidence for the existence of globins related to the Annelid extracellular globins (i.e. hexagonal bilayer globins) in Bilaterian complete genomes, other than Annelids. We argue

for a thorough reconsideration of transcriptome screens (Belato et al., 2019), as many protein globin sequences these authors found in non-Annelid Bilaterians are strikingly similar if not identical to individual Annelid sequences (as illustrated in their Figure 4). This is suggesting multiple tissue or DNA contaminations between species transcriptomes. Despite *Platynereis* blood hemoglobin having never been directly characterized by spectrometric tools, our EM and sequences similarities show that *Platynereis* must exhibit the giant assembly. In *Platynereis*, additional duplications have affected the A1 gene, making several paralogue sub-units. These sub-units, according to our *in situ* hybridization and qPCR data, may be used in producing alternative composition of circulating hemoglobin that would be used for physiological adaptations as the worm develops, grows and metamorphoses.

- The type I stem-globin evolved in a family of intracellular globins in some Annelid species, In the case of the Sedentarian worm *Capitella*, which does not have blood vessels, the hemocoel contains many red cells with intracellular globins with oxygen binding properties (Mangum et al., 1992). We have identified 17 intracellular globin paralogues, all closely related, in the genome of *Capitella* SpI and no extracellular globins. It appears as a case where the ancestral Annelid extracellular globins were lost secondarily and replaced by red blood cell globins.

- Although hemocyanin is the main respiratory pigment in Pancrustaceans (Burmester and Hankeln, 2009), Branchiopods such as *Daphnia*, a few Malacostracans and a few Insects rely on hemoglobin for respiration. *Daphnia* possess a large family of extracellular two-domains globin-coding genes. It is proposed that the hemolymph of *Daphnia* contains a crustacean hemoglobin made of 16 identical globin peptides with two hemes each (Ilan et al., 1982). For our phylogenetic analysis, we treated the two globin fold sequences independently, resulting in trees indicating that the multigenic family of *Daphnia* globins has emerged from the tandem duplications of an ancestral didomain globin gene.

However, globin gene radiations are not necessarily linked to a recruitment for a blood function. This is exemplified by Urochordates and Cephalochordates, respectively. The ascidian *Ciona intestinalis* possesses 6 type I globins, all closely related. The cephalochordate *Branchiostoma* has eight type I globins found in two clusters in this analysis. Both species have nevertheless a colorless blood and to date, no functional study has shown either oxygen binding properties or their presence in the circulatory system.

Type I globins were recruited several times for a circulatory function

All type I globins likely derive from a single ancestral type I sequence. This means that this ancestral type I globin was duplicated several times in different clades independently. This raises two additional questions:

- What was the ancestral function(s) of this type I globin that made it so well pre-adapted for a role in blood function?

- Is there a gene in current Bilaterian species that still carries this ancestral function(s)?

We can only speculate at this point because very little is known on the function of the non-circulating type I globins in the vast majority of Bilaterian species. But the existing knowledge can at least help defining the direction of future research. Part of the answer may come from Vertebrates that possess both circulatory and tissue-expressed type I globins. The “static” type I globins in Vertebrates are myoglobins and cytoglobins.

Myoglobins are monomeric type I globins present in all Vertebrates. Their function is well known and very specific: they are expressed in muscles and are capable of storing large quantities of O_2 necessary for sustained muscular effort.

Cytoglobins were first identified as distinct intracellular globins in Amniotes (Burmester et al., 2002b; Trent and Hargrove, 2002). They harbour a structural difference with hemoglobins (and myoglobin) in having a hexacoordinated iron, rather than pentacoordinated. They form homodimers, also by contrast to the former. Phylogenetically, they are firmly identified as type I globins in our study. Contrary to myoglobins, they are expressed in a large range of tissues and organs, with maybe higher expressions in cells producing large amount of cell matrix such as fibroblasts and chondroblasts (Hankeln et al, 2004). Information on the function (or functions) of cytoglobins remains scarce to this date. It is known that they bind dioxygen, carbon monoxide but also nitric oxide with high affinity (Oleksiewicz et al., 2011). It has thus been proposed that cytoglobins could play multiple roles in intracellular homeostasis. The first suggestion is of course related to dioxygen binding including roles in oxygen buffering, sensing, transport and storage. The activity of cytoglobin as nitric oxide deoxygenase, a very ancient function of globin proteins in eukaryotes, is also proposed to play an important role in the biology of the cell. Last, oxidative stress is another circumstance in which cytoglobins may play a role.

Remarkably, molecular phylogenies have demonstrated that large families of oxyphoric circulatory hemoglobins have evolved by gene duplications two times in the Vertebrates: once in the Gnathostomes and once in Agnathans (Fago et al., 2018; Schwarze et al., 2014). Interestingly, in these molecular trees, the cytoglobins of Cnathostomes and Agnathans are found as sister groups to both their large radiations of circulatory globins, suggesting that cytoglobin existed prior to the origin of red cell globins in Vertebrates. Cytoglobins have low evolutionary rates, which suggest that they have kept a conservative mode of evolution. There are good reasons to propose that cytoglobins (rather than myoglobin) may be close to this type I globin with ancestral functions, as was suggested before (Blank and Burmester, 2012).

The fruitfly *Drosophila*, like other insects, has a tracheal respiration and no circulatory globins. Yet, genome screens have revealed three globin genes in the *Drosophila* genome (Burmester et al., 2006). Two of these globins possess derived sequences. But the third one (Dme_524369; “Glob1” in Burmester et al, 2006) is found solidly clustered with type I globins in our tree. This globin is hexacoordinated and may represent a derived cytoglobin-like molecule. It is found expressed in many tissues in the developing embryo and larva (Yadav et al., 2015) but prominently in the tracheal cells. Gene knock-downs (Gleixner et al., 2016; Yadav et al., 2015) lead to reduced survival of flies under hypoxic conditions. This suggests a role in intracellular O_2 homeostasis.

Both cytoglobins and *Drosophila Glob1* may thus be close to the functions of a type I stem globin. It has to be noted that several species (Arthropods and Mollusks) have a single type I globin that may represent the unduplicated descendants of the stem globin. Few species have completely lost the type I (only the Annelid *Helobdella* and the Cephalopod *Octopus*). This may reflect the initial functional importance of type I stem globin.

Why have multiple gene duplications affected the type I stem globin each time it has been recruited for a blood function? As we have mentioned earlier, cooperativity is an important functional mechanism for a multimeric globin whose role is to store massive quantity of O_2 in the blood. It may be best developed in a heteromultimeric assemblage and thus requires duplication of the O_2 storing protein. The second factor is the developmental switch of globins. This is well illustrated by the lamprey and the Gnathostome cases. The lamprey *Petromyzon* possesses 18 hemoglobin genes that are expressed differentially at three different developmental stages, embryo, larva and adult. These developmental switches are similar to those found in Gnathostomes. Yet, all the Agnathan circulatory globins are found in a single monophyletic group completely distinct from the radiation of gnathostome hemoglobins. The lamprey globins family was convergently recruited for the very same kind of developmental as seen in Gnathostomes,. We may see to a more limited extent the same developmental switch in *Platynereis*, with different A1 globins specialized for the juvenile and the adult.

The hemogenic cells: a specialized type of blood-making cells

We demonstrate in this work the existence of a specialized category of cells in *Platynereis*, that produces the blood globins, the hemogenic cells. We analysed the development of these cells during the life cycle of *Platynereis*. They first appear in the posterior most segments in a five-segment worm. The youngest feeding juvenile worms with 3 or 4 segments probably do not have hemogenic cells and therefore no hemoglobin in their blood. These are minute worms (less than 500 µm long) that probably perform gas exchanges with the seawater by simple diffusion. In intermediate size juvenile worms (30-40

segments, less than one cm long), hemogenic cells form sheaths around lateral trunk vessels, in most segments of the worm. As the worm grows, the circulatory system structure becomes more complex and a rich network of vessels and capillaries develop inside the parapodia, the worm segmental appendages. These parapodia take on the role of gills. Hemogenic cells develop around most vessels inside the parapodia. We currently do not know whether these are the cells of the wall of existing vessels that are differentiating *in situ* to become hemogenic cells, or whether hemogenic cells colonize the parapodia by emigrating from another location. We also show that at this stage (worms larger than 40 segments), hemogenic cells within the trunk degenerate and are presumably digested by eleocytes. Worms accumulate more red blood as they get close to sexual metamorphosis. When they have a maximal quantity of blood and start metamorphosing into swarming epitoks, the hemogenic cells start to stop producing hemoglobins. It is possible as well that these cells degenerate but we have not demonstrated this phenomenon in metamorphosing worms.

To our knowledge, this is the first time the full cycle of these hemogenic cells is described in any Annelid species and this is also the first time they are found associated with the parapodia serving as gills. Hemogenic cells with very similar cytological properties have been described in other Annelid species. They have been called perivasal cells because of their position around the lumen of vessels or extravasal cells, when they are located in the mesodermal peritoneum covering the gut in earthworms (Lindner, 1965), Siboglinids (Nakahama et al., 2008) or *Arenicola* (Breton-Gorius, 1963). In some Sedentarian Annelids such as *Amphitrite* (Friedman and Weiss, 1980), a specialized organ, the heart-body, produces the hemoglobins. The heart-body is a solid mass of tissue that grows inside the lumen of the anterior dorsal vessel. There is thus a remarkable plasticity in the location of this hemogenic tissue in Annelids that will be worth studying further.

Is there a phylogenetic connection between metazoan hemogenic cells? To our opinion, this is a question that deserves investigation in the future. We propose that hemogenic cells are by definition the cells that produce the blood or hemolymph respiratory pigments, may it be hemoglobin, hemocyanin ot hemerythrin. Hemogenesis is thus quite different in definition from hematopoiesis, that is the production of the circulatory cells that make part of the blood as a sort of “liquid tissue”. By extension, the term hematopoiesis is also sometimes applied to the production of the cells that populate the coelom (coelomocytes). In evolutionary terms, this definition extension makes sense if we consider the hypothesis that hemocytes were once coelomocytes that populated the blood compartment. The term hemocytes is used for free cells both in the blood and coelom by Hartenstein (2006). Hartenstein proposes that, on the basis of developmental and molecular

similarities, hemocytes of the various Bilaterian phyla have a common evolutionary origin and derive from a common type of progenitor cell, the hemocytoblast.

What is the place of the hemogenic cells relative to hematopoiesis processes? It should be noted that, in *Platynereis* and in Annelids in general, hemogenic cells are not circulating. They are part of a mesodermal epithelium that is in contact with the coelomic cavity on the apical side and with the blood vessel lumen on the basal side. They are finely functionally polarized cells that secrete hemoglobin only on the basal side toward the vessel lumen. Besides hemogenic cells, *Platynereis* (and Annelids in general) have several classes of coelomocytes/hemocytes that are freely floating in the coelomic cavity and can cross the mesoepithelium to populate the blood. Can hemogenic cells be considered as cells produced by an extended hematopoietic process? This is not only a question of convention, when considering the diversity of hemogenic cells in Metazoans.

In the various Bilaterian groups that possess either a hemocoel or coelom/blood compartments, hemogenic cells can be either free floating or static. In Vertebrates, intracellular globins are massively produced and stored in red blood cells. Accumulating respiratory pigments in circulating cells is also the solution retained in a number of other groups. Within Annelids, the bloodworm *Glycera* have red cells filled with monomeric and polymeric hemoglobins, circulating in the coelom as the BVS is much reduced in these worms (Mangum et al., 1989). Capitellids have also a reduced BVS and red cells. Also in Annelids, Sipunculidae have pink cells filled with hemerythrin, illustrating the evolutionary pigment swapping which has occurred in several groups. Data on the existence and location of pigment producing cells in other protostome invertebrates are still patchy. In the marine Chelicerate *Limulus*, whose remarkable blue blood has been exploited for preparing bacterial endotoxin tests, the hemocyanin is secreted by hemogenic cells called cyanoblasts. Cyanoblasts are found floating in the hemocoel and burst to liberate their hemocyanin content (Fahrenbach, 1970). The tissue of origin of these cyanoblasts is unknown. In Gastropod Mollusks, hemogenic cells responsible for hemocyanin production have been identified as the pore cells or rhogocytes (Albrecht et al., 2001), abundant in the connective tissue. In pulmonate snails that have hemoglobin instead of hemocyanin, it has been shown that these rhogocytes have switched to hemoglobin production (Kokkinopoulou et al., 2015).

Conclusion

We propose a new nomenclature for Metazoan globin multigenic family based on our phylogenetic study which relied on whole genomes sequences and wide sampling of species. This phylogeny provides a new scenario on globins and blood evolution: a set of five globin genes, which we call “stem-globins” were present in *Urbilateria*. These genes derived

respectively into 5 families of globin genes present in the extant Bilaterians, which we name type I to V. In the course of Bilaterian speciations, some of these members were lost, others duplicated into paralogous subfamilies. Membrane bound globins are common in all globin types, except from the type I. All the known circulatory globins, i.e. globins entering the composition of the respiratory pigments in red blood in Bilaterians, derive from this ancestral type I stem-globin. We note that the recruitment for the respiratory function of this type I globin was each time associated with paralogous duplication events. The Annelid *Platynereis* has retained the complete set (type I to V) of ancestral globin genes. It possesses 7 type I globins which are all extracellular. This is consistent with the known presence in various Annelids of huge extracellular hemoglobins with respiratory function. Although such pigment presence was never proved on crystallography basis in *Platynereis*, our results confirm it through phylogenetic, expression and morphological evidence. We identified the cells that secrete these extracellular globins, which we call “hemogenic”. Their location and activity depends on the worm size and life stages. These cells are part of the mesothelial walls of some particular vessels: lateral vessels in juveniles, and vessels in the appendages in the maturing worms. The blood production level peaks at the onset of the maturation process, likely in preparation for the adult locomotory activity peak. The hemogenic cell morphology established with EM are similar to pigment producing tissues described in other Annelids but their location in the appendages vessels serving as gills is original.

The finding of hemogenic cells within the gill organs in a marine animal is an important step forward. *Platynereis* is easily tractable to molecular biology experiments and can be used for obtaining transgenic strains. The molecular characterization of these cells will be very useful for comparative studies and exploring the diversity of hemogenic cells in Bilaterians. One possibility would be to use single-cell RNAseq, as these hemogenic cells should be easily singled out in the data because of their massive production of extracellular globins. However, scRNAseq remains relatively expensive and shallow. Another possibility would be to develop a CRISPR-Cas9 protocol for tagging an extracellular globin gene with a GFP coding sequence. This would allow sorting hemogenic cells with flow cytometry and the bulk sequencing of their transcriptomes.

Another goal will be to understand the embryonic origin not only of the hemogenic cells but also of the other mesoepithelial cells that are involved in forming the blood vessels. Is there a developmental and lineage connection between these mesoepithelial cells of the vessels, the hemogenic cells and also the coelomocytes of *Platynereis*? In other words, do we have the equivalent of a hemangioblast lineage in the Annelid? This question is all the more important if we consider the hypothesis that the BVS evolved in the first place to distribute nutrients and was later on recruited for a gas exchange function.

Platynereis may be an excellent model in the future also to establish the functions of the "stem globins". Studying animals in which these globins have been selectively inactivated by CRISPR-Cas9 gene editing.

Tables

Table 1

This study	**Blank and Burmester, 2012**
Type I	Clade III, IV and V
Type II	GbX-like, clade II
Type III	GbX, clade II
Type IV	?
Type V	Neuroglobins, clade I
Spiralian globins	*not present*

Table legends

Table 1: Correspondence between the phylogeny-based nomenclature of metazoan globins proposed in this article and the previous most extensive nomenclature proposed (Blank and Burmester, 2012).

Figure legends

Figure 1: Consensus phylogenetic tree of the main metazoan phyla, indicating the presence or absence of a circulatory system(s) and the type of respiratory pigment used. Hb: Hemoglobin. Hc: Hemocyanin. Hm: Hemerythrin. BVS: blood vascular system. CCS: closed coelomic system.

Figure 2: The life cycle of *Platynereis dumerilii*. See text for explanations

Figure 3: Ultrastructure of *P. dumerilii* hemogenic cells. Juvenile worm, 25-30 segments. A: transversal semithin section, showing the position of hemogenic cells around the transversal blood vessel. B: closer view of the dissected transversal vessel surrounded by hemogenic cells (rectangle selection in A). C: Transmission electron micrograph of the overall hemogenic tissue organization. D-G: Transmission electron micrographs of different parts of

the hemogenic cells. D: Apical part of the cell, containing nucleus (*nu*) and rough endoplasmic reticulum (*rer*). E: Basal part of the cell with mitochondria (*mt*), rough endoplasmic reticulum (*rer*), and ferritin granules (*fg*). F: closer view of a ferritin granule (*fg*). G: Invaginations of the basal membrane (arrowheads) indicating active exocytosis process.
bv – blood vessel lumen, *coe* – coelomic cavity, *dbv* – dorsal blood vessel, *ecm* – extracellular matrix (vascular lamina), *fg* – ferritin granules, *hc* – hemogenic cells, *mt* – mitochondria, *nu* – cell nucleus, *rer* – rough endoplasmic reticulum, *vnc* – ventral nerve cord.
Scale bars : A, B: 50 microns, C: 10 microns, D: 4 microns, E-G: 2 microns

Figure 4: 3D-reconstruction of the blood vascular system (red) and hemogenic cells (blue) in body segments of *Platynereis dumerilii*. A, B: juvenile worm, 25-30 segments. C, D: pre-mature worm, about 50 segments. The dense intestinal blood network (or gut blood sinus) is not shown.
dbv – dorsal blood vessel, *hc* – hemogenic cells, *i* – intestine, *ibv* – intestinal blood vessels, *npv* – nephridial vessels, *tbv* – transversal blood vessels, *vbv* – ventral blood vessel.

Figure 5: The 17 globin genes found in *Platynereis* transcriptomes. The characteristic introns positions B12.2 and G7.0 are marked as inverted triangles, respectively pink and green. Additional introns positions in colourless triangles. Signal peptides are represented in green.

Figure 6: RNA expression patterns of a representative extracellular globin genes in *Platynereis dumerilii* WMISH on three different juvenile stages. The expression patterns were similar for all extracellular genes. Here we show the results for Pdu-Egb-A2.
A: 18 segments juvenile show expression in lateral vessels in the trunk in a series of adjacent segments. There is no expression in the rostral and caudal segments.
B and C: 21 segments juvenile and 35 segments juvenile show expression in the parapodia for central segments and expression in lateral vessels in a few segments more rostrally and caudally located. There is no expression in the segments at the extremities of the animal.
D: Schematic interpretation of the evolution of the expression pattern in time.

Figure 7: Distribution of expression levels of all extracellular globin genes combined, detected by real-time qPCR, in different stages. Metamorphing stage 1 is defined according to the following criteria: eyes starting to bulge, transforming parapodia. Metamorphing stage 2: eyes grown to maximal surface, parapodia fully transformed, but the animal is not displaying active swimming behaviour yet.

Figure 8: A: Maximum likelihood tree using LG model and a SH-like test of a dataset of 272 sequences from 24 metazoan species listed in the animals tree in B. The colours of the branches correspond to the animals phyla colour coding in B. The diamonds on the nodes represent the aLRT scores: green if between 0.75 and 0.95, and red if greater than 0.95.
The tree shows 5 groups (type I, type II, type III, type IV, type V), 4 of the 5 with well-supported nodes (type I, type II, type III, type IV), encompassing sequences of all Bilaterian phyla. An additional well- supported node groups together only Lophotrochozoan species sequences. *Platynereis* sequences are indicated in red. C: Summary of the support scores of the different groups with Maximum Likelihood and Bayesian algorithms.

Figure 9: Evolutionary scenario for Bilaterian globin genes evolution Animals tree with the 24 sampled species. Colored disk indicate the presence of representatives of different types of globins as revealed by the molecular phylogenetic tree. The inferred ancestral composition of the globin super family is indicated at some key nodes in the tree, with the same color codes. Crossed disks indicate secondary evolutionary loss. Interrogation marks indicate that the whole genome is not available for the relevant species and although we did not find globin genes of the corresponding type in such species, we cannot exclude their presence in the genome. The table recapitulates the number of sequences of each family of globin genes found in each species. Note that for *Daphnia pulex* that has a family of didomain globin genes, 17 is the total number of globin domains.

Supplementary Figure legends

Supplementary Figure 1: 3D reconstruction of the blood vascular system (red) and hemogenic cells (blue) in 4 (A-B), 5 (C-D), and 6(E-F) segments larvae of *Platynereis dumerilii*. Tbs: transversal blood vesels. In green: putative degenerative hemogenic cells

Supplementary Figure 2: Amino acids alignment of the *Platynereis* extracellular globins with the respiratory globins of *Arenicola marina*, *Lumbricus terrestris* and human hemoglobins. The highly conserved residues of globins in dark red (Motonori Ota, FEBS Letters, 1997). The conserved cysteines NA2 and H7 residues known to be involved in the formation of an intrachain disulfide chain in Annelids are shown in black (ref Globin gene family evolution and functional diversification in Annelids, Xavier Bailly et al., FEBS 2007).

Supplementary Figure 3: Transmission electron micrographs of the degenerating hemogenic cells in pre-mature *P. dumerilii* (about 50 segments). A. cross-section through the transversal blood vessel at the level of hemogenic cells. B: Closer view of A, showing the

cytoplasm, filled by electron-lucent vesicles (*arrowheads*). C: High magnification of the large electron-lucent vesicle (*asterisk*). D: Granulocyte filled with electron-dense and electron-lucent granules. Note the pseudopodia with a phagosome at the bottom-left (*arrow*).
bv – blood vessel lumen, *dmc* – muscular cells of the dissepiment, *gc* – granulocyte, *nu* – cell nucleus.
Scale bars : A: 10 microns, B-D: 4 microns.

Supplementary Figure 4: NBT/BCIP *in situ* hybridization on maturing worms parapodia for the extracellular globin genes Egb-A1c (upper right panel) and Egb-B1 (lower right panel). The expression is lining the vessels in the parapodia but sems to be in the process of decay: there are only islands of expression left along some vessels.

Supplementary Figure 5: Relative abundance of mRNA of the extracellular globin genes Egb-A1a (upper left panel), Egb-A1b (upper right panel), Egb-A1c (lower left panel), Egb-A1d (lower right panel) with respect to reference genes for each stage. For stages 50 segments, type I, type II and swimmers, the data points correspond to biological triplicates: three individuals identified as being at the same stage of development.

Supplementary Figure 6: Relative abundance of mRNA of the extracellular globin genes Egb-A2 (upper left panel), Egb-B1 (upper right panel), Egb-B2 (lower left panel) with respect to reference genes for each stage. For stages 50 segments, type I, type II and swimmers, the data points correspond to biological triplicates: three individuals identified as being at the same stage of development.

Supplementary Figure 7: Maximum likelihood tree using LG model and a SH-like test of a dataset of 272 sequences from 24 metazoan species, (simplified in Figure 8). The labels of the leaves contain the name of the species, in three letters. First letter is the initial of the genus, second and third letters are the first two letters of the species.

Supplementary Figure 8: Same tree as Suppl. Fig. 7. Arthropod, Annelid and Vertebrate blood globin radiations among the type I globins are highlighted.

Supplementary Figure 9: Same tree as Suppl. Fig. 7. OTUs branches are put in colors indicating their predicted membrane-bound properties. Branches in blue: predicted myristoylated sequences. Branches in red: extracellular globins sequences. Branches in pink: human respiratory globins. The sequences among our dataset that are predicted as

myristoylated are distributed in all types globins except from type I. Also, a single protein in type V (Ngb) is predicted membrane-bound.

Acknowledgements

We thank the Balavoine Lab Members and Biofluidics Lab members for helpful discussions during the course of this work. We acknowledge Nathalie Luciani for equipment and training for qPCR experiments and analysis. We acknowledge the ImagoSeine facility (member of the France BioImaging infrastructure supported by the French National Research Agency, ANR- 10-INSB-04, 'Investments of the future') and its experienced staff for their assistance in imaging and image analysis. S. Song obtained a PhD fellowship from the LABEX "WHO AM I?" (No.ANR-11-LABX-0071). The Balavoine Lab was funded by two grants from the ANR (METAMERE no. ANR-12- BSV2-0021 and TELOBLAST no. ANR-16- CE91-0007).

References

Albrecht, U., Keller, H., Gebauer, W. and Markl, J. (2001). Rhogocytes (pore cells) as the site of hemocyanin biosynthesis in the marine gastropod Haliotis tuberculata. *Cell Tissue Res.* **304**, 455–462.

Anisimova, M. and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552.

Arendt, D., Denes, A. S., Jékely, G. and Tessmar-Raible, K. (2008). The evolution of nervous system centralization. *Philos. Trans. R. Soc. B Biol. Sci.* **363**, 1523–1528.

Ascenzi, P., Gustincich, S. and Marino, M. (2014). *Mammalian nerve globins in search of functions*. Wiley-Blackwell.

Bailly, X., Chabasse, C., Hourdez, S., Dewilde, S., Martial, S., Moens, L. and Zal, F. (2007). Globin gene family evolution and functional diversification in annelids. *FEBS J.* **274**, 2641–2652.

Balavoine, G. (2003). The Segmented Urbilateria: A Testable Scenario. *Integr. Comp. Biol.* **43**, 137–147.

Belato, F. A., Schrago, C. G., Coates, C. J., Halanych, K. M. and Costa-Paiva, E. M. (2019). Newly Discovered Occurrences and Gene Tree of the Extracellular Globins

and Linker Chains from the Giant Hexagonal Bilayer Hemoglobin in Metazoans. *Genome Biol. Evol.* **11**, 597–612.

Blank, M., Wollberg, J., Gerlach, F., Reimann, K., Roesner, A., Hankeln, T., Fago, A., Weber, R. E. and Burmester, T. (2011). A Membrane-Bound Vertebrate Globin. *PLoS ONE* **6**, e25292.

Breton-Gorius, J. (1963). *Etude au microscopie électronique des cellules chloragogènes d'Arenicola marina L: leur rôle dans la synthèse de l'hémoglobine*. Masson.

Burmester, T. (2015). Evolution of respiratory proteins across the pancrustacea. *Integr. Comp. Biol.* **55**, 792–801.

Burmester, T. and Hankeln, T. (2009). What is the function of neuroglobin? *J. Exp. Biol.* **212**, 1423–1428.

Burmester, T., Weich, B., Reinhardt, S. and Hankeln, T. (2000). A vertebrate globin expressed in the brain. *Nature* **407**, 520–523.

Burmester, T., Ebner, B., Weich, B. and Hankeln, T. (2002b). Cytoglobin: A Novel Globin Type Ubiquitously Expressed inVertebrate Tissues. *Mol. Biol. Evol.* **19**, 416–421.

Burmester, T., Storf, J., Hasenjäger, A., Klawitter, S. and Hankeln, T. (2006). The hemoglobin genes of Drosophila. *FEBS J.* **273**, 468–480.

Christensen, A. B., Herman, J. L., Elphick, M. R., Kober, K. M., Janies, D., Linchangco, G., Semmens, D. C., Bailly, X., Vinogradov, S. N. and Hoogewijs, D. (2015). Phylogeny of echinoderm hemoglobins. *PLoS ONE* **10**,.

Couso, J. P. (2009). Segmentation, metamerism and the Cambrian explosion. *Int. J. Dev. Biol.* **53**, 1305–1316.

Dales, R. P. and Pell, J. S. (1970). Cytological aspects of haemoglobin and chlorocruorin synthesis in polychaete annelids. *Z. Für Zellforsch. Mikrosk. Anat.* **109**, 20–32.

De Robertis, E. M. (2008). Evo-Devo: Variations on Ancestral Themes. *Cell* **132**, 185–195.

Denes, A. S., Jékely, G., Steinmetz, P. R. H., Raible, F., Snyman, H., Prud'homme, B., Ferrier, D. E. K., Balavoine, G. and Arendt, D. (2007). Molecular Architecture of Annelid Nerve Cord Supports Common Origin of Nervous System Centralization in Bilateria. *Cell* **129**, 277–288.

Dröge, J., Pande, A., Englander, E. W. and Maka\lowski, W. (2012). Comparative Genomics of Neuroglobin Reveals Its Early Origins. *PLoS ONE* **7**,.

Ebner, B., Panopoulou, G., Vinogradov, S. N., Kiger, L., Marden, M. C., Burmester, T. and Hankeln, T. (2010). The globin gene family of the cephalochordate amphioxus: Implications for chordate globin evolution. *BMC Evol. Biol.* **10**, 370.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.

Erwin, D. H. and Davidson, E. H. (2002). The last common bilaterian ancestor. *Development* **129**, 3021–3032.

Fahrenbach, W. H. (1970). The cyanoblast: Hemocyanin formation in Limulus Polyphemus. *J. Cell Biol.* **44**, 445–453.

Fischer, A. and Dorresteijn, A. (2004). The polychaete Platynereis dumerilii (Annelida): A laboratory animal with spiralian cleavage, lifelong segment proliferation and a mixed benthic/pelagic life cycle. *BioEssays* **26**, 314–325.

Fischer, A. H. L., Henrich, T. and Arendt, D. (2010). The normal development of Platynereis dumerilii (Nereididae, Annelida). *Front. Zool.* **7**, 31.

Friedman, M. M. and Weiss, L. (1980). An electron microscopic study of hemoglobin synthesis in the marine annelid,Amphitrite ornata (Polychaeta: Terebellidae). *J. Morphol.* **164**, 121–138.

Gaunt, S. J. (2018). Hox cluster genes and collinearities throughout the tree of animal life. **683**, 673–683.

Gleixner, E., Ripp, F., Gorr, T. A., Schuh, R., Wolf, C., Burmester, T. and Hankeln, T. (2016). Knockdown of Drosophila hemoglobin suggests a role in O2homeostasis. *Insect Biochem. Mol. Biol.* **72**, 20–30.

Grigorian, M. and Hartenstein, V. (2013). *Hematopoiesis and hematopoietic organs in arthropods*. Springer-Verlag.

Guindon, S. and Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.* **52**, 696–704.

Hartenstein, V. and Mandal, L. (2006). The blood/vascular system in a phylogenetic perspective. *BioEssays* **28**, 1203–1210.

Hejnol, A. and Martindale, M. Q. (2008). Acoel development supports a simple planula-like urbilaterian. *Philos. Trans. R. Soc. B Biol. Sci.* **363**, 1493–1501.

Hoffmann, F. G., Opazo, J. C., Hoogewijs, D., Hankeln, T., Ebner, B., Vinogradov, S. N., Bailly, X. and Storz, J. F. (2012). Evolution of the globin gene family in deuterostomes: Lineage-specific patterns of diversification and attrition. *Mol. Biol. Evol.* **29**, 1735–1745.

Ilan, E., Weisselberg, E. and Daniel, E. (1982). Erythrocruorin from the water-flea Daphnia magna. Quaternary structure and arrangement of subunits. *Biochem. J.* **207**, 297–303.

Jhiang, S. M., Garey, J. R. and Riggs, A. F. (1986). Identification of mRNA Coding for Globin in the Chloragogen Cells of Lumbricus Terrestris. In *Invertebrate Oxygen Carriers* (ed. Linzen, B.), pp. 469–472. Springer Berlin Heidelberg.

Kakar, S., Hoffman, F. G., Storz, J. F., Fabian, M. and Hargrove, M. S. (2010). Structure and reactivity of hexacoordinate hemoglobins. *Biophys. Chem.* **152**, 1–14.

Kennedy, G. Y. and Dales, R. P. (1958). The Function of the Heart-Body in Polychaetes. *J. Mar. Biol. Assoc. U. K.* **37**, 15–31.

Klippenstein, G. L. (1980). Structural aspects of hemerythrin and myohemerythrin. *Integr. Comp. Biol.* **20**, 39–51.

Kokkinopoulou, M., Spiecker, L., Messerschmidt, C., Barbeck, M., Ghanaati, S., Landfester, K. and Markl, J. (2015). On the Ultrastructure and Function of Rhogocytes from the Pond Snail Lymnaea stagnalis. *PLOS ONE* **10**, e0141195.

Lauri, A., Brunet, T., Handberg-Thorsager, M., Fischer, A. H. L., Simakov, O., Steinmetz, P. R. H., Tomer, R., Keller, P. J. and Arendt, D. (2014). Development of the annelid axochord: Insights into notochord evolution. *Science* **345**, 1365–1368.

Le, S. Q. and Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320.

Lindner, E. (1965). Ferritin und Hemoglobin im Chloragog von Lumbriciden (Oligochaeta). *Z. Für Zellforsch. Mikrosk. Anat.* **66**, 891–913.

Mangum, C. P. and Kondon, M. (1975). The role of coelomic hemerythrin in the sipunculid worm Phascolopis gouldi. *Comp. Biochem. Physiol. A Physiol.* **50**, 777–785.

Mangum, C. P., Colacino, J. M. and Vandergon, T. L. (1989). Oxygen binding of single red blood cells of the annelid bloodwormGlycera dibranchiata. *J. Exp. Zool.* **249**, 144–149.

Mangum, C. P., Colacino, J. M. and Grassle, J. P. (1992). Red Blood Cell Oxygen Binding in Capitellid Polychaetes. *Biol. Bull.* **182**, 129–134.

Martín-Durán, J. M., De Mendoza, A., Sebé-Pedrós, A., Ruiz-Trillo, I. and Hejnol, A. (2013). A broad genomic survey reveals multiple origins and frequent losses in the evolution of respiratory hemerythrins and hemocyanins. *Genome Biol. Evol.* **5**, 1435–1442.

Maurer-Stroh, S., Eisenhaber, B. and Eisenhaber, F. (2002). N-terminal N-myristoylation of proteins: Prediction of substrate proteins from amino acid sequence. *J. Mol. Biol.* **317**, 541–557.

Nakahama, S., Nakagawa, T., Kanemori, M., Fukumori, Y. and Sasayama, Y. (2008). Direct Evidence That Extracellular Giant Hemoglobin is Produced in Chloragogen Tissues in a Beard Worm, Oligobrachia mashikoi (Frenulata, Siboglinidae, Annelida). *Zoolog. Sci.*

Oleksiewicz, U., Liloglou, T., Field, J. K. and Xinarianos, G. (2011). Cytoglobin: biochemical, functional and clinical perspective of the newest member of the globin family. *Cell. Mol. Life Sci.* **68**, 3869–3883.

Roesner, A., Fuchs, C., Hankeln, T. and Burmester, T. (2005). A Globin Gene of Ancient Evolutionary Origin in Lower Vertebrates: Evidence for Two Distinct Globin Families in Animals. *Mol. Biol. Evol.* **22**, 12–20.

Rohlfing, K., Stuhlmann, F., Docker, M. F. and Burmester, T. (2016). Convergent evolution of hemoglobin switching in jawed and jawless vertebrates. *BMC Evol. Biol.* **16**, 30.

Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Royer, W. E., Strand, K., van Heel, M. and Hendrickson, W. A. (2000). Structural hierarchy in erythrocruorin, the giant respiratory assemblage of annelids. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7107–11.

Royer, W. E., Sharma, H., Strand, K., Knapp, J. E. and Bhyravbhatla, B. (2006). Lumbricus Erythrocruorin at 3.5 Å Resolution: Architecture of a Megadalton Respiratory Complex. *Structure* **14**, 1167–1177.

Royer, W. E., Omartian, M. N. and Knapp, J. E. (2007). Low Resolution Crystal Structure of Arenicola Erythrocruorin: Influence of Coiled Coils on the Architecture of a Megadalton Respiratory Protein. *J. Mol. Biol.* **365**, 226–236.

Saudemont, A., Dray, N., Hudry, B., Le Gouar, M., Vervoort, M. and Balavoine, G. (2008). Complementary striped expression patterns of NK homeobox genes during segment formation in the annelid Platynereis. *Dev. Biol.* **317**, 430–443.

Schwarze, K., Campbell, K. L., Hankeln, T., Storz, J. F., Hoffmann, F. G. and Burmester, T. (2014). The globin gene repertoire of lampreys: Convergent evolution of hemoglobin and myoglobin in jawed and jawless vertebrates. *Mol. Biol. Evol.* **31**, 2708–2721.

Storz, J. F., Opazo, J. C. and Hoffmann, F. G. (2011). Phylogenetic diversification of the globin gene superfamily in chordates. *IUBMB Life* **63**, 313–322.

Trent, J. T. and Hargrove, M. S. (2002). A ubiquitously expressed human hexacoordinate hemoglobin. *J. Biol. Chem.* **277**, 19538–19545.

Vinogradov, S. N. and Moens, L. (2008). *Diversity of globin function: Enzymatic, transport, storage, and sensing*. American Society for Biochemistry and Molecular Biology.

Vinogradov, S. N., Hoogewijs, D., Bailly, X., Arredondo-Peter, R., Gough, J., Dewilde, S., Moens, L. and Vanfleteren, J. R. (2006). A phylogenomic profile of globins. *BMC Evol. Biol.* **6**, 1–17.

Vinogradov, S. N., Hoogewijs, D., Bailly, X., Mizuguchi, K., Dewilde, S., Moens, L. and Vanfleteren, J. R. (2007). A model of globin evolution. *Gene* **398**, 132–142.

Vinogradov, S. N., Bailly, X., Smith, D. R., Tinajero-trejo, M., Poole, R. K. and Hoogewijs, D. (2013). *Microbial Eukaryote Globins*. 1st ed. Copyright © 2013 Elsevier Ltd. All rights reserved.

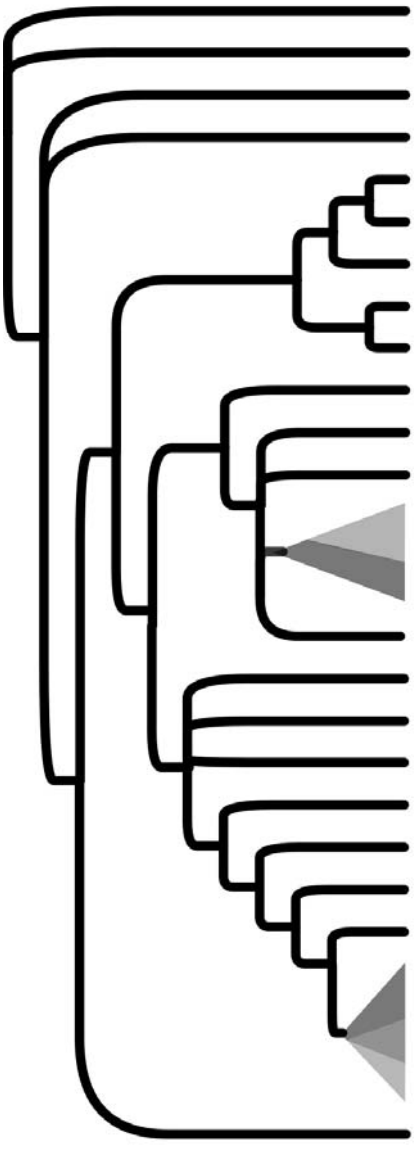
Wakasugi, K., Nakano, T. and Morishima, I. (2003). Oxidized human neuroglobin acts as a heterotrimeric Gα protein guanine nucleotide dissociation inhibitor. *J. Biol. Chem.* **278**, 36505–36512.

Yadav, R., Kundu, S. and Sarkar, S. (2015). *Drosophila glob1* expresses dynamically and is required for development and oxidative stress response: *Drosophila glob1* Expresses Dynamically. *genesis* **53**, 719–737.

Zantke, J., Bannister, S., Veedin Rajan, V. B., Raible, F. and Tessmar-Raible, K. (2014). Genetic and genomic tools for the marine annelid Platynereis dumerilii. *Genetics* **197**, 19–31.

Zhang, J. H. and Kurtz, D. M. (1991). Two distinct subunits of hemerythrin from the brachiopod Lingula reevii: an apparent requirement for cooperativity in oxygen binding. *Biochemistry* **30**, 9121–9125.

Zuckerkandl, E. and Pauling, L. B. (1962). Molecular disease, evolution, and genetic heterogeneity. In *Horizons in Biochemistry* (ed. Kasha, M.) and Pullman, B.), pp. 189–225. New York: Academic Press.



	Circulatory Pigment	Circulatory systems
Porifera	-	-
Ctenophora	-	-
Placozoa	-	-
Cnidaria	-	-
Vertebrata	HG	Closed BVS + Coelom
Urochordata	-	Open BVS + Hemocoel
Cephalochordata	-	Closed BVS + Coelom
Hemichordata	-	Open BVS + Hemocoel
Echinodermata	-	BVS + CCS
Platyhelminthes	-	-
Brachiopoda	-	Open BVS
Nemertea	-	CCS
Cephalopoda	HC	Closed BVS
other Mollusca	HC sometimes HG	Open BVS + Hemocoel
Annelida	HG sometimes HE	Closed BVS + Coelom
Priapulida	HE	Pseudocoelom
Kinorhyncha	-	Pseudocoelom
Nematoda	-	Pseudocoelom
Tardigrada	-	-
Onychophora	-	Open BVS + Hemocoel
Chelicerata	HC	Open BVS + Hemocoel
Myriapoda	HC	Open BVS + Hemocoel
other Crustacea	HC or HG	Open BVS + Hemocoel
Branchiopoda	HG	Open BVS + Hemocoel
Insecta	-	Open BVS + Hemocoel
Acoelomorpha	-	-

Figure 1

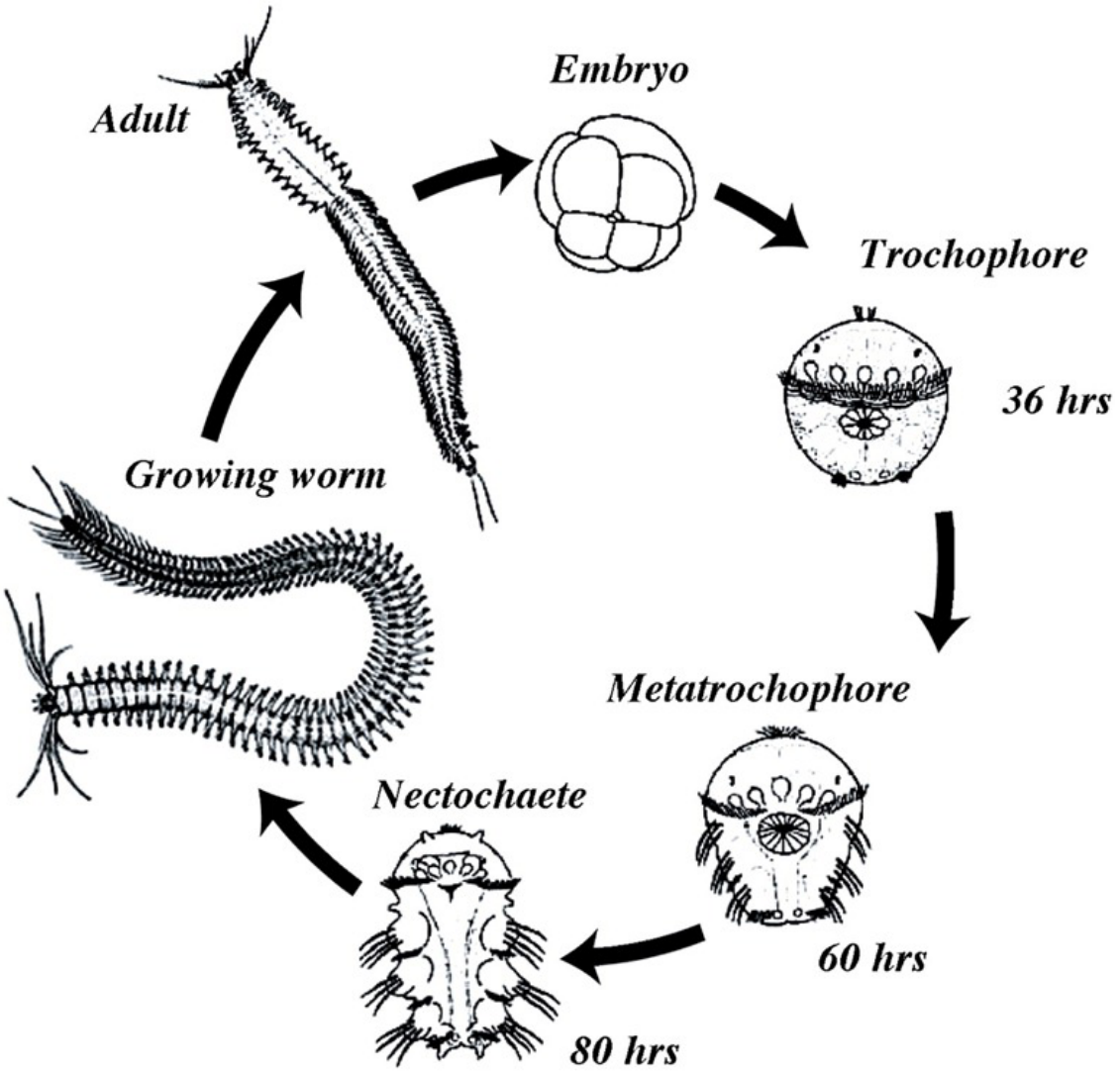

Adult
Embryo
Trochophore
36 hrs
Growing worm
Metatrochophore
Nectochaete
60 hrs
80 hrs

Figure 2

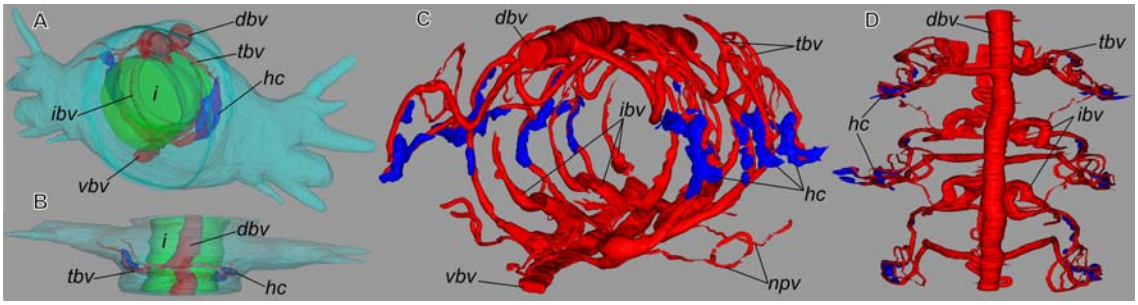

A
dbv
tbv
hc
i
ibv
vbv
B
i
dbv
tbv
hc
C
dbv
tbv
ibv
hc
vbv
npv
D
dbv
tbv
hc
ibv

Figure 3

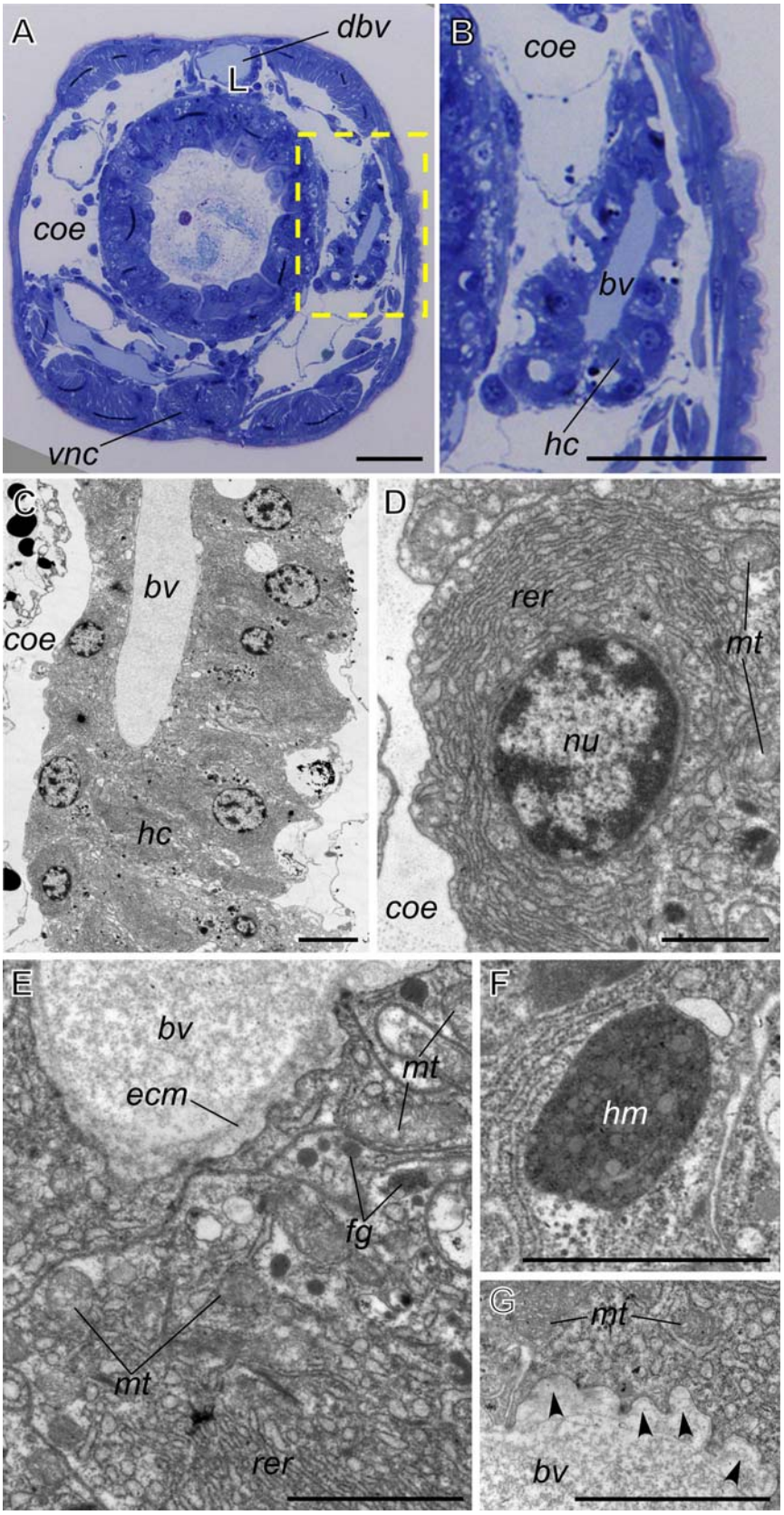

A
dbv
L
coe
vnc
B
coe
bv
hc
C
bv
coe
hc
D
rer
mt
nu
coe
E
bv
ecm
mt
fg
mt
rer
F
hm
G
mt
bv

Figure 4

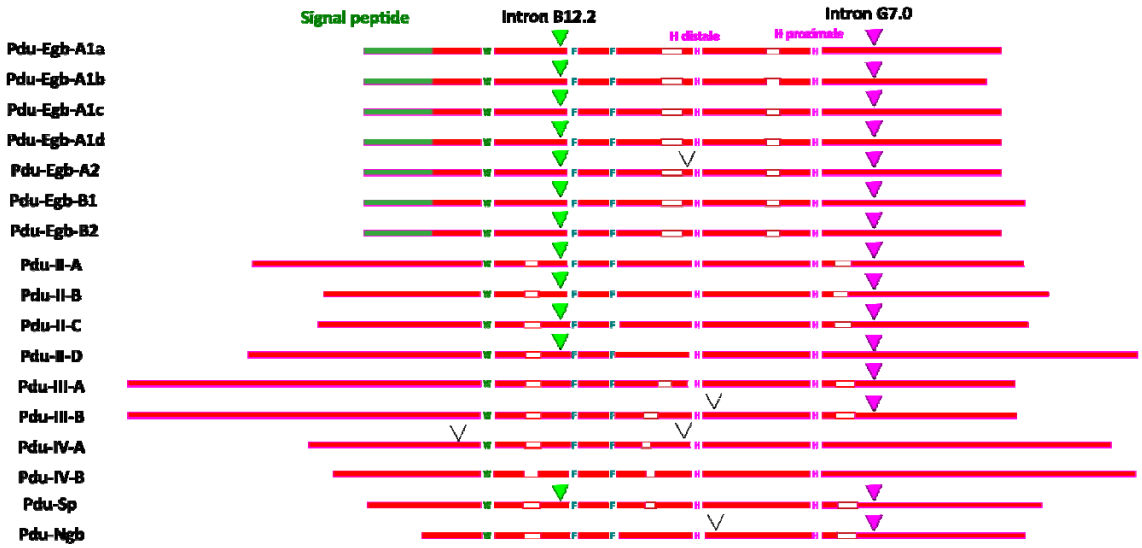

Signal peptide
Intron B12.2
Intron G7.0
H distale
H proximale
Pdu-Egb-A1a
Pdu-Egb-A1b
Pdu-Egb-A1c
Pdu-Egb-A1d
Pdu-Egb-A2
Pdu-Egb-B1
Pdu-Egb-B2
Pdu-II-A
Pdu-II-B
Pdu-II-C
Pdu-II-D
Pdu-III-A
Pdu-III-B
Pdu-IV-A
Pdu-IV-B
Pdu-Sp
Pdu-Ngb

Figure 5

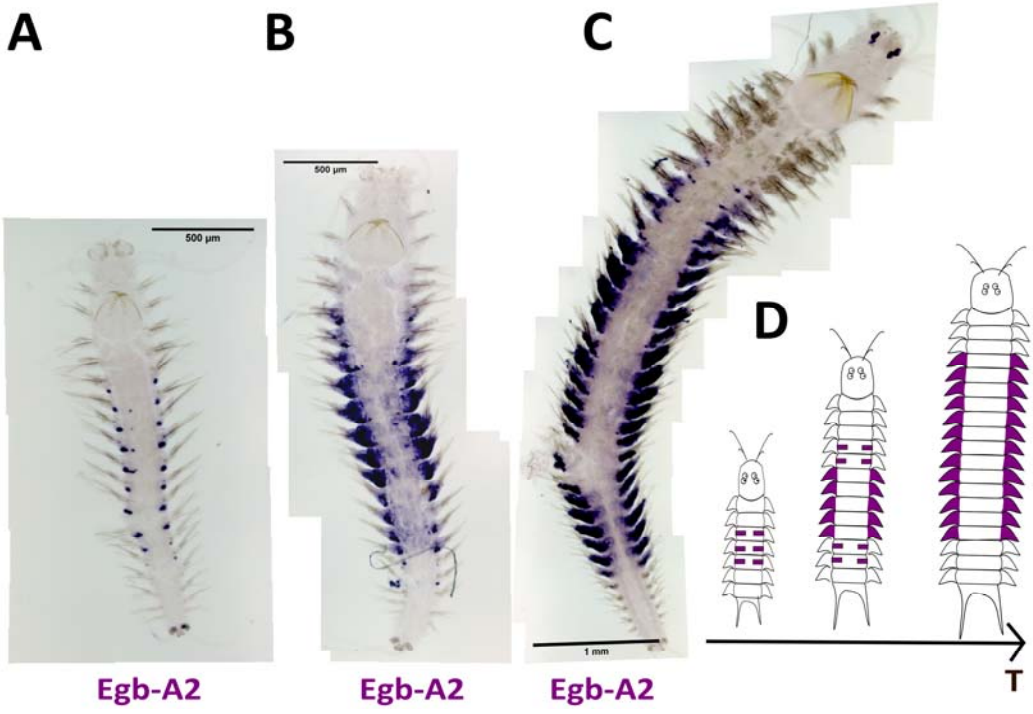


Figure 6

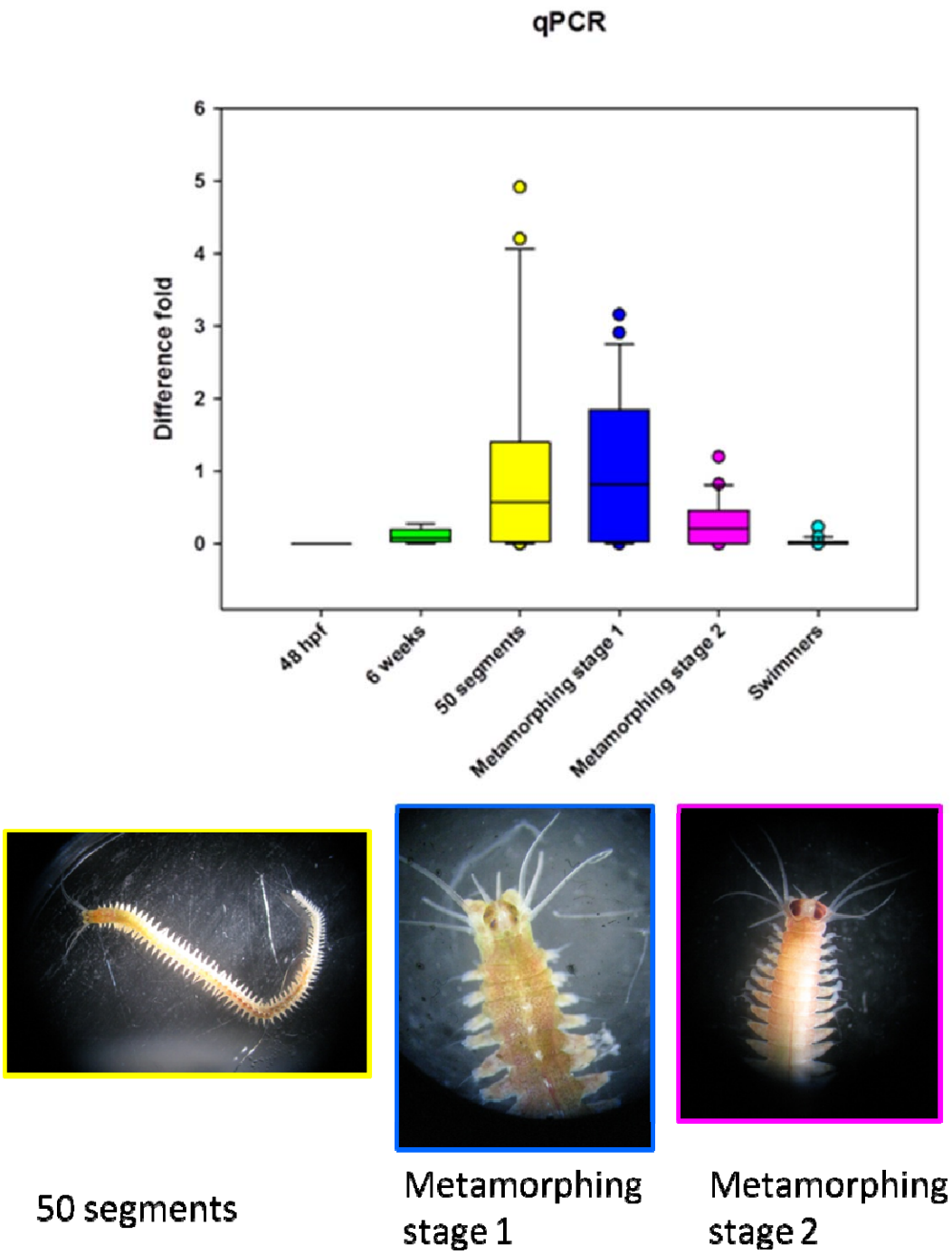
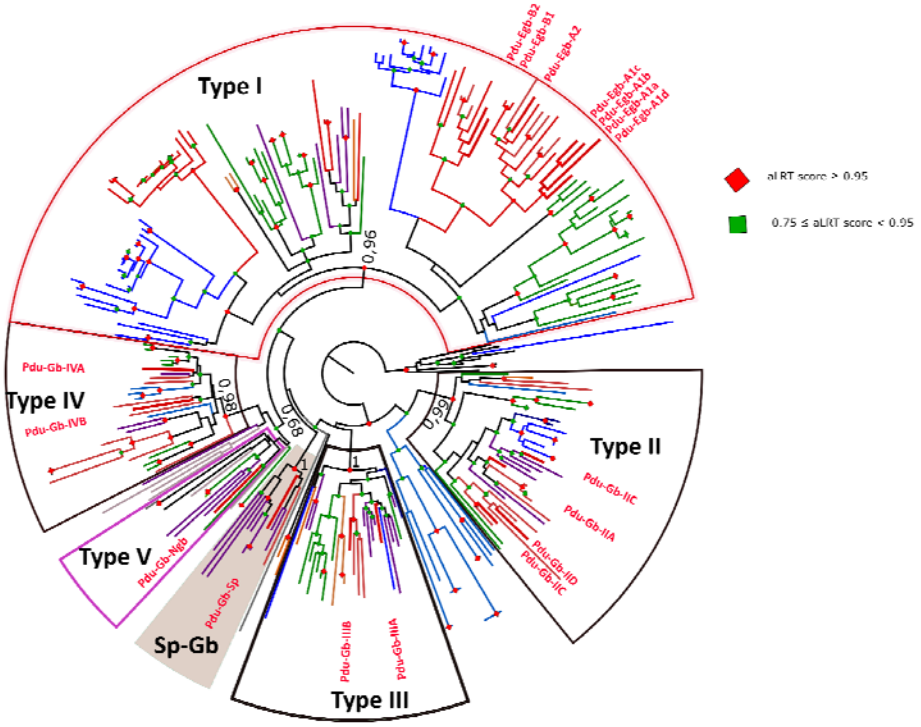
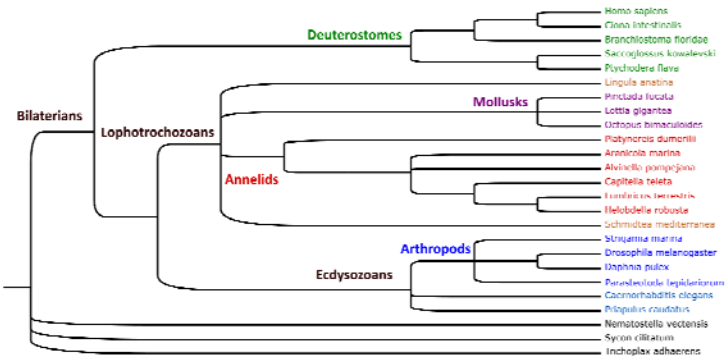

qPCR
Difference fold
6
5
4
3
2
1
0
48 hpf
6 weeks
50 segments
Metamorphing stage 1
Metamorphing stage 2
Swimmers
50 segments
Metamorphing stage 1
Metamorphing stage 2

Figure 7

A



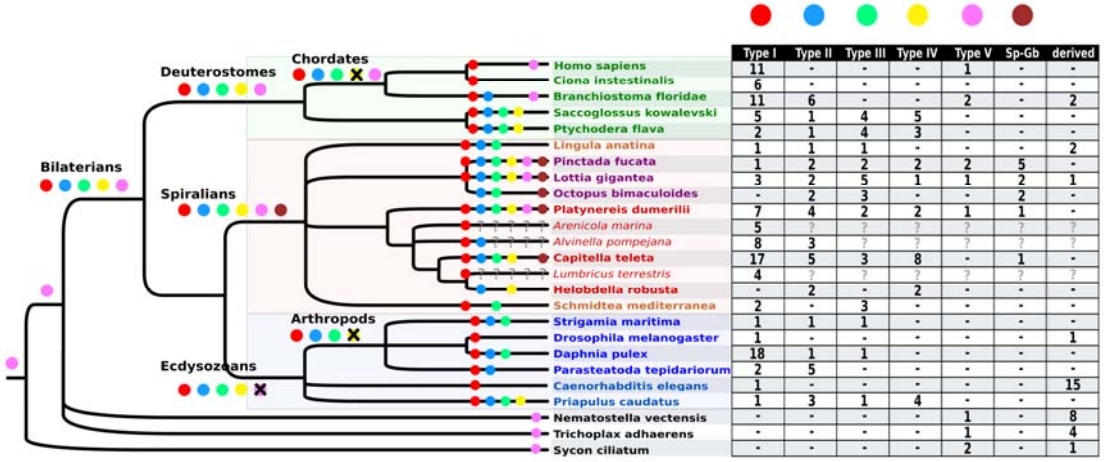
B



C

	Type I	Type II	Type III	Type IV	Type V	Sp-Gb
Max Likelihood LG aLRT	0,96	0,99	1	0,98	0,68	1
Bayesian LG posterior clade probability	0,96	0,97	1	0,82	not found	1
Bayesian WAG posterior clade probability	0,55	0,89	1	0,99	not found	1

Figure 8



	Type I	Type II	Type III	Type IV	Type V	Sp-Gb	derived
Homo sapiens	11	-	-	-	1	-	-
Ciona instestinalis	6	-	-	-	-	-	-
Branchiostoma floridae	11	6	-	-	2	-	2
Saccoglossus kowalevski	5	1	4	5	-	-	-
Ptychodera flava	2	1	4	3	-	-	-
Lingula anatina	1	1	1	-	-	-	2
Pinctada fucata	1	2	2	2	2	5	-
Lottia gigantea	3	2	5	1	1	2	1
Octopus bimaculoides	-	2	3	-	-	2	-
Platynereis dumerilii	7	4	2	2	1	1	-
Arenicola marina	5	?	?	?	?	?	?
Alvinella pompejana	8	3	?	?	?	?	?
Capitella teleta	17	5	3	8	-	1	-
Lumbricus terrestris	4	?	?	?	?	?	?
Helobdella robusta	-	2	-	2	-	-	-
Schmidtea mediterranea	2	-	3	-	-	-	-
Strigamia maritima	1	1	1	-	-	-	-
Drosophila melanogaster	1	-	-	-	-	-	1
Daphnia pulex	18	1	1	-	-	-	-
Parasteatoda tepidariorum	2	5	-	-	-	-	-
Caenorhabditis elegans	1	-	-	-	-	-	15
Priapulus caudatus	1	3	1	4	-	-	-
Nematostella vectensis	-	-	-	-	1	-	8
Trichoplax adhaerens	-	-	-	-	1	-	4
Sycon ciliatum	-	-	-	-	2	-	1

Figure 9