Novel insights on obligate symbiont lifestyle and adaptation to chemosynthetic environment as revealed by the giant tubeworm genome

André Luiz de Oliveira[1], Jessica Mitchell[2], Peter Girguis[2], Monika Bright[1]

[1] Department of Functional and Evolutionary Ecology, University of Vienna, Austria

[2] Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

Abstract

The mutualism between the giant tubeworm *Riftia pachyptila* and its endosymbiont *Candidatus* Endoriftia persephone has been extensively researched over the past 40 years. However, the lack of the host whole genome information has impeded the full comprehension of the genotype/phenotype interface in *Riftia*. Here we described the high-quality draft genome of *Riftia*, its complete mitogenome, and tissue-specific transcriptomic data. The *Riftia* genome presents signs of reductive evolution, with gene family contractions exceeding expansions. Expanded gene families are related to sulphur metabolism, detoxification, anti-oxidative stress, oxygen transport, immune system, and lysosomal digestion, reflecting evolutionary adaptations to the vent environment and endosymbiosis. Despite the derived body plan, the developmental gene repertoire in the gutless tubeworm is extremely conserved with the presence of a near intact and complete Hox cluster. Gene expression analyses establishes that the trophosome is a multi-functional organ marked by intracellular digestion of endosymbionts, storage of excretory products and haematopoietic functions. Overall, the plume and gonad tissues both in contact to the environment harbour highly expressed genes involved with cell cycle, programmed cell death, and immunity indicating a high cell turnover and defence mechanisms against pathogens. We posit that the innate immune system plays a more prominent role into the establishment of the symbiosis during the infection in the larval stage, rather than maintaining the symbiostasis in the trophosome. This genome bridges four decades of physiological research in *Riftia*, whilst simultaneously provides new insights into the development, whole organism functions and evolution in the giant tubeworm.

Main. The discovery of the giant tubeworm *Riftia pachyptila* Jones, 1981 at deep-sea hydrothermal vents on the Galapagos Spreading centre in 1977 (Corliss et al. 1979) has initiated the onset of a continuous torrent of studies (Childress and Fisher 1992; Nelson and Fisher 1995; Stewart and Cavanaugh 2006; Bright and Lallier 2010; Childress and Girguis 2011; Hilário et al. 2011). With its enormous size (Fisher et al. 1988a; Hessler et al. 1988; Shank et al. 1998), rapid cell proliferation (Pflugfelder et al. 2009), seemingly fast growth (Lutz et al. 1994; Lutz et al. 2001), but short life (Klose et al. 2015) one of the most puzzling findings was the lack of a digestive system in an animal with a highly unusual body plan (Jones 1981). Descriptions of mouth- and gutless pogonophoran relatives go back a century (Caullery 1914). The first vestimentiferans *Lamellibrachia barhami* Webb, 1969 and *L. luymesi* van der Land and Nørrevang, 1975 were described already a few years earlier than *Riftia.* However, it was the discovery of *Riftia,* thriving in an apparently poisonous hydrothermal vent environment, which sparked the discovery of the first-described chemosynthetic animal-microbe symbiosis (Cavanaugh et al. 1981); an association in which *Riftia*, without a mouth or a gut, relies on the sulphide oxidizing chemoautotrophic symbionts for nutrition (Cavanaugh et al. 1981; Felbeck 1981; Rau 1981a; Rau 1981b) (Arp and Childress 1981; Arp and Childress 1983).

Despite the fact that neither the animal host, nor the symbiont, nor the intact association are amenable to long-term cultivation, *Riftia* is easily one of the best studied deep-sea animals which have consistently led to major discoveries (reviewed by Bright and Lallier 2010). Crucial was the development of various devices to measure chemical and physical parameters directly in the deep sea to understand the abiotic conditions under which this tubeworm thrives at vigorous diffuse vent flow (Hessler et al. 1988; Shank et al. 1998; Luther et al. 2001; Le Bris et al. 2003; Mullineaux et al. 2003; Le Bris, Govenar, et al. 2006; Le Bris, Rodier, et al. 2006). Unprecedented and equally important was the development of high-pressure flow-through systems to simulate in situ conditions in the lab (Quetin and Childress 1980; Girguis et al. 2000).There has been probably no deep-sea animal with more resourceful experimental approaches applied *in situ* and *ex situ* than *Riftia,* e.g. catheterised tubeworms under flow-through pressure (Felbeck and Turner 1995), artificial insemination and developmental studies under pressure (Marsh et al. 2001), predation experiments with mesh cages *in situ* (Micheli et al. 2002), hydraulically actuated collection devices of tubeworm aggregations (Hunt et al. 2004; Govenar et al. 2005), artificial plastic tube deployments (Govenar and Fisher 2007), pressurized experiments (Goffredi et al. 1997; Shillito et al. 1999; Girguis et al. 2000; Girguis et al. 2002), and finally, various *in situ* settlement devices for tubeworm larvae (Mullineaux et al. 2000; Nussbaumer et al. 2006; Mullineaux et al. 2020). These innovative experiments associated with four decades of research taught us about many aspects of *Riftia's* evolution and biology.

After many microanatomical studies accompanied by heated, highly controversial phylogenetic discussions, the question of who the closest relatives of *Riftia* are was ultimately solved by traditional cladistic and novel molecular analyses (Fauchald and Rouse 1997; McHugh 1997; Halanych et al. 2001; Rouse 2001; Schulze 2002). They showed that vestimentiferans are lophotrochozoan polychaetae worms within Annelida (Fig. 1A) (Polychaeta, Siboglinidae, Vestimentifera) (Pleijel et al. 2009). Similar to many other polychaetes, *Riftia* is gonochoristic with internal fertilization and undergoes a biphasic life cycle with a pelagic phase including indirect development through spiral cleavage and a trochophore larvae (Marsh et al. 2001). The benthic phase is marked by the uptake of the symbiont into the metatrochophore larvae and growth into an adult, which completely reduces its mouth, gut, and anus. Instead, a unique mesodermal nutritional organ, the trophosome, functionally replaces the digestive system (Nussbaumer et al. 2006; Bright et al. 2013). The adult body is organized into four distinct regions, the obturacular region, the vestimentum, the trunk and the opisthosoma (Fig. 1B). The anterior obturacular region of the animal projects a vascularised branchial plume, which is responsible for the sequestration of nutrients and gas exchange, followed by the vestimentum, a muscular head region enclosing the heart, brain, the excretory organ, and the gonopores. The trunk region, the single elongated first segment, harbours the trophosome and the gonads. The posterior part, the opisthosoma, contains

a typical segmented annelid region with serially arranged chaetae (Bright et al. 2013). It is so far unknown how this unusual body plan lacking the entire digestive system is reflected in their developmental genes and signalling pathways. Gutless parasitic tapeworms, e.g. have lost many developmental genes including all ParaHox genes (Tsai et al. 2013).

The trophosome of *Riftia*, a soft multi-lobed and highly vascular tissue, houses a polyclonal endosymbiotic population dominated by one genotype of *Candidatus* Endoriftia persephone, a chemoautotrophic gammaproteobacteria (Robidart et al. 2008; Gardebrecht et al. 2012; Polzin et al. 2019) that oxidizes sulphur compounds via oxygen and nitrate and, in turn, harnesses that energy to fix dissolved inorganic carbon (or DIC, which includes carbon dioxide and bicarbonate) to organic matter. Briefly, the trophosome is far removed from the external environment, so the host presumably provides all of the inorganic nutrients to the symbionts. This primarily occurs via the highly vascular brachial plume, which takes up oxygen and hydrogen sulphide (H_2S) from the external environment and transports these to the trophosome via a complex and unique complement of haemoglobins (Arp and Childress 1981; Arp and Childress 1983; Arp et al. 1987; Zal et al. 1996; Zal et al. 1997; Bailly et al. 2002; Flores et al. 2005). DIC is also taken up by *Riftia*, which is unusual as carbon dioxide is an animal respiratory waste product. However, in this case the worm must provide additional DIC to the symbionts for *net* carbon fixation, and does so by accumulating DIC in the blood (e.g. Goffredi et al. 1997; Goffredi et al. 1999). Moreover, physiological studies have shown that *Riftia* also takes up nitrate (also unusual for an animal), and in turn the symbionts reduce it to organic nitrogen (e.g. Hentschel et al. 1993; Girguis et al, 2000). In return, the host is nourished through the symbiont releasing organic matter and symbiont digestion, which occurs prior to bacteriocyte death in the periphery of the trophosome lobules (Felbeck 1985; Hand 1987; Felbeck and Jarchow 1998; Bright et al. 2000; Hinzke et al. 2019). Despite four decades of research, key questions about trophosome function remain, including but not limited to A) which of the two nutritional modes is more important (organic matter release or symbiont digestion; Bright et al. 2000) and B) the mechanisms that underlie organic nitrogen synthesis and distribution between the symbionts and the host.

Despite the highly derived annelid body plan, symbiotic lifestyle, and over forty years of extensive physiological research, whole genome information of *Riftia* has been lacking. Here, we generated a high-quality genome draft and distinct tissue-specific transcriptomes of the giant gutless tubeworm *Riftia*. By analysing the genome and transcriptomes of *Riftia* in a comparative framework, we highlight many evolutionary adaptations related to the obligate symbiotic lifestyle and survival in the deep-sea hydrothermal vent environment. The *Riftia* genome, together with a transcriptome and proteome study (Hinzke et al. 2019), a transcriptome study on the close relative *Ridgeia piscesae* (Nyholm et al. 2012), the genomic resources available for another close relative *Lamellibrachia luymesi* (Li et al. 2019) (short *Lamellibrachia*), and an extensive body of research

broadens our understanding of one of the most conspicuous models for host-symbiont interaction and of the biology of Vestimentifera. Most importantly, we show that the developmental gene repertoire is conserved, and that besides the well-known nutritional aspect of the trophosome, its mesodermal origin brought an inherited suite of functions such as, haematopoiesis, endosomal digestion of endosymbionts, and storage of excretory products likely adapted to serve host-symbiont physiological interactions. While the innate immune system apparently is little upregulated in the presence of the symbiont, it is highly active in the remaining body directly exposed, or connected through openings to, to the environment.

Results and discussion

***Riftia* represents the most complete annelid genome to date including a complete mitogenome.** To assess the whole genome content of the giant tubeworm (Jones 1981), we sequenced a single individual from the hydrothermal vent site Tica, East Pacific Rise 9° 50' N region, with ~87-fold coverage using Pacific Biosciences Sequel system (Supplementary Figures 1-3; Supplementary Table 1). We found the haploid genome size (560,7Mb with a N50 length of ~2,8Mb) to be smaller than previous genome-size estimates (Bonnivard et al. 2009) (Table 1-Supplementary Figure 4). The *Riftia* GC value is 40.49%, and the repeat content accounts for 29.99% of the total length of the genome with most of the repetitive landscape dominated by interspersed and unclassified lineage-specific elements (35.2%) (Supplementary Figure 5). After genome post-processing, we identified a total of 25,984 protein coding genes with homologue, transcriptome, *ab-initio,* and gene expression evidence. The BUSCO4 (Simão et al. 2015) genome completeness score is 99.37%. These numbers render *Riftia* the most complete annelid genome to date (Simakov et al. 2013; Li et al. 2019; Martín-Durán et al. 2020) (Supplementary Figure 6).

The complete reconstruction of siboglinid mitochondrial genomes including the AT-rich control region has been notoriously difficult (Li et al. 2015). In this case, we were able to obtain it due to deep long read sequencing. The 15,406 bp circular mitochondrial genome contains all expected 13 coding sequence genes, two ribosomal RNA genes and the 22 tRNAs, typical of bilaterian mitogenomes (Fig. 1C – Supplementary Figure 7) (Boore 1999). In contrast to two other *Riftia* reference mitogenomes (Jennings and Halanych 2005; Li et al. 2015), we recovered the full control region (D-loop), yielding a mitochondrial genome longer than those previously reported. The gene order and the number of genes are conserved among all three *Riftia* and other siboglinids reference mitogenomes, though there are size differences that are most likely due to the incomplete nature of previously published genomes.

The developmental gene repertoire in gutless *Riftia* is conserved. Because of the lack of molecular information on the development of cell types and the evolution of the vestimentiferan body plan, we identified and annotated a suite of key developmental genes and signalling pathway-

related genes in the giant tubeworm genome. We found that key genes involved in the development of the digestive tract in metazoans (Hejnol and Martín-Durán 2015; Nielsen et al. 2018), such as *goosecoid, brachyury*, *foxA* and all three ParaHox genes, *xlox*, *cdx* and *gsx*, present in the *Riftia* and *Lamellibrachia* genomes (Supplementary Figure 8-10). The conservation of these genes in vestimentiferans is apparently not only crucial for developmental processes but also serves the microphagous nutrition in settled larvae until nourishment by the symbionts takes over in juveniles (Nussbaumer et al. 2006).

The Hox cluster (~578kb in size – Fig. 2A), homeodomain-containing transcription factors with roles in anterior-posterior axial identity in metazoans (Pearson et al. 2005; Duboule 2007), is nearly intact and complete in the giant tubeworm genome (Supplementary Figures 8-9, Supplementary note 2). We did not identify *hox7* in *Riftia*, indicating a secondary loss of this gene in the giant tubeworm, a pattern also observed in other lophotrochozoan representatives such as phoronids (Luo et al. 2018) and bivalves (Gerdol et al. 2015; Calcino et al. 2019). *Hox7*, *lox*2 and *lox5* are missing from *Lamellibrachia* genome suggesting a possible loss of the central Hox cluster elements (Fig. 2B) (Li et al. 2019). The Hox-like elements, homeotic genes equally important for body plan specification and developmental processes, *gbx*, *evx*, *mox*, *mnx, en* and *dlx* were also found in the giant tubeworm genome. *Engrailed* (*En*) and *even-skipped* (*Evx*) have 2 and 4 copies, respectively (Supplementary Figure 8).

Few signalling pathways are required to control cell-to-cell interactions and produce the plethora of cell types and tissues in Metazoa (Pires-daSilva and Sommer 2003) among them TGFβ, Wnt, Notch and Hedgehog (Moustakas and Heldin 2009; Ingham et al. 2011; Holstein 2012; Massagué 2012; Niehrs 2012; Gazave et al. 2017) (Supplementary Figures 11-15) The *Riftia* genome contains 14 TGFβ genes, including *nodal* and its antagonist *lefty*, the latter previously assumed to be a deuterostome innovation (Simakov et al. 2015). *Notch* and *hedgehog* are present as single copy genes in the *Riftia* genome as well as in *Lamellibrachia*, however, the *notch* receptor *jagged* is missing from both tubeworms. *Jagged* is present in the annelids *Capitella teleta*, *Helobdella robusta* and *Platynereis dumerilii* (Gazave et al. 2017), suggesting a secondary loss in Vestimentifera. Patched and dispatched genes, membrane receptors for the hedgehog ligand (Ingham et al. 2011) are present in *Riftia* with the dispatched genes expanded in vestimentiferans. In *Riftia*, we identified the 12 expected Wnt ligands (*Wnt3* has been shown to be lost in the Protostomia lineage) and their receptors frizzled, smoothened and sFRP (Holstein 2012). There is a genetic linkage of *Wnt1*,*6, 9* and 10 in *Riftia* akin to the gastropod *Lottia gigantea* and the fruit fly *Drosophila melanogaster* (Cho et al. 2010), reaffirming the ancient protostomian ancestral conserved linkage. The remaining eight Wnt genes in *Riftia* are disorganized on eight different scaffolds. Overall, despite the highly derived body plan, *Riftia* presents a deep conservation of the developmental gene toolkit akin to many distinct bilaterian animals.

The *Riftia* genome is characterised by reductive evolution. Multiple lines of evidence point to a relatively small genome, with gene family contractions exceeding expansions in *Riftia*, indicative of reductive evolution. The giant tubeworm genome is ~168Mb smaller than *Lamellibrachia,* its relative from cold hydrocarbon seeps whose genome is ~688 Mb with a N50 of 373kb (Li et al. 2019)). The difference can be attributed to the increased number of repeat elements and protein coding genes in the cold seep tubeworm (38,998 gene models and repetitive content of 36.92%; Supplementary Note 1; Supplementary Figure 16).

To identify clusters of orthologous genes shared among the two vestimentiferans *Riftia* and *Lamellibrachia,* the polychaete *Capitella teleta* (herein called *Capitella*), and the clitellid *Helobdella robusta* (herein called *Helobdella*) (Simakov et al. 2013), we employed tree-based orthology inferences (Emms and Kelly 2019). The annelid core genome, the collection of orthogroups shared among the four annelids, contains 6,349 cluster of orthologous genes. Less than half of them represent the vestimentiferan core genome (2,883 orthogroups) shared between *Riftia* and *Lamellibrachia*. Interestingly, the number of shared orthogroups between the *Riftia-Helobdella* (17) and -*Capitella* (116) pairs are smaller than those between *Lamellibrachia-Helobdella* (89) and -*Capitella* (349) pairs, indicating that *Riftia* contains a more derived gene repertoire than its close relative *Lamellibrachia*.

To further investigate the important processes of gene losses and gains, known to shape animal evolution (Fernández and Gabaldón 2020; Guijarro-Clarke et al. 2020), identify expanded protein domains, taxonomically restricted genes, and positively selected genes in *Riftia*, we employed multi-level comparative approaches involving statistical analysis, taxon rich orthology inferences (N=36), and sensitive similarity searches (Supplementary Table 2; Supplementary Note 3). The *Riftia* genome shows a net reduction of gene numbers with only 734 expanded but 1,897 contracted gene families, whereas the evolutionary history of *Lamellibrachia* is characterised by gene gains. Notably, the average expansion value of gene families in *Riftia* is the lowest among the four selected annelids herein analysed (Supplementary Figure 17). A total of 8,629 lineage-specific genes (~33.21% of the total) were identified in *Riftia*. Compared to the giant tubeworm, *Lamellibrachia* contains more lineage-specific genes (10,262 – 26.31% of the total).

The contracted gene families are not restricted to any specific biological process, as revealed by our gene ontology (GO) term enrichment analysis (N=18) (Supplementary Figure 18). Rather it appears that the giant tubeworm genome is undergoing a broad reduction in gene content (i.e., reductive evolution). Among the contracted gene families are genes controlling the transcriptional machineries (Supplementary Figure 19; Supplementary Table 3). Transcription factors (TFs) are proteins with sequence specific DNA-binding domains that control gene transcription and tissue

identity (Schmitz et al. 2016). To gain understanding into the repertoire of TFs in *Riftia*, we annotated and classified genes in the tubeworm genome present in five major groups of TFs (bzip, homeobox, nuclear factor, bHLH and zinc-finger) with sensitive similarity searches. The giant tubeworm presents the lowest number of TFs within the analysed annelids (414), supporting our gene family analysis (discussed below). The cold-seep tubeworm genome contains a similar complement size as *Riftia* (423), with *Capitella* (551) and *Helobdella* (568) presenting a higher number of TF genes, comparatively. These results point to pervasive TF losses in the Vestimentifera lineage (Supplementary note 3).

Expanded and lineage specific gene families in *Riftia.* Despite overall genome reduction, the *Riftia* genome exhibits, there is also a variety of expanded gene families (Supplementary note 3; Supplementary Figure 20; Supplementary Table 4). These expanded families are enriched with GO terms associated with sulphur metabolism, membrane transport and detoxification of xenobiotic, e.g. foreign substances (xenobiotic transmembrane transporter activity, galactosylceramide sulfotransferase activity, CoA-transferase activity) (Gamage et al. 2006), detoxification of hydrogen peroxide as anti-oxidative stress response (glutathione catabolic and biosynthetic processes) (Espinosa-Diez et al. 2015), neurotransmitter- and ion channel-related functions (sodium symporter activity), oxygen transport (oxygen binding, haemoglobin complex), endosomal degradation (lysozyme activity), and secretion of chitin (chitin binding, protein glycosylation) (discussed with more details later).

Genes involved in the production of extracellular components of vestimentiferans such as the cuticle and the basal matrixes as well as the tube and chaetae (Gardiner and Jones 1994) were found in expanded families of *Riftia* (as well as *Lamellibrachia*), some of which are specific to either *Riftia* or *Lamellibrachia* (Supplementary note 3; Supplementary Figures 21-24; Supplementary Tables 5-6). The *Riftia* genome contains expanded protein domains related to several high-molecular mass proteins such as laminin, nidogen and collagen. These proteins are part of extracellular matrix secreted basally from epithelia, also known to regulate cellular activity and growth in other animals (Timpl and Brown 1996). In *Riftia*, extensive short collagen fibres are found below the epidermis, extending between muscles cells, and building the matrix of the obturaculum. In addition, long helically arranged collagen fibres are the main component of the cuticle apically secreted from the epidermis (Gardiner and Jones 1994). Importantly, many genes involved in chitin production, a biopolymer part of the hard protective tube secreted from pyriform glands of the vestimentum, trunk, body wall, and opisthosoma (Gardiner and Jones 1993), are taxonomically restricted to the *Riftia* lineage. Expectedly, we identified in the vestimentum and body wall tissues of *Riftia* several tissue-specific genes (TSGs) involved in the chitin metabolism responsible for the tube production as well as dissolution (Supplementary Figures 25-26). Although the specific gland type responsible for dissolution of tube material has yet to be identified, we suggest that in *Riftia*

the straight tube, that can reach up to three metres in length and five centimetres in diameter (Gaill and Hunt 1986; Grassle 1987; Fisher et al. 1988b), can only widen in diameter to accommodate growth of the worm when tube material is dissolved and newly secreted, which agrees with the distribution of many TSG involved with tube biosynthesis. Overall, our findings of these expanding gene families as well as gene expression patterns underline the importance of chitin in *Riftia,* which is considered one of the fasted growing invertebrates (Lutz et al. 1994. Lutz et al. 2001). In order to achieve these high growth rates, *Riftia* needs to both digest and remodulate its own tube with astonishing speed.

Furthermore, the multi-level comparative analyses revealed an enrichment of GO terms in the lineage specific *Riftia* genes involved with the control of the chromosome condensation and nucleosome assembly, and positively selected genes related to tumour suppression (*PIN2/TERF1*-interacting telomerase inhibitor) and transcription initiation (*TFIIB*- and *-D*) (Roeder 1996; Zhou and Lu 2001). Interestingly, in *Lamellibrachia smad4* (Li et al. 2019), which is a tumour suppressor and transcription factor, is under positive selection, suggesting a common vestimentiferan evolutionary adaption responsible for controlling the chromatin-remodelling events and the extraordinarily cell proliferation rates in these two tubeworms (Supplementary Table 7; Supplementary note 3) (Pflugfelder et al. 2009).

The protein annotation of the rapidly evolving expanded gene families in *Riftia* identified members of the complement system involved in innate immunity and self-, non-self-recognition (sushi repeat domain-containing protein) (Kirkitadze and Barlow 2001). *Riftia* contains the greatest number of sushi-domain containing proteins among lophotrochozoans, presenting a total of 42 copies which are organised either in genomic clusters or dispersed as single elements throughout the genome (Supplementary Figure 4; Supplementary Figures 27). Of these, only 40 are shared with the cold seep tubeworm *Lamellibrachia*, pointing to a lineage-specific expansion at the base of Vestimentifera. Sushi genes, a common component of haemocytes (i.e., immune cells with phagocytic function (Pila et al. 2016), have been implicated in the mediation of the host-symbiont tolerance in the bobtail squid *Euprymna scolopes* – bioluminescent *Aliivibrio fischeri* association (McAnulty and Nyholm 2017). Although the rapid evolution of these proteins in *Riftia* and *Lamellibrachia* suggests similar evolutionary adaptations to the tubeworm/endosymbiont mutualism, the absence of any significant expressions in adult tissues rather point to their involvement in recognition of the symbiont during transmission in the larval stage or to potential pathogen recognition upregulated upon exposure.

Substrate transport for energy conservation and biosynthesis is supported by lineage-specific adaptations and parallel evolutionary events in *Riftia*. As an adaptation to the sulphidic vent environment, and in support of a symbiotic lifestyle, the respiratory pigments in

Riftia, and other vestimentiferans such as *Lamellibrachia*, bind non-competitively and reversibly to oxygen and sulphide, simultaneously providing a key substrate for chemosynthesis by the symbionts while also averting the sulphidic inhibition of the hosts' mitochondrial oxidative chain reactions (Arp and Childress 1983; Terwilliger et al. 1985). Our previous gene family evolution analysis identified an expansion of haemoglobins in the giant tubeworm genome compared to other non-vestimentiferan lophotrochozoans (Supplementary Figure 21; Supplementary Table 6; Fig. 3A). Additionally, a recent genome study found a massive expansion of β1-haemoglobin in the cold seep tubeworm *Lamellibrachia* (Li et al. 2019). To gain better insights in the evolution of Hb and linker genes in the Vestimentifera lineage, we employed thorough comparative genomics, phylogenetics, domain composition and gene quantification analyses. The genomic arrangement of the giant tubeworm Hb genes indicates that they were originated through a series of tandem duplications, totalling seven distinct genomic clusters (Fig. 3B). We annotated 26 extracellular Hbs and six linker genes in the *Riftia* genome (Supplementary Figures 28-29; Supplementary Note 4). Twenty-two Hb genes were phylogenetically placed in the β1-Hb group, surpassing previous estimates of the β1-Hb complement in the giant tubeworm (Bailly et al. 2002; Sanchez et al. 2007; Hinzke et al. 2019). α2- and β2-Hbs are found as single copy genes, whereas α1-Hb group contains two paralogous genes. The sulphide-binding ability of the *Riftia* Hbs is associated with the occurrence of free cysteine residues in one α2 and one β2 Hb genes (Bailly et al. 2002), as well as the formation of persulphide groups on linker chains (Zal et al. 1998; Bailly et al. 2002). Our results show that seven additional paralogous genes belonging to the β1-Hb group contain the putative free-cysteine residues, which were confirmed through multiple sequence alignments and homology model generation (Supplementary Figures 30-31, Supplementary note 4). Additionally, it has been hypothesized that zinc ions, rather than free-cysteine residues, are responsible for the H2S binding and transport on vestimentiferan α2 chains (Flores et al. 2005). We identified the three conserved histidine residues (B12, B16 and G9), predicted to bind zinc moieties, in *Riftia* Hb genes. However, we observed variations within the *Lamellibrachia* α2 genes. A broader comparison of α2-Hb genes belonging to different annelid taxa challenged the hypothesis of zinc sulphide-binding mechanisms for H2S in siboglinids and vestimentiferans (Li et al. 2019). Our results, solely based on the conservation of histidine residues, corroborate Flores et al. (2005) hypothesis that zinc residues may be involved in the sequestration and transport of hydrogen sulphide at least on the giant tubeworm.

To investigate the gene expression dynamics of the newly and previously identified Hb paralogs in *Riftia*, we analysed published transcriptomes sampled from *Riftia's* trophosomes containing sulphur-rich to sulphur-depleted symbionts (Hinzke et al. 2019). Hb gene expression showed great variation, indicating a more specialised role of the Hbs according to the environmental chemical fluctuations in the unstable deep-vent ecosystem (Fig. 3C, Supplementary note 4; Supplementary Table 8). Taken together, these results suggest a more complex system coordinating oxygen-

sulphide sequestration and distribution in the giant tubeworm tissues. The Hb complement of *Riftia* and *Lamellibrachia* are similar and unique among annelids and lophotrochozoans, in respect to gene numbers and distribution, indicating a Vestimentifera synapomorphy.

As the endosymbionts require carbon dioxide (CO_2) for fixing inorganic carbon, the transport of CO_2 and the conversion of its alternative forms (e.g., bicarbonate; $HCO3^-$) is mediated by another class of enzymes, the carbonic anhydrases (CAs) (Shively et al. 1998; Cian et al. 2003). We found ten carbonic anhydrase genes in the *Riftia* genome, from which seven are tandemly arrayed in two genomic clusters (Supplementary Figure 32). A similar CA complement in *Riftia* (nine genes) was found in a recent study (Hinzke et al. 2019). To better understand the diversity of CA genes we analysed tissue-specific transcriptomes and found at least five CA genes are membrane bound with three of them moderately/highly expressed in the trophosome, indicating that $HCO3^-$ conversion to CO2 and diffusion across the bacteriocyte membrane might be a common process in the trophosome, as suggested previously (Sanchez et al. 2007; Bright and Lallier 2010; Hinzke et al. 2019). Tandem duplications and tissue-specific CA expression linked to the intracellular supply of CO_2 to endosymbionts have been recently reported in deep-sea bivalves (Ip et al. 2021), showing remarkable resemblance to our findings. Taken together, our results show that the transport of essential compounds to the chemoautotrophic endosymbionts and the maintenance of the mutualistic relationship is driven by lineage-specific and parallel evolutionary events.

Trait and gene loss is compensated by the endosymbionts. The loss of the digestive system requires nourishment through the symbiont. The mechanisms of carbon transfer between the endosymbiont and *Riftia* were shown to be through the fast release of fixed carbon from the symbiont and uptake into host tissue, as well as, through symbiont digestion prior death of the bacteriocytes (Felbeck 1985; Hand 1987; Felbeck and Jarchow 1998; Bright and Lallier 2010; Hinzke et al. 2019). We found corroborating evidence for the uptake of released organic carbon from the symbiont based on the enrichment of GO terms and tissue specificity of succinate-semialdehyde complex genes and nuclear-encoded proteins of the inner mitochondrial membranes (including the tricarboxylate mitochondrial carrier responsible for the transport of succinate) (Majd et al. 2018) in the trophosome (Supplementary Figure 33; Supplementary Table 9). These results suggest an increased movement of cytosolic succinate through the mitochondrial membrane, possibly increasing the ATP production via the oxidative metabolism. These findings corroborate previous findings and support the involvement of this molecule for nourishment in *Riftia* from its endosymbiont (Felbeck and Jarchow 1998).

Evidence of digestion was revealed with tissue-specific transcriptome analyses which allowed us to identify the genes involved in the successive stages of lysosomal-associated degradation of symbionts (Supplementary Figure 33; Supplementary Tables 8). The expression of genes

associated with endosomal activity, the expression of several lysosomal-associated hydrolases (Supplementary Figures 34-35), vacuolar ATPases, and small Ras-related GTPases (rab genes) (Supplementary Figure 36; Supplementary Note 5) in the trophosome, is indicative of lysosomal-associated degradation of symbionts, as suggested earlier in ultrastructural studies, which describe the presence of primary lysosomes and symbionts in different lytic stages in *Riftia* (Bright et al. 2000; Bright and Sorgo 2003; Hinzke et al. 2019). In addition, we detected specific genes in the trophosome which are associated with actin cytoskeleton dynamics (ARP2/3) known to be essential for endosomal dynamics (Kast and Dominguez 2017) (Supplementary Figure 33; Supplementary Table 9). Recent *de novo* tissue-specific transcriptomes and gene expression quantification of the host and symbiont support the digestive route of nutrition in *Riftia*/Endoriftia symbiosis (Hinzke et al., 2019).

Furthermore, genes involved in the transport of fatty acyl units into the mitochondrial matrix (e.g., carnitine/acylcarnitine carrier) (Indiveri et al. 2011), tricarboxylic acid (TCA) cycle, oxidative phosphorylation, antioxidant systems (i.e., superoxide dismutase II genes and methionine sulfoxide reductases) (Supplementary Figure 38), and key players of the fatty acid β-oxidation (Fig. 4A, B and C), showed tissue specificity in the trophosomal tissue (Supplementary Table 9; Supplementary note 5). Fatty acid β-oxidation is a central and deeply conserved energy-yielding process that fuels the TCA cycle and oxidative phosphorylation (Houten et al. 2016). As *Riftia* relies solely on its endosymbionts for sustenance, the metabolism of fatty acids in the trophosome is certainly linked to the bacterial digestion in this tissue, which is corroborated by a previous proteomic study (Hinzke et al. 2019). Altogether, the results point to different modes of nutrient transfer in the trophosome involving the translocation of released nutrients from symbiont to host through succinate, and the digestion of the symbionts by lysosomal enzymes followed by the degradation of fatty acids using the mitochondrial β-oxidation pathway.

To further explore the extent to which degree *Riftia* is dependent on its endosymbiont for nutrition, we screened the genome of giant tubeworm and selected annelids for key enzymes related to amino acid biosynthesis (Supplementary Table 10). We found that *Riftia*, together with cold-seep tubeworm *Lamellibrachia* and the parasitic leech *Helobdella*, lacks many key enzymes related to amino acid biosynthesis when compared to close free-living polychaete relative *Capitella* (Fig. 4D). Genes involved with amino acid biosynthesis are constitutively expressed across the tubeworm tissues, with enzymes related to arginine and glycine metabolism highly expressed in the trophosome (Supplementary Figure 39). These findings suggest that loss of key enzymes in mutualistic vestimentiferans as well as a parasitic leech may be due to the beneficial and parasitic relationships, respectively allowing for compensated gene loss, compared to free-living polychaetes.

Overall, endosomal-associated digestion of endosymbionts seems to be a hallmark of intracellular digestion accomplished in the mesodermal trophosome of vestimentiferans, such as *Riftia* and *Lamellibrachia* (Nussbaumer et al. 2006; Hinzke et al. 2019; Li et al. 2019) (see also below). This process serves the host nutrition as well as the control of the symbiont population density during host growth, known from many other symbioses (Angela E. Douglas 2010). In addition, the symbiont provides the host with released organic carbon. While we do not know yet which partner controls this mode of nutritional translocation, both the evolutionary adaptation of endosymbiont digestion in a mesodermal tissue as well as carbon release contributed to trait loss in one partner compensated by the other (Ellers et al. 2012), and consequently has made *Riftia* obligatorily associated with its symbiont.

Haematopoiesis operates in the trophosome of *Riftia*. Haematopoiesis, the production of blood cells and pigments is still a poorly understood process in vestimentiferans. The heart body, a mesodermal tissue in the dorsal blood vessel of vestimentiferans has been hypothesized to be the site of haemoglobin biosynthesis (Schulze 2002). The presence of many TSGs in the trophosome related to 5-aminolevulinate synthase, porphyrin metabolism, and metal ion binding indicate that this tissue harbours the enzymatic machinery necessary for haem biosynthesis. Haem is an integral part of haemoglobin molecules, which is synthesized in a seven multistep pathway that begins and ends in the mitochondrion. To fully characterise the haem biosynthesis pathway in the giant tubeworm, we screened the *Riftia* genome for the presence of the seven universal enzymes required to synthesize the haem (Supplementary Figure 40; Supplementary note 5). The giant tubeworm contains all the seven enzymes present as single copy in its genome with recognisable orthologs in the annelids *Lamellibrachia*, *Capitella* and *Helodbdella.* Gene expression analysis showed that the key enzymes present in the haem biosynthetic pathway are moderately/highly expressed in the trophosome, supporting the GO enrichment analysis (Supplementary Figure 40; Supplementary Figure 33). The haem biosynthesis in the trophosome is further corroborated by the presence of TSGs in this tissue related to phosphoserine aminotransferase and the mitochondrial coenzyme A transporter, which act as an important co-factor in the final step of haem synthesis and in the transport of coenzyme A into the mitochondria, respectively (Schneider et al. 2000; Fiermonte et al. 2009). These findings confirm the involvement of the mesodermal trophosome in haemoglobin metabolism and suggests that this organ is the site haematopoiesis. In the frenulate *Oligobrachia mashikoi*, the visceral mesoderm also strongly expresses globin subunits based on *in-situ* hybridisation and semi-quantitative RT-PCR (Nakahama et al. 2008), but in this siboglinid the visceral mesoderm is organized as simple peritoneum surrounding the endodermal trophosome (Southward 1993).

Excretory products are stored in the trophosome of *Riftia*. The finding of TSGs in the trophosome related to the biosynthesis of nitrogen-containing compounds and in the transport of

ornithine (Supplementary Tables 9 and 11) agrees with the high levels of uric acid and urease activity in this host tissue, as previously reported (Cian et al. 2000; Minic and Hervé 2003).To explore the nitrogen metabolism pathways in *Riftia*, we identified and quantified the gene expression of several enzymes related to the purineolytic/uricolytic, purine/pyrimidine, taurine, and the polyamine pathways, as well as the urea and ammonia cycles (Supplementary Figures 41-45; Supplementary Note 6).

Most of the identified genes are found as single copy in the giant tubeworm genome (Supplementary Note 6; Supplementary Figures 41-45), however, we identified the presence of three chromosomal clusters harbouring glutamine synthetase, cytoplasmatic taurocyamine kinase and xanthine dehydrogenase/oxidase genes (Fig 5A). These enzymes are involved in the ammonia, urea and uricolytic pathways, respectively. *Riftia* and *Lamellibrachia* contain the highest number of glutamine synthetase genes in the herein investigated lophotrochozoan genomes (Fig. 5B). Seven out of the nine glutamine synthetase genes present in *Riftia* belong to the group I and the remaining to the group II (Fig. 5C). Interestingly, only annelid orthologs are found to be phylogenetically close to the prokaryotic group I, indicating a secondary loss of this genes in the remaining lophotrochozoan lineages (see Supplementary Figure 43 for the expanded version of the phylogenetic tree). An expanded set of lengsin genes, an ancient class I glutamine synthetase family (Wyatt et al. 2006), is present in Vestimentifera (seven copies in *Riftia* and 13 in *Lamellibrachia*). Some members of the newly identified lengsins are also highly expressed in the trophosome, suggesting that these enzymes might play a role in mitigating toxicity of urea, ammonia, and other nitrogenous compounds (Wyatt et al. 2006).

The identification of five cytoplasmatic taurocyamine kinase genes, four organised in a genomic cluster (Fig. 5A), surpasses previous reports (in which only one cytoplasmatic gene was identified; Supplementary Figure 44) (Uda et al. 2005). In accordance with a recent study (Hinzke et al. 2019) and contrasting previous biochemicals investigations on the *de novo* pyrimidine and polyamine biosynthesis in *Riftia* (Minic et al. 2001; Minic and Hervé 2003), we identified the trifunctional CAD protein in the genome of the tubeworm reinforcing the notion that *Riftia* can catalyse the first steps of the pyrimidine synthesis independently of its endosymbionts (Supplementary Note 6). These results are not unforeseen, since during the aposymbiotic phase (i.e., *Riftia*'s fertilised egg until the settled larva) pyrimidine metabolism plays a fundamental role in the development and growth of the animal.

We also found that the key enzymes of the uricolytic pathway and urea cycle are highly active in the trophosome (Supplementary Figure 41; Supplementary Table 8). These results are consistent with enzymatic/light micrograph studies, which show that the trophosome contains high concentration of ammonia, urea, creatinine, and uric acid crystals in the periphery of the lobules

(Cian et al. 2000), and with a more recent transcriptomic and metaproteomic study (Hinzke et al. 2019). Surprisingly, we only identified in closed (De Oliveira, in review) and previous endosymbiont genome drafts the subunit-A of the urea transporter (*urtA*), with the four remaining subunits missing (*urtBCDE*) (Veaudor et al. 2019). Since all five subunits are required for a proper function of the urea transporter, these results challenge the idea of an active shuttle of urea from the host to the endosymbiont (Robidart et al. 2008).

Cell proliferation and cell death interplay with innate immunity in *Riftia*. To better understand how fast growth (Lutz et al. 1994) fuelled by high proliferation rates (Pflugfelder et al. 2009) and innate immunity act in *Riftia* in tissues exposed to the environment and in the endosymbiont-housing trophosome, we characterised the key molecular components, and their gene expressions, of important pathways related to cell cycle signalling (Supplementary Figure 46), Toll-like receptor/MyD88 (Supplementary Figures 47-49), as well as the apoptotic (Supplementary Figures 50-57) and autophagic (i.e. macroautophagy) cell death events (Supplementary Figures 58). The *Riftia* genome similar to other investigated lophotrochozoans and closely related annelids harbours all the key components of these conserved pathways (Supplementary Note 7) (G. Zhang et al. 2012; Sun et al. 2017; Luo et al. 2018; Li et al. 2019; Ip et al. 2021). We did not identify any extensive remodelling (i.e., gene family expansions and contractions) of the immune (with the exception of sushi genes) and programmed cell death components in the giant tubeworm genome, as shown to be important in the maintenance of host-symbiont interactions in deep-sea mussels and clams (Sun et al. 2017; Ip et al. 2021).

Overall, *Riftia's* gonad and plume tissues are highly active in cell proliferation and programmed cell death. Subject to potential pathogen infections through the gonopore opening and direct contact to the vent water (Jones 1981), respectively, these tissues show the entire suite of genes involved in the innate immunity recognition with TLRs, downstream cellular immune responses, as well as apoptosis, autophagy and endosomal-related genes. These results were additionally supported by the GO enrichment analyses in the female gonad and plume tissues (Supplementary Figures 59-60). The trophosome, in contrast, despite the remarkably high bacterial population density (Powell and Somero 1986; Bright and Sorgo 2003) does not show any striking upregulation of TLR for endosymbiont recognition, nor cell proliferation, nor programmed cell death pathways (at least not in the classical sense; see Hintzke et al., 2019). Instead, we found few moderately/highly expressed genes present in the immune system (*irak2* and 4, *tab1, tak1*, *mkk3/6*), cell cycle (cyclin A, B2, D2, *cdk4*), apoptotic (*cas2*, *cas8*, *birc8*), and autophagic (*becn1*, *atg2b-7-8-16*) pathways in the trophosome of *Riftia*. Interestingly, a previous study suggested that immune-related genes were significantly more expressed in the trophosome in relation to other symbiont-free tissues in the siboglinid *Ridgeia piscesae*, positing a more important role of the immune system in the host-endosymbiont homeostasis (Nyholm et al. 2012).

Few other individual components of the innate immune system, i.e., bactericidal permeability-increasing proteins and pattern recognition receptors, have been implicated in symbiont population control in tubeworms (Nyholm et al. 2012; Hinzke et al. 2019). However, based on our broad gene expression analyses, we argue that the host immune system does not play a major role in taming the endosymbiont population in the trophosome, as previously suggested (Hinzke et al. 2019). Furthermore, immunohistochemical and ultrastructural cell cycle analyses identified apoptotic and proliferative events in the trophosome (Pflugfelder et al. 2009), indicating that despite the overall low expression of gene markers related to these pathways described herein, these events occur in this tissue. In which extent these different pathways interact to shape the host/symbiont interactions and to maintain tissue homeostasis remains to be shown, however, it is clear that multiple and not mutually exclusive programmed cell-death, immune-related, and proliferative events (Supplementary note 7) are acting on the trophosome.

From phenotype to genotype and back

After 40 years of intensive research, we are now finally able to integrate the obtained genome and tissue-specific transcriptome information with the current body of knowledge on the phenotype to better understand the genotype-phenotype interplay in the giant tubeworm. The *Riftia pachyptila* genome is characterized by reductive evolution with broad gene family contractions exceeding gene family expansions. Compared to the close relative *Lamelibrachia luymesi* (*Lamellibrachia* live at longer-lived and less physiologically-taxing hydrocarbon seeps), *Riftia* exhibits a more derived gene repertoire for important traits related to symbiosis and the highly disturbed and stressful hydrothermal vent habitat they inhabit in the deep sea.

The mutualism between *Riftia* and its symbiont has not transited from individuality of symbiotic partners to a new integrated organism (Szathmáry and Smith 1995) because it lacks mutual dependency (Kiers and West 2015; West et al. 2015). *Riftia* is, in fact, one of the few examples known in which dependency is asymmetric with a facultative horizontally transmitted symbionts, which have the capacity to live with or without the host. The *Riftia* host, however, is obliged to partner with the symbiont or else they cannot thrive. Therefore, the host's fitness is strictly tied to the persistence of this association over ecological and evolutionary time scales. The genome data now clearly shows the peculiarities and divergencies in *Riftia's* genotype compared to closely related free-living annelids and other lophotrochozoans, as well as which evolutionary adaptations of the host genotype ensure the maintenance of the association.

We found that despite the drastic morphological remodelling during its early development leading to the mouth-, gutless adult animal, *Riftia* retained the highly conserved developmental gene repertoire present in other lophotrochozoans and distant related animals. These results can be

interpreted as counterintuitive considering that the adult body plan alone provides little unambiguous evidence of the vestimentiferan phylogenetic relationship. These animals were initially compared to deuterostomes (Caullery 1914) and considered related to hemichordates (Beklemishev 1944) as well as protostomes, but so unique that new phyla were erected to accommodate them (i.e., Pogonophora and Vestimentifera) (reviewed by (Rouse 2001; Pleijel et al. 2009)). The conservation of the developmental gene toolkit probably reflects the developmental constraints into the necessary to go step by step through deterministic stereotypic spiral cleavage and larval development (Nielsen 2004). Akin to other polychaetes, the endoderm is necessary not only later for feeding functions, also seen in the metatrochophore larvae prior symbiont infection (Nussbaumer et al. 2006), but also to develop most mesodermal tissue.

Combining the genomic information with tissue-specific transcriptomes allows us to hypothesize that the mesodermal trophosome (Nussbaumer et al. 2006; Bright et al. 2013) (Nussbaumer et al. 2006; Bright et al. 2013) is a multi-functional organ with ancestral inherited functions such as haematopoiesis. This trait, we hypothesize belongs to the functional repertoire known from mesodermal chloragogen (extravasal tissue surrounding the gut and blood vessels) derived from the visceral mesoderm in annelids like the trophosome in vestimentiferans (Nussbaumer et al. 2006; Bright et al. 2013). In fact, van der Land and Nørrevang suggested already in 1975, long before the symbionts were detected, that the trophosome in *Lammelibrachia luymesi* is the nutritive chloragogen tissue (Van der Land 1975). Although, overall knowledge is fragmentary it has been suggested that haematopoiesis in annelids is carried out by visceral as well as somatic mesoderm (Hartenstein 2006; Grigorian and Hartenstein 2013). In various polychaete species it was localized in particular in the (extravasal) chloragogen tissue, the (intravasal) heart body (Potswald 1969; Friedman and Weiss 1980; Braunbeck and Dales 1984; Fischer 1993) or the somatic peritoneum (Eckelbarger 1976). Our data unambiguously support the production of haemoglobin in the trophosome. Whether coelomocytes, known to be the immunocompetent cells of eucoelomates (Vetvicka and Sima 2009) including annelids (Dales 1964; Salzet et al. 2006; Cuvillier-Hot et al. 2014), and the haemocytes also develop from trophosomal tissue appears to be likely but remains to be verified.

The trophosome, however, further shows adaptations to new functions such as the well-known intracellular digestion through endosomal-like maturation of symbiosomes, as well as the processing of ammonia and storage of nitrogen waste analogous to the vertebrate liver. Most aquatic invertebrates, including annelids, are virtually ammoniotelic secreting ammonia (Larsen et al. 2011). Surprisingly, *Riftia* employs an additional ureotelic metabolism similar to terrestrial invertebrates and vertebrates converting toxic ammonia to urea/and or uric acid. Specifically, we found the entire set of genes for a complete urea cycle known to detoxify ammonia in the *Riftia* genome, with most of them upregulated in the trophosome. Therefore, we hypothesize that the

trophosome share similar functions to the liver of vertebrates: Instead of secreting nitrogenous waste products through kidneys like in vertebrates or nephridia in annelids, the trophosome was found to store large amounts of uric acid and urea. Uric acid and urea can be utilized as a bioavailable source of N via the catabolic arm of the urea cycle yielding NH_4^+ and CO_2. Given the lack of urea transporters in the symbiont's genome, and the presence of active ureases in the trophosome host tissue, this suggests that both the synthesis and breakdown of uric acid and urea is under host control.

What factor(s) might lead to the evolution of this physiological capacity to sequester and metabolize urea and uric acid? It has been shown in other symbioses that the exchange of bioavailable N between symbiotic partners plays an important role in recycling bioavailable N, such as in coral-dinoflagellate symbiosis that show an almost complete retention of bioavailable N (Tanaka et al. 2018). At many deep-sea vents, including those where *Riftia* thrive, bioavailable N is limited as ammonium and free amino acids are found in pM concentrations (Johnson et al. 1988). Moreover, *Riftia* are unable to ingest particulate matter so they cannot derive nitrogen from detritus. However, an abundant source of N is nitrate, which is found in deep seawater and can be reduced to ammonium by some microbes (Girguis et al. 2000). Previous studies (Hentschel et al. 1993; Girguis et al. 2000) found that *Riftia* take up nitrate from their environment, and the symbionts reduce nitrate to ammonium for symbiont and host growth and biosynthesis. However, the *Riftia* host's ability to produce urea means that if can sequester bioavailable N that is only available to the host. At first glance, limiting symbiont access to N might be considered a way to control symbiont growth, as seen in cnidarian-Symbiodiniaceae (Xiang et al. 2020). This latter scenario, however, seems unlikely as there is ample bioavailable N (in the form of ammonium) throughout the trophosome in both freshly collected and experimentally tested worms (De Cian et al. 2000; Girguis et al. 2000). Rather, it seems plausible that *Riftia*'s production of urea allows the host to store and sequester N in a stable, largely nontoxic form. Whether urea is mobilized and provided to the host and symbionts during time of low N availability has yet to be experimentally tested, but this physiological capacity is another example of the remarkable adaptions found within that host, which allow it to modulate the rapid environmental changes found at vents and continue to provide for its own and the symbionts' metabolic demands.

While the physiological and evolutionary aspects of tubeworm endosymbiosis have been sufficiently addressed over the past 40 years, the molecular mechanisms regulating host and symbiont interactions in siboglinids are still not fully understood. An immuno-centric view has been explored to explain the maintenance and regulation of the endosymbiont population in the giant tubeworm trophosome (Nyholm et al. 2012; Hinzke et al. 2019). Our results, contrary to the expectations, indicate that genes involved with the innate immune responses are downregulated in the trophosome (e.g., Toll-like receptor/MyD88) or in adult tubeworm tissues (e.g., sushi). These

results suggest that the innate immune system plays a more prominent role into the establishment of the symbiosis during the infection in the larval stage, rather than preservation of the mutualism during the juvenile/adult life cycle. The control of the endosymbiont population in the trophosome is mainly achieved by the upregulation of endosomal and lysosomal hydrolases resulting in the active digestion of the endosymbionts (a “mowing” process as described by Hinzke et al. 2019).

The giant tubeworm genome establishes a unique and unprecedent hallmark bridging more than four decades of physiological research in *Riftia*, whilst it simultaneously provides new insights into the development, whole organism function and evolution of one of the most studied models for metazoan-symbiont interaction. We envisage that the resources generated herein foster many hypothesis-driven research pointing towards a more complete understanding of the genotype/phenotype interface in the *Riftia* and closely related taxa.

Methods

A detailed methods section is available in Supplementary Material and methods and in Supplementary Figure 61. A brief overview of the bioinformatics pipeline follows below.

Biological material and sequencing

Riftia genome DNA was obtained from a piece of vestimentum tissue belonging to single worm collected at the hydrothermal vent site Tica, East Pacific Rise (Alvin dive 4839, 9º 50.398 N, 104º 17.506 W, 2514 m depth, 2016) (Supplementary Figures 1, 2). PacBio libraries were generated with Sequel technology using the purified *Riftia* DNA. Tissue-specific transcriptomes were obtained from two specimens collected at Guaymas Basin, one female from the vent site Rebecca’s Roost (SuBastian dive 231, 27º 0.645 N, 111º 24.418 W, 2012 m depth, 2019) and one male from a vent site close to Big Pagoda (SuBastian dive 233, 27º 0.823 N, 111º 24.663 W, 2015 m depth, 2019) (Supplementary Figures 1, 2). The eight stranded paired-end tissue-specific transcriptomes (2x150 pb) were sequenced using Illumina NovaSeq SP technology.

Genome assembly and processing

The five *Riftia* PacBio libraries were mapped against a custom database built with *Riftia* mitochondrial genome and its complete Endoriftia genome using minimap v2.17-r941 (Li 2018). Genome assembly was performed with canu v1.8 (Koren et al. 2017) with optimised parameters. Genome pre-processing, polishing, haplotig removal and contamination screening, was performed with arrow v2.3.3 (https://github.com/pacificbiosciences/genomicconsensus), purge_dups (https://github.com/dfguan/purge_dups) and blobtools v1.1.1 (Laetsch and Blaxter 2017), respectively. Mitochondrial and endosymbiont genome assemblies were executed with flye v2.5 (Kolmogorov et al. 2019). Annotation of the mitochondrial genome was performed with MITOS2 and GeSeq (Bernt et al. 2013; Tillich et al. 2017).

691

692 **Transcriptome assembly and processing**

693 The removal of adapter sequences and quality filtering of reads from the raw transcriptome
694 databases was performed with bbduk v38.42 (https://sourceforge.net/projects/bbmap/). *De novo*
695 and genome-guided transcriptome assemblies were performed with transabyss v.2.0.1 and STAR
696 v2.7.1a – Stringtie v2.0.6, respectively (Robertson et al. 2010; Dobin et al. 2013; Kovaka et al.
697 2019). Possible Endoriftia contamination was removed from the transcriptomes using blastn
698 v2.8.1+ (Camacho et al. 2009). A global *de novo* transcriptome was generated with corset and
699 Lace (https://github.com/Oshlack/Lace) (Davidson and Oshlack 2014).

700

701 **Gene prediction and annotation**

702 The repeat landscape of *Riftia* genome was identified combining a custom giant tubeworm
703 RepeatModeler v2.0 library followed by the masking of the repetitive elements with RepeatMasker
704 v4.0.9 (A.F.A Smit, R. Hubley & P. Green, *RepeatMasker Open-4.0*). *Ab initio* gene prediction was
705 performed with Augustus v3.3.3 aided with hint files (Stanke and Morgenstern 2005; Hoff and
706 Stanke 2018). Only gene models with homology, orthology and gene expression evidence were
707 kept. Protein annotation was performed with Interproscan v5.39-77.0, RNAscan-SE 2.0.5, signalP
708 v5.0b and pfam_scan.pl (Jones et al. 2014; Lowe and Chan 2016; Almagro Armenteros et al.
709 2019).

710

711 **Identification of gene toolkits in *Riftia***

712 *Riftia* protein sequences were searched against well-curated catalog of developmental genes,
713 amino and fatty acid biosynthesis, endocytosis-, apoptosis, autophagy- and immune-related genes
714 using blastp v2.8.1+ and KEGG Automatic Annotation server (https://www.genome.jp/kegg/kaas/)
715 (Moriya et al. 2007). Additionally, protein domain information was retrieved from pfam_scan.pl and
716 Interpro results. Homology of the identified genes was confirmed through phylogenetic inferences
717 using iqtree v1.6.11 combining ModelFinder, tree search, 1000 ultra-fast bootstrap and SH-aLRT
718 test replicates (Nguyen et al. 2015; Kalyaanamoorthy et al. 2017; Hoang et al. 2018). The protein
719 diagrams were drawn using IBS v.1.0.3 software (Liu et al. 2015), and the clustered heatmaps
720 generated with the R package pheatmap (v1.0.12). Quantification of the gene expression levels
721 was performed with kallisto v0.46.1 (Bray et al. 2016).

722

723 **Orthology, gene family analysis and positively selected genes**

724 To assess *Riftia, Lamellibrachia* and Annelida lineage-specific genes orthology inferences using
725 selected non-bilaterian, deuterostome, lophotrochozoan, ecdysozoans, representatives (N=36)
726 were performed with Orthofinder v2.3.8 (Emms and Kelly 2019). To identify statistically significant

gene family expansions/contractions in *Riftia* compared to other lophotrochozoans a second round of orthology was performed using 18 lophotrochozoan representatives and *Tribolium castaneum* as outgroup. Finally, to identify the gene family core within Annelida a last instance of orthofinder v2.3.8 was invoked using the *C. teleta, H. robusta, L. luymesi* and *R. pachyptila.* Only the longest isoform for each gene was used in the analysis. Non-synonymous (Ka) and synonymous (Ks) substitution rates were calculated with the stand-alone version of KaKs_calculator v.2 and HyPhy v. 2.5.15 (Pond et al. 2005; Wang et al. 2010; Z. Zhang et al. 2012). Only single-copy genes (1:1 orthologs) without any inconsistencies between the nucleotide and protein sequences were used in the analyses. Contracted and expanded gene families in the giant tubeworm genome were identified using CAFE v4.2.1 (De Bie et al. 2006; Han et al. 2013) using a calibrated starting tree produced by Phylobayes v4.1b (Lartillot et al. 2013). The contracted/expanded gene families were annotated with Interproscan v5.39-77.0 and the enrichment analysis for Gene Ontology was performed with topGO v2.36.0 using Fisher's exact test against the *R. pachyptila* background (i.e., complete set of *Riftia* genes) coupled with weight01 algorithm. Rapidly evolving gene families in *Riftia* were annotated using PANTHER HMM scoring tool v2.2 with PANTHER_hmmscore database v15 (Mi et al. 2017). Protein domain contractions and expansions were found using iterative two-tailed Fisher's exact (Supplementary File 2) test applied to pfam_scan.pl results. The obtained p-values were corrected using Benjamini and Hochberg method (Benjamini and Hochberg 1995) and only domains with a significant p-value of < 0.01 were further investigated.

Haemoglobin evolution

The predicted *Riftia* haemoglobin (Hb) protein sequences were interrogated for the presence of the globin domain (PF00042) with hmmalign v3.1b2 (Mistry et al. 2013) and proteins without a hit were excluded from the analyses. Manual inspection and characterisation of the signature diagnostic residues/motifs in the haemoglobin chain and linker sequences were performed following previous works (Belato et al. 2019). Phylogenetic analyses were carried out as described in "Identification of gene toolkits in *Riftia*". The resulting trees were midpoint rooted using Figtree (http://tree.bio.ed.ac.uk/software/figtree/). Additionally, to investigate the haemoglobin gene expression across different environmental conditions (sulphur rich, sulphur depleted and medium) we downloaded six publicly available trophosome transcriptomes from SRA (https://www.ncbi.nlm.nih.gov/sra) (accession numbers: SRR8949066 to SRR8949071). The transcriptome libraries were pre-processed as described in "Transcriptome assembly and processing". *Riftia* Hb sequence was modelled using the Prime program implemented in the Schrödinger Drug Discovery (v2020.2) software suite. All illustrations of structures were made with PyMol v2.4 (https://pymol.org/2/).

Comparative tissue-specific transcriptome

The *Riftia* transcriptome libraries were pseudoaligned against the merged filtered AUGUSTUS gene models with kallisto v.0.46.1 (Bray et al. 2016) to collect the gene expression data expressed as TPM counts (transcripts per million). Normalisation within and across tissues were independently performed before calculating the tissue specificity tau values (see https://rdrr.io/github/roonysgalbi/tispec/f/vignettes/). To mitigate possible sex-specific differences in the gene expression levels, tau calculations were performed using only the tubeworm female tissues. The absolutely tissue-specific genes (genes expressed only in a single tissue defined by a tau value of 1) were submitted to enrichment analyses for Gene Ontology with topGO as mentioned in "Orthology, gene family analysis and positively selected genes".

Data availability

The raw long and short reads used to generate the draft genome and tissue-specific transcriptomes, respectively, are available in the SRA database under the BioProject number PRJNA754493 (Supplementary Table 12).

Acknowledgments

We thank Christian Baranyi (University of Vienna) and Jennifer Delaney (Harvard University) for the technical support on the RNA/DNA extraction and purification protocols, respectively. We also thank the Vienna BioCenter Core Facilities and the Genomics Center of the University of Minnesota for the generation of *Riftia*'s "omics" data. Andre Luiz de Oliveira (ALO) thanks Thales Kronenberger (Universität Klinikum Tübingen) for the help with the haemoglobin 3D homology modelling. ALO also thanks Andrew Calcino and Salvador Espada Hinojosa for their constructive conversations during the development of this work, and finally Bruna Yuri Pinheiro Imai for her continuous support to the first author. This work was funded by the Austrian FWF project P 31543 granted to M.B.

Author contributions

ALO, MB and PG designed the project. ALO generated, implemented, and executed all the bioinformatic pipelines. All data analysis was performed by ALO with input from MB and PG, and JM. ALO wrote the first complete draft of this manuscript and all authors read, commented on, and approved the final version.

References

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37:420–423.

800 Angela E. Douglas. 2010. The Symbiotic Habit. Princeton University Press Available from:
801 https://press.princeton.edu/books/paperback/9780691113425/the-symbiotic-habit

802 Arp AJ, Childress JJ. 1981. Blood Function in the Hydrothermal Vent Vestimentiferan Tube Worm.
803 *Science* 213:342–344.

804 Arp AJ, Childress JJ. 1983. Sulfide Binding by the Blood of the Hydrothermal Vent Tube Worm
805 Riftia pachyptila. *Science* 219:295–297.

806 Arp AJ, Childress JJ, Vetter RD. 1987. The Sulphide-Binding Protein in the Blood of the
807 Vestimentiferan Tube-Worm, Riftia Pachyptila, is the Extracellular Haemoglobin. *J. Exp.*
808 *Biol.* 128:139–158.

809 Bailly X, Jollivet D, Vanin S, Deutsch J, Zal F, Lallier F, Toulmond A. 2002. Evolution of the Sulfide-
810 Binding Function Within the Globin Multigenic Family of the Deep-Sea Hydrothermal Vent
811 Tubeworm Riftia pachyptila. *Mol. Biol. Evol.* 19:1421–1433.

812 Bartolomaeus Thomas, Quast B. 2005. Structure and development of nephridia in Annelida and
813 related taxa. In: Bartolomaeus T., Purschke G, editors. Morphology, Molecules, Evolution
814 and Phylogeny in Polychaeta and Related Taxa. Developments in Hydrobiology. Dordrecht:
815 Springer Netherlands. p. 139–165. Available from: https://doi.org/10.1007/1-4020-3240-4_9

816 Beklemishev VN. 1944. Osnovy sravintel'noi anatomii bespozvonochnykh. [Principles of the
817 comparative anatomy of invertebrates. *Mosc. Akad. Nauk* [Internet] 698. Available from:
818 https://www.marinespecies.org/aphia.php?p=sourcedetails&id=66718

819 Belato FA, Schrago CG, Coates CJ, Halanych KM, Costa-Paiva EM. 2019. Newly Discovered
820 Occurrences and Gene Tree of the Extracellular Globins and Linker Chains from the Giant
821 Hexagonal Bilayer Hemoglobin in Metazoans.Eyre-Walker A, editor. *Genome Biol. Evol.*
822 11:597–612.

823 Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful
824 Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* 57:289–300.

825 Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler
826 PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol.*
827 *Phylogenet. Evol.* 69:313–319.

828 Bonnivard E, Catrice O, Ravaux J, Brown SC, Higuet D. 2009. Survey of genome size in 28
829 hydrothermal vent species covering 10 families. *Genome* 52:524–536.

830 Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27:1767–1780.

831 Braunbeck T, Dales RP. 1984. The role of the heart-body and of the extravasal tissue in disposal of
832 foreign cells in two polychaete annelids. *Tissue Cell* 16:557–563.

833 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq
834 quantification. *Nat. Biotechnol.* 34:525–527.

835 Bright M, Eichinger I, von Salvini-Plawen L. 2013. The metatrochophore of a deep-sea
836 hydrothermal vent vestimentiferan (Polychaeta: Siboglinidae). *Org. Divers. Evol.* 13:163–
837 188.

838 Bright M, Keckeis H, Fisher CR. 2000. An autoradiographic examination of carbon fixation, transfer
839 and utilization in the Riftia pachyptila symbiosis. *Mar. Biol.* 136:621–632.

840 Bright M, Lallier F. 2010. The biology of vestimentiferan tubeworms. In: Gibson, RN, Atkinson, RJA,
841 Gordon, JDM, editors. Oceanography and Marine Biology: An animal review, Vol 48. Vol.
842 48. Oceanography and Marine Biology. CRC PRESS-TAYLOR & FRANCIS GROUP. p.
843 213–265. Available from: https://hal.archives-ouvertes.fr/hal-01250932

844 Bright M, Sorgo A. 2003. Ultrastructural reinvestigation of the trophosome in adults of Riftia
845 pachyptila (Annelida, Siboglinidae). *Invertebr. Biol.* 122:347–368.

846 Calcino AD, de Oliveira AL, Simakov O, Schwaha T, Zieger E, Wollesen T, Wanninger A. 2019. The
847 quagga mussel genome and the evolution of freshwater tolerance. *DNA Res.* 26:411–422.

848 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
849 BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.

850 Caullery M. 1914. Sur les Siboglinidae, type nouveau d'invertébrés recueillis par l'expédition du
851 Siboga. *Bull. Société Zool. Fr.*:350–353.

852 Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB. 1981. Prokaryotic Cells in
853 the Hydrothermal Vent Tube Worm Riftia pachyptila Jones: Possible Chemoautotrophic
854 Symbionts. *Science* 213:340–342.

855 Childress JJ, Fisher CR. 1992. The biology of hydrothermal vent animals: physiology, biochemistry,
856 and autotrophic symbioses. *Unkn. J.*:337–441.

857 Childress JJ, Girguis PR. 2011. The metabolic demands of endosymbiotic chemoautotrophic
858 metabolism on host physiological capacities. *J. Exp. Biol.* 214:312–325.

859 Cho S-J, Vallès Y, Giani VC, Seaver EC, Weisblat DA. 2010. Evolutionary Dynamics of the wnt
860 Gene Family: A Lophotrochozoan Perspective. *Mol. Biol. Evol.* 27:1645–1658.

861 Cian M-CD, Bailly X, Morales J, Strub J-M, Dorsselaer AV, Lallier FH. 2003. Characterization of
862 carbonic anhydrases from Riftia pachyptila, a symbiotic invertebrate from deep-sea
863 hydrothermal vents. *Proteins Struct. Funct. Bioinforma.* 51:327–339.

864 Cian MD, Regnault M, Lallier FH. 2000. Nitrogen metabolites and related enzymatic activities in the
865 body fluids and tissues of the hydrothermal vent tubeworm Riftia pachyptila. *J. Exp. Biol.*
866 203:2907–2920.

867 Corliss JB, Dymond J, Gordon LI, Edmond JM, von Herzen RP, Ballard RD, Green K, Williams D,
868 Bainbridge A, Crane K, et al. 1979. Submarine Thermal Springs on the Galápagos Rift.
869 *Science* 203:1073–1083.

870 Cuvillier-Hot V, Boidin-Wichlacz C, Tasiemski A. 2014. Polychates as annelid models to study
871 ecoimmunology of marine organisms.

872 Dales RP. 1964. The Coelomocytes of the Terebellid Polychaete Amphitrite Johnstoni. *J. Cell Sci.*
873 s3-105:263–279.

874 Davidson NM, Oshlack A. 2014. Corset: enabling differential gene expression analysis for de
875 novoassembled transcriptomes. *Genome Biol.* 15:410.

876 De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of
877 gene family evolution. *Bioinformatics* 22:1269–1271.

878 De Cian M-C, Regnault M, Lallier F. 1997. Nitrogenous metabolites in tissues and circulating fluids
879 of Riftia pachyptila. *Cah. Biol. Mar.* [Internet]. Available from:
880 http://www.vliz.be/en/imis?refid=66292

881 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
882 TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

883 Duboule D. 2007. The rise and fall of Hox gene clusters. *Development* 134:2549–2560.

884 Eckelbarger KJ. 1976. Larval Development and Population Aspects of the Reef-Building
885 Polychaete Phragmatopoma lapidosa from the East Coast of Florida. *Bull. Mar. Sci.*
886 26:117–132.

887 Ellers J, Kiers ET, Currie CR, McDonald BR, Visser B. 2012. Ecological interactions drive
888 evolutionary loss of traits. *Ecol. Lett.* 15:1071–1082.

889 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative
890 genomics. *Genome Biol.* 20:238.

Espinosa-Diez C, Miguel V, Mennerich D, Kietzmann T, Sánchez-Pérez P, Cadenas S, Lamas S. 2015. Antioxidant responses and cellular adjustments to oxidative stress. *Redox Biol.* 6:183–197.

Fauchald K, Rouse G. 1997. Polychaete systematics: Past and present. *Zool. Scr.* 26:71–138.

Felbeck H. 1981. Chemoautotrophic Potential of the Hydrothermal Vent Tube Worm, Riftia pachyptila Jones (Vestimentifera). *Science* 213:336–338.

Felbeck H. 1985. CO_2 Fixation in the Hydrothermal Vent Tube Worm Riftia pachyptila (Jones). *Physiol. Zool.* 58:272–281.

Felbeck H, Jarchow J. 1998. Carbon release from purified chemoautotrophic bacterial symbionts of the hydrothermal vent tubeworm Riftia pachyptila. *Physiol. Zool.* 71:294–302.

Felbeck H, Turner PJ. 1995. CO2 transport in catheterized hydrothermal vent tubeworms,Riftia pachyptila (vestimentifera). *J. Exp. Zool.* 272:95–102.

Fernández R, Gabaldón T. 2020. Gene gain and loss across the metazoan tree of life. *Nat. Ecol. Evol.* 4:524–533.

Fiermonte G, Paradies E, Todisco S, Marobbio CMT, Palmieri F. 2009. A Novel Member of Solute Carrier Family 25 (SLC25A42) Is a Transporter of Coenzyme A and Adenosine 3′,5′-Diphosphate in Human Mitochondria *. *J. Biol. Chem.* 284:18152–18159.

Fischer E. 1993. The myelo-erythroid nature of the chloragogenous-like tissues of the annelids. *Comp. Biochem. Physiol. A Physiol.* 106:449–453.

Fisher CR, Childress JJ, Arp AJ, Brooks JM, Distel D, Favuzzi JA, Macko SA, Newton A, Powell MA, Somero GN, et al. 1988a. Physiology, morphology, and biochemical composition of Riftia pachyptila at Rose Garden in 1985. *Deep Sea Res. Part Oceanogr. Res. Pap.* 35:1745–1758.

Fisher CR, Childress JJ, Arp AJ, Brooks JM, Distel D, Favuzzi JA, Macko SA, Newton A, Powell MA, Somero GN, et al. 1988b. Physiology, morphology, and biochemical composition of Riftia pachyptila at Rose Garden in 1985. *Deep Sea Res. Part Oceanogr. Res. Pap.* 35:1745–1758.

Flores JF, Fisher CR, Carney SL, Green BN, Freytag JK, Schaeffer SW, Royer WE. 2005. Sulfide binding is mediated by zinc ions discovered in the crystal structure of a hydrothermal vent tubeworm hemoglobin. *Proc. Natl. Acad. Sci.* 102:2713–2718.

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

923 Gaill F, Hunt S. 1986. Tubes of deep sea hydrothermal vent worms Riftia pachyptila
924 (Vestimentifera) and Alvinella pompejana (Annelida). *Mar. Ecol. Prog. Ser.* 34:267–274.

925 Gamage N, Barnett A, Hempel N, Duggleby RG, Windmill KF, Martin JL, McManus ME. 2006.
926 Human Sulfotransferases and Their Role in Chemical Metabolism. *Toxicol. Sci.* 90:5–22.

927 Gardebrecht A, Markert S, Sievert SM, Felbeck H, Thürmer A, Albrecht D, Wollherr A, Kabisch J,
928 Le Bris N, Lehmann R, et al. 2012. Physiological homogeneity among the endosymbionts
929 of Riftia pachyptila and Tevnia jerichonana revealed by proteogenomics. *ISME J.* 6:766–
930 776.

931 Gardiner SL, Jones ML. 1993. Microscopic anatomy of invertebrates. Wiley-Liss New York

932 Gardiner SL, Jones ML. 1994. On the Significance of Larval and Juvenile Morphology for
933 Suggesting Phylogenetic Relationships of the Vestimentifera. *Am. Zool.* 34:513–522.

934 Gazave E, Lemaître QIB, Balavoine G. 2017. The Notch pathway in the annelid Platynereis:
935 insights into chaetogenesis and neurogenesis processes. *Open Biol.* [Internet] 7. Available
936 from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5356439/

937 Gerdol M, Venier P, Pallavicini A. 2015. The genome of the Pacific oyster Crassostrea gigas brings
938 new insights on the massive expansion of the C1q gene family in Bivalvia. *Dev. Comp.*
939 *Immunol.* 49:59–71.

940 Girguis PR, Lee RW, Desaulniers N, Childress JJ, Pospesel M, Felbeck H, Zal F. 2000. Fate of
941 nitrate acquired by the tubeworm *Riftia pachyptila*. *Appl. Environ. Microbiol.* 66:2783–2790.

942

943 Girguis PR, Childress JJ, Freytag JK, Klose K, Stuber R. 2002. Effects of metabolite uptake on
944 proton-equivalent elimination by two species of deep-sea vestimentiferan tubeworm, Riftia
945 pachyptila and Lamellibrachia cf luymesi: proton elimination is a necessary adaptation to
946 sulfide-oxidizing chemoautotrophic symbionts. J. Exp. Biol. 205:3055–3066.

947 Goffredi S, Childress J, Desaulniers N, Lee R, Lallier F, Hammond DO. 1997. Inorganic carbon
948 acquisition by the hydrothermal vent tubeworm *Riftia pachyptila* depends upon high
949 external PCO2 and upon proton-equivalent ion transport by the worm. The Journal of
950 Experimental Biology. 200(5):883-96.

951 Goffredi SK, Girguis PR, Childress JJ, Desaulniers NT. 1999. Physiological functioning of carbonic
952 anhydrase in the hydrothermal vent tubeworm *Riftia pachyptila*. The Biological Bulletin.
953 196(3):257-264.

954 Govenar B, Fisher CR. 2007. Experimental evidence of habitat provision by aggregations of *Riftia*
955 *pachyptila* at hydrothermal vents on the East Pacific Rise. *Mar. Ecol.* 28:3–14.

956 Govenar B, Le Bris N, Gollner S, Glanville J, Aperghis A, Hourdez S, Fisher C. 2005. Epifaunal
957 community structure associated with *Riftia pachyptila* aggregations in chemically different
958 hydrothermal vent habitats. *Mar. Ecol. Prog. Ser.* 305:67–77.

959 Grassle JF. 1987. The Ecology of Deep-Sea Hydrothermal Vent Communities. In: Blaxter JHS,
960 Southward AJ, editors. Advances in Marine Biology. Vol. 23. Academic Press. p. 301–362.
961 Available from: https://www.sciencedirect.com/science/article/pii/S0065288108601108

962 Grigorian M, Hartenstein V. 2013. Hematopoiesis and hematopoietic organs in arthropods. *Dev.*
963 *Genes Evol.* 223:103–115.

964 Guijarro-Clarke C, Holland PWH, Paps J. 2020. Widespread patterns of gene loss in the evolution
965 of the animal kingdom. *Nat. Ecol. Evol.* 4:519–523.

966 Halanych KM, Feldman RA, Vrijenhoek RC. 2001. Molecular Evidence that Sclerolinum brattstromi
967 Is Closely Related to Vestimentiferans, not to Frenulate Pogonophorans (Siboglinidae,
968 Annelida). *Biol. Bull.* 201:65–75.

969 Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating Gene Gain and Loss Rates
970 in the Presence of Error in Genome Assembly and Annotation Using CAFE 3. *Mol. Biol.*
971 *Evol.* 30:1987–1997.

972 Hand SC. 1987. Trophosome ultrastructure and the characterization of isolated bacteriocytes from
973 invertebrate-sulfur bacteria symbioses. *Biol. Bull.* 173:260–276.

974 Hartenstein V. 2006. Blood Cells and Blood Cell Development in the Animal Kingdom. *Annu. Rev.*
975 *Cell Dev. Biol.* 22:677–712.

976 Hejnol A, Martín-Durán JM. 2015. Getting to the bottom of anal evolution. *Zool. Anz. - J. Comp.*
977 *Zool.* 256:61–74.

978 Hentschel U, Felbeck H. 1993, Nitrate respiration in the hydrothermal vent tubeworm *Riftia*
979 *pachyptila*. Nature. 366(6453):338-40.

980 Hessler RR, Smithey WM, Boudrias MA, Keller CH, Lutz RA, Childress JJ. 1988. Temporal change
981 in megafauna at the Rose Garden hydrothermal vent (Galapagos Rift; eastern tropical
982 Pacific). *Deep Sea Res. Part Oceanogr. Res. Pap.* 35:1681–1709.

983 Hilário A, Capa M, Dahlgren TG, Halanych KM, Little CTS, Thornhill DJ, Verna C, Glover AG. 2011.
984 New Perspectives on the Ecology and Evolution of Siboglinid Tubeworms.Laudet V, editor.
985 *PLoS ONE* 6:e16309.

986 Hinzke T, Kleiner M, Breusing C, Felbeck H, Häsler R, Sievert SM, Schlüter R, Rosenstiel P,
987 Reusch TBH, Schweder T, et al. 2019. Host-Microbe Interactions in the Chemosynthetic
988 Riftia pachyptila Symbiosis. 10:20.

989 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the
990 Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35:518–522.

991 Hoff KJ, Stanke M. 2018. Predicting Genes in Single Genomes with AUGUSTUS. *Curr. Protoc.*
992 *Bioinforma.*:e57.

993 Holstein TW. 2012. The Evolution of the Wnt Pathway. *Cold Spring Harb. Perspect. Biol.* [Internet]
994 4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3385961/

995 Houten SM, Violante S, Ventura FV, Wanders RJA. 2016. The Biochemistry and Physiology of
996 Mitochondrial Fatty Acid β-Oxidation and Its Genetic Disorders. *Annu. Rev. Physiol.* 78:23–
997 44.

998 Hunt HL, Metaxas A, Jennings RM, Halanych KM, Mullineaux LS. 2004. Testing biological control
999 of colonization by vestimentiferan tubeworms at deep-sea hydrothermal vents (East Pacific
1000 Rise, 9°50′N). *Deep Sea Res. Part Oceanogr. Res. Pap.* 51:225–234.

1001 Indiveri C, Iacobazzi V, Tonazzi A, Giangregorio N, Infantino V, Convertini P, Console L, Palmieri F.
1002 2011. The mitochondrial carnitine/acylcarnitine carrier: Function, structure and
1003 physiopathology. *Mol. Aspects Med.* 32:223–233.

1004 Ingham PW, Nakano Y, Seger C. 2011. Mechanisms and functions of Hedgehog signalling across
1005 the metazoa. *Nat. Rev. Genet.* 12:393–406.

1006 Ip JC-H, Xu T, Sun J, Li R, Chen C, Lan Y, Han Z, Zhang H, Wei J, Wang H, et al. 2021. Host-
1007 Endosymbiont Genome Integration in a Deep-Sea Chemosymbiotic Clam. *Mol. Biol. Evol.*
1008 [Internet]. Available from: https://academic.oup.com/mbe/advance-
1009 article/doi/10.1093/molbev/msaa241/5909661

1010 Jennings RM, Halanych KM. 2005. Mitochondrial Genomes of Clymenella torquata (Maldanidae)
1011 and Riftia pachyptila (Siboglinidae): Evidence for Conserved Gene Order in Annelida. *Mol.*
1012 *Biol. Evol.* 22:210–222.

1013 Jones ML. 1981. Riftia pachyptila Jones: Observations on the Vestimentiferan Worm from the
1014 Galápagos Rift. *Science* 213:333–336.

1015 Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A,
1016 Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification.
1017 *Bioinformatics* 30:1236–1240.

1018 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast
1019 model selection for accurate phylogenetic estimates. *Nat. Methods* 14:587–589.

1020 Kast DJ, Dominguez R. 2017. The Cytoskeleton–Autophagy Connection. *Curr. Biol.* 27:R318–
1021 R326.

1022 Kiers ET, West SA. 2015. Evolving new organisms via symbiosis. *Science* 348:392–394.

1023 Kirkitadze MD, Barlow PN. 2001. Structure and flexibility of the multiple domain proteins that
1024 regulate complement activation. *Immunol. Rev.* 180:146–161.

1025 Klose J, Polz MF, Wagner M, Schimak MP, Gollner S, Bright M. 2015. Endosymbionts escape dead
1026 hydrothermal vent tubeworms to enrich the free-living population. *Proc. Natl. Acad. Sci.*
1027 112:11300–11305.

1028 Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat
1029 graphs. *Nat. Biotechnol.* 37:540–546.

1030 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and
1031 accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. *Genome*
1032 *Res.* 27:722–736.

1033 Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019. Transcriptome
1034 assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20:278.

1035 Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies. *F1000Research*
1036 6:1287.

1037 Larsen EH, Deaton LE, Onken H, O'Donnell M, Grosell M, Dantzler WH, Weihrauch D. 2011.
1038 Osmoregulation and excretion. *Compr. Physiol.* 4:405–573.

1039 Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: Phylogenetic Reconstruction
1040 with Infinite Mixtures of Profiles in a Parallel Environment. *Syst. Biol.* 62:611–615.

1041 Le Bris N, Govenar B, Le Gall C, Fisher CR. 2006. Variability of physico-chemical conditions in
1042 9°50′N EPR diffuse flow vent habitats. *Mar. Chem.* 98:167–182.

1043 Le Bris N, Rodier P, Sarradin P-M, Le Gall C. 2006. Is temperature a good proxy for sulfide in
1044 hydrothermal vent habitats? *Cah. Biol. Mar.* 47:465–470.

1045 Le Bris N, Sarradin P-M, Caprais J-C. 2003. Contrasted sulphide chemistries in the environment of
1046 13°N EPR vent fauna. *Deep Sea Res. Part Oceanogr. Res. Pap.* 50:737–747.

1047 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences.Birol I, editor. *Bioinformatics*
1048 34:3094–3100.

1049 Li Y, Kocot KM, Schander C, Santos SR, Thornhill DJ, Halanych KM. 2015. Mitogenomics reveals
1050 phylogeny and repeated motifs in control regions of the deep-sea family Siboglinidae
1051 (Annelida). *Mol. Phylogenet. Evol.* 85:221–229.

1052 Li Y, Tassia MG, Waits DS, Bogantes VE, David KT, Halanych KM. 2019. Genomic adaptations to
1053 chemosymbiosis in the deep-sea seep-dwelling tubeworm Lamellibrachia luymesi. *BMC*
1054 *Biol.* 17:91.

1055 Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, Lahrmann U, Zhao Q, Zheng Y, Zhao Y, et al. 2015. IBS:
1056 an illustrator for the presentation and visualization of biological sequences. *Bioinformatics*
1057 31:3359–3361.

1058 Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of
1059 transfer RNA genes. *Nucleic Acids Res.* 44:W54–W57.

1060 Luo Y-J, Kanda M, Koyanagi R, Hisata K, Akiyama T, Sakamoto H, Sakamoto T, Satoh N. 2018.
1061 Nemertean and phoronid genomes reveal lophotrochozoan evolution and the origin of
1062 bilaterian heads. *Nat. Ecol. Evol.* 2:141–151.

1063 Luther GW, Rozan TF, Taillefert M, Nuzzio DB, Di Meo C, Shank TM, Lutz RA, Cary SC. 2001.
1064 Chemical speciation drives hydrothermal vent ecology. *Nature* 410:813–816.

1065 Lutz RA, Shank TM, Evans R. 2001. Life After Death in the Deep Sea: Following immolation by
1066 volcanic eruption, the community around a hydrothermal vent recovers spectacularly. *Am.*
1067 *Sci.* 89:422–431.

1068 Lutz RA, Shank TM, Fornari DJ, Haymon RM, Lilley MD, Von Damm KL, Desbruyeres D. 1994.
1069 Rapid growth at deep-sea vents. *Nature* 371:663–664.

1070 Majd H, King MS, Smith AC, Kunji ERS. 2018. Pathogenic mutations of the human mitochondrial
1071 citrate carrier SLC25A1 lead to impaired citrate export required for lipid, dolichol,
1072 ubiquinone and sterol synthesis. *Biochim. Biophys. Acta BBA - Bioenerg.* 1859:1–7.

1073 Marsh AG, Mullineaux LS, Young CM, Manahan DT. 2001. Larval dispersal potential of the
1074 tubeworm Riftia pachyptila at deep-sea hydrothermal vents. *Nature* 411:77–80.

1075 Martín-Durán JM, Vellutini BC, Marlétaz F, Cetrangolo V, Cvetesic N, Thiel D, Henriet S, Grau-
1076 Bové X, Carrillo-Baltodano AM, Gu W, et al. 2020. Conservative route to genome
1077 compaction in a miniature annelid. *Nat. Ecol. Evol.*

1078 Massagué J. 2012. TGFβ signalling in context. *Nat. Rev. Mol. Cell Biol.* 13:616–630.

1079 McAnulty SJ, Nyholm SV. 2017. The Role of Hemocytes in the Hawaiian Bobtail Squid, Euprymna
1080 scolopes: A Model Organism for Studying Beneficial Host–Microbe Interactions. *Front.*
1081 *Microbiol.* [Internet] 7. Available from:
1082 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5216023/

1083 McHugh D. 1997. Molecular evidence that echiurans and pogonophorans are derived annelids.
1084 *Proc. Natl. Acad. Sci. U. S. A.* 94:8006–8009.

1085 Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. 2017. PANTHER version
1086 11: expanded annotation data from Gene Ontology and Reactome pathways, and data
1087 analysis tool enhancements. *Nucleic Acids Res.* 45:D183–D189.

1088 Micheli F, Peterson CH, Mullineaux LS, Fisher CR, Mills SW, Sancho G, Johnson GA, Lenihan HS.
1089 2002. Predation Structures Communities at Deep-Sea Hydrothermal Vents. *Ecol. Monogr.*
1090 72:365–382.

1091 Minic Z, Hervé G. 2003. Arginine Metabolism in the Deep Sea Tube Worm *Riftia pachyptila* and Its
1092 Bacterial Endosymbiont. *J. Biol. Chem.* 278:40527–40533.

1093 Minic Z, Simon V, Penverne B, Gaill F, Hervé G. 2001. Contribution of the Bacterial Endosymbiont
1094 to the Biosynthesis of Pyrimidine Nucleotides in the Deep-sea Tube Worm *Riftia pachyptila*.
1095 *J. Biol. Chem.* 276:23777–23784.

1096 Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search:
1097 HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41:e121–
1098 e121.

1099 Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome
1100 annotation and pathway reconstruction server. *Nucleic Acids Res.* 35:W182–W185.

1101 Moustakas A, Heldin C-H. 2009. The regulation of TGFβ signal transduction. *Development*
1102 136:3699–3714.

1103 Mullineaux LS, Fisher CR, Peterson CH, Schaeffer SW. 2000. Tubeworm succession at
1104 hydrothermal vents: use of biogenic cues to reduce habitat selection error? *Oecologia*
1105 123:275–284.

1106 Mullineaux LS, Mills SW, Le Bris N, Beaulieu SE, Sievert SM, Dykman LN. 2020. Prolonged
1107 recovery time after eruptive disturbance of a deep-sea hydrothermal vent community. *Proc.*
1108 *R. Soc. B Biol. Sci.* 287:20202070.

1109 Mullineaux LS, Peterson CH, Micheli F, Mills SW. 2003. Successional Mechanism Varies Along a
1110 Gradient in Hydrothermal Fluid Flux at Deep-Sea Vents. *Ecol. Monogr.* 73:523–542.

1111 Nakahama S, Nakagawa T, Kanemori M, Fukumori Y, Sasayama Y. 2008. Direct evidence that
1112 extracellular giant hemoglobin is produced in chloragogen tissues in a beard worm,
1113 Oligobrachia mashikoi (Frenulata, Siboglinidae, Annelida). *Zoolog. Sci.* 25:1247–1252.

1114 Nelson D, Fisher C. 1995. Chemoautotrophic and methanotrophic endosymbiotic bacteria at deep-
1115 sea vents and seeps. In: The Microbiology of Deep-Sea Hydrothermal Vents. Karl DM.
1116 Boca Raton, FL: CRC. p. 125–167.

1117 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective
1118 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.*
1119 32:268–274.

1120 Niehrs C. 2012. The complex world of WNT receptor signalling. *Nat. Rev. Mol. Cell Biol.* 13:767–
1121 779.

1122 Nielsen C. 2004. Trochophora larvae: cell-lineages, ciliary bands, and body regions. 1. Annelida
1123 and Mollusca. *J. Exp. Zoolog. B Mol. Dev. Evol.* 302:35–68.

1124 Nielsen C, Brunet T, Arendt D. 2018. Evolution of the bilaterian mouth and anus. *Nat. Ecol. Evol.*
1125 2:1358–1376.

1126 Nussbaumer AD, Fisher CR, Bright M. 2006. Horizontal endosymbiont transmission in
1127 hydrothermal vent tubeworms. *Nature* 441:345–348.

1128 Nyholm SV, Song P, Dang J, Bunce C, Girguis PR. 2012. Expression and Putative Function of
1129 Innate Immunity Genes under in situ Conditions in the Symbiotic Hydrothermal Vent
1130 Tubeworm Ridgeia piscesae.Leulier F, editor. *PLoS ONE* 7:e38267.

1131 Pearson JC, Lemons D, McGinnis W. 2005. Modulating Hox gene functions during animal body
1132 patterning. *Nat. Rev. Genet.* 6:893–904.

1133 Pflugfelder B, Cary SC, Bright M. 2009. Dynamics of cell proliferation and apoptosis reflect
1134 different life strategies in hydrothermal vent and cold seep vestimentiferan tubeworms. *Cell*
1135 *Tissue Res.* 337:149–165.

1136 Pila EA, Sullivan JT, Wu XZ, Fang J, Rudko SP, Gordy MA, Hanington PC. 2016. Haematopoiesis
1137 in molluscs: A review of haemocyte development and function in gastropods, cephalopods
1138 and bivalves. *Dev. Comp. Immunol.* 58:119–128.

1139 Pires-daSilva A, Sommer RJ. 2003. The evolution of signalling pathways in animal development.
1140 *Nat. Rev. Genet.* 4:39–49.

1141 Pleijel F, Dahlgren TG, Rouse GW. 2009. Progress in systematics: from Siboglinidae to
1142 Pogonophora and Vestimentifera and back to Siboglinidae. *C. R. Biol.* 332:140–148.

1143 Polzin J, Arevalo P, Nussbaumer T, Polz MF, Bright M. 2019. Polyclonal symbiont populations in
1144 hydrothermal vent tubeworms and the environment. *Proc. R. Soc. B Biol. Sci.*
1145 286:20181281.

1146 Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies.
1147 *Bioinformatics* 21:676–679.

1148 Potswald HE. 1969. Cytological observations on the so-called neoblasts in the serpulid Spirorbis.
1149 *J. Morphol.* 128:241–259.

1150 Powell MA, Somero GN. 1986. Adaptations to Sulfide by Hydrothermal Vent Animals: Sites and
1151 Mechanisms of Detoxification and Metabolism. *Biol. Bull.* 171:274–290.

1152 Quetin LB, Childress JJ. 1980. Observations on the swimming activity of two bathypelagic mysid
1153 species maintained at high hydrostatic pressures. *Deep Sea Res. Part Oceanogr. Res.*
1154 *Pap.* 27:383–391.

1155 Rau GH. 1981a. Low 15N/14N in hydrothermal vent animals: ecological implications. *Nature*
1156 289:484–485.

1157 Rau GH. 1981b. Hydrothermal Vent Clam and Tube Worm 13C/12C: Further Evidence of
1158 Nonphotosynthetic Food Sources. *Science* 213:338–340.

1159 Rimskaya-Korsakova NN, Karaseva NP, Temereva EN, Malakhov VV. 2018. Protonephridial
1160 Excretory System in Vestimentifera (Siboglinidae, Annelida). *Dokl. Biol. Sci.* 478:22–25.

1161 Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM,
1162 Qian JQ, et al. 2010. De novo assembly and analysis of RNA-seq data. *Nat. Methods*
1163 7:909–912.

1164 Robidart JC, Bench SR, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T, Allen EE,
1165 Felbeck H. 2008. Metabolic versatility of the Riftia pachyptila endosymbiont revealed
1166 through metagenomics. *Environ. Microbiol.* 10:727–737.

1167 Roeder RG. 1996. The role of general initiation factors in transcription by RNA polymerase II.
1168 *Trends Biochem. Sci.* 21:327–335.

1169 Rouse GW. 2001. A cladistic analysis of Siboglinidae Caullery, 1914 (Polychaeta, Annelida):
1170 formerly the phyla Pogonophora and Vestimentifera. *Zool. J. Linn. Soc.* 132:55–80.

1171 Ruppert EE, Smith PR. 1988. The Functional Organization of Filtration Nephridia. *Biol. Rev.*
1172 63:231–258.

1173 Salzet M, Tasiemski A, Cooper E. 2006. Innate Immunity in Lophotrochozoans: The Annelids. *Curr.*
1174 *Pharm. Des.* 12:3043–3050.

1175 Sanchez S, Hourdez S, Lallier FH. 2007. Identification of proteins involved in the functioning of
1176 Riftia pachyptila symbiosis by Subtractive Suppression Hybridization. *BMC Genomics*
1177 8:337.

1178 Schmitz JF, Zimmer F, Bornberg-Bauer E. 2016. Mechanisms of transcription factor evolution in
1179 Metazoa. *Nucleic Acids Res.* 44:6287–6297.

1180 Schneider G, Käck H, Lindqvist Y. 2000. The manifold of vitamin B6 dependent enzymes. *Structure*
1181 8:R1–R6.

1182 Schulze A. 2001. Comparative anatomy of excretory organs in vestimentiferan tube worms
1183 (Pogonophora, Obturata). *J. Morphol.* 250:1–11.

1184 Schulze A. 2002. Histological and ultrastructural characterization of the intravasal body in
1185 Vestimentifera (Siboglinidae, Polychaeta, Annelida). Available from: http://application.sb-
1186 roscoff.fr/cbm/doi/10.21411/CBM.A.CBEBBD63

1187 Shank TM, Fornari DJ, Von Damm KL, Lilley MD, Haymon RM, Lutz RA. 1998. Temporal and
1188 spatial patterns of biological community development at nascent deep-sea hydrothermal
1189 vents (9°50′N, East Pacific Rise). *Deep Sea Res. Part II Top. Stud. Oceanogr.* 45:465–515.

1190 Shillito B, Ravaux J, Gaill F, Delachambre J, Thiébaut E, Childress JJ. 1999. Preliminary data on
1191 carbon production of deep-sea vent tubeworms. *Mar. Ecol. Prog. Ser.* 183:275–279.

1192 Shively JM, van Keulen G, Meijer WG. 1998. SOMETHING FROM ALMOST NOTHING: Carbon
1193 Dioxide Fixation in Chemoautotrophs. *Annu. Rev. Microbiol.* 52:191–230.

1194 Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, Bredeson J,
1195 Shoguchi E, Gyoja F, et al. 2015. Hemichordate genomes and deuterostome origins.
1196 *Nature* 527:459–465.

1197 Simakov O, Marletaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo D-H, Larsson T,
1198 Lv J, Arendt D, et al. 2013. Insights into bilaterian evolution from three spiralian genomes.
1199 *Nature* 493:526–531.

1200 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing
1201 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*
1202 31:3210–3212.

1203 Southward E. 1993. Pogonophora. In: Microscopic Anatomy of Invertebrates, Onychophora,
1204 Chilopoda and Lesser Protostomata. F.W. Harrison and M. E. Rice. Wiley-Liss New York. p.
1205 327–369.

1206 Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that
1207 allows user-defined constraints. *Nucleic Acids Res.* 33:W465–W467.

1208 Stewart FJ, Cavanaugh CM. 2006. Symbiosis of Thioautotrophic Bacteria with Riftia pachyptila. In:
1209 Overmann J, editor. Molecular Basis of Symbiosis. Vol. 41. Progress in Molecular and
1210 Subcellular Biology. Berlin/Heidelberg: Springer-Verlag. p. 197–225. Available from:
1211 http://link.springer.com/10.1007/3-540-28221-1_10

1212 Sun J, Zhang Yu, Xu T, Zhang Yang, Mu H, Zhang Yanjie, Lan Y, Fields CJ, Hui JHL, Zhang W, et
1213 al. 2017. Adaptation to deep-sea chemosynthetic environments as revealed by mussel
1214 genomes. *Nat. Ecol. Evol.* 1:0121.

1215 Szathmáry E, Smith JM. 1995. The major evolutionary transitions. *Nature* 374:227–232.

1216 Tanaka Y, Suzuki A, Sakai K. 2018. The stoichiometry of coral-dinoflagellate symbiosis: carbon and
1217 nitrogen cycles are balanced in the recycling and double translocation system. ISME J.
1218 12:860–868.

1219 Terwilliger R, Terwilliger N, Bonaventura C, Bonaventura J, Schabtach E. 1985. Structural and
1220 functional properties of hemoglobin from the vestimentiferan Pogonophora, Lamellibrachia.
1221 *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* 829:27–33.

1222 Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq -
1223 versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45:W6–W11.

1224 Timpl R, Brown JC. 1996. Supramolecular assembly of basement membranes. *BioEssays* 18:123–
1225 132.

1226 Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, Brooks KL, Tracey A, Bobes
1227 RJ, Fragoso G, Sciutto E, et al. 2013. The genomes of four tapeworm species reveal
1228 adaptations to parasitism. *Nature* 496:57–63.

1229 Uda K, Tanaka K, Bailly X, Zal F, Suzuki T. 2005. Phosphagen kinase of the giant tubeworm Riftia
1230 pachyptila: Cloning and expression of cytoplasmic and mitochondrial isoforms of
1231 taurocyamine kinase. *Int. J. Biol. Macromol.* 37:54–60.

1232 Van der Land J. 1975. The systematic position of Lamellibrachia (Annelida, Vestimentifera). *Z.*
1233 *Zool. Syst. Evol. Sonderh.* 1975:86–101.

1234 Veaudor T, Cassier-Chauvat C, Chauvat F. 2019. Genomics of Urea Transport and Catabolism in
1235 Cyanobacteria: Biotechnological Implications. *Front. Microbiol.* [Internet] 10. Available from:
1236 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6737895/

1237 Vetvicka V, Sima P. 2009. Origins and functions of annelide immune cells: the concise survey.
1238 *Invertebr. Surviv. J.* 6:138–143.

1239 Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: A Toolkit Incorporating
1240 Gamma-Series Methods and Sliding Window Strategies. *Genomics Proteomics*
1241 *Bioinformatics* 8:77–80.

1242 West SA, Fisher RM, Gardner A, Kiers ET. 2015. Major evolutionary transitions in individuality.
1243 *Proc. Natl. Acad. Sci.* 112:10112–10119.

1244 Wyatt K, White HE, Wang L, Bateman OA, Slingsby C, Orlova EV, Wistow G. 2006. Lengsin is a
1245 survivor of an ancient family of class I glutamine synthetases in eukaryotes that has
1246 undergone evolutionary re-engineering for a role in the vertebrate eye lens. *Struct. Lond.*
1247 *Engl. 1993* 14:1823–1834.

1248 Xiang T, Lehnert E, Jinkerson RE, Clowez S, Kim RG, DeNofrio JC, Pringle JR, Grossman AR.
1249 2020. Symbiont population control by host-symbiont metabolic interaction in
1250 Symbiodiniaceae-cnidarian associations. *Nat. Commun.* 11:108.

1251 Zal F, Lallier FH, Green BN, Vinogradov SN, Toulmond A. 1996. The Multi-hemoglobin System of
1252 the Hydrothermal Vent Tube Worm Riftia pachyptila II. COMPLETE POLYPEPTIDE CHAIN

1253 COMPOSITION INVESTIGATED BY MAXIMUM ENTROPY ANALYSIS OF MASS
1254 SPECTRA. *J. Biol. Chem.* 271:8875–8881.

1255 Zal F, Leize E, Lallier FH, Toulmond A, Dorsselaer AV, Childress JJ. 1998. S-Sulfohemoglobin and
1256 disulfide exchange: The mechanisms of sulfide binding by Riftia pachyptila hemoglobins.
1257 *Proc. Natl. Acad. Sci.* 95:8997–9002.

1258 Zal F, Suzuki T, Kawasaki Y, Childress JJ, Lallier FH, Toulmond A. 1997. Primary structure of the
1259 common polypeptide chain b from the multi-hemoglobin system of the hydrothermal vent
1260 tube worm Riftia pachyptila: An insight on the sulfide binding-site. *Proteins Struct. Funct.*
1261 *Bioinforma.* 29:562–574.

1262 Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, et al. 2012. The
1263 oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490:49–
1264 54.

1265 Zhang Z, Xiao J, Wu J, Zhang H, Liu G, Wang X, Dai L. 2012. ParaAT: A parallel tool for
1266 constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.*
1267 419:779–781.

1268 Zhou XZ, Lu KP. 2001. The Pin2/TRF1-Interacting Protein PinX1 Is a Potent Telomerase Inhibitor.
1269 *Cell* 107:347–359.

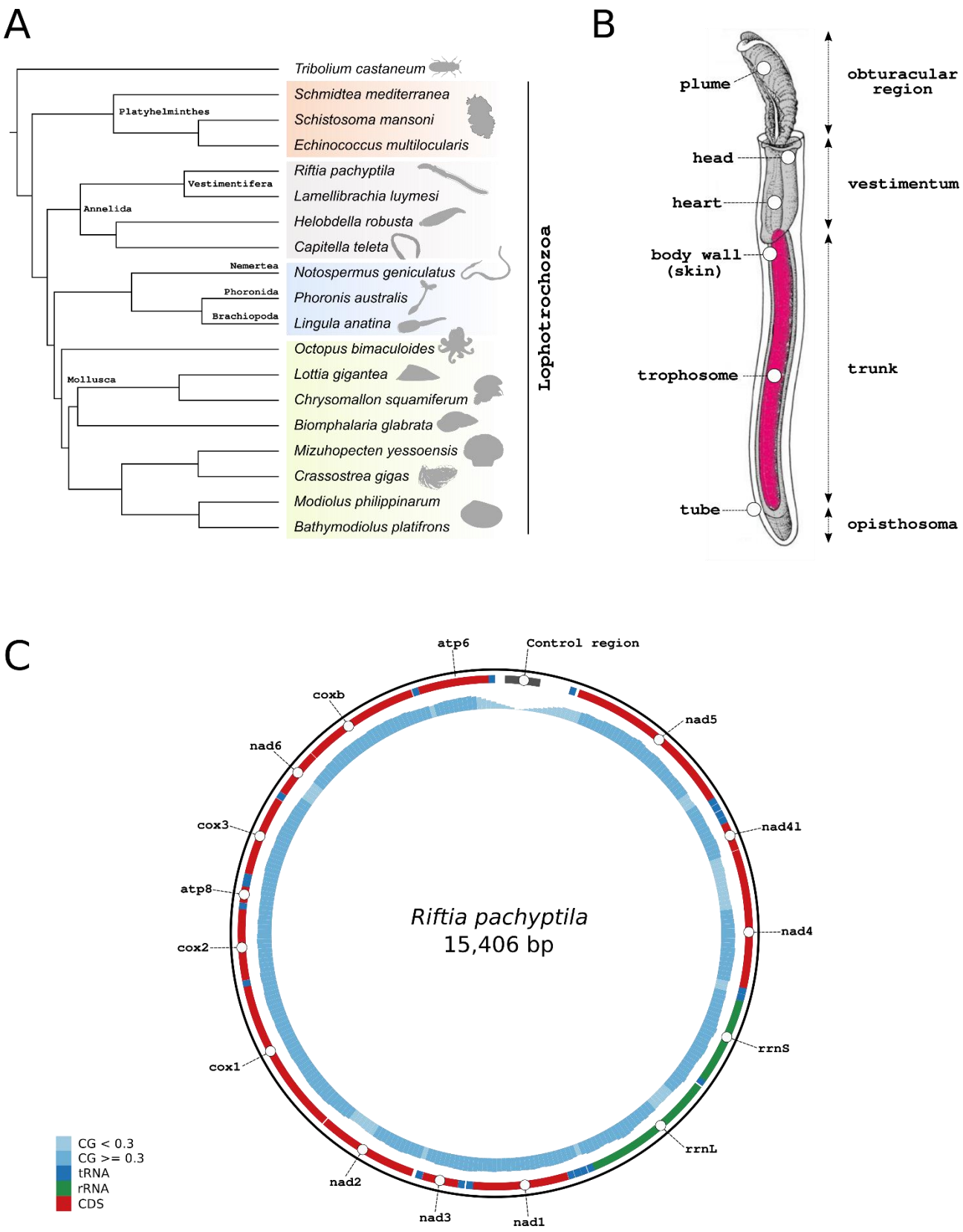


Figure 1 – Overview of *Riftia pachyptila* body plan anatomy, phylogenetic placement and mitochondrial genome. **A**, Phylogetic placement of *Riftia pachyptila* within Lophotrochozoa. *Riftia* together with *Lamellibrachia* forms the clade Vestimentifera, a clade of marine animals living in chitinous tubes and lacking a digestive tract. Animal silhouettes were download from http://phylopic.org/. Tree topology was obtained through phylogenomic analysis. **B**, Schematic

drawing of *Riftia pachyptila* adult. The first part of the body, the obturacular region, contains the highly vascularised plume, whereas the head, heart and gonads are located in the second body part, the vestimentum. The trunk region and third body part harbors the trophosome (organ that houses the symbiotic bacteria), body wall (skin). The posterior part, the opisthosoma is the fourth and last body region of the tubeworm. Schematic drawing was modified from Nussbaumer et al. (2006). **C**, Schematic representation of *Riftia pachyptila* mitochondrial genome, including the complete control region. CG-content and tRNA genes are represented by the blue histograms and boxes, respectively.

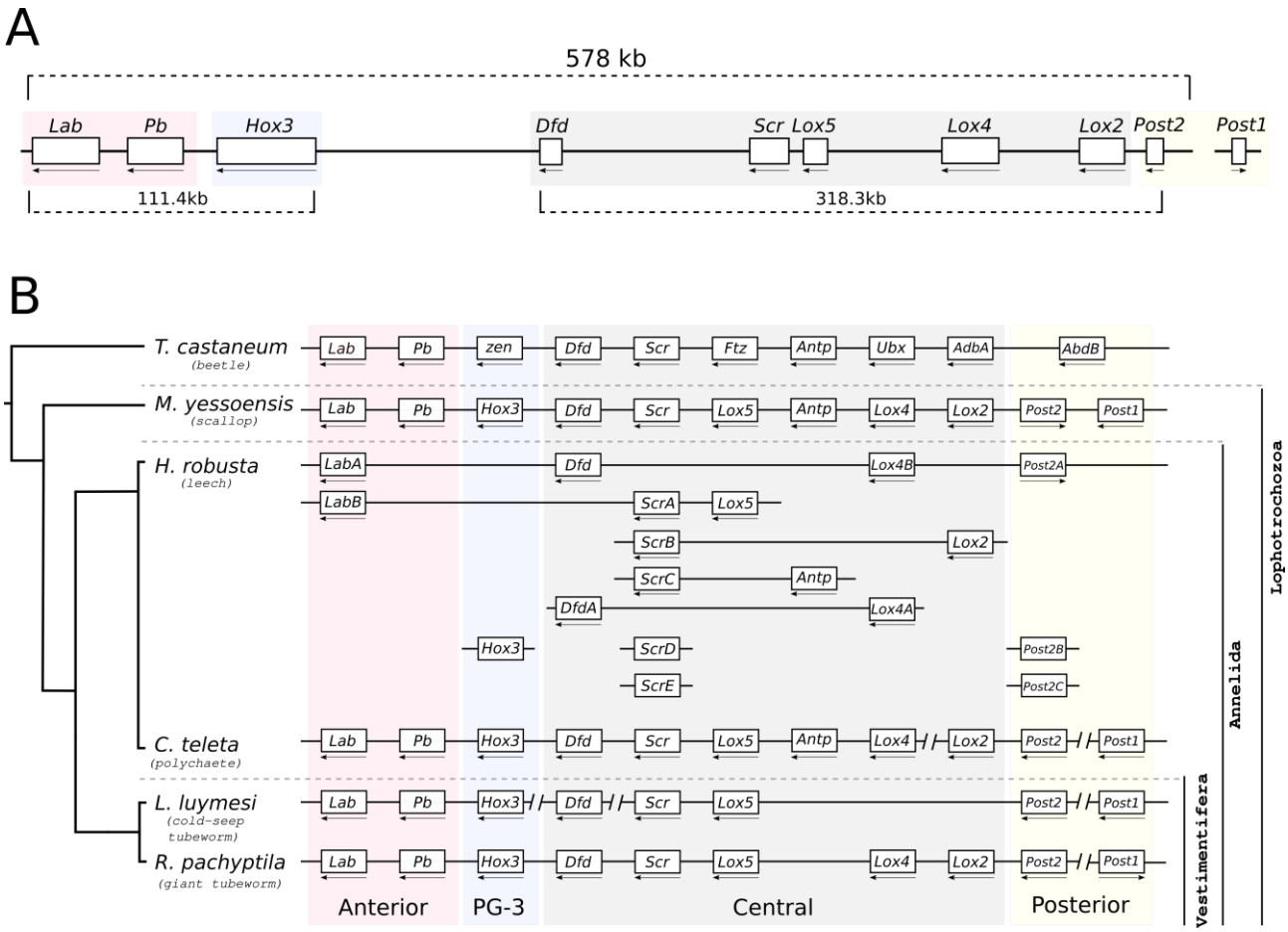
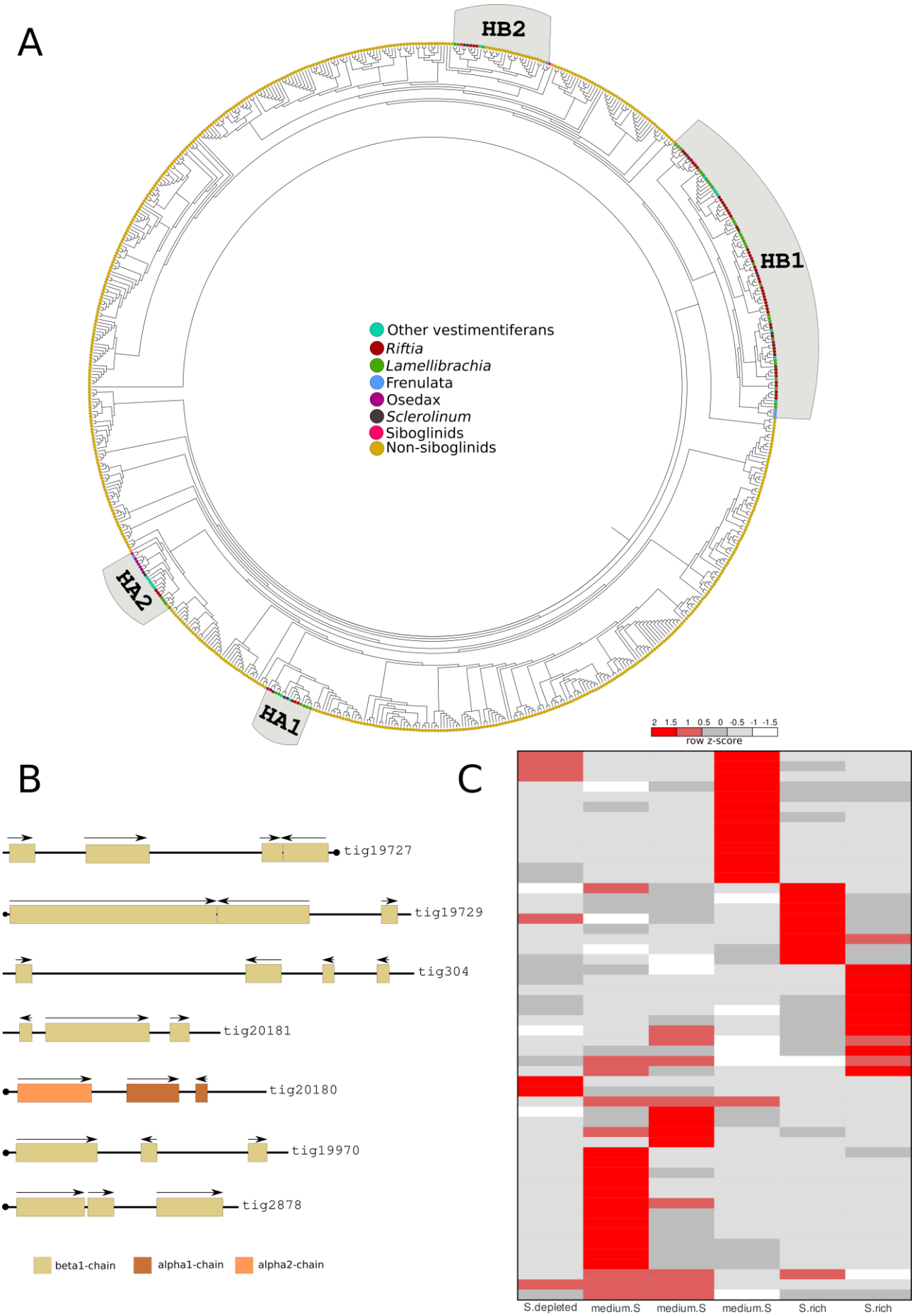


Figure 2 – *The* Hox gene complement of *Riftia pachyptila* and selected metazoans. **A**, Hox cluster organisation in the genome of *Riftia pachyptila*. Nine out of the ten Hox genes are located in one single genomic scaffold. *Hox7* is missing from the giant tubeworm genome. Arrows indicate direction of transcription. Only the longest gene model is shown. **B**, Hox cluster present in selected metazoans. *Riftia* presents the most intact Hox cluster among annelids. Part of the central Hox class is missing from the cold-seep tubeworm *Lamellibrachia*. *Helobdella* and *Capitella* cluster are adapt from Simakov et al. (2013), whereas *Mizuhopecten* cluster is from Wang et al. (2017).



1292 Figure 3 – Expanded haemoglobin complement in *Riftia pachyptila.* **A**, Midpoint rooted phylogeny
1293 of 693 *Riftia*, annelid and metazoan haemoglobin genes, using Paiva et al. (2019) as backbone.

Coloured circles correspond to different annelid taxa and metazoans. **B**, Seven genomic clusters of haemoglobin genes in *Riftia* genome. Arrows indicate the direction of transcription. Scaffolds with a circle on the end indicate the presence of Hb genes in the terminal end of the scaffolds. Only the longest gene models are shown. Colours represent the different haemoglobin chains. Two β1- and one β2-Hbs genes are located in three separate scaffolds (tig3224, 19723 and 19768). **C**, Heat map expression of haemoglobins in the trophosome under three experimental conditions: medium sulphide (medium.S), sulphide rich (S.rich) and sulphide depleted (S.depleted).

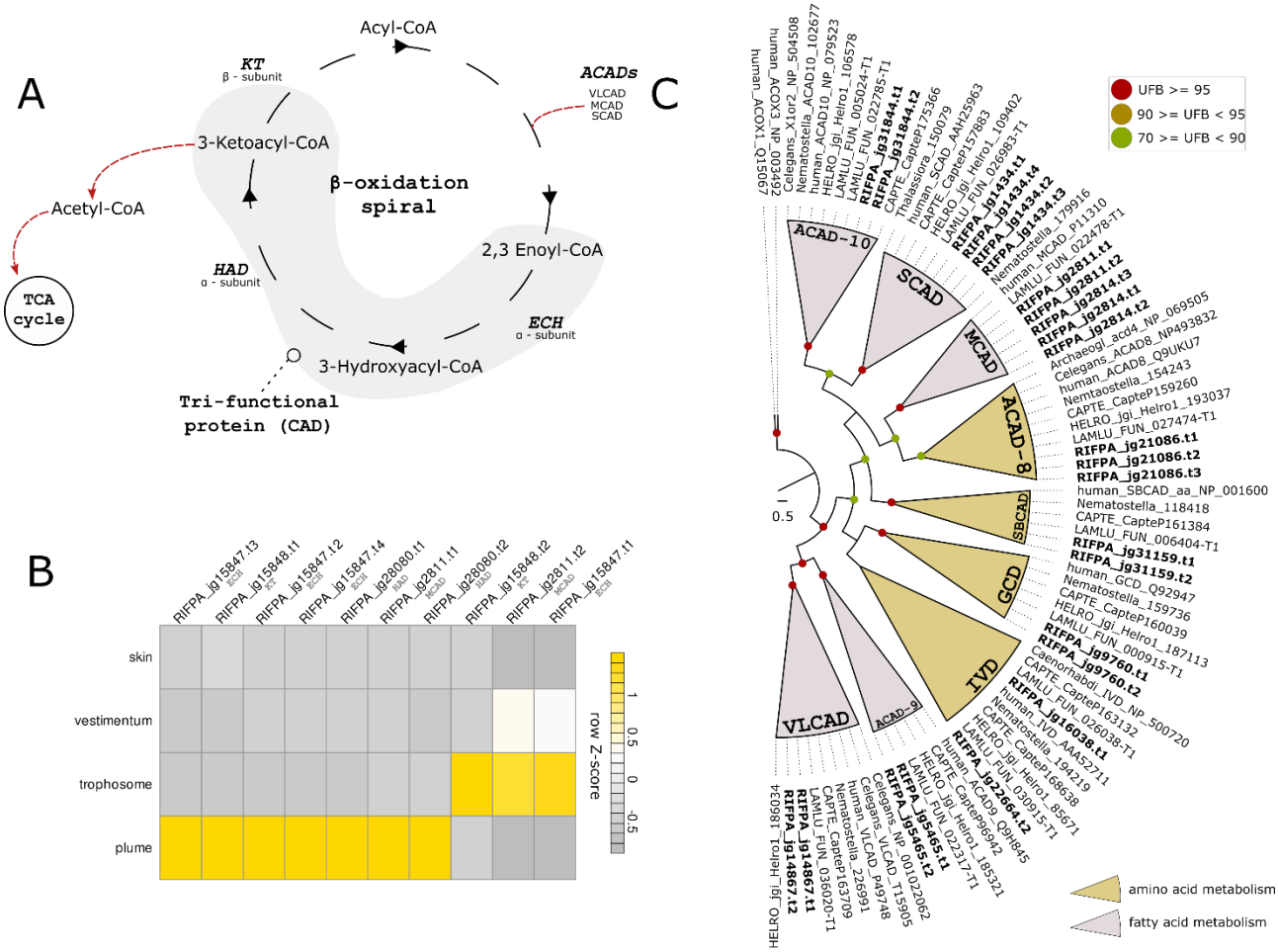


Figure 4 – Amino acid and fatty acid biosynthesis in *Riftia***A,** A schematic representation of mitochondrial fatty acid β-oxidation (FAO). The fatty acid degradation is performed in four

enzymatic steps involving the membrane bound mitochondrial trifunctional protein and acyl-CoA dehydrogenases. The resulting acetyl-CoA is further oxidised in the TCA cycle. **B**, Expression profile of FAO genes. Colour coding reflects the expression patterns based on row Z-score calculations. The FAO pathway is activated in the trophosome and plume tissues. **C,** Maximum-likelihood phylogenetic tree inference of the ACAD genes using 1000 rapid bootstrap replicates. The branch support values are represented by the coloured circles in the tree nodes. Accession numbers for NCBI database are displayed after the species names and homologs were retrieved from a previous study (Swigoňová et al. 2009). *Capitella, Helobdella* and *Lamellibrachia* gene identification are derived from the publicly available annotated genomes. **D**, Key enzymes related to amino acid biosynthesis identified in *Riftia*, selected annelids and two tubeworm endosymbiont genomes. Identification of the genes was performed with KEGG and reconfirmed with similarity searches against publicly protein dabases. *Riftia* and *Lamellibrachia* lack many amino acid biosynthesis genes indicating nutritional dependence on their endosymbionts. Stars represent genes present in the free-living polychaete *Capitella* and Endoriftia but absent in *Riftia* (based on Li et al. (2019) scheme).

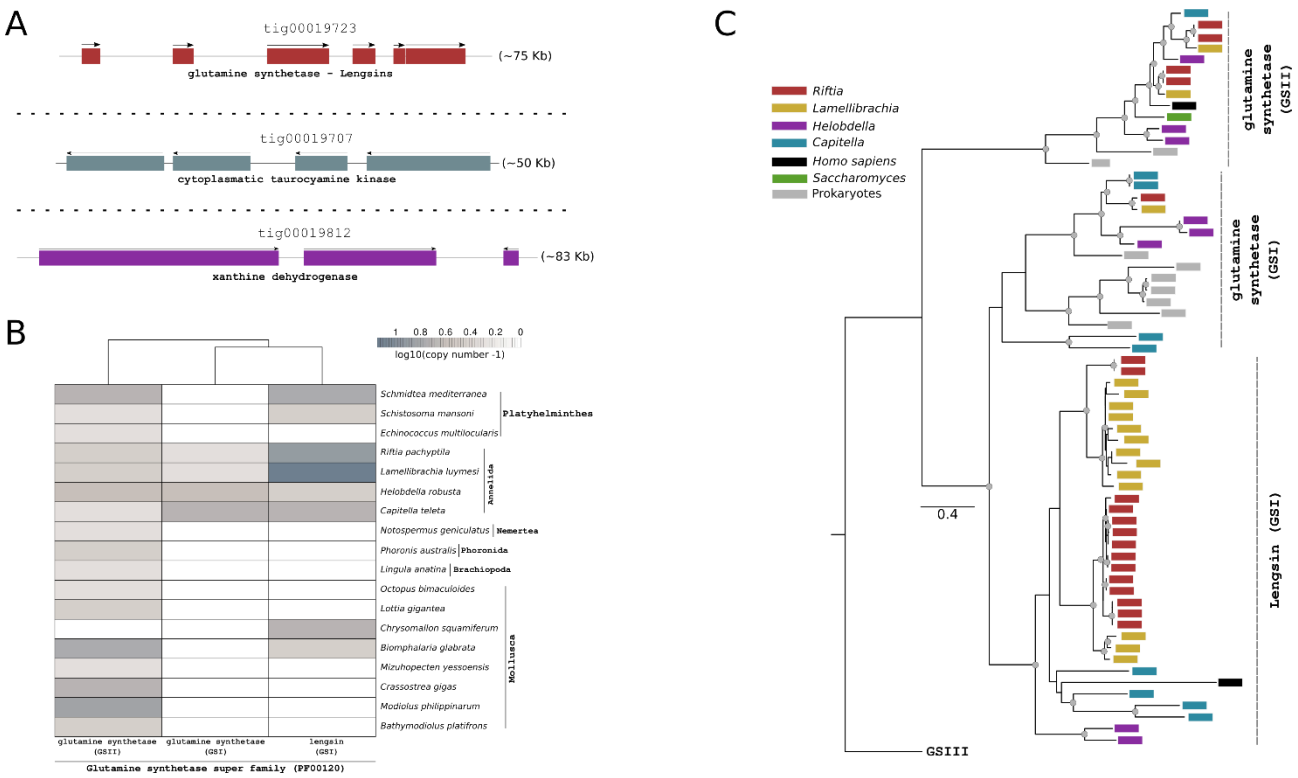


Figure 5 – Genomic clusters of important genes related to the nitrogen metabolism in *Riftia*, and distribution of phylogeny of glutamine synthetase genes among selected metazoans. **A,** Genomic organisation of important genes related to nitrogen metabolism in *Riftia*. Arrows indicate the direction of transcription. **B,** Distribution of glutamine synthase-related genes in the giant tubeworm, closely related annelids, and selected lophotrochozoans. **C,** Maximum likelihood

phylogenetic tree inference of members of the glutamine synthetase superfamily using 1000 ultrafast bootstrap replicates. Coloured boxes correspond to different annelids, vertebrates, yeast, and prokaryotes.

Additional Information

Supplementary tables

Supplementary Table 1 – *Riftia* genome and transcriptome pre-processing and annotation (.xlsx document). **A-B**, Overview of the *Riftia* PacBio libraries and genome statistics for the different pre-processed genome drafts. **C**, RepeatMasker results indicating the distribution of repeat elements in the *Riftia* genome. **D**, Overview of the eight tissue-specific *Riftia* libraries, trimming statistics and *de novo* assemblies. **E**, Mapping statistics of the individual tissue-specific transcriptomes using StringTie (full length transcripts x draft genome) and Bowtie2 (transcriptome library X *de novo* transcriptome) .**F**, Proteome prediction and BUSCO4 scores of the combined *de novo* and reference-based transcriptomes. The predicted proteomes were also mapped against the nr database.

Supplementary Table 2 – Databases used in the orthoFinder analysis (.xlsx document). **A**, Metazoan databases used in the orthoFinder analysis. **B**, Phyletic distribution of the orthogroups found in selected annelid genomes.

Supplementary Table 3 – Overview and quantification of transcription factor families in selected metazoans. Protein domain identification was performed with pfamscan. General overview of the transcription factors identified in annelids, molluscs, flatworms, phoronids, brachiopods and nemerteans.

Supplementary Table 4 – Gene family analysis with CAFE. **A**, Average expansions rates calculated by CAFE using lophotrochozoan orthogroups (N=18). Annelid values are highlighted in light red. **B-D**, Expanded, contracted and rapidly evolving gene families identified by CAFE in the giant tubeworm genome. Genes were annotated with Panther scoring tool. GO enrichment analyses using lineage-specific genes were performed with topGO in the three distinct groups. The enriched GO terms found in the three main ontologies are shown (Biological process, molecular function, and cellular component). In **D**, light purple and light red rows indicated expanded and contracted rapidly evolving orthogroups, respectively.

Supplementary Table 5 – Lineage-specific gene annotation and GO enrichment analyses. **A-C**, *Riftia*-, *Lamellibrachia*- and Siboglinidae-specific genes obtained through orthology analyses. Genes were annotatated with Panther scoring tool. GO enrichment analyses using lineage-specific genes were performed with topGO in the three distinct groups. The enriched GO terms found in the three main ontologies are shown (Biological process, molecular function, and cellular component).

Supplementary Table 6 – Contracted and expanded PFAM analysis in selected lophotrochozoans using two-tailed Fisher's exact test with Bonferroni correction. **A-D**, PFAM domain quantification using four different sets of organisms. **A-** All lophotrochozoans and PFAM domains. **B –** All lophotrochozoans without transposase- and DUF-associated domains. **C** – All lophotrochozoans without transposase- and DUF-associated domains, except the siboglinids *Riftia* and *Lamellibrachia*. **D-** All lophotrochozoans without transposase- and DUF-associated domains, except deep-vent symbiotic animals. **E1-2**, Two-tailed Fisher's exact test with Bonferroni correction using PFAM domains of *Riftia*/*Lamellibrachia,* and *Riftia*/average non-siboglinid lophotrochozoans pairs. **F,** Two-tailed Fisher's exact test with Bonferroni correction using PFAM domains of *Lamellibrachia* and average non siboglinid lophotrochozoans. **G1-4**, Pairwise two-tailed Fisher's exact test with Bonferroni correction between deep-sea symbiotic animals (*Riftia*, *Lamellibrachia*, *Bathymodiolus*, *Chrysomallon*) and the average non-deep-sea-symbiotic lophotrochozoans. **H**, Overlapping contracted/expanded domains found in symbiotic deep-sea symbiotic lophotrochozoans.

Supplementary Table 7 – Positively selected genes identified by HyPhy and KaKs calculator. **A-B**, Positively selected genes using two distinct methods. Gray rows correspond to positively selected genes identified in both methods. Annotation of positively selected genes was performed with Panther scoring tool.

Supplementary Table 8 – Gene expression quantification of selected proteins and protein families. **A-Y**, TPM (transcripts per million) values of selected proteins and protein families found in the eight different tissue-specific transcriptomes of *Riftia*. Colour scale (red – minimum; yellow – percentile 50; green -max) depicts the TPM values of genes found in the different *Riftia* tissues.

Supplementary Table 9 – Tau specific genes and GO enrichment analyses for the female tissue-specific transcriptomes (.xlsx document). Annotation of tau genes was performed with Panther

scoring tool. GO enrichment analyses using tau specific genes were performed with topGO in the three distinct groups. The enriched GO terms found in the three main ontologies are shown (Biological process, molecular function, and cellular component).

Supplementary Table 10 – Distribution of important enzymes related to the biosynthesis of amino acids in *Riftia, Lamellibrachia* and their endosymbionts. For comparison other two annelids were included in the analysis (*Capitella*, *Hellobdela*). **A,** Distribution of key enzymes related to amino acids biosynthesis based on the KEGG pathways. Colour scale (red – minimum; yellow – percentile 50; green -max) depicts the number of genes found in the different annelid genomes. **B,** Gene identifier of key enzymes related to the biosynthesis of amino acids in the *Capitella*, *Helobdella*, *Riftia* and *Lamellibrachia* genomes. **C**, Distribution of Endoriftia genes found in the secretion system type II pathway, as presented in KEGG.

Supplementary Table 11 – Mitochondrial carrier proteins identified in the *Riftia* genome and their PANTHER and blastp annotations. Genes highlighted in grey are highly expressed in the trophosome tissue.

Supplementary Table 12 – SRA accession numbers for the genomic and transcriptomic data generated in this study.

Supplementary files

Supplementary File 1: Supplementary information (.pdf).

Supplementary File 2: Rscript files, CAFE codes, *Riftia´s* gene model created with Augustus and the giant tubeworm repeat database generated with RepeatModeler and RepeatMasker(.zip).