BMC Genomics

Organ transcriptomes of the lucinid clam *Loripes orbiculatus* (Poli, 1791) provide insights into their specialised roles in the biology of a chemosymbiotic bivalve

Benedict Yuen[*] , Julia Polzin and Jillian M. Petersen

Abstract

Background: The lucinid clam *Loripes orbiculatus* lives in a nutritional symbiosis with sulphur-oxidizing bacteria housed in its gills. Although our understanding of the lucinid endosymbiont physiology and metabolism has made significant progress, relatively little is known about how the host regulates the symbiosis at the genetic and molecular levels. We generated transcriptomes from four *L. orbiculatus* organs (gills, foot, visceral mass, and mantle) for differential expression analyses, to better understand this clam's physiological adaptations to a chemosymbiotic lifestyle, and how it regulates nutritional and immune interactions with its symbionts.

Results: The transcriptome profile of the symbiont-housing gill suggests the regulation of apoptosis and innate immunity are important processes in this organ. We also identified many transcripts encoding ion transporters from the solute carrier family that possibly allow metabolite exchange between host and symbiont. Despite the clam holobiont's clear reliance on chemosynthesis, the clam's visceral mass, which contains the digestive tract, is characterised by enzymes involved in digestion, carbohydrate recognition and metabolism, suggesting that *L. orbiculatus* has a mixotrophic diet. The foot transcriptome is dominated by the biosynthesis of glycoproteins for the construction of mucus tubes, and receptors that mediate the detection of chemical cues in the environment.

Conclusions: The transcriptome profiles of gills, mantle, foot and visceral mass provide insights into the molecular basis underlying the functional specialisation of bivalve organs adapted to a chemosymbiotic lifestyle.

Keywords: Chemosynthetic symbiosis, RNA-Seq, Bivalve, Lucinid, Innate immunity, Nutritional symbiosis

Background

Associations between invertebrates and chemoautotrophic bacteria contributed to the evolutionary success of diverse animal lineages, and are also fundamental to the functioning of marine ecosystems [1, 2]. One prominent example is chemosynthetic symbioses that are found in at least seven phyla of metazoan invertebrates [2]. In the deep sea, these symbioses underpin primary productivity that supports an unexpectedly large biomass (hydrothermal vent mussels and tubeworms) in a food scarce environment, thus serving as the foundation of entire ecosystems [1, 2]. Invertebrates that host

chemoautotrophic bacteria are also ubiquitous in shallow water habitats and lucinid clams, in particular, are highly abundant members of the macrofaunal community in shallow water sediments. Indeed, lucinids are the most species rich and widely distributed group of chemosynthetic bivalves [3], and their ecological and evolutionary success can be attributed to their highly specific association with a single phylotype of sulphur-oxidising gammaproteobacteria.

The symbiosis between lucinid clams and sulphur-oxidising bacteria is a nutritional strategy; the symbionts use reduced sulphur compounds – produced by organic matter decomposing in anoxic sediment – as electron donors and autotrophically fix carbon dioxide to synthesise organic compounds that are transferred to the host [1].

* Correspondence: yuen@microbial-ecology.net
Centre for Microbiology and Environmental Systems Science, Division of Microbial Ecology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria

Lucinid clams possess multiple atypical morphological traits associated with the peculiar nature of this nutritional strategy. To house significant bacterial biomass, the gills of lucinids are modified into a plump and conspicuous symbiont-housing organ with a large surface area extending beyond the limits of the visceral mass [4, 5]. The gill symbionts are enclosed in vacuoles of specialised cells known as bacteriocytes and the gills can constitute up to a third of the host's weight in soft tissue [6]. This heavy reliance on symbionts for nutrition is reflected in the morphological simplification of the lucinid feeding structures and digestive system; poorly developed labial palps, a truncated stomach and digestive glands with a reduced number of tubules [4]. The lucinid foot is also modified to become vermiform and highly extensible (up to 5 times body length) [4]. In addition to locomotory and burrowing activities, the foot is responsible for building a tube connecting the clam to the sediment-water interface, which is a functional replacement for the loss of their inhalant siphon [4]. The lucinid foot also constructs tubes burrowing into the sediments below the animal to mine pockets of dissolved sulphide associated with decaying organic matter or the insoluble iron sulphides attached to sediment, in order to support the sulphide-requiring metabolism of their endosymbionts [7]. These unusual morphological adaptations to a chemosymbiotic lifestyle have allowed lucinids to successfully exploit the deep layers of sulphide-rich sediment, an ecological niche inaccessible to most other animals.

A wealth of literature documents the taxonomy and ecology of lucinid symbioses [4, 8–10], and we are beginning to unravel metabolic and physiological capabilities of the endosymbionts [11–13]. However, we still have little understanding of how the host regulates symbiotic interactions at the molecular level. Previous studies suggest a high degree of control over symbionts in the gills, as lucinids appear to inhibit symbiont cell division and are able to selectively digest symbionts or reacquire the same strain from the environment [14–16]. These studies suggest lucinid-symbiont crosstalk is not limited to nutrient exchange and may be regulated by other processes including host immunity. Indeed, immunity has been implicated in regulating interactions between deep sea mussels and the endosymbiotic bacteria harboured in their gills [17–19]. Additionally, non-symbiotic lucinid organs can also influence lucinid-symbiont interactions by controlling the availability of essential resources required for metabolism. Examples include filter feeding and digestion in the digestive tract [20–22], and sulphide and oxygen acquisition by the foot [7]. These examples show that the functions and physiology of the non-symbiotic host organs also contribute to the functioning of the chemosynthetic holobiont. Therefore, transcriptional profiles of symbiotic and non-symbiotic organs will

provide more complete insights into how the lucinid host regulates this symbiosis.

We de novo assembled and annotated a reference transcriptome using RNA-Seq reads from four different organs (gills, mantle, visceral mass, and foot) of *Loripes orbiculatus* (Fig. 1 a), a lucinid clam widely distributed throughout the Mediterranean Sea. This served as the basis for differential gene expression analyses and the subsequent characterisation of organ-specific transcriptome profiles. We discuss the functional implications of organ-specific gene clusters, focusing on the symbiont-housing gills and the contribution of the non-symbiotic organs to regulating lucinid-symbiont interactions. These data are a valuable addition to the existing genomic resources for studying the lucinid symbiosis and have the potential to provide new insights into the molecular basis of animal-bacteria interactions.
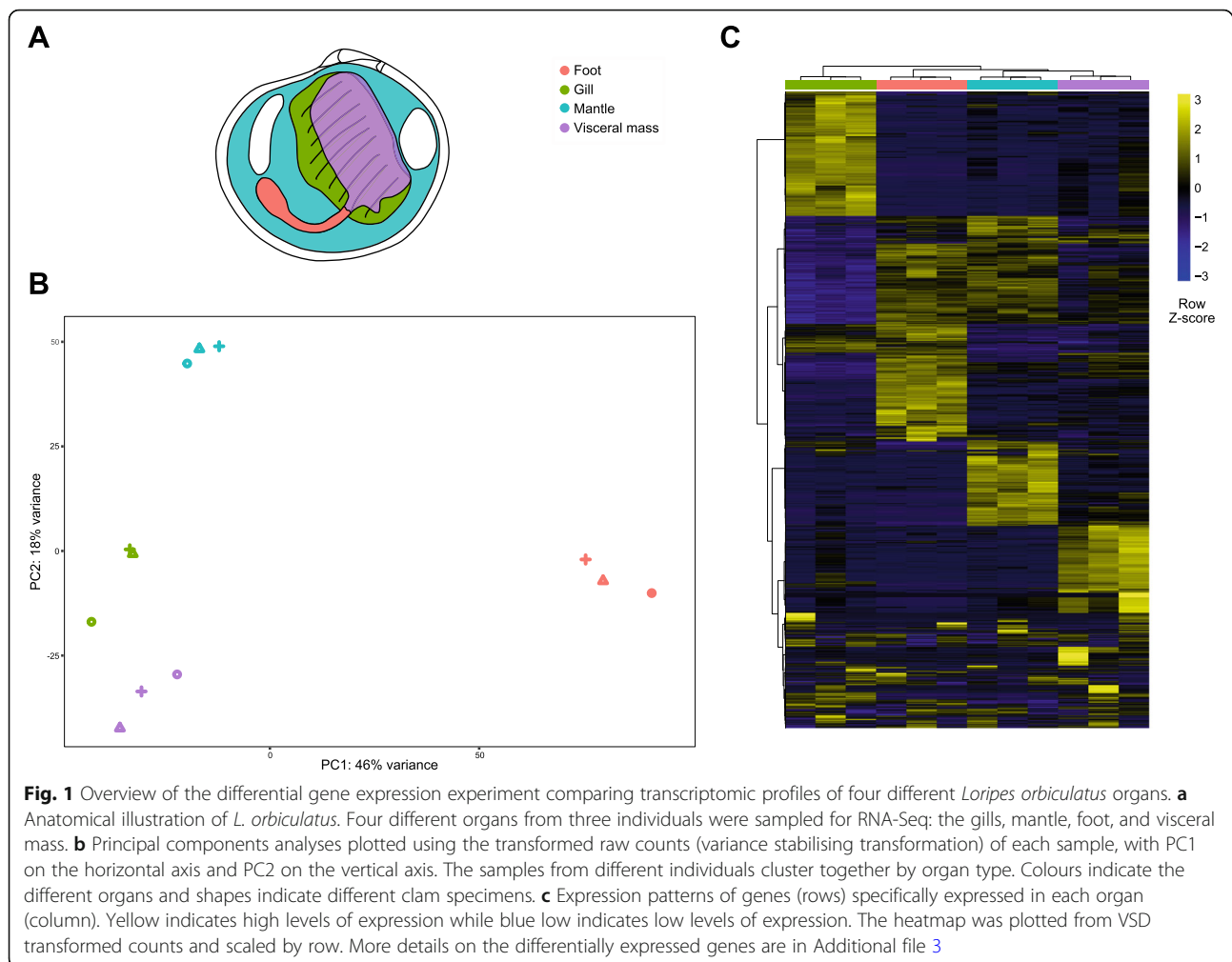
Results & discussion
Assembly of a comprehensive lucinid transcriptome
We sequenced 12 RNA-Seq libraries from three replicates each of four organ types: the gills, foot, mantle, and visceral mass. After error correction, filtering, and trimming, each library comprised an average of 22,186,325 million read pairs that were de novo assembled into transcriptomes. These were combined with previously sequenced gill RNA-Seq libraries into a reference transcriptome of 104, 568 contigs with an N50 of 1365 bp, and 1996 contigs above 1 kbp in size (assembly statistics in Additional file 1). Analysis with BUSCO revealed 99% (97.1% complete and 1.8% partial) of the metazoan BUSCO gene set to be present; a higher level of completeness than most other published bivalve transcriptomes [23, 24]. This transcriptome is expected to be a valuable addition to the existing genetic resources for studying the molecular underpinnings of lucinid symbioses [11, 12]. Despite the large number of contigs in the final assembly, only 2.5% of the BUSCOs were identified as duplicates while 20,163 transcripts were functionally annotated with gene ontology terms. This could be due to a number of factors that might include a high frequency of uncharacterised lineage-specific genes and non-coding transcripts. Indeed, this large number of predicted gene models is not unprecedented as similar trends have been reported in other members of the Imparidentia; the draft genomes of *Ruditapes phillippinarum* and *Dreissena polymorpha* are predicted to encode over 100, 000 and 60, 000 genes, respectively [25, 26].

Each organ is characterised by a unique transcriptome profile
We used this highly complete transcriptome as a reference to test for genes differentially expressed amongst the gill, visceral mass, mantle and foot (Fig. 1). Identifying clusters

Fig. 1 Overview of the differential gene expression experiment comparing transcriptomic profiles of four different *Loripes orbiculatus* organs. **a** Anatomical illustration of *L. orbiculatus*. Four different organs from three individuals were sampled for RNA-Seq: the gills, mantle, foot, and visceral mass. **b** Principal components analyses plotted using the transformed raw counts (variance stabilising transformation) of each sample, with PC1 on the horizontal axis and PC2 on the vertical axis. The samples from different individuals cluster together by organ type. Colours indicate the different organs and shapes indicate different clam specimens. **c** Expression patterns of genes (rows) specifically expressed in each organ (column). Yellow indicates high levels of expression while blue low indicates low levels of expression. The heatmap was plotted from VSD transformed counts and scaled by row. More details on the differentially expressed genes are in Additional file 3

of organ-specific genes provides informative baseline data on the functions and physiology of each organ, and thus how both symbiotic and non-symbiotic organs may contribute to the functioning of the lucinid holobiont. A principal component analysis (PCA) of the variance stabilised dataset showed clear separation amongst the organs without any overlaps, indicating each organ has a clearly distinct transcriptomic profile (Fig. 1 b). This is supported by differential gene expression analyses, which showed that 1917, 1064, 823, and 999 corset unigenes were most highly expressed in the gill, visceral mass, mantle, and foot, respectively (Fig. 1 c). We then used gene set enrichment analyses of gene ontology (GO) terms and PFAM domains, as well as manual curation to gain deeper insights into the functional implications of these organ-specific gene clusters.

## Immunity and cell death are important processes in the symbiont-housing organ

To identify candidate immune genes involved in mediating lucinid clam-symbiont crosstalk, we used KEGG term annotation and conserved Interpro domains to analyse the reference transcriptome for metazoan immune pathways and receptors. The key components of the main immune signalling pathways conserved across most metazoan lineages were present in the transcriptome of *L. orbiculatus* (Additional file 2). We also investigated the repertoire of different pattern recognition receptor (PRR) gene families in the *L. orbiculatus* transcriptomes. PRRs of the evolutionarily ancient innate immune system detect Microbe-Associated Molecular Patterns, the non-self molecules that illicit the animal immune response [27, 28]. All major intracellular and extracellular PRR families typically observed in bivalve genomes are expressed by *L. orbiculatus* and we found no evidence massive immune gene family shrinkage nor loss of immune signalling pathways (Additional file 2). Some of these PRR families have undergone expansions similar to those in non-symbiotic bivalve genomes: the Toll-like receptors (TLRs), Scavenger Receptor Cysteine-Rich domain-containing genes (SRCRs), C1q-domain containing genes, and C-terminal fibrinogen-related domain-containing proteins (FReDs) (Additional file 2).

Yuen *et al. BMC Genomics*   (2019) 20:820

Page 4 of 14

To gain insights into their functions, we analysed the expression patterns of these *L. orbiculatus* PRRs across the different organs and found that different PRR families have organ-specific expressions profiles. For example, more TIR domain containing unigenes (including TLRs) are more abundant in the gills than in other organs, while SRCRs and C-type lectins are more abundant in the mantle and visceral mass, respectively (Fig. 2 a). The organ specificity of these PRR families suggests that each organ represents a unique immunological environment with challenges specific to the physiology and function of each organ [29]. We then focussed on the symbiont-housing gills to identify candidate innate immune genes potentially involved in regulating lucinid-symbiont interactions. An array of unigenes encoding diverse PRRs are more abundant in the gills, including eight TIR domain-containing (PF01582) unigenes, four SRCRs,

one Peptidoglycan Recognition Protein, and three C-type lectins (Additional file 3). Based on their conserved structural organisation, seven of the nine TIR domain-containing unigenes can be considered true TLRs (Fig. 2 b). One of these is a "twin-TIR TLR" marked by the presence of two intracellular TIR domains, an unusual structure previously reported only in other bivalves [30]. The conserved role of TLRs in mediating beneficial host-microbe interactions across metazoan lineages [27], in conjunction with their potential to be localised on endosomal membranes, makes the TLRs excellent candidates for mediating lucinid-symbiont crosstalk.

TLRs recruit proteins downstream to transduce their signals through the dimerization of the cytosolic TIR domain with the TIR domain of adaptor proteins [31]. The remaining TIR domain-containing unigene highly expressed in the gills is unlikely to mediate MAMP recognition
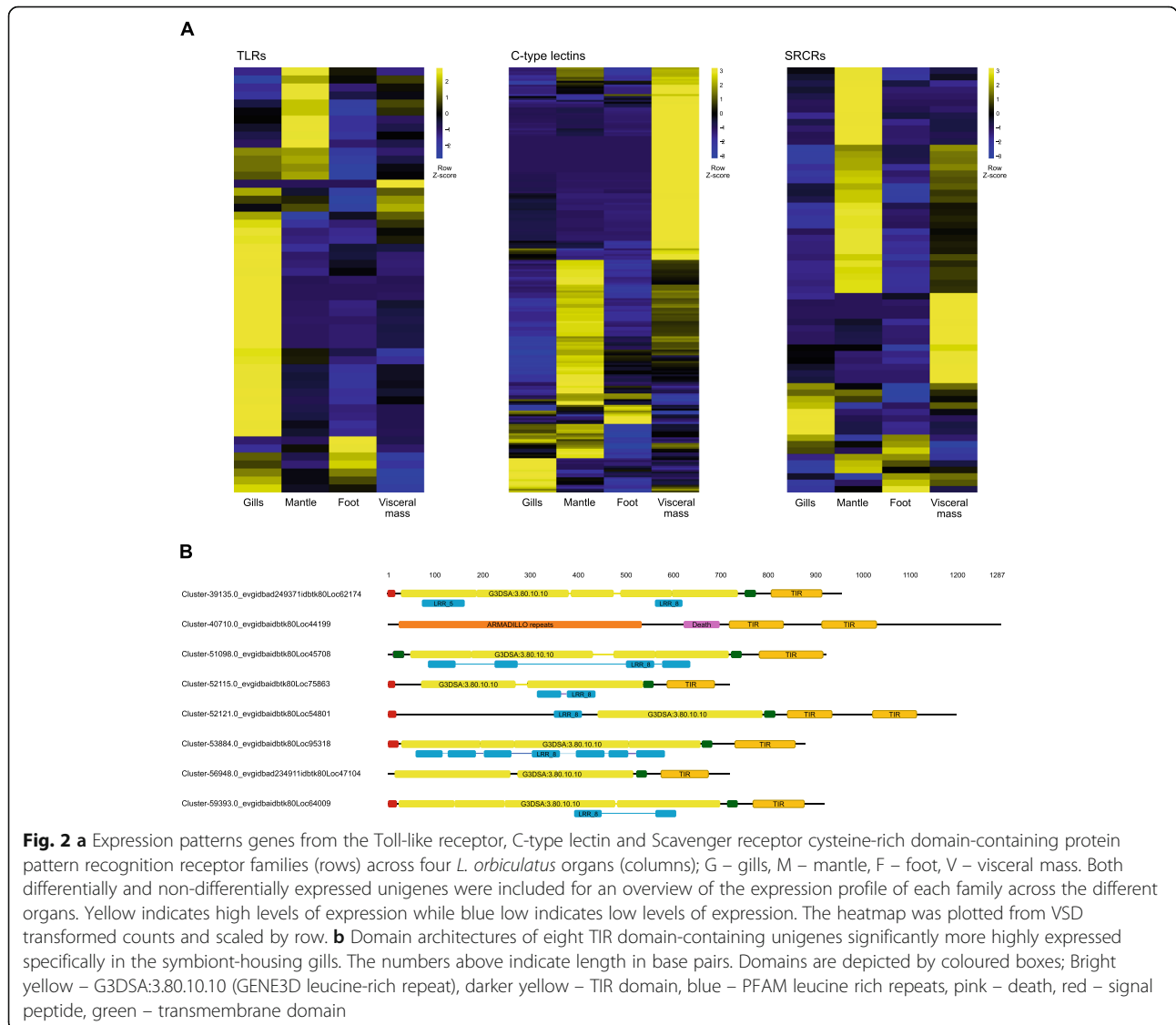
Fig. 2 a Expression patterns genes from the Toll-like receptor, C-type lectin and Scavenger receptor cysteine-rich domain-containing protein pattern recognition receptor families (rows) across four *L. orbiculatus* organs (columns); G – gills, M – mantle, F – foot, V – visceral mass. Both differentially and non-differentially expressed unigenes were included for an overview of the expression profile of each family across the different organs. Yellow indicates high levels of expression while blue low indicates low levels of expression. The heatmap was plotted from VSD transformed counts and scaled by row. **b** Domain architectures of eight TIR domain-containing unigenes significantly more highly expressed specifically in the symbiont-housing gills. The numbers above indicate length in base pairs. Domains are depicted by coloured boxes; Bright yellow – G3DSA:3.80.10.10 (GENE3D leucine-rich repeat), darker yellow – TIR domain, blue – PFAM leucine rich repeats, pink – death, red – signal peptide, green – transmembrane domain

because it lacks the conserved organisation of TLRs and is further predicted to be localised to the cytosol as evidenced by the absence of a transmembrane domain or signal peptide motif (Fig. 2 b). However, it could regulate lucinid-symbiont interactions by transducing signals from TLRs through their TIR domains. This gene comprises a long N-terminal region containing ARMADILLO repeat-like motifs, a central Death domain, and two N-terminal TIR domains (Fig. 2 b); an unusual domain composition characteristic of the ecTIR-DC family 6 genes reported only in other invertebrates [30]. The Death and TIR domain combination is characteristic of MyD88, the adaptor protein that transduces signals from ligand-bound TLRs [31], while ARMADILLO repeats are also found on proteins that regulate immune signalling [32]. The functions of these domains and the co-expression of this unigene with TLRs in the gills leads us to speculate it could be involved in transducing TLR signals.

Biological Process (BP) GO terms associated with apoptosis regulation are enriched amongst the unigenes expressed in the gills (Fig. 3). One of these encodes a tumour necrosis factor receptor (TNFR) with an intracellular Death domain (Fig. 4), which is a structure typical of the type I TNFRs responsible for activating apoptosis and inflammatory signalling in mammals [33]. However, the majority of these apoptosis-related unigenes simply contain CARD (PF00619) or Death domains (PF00531), which are usually associated with adaptor proteins that mediate cell death signalling by recruiting caspases through homotypic interactions [34]. The importance of regulating apoptosis in the gills is further emphasised by the expression of unigenes encoding B-cl2-like and Bax-inhibitor 1-like proteins that supress cell death (Fig. 4, Additional file 3) [35, 36]. It is noteworthy that an expanded repertoire of Death domain-containing genes was reported in the genome of *Bathymodiolus platifrons*, a deep sea mussel that also hosts chemoautotrophic endosymbionts [19]. Interestingly, TUNEL staining of the gill tissues of lucinid and chemosynthetic deep sea mytilids indicates cell death is prevalent in the ciliated zone of both lucinid and symbiotic deep sea mytilids but not in the bacteriocyte zone [37, 38]. These lines of evidence suggest that apoptosis regulation is a key process in the symbiont housing organs of chemosynthetic bivalves in general. High rates of chemoautotrophic metabolism imposes an oxygen demand upon animal hosts that is typically much higher than for its non-symbiotic relatives [39]. Therefore, one possible explanation is that the high prevalence of apoptosis in ciliated cells and enrichment of apoptosis-related GO terms in the lucinid gills are due to a high cell turnover rate caused by greater oxidative stress. However, recent work has shown that genes involved in regulating apoptosis are more highly expressed in bathymodiolin mussels when their symbiont load is experimentally depleted [40], which suggests the role of

apoptosis in chemosynthetic symbioses may be more complicated.

The expression of these diverse molecules involved in various arms of innate immunity suggests cells in the lucinid gills have the capacity to detect and raise a complex response to bacteria. We cannot rule out the possibility that the PRRs, immune signalling and effector molecules expressed in the gills are used to defend the host against harmful bacteria because the gills are constantly exposed to the water from the environment. An important caveat to note is that the KEGG immune signalling pathway annotations are biased towards classical model organisms and little conclusive functional evidence exists to support the involvement of proteins containing the domains discussed (e.g. C-type lectin, Death and SRCR) in bivalve immunity. Nevertheless, these data allow to shortlist candidates for future studies investigating the role of innate immunity in lucinid-symbiont crosstalk as well as candidates for investigating the roles of cell death in regulating endosymbiont abundance or bacteriocyte turnover dynamics.

Unigenes expressed in the gill facilitate the chemoautotrophic metabolism of the endosymbiotic bacteria

The beating of the lateral cilia on bivalve gills generates the water currents that facilitate gas exchange and filter-feeding, and previous studies indicate this activity is under neurohormonal control [41]. This is consistent with the enrichment of BP GO terms associated with ion transmembrane transport in the gills, because half of the unigenes contributing to this GO term encode ion channels typical of the animal nervous system known as nicotinic acetylcholine receptors (Fig. 3; Additional file 3). Unigenes associated with the dynein complex are also enriched amongst the cellular component GO terms (Additional file 4); dyneins are critical to ciliary movement because they convert ATP to mechanical motion [42]. Nicotinic acetylcholine receptors and dynein complex genes are similarly over-represented in the gills of non-chemosymbiotic bivalves [23], which indicates that these genes are unlikely to be directly involved in the unique symbiont housing role of lucinid gills. Nevertheless, their functions influence host-symbiont interactions because the water flow driven by ciliary beating ensures symbionts are supplied with the oxygen and sulphide necessary for chemoautotrophic metabolism.
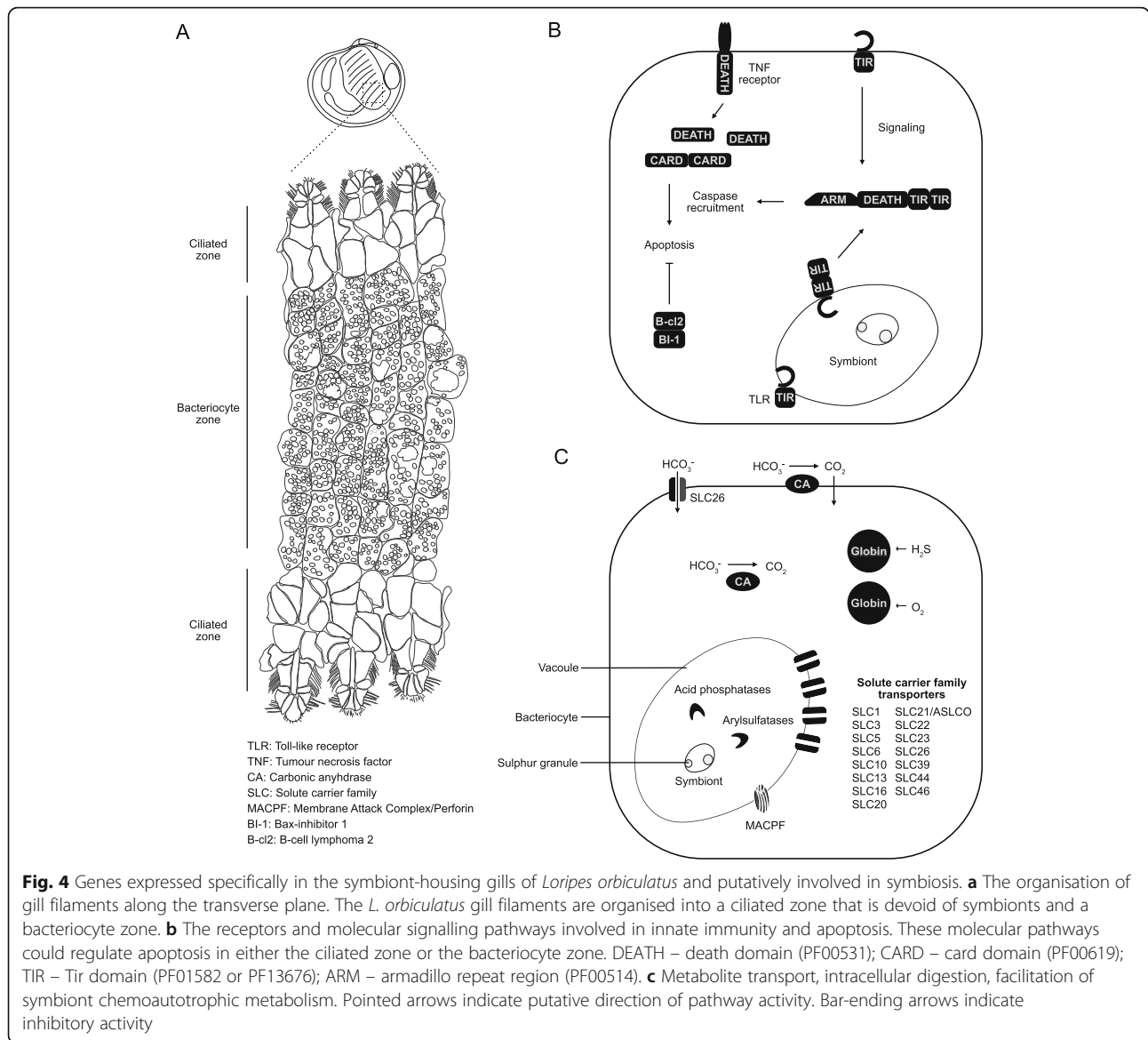
In order to supply their symbionts with resources for chemoautotrophic metabolism, the haemoglobins of thioautotrophic invertebrates have evolved to bind sulphide and oxygen while preventing the rapid oxidation of sulphides [39]. Two globin domain-containing unigenes are expressed in the gills, one of which groups with the sulphide-binding hemoglobin HBI of the lucinid
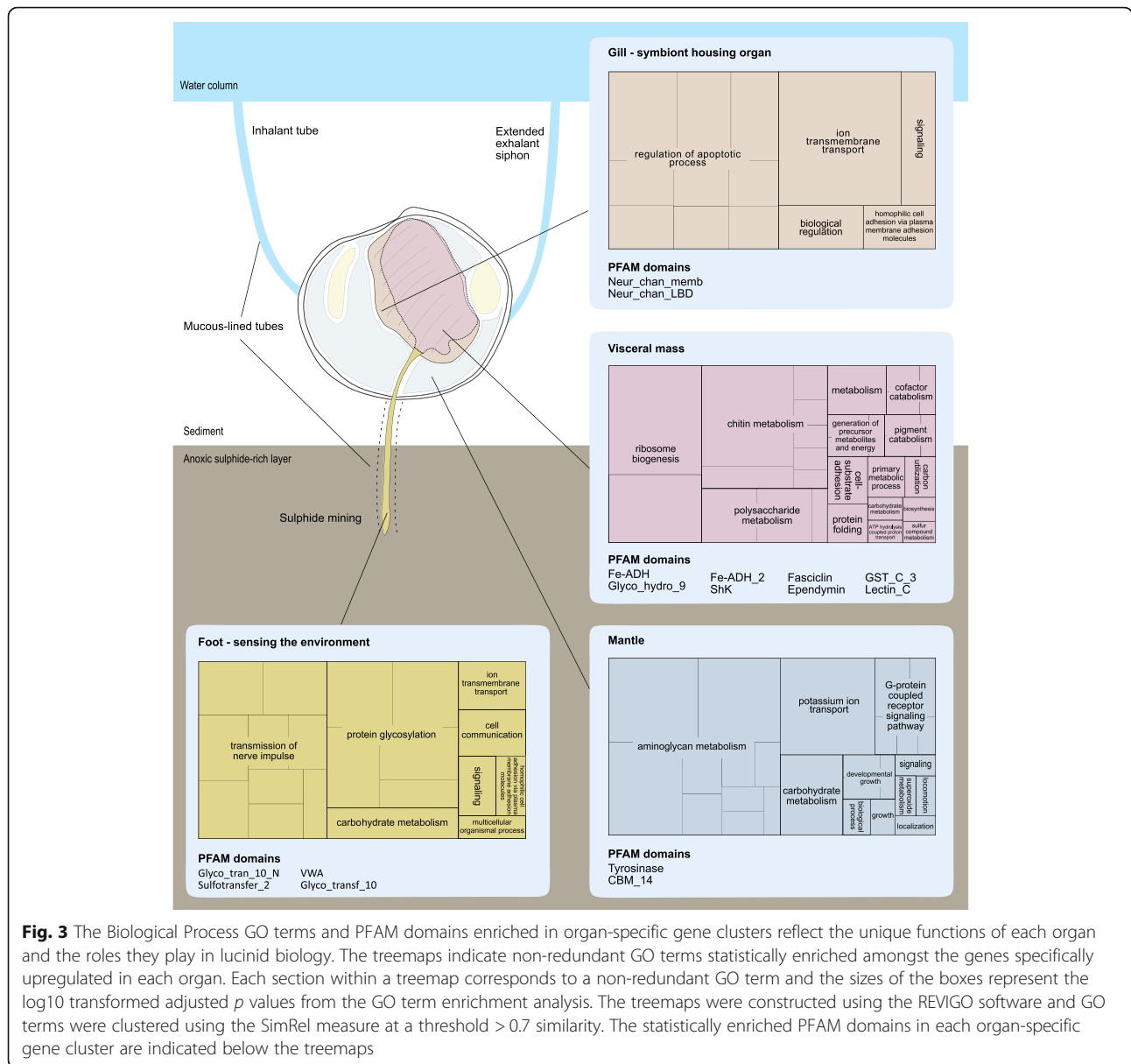
Fig. 3 The Biological Process GO terms and PFAM domains enriched in organ-specific gene clusters reflect the unique functions of each organ and the roles they play in lucinid biology. The treemaps indicate non-redundant GO terms statistically enriched amongst the genes specifically upregulated in each organ. Each section within a treemap corresponds to a non-redundant GO term and the sizes of the boxes represent the log10 transformed adjusted *p* values from the GO term enrichment analysis. The treemaps were constructed using the REVIGO software and GO terms were clustered using the SimRel measure at a threshold > 0.7 similarity. The statistically enriched PFAM domains in each organ-specific gene cluster are indicated below the treemaps

Phacoides pectinatus, while the other forms a clade with *P. pectinatus* HBII and HBIII, which are responsible for oxygen binding (Additional file 5). These findings suggest that like other invertebrates associated with thioautrophic symbionts, *L. orbiculatus* uses specialised haemoglobins to supply its symbionts with sulphide and oxygen. Through the oxidation of sulphur, endosymbionts produce energy that is used to fixed inorganic carbon in the form of CO_2 into sugars [1]. However, dissolved CO_2 levels are generally low in seawater and the spontaneous generation of CO_2 from HCO_3^- is minimal due to the high alkalinity of sea water. To solve this physiological challenge, all known thioautotrophic symbioses rely on carbonic anhydrases (CAs) to catalyse the reversible dehydration of bicarbonate to carbon dioxide,

which diffuses easily across cell membranes and facilitates inorganic carbon uptake [43]. Two CA unigenes are expressed in the gills (Fig. 4, Additional file 3, Additional file 4), both of which have signal peptides that indicate they are secreted. However, the larger of the two has a predicted GPI anchor site (Serine position 432) that could mediate the attachment of the protein to the cell membrane (Additional file 5). A similar secreted CA bearing a GPI anchor site is expressed in bacteriocytes of the chemosynthetic vesicomyid clam *Phreagena okutanii* and is probably responsible for catalysing the conversion of HCO_3^- in the hemolymph into CO_2 that can diffuse into the bacteriocytes [44]. The CAs in bivalves regulate multiple physiological functions including respiration, calcification and mineralization [45, 46]. However, in addition

Fig. 4 Genes expressed specifically in the symbiont-housing gills of *Loripes orbiculatus* and putatively involved in symbiosis. **a** The organisation of gill filaments along the transverse plane. The *L. orbiculatus* gill filaments are organised into a ciliated zone that is devoid of symbionts and a bacteriocyte zone. **b** The receptors and molecular signalling pathways involved in innate immunity and apoptosis. These molecular pathways could regulate apoptosis in either the ciliated zone or the bacteriocyte zone. DEATH – death domain (PF00531); CARD – card domain (PF00619); TIR – Tir domain (PF01582 or PF13676); ARM – armadillo repeat region (PF00514). **c** Metabolite transport, intracellular digestion, facilitation of symbiont chemoautotrophic metabolism. Pointed arrows indicate putative direction of pathway activity. Bar-ending arrows indicate inhibitory activity

to these other physiological roles, we hypothesise that the CAs expressed in the gills of *L. orbiculatus* could be localised to the bacteriocyte cell membrane or secreted into the gill hemolymph to make inorganic carbon available of to the endosymbionts. An alternative but non-mutually exclusive route for providing the symbionts with inorganic carbon is to transport HCO_3^- directly into the bacteriocytes. Diatoms rely on the SLC4 family of carbonate anion transporters to obtain the inorganic carbon required for photosynthesis [47]. Unigenes encoding SLC4 were not more abundant in the gills compared to the other organs analysed, but we identified a unigene from the SLC26 family of transporters more highly expressed in the gills (Fig. 4, Additional file 4, Additional file 5). This SLC26 unigene has a highly conserved domain architecture and forms a clade with *Drosophila melanogaster* and *Anopheles*

gambiae genes encoding Prestin (Additional file 5), a protein with the capacity to transport bicarbonate, oxalate, and sulphate anions [48]. The role of SLC26 in transporting bicarbonate into the bacteriocytes is worth further investigation as the symbiont genome encodes multiple carbonic anhydrases that could convert bicarbonate into CO_2 in bacteriocytes [12].

The lucinid metabolism relies on both nutrient translocation and heterotrophy

Previous studies have inferred that lucinids acquire nutrients from their symbionts through a carbon transfer strategy known as "milking", whereby nutrients are 'leaked' or translocated to the host following carbon fixation [49]. This was supported by microscopic evidence indicating low rates of symbiont digestion in the gills

and that symbiont cell division is inhibited [15, 50]. Expressed in the gills are 20 unigenes encoding transporters belonging the to the Solute Carrier (SLC), and folate (Additional file 5). Indeed, the SLC family has been implicated in regulating nutrient transfer in family that contribute to the enrichment of BP GO terms associated with ion transmembrane transport (Fig. 4, Additional file 5). There are a total 46 SLC families that are united by a common mode of function – they use the ion gradients across cell membranes to drive the transmembrane transport of solutes. Fifteen different SLC families are expressed in the gills, with specificities for a diversity of substrates including amino acids, glucose, phosphate, monocarboxylates, organic cations and anions, zinc, ascorbic acidother nutritional symbioses. The SLC5 and SLC46A3 families are thought to regulate nutrient transfer between cnidarians and endosymbiotic algae [51, 52], while members of SLC families 5, 6, 13, and 16 are enriched in the symbiont-housing root of the bone eating worm *Osedax japonicas* [53]. The diverse transporters expressed in the *L. orbiculatus* gill therefore likely facilitate lucinid-endosymbiont nutrient exchange and supports the inference that the host obtains nutrients through "milking", although intracellular digestion may also play a role (see below).

We also identified unigenes from other transporter superfamilies in the gills, including monocarboxylate and organic cation transporters from the Major Facilitator Superfamily, and a DUR3-like unigene encoding a urea-proton symporter (Fig. 4, Additional file 5). DUR3 is thought to be involved in nitrogen transfer between invertebrate hosts (Hydra and the giant clam) and their algal symbionts [54, 55], but the expression of a urea transporter in the lucinid gills is somewhat unexpected because *Candidatus* Thiodiazotropha endoloripes, the endosymbiont of *L. orbiculatus*, has the capacity to fix nitrogen. Nevertheless, the endosymbiont genome also encodes the transporters and enzymes (Additional file 5) necessary for metabolising urea [12], which suggests the *L. orbiculatus* DUR3 could be used to provide symbionts with a nitrogen source in the form of urea. The recycling of nitrogenous compounds like urea could be a symbiont strategy for coping with nitrogen-limitation while also helping the host remove metabolic waste. An alternative possibility is that the lucinid host could play an important role in controlling symbiont nitrogen metabolism. Indeed, the pea plant prevents rhizobium from assimilating ammonium by providing the bacteria with amino acids, thereby regulating symbiotic nitrogen fixation [56]. It is thus interesting to speculate that metabolite transporters could also serve as a mechanism for controlling symbionts by regulating access to nutrients.

Metabolic flexibility in both *L. orbiculatus* and its symbionts is likely important in the coastal sediments of Elba where the waters are oligotrophic and the availability of sulphide fluctuates in time and space [57, 58]. Biochemical signals of acid phosphatase and arylsulfatase activity, both markers of lysosomal activity, were detected in freshly collected clams indicating that lucinids also directly digest their symbionts [50]. This is consistent with the expression of unigenes encoding proteins involved in lysosomal activity in the gills, including two acid phosphatases, two arylsulfatases, and one Cathepsin L (Fig. 4, Additional file 3). We also identified two unigenes from the Membrane Attack Complex/Perforin (MACPF) superfamily (PF01823) expressed in the gills. Proteins from this superfamily are pore forming toxins involved in immunity that form holes in the membrane of bacteria leading to cell lysis [59]. A MACPF gene in the oyster *Crassostrea gigas* is upregulated under bacterial exposure, localised to late endosomes, and is highly expressed in the gills, digestive gland and gonads [60]. While it is possible the lucinid MACPF proteins are involved primarily in immune defence, these genes could potentially be co-opted to facilitate the digestion of endosymbiont bacteria.

Gill endosymbiont digestion is, however, unlikely to play a significant role in nutrient acquisition except under exceptional circumstances such as prolonged sulphide limitation [50]. In fact, many more enzymes typically associated with metazoan digestion are highly expressed in the lucinid visceral mass rather than the gills (Additional file 3). This includes carboxypeptidases, papain and pepsin family aspartic peptidases such as cathepsins (D, F, and L), as well as chymotrypsin and subtilisin family serine peptidases (Additional file 3). The importance of digestion in the visceral mass is further reflected in the enrichment of BP GO terms with chitin and polysaccharide metabolism, as well as the Glycosyl hydrolase family 9 PFAM domain (PF00759) (Fig. 3). These expression patterns suggest the lucinid digestive system is able to digest complex polysaccharides that are typical constituents of phytoplankton in the marine environment [61]. This correlates with the enrichment of lectin C-type domains (PF00059) – corresponding to 25 C-type lectin unigenes – in the visceral mass, suggesting that carbohydrate recognition is important in the lucinid gut. C-type lectins are characterised by having one or more copies of the C-type lectin domain, and they fulfil diverse immunological and physiological roles through the recognition of an array of primarily sugar-based ligands [62]. In the mosquito gut, C-type lectins bind to microbes to enable antimicrobial peptide evasion, thereby allowing microbial colonisation and gut homeostasis [63]. It is likely that the lucinid digestive track is also associated with a thus far unstudied microbiota, and the visceral mass C-type lectins could thus play a role in gut immunity and

homeostasis. Alternatively, the lucinid C-type lectins could also regulate food sorting as in *Crassostrea gigas*, where they are expressed on the mucosal surfaces of feeding organs [64]. Regardless of their precise roles, these lines of evidence suggest that heterotrophy remains an important strategy for acquiring nutrients despite the morphological simplification of the lucinid digestive system and its symbiosis with chemosynthetic bacteria. Our findings thus provide molecular support for previous observations that lucinids are able to capture and ingest particulate organic matter [20].

The foot as a sensor and regulator of interactions with the environment

The BP GO terms enriched in the foot include transmission of nerve impulse, protein glycosylation and carbohydrate metabolism (Fig. 3). These GO terms correlate with the enrichment of Glycosyl transferase (PF17039 and PF00852), sulfotransferase (PF03567), and von wildebrand A (PF00092) PFAM domains in the foot (Fig. 3). VWA domains are typically found in glycoproteins that form the basis of mucus and are secreted into the extracellular matrix [65, 66]. These proteins are often post-translationally modified with glycan attachment in a process known as glycosylation involving enzymes called glycosyltransferases. Further post-translational modifications are carried out by various enzymes such as sulfotransferases and fucosyltransferases [67]. Therefore, the enrichment of GO terms and PFAM domains associated with glycosylation indicates a battery of genes expressed in the lucinid foot are dedicated to mucus biosynthesis and is consistent with the foot's ecological and physiological functions. Unlike most other bivalves, lucinids use the foot to build a mucus-lined inhalant tube to access the oxygenated water column while they inhabit the deeper sulphide-rich sediment layers inaccessible to other bivalves [4]. They also dig mucus-lined sulphide-mining tubes ventrally to probe deeper sediments for sulphide to provide to their symbionts [5, 7]. By contrast, gene expression in the non-symbiotic mussel foot is instead dominated by byssal cuticle proteins and other proteins that contribute to connective tissue structures [23], which reflects the mussel foot's primary role in secreting byssal threads that facilitate attachment to hard substrates.

The chemosynthetic invertebrates inhabiting deep sea hydrothermal vent ecosystems are constantly bathed in dissolved sulphide, while lucinids, on the other hand, inhabit sediments where sulphides levels may be low or patchy [5, 7]. Consequently, the lucinid foot has become a specialised probe for locating and acquiring sulphide in the clam's surrounding environment [5, 7]. The enrichment of GO terms associated with nerve impulse transmission indicates that the lucinid nervous system plays an important role in coordinating this behaviour (Fig. 3); unigenes contributing to this GO term include both nicotinic acetylcholine receptors and ionotropic glutamate receptors (iGluRs). In muscles, the former are located at neuromuscular junctions where electrical signals from neurons control the contraction of muscle cells [68]. Three iGluRs unigenes are expressed in the foot, one of which belongs to the NMDA2 subfamily while the remaining two are members of the AKDF subfamily (Additional file 3). iGluRs similarly mediate synaptic transmission throughout the nervous system, and while they do not directly bind H_2S, it is interesting that NMDA2 receptor signalling is enhanced by physiological levels of H_2S in the mammalian nervous system [69, 70]. Although little is known about the receptors specific for H_2S, it is interesting to speculate that the *iGluRs* expressed in the foot could be involved in coordinating the clam's behavioural response to sulphide in the environment.

G-protein coupled receptors (GPCRs) are also more abundant in the foot (16 unigenes) than in all other organs, except the mantle (18 unigenes), which is consistent with the mantle's role in sensing and initiating responses to environmental cues [71]. Based on the Interpro annotations of the unigenes, most of the GPCRs expressed in the foot contain a 7tm_1 domain (PF0001) and have short N-termini regions, which indicate that they belong to the rhodopsin family of GPCRs [72]. GPCRs belonging to this family interact with a broad spectrum of ligands and their expression in the lucinid foot suggests they could play a role in enabling the foot to detect cues in the immediate environment. Furthermore, the deployment of non-overlapping subsets of GPCRs in the mantle and foot also suggests the molecular set-up for sensing environmental cues is specialised to the distinct functions of each organ (Additional file 3). In addition to receptors for chemical stimuli, a unigene encoding a piezo-type mechanosensitive ion channel component is also expressed in the foot; this unigene is likely to be important for mechanical signal transduction as the foot probes the sediment [73]. These findings indicate the lucinid foot possesses an intricate array of receptors deployed for probing the environment and we hypothesise these chemosensory abilities play an important role in detecting and acquiring resources for the symbiosis.

Conclusions

The global transcriptomic profiles of the *L. orbiculatus* organs are broadly consistent with the expected functions of each organ based on our current understanding of lucinid biology and ecology. However, our findings also allow us to generate new questions to address in future studies. For example, in addition to detecting resources in the environment, does the foot also play a

role in replenishing symbiont abundance by secreting mucus that serves as a chemoattractant and first point of contact a potential symbiont might have with the host? Given the importance of heterotrophy, how do seasonal changes to resource availability in the environment (sediment sulphide and phytoplankton abundance) influence the dynamics of nutritional and immune exchanges in the symbiont-housing gills? The transcriptome we have assembled and the accompanying insights it has provided into molecular basis each lucinid organ's functions, set up the foundation for future studies to build a picture of how lucinids use their molecular toolkit to orchestrate the tripartite interactions linking host, symbiont and the environment.

Methods

Sample collection and processing

L. orbiculatus (roughly 1.5 cm in length) were dug up from 10 to 20 cm deep sediment in the Bay of Fetovaia, Elba, Italy (May 2017). The clams were brought back to the Hydra Institute where they were kept in sediment and aerated sea water from the bay for no longer than 3 h before they were fixed in RNAlater (Thermo Fisher Scientific, Vilnius, Lithuania) and stored at − 20 °C. The gills, mantle, visceral mass, and foot from three individuals were dissected, placed in a tube containing sterile glass beads and TRIzol (Thermo Fisher Scientific), and subsequently homogenised using a pellet pestle. Total RNA was extracted following the manufacturer's instructions. The extracted RNA were quantified using a Qubit 3.0 fluorometer (Thermo Fisher Scientific) and their quality assessed using the RNA 6000 Nano Kit on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA).

To generate a comprehensive reference assembly, we also included 5 additional gill cDNA libraries that were previously sequenced by the group. These gill samples were excised from freshly collected specimens from the same location Elba in October 2015 and from the same batch of specimens that were brought back to the University of Vienna and kept in oxygenated water and Elba sediment until December 2016. The samples were fixed and RNA was extracted as described above. Each RNA extract from these samples was DNase-treated with the Turbo DNA-free Kit (Thermo Fisher Scientific, Vilnius, Lithuania) following the manufacturer's instructions. The treated RNA was ethanol precipitated by adding 150 μl DEPC water, 20 μl sodium acetate (3 M, pH 5.2), 2 μl glycogen (20 mg/ml), 660 μl ethanol and precipitated at − 80 °C for 30 min. The precipitated RNA was then pelleted at 21,000 x g for 15 min at 4 °C, washed with 70% ethanol, centrifuged at 21,000 x g for 10 min at 4 °C, air dried, and eluted in 30 μl DEPC water.

Library construction and sequencing, sequence pre-processing, and filtering

The RNA samples were poly (A) enriched and cDNA libraries were prepared using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA) by the Vienna Biocenter Core Facilities GmbH (VBCF, Vienna, Austria). A total of 12 stranded RNA-Seq libraries (4 tissues × 3 biological replicates per tissue) were multiplexed and sequenced with TruSeq V4 chemistry by the VBCF in one lane of a flow cell on the Illumina HiSeq 2500 to generate 125 bp paired-end reads. Predicted errors in the raw reads were corrected using Rcorrector under default settings and the uncorrectable reads were removed [74]. SortMeRNA was used to remove rRNA sequences from the libraries by searching against the 5S, 5.8S, 16S, 23S, 18S, and 28S rRNA sequences in the package's default rRNA database [75]. The filtered and corrected reads were then trimmed using Trimmomatic v0.36 with the following settings "SLIDINGWINDOW:4:15 MINLEN:36" [76]. Illumina adapter sequences were also removed from the paired-end libraries with the option ILLUMINACLIP:adapters/TruSeq3-PE-2.fa:2:30. The quality of the trimmed and filtered reads were assessed using FastQC v0.11.5, taking into consideration per base quality scores, GC-content, and read length (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc).

For the additional 5 gill libraries that were added to our comprehensive reference dataset, the RNA samples were processed with a ribosomal RNA depletion step using the Ribo-Zero Gold rRNA removal Kit (Epidemiology) prior to library preparation. Library preparation and sequencing were then carried out by the VBCF as described above. These libraries were processed using the same workflow as the other libraries sequenced for the present study but with an additional step to remove symbiont-derived reads. Briefly, after filtering out rRNA reads using SortMeRNA, the remaining reads were mapped under default parameters to the symbiont genome (accession IDs LVJW00000000 and LVKA00000000; genomes available on NCBI in in BioProject PRJNA314435), using Bowtie2 v2.3.3.1 [12, 77]. The unmapped reads were retained for trimming and processing as described above.

De novo transcriptome assembly, quality assessment

Each filtered library of reads was individually assembled into contiguous cDNA sequences with IDBA_tran v1.1.3 with the parameters -max_isoforms 100 --mink 20 --maxk 80 --step 5 [78]. The assembled libraries were then concatenated and CD-HIT-EST v4.7 was used to cluster the contigs and stringently remove redundancies with the parameters -c 1.0 -G 0 -aS 1.0 -aL 0.005 [79]. The Evidential Gene tr2aacds pipeline was used to predict the coding sequence regions within the transcripts

(https://sourceforge.net/projects/evidentialgene/), and highly similar peptide sequences (90% identity) were clustered using CD-HIT v4.7 (–c 0.90) to create the final list of transcripts [79]. We used TransRate to assess the assembled transcriptome with variety of metrics, including number of transcripts, total number of bases, N50 values and GC content using [80]. Assembled transcripts with predicted open reading frames shorter than 80 amino acids were filtered out. Although small open reading frames and other non-coding transcripts are likely to have important albeit unknown biological functions, this arbitrary threshold was chosen to reduce the number of transcripts in the final assembly to reduce the computational resources required for subsequent analyses. We used the BUSCO program v3.0.2 to assess the completeness of the final reference transcriptome through a search against a set of 978 metazoan BUSCOs (Benchmarking set of Universal Single-Copy Orthologs) [81, 82].

Functional annotation of the predicted proteins

The predicted peptide sequences were annotated with protein domain and conserved motif information using InterProScan (5.28–67.0) [83] and hmmscan against the PFAM-A database [84]. We also annotated the predicted proteins by blastp searching against the SwissProt database (–max_target_seqs 1 -evalue 1e-6). The BLAST and InterProScan outputs were combined and then annotated with GO terms using BLAST2GO v5.0.21 [85, 86]. The predicted peptide sequences were also annotated with KEGG terms using the BlastKOALA webserver (https://www.kegg.jp/blastkoala/) and then mapped to the KEGG pathways database to identify unigenes involved in immune signalling [87, 88].

Identification of differentially expressed genes and functional enrichment of organ-specific expression profiles

Only the libraries sequenced from the organs collected in May 2017 were used for differential expression analyses. We used a combination of Salmon and Corset to map reads to each transcript and cluster transcripts into putative unigenes [89, 90]; each unigene is assembled from transcripts that appear to originate from the same transcriptional locus. A reference transcriptome index was constructed using Salmon v0.9.1 with the parameters "--type quasi -k 31", and we used the "pseudo-alignment" algorithm to map the reads from each library to the index with default settings [89]. The numbers of reads mapped to each transcript were generated using Corset v1.07 under default parameters but with '-m 0' applied to retain all transcripts [90]. Corset uses the Fragment equivalence classes output by Salmon to cluster transcripts into putative genes based on sequence similarity and read mapping [90]. A Principal component analysis

(PCA) of the vsd-transformed (variance stabilizing transformation) counts was used to explore the global transcriptomic variation across samples; implemented through DESeq2 version v1.18.1 according to instructions in the DESeq2 manual [91] and visualised using the R package "ggplot2" [92].

Analyses to identify unigenes differentially expressed across the organs were conducted through pairwise comparisons between each organ using DESeq2 v1.18.1 [91]. TXimport v1.6.0 was used to import the read count data from the Salmon output [93]. We excluded unigenes with fewer than 10 reads mapped across all samples to reduce computational demands. A false discovery rate (FDR) of 10% was implemented using the Benjamini–Hochberg method. An FDR-adjusted p-value ≤0.01 and log2 fold change threshold of 1 (equates to a minimum 2-fold change) were used to produce the final list of differentially expressed genes. Organ-specific clusters of unigenes upregulated in one organ versus all others were identified with a Venn diagram (http://bioinfogp.cnb.csic.es/tools/venny).

A hypergeometric test, implemented through the Cytoscape plug-in BiNGO v3.0.3, was used to test for statistical enrichment (FDR-adjusted p-value < 0.05) of GO terms in the lists of organ-specific unigenes [94]. The GO terms enriched in each organ were visually summarised using Revigo [95]. GO terms were clustered using semantic similarity measure (SimRel) in REVIGO and GO terms with > 0.7 redundancy were collapsed [95]. We also tested for statistically enriched PFAM domains amongst the different organs using a hypergeometric test and the p-values were adjusted for an FDR of 10%. The PFAM domain enrichment analyses were carried out in R using the script provided in Supplemental Text 1 of Chandran et al. (2009) [96].

Supplementary information

> **Additional file 1.** Assembly statistics – Assembly statistics for reference transcriptome
>
> **Additional file 2.** Genes involved in immunity – Tables summarising the genes putatively involved in *L. orbiculatus* immunity
>
> **Additional file 3.** Differentially expressed genes lists – Lists of genes differentially expressed in pairwise comparisons of the four organs analysed in this study
>
> **Additional file 4.** GO term enrichments – Results of the gene ontology term enrichment analyses of genes highly expressed in each organ
>
> **Additional file 5.** Trees and tables – Phylogenetic trees, table of the transporter genes annotated in the *L. orbiculatus* transcriptome, and table of the urease metabolism genes in the genome of *Loripes orbiculatus* endosymbiont *candidatus* Thiodiazotropha endoloripes

Abbreviations

ATP: Adenosine triphosphate; BP: Biological Process; BUSCO: Benchmarking Universal Single-Copy Ortholog; CA: Carbonic anhydrase; DEPC: Diethyl

pyrocarbonate; FDR: False discovery rate; FReD: C-terminal fibrinogen-related domain; GO: Gene ontology; GPCR: G protein-coupled receptor; GPI: Glycosylphosphatidylinositol; iGluR: Ionotropic glutamate receptor; MACPF: Membrane attack complex/Perforin; MAMP: Microbe Associated Molecular Pattern; PCA: Principal component analysis; PRR: Pattern Recognition Receptor; RNA: Ribonucleic acid; SLC: Solute carrier; SRCR: Scavenger receptor cysteine-rich; TIR: Toll/interleukin-1 receptor; TLR: Toll-like receptor; TNFR: Tumor necrosis factor receptor

References
1. Cavanaugh CM, McKiness ZP, Newton ILG, Stewart FJ. Marine chemosynthetic symbioses. In: Rosenberg E, EF DL, Lory S, Stackebrandt E, Berlin TF, editors. The Prokaryotes: Prokaryotic Biology and Symbiotic Associations. Heidelberg: Springer Berlin Heidelberg; 2013. p. 579–607.
2. Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat Rev Microbiol. 2008;6(10):725–40.
3. Roeselers G, Newton IL. On the evolutionary ecology of symbioses between chemosynthetic bacteria and bivalves. Appl Microbiol Biotechnol. 2012;94(1):1–10.
4. Allen JA. On the basic form and adaptations to habitat in the Lucinacea (Eulamellibranchia). Philos Trans R Soc Lond Ser B Biol Sci. 1958;241(684):421.
5. Dando PR, Southward AJ, Southward EC, Bone Q. Chemoautotrophic symbionts in the gills of the bivalve mollusc *Lucinoma borealis* and the sediment chemistry of its habitat. Proc R Soc Lond B. 1986;227(1247):227–47.
6. Distel DL, Felbeck H. Endosymbiosis in the lucinid clams *Lucinoma aequizonata*, *Lucinoma annulata* and *Lucina floridana*: a reexamination of the functional morphology of the gills as bacteria-bearing organs. Mar Biol. 1987;96(1):79–86.
7. Dando P, Ridgway SA, Spiro B. Sulphide 'mining' by lucinid bivalve molluscs - demonstrated by stable Sulphur isotope measurements and experimental models. Mar Ecol Prog Ser. 1994;107(1–2):169–75.
8. van der Heide T, Govers LL, de Fouw J, Olff H, van der Geest M, van Katwijk MM, Piersma T, van de Koppel J, Silliman BR, Smolders AJP, et al. A three-stage symbiosis forms the foundation of seagrass ecosystems. Science. 2012;336(6087):1432.
9. Higgs ND, Newton J, Attrill MJ. Caribbean spiny lobster fishery is underpinned by trophic subsidies from chemosynthetic primary production. Curr Biol. 2016;26(24):3393–8.
10. Taylor JD, Glover EA, Smith L, Ikebe C, Williams ST. New molecular phylogeny of Lucinidae: increased taxon base with focus on tropical Western Atlantic species (Mollusca: Bivalvia). Zootaxa. 2016;4196(3):381–98.
11. König S, Gros O, Heiden SE, Hinzke T, Thurmer A, Poehlein A, Meyer S, Vatin M, Mbeguie AMD, Tocny J, et al. Nitrogen fixation in a chemoautotrophic lucinid symbiosis. Nat Microbiol. 2016;2:16193.
12. Petersen JM, Kemper A, Gruber-Vodicka H, Cardini U, van der Geest M, Kleiner M, Bulgheresi S, Mußmann M, Herbold C, Seah BKB, et al. Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. Nat Microbiol. 2016;2:16195.
13. Lim SJ, Davis BG, Gill DE, Walton J, Nachman E, Engel AS, Anderson LC, Campbell BJ. Taxonomic and functional heterogeneity of the gill microbiome in a symbiotic coastal mangrove lucinid species. ISME J. 2019; 13:902–20.
14. Caro A, Got P, Bouvy M, Troussellier M, Gros O. Effects of long-term starvation on a host bivalve (*Codakia orbicularis*, lucinidae) and its symbiont population. Appl Environ Microbiol. 2009;75(10):3304–13.
15. Caro A, Gros O, Got P, De Wit R, Troussellier M. Characterization of the population of the sulfur-oxidizing symbiont of *Codakia orbicularis* (Bivalvia, Lucinidae) by single-cell analyses. Appl Environ Microbiol. 2007;73(7):2101.
16. Gros O, Elisabeth NH, Gustave SD, Caro A, Dubilier N. Plasticity of symbiont acquisition throughout the life cycle of the shallow-water tropical lucinid *Codakia orbiculata* (Mollusca: Bivalvia). Environ Microbiol. 2012;14(6):1584–95.
17. Bettencourt R, Pinheiro M, Egas C, Gomes P, Afonso M, Shank T, Santos RS. High-throughput sequencing and analysis of the gill tissue transcriptome from the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus*. BMC Genomics. 2010;11(1):559.
18. Zheng P, Wang M, Li C, Sun X, Wang X, Sun Y, Sun S. Insights into deep-sea adaptations and host-symbiont interactions: a comparative transcriptome study on *Bathymodiolus* mussels and their coastal relatives. Mol Ecol. 2017;19:5133–48.
19. Sun J, Zhang Y, Xu T, Zhang Y, Mu H, Zhang Y, Lan Y, Fields CJ, Hui JHL, Zhang W, et al. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. Nature Ecology & Evolution. 2017;1:0121.
20. Duplessis MR, Dufour SC, Blankenship LE, Felbeck H, Yayanos AA. Anatomical and experimental evidence for particulate feeding in *Lucinoma aequizonata* and *Parvilucina tenuisculpta* (Bivalia: Lucinidae) from the Santa Barbara Basin. Mar Biol. 2004;145(3):551–61.
21. van der Geest M, Sall AA, Ely SO, Nauta RW, van Gils JA, Piersma T. Nutritional and reproductive strategies in a chemosymbiotic bivalve living in a tropical intertidal seagrass bed. Mar Ecol Prog Ser. 2014;501:113–26.
22. Cardini U, Bartoli M, Lücker S, Mooshammer M, Polzin J, Lee RW, Micić V, Hofmann T, Weber M, Petersen JM. Chemosymbiotic bivalves contribute to the nitrogen budget of seagrass ecosystems. ISME J. 2019. https://www.nature.com/articles/s41396-019-0486-9.
23. Gerdol M, Fujii Y, Hasan I, Koike T, Shimojo S, Spazzali F, Yamamoto K, Ozeki Y, Pallavicini A, Fujita H. The purplish bifurcate mussel *Mytilisepta virgata* gene expression atlas reveals a remarkable tissue functional specialization. BMC Genomics. 2017;18(1):590.
24. Ghiselli F, Iannello M, Puccio G, Chang PL, Plazzi F, Nuzhdin SV, Passamonti M. Comparative transcriptomics in two bivalve species offers different perspectives on the evolution of sex-biased genes. Genome Biol Evol. 2018; 10(6):1389–402.
25. Mun S, Kim Y-J, Markkandan K, Shin W, Oh S, Woo J, Yoo J, An H, Han K. The whole-genome and transcriptome of the manila clam (*Ruditapes philippinarum*). Genome Biol Evol. 2017;9(6):1487–98.
26. McCartney MA, Auch B, Kono T, Mallez S, Zhang Y, Obille A, Becker A, Abrahante JE, Garbe J, Badalamenti JP, et al. The genome of the zebra

mussel, *Dreissena polymorpha*: A resource for invasive species research. bioRxiv. 2019:696732.

27. Chu H, Mazmanian SK. Innate immune recognition of the microbiota promotes host-microbial symbiosis. Nat Immunol. 2013;14(7):668–75.

28. Janeway JCA, Medzhitov R. Innate immune recognition. Annu Rev Immunol. 2002;20(1):197–216.

29. Allam B, Pales Espinosa E. Bivalve immunity and response to infections: are we looking at the right place? Fish Shellfish Immunol. 2016;53:4–12.

30. Gerdol M, Venier P, Edomi P, Pallavicini A. Diversity and evolution of TIR-domain-containing proteins in bivalves and Metazoa: new insights from comparative genomics. Dev Comp Immunol. 2017;70:145–64.

31. Guven-Maiorov E, Keskin O, Gursoy A, VanWaes C, Chen Z, Tsai C-J, Nussinov R. The architecture of the TIR domain signalosome in the toll-like receptor-4 signaling pathway. Sci Rep. 2015;5:13128.

32. Tewari R, Bailes E, Bunting KA, Coates JC. Armadillo-repeat protein functions: questions for little creatures. Trends Cell Biol. 2010;20(8):470–81.

33. Cabal-Hierro L, Lazo PS. Signal transduction by tumor necrosis factor receptors. Cell Signal. 2012;24(6):1297–305.

34. Park HH, Lo YC, Lin SC, Wang L, Yang JK, Wu H. The death domain superfamily in intracellular signaling of apoptosis and inflammation. Annu Rev Immunol. 2007;25:561–86.

35. Xu Q, Reed JC. Bax inhibitor-1, a mammalian apoptosis suppressor identified by functional screening in yeast. Mol Cell. 1998;1(3):337–46.

36. Hata AN, Engelman JA, Faber AC. The BCL2 family: key mediators of the apoptotic response to targeted anticancer therapeutics. Cancer Discov. 2015;5(5):475–87.

37. Elisabeth NH, Gustave SD, Gros O. Cell proliferation and apoptosis in gill filaments of the lucinid *Codakia orbiculata* (Montagu, 1808) (Mollusca: Bivalvia) during bacterial decolonization and recolonization. Microsc Res Tech. 2012;75(8):1136–46.

38. Piquet B, Shillito B, Lallier FH, Duperron S, Andersen AC. High rates of apoptosis visualized in the symbiont-bearing gills of deep-sea *Bathymodiolus* mussels. PLoS One. 2019;14(2):e0211499.

39. Childress JJ, Girguis PR. The metabolic demands of endosymbiotic chemoautotrophic metabolism on host physiological capacities. J Exp Biol. 2011;214(2):312.

40. Détrée C, Haddad I, Demey-Thomas E, Vinh J, Lallier FH, Tanguy A, Mary J. Global host molecular perturbations upon in situ loss of bacterial endosymbionts in the deep-sea mussel *Bathymodiolus azoricus* assessed using proteomics and transcriptomics. BMC Genomics. 2019;20(1):109.

41. Carroll MA, Catapane EJ. The nervous system control of lateral ciliary activity of the gill of the bivalve mollusc, *Crassostrea virginica*. Comp Biochem Physiol A Mol Integr Physiol. 2007;148(2):445–50.

42. Roberts AJ, Kon T, Knight PJ, Sutoh K, Burgess SA. Functions and mechanics of dynein motor proteins. Nat Rev Mol Cell Biol. 2013;14(11):713–26.

43. Kochevar RE, Childress JJ. Carbonic anhydrase in deep-sea chemoautotrophic symbioses. Mar Biol. 1996;125(2):375–83.

44. Hongo Y, Ikuta T, Takaki Y, Shimamura S, Shigenobu S, Maruyama T, Yoshida T. Expression of genes involved in the uptake of inorganic carbon in the gill of a deep-sea vesicomyid clam harboring intracellular thioautotrophic bacteria. Gene. 2016;585(2):228–40.

45. Wilbur KM, Saleuddin ASM. 6 - Shell Formation. In: ASM S, Wilbur KM, editors. *The Mollusca*: Academic Press; 1983. vol. 4. p. 235–87.

46. Cudennec B, Rousseau M, Lopez E, Fouchereau-Peron M. CGRP stimulates gill carbonic anhydrase activity in molluscs via a common CT/CGRP receptor. Peptides. 2006;27(11):2678–82.

47. Nakajima K, Tanaka A, Matsuda Y. SLC4 family transporters in a marine diatom directly pump bicarbonate from seawater. Proc Natl Acad Sci. 2013; 110(5):1767.

48. Hirata T, Czapar A, Brin L, Haritonova A, Bondeson DP, Linser P, Cabrero P, Thompson J, Dow JA, Romero MF. Ion and solute transport by Prestin in *Drosophila* and *Anopheles*. J Insect Physiol. 2012;58(4):563–9.

49. Schweimanns M, Felbeck H. Significance of the occurrence of chemoautotrophic bacterial endosymbionts in lucinid clams from Bermuda. Mar Ecol Prog Ser. 1985;24:113–20.

50. König S, Le Guyader H, Gros O. Thioautotrophic bacterial endosymbionts are degraded by enzymatic digestion during starvation: case study of two lucinids *Codakia orbicularis* and *C. orbiculata*. Microsc Res Tech. 2015;78(2):173–9.

51. Lehnert EM, Mouchka ME, Burriesci MS, Gallo ND, Schwarz JA, Pringle JR. Extensive differences in gene expression between symbiotic and aposymbiotic cnidarians. G3. 2014;4(2):277–95.

52. Bertucci A, Forêt S, Ball EE, Miller DJ. Transcriptomic differences between day and night in *Acropora millepora* provide new insights into metabolite exchange and light-enhanced calcification in corals. Mol Ecol. 2015;24(17): 4489–504.

53. Miyamoto N, Yoshida M-a, Koga H, Fujiwara Y. Genetic mechanisms of bone digestion and nutrient absorption in the bone-eating worm *Osedax japonicus* inferred from transcriptome and gene expression analyses. BMC Evol Biol. 2017;17(1):17.

54. Hamada M, Schröder K, Bathia J, Kürn U, Fraune S, Khalturina M, Khalturin K, Shinzato C, Satoh N, Bosch TCG. Metabolic co-dependence drives the evolutionarily ancient *Hydra–Chlorella* symbiosis. eLife. 2018;7:e35122.

55. Chan CYL, Hiong KC, Boo MV, Choo CYL, Wong WP, Chew SF, Ip YK. Light exposure enhances urea absorption in the fluted giant clam, *Tridacna squamosa*, and up-regulates the protein abundance of a light-dependent urea active transporter, DUR3-like, in its ctenidium. J Exp Biol. 2018;221(Pt 8).

56. Lodwig EM, Hosie AHF, Bourdès A, Findlay K, Allaway D, Karunakaran R, Downie JA, Poole PS. Amino-acid cycling drives nitrogen fixation in the legume–*Rhizobium* symbiosis. Nature. 2003;422:722.

57. Giovanardi F, Vollenweider R. Trophic conditions of marine coastal waters: experience in applying the trophic index TRIX to two areas of the Adriatic and Tyrrhenian seas. J Limnol. 2004;63(2):199–218.

58. Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, Chang Y-J, Shah M, VerBerkmoes NC, Zarzycki J, et al. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. Proc Natl Acad Sci. 2012;109(19):E1173–82.

59. McCormack R, Podack E. Perforin-2/Mpeg1 and other pore-forming proteins throughout evolution. J Leukoc Biol. 2015;98:761–8.

60. He X, Zhang Y, Yu Z. An MPEG (macrophage expressed gene) from the Pacific oyster *Crassostrea gigas*: molecular characterization and gene expression. Fish Shellfish Immunol. 2011;30(3):870–6.

61. Mühlenbruch M, Grossart H-P, Eigemann F, Voss M. Mini-review: phytoplankton-derived polysaccharides in the marine environment and their interactions with heterotrophic bacteria. Environ Microbiol. 2018;20(8): 2671–85.

62. Brown GD, Willment JA, Whitehead L. C-type lectins in immunity and homeostasis. Nat Rev Immunol. 2018;18(6):374–89.

63. Pang X, Xiao X, Liu Y, Zhang R, Liu J, Liu Q, Wang P, Cheng G. Mosquito C-type lectins maintain gut microbiome homeostasis. Nat Microbiol. 2016;1: 16023.

64. Pales Espinosa E, Allam B. Reverse genetics demonstrate the role of mucosal C-type lectins in food particle selection in the oyster *Crassostrea virginica*. J Exp Biol. 2018;221:jeb174094.

65. Hynes RO, Naba A. Overview of the matrisome - an inventory of extracellular matrix constituents and functions. Cold Spring Harb Perspect Biol. 2012;4(1):a004903.

66. Hynes R. The evolution of metazoan extracellular matrix. J Cell Biol. 2012; 196:671–9.

67. Brockhausen I, Stanley P. O-GalNAc Glycans. 2017. In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology [Internet]. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017. Chapter 10. Available from: https://www.ncbi.nlm.nih.gov/books/NBK453030/.

68. Fagerlund MJ, Eriksson LI. Current concepts in neuromuscular transmission. Br J Anaesth. 2009;103(1):108–14.

69. Abe K, Kimura H. The possible role of hydrogen sulfide as an endogenous neuromodulator. J Neurosci. 1996;16(3):1066.

70. Marutani E, Kosugi S, Tokuda K, Khatri A, Nguyen R, Atochin DN, Kida K, Van Leyen K, Arai K, Ichinose F. A novel hydrogen sulfide-releasing N-methyl-D-aspartate receptor antagonist prevents ischemic neuronal death. J Biol Chem. 2012;287(38):32124–35.

71. Schmidt-Rhaesa A, Harzsch S, Purschke G. Structure and evolution of invertebrate nervous systems. Oxford: Oxford University press; 2016.

72. Krishnan A, Schioth HB. The role of G protein-coupled receptors in the early evolution of neurotransmission and the nervous system. J Exp Biol. 2015; 218(4):562–71.

73. Wu J, Lewis AH, Grandl J. Touch, tension, and transduction - the function and regulation of piezo ion channels. Trends Biochem Sci. 2017;42(1):57–71.

74. Song L, Florea L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. GigaScience. 2015;4:48.

75. Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24): 3211–7.

76. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

77. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9(4):357–9.

78. Peng Y, Leung HCM, Yiu S-M, Lv M-J, Zhu X-G, Chin FYL. IDBA-Tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. Bioinformatics. 2013;29(13):i326–34.

79. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150–2.

80. Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S. TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome Res. 2016;26(8):1134–44.

81. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.

82. Waterhouse RM, Kriventseva EV, Simão FA, Klioutchnikov G, Seppey M, Manni M, Ioannidis P, Zdobnov EM. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 2017; 35(3):543–8.

83. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40.

84. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:W29–37.

85. Conesa A, Terol J, García-Gómez JM, Talón M, Robles M, Götz S. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674–6.

86. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36(10):3420–35.

87. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44(D1):D457–62.

88. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428(4):726–31.

89. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017; 14(4):417–9.

90. Davidson NM, Oshlack A. Corset: enabling differential gene expression analysis for de novoassembled transcriptomes. Genome Biol. 2014;15(7):410.

91. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

92. Wickham H. ggplot2 Elegant graphics for data analysis. New York: Springer; 2009.

93. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research. 2015;4:1521.

94. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 2005;21(16):3448–9.

95. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6(7):e21800.

96. Chandran D, Tai YC, Hather G, Dewdney J, Denoux C, Burgess DG, Ausubel FM, Speed TP, Wildermuth MC. Temporal global expression data reveal known and novel salicylate-impacted processes and regulators mediating powdery mildew growth and reproduction on *Arabidopsis*. Plant Physiol. 2009;149(3):1435–51.

Publisher's Note