

RESEARCH

Open Access



Genomic analysis of the tryptome reveals molecular mechanisms of gland cell evolution

Leslie S. Babonis^{1*} , Joseph F. Ryan^{1,2}, Camille Enjolras¹ and Mark Q. Martindale^{1,2}

Abstract

Background: Understanding the drivers of morphological diversity is a persistent challenge in evolutionary biology. Here, we investigate functional diversification of secretory cells in the sea anemone *Nematostella vectensis* to understand the mechanisms promoting cellular specialization across animals.

Results: We demonstrate regionalized expression of gland cell subtypes in the internal ectoderm of *N. vectensis* and show that adult gland cell identity is acquired very early in development. A phylogenetic survey of trypsins across animals suggests that this gene family has undergone numerous expansions. We reveal unexpected diversity in trypsin protein structure and show that trypsin diversity arose through independent acquisitions of non-trypsin domains. Finally, we show that trypsin diversification in *N. vectensis* was effected through a combination of tandem duplication, exon shuffling, and retrotransposition.

Conclusions: Together, these results reveal the numerous evolutionary mechanisms that drove trypsin duplication and divergence during the morphological specialization of cell types and suggest that the secretory cell phenotype is highly adaptable as a vehicle for novel secretory products.

Keywords: Trypsin, Novelty, Organelle evolution, Secretion, Cnidarian, Domain evolution, Cell-type specialization

Introduction

The development of new tissue layers provides the opportunity to spatially segregate cell types enabling the compartmentalization of different functions. Cnidarians are diploblasts, comprised of an internal endodermal epithelium separated from an external ectodermal epithelium by a largely acellular matrix called mesoglea. Anthozoans (corals, sea anemones, and their kin) are unusual among cnidarians in their possession of internal tissues (pharynx and mesenteries) that arise by secondary epithelial fold morphogenesis following completion of gastrulation [1]. Additional growth and differentiation of both internalized layers result in the morphogenesis of the pharynx and mesenteries and in an adult form quite different from that of medusozoans. In anthozoans, both layers (endoderm and ectoderm) are in contact with the gastric cavity; whereas in medusozoans (and, indeed,

most other animals), the gastrovascular cavity is lined only by endoderm. The secondary internalization of both ectoderm and endoderm in anthozoans provided a new opportunity for compartmentalization of cell functions and may have facilitated the expansion of novel cell types through regionalized cell-type specialization.

Nematostella vectensis, the starlet sea anemone, has become a valuable model for studies of animal body plan evolution [2–6]; yet, little is known about the extent of cell diversity in the tissues that comprise the pharynx and mesenteries. The endodermal component of the mesenteries houses the germ cell precursors and two types of muscle cells, and the few recent studies of the mesenteries in *Nematostella* have focused largely on these endodermal functions [7–9]. The ectodermal component of the mesenteries is known to be populated by cnidocytes and gland cells [10] and two recent studies demonstrated the expression of multiple proteases in the mesenteries of *N. vectensis* [11, 12]. Trypsins are the largest family of proteases, and although they have diverse functions, most trypsins are secreted to the extracellular environment and are, therefore, expressed

*Correspondence: babonis@whitney.ufl.edu

¹ Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, FL 32080, USA

Full list of author information is available at the end of the article



in zymogen-type gland cells [13]. A previous study cataloging trypsin diversity from prokaryotes and eukaryotes identified 75 trypsins in the genome of *N. vectensis* [14], suggesting that the few cell types identified anatomically as zymogen gland cells [10] may belie the digestive capacity of the mesenteries.

We sought to understand the evolutionary mechanisms promoting functional diversification at the cell and tissue levels in the mesenteries of *N. vectensis*, and to characterize the evolutionary history of a large (super)family of proteases expressed abundantly in the mesenteries. Building on a previous study using RNA-seq to characterize the expression profile of the mesenteries in *N. vectensis* [11], we show that the continuous epithelium comprising the internal ectoderm in *N. vectensis* is partitioned into different regions associated with distinct morphologies and functions. Additionally, we show numerous lineage-specific expansions of trypsins and that trypsin diversification arises through novel domain acquisition. Finally, we propose a model by which the expansion of trypsins may have promoted specialization of gland cell subtypes in cnidarians.

Results

Morphology and function of the internal ectoderm

We examined the fine structure of the internal and external ectoderm in the region of the mouth of *N. vectensis* during feeding for evidence of morphological and functional variation (Fig. 1). Cells in the external ectoderm around the mouth are organized into a low cuboidal-type epithelium that covers the closed mouth between feeding events (Fig. 1a–d). In the presence of prey, the pharynx is partially everted, exposing the tall columnar epithelium of the pharyngeal ectoderm (Fig. 1e–g). After passing through the pharynx (Fig. 1h), ingested prey remains in contact with the ectodermal portion of the mesenteries, which is populated by cnidocytes and gland cells (Fig. 1i–k).

The pharyngeal ectoderm contains numerous distinct electron dense (zymogen-secreting) and electron lucent (mucus-secreting) gland cells (Fig. 2a–f). The adjacent non-secretory cells in this epithelium have distinctive apical electron-dense vesicles (Fig. 2f). The proximal region of the mesentery (adjacent to the body wall) is comprised of endoderm, while the distal portion (the free edge) is comprised of ectoderm (Fig. 2g) [10, 12, 15]. The

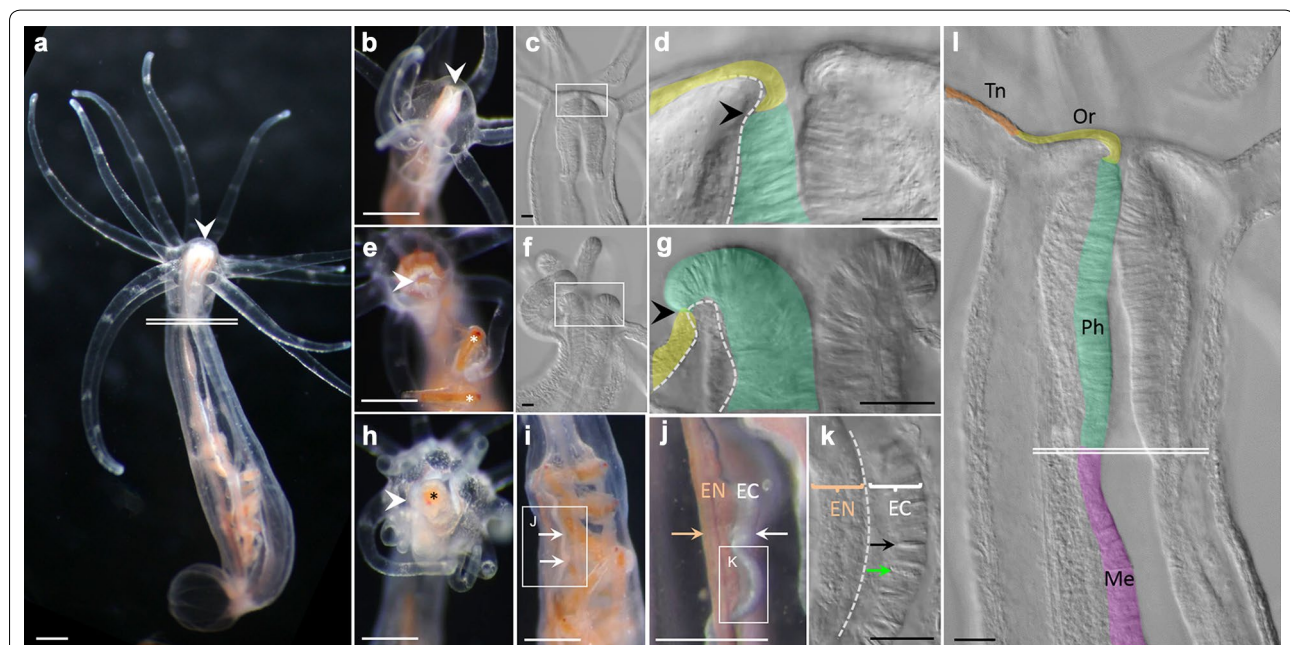


Fig. 1 Morphology of the pharynx and mesenteries. **a** Adult polyp. **b–d** Polyps at rest; the pharyngeal ectoderm (green) is retracted inside the oral ectoderm (yellow). **e–g** Partial eversion of pharynx occurs during capture/handling of prey (*Artemia* sp., indicated by *). **h–j** Ingested prey passes through the mouth and pharynx and remains in contact with the ectoderm (white arrows) of the mesenteries during digestion; colored arrow indicates endoderm of mesenteries (pigmentation from consumption of *Artemia*). **k** Cnidocytes (black arrow) and gland cells (green arrow) are restricted to the ectoderm of the mesenteries. **l** The external ectoderm of the tentacles (Tn, orange) and oral region (Or, yellow) is continuous with the internal ectoderm of the pharynx (Ph, green) and mesenteries (Me, purple). **c, d, f, g, k, l** are DIC micrographs. **d, g, l** are false colored. **e, h** are oral views, the remaining images are lateral/oblique views. Dotted lines in **d, g, k** denote position of mesoglea. White arrowheads point to the mouth, black arrowheads denote transition from external to internal ectoderm. Double white lines in **a, l** denote the transition from pharynx to mesentery. White scale bars = 2.5 mm; black scale bars = 20 μ m

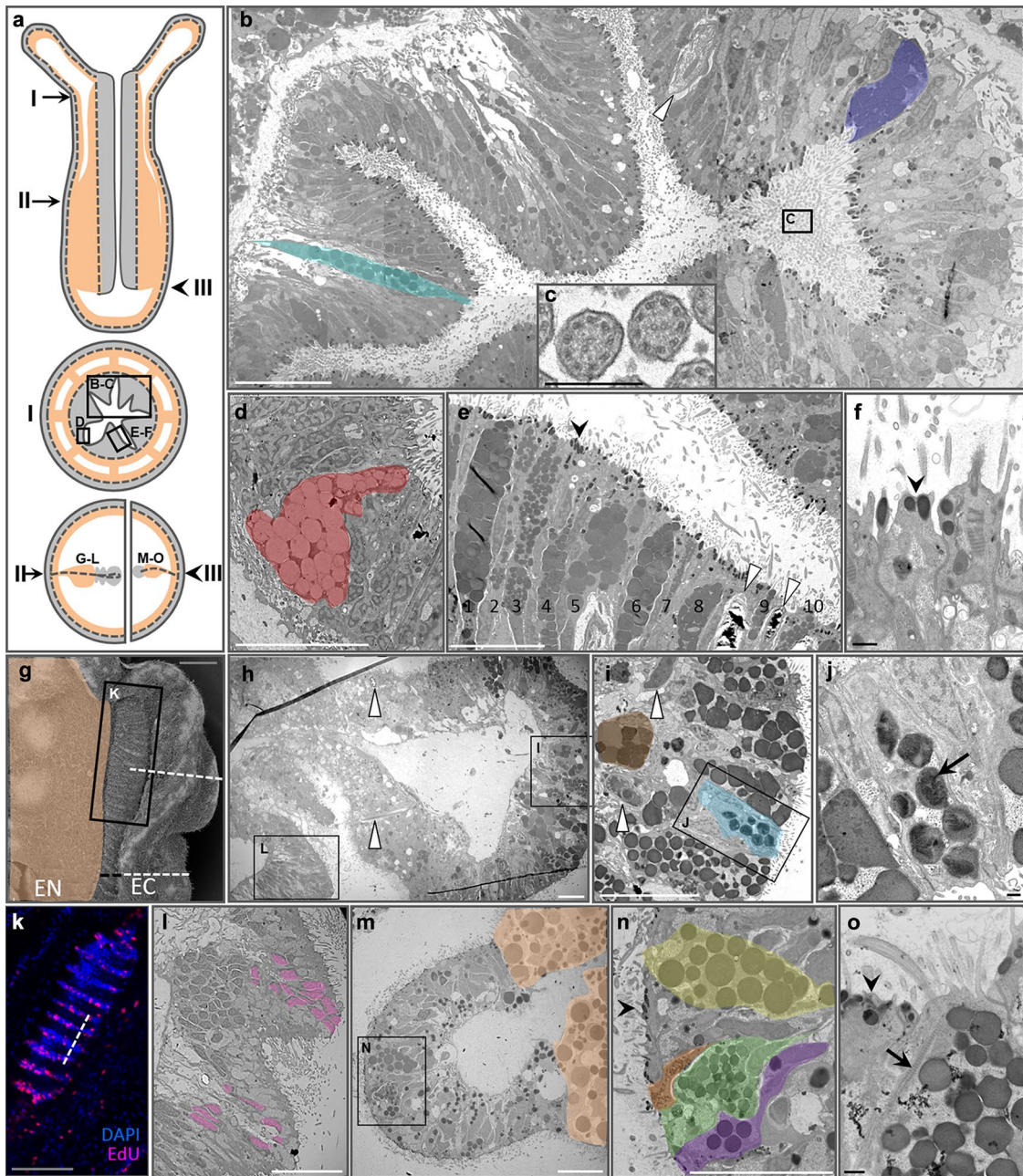


Fig. 2 Fine structure of the pharyngeal ectoderm and ectodermal mesenteries. **a** Diagrammatic representation of regions where thin sections were examined through the pharynx (I) and at two positions in the mesenteries (II, III). Top—sagittal view of a polyp, middle—cross section at position I, bottom—partial cross sections at positions II (left) and III (right). **b–f** TEMs of a region of the pharyngeal ectoderm (I) indicated by the box in A. **b** Two zymogen gland cells are false colored for emphasis. **c** Cross sections of cilia emerging from the pharyngeal ectoderm into the pharyngeal canal. **d** Mucus-secreting cell, false colored. **e** Ten zymogen-type gland cells and two cnidocytes. **f** Electron-dense vesicles (black arrowheads in **e**, **f**, **n**, **o**) in the apex of non-glandular ectodermal cells. **g** SEM of a mesentery from position II in panel **a**. The ectodermal (EC) portion has two parts: the ciliated tract (black line) and cnidoglandular tract (white line); the endodermal (EN) portion is false colored. **h–j** TEMs of the cnidoglandular tract of a mesentery from position II in A. Sections correspond to the position of the dotted line in **g**. **i** Two zymogen-type gland cells are false colored. **j** Some zymogen vesicles have heterogeneous contents (black arrow). **k** 3D reconstruction of a confocal z-stack through the ciliated tract of a mesentery from position II in A; pink—nuclei of proliferating cells (EdU), blue—quiescent nuclei (DAPI). **l** ITEM section through the ciliated tract at the position indicated by the dotted line in **k** and by the box in **h**. Nuclei corresponding to panel **k** are false colored pink. **m–o** TEMs of a mesentery lacking a ciliated tract, position III in panel A. **m** Endoderm is false colored. **n** Zymogen gland cells are false colored. **o** Ciliary rootlet (black arrow) in the apex of a zymogen gland cell. Scale bars: white line—10 μ m, black line—500 nm, gray line (panels **g**, **k**)—50 μ m. Panels **b** and **h** are composites of multiple micrographs. White arrowheads indicate cnidocytes throughout

ectodermal region gives rise to both the cnidoglandular tract at the most distal extent (Fig. 2h–j, m–o) and the ciliated tract more proximally (Fig. 2k, l). Thin sections of the ectodermal mesentery in the oral region (near the pharynx) show abundant zymogen gland cells (Fig. 2h–j), some of which contain secretory vesicles with heterogeneous contents (Fig. 2j). Ciliated tracts are short and are present only in the oral end of each mesentery. Cells of the ciliated tract are highly proliferative and have apical motile cilia but do not have other distinguishing features (Fig. 2k, l). The aboral mesentery lacks a ciliated tract but the cnidoglandular tract still contains numerous distinct zymogen gland cells, some with motile apical cilia (Fig. 2m–o). Mucus-secreting cells were found in the pharyngeal ectoderm (Fig. 2d) and in the external ectoderm of the body wall and tentacles (Additional file 1), but never in the endoderm.

Proteolytic enzymes are expressed in the developing mesenteries

We previously identified numerous genes encoding different classes of proteases to be upregulated in the adult mesentery of *N. vectensis* [11]. Using in situ hybridization, we examined the spatial and temporal expression of various classes of proteases identified from this study during early development of the pharynx and mesenteries to understand the ontogeny of digestive function and the onset of terminal gut cell differentiation. All genes examined were expressed in individual ectodermal cells of the mesenteries at the primary polyp stage, just after metamorphosis (Fig. 3a, b); two protease genes (NVJ_82725 and NVJ_83864) were also expressed in the pharyngeal ectoderm of the primary polyp. There was surprisingly little variation in the onset of protease expression, although serine proteases (trypsins) consistently exhibited expression in the early planula stage before differentiation of the presumptive pharynx and mesenteries (Fig. 3b). Double fluorescent in situ hybridization for two metalloprotease genes (NVJ_88668 and NVJ_2109) indicates both co-expression of these two enzymes in few cells at the aboral end of the pharynx and independent expression of the two genes in distinct cells of the ectodermal mesenteries in the late tentacle bud stage (Fig. 3c). These results suggest that adult gland cell identity is acquired very early in development, coincident with the morphogenesis of the pharynx and mesenteries.

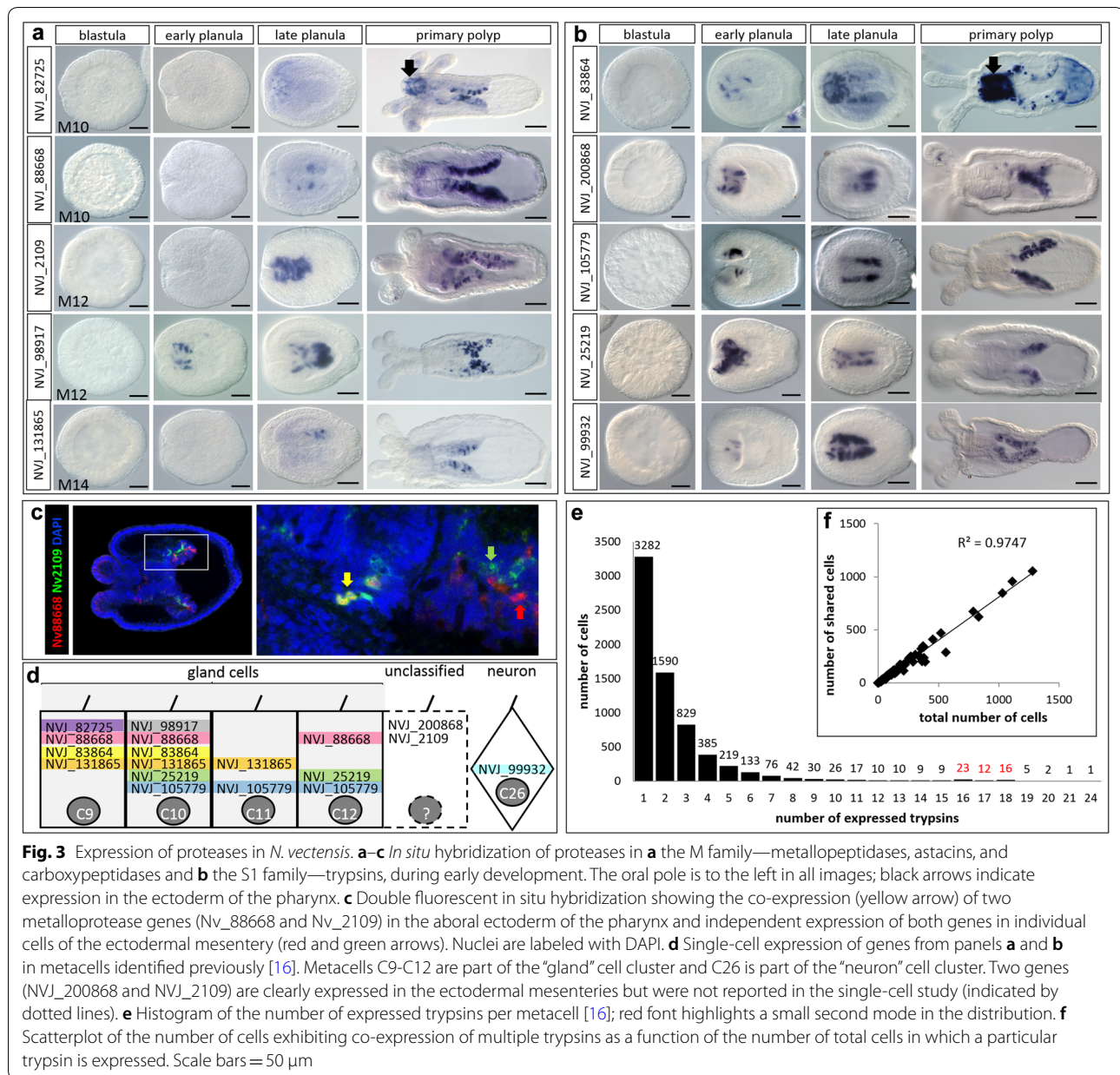
The surprising lack of any obvious spatial segregation in protease expression led us to hypothesize that many proteases may be co-expressed together in the few anatomically distinguishable gland cells identified above (Fig. 2). Using the raw data from a single-cell RNA-Seq study published previously [16], we show co-expression of 6 of the 10 proteases we studied by in situ

hybridization in a single putative gland cell (Fig. 3d). Using the raw data from the same study and a very low cutoff for gene expression ($N \geq 1$ read), we examined more fully the co-expression of the large superfamily of trypsin proteases and found 6727 cells expressing at least one trypsin gene. Nearly, 50% of the trypsin-expressing cells (3282/6727) appear to express only a single trypsin, while the remaining cells exhibited co-expression of up to 24 trypsins (Fig. 3e). For each trypsin, we then examined the relationship between the ubiquity of expression (the total number of cells in which that trypsin is expressed) and the number of cells in which it is co-expressed with other trypsins and found a strong positive correlation (Fig. 3f), confirming that the trypsins with the broadest expression profiles were most likely to be co-expressed with other trypsins.

The tryptome of *N. vectensis* is unique

To characterize the tryptome (all proteins with a trypsin domain) of *N. vectensis*, we searched the JGI gene models (<https://genome.jgi.doe.gov/Nemve1/Nemve1.home.html>) for all sequences containing a significant Trypsin or Trypsin_2 domain using *hmmsearch* (HMMER 3.1b2; <http://hmmer.org>) and constructed domain architecture diagrams for each protein (Fig. 4). Of the 72 trypsin gene models that remained after curation (see “Methods”), 28 encode a trypsin domain but lack any other conserved domains and the other 44 encode a trypsin domain and at least one additional conserved domain. In total, trypsin domains were found in association with 24 other domains in *N. vectensis*. To determine if any of these associated domains were overrepresented in the tryptome, we compared the abundance of trypsin-associated domains in the tryptome and in the proteins predicted from the JGI gene models ($N = 27,273$ protein predictions). Six domains were found to be represented in high abundance ($\geq 10\%$) in the tryptome: DIM, ShK, Lustrin_cystein, Sushi, MAM and SRCR (Fig. 4a). The DIM and Lustrin_cystein domains are present in low abundance throughout the predicted proteome (1 and 4 total domains, respectively), artificially inflating their perceived abundance in the tryptome. For ShK, Sushi, MAM, and SRCR, $\geq 15\%$ of the domains found in the proteome were associated with trypsins, suggesting the association between trypsin and each of these domains provides a strong selective advantage in the biology of *N. vectensis*.

To determine whether the makeup of the tryptome was unique to *N. vectensis*, we searched for proteins with these same domain architectures in representatives from all domains of life (other cnidarians, bilaterians, non-metazoan eukaryotes, and a selection of prokaryotes). Two domain architectures were found to be present across taxa: those



with only a trypsin domain, and those with a trypsin and a PDZ domain (Fig. 4b). Trypsin diversity appears to have expanded considerably with the evolution of multicellular animals, as both choanoflagellate lineages had fewer than 5 trypsins but the ctenophore *Mnemiopsis leidyi* and the placozoan *Trichoplax adhaerens* (representing two of the earliest diverging animal lineages) both have at least 20. Surprisingly, there was little conservation in trypsin domain architecture across animals. The tryptome of *N. vectensis* had more trypsin domain architectures in common with other actinarians (sea anemones) than with any other animal group; however, we still identified 3 trypsin architectures

unique to *N. vectensis* that were absent even from *Edwardsiella lineata* (a representative of the genus sister to *Nematostella*). Two of these (NVJ_105271 and NVJ_199428) represent unique associations between trypsin and other conserved domains (WSC and DIM, respectively) and the other (NVJ_105548) exhibits a novel arrangement of trypsin and its associated MAM domains (Fig. 4b).

Trypsins diversified independently in cnidarians and bilaterians

To characterize the diversification of animal trypsins, we built a phylogeny of trypsin domains from taxa

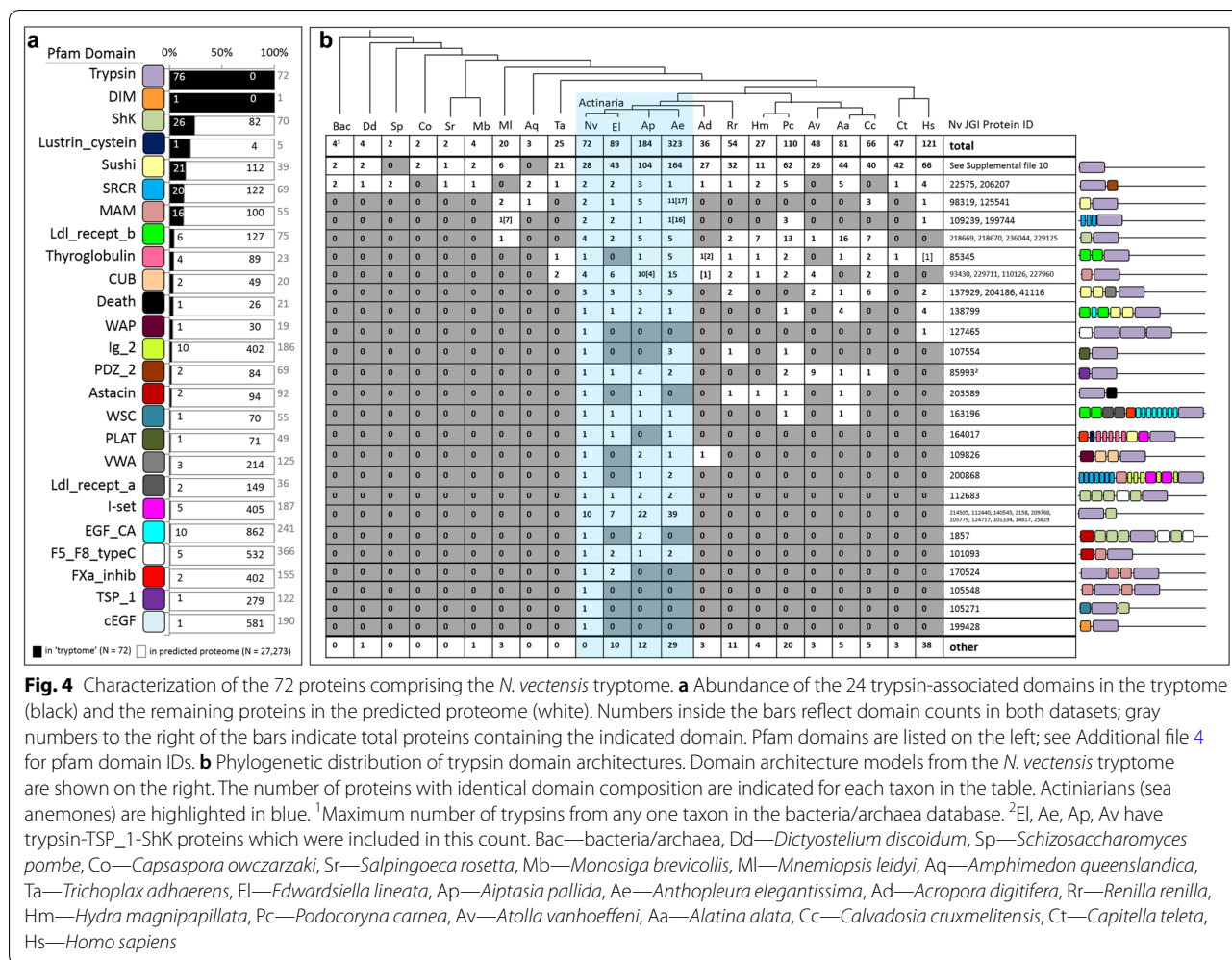


Fig. 4 Characterization of the 72 proteins comprising the *N. vectensis* tryptome. **a** Abundance of the 24 trypsin-associated domains in the tryptome (black) and the remaining proteins in the predicted proteome (white). Numbers inside the bars reflect domain counts in both datasets; gray numbers to the right of the bars indicate total proteins containing the indicated domain. Pfam domains are listed on the left; see Additional file 4 for pfam domain IDs. **b** Phylogenetic distribution of trypsin domain architectures. Domain architecture models from the *N. vectensis* tryptome are shown on the right. The number of proteins with identical domain composition are indicated for each taxon in the table. Actinarians (sea anemones) are highlighted in blue. ¹Maximum number of trypsins from any one taxon in the bacteria/archaea database. ²El, Ae, Ap, Av have trypsin-TSP_1-ShK proteins which were included in this count. Bac—bacteria/archaea, Dd—*Dictyostelium discoïdum*, Sp—*Schizosaccharomyces pombe*, Co—*Capsaspora owczarzewski*, Sr—*Salpingoeca rosetta*, Mb—*Monosiga brevicollis*, Ml—*Mnemiopsis leidyi*, Aq—*Amphimedon queenslandica*, Ta—*Trichoplax adhaerens*, El—*Edwardsiella lineata*, Ap—*Aiptasia pallida*, Ae—*Anthopleura elegantissima*, Ad—*Acropora digitifera*, Rr—*Renilla renilla*, Hm—*Hydra magnipapillata*, Pc—*Podocoryna carnea*, Av—*Atolla vanhoeffeni*, Aa—*Alatina alata*, Cc—*Calvadosia cruxmelitensis*, Ct—*Capitella teleta*, Hs—*Homo sapiens*

representing each of the 5 major animal lineages: bilaterians, cnidarians, placozoans, sponges, and ctenophores. Using this tree, we identify 6 clades of trypsins and classify them by their function in human: a non-catalytic group, the intracellular trypsins, tryptases and transmembrane trypsins, trypsins involved in coagulation and immune response, chymotrypsins, and the clade including granzymes, pancreatic trypsins, kallikreins, hepatocyte growth factors, and elastases (Fig. 5a). Each of these includes representatives from bilaterians, cnidarians, and at least one placozoan, sponge, or ctenophore and likely represents the suite of trypsin clades present in the last common ancestor of animals. The *N. vectensis* tryptome includes representatives of 5 of 6 clades likely present in the common ancestor of animals; *N. vectensis* may have lost representatives of the tryptase/transmembrane clade as this these trypsins appear to be present in *M. leidyi*, *A. digitifera*, and bilaterians (Fig. 5a, Additional file 2).

We compared the distribution of conserved domains from different clades of trypsins in *N. vectensis* and *H. sapiens* (Fig. 5b). In *N. vectensis*, domain diversity is greatest among the trypsins that group with human chymotrypsins ($N=14$), followed by trypsins in the immune/coagulation group ($N=10$), the “pancreatic” group (including granzymes, kallikreins, HGF, and elastase) ($N=5$), and intracellular trypsins ($N=2$). Trypsins from the non-catalytic clade lack associated domains completely. Four trypsin-associated domains (Sushi, EGF_CA, CUB, and FXa_inhibition) were found in the immune/coagulation clades from both *N. vectensis* and *H. sapiens*, the CUB domain was found in chymotrypsins from both taxa, and the PDZ domain is restricted to the intracellular clade of trypsins in both taxa; surprisingly, there were no other domains found in common between *N. vectensis* and *H. sapiens* trypsins from the same clade (see Additional file 3 for distribution of human trypsin domain architectures).

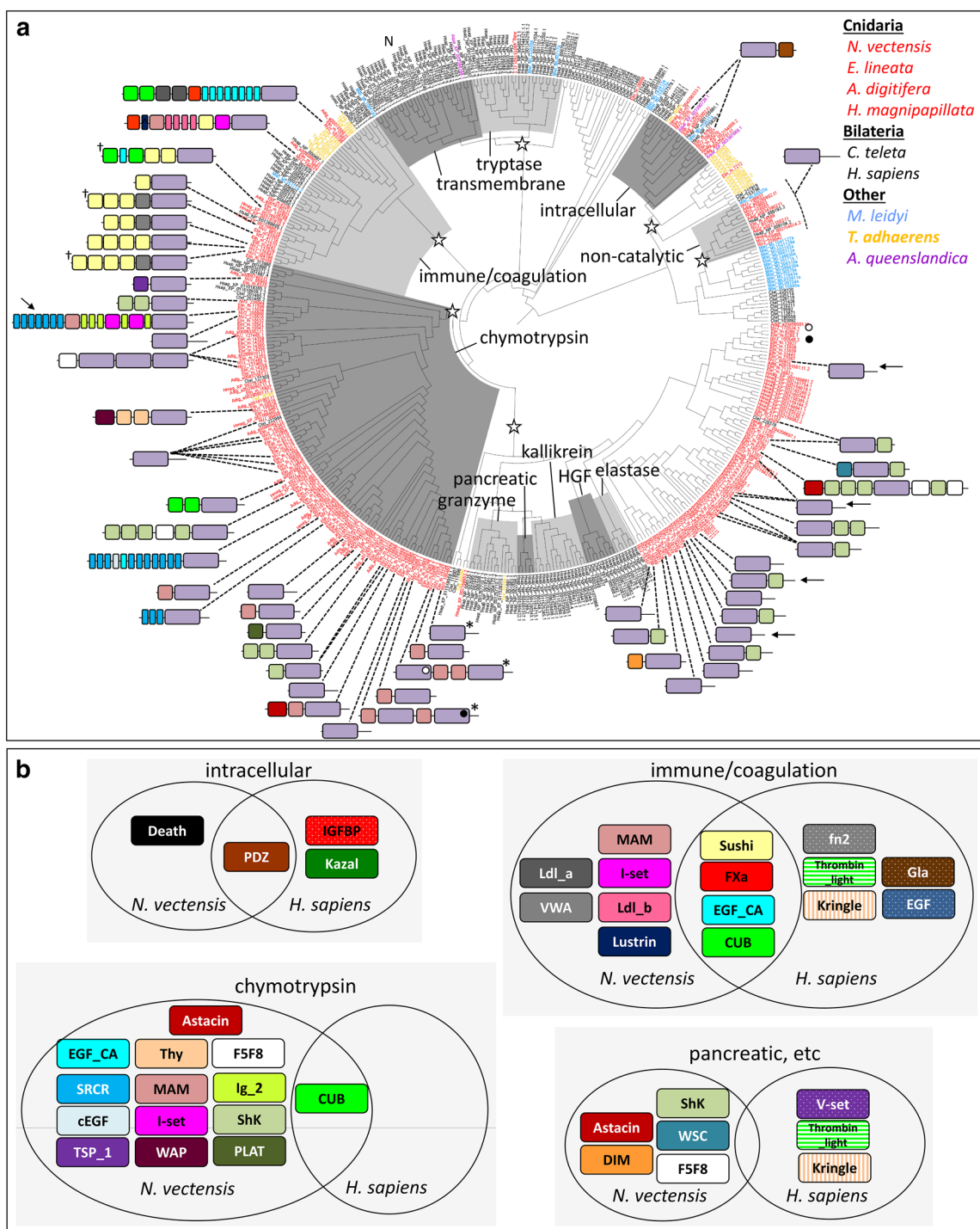
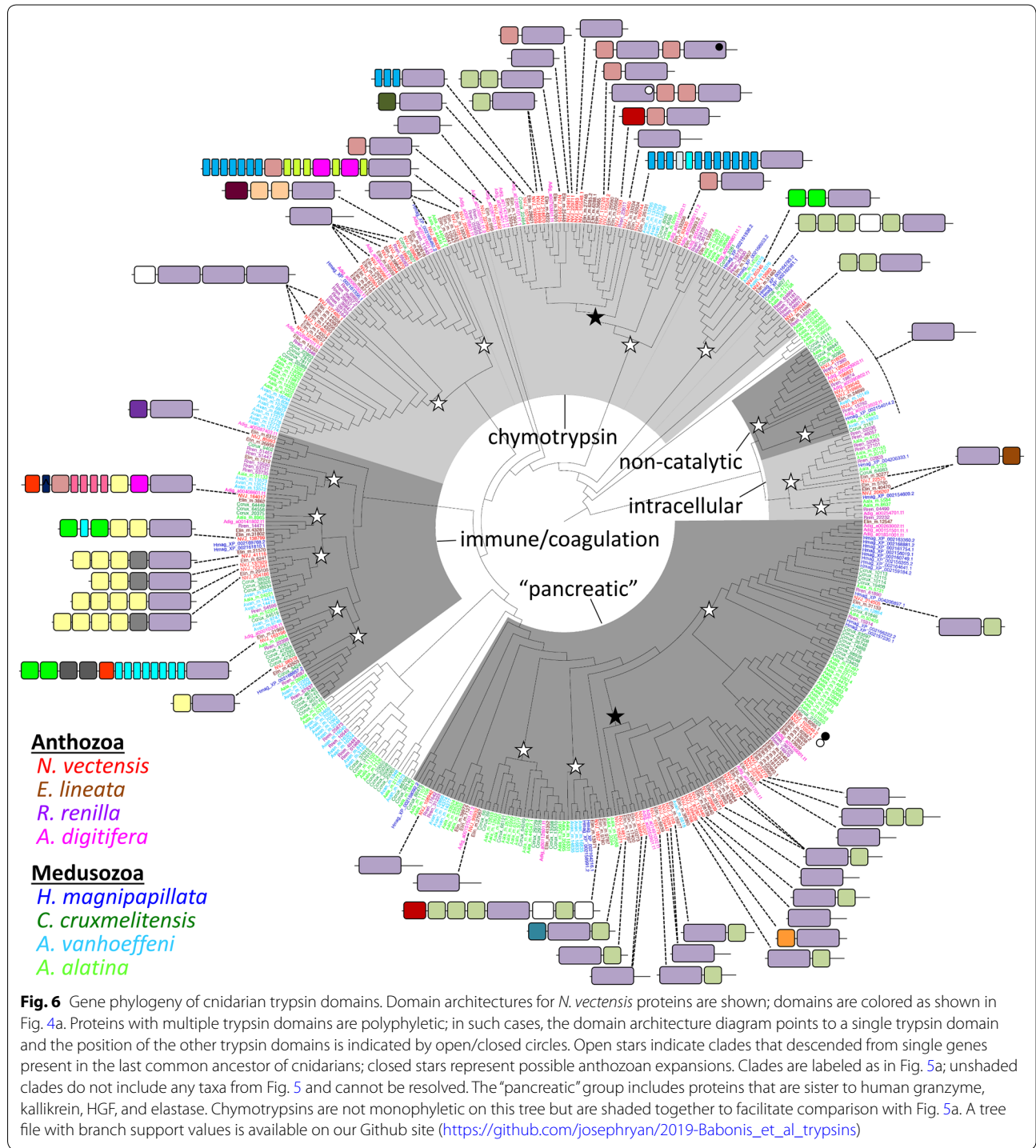


Fig. 5 Evolutionary history of animal trypsins. **a** Phylogeny of animal trypsins; clades are named after the orthologs from *H. sapiens*. Domain architectures for *N. vectensis* proteins are shown. Proteins with multiple trypsin domains are polyphyletic; in such cases, the domain architecture diagram points to a single trypsin domain and the position of the other trypsin domains is indicated by open/closed circles. Open stars indicate clades that descended from single genes present in the last common ancestor of animals. “N” indicates the position of human neurotrypsin (NP_003610.2). The trypsin-Death protein groups with trypsin-PDZ proteins but is not found on this tree. *N. vectensis* proteins characterized by in situ hybridization are indicated: arrows (this study), †[34], or * [12]. A tree file with branch support values is available on our Github site (https://github.com/josephryan/2019-Babonis_et_al_trypsins). **b** Comparison of the domain content of trypsins from *N. vectensis* and *H. sapiens*. The non-catalytic trypsins do not have associated domains and are not included. The tryptase/transmembrane trypsins are not represented in *N. vectensis* and are also not included. The “pancreatic” group includes granzyme, kallikrein, HGF, and elastase

To determine whether the tryptome diversity of *N. vectensis* is reflective of other cnidarians, we built a phylogeny using representatives of each class within Cnidaria (Fig. 6). We identify 16 clades of trypsins that include representatives of at least two lineages of anthozoans and two lineages of medusozoans,

suggesting that these clades may have been present in the stem cnidarian. Two clades (the trypsin-MAM and trypsin-ShK clades) seem to have undergone further expansion in anthozoans after their divergence from medusozoans.



The *Nematostella* tryptome diversified through numerous mechanisms

To understand the mechanisms generating trypsin diversity in *N. vectensis*, we examined the evolutionary relationships of the 72 trypsin proteins in the tryptome (Fig. 7a). Among the 72 predicted proteins, 85% (61/72) had all three conserved residues constituting the catalytic triad and are likely to function as proteases, 79% (57/72) were predicted to have a signal peptide and are presumably secreted, and 7% (5/72) were predicted to have a transmembrane domain (see Additional file 4). The trypsin superfamily, therefore, exhibits evidence of functional specialization through protein primary structure modification, directing protein localization to specific sub-cellular compartments. Furthermore, 4 of the 5 clades of trypsins from *N. vectensis* (excluding the intracellular clade) include secreted trypsins, membrane-bound trypsins, and trypsins with divergent sequence that have likely lost their catalytic function, suggesting that spatial and functional specialization has evolved multiple times in different lineages of trypsins.

Numerous trypsins from the “pancreatic” and chymotrypsin clades were associated with ShK domains. Likewise, over 30% (26/82) of the ShK domains in *N. vectensis* are associated with trypsins (Fig. 4a). To determine if the combination of the trypsin and ShK domains may have duplicated together, we built a phylogeny of all 108 ShK domains from the *N. vectensis* proteins predicted from gene models (Fig. 7b). Despite the abundance of trypsin-ShK associations, the ShK domains from sister trypsins were almost never monophyletic, suggesting this domain is gained and lost easily. Consistent with this, every ShK domain in the tryptome of *N. vectensis* was encoded by only a single exon (Additional file 5), supporting the rapid evolution of the tryptome through exon shuffling. Two trypsin-ShK proteins (NVJ_218669 and NVJ_218670) were found to be sister in both phylogenies, suggesting they arose by duplication of the combined domains. These two genes are encoded on the same scaffold and are separated by approximately 1 kb of genomic DNA; thus, they are likely the result of a recent tandem duplication event. The ShK domain is a short peptide found in a K-channel inhibitor originally isolated from the sea anemone *Stichodactyla helianthus* [17]. What role the ShK domain plays when it is paired with the trypsin domain is not known but the overabundance of these two combined domains in cnidarian tryptomes (Additional file 6) combined with the multiple independent origins of this domain combination in *N. vectensis* (Fig. 7b) suggests that the pairing provides a strong selective advantage in the biology of cnidarians.

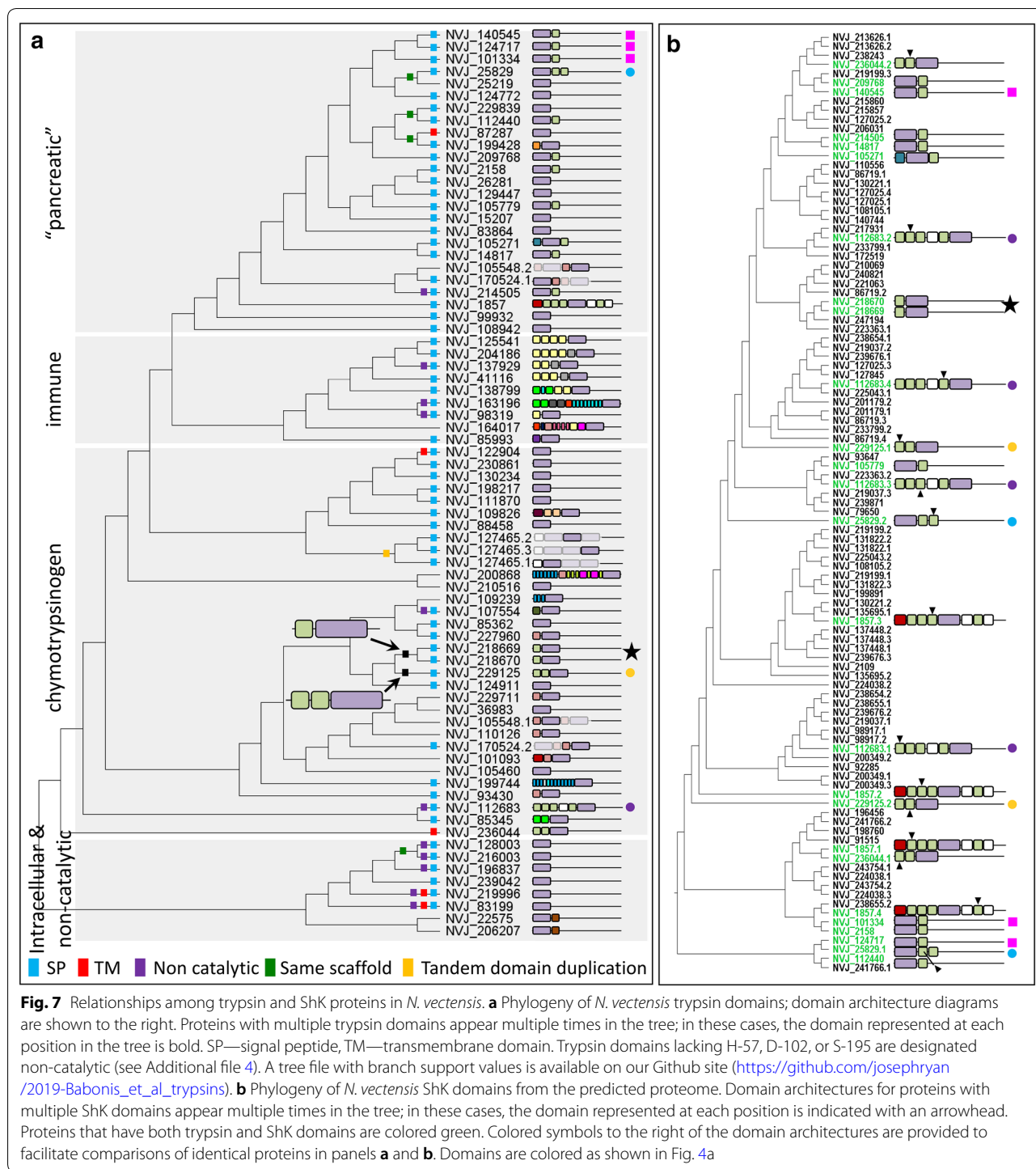
Multidomain proteins are more common than proteins with only a single domain as domain recombination

increases versatility in protein function [18]. Selection to maintain the catalytic activity of the trypsin domain while allowing the context in which this domain is expressed to vary was a critical component of diversification in this gene superfamily. In support of this, we found surprisingly little conservation in trypsin-associated domains across animals, even among cnidarians (Fig. 4, Additional file 6), suggesting that the associated domains have been continuously gained and lost in each lineage. Furthermore, nearly 40% (28/72) of the proteins comprising the *N. vectensis* tryptome have only a trypsin domain (Fig. 7); yet, these trypsin-only proteins did not form a monophyletic group (Figs. 5, 6), suggesting that trypsin domains themselves may be rapidly gained and lost from evolutionarily unrelated proteins. Indeed, trypsin diversification does occur independent of the acquisition of associated domains. One gene from the tryptome of *N. vectensis* (NVJ_127465) encodes three trypsin domains, all of which form a monophyletic group suggesting this gene structure arose through tandem duplication of the trypsin domain (Fig. 5). The tryptome from *H. sapiens* also includes two proteins with three trypsin domains each (Additional file 6). While these 6 trypsin domains from *H. sapiens* are found in the tryptase/transmembrane clade (Additional file 3), the three domains in NVJ_127465 group with chymotrypsins (Fig. 5). Thus, despite their similar domain architecture, triple-trypsin domain proteins appear to have evolved multiple times.

Several other mechanisms contributed to diversification of trypsins in *N. vectensis*. We identified four cases where sister trypsins are found on the same scaffold (Fig. 7a), suggesting tandem gene duplication. Furthermore, while most (70/72) of the trypsin domains were encoded across multiple exons (Additional file 5), two genes (NVJ_128003 and NVJ_216003) lack introns completely, and likely arose through recent retrotransposition. These two genes are also on the same scaffold, suggesting that retrotransposition may have been followed by tandem gene duplication.

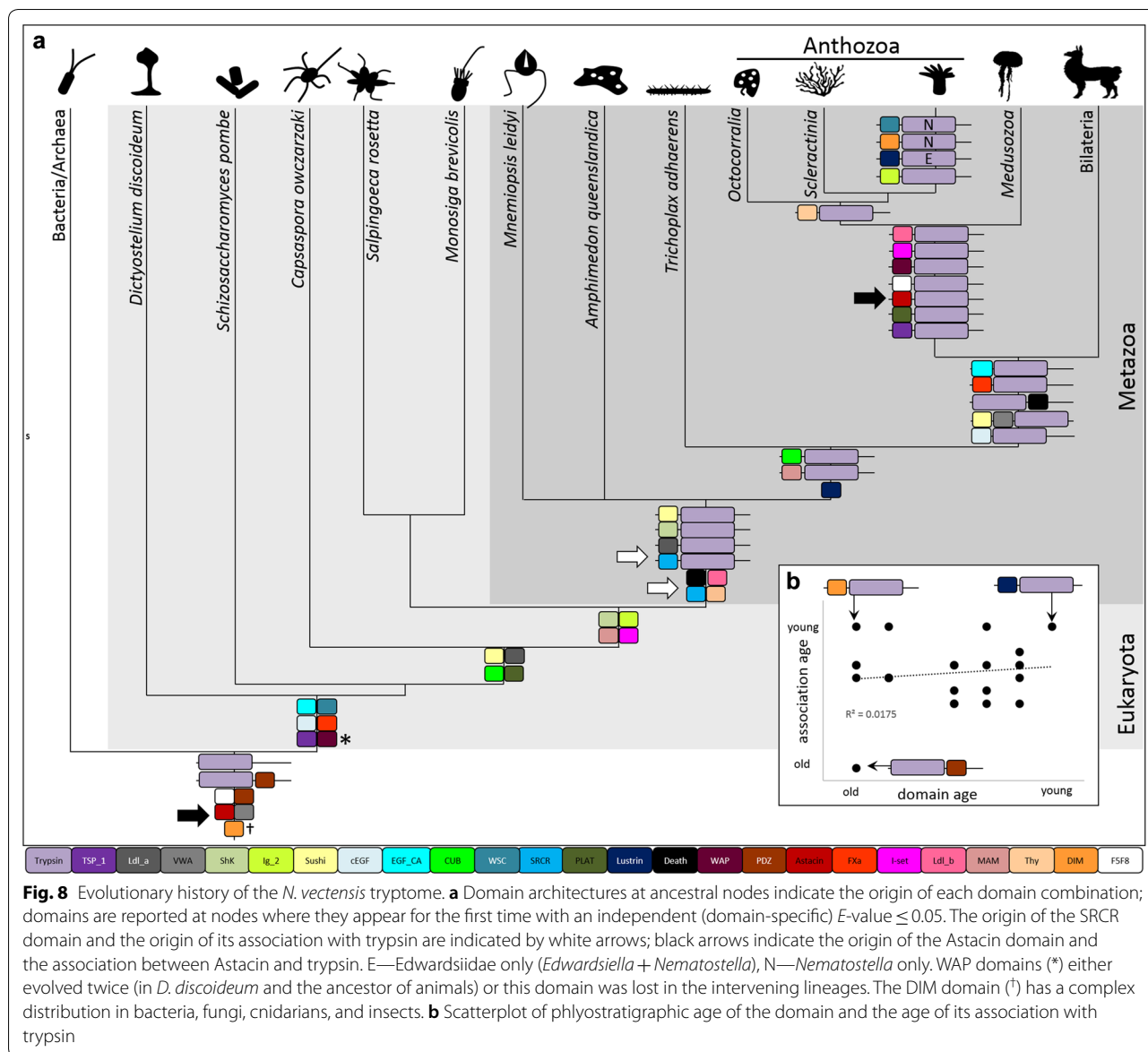
Trypsin diversity increases through new associations with old domains

Gene age can be estimated using a phylostratigraphic approach; in such analyses, the minimum age of a gene is inferred by identifying the last common ancestor in which the gene is present [19, 20]. We examined the age of the trypsins found in *N. vectensis* and the age of each associated domain across all domains of life to understand the evolution of trypsin diversity. Trypsin-PDZ and a subset of the trypsin-only proteins likely arose before bacteria/archaea split from eukaryotes, over 2 billion years ago (Fig. 8). While trypsin-only proteins are present in every lineage examined, trypsin-PDZ



proteins appear to have been lost in several taxa including *C. owczarzaki*, *M. leidyi*, *A. vanhoeffeni*, and *C. cruxmelitensis* (Fig. 4). All other associations between trypsin and other conserved domains appear to have originated after the stem metazoan diverged from the

rest of life (~800 million years ago) [21]. Many of the trypsin-associated domains originated long before they became associated with trypsin; for example, the Astacin domain was present in the ancestor of all life but the trypsin- Astacin association likely did not arise



until the origin of Cnidaria (Fig. 8a). By contrast, the SRCR domain and its association with trypsin likely arose in the stem metazoan as trypsin-SRCR proteins were found in *M. leidy* (Additional file 6).

There is no relationship between the age of the domain and the origin of its association with trypsin (Fig. 8b). Two trypsin associations were found only in *N. vectensis*: trypsin-DIM (NVJ_199428) and trypsin-WSC (NVJ_105271), and one association was found only in *Edwardsiidae* (*Nematostella* + *Edwardsiella*): trypsin-Lustrin_cystein (NVJ_164017). The WSC domain is present throughout eukaryotes (Fig. 8a) but was associated with trypsin only in *N. vectensis*. The Lustrin_cystein domain seems to have arisen

in the last common ancestor of parahoxozoa (Placozoa + Cnidaria + Bilateria). These two associations represent extreme cases whereby trypsin diversity in *N. vectensis* arose through acquisition of both young (Lustrin_cystein) and old (WSC) domains.

Discussion

The not-so-simple cnidarian ectoderm

Although cnidarian body plans develop from only two tissue layers, morphological diversity varies widely across taxa. Similarly, only a dozen or so morphologically unique cell types have been described [10, 22], but cnidarian genomic and functional diversity rival that of any bilaterian lineage [4, 23]. While the ectodermal

layer comprising the external and pharyngeal epithelia may be contiguous, these regions are morphologically and functionally distinct in *N. vectensis* (Fig. 1) [10, 12]. In this study, we further demonstrate that the continuous layer of internal ectoderm from the pharynx through the mesenteries is equally heterogeneous. The pharyngeal ectoderm houses numerous zymogen and mucous cells, while the ectoderm of the mesenteries houses only the former (Fig. 2). This anatomical heterogeneity is supported by variable gene expression: some proteases are expressed throughout the pharyngeal and mesentery ectoderm, while others are restricted only to the mesentery ectoderm (Fig. 3a, b) (also see [12, 24]). Furthermore, the combinatorial expression of only two proteases can result in the development of at least three distinct cell types (Fig. 3c) and some cells express over 20 different trypsins (Fig. 3e). Together, the combination of a diverse tryptome and extensive trypsin co-expression suggests that cell functional diversity in cnidarians may well exceed historical expectations.

We found no evidence of endodermal gland cells (zymogen type or mucous type) in our TEM or in situ hybridization results (Figs. 2, 3, Additional file 1). Indeed, all non-neuronal secretory cells (including mucous cells, zymogen cells, and cnidocytes), are restricted to the ectoderm in *N. vectensis* but their distribution is heterogeneous. Zymogen cell diversity, for example, is much higher in the internal than the external ectoderm (Fig. 2, Additional file 1). This is consistent with the histological analyses of Frank and Bleakney [10] but seems to be in contrast with the distribution of gland cells in medusozoans. In *Hydra*, for example, zymogen gland cells are found exclusively in the endoderm [25]. These observations suggest that the internalization of the ectoderm in anthozoans was a pivotal event in the diversification of specialized zymogen cells. Cell products secreted from the tentacle ectoderm may quickly become diluted in the water column, whereas the closed environment of the gastrovascular cavity limits the space over which secreted products can diffuse; thus, internalization created distinct selective pressures in different regions of the ectoderm. Indeed, selection for secretion of digestive enzymes into the enclosed gastrovascular cavity may have driven the development of gland cells in the internal ectoderm of anthozoans and the endoderm of medusozoans (and many bilaterians). As such, we see no reason to homologize the ectoderm of anthozoan mesenteries and the endodermal lining of the vertebrate midgut/pancreas [12]. We consider it more likely that these tissues have converged on similar morphologies and gene expression profiles in response to similar selection pressures associated with extracellular digestion.

***Nematostella vectensis* trypsins have many putative functions**

The trypsin domain catalyzes the cleavage of polypeptides at internal amino acid residues and is therefore essential for processing large proteins into smaller peptide chains. Digestive trypsins are synthesized in secretory cells with zymogen-type secretory granules where they are packaged into vesicles for release into the gut. We show that there are at least ten morphologically distinct zymogen gland cell types in the pharyngeal and mesentery ectoderm of *N. vectensis* (Fig. 2), that numerous proteases are expressed in these tissues (Fig. 3), and that the vast majority of trypsins in *N. vectensis* encode a signal peptide (Fig. 7a). Using published single-cell expression data [16], we identified 10 putative gland cells that express trypsins, at least two of which also express synaptotagmin (Additional file 4), which facilitates fusion of the vesicle with the cell membrane during regulated secretion. These combined results strongly support a role for the internal ectoderm in extracellular protein degradation in *N. vectensis*.

Numerous trypsins were expressed outside of the putative gland cells identified by Sebe-Pedros et al. [16]. At least 20 cells categorized by these authors as neurons exhibited trypsin expression but unlike gland cells, the maximum number of trypsins expressed by any putative neuron is three (Additional file 4). We show trypsin-expressing cells differentiating very early in development, in the invaginating pharynx/mesenteries (Fig. 3), where several neurons (including those expressing RFamide and Elav) are also undergoing terminal differentiation [22, 26]. Indeed, the trypsin protease NVJ_99932 (Fig. 3) is co-expressed with two other trypsins (NVJ_230861 and NVJ_130234) in a putative neuron expressing GABA and dopamine receptors (Additional file 7). In vertebrates, secretion of neurotrypsin from the pre-synaptic membrane facilitates degradation of the extracellular matrix during synaptic plasticity and axon guidance [27], providing clues to the potential function of these neutrally expressed trypsins in *N. vectensis*. Although 17 different trypsins were expressed in putative neurons, none of the trypsins from *N. vectensis* clustered with human neurotrypsin (Fig. 5); as such, these functions may have been acquired independently from different ancestral trypsins in cnidarians and bilaterians.

Trypsins are important regulators of tissue remodeling, and upregulation of trypsins and other proteases often coincides with wound healing and tissue regeneration [28]. Recent studies of regeneration in *N. vectensis* demonstrated that a new pharynx will regenerate from the oral ends of the mesenteries after amputation [29] and that many proteases are expressed abundantly during this process [30]. Thus, the mesenteries appear

to play an important role in directing the tissue remodeling process in *N. vectensis*. In support of this, a study of wound healing in response to a body wall injury demonstrated that the mesenteries come into direct contact with damaged tissue during the healing process [31]. This study also showed that two trypsins (NVJ_107554 and NVJ_112683) are among the top genes undergoing upregulation during wound healing in *N. vectensis*. While NVJ_112683 was not reported in the single-cell dataset, NVJ_107554 is expressed in two putative gland cells (metacells C12 and C19; Fig. 3, Additional file 4). Thus, mesentery-expressed trypsins play important roles in the cell and tissue biology of *N. vectensis* during wound healing and regeneration and these roles may vary through ontogeny.

Beyond their roles in digestion and tissue remodeling, trypsins are an important component of the innate immune system. In vertebrates, immune trypsins play a role in blood coagulation and are part of the complement system which recognizes foreign particles [32]. In symbiotic cnidarians, immune trypsins play a role in the beneficial interaction between the host and the alga [33]. While *N. vectensis* does not host symbiotic algae, a previous study aimed at understanding the origin of the innate immune system reported the expression of three immune system trypsins in *N. vectensis*: MASP (NVJ_138799) and two paralogs of Factor B (NVJ_41116, NVJ_204186), each of which was expressed in the endoderm (gastrodermis) of juvenile polyps [34]. We found that the two factor B orthologs were also co-expressed in a single putative gastrodermal cell (Additional file 4) further supporting a role for the endoderm in the immune response of *N. vectensis*. One trypsin (NVJ_127465) was not reported in the single-cell dataset [16] but was among the genes found to be significantly upregulated in the tissue-specific transcriptome of nematosomes, which may also play a role in the immune system of *N. vectensis* [11]. This gene clustered with human chymotrypsin genes, not the immune system trypsins (Fig. 5), suggesting it acquired a role in the immune system secondarily.

Trypsin functional diversity has undergone numerous expansions

Our phylogeny of animal trypsins suggests that the last common ancestor of animals may have had at least six major groups of trypsins (Fig. 5), and extensive lineage-specific trypsin duplication occurred thereafter. Sponges are unusual among animals in that they have only three trypsins—two trypsin-PDZ paralogs and a trypsin-Sushi protein (Additional file 6). This suggests either extensive loss of trypsins in Porifera or independent diversification of trypsins in ctenophores and in the stem of paraxozoa. The evolutionary history of trypsin domain

architectures sheds little light on this topic. While trypsin-Sushi, trypsin-SRCR, and trypsin-ShK proteins are found in ctenophores, the patchy distribution of these proteins across animals makes it difficult to determine whether this pattern has resulted from multiple gains or multiple losses (Figs. 4, 8, Additional file 6). Given that the association between trypsin and ShK seems to have arisen multiple times in *N. vectensis* (Fig. 7), we think that rapid independent gains of beneficial domain associations (including trypsin-Sushi and trypsin-SRCR) was a primary driver of trypsin diversification throughout the evolution of animals.

The ancestral cnidarian seems to have had a far more diverse suite of trypsins than the ancestral animal. Indeed, our data suggest there were at least 17 lineages of trypsins present in the last common cnidarian ancestor (Fig. 6) and 12 of the associations between trypsin and another conserved domain in *N. vectensis* are specific to cnidarian lineages (Fig. 8). There was extensive divergence in the trypsin gene superfamily during the diversification of cnidarians but anthozoans seem to have undergone additional radiations in at least two trypsin clades (Fig. 6). Anthozoans are the most speciose group of cnidarians and are largely sessile; thus, selection for trophic specialization and sympatric niche diversification may be stronger among anthozoans than medusozoans. Diversification of the trypsin superfamily was facilitated by gene duplication followed by the acquisition of additional domains (Fig. 8); however, we found no relationship between domain age and the age of its association with trypsin (Fig. 8b). Therefore, trypsin domain architectures diversify continuously and are not dependent on the origin of novel domains.

Secretory cells and the evolution of cnidarian body plans

Resolving the embryological origin of cnidarian gland cells will be important for understanding the evolution of life history in Cnidaria. If the anthozoan polyp body plan is ancestral to all cnidarians [35], then the origin of strobilation (medusa formation) and its associated tissue remodeling in the stem medusozoan may have necessitated the sacrifice of the internalized tissue layers of the ancestral pharynx and mesenteries. In this case, the stem medusozoan may have overcome this loss by shifting the development of their gland cell population to the endoderm without sacrificing the selective advantage of secreting their products into the gastrovascular cavity. In support of this hypothesis, gland cells in *Hydra* are known to undergo differentiation in a location-specific manner, suggesting the identity of this cell lineage is highly sensitive to positional cues from other cells in their environment [36]. Furthermore, a recent study of single-cell dynamics in *Hydra* demonstrated that gland

cells acquire their identity in the endoderm only after their precursor migrates out of the ectoderm and across the mesoglea [37]. Both of these studies point to the highly plastic nature of gland cell identity in *Hydra* but similar analyses in more medusozoans are needed to understand the relationship between gland cell development and cnidarian life history evolution.

Conclusions

The transition from unicellular to multicellular life was marked by many transitions that enabled functional specialization. Unicellular taxa used trypsins for intracellular protein regulation, but the origin of the regulated secretion system created new opportunities for protease activity in multiple tissue compartments. Secretion of molecules to the extracellular space enabled the development of the nervous, endocrine, immune, and digestive systems, and permitted spatial and temporal separation of multiple functions performed by a single cell. The diversification of animals was associated with a large

expansion of trypsins. Trypsins with transmembrane domains first appear in the choanoflagellates but trypsins with signal peptides did not appear until the origin of animals. Subsequent duplication and divergence (e.g., through exon shuffling and retrotransposition) of genes encoding secreted proteases enabled nuanced variation in the function of these secretory cells before the increase in anatomical diversity (Fig. 9).

Methods

Electron microscopy, cell proliferation assay, and in situ hybridization

Adult polyps were immobilized for 10 min in 7.5% MgCl₂ and processed for transmission electron microscopy as described previously [38]. Samples were imaged on a Hitachi HT7700 at the University of Hawaii’s Biological Electron Microscopy facility. To identify proliferating nuclei, live adult polyps were incubated in 100 μM EdU (in 1/3× seawater) for 30 min at room temperature. Animals were then immobilized and fixed briefly (1.5 min) at

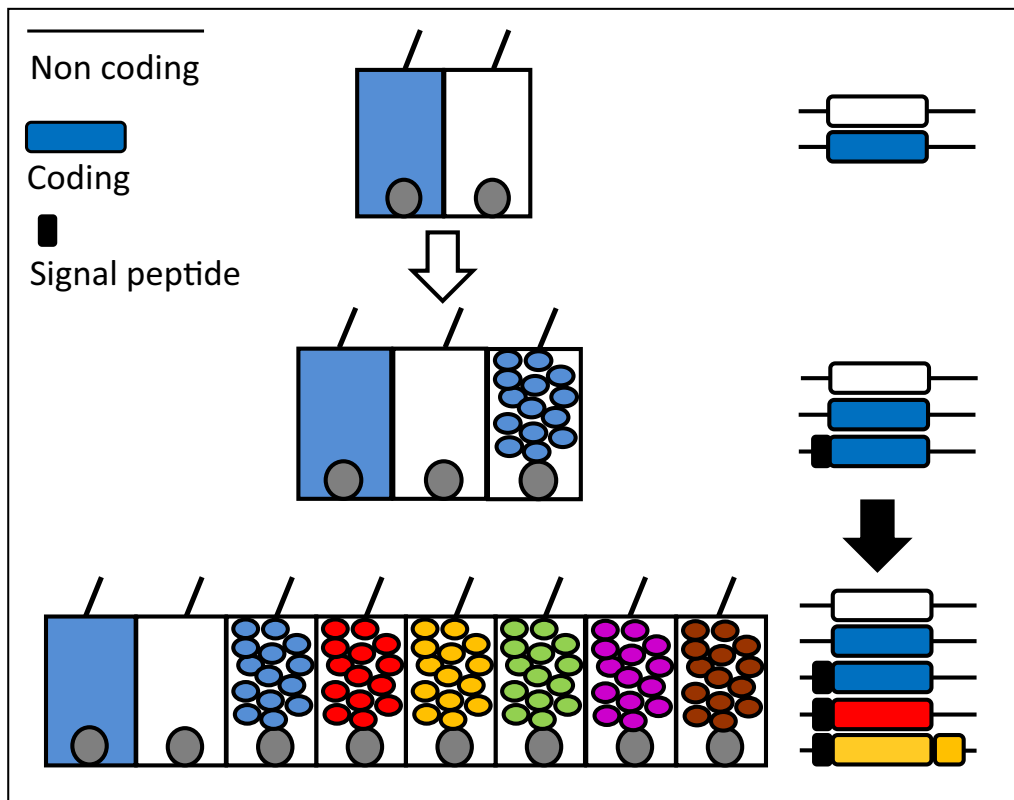


Fig. 9 Secretory vesicles permit functional expansion without anatomical variation. Blue and white cells reflect the intracellular expression of blue and white gene products. The origin of the signal peptide directing proteins to the regulated secretion system/secretory vesicle (white arrow) permitted segregation of gene products into two distinct compartments: intracellular and intravesicular. The subsequent duplication and divergence of the blue gene (black arrow) could result in the acquisition of numerous new cell types through unique and combinatorial gene expression. Green—co-expression of blue and yellow, purple—co-expression of blue and red, brown—co-expression of blue, yellow, and red

25 °C in 4% paraformaldehyde with 0.2% glutaraldehyde in phosphate buffered saline with 0.1% Tween-20 (PTw) followed by a long fixation (60 min) in 4% paraformaldehyde in PTw at 4 °C. Fixed tissues were analyzed using the Click-IT EdU kit (#C10340, Invitrogen, USA) following the manufacturer's protocol. Nuclei were counter stained in a 30-min incubation in DAPI at room temperature and samples were imaged on a Zeiss 710 confocal microscope at the Whitney Lab for Marine Bioscience. To characterize the localization of target genes, we performed in situ hybridization following a standard protocol for *N. vectensis* [39].

Protein domain analysis

To identify trypsin-domain proteins from *N. vectensis*, we first searched the JGI protein models (indicated throughout by NVJ_X) using the default settings with hmmsearch (HMMER 3.1b2; <http://hmmmer.org/>) and two target HMMs: Trypsin (PF00089) and Trypsin_2 (PF13365). This approach yielded 99 putative trypsin-domain containing proteins with an E -value $\leq 1e-05$ [40]. Where multiple partial non-overlapping trypsin domains were identified from the same protein, we assumed these represented one single contiguous domain [41]. Based on a reciprocal BLAST comparison with transcriptome data available publicly [11], we found 68/99 of the JGI gene models coding for trypsin proteins were incomplete. We manually corrected these sequences using the transcriptome data and used these corrected sequences for downstream analyses. We then used the transcriptome data to search protein models for evidence of pseudogenes (with premature stop codons) using the translation and alignment features in Geneious v 7.1.8 (<https://www.geneious.com>) and manually examined models for duplicate predictions using the JGI genome viewer. Based on these analyses, we removed 27 sequences, resulting in a final set of 72 curated trypsin protein models (FASTA file available at: https://github.com/josephryan/2019-Babonis_et_al_trypsins).

We examined the domain architecture of trypsin proteins from *N. vectensis* by searching for non-Trypsin domains in the amino acid sequences using hmmscan (HMMER 3.1b2) and the complete Pfam-A database (downloaded Oct 27, 2017). Hmmscan identifies regions of similarity between protein queries and domain models (protein profiles) derived from numerous proteins within the family from a range of animals [42]. Following the protocol of Koch et al. [40], we ran hmmscan using the default parameters and report only those domains with an independent (domain-specific) E -value ≤ 0.05 that were found in a protein containing a significant Trypsin (or Trypsin_2) domain. Domains that overlapped by $\leq 20\%$ were both retained; when the overlap was $> 20\%$

the domain with the lower E -value was retained. In addition to domain analysis, we manually searched an alignment of the corrected set of trypsin protein models from *N. vectensis* for the conserved residues that comprise the trypsin catalytic triad (necessary for inferring protease activity): H-57, D-102, or S-195. Finally, we searched the corrected amino acid sequences for signal peptides and transmembrane domains using SignalP v4.1 [43] and TMHMM v2.0 [44], respectively.

To characterize the origin of trypsin domain architecture, we used hmmscan with the same approach described above to identify and characterize trypsins from representatives across all domains of life (trypsin protein IDs for all taxa provided in Additional file 7). We sampled three bilaterians (*Capitella teleta*, *Branchiostoma floridae*, *Homo sapiens*), 10 cnidarians (*Nematostella vectensis*, *Edwardsiella lineata*, *Aiptasia pallida*, *Anthopleura elegantissima*, *Acropora digitifera*, *Renilla renilla*, *Hydra magnipapillata*, *Calvadosia cruxmelitensis*, *Atolla vanhoeffeni*, *Alatina alata*), three non-planulozoan animals (*Mnemiopsis leidyi*, *Amphimedon queenslandica*, *Trichoplax adhaerens*), five non-metazoan eukaryotes (*Dictyostelium discoideum*, *Schizosaccharomyces pombe*, *Capsaspora owczarzaki*, *Monosiga brevicolis*, *Salpingoeca rosetta*) and a combined database of representative archaea and bacteria (*Candidatus aquiluna*, *Candidatus nitrosopumilus*, *Candidatus pelagibacter*, *Glaciecola pallidula*, *Marinobacter adhaerens*, a marine gamma proteobacterium, and a marine group I thaumarchaeote). Protein models were predicted from transcriptome data previously for *N. vectensis*, *E. lineata*, *A. pallida*, *A. elegantissima*, *A. alatina*, *A. vanhoeffeni*, and *P. carnea* [11]. Proteomes for *R. renilla* and *C. cruxmelitensis* were predicted from the transcriptome data reported by Kayal et al. [35] using the same methods. For all other taxa, protein models were downloaded directly (commands available at: https://github.com/josephryan/2019-Babonis_et_al_trypsins).

Phylotocol (phylogenetic transparency)

All phylogenetic investigations were planned prior to running any analyses and all are reported in this manuscript. In most cases, these analyses were outlined beforehand in a phylotocol [45] that is posted on our GitHub site: https://github.com/josephryan/2019-Babonis_et_al_trypsins. Any analyses performed prior to being added to our phylotocol were later added to the document and justified.

Phylogenetics

To understand the diversification of animal trypsins, we built a phylogeny using predicted proteins from *M. leidyi*, *A. queenslandica*, *T. adhaerens*, *N. vectensis*, *E.*

lineata, *H. magnipapillata*, *C. teleta*, *B. floridae*, and *H. sapiens*. First, we used a custom script to generate alignments from these protein files using the Trypsin HMM (commands available at: https://github.com/josephryan/2019-Babonis_et_al_trypsins). All trees were constructed using a maximum likelihood framework with RAxML and IQ-TREE [46–48]. We used the model finder function with IQ-TREE (-m MF) to determine the best substitution model for the alignment and then ran three approaches in parallel: RAxML with 25 parsimony starting trees, RAxML with 25 random starting trees, and a single run with IQ-TREE (which, by default, uses a broad sampling of initial starting trees). We selected the best tree by comparing the maximum likelihood scores of all three approaches. To assay branch support, we ran 1000 bootstraps using the rapid bootstrapping function with RAxML(-x); tree files with branch support are available on our Github site (https://github.com/josephryan/2019-Babonis_et_al_trypsins).

Using the Trypsin HMM, we recovered 97% (70/72) of the curated trypsin proteins from *N. vectensis*. The two remaining trypsin proteins (NVJ_23745 and NVJ_203589) were recovered using the Trypsin_2 HMM (note: the Trypsin_2 HMM recovered only 89% (64/72) of the curated trypsins). To understand the evolutionary relationships of these two proteins to the rest of the trypsin family, we generated another phylogeny using the same procedure as above and an alignment built using the Trypsin_2 HMM. This best tree recovered using the Trypsin_2 HMM is provided in Additional file 2. After inspecting both trees, we removed sequences from *B. floridae* for ease of viewing and reran the full analyses. All tree files and alignment files are available on our Github site (https://github.com/josephryan/2019-Babonis_et_al_trypsins).

To evaluate whether *N. vectensis* has undergone lineage-specific expansion of trypsins or if the common ancestor of all cnidarians had an equally diverse tryptome, we built a phylogeny of trypsin proteins from cnidarians only using a subset of the proteomes listed above. Specifically, we used four species of anthozoans (*N. vectensis*, *E. lineata*, *R. renilla*, *A. digitifera*) and four medusozoans (*H. magnipapillata*, *C. cruxmelitensis*, *A. vanhoeffeni*, *A. alata*). We then pruned all non-*Nematostella* taxa from this tree using Phyutility v.2.2.6 [49] to generate a tree for *N. vectensis* trypsins only. To examine the evolutionary history of ShK domains from *N. vectensis*, we used hmmsearch with the ShK HMM and a custom script (as above) to identify and align all ShK domains from the predicted proteome. We then used the approach described above to produce a phylogeny of ShK domains.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13227-019-0138-1>.

Additional file 1. Gland cells of the external ectoderm. **a–c** Mucus cells (false colored yellow) in the tentacle ectoderm. **d** Zymogen cell (false colored green) in the body wall ectoderm. White arrows point to cnidocytes, black arrowheads point to electron dense apical vesicles in cells adjacent to gland cells. **a, b** SEM, **c, d** TEM. Scale bars: black—5 μ m, white—10 μ m.

Additional file 2. Phylogeny of Trypsin_2 domains across animals. Protein models for *N. vectensis* are shown. NVJ_203589 and NVJ_23745 were not detected by the trypsin HMM and do not appear in Fig. 5. Colors as in Fig. 4a.

Additional file 3. Human trypsin domain architecture mapped on the animal trypsin phylogeny. Proteins with multiple trypsin domains are polyphyletic; in such cases, the diagram points to a single trypsin domain and the position of the other trypsin domains is indicated by symbols. *N. vectensis* proteins characterized by in situ hybridization are indicated by arrows (this study), †[34], or *[12].

Additional file 4. Excel file tabulating: presence/absence of signal peptides and transmembrane domains in *N. vectensis* trypsins, amino acid sequences for trypsin catalytic domains for all taxa, pfam IDs for all domains, and a summary of single-cell expression of trypsins published previously [16].

Additional file 5. All trypsin domains are encoded by multiple exons in *N. vectensis* (excluding the NVJ_128003 and NVJ_216003) but many of the associated domains are encoded by a single exon (indicated by triangle). Domains that span intron/exon boundaries by ten or fewer nucleotides were considered to be encoded by a single exon.

Additional file 6. Trypsin protein domain architectures from all taxa examined in this study.

Additional file 7. Trypsin protein IDs from all taxa examined in this study.

Acknowledgements

We are grateful to Tina Carvahlo and Dr. Marilyn Dunlap of the Biological Electron Microscopy Facility at the University of Hawai'i Manoa for their assistance with electron microscopy.

Authors' contributions

Study design/concept: LSB, MQM, JFR; animal/tissue methods: LSB, CE; phylogenetics: LSB, JFR; other analyses: LSB; writing: LSB; editing: MQM, JFR, CE. All authors read and approved the final manuscript.

Funding

This work was supported by the National Aeronautics and Space Administration (Grant NNX14AG70G to MQM) and the National Science Foundation (Grant 1542597 to JFR).

Availability of data and materials

All data generated or analyzed during this study are included in this published article (and its additional files).

Ethics approval and consent to participate

No human subjects were used in this study and the invertebrate animals used in this study are not subject to regulation by animal ethics committees.

Consent for publication

Not applicable.

Competing interests

The authors declare that have no competing interests.

Author details

¹ Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, FL 32080, USA. ² Department of Biology, University of Florida, Gainesville, FL 32611, USA.

Received: 20 July 2019 Accepted: 13 September 2019

Published online: 30 September 2019

References

- Magie CR, Daly M, Martindale MQ. Gastrulation in the cnidarian *Nematostella vectensis* occurs via invagination not ingression. *Dev Biol*. 2007;305(2):483–97.
- Fritzenwanker JH, Genikhovich G, Kraus Y, Technau U. Early development and axis specification in the sea anemone *Nematostella vectensis*. *Dev Biol*. 2007;310(2):264–79.
- Finnerty JR, Pang K, Burton P, Paulson D, Martindale MQ. Origins of bilateral symmetry: hox and dpp expression in a sea anemone. *Science*. 2004;304(5675):1335–7.
- Kusserow A, Pang K, Sturm C, Hrouda M, Lentfer J, Schmidt HA, et al. Unexpected complexity of the Wnt gene family in a sea anemone. *Nature*. 2005;433(7022):156–60.
- Wijesena N, Simmons DK, Martindale MQ. Antagonistic BMP-cWNT signaling in the cnidarian *Nematostella vectensis* reveals insight into the evolution of mesoderm. *P Natl Acad Sci USA*. 2017;114(28):E5608–15.
- Leclere L, Rentzsch F. RGM regulates BMP-mediated secondary axis formation in the sea anemone *Nematostella vectensis*. *Cell Rep*. 2014;9(5):1921–30.
- Moiseeva E, Rabinowitz C, Paz G, Rinkevich B. Histological study on maturation, fertilization and the state of gonadal region following spawning in the model sea anemone, *Nematostella vectensis*. *Plos One*. 2017;12(8):e0182677.
- Jahnel SM, Walzl M, Technau U. Development and epithelial organisation of muscle cells in the sea anemone *Nematostella vectensis*. *Front Zool*. 2014;11:44.
- Renfer E, Amon-Hassenzahl A, Steinmetz PRH, Technau U. A muscle-specific transgenic reporter line of the sea anemone, *Nematostella vectensis*. *P Natl Acad Sci USA*. 2010;107(1):104–8.
- Frank P, Bleakney J. Histology and sexual reproduction of the anemone *Nematostella vectensis* Stephenson 1935. *J Nat Hist*. 1976;10(4):441–9.
- Babonis LS, Martindale MQ, Ryan JF. Do novel genes drive morphological novelty? An investigation of the nematosomes in the sea anemone *Nematostella vectensis*. *Bmc Evol Biol*. 2016;16:114.
- Steinmetz PRH, Aman A, Kraus JEM, Technau U. Gut-like ectodermal tissue in a sea anemone challenges germ layer homology. *Nat Ecol Evol*. 2017;1(10):1535–42.
- Page MJ, Di Cera E. Serine peptidases: classification, structure and function. *Cell Mol Life Sci*. 2008;65(7–8):1220–36.
- Page MJ, Di Cera E. Evolution of peptidase diversity. *J Biol Chem*. 2008;283(44):30010–4.
- Carlgren O. Actinaria Part I. The Danish In-golf Expedition. Copenhagen. 1921;5(9):1–241.
- Sebe-Pedros A, Saudemont B, Chomsky E, Plessier F, Mailhe MP, Renno J, et al. Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-Seq. *Cell*. 2018;173(6):1520.
- Castaneda O, Harvey AL. Discovery and characterization of cnidarian peptide toxins that affect neuronal potassium ion channels. *Toxicon*. 2009;54(8):1119–24.
- Vogel C, Teichmann SA, Pereira-Leal J. The relationship between domain duplication and recombination. *J Mol Biol*. 2005;346(1):355–65.
- Domazet-Lošo T, Brajkovic J, Tautz D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet*. 2007;23(11):533–9.
- Long M, Betran E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. *Nat Rev Genet*. 2003;4(11):865–75.
- Dohrmann M, Worheide G. Dating early animal evolution using phylogenomic data. *Sci Rep*. 2017;7:3599.
- Marlow HQ, Srivastava M, Matus DQ, Rokhsar D, Martindale MQ. Anatomy and development of the nervous system of *Nematostella vectensis*, an anthozoan cnidarian. *Dev Neurobiol*. 2009;69(4):235–54.
- Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science*. 2007;317(5834):86–94.
- Moran Y, Praher D, Schlesinger A, Ayala A, Tal Y, Technau U. Analysis of soluble protein contents from the nematocysts of a model sea anemone sheds light on venom evolution. *Mar Biotechnol*. 2013;15(3):329–39.
- Haynes JF, Davis LE. Ultrastructure of zymogen cells in *Hydra viridis*. *Z Zellforsch Mik Ana*. 1969;100(2):316.
- Nakanishi N, Renfer E, Technau U, Rentzsch F. Nervous systems of the sea anemone *Nematostella vectensis* are generated by ectoderm and endoderm and shaped by distinct mechanisms. *Development*. 2012;139(2):347–57.
- Almonte AG, Sweatt JD. Serine proteases, serine protease inhibitors, and protease-activated receptors: roles in synaptic function and behavior. *Brain Res*. 2011;1407:107–22.
- Daley WP, Peters SB, Larsen M. Extracellular matrix dynamics in development and regenerative medicine. *J Cell Sci*. 2008;121(3):255–64.
- Amiel AR, Johnston HT, Nedoncelle K, Warner JF, Ferreira S, Rottinger E. Characterization of morphological and cellular events underlying oral regeneration in the sea anemone, *Nematostella vectensis*. *Int J Mol Sci*. 2015;16(12):28449–71.
- Schaffer AA, Bazarsky M, Levy K, Chalifa-Caspi V, Gat U. A transcriptional time-course analysis of oral vs aboral whole-body regeneration in the sea anemone *Nematostella vectensis*. *Bmc Genomics*. 2016;17:718.
- Dubuc TQ, Traylor-Knowles N, Martindale MQ. Initiating a regenerative response; cellular and molecular features of wound healing in the cnidarian *Nematostella vectensis*. *Bmc Biol*. 2014;12:24.
- Heutinck KM, ten Berge IJM, Hack CE, Hamann J, Rowshani AT. Serine proteases of the human immune system in health and disease. *Mol Immunol*. 2010;47(11–12):1943–55.
- Poole AZ, Kitchen SA, Weis VM. The role of complement in cnidarian-dinoflagellate symbiosis and immune challenge in the sea anemone *Aiptasia pallida*. *Front Microbiol*. 2016;7:519.
- Kimura A, Sakaguchi E, Nonaka M. Multi-component complement system of Cnidaria: c3, Bf, and MASP genes expressed in the endodermal tissues of a sea anemone, *Nematostella vectensis*. *Immunobiology*. 2009;214(3):165–78.
- Kayal E, Bentlage B, Pankey MS, Ohdera AH, Medina M, Plachetzki DC, et al. Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. *Bmc Evol Biol*. 2018;18:68.
- Siebert S, Anton-Erxleben F, Bosch TCG. Cell type complexity in the basal metazoan Hydra is maintained by both fly stem cell based mechanisms and transdifferentiation. *Dev Biol*. 2008;313(1):13–24.
- Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, Juliano CE. Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science*. 2019;365(6451):9314.
- Dubuc TQ, Dattoli AA, Babonis LS, Salinas-Saavedra M, Rottinger E, Martindale MQ, et al. *In vivo* imaging of *Nematostella vectensis* embryogenesis and late development using fluorescent probes. *Bmc Cell Biol*. 2014;15:44.
- Wolenski FS, Layden MJ, Martindale MQ, Gilmore TD, Finnerty JR. Characterizing the spatiotemporal expression of RNAs and proteins in the starlet sea anemone, *Nematostella vectensis*. *Nat Protoc*. 2013;8(5):900–15.
- Koch BJ, Ryan JF, Baxevas AD. The diversification of the LIM superclass at the base of the Metazoa increased subcellular complexity and promoted multicellular specialization. *PLoS ONE*. 2012;7(3):e33261.
- Triant DA, Pearson WR. Most partial domains in proteins are alignment and annotation artifacts. *Genome Biol*. 2015;16:99.
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, et al. The Pfam protein families database. *Nucleic Acids Res*. 2004;32:D138–41.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011;8(10):785–6.
- Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305(3):567–80.
- Ryan JF, Debiase MB. Phylotocol: Promoting transparency and overcoming bias through publicly posted, a priori methodological protocols in phylogenetics. *Integr Comp Biol*. 2018;58:410.

46. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587.
47. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–74.
48. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3.
49. Smith SA, Dunn CW. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics*. 2008;24(5):715–6.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

