 doi:10.20944/preprints202102.0221.v1

The *Daphnia* carapace and the origin of novel structures

Authors: Heather S. Bruce

Affiliations: Marine Biological Laboratory, Woods Hole, MA.
*Correspondence to: hbruce@mbl.edu

Abstract

Understanding how novel structures arise is a central question in evolution. The carapace of the waterflea *Daphnia* is a bivalved "cape" of exoskeleton that surrounds the animal, and has been proposed to be one of many novel structures that arose through repeated co-option of genes that also pattern insect wings. To determine whether the *Daphnia* carapace is a novel structure, the expression of *pannier*, the *Iroquois* gene *aurucan*, and *vestigial* are compared between *Daphnia*, *Parhyale*, and *Tribolium*. The results suggest that the *Daphnia* carapace did not arise by cooption, but instead represents an elongated ancestral exite (lateral lobe or plate) that emerges from a proximal leg segment that was incorporated into the *Daphnia* body wall. The *Daphnia* carapace therefore appears to be homologous to the *Parhyale* tergal plate and the insect wing. In addition, the vg-positive region that gives rise to the *Daphnia* carapace also appears to be present in *Parhyale* and *Tribolium*, which do not form a carapace. Thus, rather than a novel structure resulting from gene co-option, the carapace appears to have arisen from an ancient, conserved developmental field that persists in a cryptic state in other arthropod lineages, but in *Daphnia* became elaborated into the carapace. Cryptic persistence of serially homologous developmental fields may thus be a general solution for the origin of many novel structures.

Keywords: novelty, arthropod, appendages, Parhyale, Tribolium, Daphnia, carapace

Introduction

Understanding how novel structures arise is a central question in evolution. Novelties are often defined as structures that are not homologous to any structure in the ancestor nor to any other structure in the same organism(1). Co-option of genetic pathways has become the dominant explanation for the origin of novel structures within the field of evolutionary developmental biology (evo-devo). But what if that assumption is incorrect? And if it is incorrect, where do these structures actually come from?

Arthropods are an ideal system in which to study evolutionary novelties, as their legs and bodies are highly modular and are decorated with structures that have a diversity of morphologies and perform a variety of functions, many of which have been proposed to be novel

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0221.v1

structures. There are four main groups of arthropods: insects (beetles, butterflies, etc), crustaceans (shrimps, crabs, etc), myriapods (centipedes, millipedes, etc), and chelicerates (spiders, scorpions, etc). Bruce 2020(*2*) and Bruce 2021(*4*) showed that the leg segments of all four groups of arthropods can be aligned in a one-to-one fashion (Fig. 1). This analysis, drawing on over a century of morphological and embryological studies as well as gene expression and loss-of-function studies, generated a model for understanding the homologies of any structure on any arthropod leg. The ancestral number of leg segments in arthropods appears to be 8 (*5–8*). However, most arthropods appear to have incorporated one or two ancestral proximal leg segments into the body wall (*2*, *4*, *7*, *9*, *10*), i.e. leg segments 7 and 8, counting from the terminal claw as segment 1 (Fig. 1). Bruce 2020(*2*) and Bruce 2021(*4*) found that the division between true body wall and the incorporated leg segments that now function as body wall can be distinguished by the expression of *pannier* (*pnr*) and *Iroquois* complex genes such as *aurucan* (*ara*). In the embryos of all arthropods examined to date in three of the four arthropod clades (*Drosophila*, insect; *Tribolium*, insect; *Parhyale*, crustacean; and *Acanthoscurria*, chelicerate), *ara* expression brackets the hypothesized incorporated 8th leg segment, while *pnr* is expressed in the dorsal-most tissue and appears to mark the "true" body wall. I sought to apply this leg segment alignment model to understand the homology of an iconic crustacean structure, the carapace of the water flea, *Daphnia*.

The *Daphnia* carapace has been proposed to be a novel structure that arose through gene co-option(*11*). In the water flea *Daphnia magna*, the carapace is a bivalved "cape" of exoskeleton that surrounds the animal (Fig. 2). The beautiful work from Shiga 2017 (*11*) showed that the *Daphnia* carapace emerges from the dorsal region of the head (maxilla 2), and, like insect wings, is composed of a bilayered sheet of ectodermal cells (*11*). They also showed that the carapace expresses and requires the "wing" genes *vestigial* and *scalloped*. Shiga 2017 propose that the *Daphnia* carapace and other flat, lateral lobes in arthropods arose by multiple instances of co-option of wing patterning genes. However, the analysis in Bruce 2020 suggests an alternative hypothesis. This study showed that incorporated leg segments that now function as body wall can nevertheless express exites - lateral lobes that express "wing" genes, such as the insect wing and the *Parhyale* tergal plate (Fig. 1). These exites emerge from what appears to be dorsal-lateral body wall. Based on this model, the morphological and molecular data in Shiga 2017 suggests that the *Daphnia* carapace did not arise by co-option, but instead represents an

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0221.v1

exite on an incorporated 8th leg segment. The *Daphnia* carapace would therefore be homologous to the *Parhyale* tergal plate(*2*, *3*) and the insect wing(*2*).

Results

To test the proximal-distal register of the *Daphnia* carapace, the expression of *pannier*, an *Iroquois* gene, and the wing/exite patterning gene *vg* was examined in *Daphnia magna* embryos. A single *pnr* gene was identified in *Daphnia* which was the reciprocal best blast hit of *Drosophila*, *Tribolium*, and *Parhyale pnr* (*2*). A single *Iroquois* complex gene was identified in *Daphnia* which was the reciprocal best blast hit of *Drosophila*, *Tribolium*, and *Parhyale ara*, so this *Daphnia* gene is hereafter referred to as *ara* (*2*). *Daphnia vg* was identified previously by Shiga 2017(*11*).

If the *Daphnia* carapace is the exite of the incorporated 8th leg segment, then it will be bracketed dorsally and ventrally by *ara* expression domains, and *pnr* will be expressed dorsal to the carapace. Alternatively, if the carapace is a dorsal, non-leg structure, then *pnr* expression should extend into the carapace, and the two domains of *ara* should be located ventral to the carapace. In either case, *vg* will be expressed around the perimeter of the carapace.

Consistent with the hypothesis that the carapace is an exite on leg segment 8, *Daphnia vg* is expressed in the perimeter of the carapace, *pnr* is restricted to a single, dorsal stripe above the carapace, and the carapace is bracketed by a dorsal and a ventral *ara* expression domain (Figs. 3 and 4). Notably, in the head where the carapace forms, the dorsal *ara* expression domain is greatly expanded, covering the proximal half of the carapace (Fig. 3b, 4c, f). This is expected if the carapace is an elongated expansion of the 8th leg segment. Conversely, in the legs of the thorax, where no carapace forms, the dorsal *ara* expression is not expanded (Fig. 4i), and looks like the dorsal *ara* domain of arthropods which lack a carapace, including *Tribolium* (Fig. 4g), *Parhyale* (Fig. 4h), and *Acanthoscurria* (*4*).

If the *Daphnia* carapace is the exite of the 8th leg segment of a modified leg (mouthpart) on the head, there may be previously unappreciated vestiges of these exites on the heads of arthropods that do not have a carapace. In support of this hypothesis, *vg* is expressed in the head of *Tribolium* in the dorsal/proximal maxilla and antenna (Fig. 5). This *vg* domain is bracketed by *ara* expression, just like the insect wing, the *Parhyale* tergal plate, and the *Daphnia* carapace. This region is therefore is presumably serially homologous to the 8th leg segment. The antenna

 doi:10.20944/preprints202102.0221.v1

vg domain may pattern a small shelf-like protrusion called the gena, which forms the characteristic indentation on the *Tribolium* eye. However, the maxilla *vg* domain does not pattern any obvious structure. In *Parhyale*, *vg* patterns the flanges that protect the mouthparts, because these flanges are reduced when *vg* is knocked out (compare Figs. 1f, j in Clark-Hachtel 2020 (*3*). These flanges are serially homologous to the 8th leg segment, because they are bracketed by *ara* expression. Thus, rather than new, co-opted domains of *vg* expression, these *vg* head domains are ancient and conserved.

Discussion

Many arthropods have lateral structures (carapaces, plates, gills, glands, etc), the homologies of which are frequently debated(*8*, *12*). The work here demonstrates that the proximal-distal identity of these structures can be elucidated using the expression of *pnr* and *Iroquois* genes such as *ara*, and further pinpointed with leg segment joint markers like odd-skipped and distal leg markers such as *Dll*, and exite genes such as *vg*. This molecular triangulation strategy can illuminate the homologies of long-debated proximal structures. Based on this model, predictions can be made about the expected gene expression and homologies of any leg- or body wall-associated arthropod structure.

In contrast to repeated gene co-option (a dominant narrative for the origin of novel structures), the work presented here, and in Bruce 2020(2) and Bruce 2021(*4*), provides an alternative hypothesis: these leg-associated structures (and perhaps other novel structures too) are better explained as arising at previously unappreciated serially homologous locations from already existing developmental fields (i.e. crustacean leg exites). In this model, developmental, morphogenetic fields can persist in a cryptic, morphologically unrecognizable form (such as silverfish paranotal lobes) in intermediate lineages and become elaborated again in later lineages (such as insect wings), such that they may no longer be recognizable as the ancestral structure and appear to pop out of nowhere. In this model, structures would not have to be continuously present in a morphologically obvious state from ancestor to descendant in order to be homologous. Rather than de novo co-options, these developmental fields are always there, persisting in a dormant, truncated, or highly modified state, and de-repressed in various lineages to form apparently novel structures. Cryptic persistence of developmental fields therefore

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0221.v1

provides a more mechanistically satisfying explanation for the origin of novel structures, in contrast to the current, rather fuzzy theory of repeated gene co-option.

In fact, cryptic persistence is hinted at in the definition of crustacean exites. In crustacean morphology, exites are defined as lateral lobes that may occur on any of the three proximal leg segments (leg segment 7 – 8, “protopod”) and on any post-mandibular leg along the anterior-posterior axis(*8*, *12*). However, it is widely appreciated that not all possible exites occur on all proximal leg segments on all legs in all crustacean taxa. Exites occur variably across different crustacean taxa: they may appear in some lineages but not in others, on some legs and not others, and on some leg segments but not others (*8*, *12*, *13*). These observations imply that there is some latent ability to form a structure at these locations: sometimes the pathway is activated and sometimes it is repressed, yet all these structures are considered exites, rather than repeatedly novel structures. The concept of cryptic persistence makes this implicit assumption explicit, and furthermore provides a mechanistic, molecular developmental explanation for the variable occurrence of exites.

For decades, researchers have been finding that many exciting structures on the legs and body wall of many insects express the same genes (the “wing gene network”). These were interpreted to be wing serial homologs or partial wing homologs, and have often been interpreted as evidence of gene co-option. This has contributed to a widespread narrative that co-option is a common mechanism for the generation of novel structures. However, given that insects have incorporated two leg segments into the body wall, each with an exite(*2*), and that crustacean exites may occur on any leg segment along the anterior-posterior axis(*12*, *13*), all of these insect structures may be exites inherited from their crustacean ancestors. Thus, the lack of a molecular, developmental model from the ancestral crustacean group gave the appearance that insect structures had repeatedly arisen de novo, and spurring a mistaken foundation for one of the most widely discussed notions in evo devo. This is a clarion call for a diversity of research model organisms in every major clade, so that structures no longer appear to be novel simply because the ancestral state has not yet been investigated.

In addition to the phylogenetically unequal distribution of research organisms, a second reason for the dominance of the gene co-option model in explaining the origin of novel structures may be the poor naming of "master regulator". In many discussions surrounding the origin of novel structures, authors propose that a novel structure came about when an entire genetic

 doi:10.20944/preprints202102.0221.v1

pathway was co-opted(*11, 14*). The underlying assumption seems to be that this structure and its genetic pathway was not present at that location before, and that the tissue doing the co-opting is unrelated to the tissue it is co-opting from, i.e. they are different developmental fields. The related assumption seems to be that it is relatively straightforward for a master regulator to become misexpressed in an unrelated tissue (i.e. outside of that master regulator's developmental field at serially homologous locations, i.e. a tissue not previously primed with dozens of other TFs) in order to activate the entire structure in an entirely novel, non-serially homologous location. For example, the idea that dorsal non-leg tissue co-opted leg exite genes to form wings on the back. When a more detailed molecular explanation is written out for conceptual terms like co-option or cross-wiring, they become more dubious.

Perhaps the most thoroughly explored novel structure from a regulatory standpoint is spots on fly wings(*15*). Rather than recruiting a small number of large effect TFs, it was found that hundreds of genes contributed to the evolution of pigment spots on the wings. Wing spots are a relatively simple character, compared to something such as a wing that is a large, three dimensional structure with multiple tissue and cell types. Thus, if novel wing spots require hundreds of genes, then more complex structures would presumably require even more genes. This example makes the idea of pathway co-option less plausible.

A better model for the origin of apparently novel structures may be the Hox genes, where serially homologous tissues may switch between alternate identities. In this model, misexpression of “master regulators” or selector genes at previously unappreciated serially homologous locations would be *expected* to transform tissue towards the structures it patterns. Unique phenotypes may arise due to the selector gene being expressed in a different Hox or proximal-distal regulatory environment, or because the structures have not been expressed at this body position for millions of years and will thus communicate differently with the genetic pathways at this position.

Criteria should be formed to distinguish structures that have been de-repressed at a serially homologous location. For example, de-repressed serial homologs would be expected to appear in the same developmental register, to be composed of the same germ layer and similar tissue type (which in itself may be difficult to determine with certainty), and to express most of the genes in the pathway. In contrast to pathway co-option, gene co-option may be plausible, and would be expected to take genes piece-meal from different pathways, like butterfly eyespots,

 doi:10.20944/preprints202102.0221.v1

rather than entire pathways, which would require co-opting multiple regulatory genes simultaneously.

Many lateral, leg-associated structures across arthropods may be exites. Examples include the lateral organs of arachnids, glands of pycnogonid larvae, trachea of myriapods and insects, abdominal gills of insects, gin traps of insects, thoracic styli of jumping bristletails, scarab beetle support structures, and oostegites of amphipod crustaceans. Rather than being repeated co-options of "wing" genes, or "partial wing homologs", these structures may simply be different kinds of exites, and the observed differences in gene expression profiles would be due to differences in the type and therefore function of the exite (i.e. respiratory gill/trachea, secretory gland, rigid plate, etc).

Developmental fields are related to vestigial structures in the sense that vestigial structures are generated from developmental fields. However, a developmental field can exist in a dormant, repressed state where it does not form a structure at all, vestigial or otherwise, and yet it is still capable of forming a structure in the future if it becomes de-repressed. Cryptic persistence of developmental fields therefore reframes the discussion to focus on the molecular networks that generate structures. The concept therefore encompasses vestigial structures, when the field forms a structure, but expands the idea by pointing out that not all fields generate an actual structure, and also explains why structures can re-appear at serially homologous locations. Developmental fields occur at serially repeated locations, and not all of these locations will form a structure, vestigial or otherwise. But these locations should be identified and examined in order to discover dormant fields so that homologies with structures of interest may be established. Identifying all relevant serially homologous developmental fields within an organism is the only way to know for sure whether a structure is expressed at a novel or a serially homologous position.

The work presented here calls into question whether gene co-option should be invoked to explain the origin of novel structures as often as it is. On a final note, it is interesting to consider that if cryptic persistence is commonplace, and most "novel" structures evolved from existing structures, then many familiar developmental fields may be far more evolvable, and have far more ancient origins, than currently thought.

References

1. G. B. Müller, G. P. Wagner, Novelty in evolution: restructuring the concept. *Annu Rev Ecol Syst* (1991) (available at https://www.annualreviews.org/doi/pdf/10.1146/annurev.es.22.110191.001305).

2. H. S. Bruce, N. H. Patel, Knockout of crustacean leg patterning genes suggests that insect wings and body walls evolved from ancient leg segments. *Nat Ecol Evol*. **4**, 1703–1712 (2020).

3. C. M. Clark-Hachtel, Y. Tomoyasu, Two sets of candidate crustacean wing homologues and their implication for the origin of insect wings. *Nat Ecol Evol* (2020), doi:10.1038/s41559-020-1257-8.

4. H. S. Bruce, How to align arthropod leg segments. *BioRxiv*, 43 (2021).

5. J. Yang, J. Ortega-Hernández, D. A. Legg, T. Lan, J. Hou, X. Zhang, Early Cambrian fuxianhuiids from China reveal origin of the gnathobasic protopodite in euarthropods. *Nat Commun*. **9**, 1–9 (2018).

6. J. W. Shultz, Morphology of locomotor appendages in Arachnida: evolutionary trends and phylogenetic implications. *Zool J Linn Soc*. **97:1–56** (1989) (available at http://onlinelibrary.wiley.com/doi/10.1111/j.1096-3642.1989.tb00552.x/full).

7. R. E. Snodgrass, *Morphology and mechanism of the insect thorax* (City of Washington, Smithsonian institution, 1927), vol. 80.

8. G. A. Boxshall, The evolution of arthropod limbs. *Biol. Rev.* **79**, 253–300 (2004).

9. Y. Kobayashi, K. Niikura, Y. Oosawa, Y. Takami, Embryonic development of Carabus insulicola (Insecta, Coleoptera, Carabidae) with special reference to external morphology and tangible evidence for the subcoxal theory. *J. Morphol.* **274**, 1323–1352 (2013).

10. R. Matsuda, Morphology and evolution of the insect thorax. *Memoirs of the Entomological Society of Canada*. **Volume 102**, 5–431 (1970).

11. Y. Shiga, Y. Kato, Y. Aragane-Nomura, T. Haraguchi, T. Saridaki, H. Watanabe, T. Iguchi, H. Yamagata, M. Averof, Repeated co-option of a conserved gene regulatory module underpins the evolution of the crustacean carapace, insect wings and other flat outgrowths. **120**, 621–24 (2017).

12. G. A. Boxshall, D. Jaume, Exopodites, epipodites and gills in Crustaceans. *Arthropod Systematics & Phylogeny*, 1–27 (2009).

13. F. R. Schram, *Crustacea* (Oxford University Press, USA, 1986).

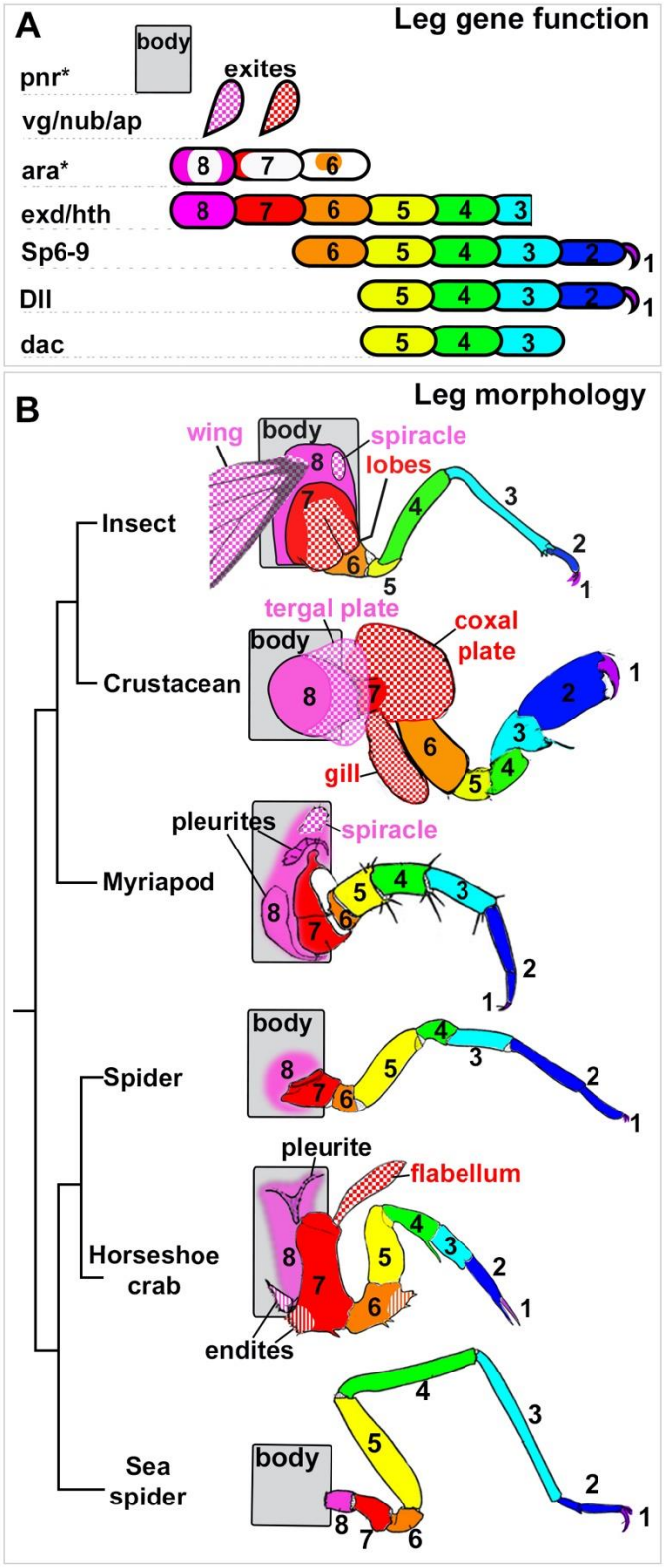
14. J. H. Bowsher, H. F. Nijhout, Partial co-option of the appendage patterning pathway in the development of abdominal appendages in the sepsid fly Themira biloba. *Dev Genes Evol*. **219**, 577–587 (2010).

15. Y. Fukutomi, S. Kondo, A. Toyoda, S. Shigenobu, S. Koshikawa, Transcriptome analysis reveals *wingless* regulates neural development and signaling genes in the region of wing pigmentation of a polka-dotted fruit fly. *FEBS J*. **288**, 99–110 (2021).

16. J. F. Coulcher, G. D. Edgecombe, M. J. Telford, Molecular developmental evidence for a subcoxal origin of pleurites in insects and identity of the subcoxa in the gnathal appendages. *Scientific reports*. **5**, 1–8 (2015).

17. D. M. Linz, Y. Tomoyasu, Dual evolutionary origin of insect wings supported by an investigation of the abdominal wing serial homologs in Tribolium. **115**, E658–E667 (2018).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0221.v1



 doi:10.20944/preprints202102.0221.v1

Fig. 1. Model of how to align all arthropod legs from Bruce 2021 (*4*). A. Schematic of which structures require each gene in chelicerates, crustaceans, insects provides a model for how to align crustacean and insect legs. Based on the function of *exd*, *hth*, *Dll*, *Sp6-9*, and *dac*, the six distal leg segments (leg segment 1 through leg segment 6) of chelicerates, crustaceans, and insects correspond with each other in a one-to-one fashion. The alignment of the two proximal leg segments is based on expression of *pnr* and *ara* in chelicerates, crustaceans, and insects, and also function of wing/exite genes in insects and crustaceans B. Morphology and proposed homologies of arthropod leg segments. Colors and patterns indicate proposed homologies. Exites (checker pattern); endites (stripe pattern). Drawings in B modified from Snodgrass 1952.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0221.v1

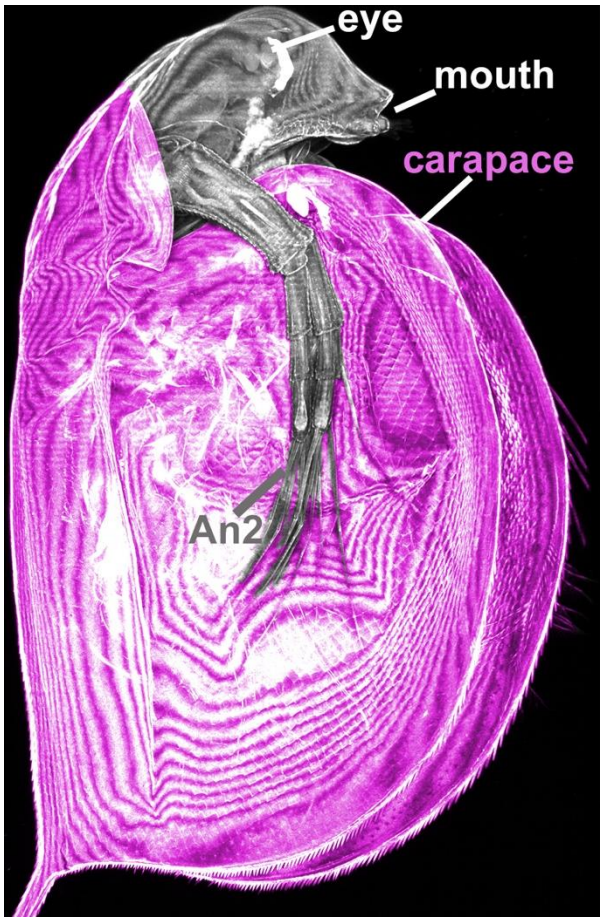


Fig. 2. *Daphnia magna* carapace (pink) is a bivalved "cape" of exoskeleton that surrounds the animal Shiga 2017 (*11*) showed that the *Daphnia* carapace emerges from the dorsal region of the head (maxilla 2), and, like insect wings, is composed of a bilayered sheet of ectodermal cells. Antenna 2 (An2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0221.v1

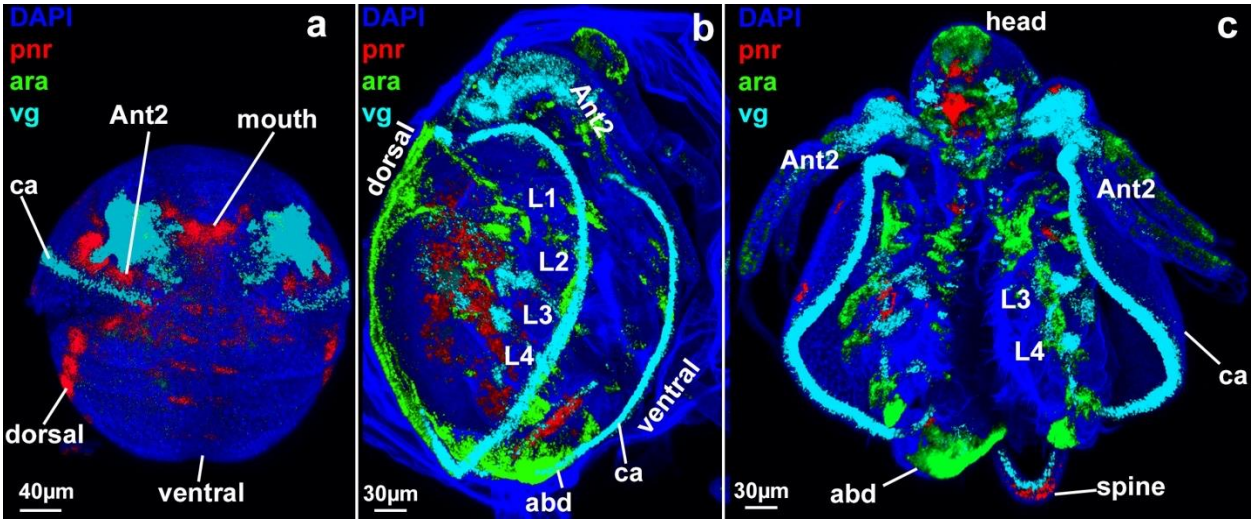


Fig. 3. Expression of *pannier*, *aurucan*, and *vestigial* in *Daphnia* embryos. *pnr* (red), *ara* (green), *vg* (light blue), DAPI (dark blue). a, early embryo. *pnr* is expressed dorsally, *ara* is expressed in rows of dots within the *vg* domain adjacent to the mouth, *vg* is expressed in the presumptive carapace (ca) and head. b - c, late embryo. b, ventral-lateral view, c ventral view. *pnr* expression is dorsal but in b is visible through the legs. *ara* is expressed in three domains: in the distal tips of the legs, in a dorsal region near *pnr*, and at the base of the free leg segments. *vg* is expressed in the perimeter of the carapace, in the mesoderm of antenna 2 (Ant2), and in legs 3 and 4 (L3 and L4). Spine of carapace (sp), abdomen (abd).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0221.v1

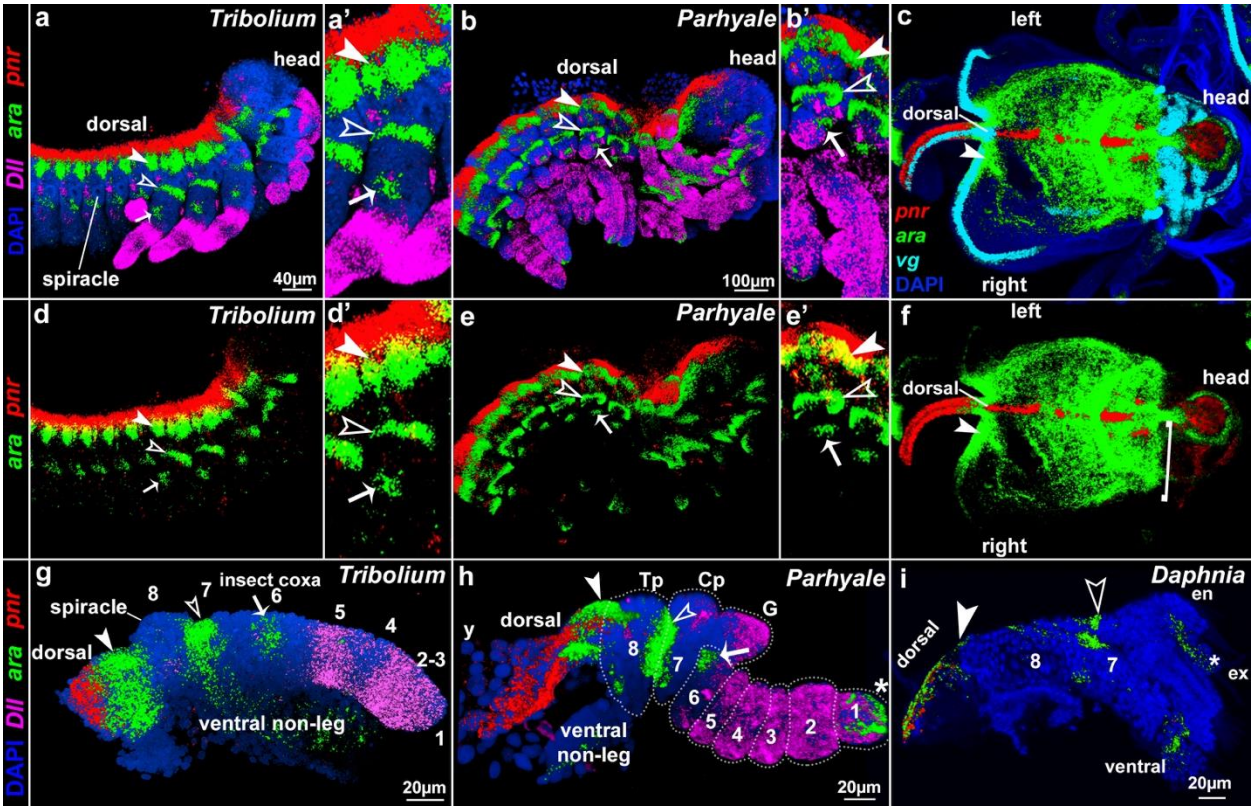


Fig. 4. Expression of *pnr* and *ara* elucidates the proximal leg segments. (a, a', d, d') dissected right half of *Tribolium* embryo. a' and d' are zoomed in images of a and d respectively. (b, b', e, e'), dissected right half of *Parhyale* embryo. b' and e' are zoomed in images of b and e respectively. (c, f) dorsal view of *Daphnia* embryo. (g) dissected leg of *Tribolium* embryo. (h) dissected T3 leg of *Parhyale* embryo. Large cells dorsal to *pnr* expression in b and f are extra-embryonic yolk (y) that exist prior to dorsal closure. (i) dissected leg of *Daphnia* embryo. Exopod (ex) and endopod (en) are labelled. In all three arthropods, the two *ara* (green) armband domains bracket a region proximal to leg segment 7. In all three arthropods, *pnr* (red) marks the most dorsal domain and is adjacent and partially overlapping the dorsal *ara* domain. In g, h, i, *ara* is expressed in the tips of the legs (*) and in a smattering of ventral non-leg cells. *Tribolium* larvae have a fused tibia and tarsus, the tibiotarsus, here labelled 2-3 . In both *Tribolium* (g) and *Parhyale* (h) legs, a muscle expressing *pnr* and *ara* that extends the length of leg segments 7 and 8 was masked to clearly show the ectodermal domains. Tp, tergal plate. Cp, coxal plate. G, gill. Images in a, a', b, b', d, d', e, e', g, and h from Bruce 2020 (*2*).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0221.v1

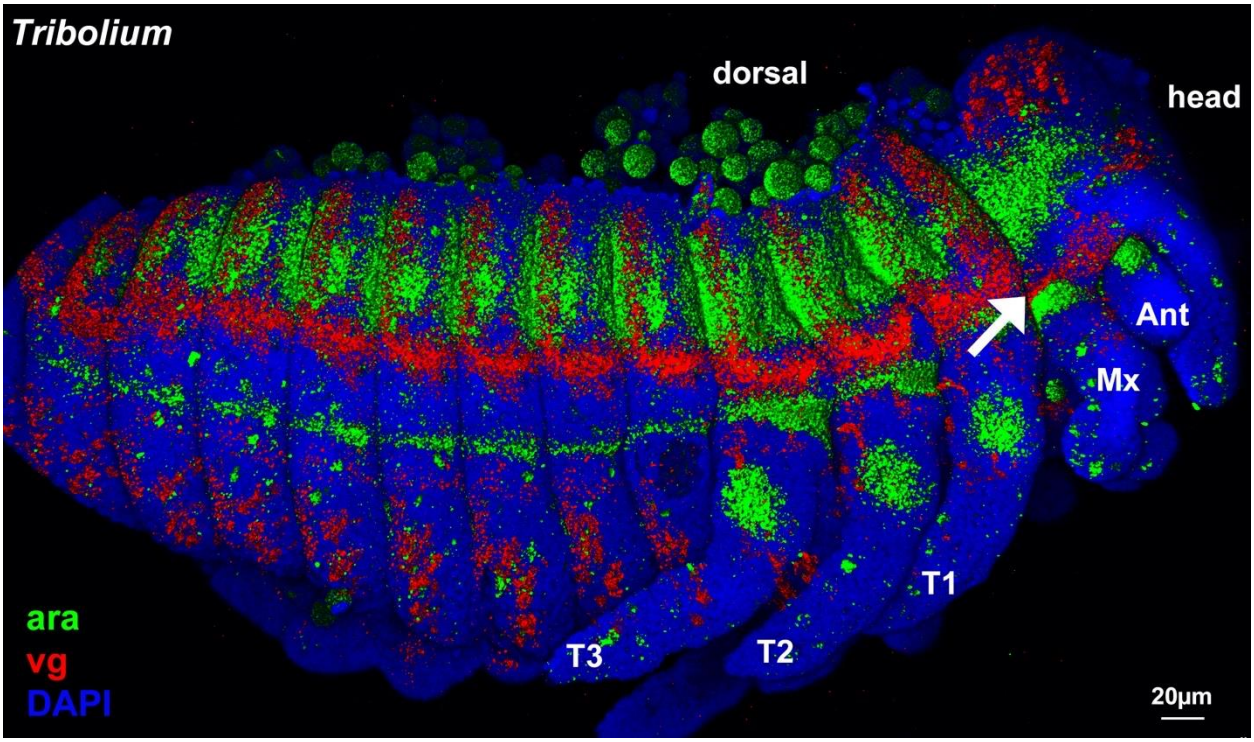


Fig. 5. Expression of *vg* in *Tribolium*. *vg* is expressed at the same register in all visible body segments. In all visible body segments, *vg* is bracketed dorsally and ventrally by *ara* expression domains. In thoracic legs T2 and T3, these *vg* expression domains pattern wings (the exites of the 8th leg segment). In the abdominal segments, *vg* patterns the gin traps, which are serially homologous to the wings(*17*), and therefore also exites. In the head, as in the other body segments, *vg* is expressed in the same register and is bracketed by ara. Here, *vg* is expressed in the dorsal/proximal maxilla (Mx, arrow) and antenna (Ant). The *vg* domain in the dorsal/proximal antenna may pattern a shelf-like structure called the gena, which forms the characteristic indent in the *Tribolium* eye. However, the *vg* domain in the dorsal/proximal maxilla does not pattern any obvious outgrowth.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 February 2021 doi:10.20944/preprints202102.0221.v1

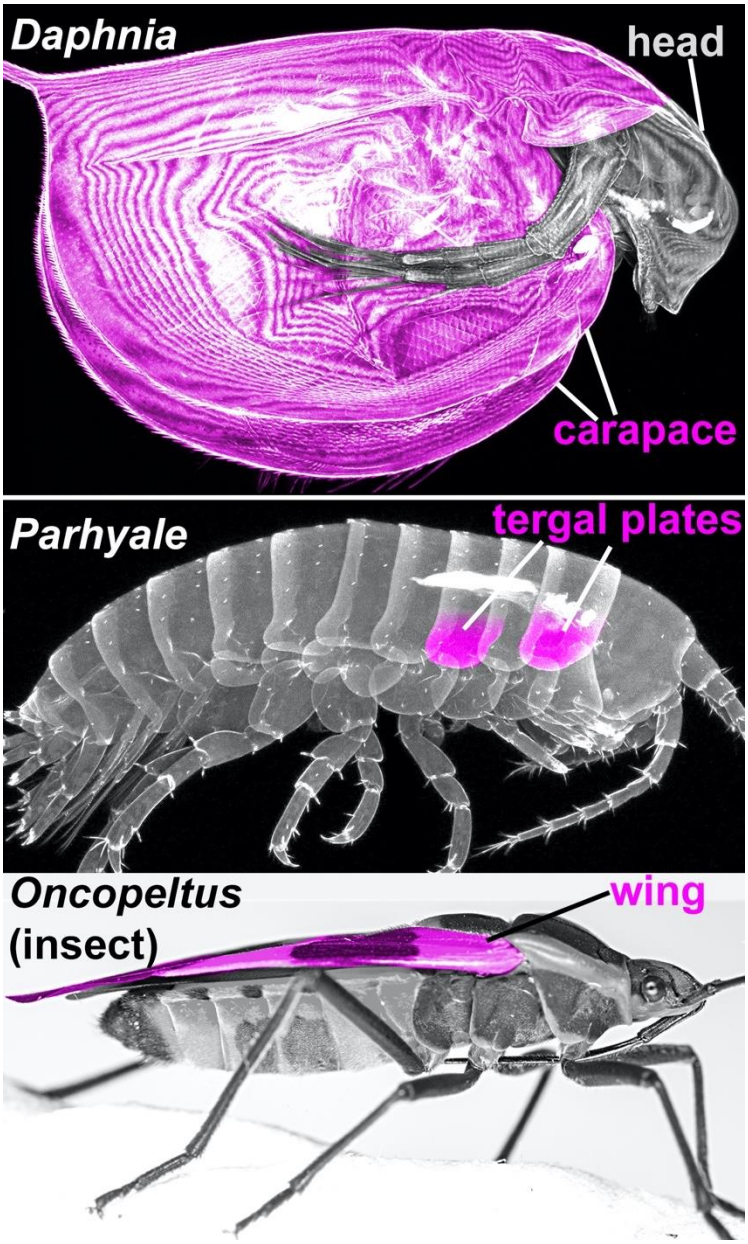


Fig. 6. *Daphnia* carapace appears to be the exite of the incorporated 8th leg segment, homologous to the *Parhyale* tergal plate and insect wing. In *Parhyale*, only two examples of tergal plates are highlighted.