universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„How to build a bivalve: Lessons from gene expression and single-cell RNA sequencing analyses“

verfasst von / submitted by

David Andrés Salamanca Díaz

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 794 685 437
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Biologie
Betreut von / Supervisor:	Univ.-Prof. DDr. Andreas Wanninger

Acknowledgements

I want to first thank my supervisor Andreas Wanninger, for taking me as his student, conceptualizing the project, his supervision, scientific input, and support along my PhD. He enabled me during this time to receive valuable feedback and exchange retroactively my ideas with many peers to make great collaborations with incredible researchers and improve my thesis.

I thank Tim Wollesen for being my personal researcher's role model, his indirect supervision, guidance in the lab and countless scientific discussions. To Sonia Rodriguez for being a patient tutor in the lab and an unconditional friend.

To all current and former members of the Integrative Zoology department, especially Elisabeth Zieger, Andrew Calcino, Stephan Schulreich, Thomas Schwaha, Julian Bibemair, Sebastian Decker, Attila Sachslehner, André de Oliveira, Emmanuel Redl, Lisa Maria Schmidt, Andy Sombke, Iulia Barutia for their help, scientific discussions, advice on lab techniques and the sporadic beers. Thank you all for being such an amazing lab and office partners.

I thank all members of the Technau and Genikhovich groups, for their scientific input and social evenings. To all participants of the 2019 embryology course at the Marine Biological Laboratory in Woods Hole. To all members of the Simakov lab, especially Hannah Schmidbaur and Elena Ritschard, the most supportive collaborators, colleagues, and wonderful friends. These years would not have been simply the same without any of you.

To my girlfriend Marlene Berl, for being such an incredible support, and having infinite patient with me in the last stages of my PhD and during pandemic times. I want to thank my friends Juan Pablo Novoa, Carolina Urbina, Manuel Fonseca, Yiselle Cano and Luisa Castellanos. To my family Claudia Díaz, John Salamanca, Angela Salamanca,

Juan Pablo Salamanca and Sofía Salamanca for their love and support. I would not be where I am right now without you.

My research was supported by the Austrian Science Fund (FWF) grant P29455-B29 and the additional support of the completion grant from the Vienna Doctoral School of Ecology and Evolution (VDSEE).

Table of Contents

Abstract

Mollusks are known for their highly diverse repertoire of body plans especially in the Conchifera, one of the two major lineages that comprises those taxa that originated from a uni-shelled ancestor (Monoplacophora, Gastropoda, Cephalopoda, Scaphopoda, Bivalvia). For most clades, and bivalves in particular, data concerning the molecular basis essential to this morphological variation is scarce. In this thesis I initially investigated Hox expression during development of the quagga mussel, *Dreissena rostriformis*, to elucidate to which degree they might contribute to specific phenotypic traits as in other conchiferans. Hox genes are key developmental regulators that are involved in establishing morphological features during animal ontogeny. Hox expression in *Dreissena* is first detected in the gastrula with widely overlapping expression domains of most genes, shifting towards more compact, largely mesodermal domains in the trochophore larva. The non-collinear mode of Hox expression in *Dreissena* might be a result of the low degree of body plan regionalization along the bivalve anterior-posterior axis as exemplified by the lack of key morphological traits such as a distinct head, cephalic tentacles, radula apparatus, and a simplified central nervous system. Secondly, the transcriptional complexity of a larval stage from *Dreissena* were investigated. For this, we profiled single cell transcriptomes of distinct cell populations from the trochophore larva. To annotate these cell populations, gene sets underlying the clustering were examined and compared to expression data of their orthologs in other lophotrochozoans. Genes expected to be specific to certain tissues, such as prototrochal cells, muscle cells, among others, corroborate that the recovered cell clusters reflect the larval body composition of our animal. In the third chapter we compared and analyzed genes expressed in the shell field. Our comparative gene

analyses in this cluster indicate that there is a significant emergence of novel genes shaping the unique transcriptomic signatures of shell field cells from *Dreissena*. Alternating temporal expression dynamics of these genes suggests a possible independent evolutionary origin (and thus argue against ontogenetic homology) of the embryonic shell (protoconch I) and the larval shell (protoconch II). Altogether, this thesis provides a molecular atlas of developmental processes and tempo-spatial gene expression profiles underlying bivalve development that significantly expands our knowledge regarding the ontogenetic factors that drive molluscan morphological diversity.

Zusammenfassung

Mollusken sind bekannt für ihre grosse morphologische Vielfalt, insbesondere innerhalb der Conchifera, jene der beiden Hauptlinien innerhalb des Tierstammes, welche von einem einschaligen Vorfahren abstammt und die Klassen Monoplacophora, Gastropoda, Cephalopoda, Scaphopoda und Bivalvia umfasst. Für die meisten dieser Klassen, insbesondere der Bivalvia, sind molekuare Daten zur Entwicklung wesentlicher morphologischer Strukturen nur ansatzweise vorhanden. In dieser Dissertation untersuchte ich zunächst die Hox-Expression während der Entwicklung der Quagga-Muschel, *Dreissena rostriformis*, um herauszufinden, inwieweit sie zur Ontogenie bestimmter phänotypischer Merkmale beitragen. Hox-Gene sind wichtige Entwicklungsregulatoren, die eine Vielzahl an Aufgaben während der Entwicklung erfüllen. In *Dreissena* werden die Hox Gene zuerst im Gastrulastadium in überlappenden Domänen exprimiert, die sich in Richtung kompakterer, größtenteils mesodermaler Domänen in der Trochophora-Larve verschieben. Der nicht-kollineare Modus der Hox-Expression bei *Dreissena* könnte auf den geringen Grad der Regionalisierung des Körpers entlang der anterior-posterioren Achse der Muscheln zurückzuführen sein, was durch das Fehlen wichtiger morphologischer Merkmale wie einem ausgeprägten Kopf, cephaler Tentakel, Radula-Apparat und ein vereinfachtes Nervensystems charakterisiert ist. In einer zweiten Studie wurden Einzelzell-Transkriptome im Trochophora-Stadium mittels single cell RNA Sequenzierung untersucht. Dadurch konnten distinkte Zellpopulationen basierend auf ihrem Genexpressionsprofil identifiziert werden. Um diese Zellpopulationen zu annotieren, wurden die der jeweiligen Cluster zugrunde liegenden Expressionsprofile untersucht und diese mit der Expression bekannter Genorthologe innerhalb der Lophotrochozoen verglichen. Gene, von denen erwartet wird, dass sie spezifisch für die Ausbildung bestimmter Gewebe bzw. Zelltypen verantwortlich sind, wie z. B. Prototrochzellen, Muskelzellen oder Nervenzellen, konnten

hierbei auch im Trochophorastadium von *Dreissena* identifiziert werden. In der dritten Studie wurden jene Gene näher untersucht, die im Schalenfeld der *Dreissena*-Trochophora exprimiert werden. Hierbei wurde eine hohe Anzahl an Genen identifiziert, für die keine entsprechenden Orthologe innerhalb der Metazoa gefunden werden konnten. Insgesamt liefert diese Arbeit einen ersten molekularen Atlas der Entwicklungsprozesse, die der Entwicklung eines Vertreters der Bivalvia zugrunde liegt, und schafft damit eine wesentliche Basis für weitere vergleichende Studien zur Evolution der morphologischen Diversität innerhalb der Mollusca.

1. General Introduction

1.1. Who are the Mollusks?

Among all extant clades of animals, Mollusca is without doubt one of the most diverse phyla. It constitutes the second most specious phylum after Arthropoda with around 85000 recognized living species (Rosenberg, 2014), without counting the fossil record that estimates between 60000 and 100000 more species (Taylor & Lewis, 2007). Many of these species play an important role to humans. In recent years mollusks have been gaining commercial, scientific and cultural value in the fields of gastronomy (e.g. clams, mussels, snails, squids, octopuses), jewelry (e.g. shell currency and pearls), invasive species ecology (e.g. garden slugs, quagga and zebra mussels), and biomedical research (e.g. analgesic development based on cone snails toxin and studies related to schistosomiasis) (reviewed in Wanninger and Wollesen, 2019).

According to recent phylogenomic studies, the phylum Mollusca is composed of two sublineages, Aculifera and Conchifera. The Aculifera includes the vermiform, spicule-bearing Solenogastres (Neomeniomorpha) and Caudofoveata (Chaetodermomorpha) as well as the dorso-ventrally flattened Polyplacophora with eight shell plates, while the primarily single-shelled Conchifera contains the Monoplacophora, Cephalopoda, Scaphopoda, Gastropoda, and Bivalvia (Kocot et al., 2011; Smith et al., 2011; Wanninger & Wollesen, 2015). Most class-level lineages emerged in the Cambrian around 500 million years ago (mya) from a common ancestor (Vinther, 2015; Parkhaev, 2017).

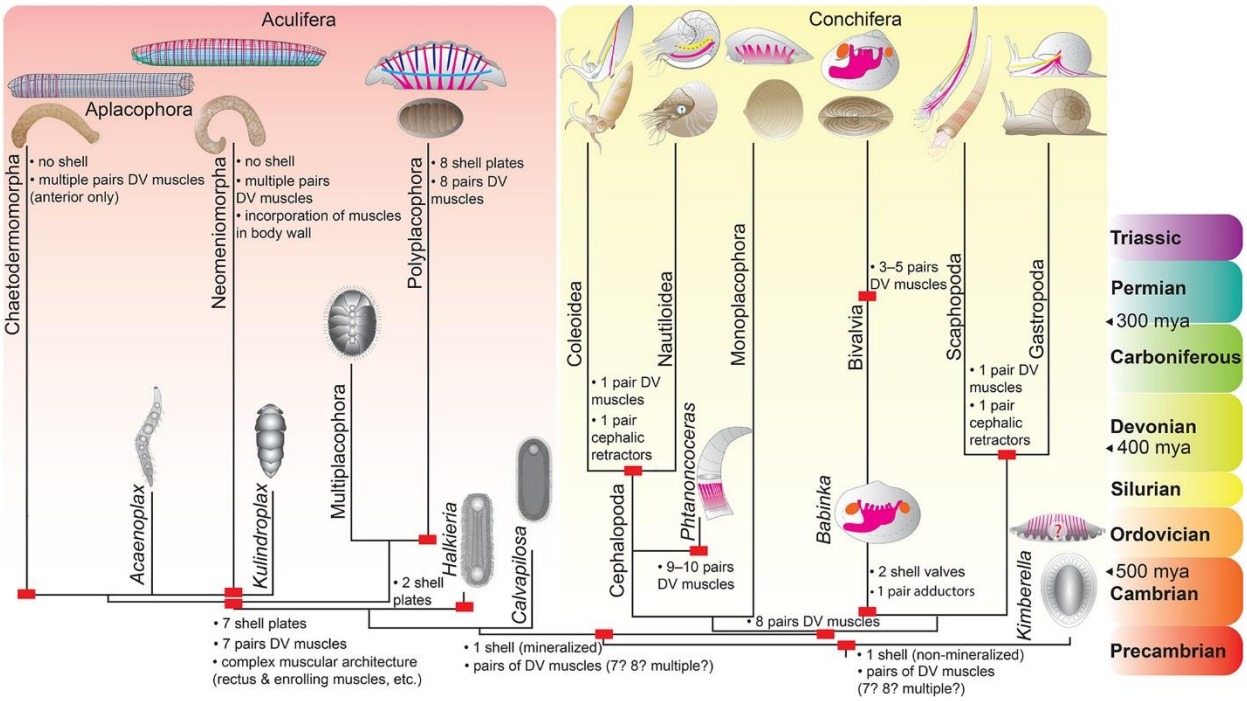


Figure 1. Molluscan intraphyletic relationships and hypothesized evolutionary pathways of major exoskeletal and muscular subsets modified from Wanniger & Wollesen, 2019. Only changes in character states are indicated (red rectangles). Phylogeny based on Smith et al. (2011) and Vinther et al. (2017). Ventro-lateral muscle (green), enrolling muscle (light blue), rectus muscle (red), ring muscles (dark blue), bivalvian adductor muscles (orange).

The dramatic variations in morphological sizes and shapes have allowed mollusks to conquer a large spectrum of habitats and evolve a variety of life history strategies. They may swim in the water column actively or passively, can have a sessile or motile lifestyle, have predatory, foraging or filtering behaviors, and possess different life cycle stages (e.g., trochophore, veliger, glochidium larval types). Due to this variety in morphology and lifestyles, Mollusca has become a group of interest for comparative evolutionary studies on the macroscopic, microscopic, and molecular level.

1.2. Integrative tools for understanding mollusks

The field of evolutionary developmental biology (Evo-Devo) is a discipline with the goal of gaining evolutionary insights from developmental processes involved in morphological diversity (Müller, 2007; Carroll, 2008). Accordingly, mollusks have recently entered the limelight of studies aiming at understanding the molecular mechanisms underlying their phenotypic variation. Recent gene expression profiling on a variety of molluscan taxa by *in situ* hybridization has provided detailed insights into differences and commonalities on the ontogeny of tissues and organ systems (e.g., Hinman et al., 2003; Lee et al., 2003; Wollesen et al., 2015a; 2015b; Redl et al., 2016; Wang et al., 2017; Wollesen et al., 2018; this thesis). However, this approach is obstructed by the fact that only a fraction of genes is known to be active in certain regions and developmental stages of the respective target species. At the same time, more traditional studies have focused on the ontogeny of different tissues by tracing the mitotic history of a cell and generating so-called fate maps, resulting in the reconstruction of so-called cell lineages (Dictus & Damen, 1997; Farrell et al., 2018; Hejnol et al., 2007; Henry et al., 2004; Lillie, 1895; Luetjens & Dorresteijn, 1995; Lyons et al., 2017; Meisenheimer, 1900; Render, 1997). However, some lineage diagrams do not assign fates to all terminal ends of the cell lineage tree, hindering broad comparisons across taxa (Luetjens & Dorresteijn, 1995; Meisenheimer, 1900).

Next generation sequencing has allowed for *in silico* methods following computational pipelines attempting to classify tissues and test for gene families based on expression profiles using RNA-sequencing (Kocot et al., 2011; Zhang et al., 2012; Paps et al., 2015; De Oliveira et al., 2016; Takeuchi et al., 2016; de Oliveira et al., 2019; Varney et al., 2020). Nevertheless, these bioinformatic methods present limitations since a tissue can be composed of many different cell types. Thus, while this approach can handle many

genes, by itself does not provide information on spatial resolution of gene transcript expression in the body. Therefore, this method alone lacks the necessary resolution to assign cell type domains during development.

Single-cell RNA sequencing (scRNA-seq) has taken a step forward by revealing comprehensive lists of genes that are expressed during ontogeny, potentially in any given cell at any given point in the life cycle of a specimen. It provides methods for sequencing live, dissociated cells followed by in silico methods using computational pipelines which allow for characterization of transcriptomic profiles related to cell states (Achim et al., 2015; Satija et al., 2015; Svensson et al., 2018; Tang et al., 2010; Trapnell, 2015; Trapnell et al., 2014; Vergara et al., 2017; Zhong et al., 2018). Accordingly, recent studies using scRNA-seq pipelines have uncovered the presence of previously uncharacterized cell types and an unexpected transcriptomic heterogeneity amongst cell populations and have revealed numerous genes with hitherto unknown function (Achim et al., 2018; Briggs et al., 2018; Duruz et al., 2020; Farrell et al., 2018; Fincher et al., 2018; Foster et al., 2019; Karaiskos et al., 2017; Plass et al., 2018; Sebé-Pedrós et al., 2018; Siebert et al., 2019; Sur & Meyer, 2021; Wagner et al., 2018; Wendt et al., 2020). Altogether, these studies represent a framework for comparative analyses of genes underlying the formation of key molluscan features such as the shell field, nervous system, foot, sensory structures, and others (Wanninger & Wollesen, 2019). Nevertheless, many of these studies so far lack the integrative use of methods to make a comparative approach within and outside of Mollusca. While all these tools have provided new perspectives on developmental gene expression, there is a substantial lack of data to make broad evolutionary and developmental parallels.

1.3. The invasive quagga mussel

Dreissena rostriformis is a member of the dreissenid bivalves. It is a freshwater bivalve that attaches to hard substrates using byssus threads and can form high population densities (A. Karatayev et al., 2002). Being native to the Dniepr drainage of the Ukrainian Black Sea region (Nalepa & Schloesser, 2019), it has managed to rapidly spread across European and North American freshwater bodies (Aldridge et al., 2014; Heiler et al., 2013; A. Y. Karatayev et al., 2014; Mills et al., 1993; Nalepa & Schloesser, 2019; Wakida-Kusunoki et al., 2015). Together with its closest relative, the zebra mussel *D. polymorpha*, it may form dense populations which consume large amounts of phytoplankton with often damaging impact on native limnic populations (A. Y. Karatayev et al., 2014; McNickle et al., 2006; Nalepa & Schloesser, 2019; Raikow, 2004; Strayer et al., 2004). This makes the quagga and zebra mussel one of the most widespread and effective invasive aquatic mollusks (A. Y. Karatayev et al., 2007).

Figure 2. Global distribution (up to 2019) of the quagga mussel (*D. rostriformis*) highlighting native (blue) and colonized (red) habitats, modified from Calcino et al., 2019.

Their free-swimming larval stages enhance the invasive capabilities and allows rapid spread into the water body they are introduced (e.g., trochophore and veliger states) (A. Karatayev et al., 2002, 2003, 2014). This, together with their conserved developmental mode (spiral cleavage, presence of a trochophore- and veliger-type larva) make the

quagga mussel an ideal candidate for comparative studies across Mollusca and Lophotrochozoa. Efforts to understand and manipulate the rate of colonization and dispersal for these species have mainly focused on the adult stage when the population is already established, while data on the dispersive larval stages, the trochophore and veliger, are still absent (McCartney et al., 2019). Hence, molecular pathways causing the morphological diversification across the life cycle of this invasive organism remain largely unresolved.

1.4. Thesis aims

Evolutionary developmental studies in Mollusca have focused mostly either on *in silico* or *in situ* research alone, leaving aside the possibility of combining both approaches and include *in vivo* studies to get a broader perspective into molluscan morphological variation. This dissertation combines gene expression analyses of key developmental regulators as well as cell type specification by single cell RNA transcriptomics. Specifically, the following issues were addressed:

i. Identification and characterization of Hox/ParaHox expression throughout larval development of a bivalve (Manuscript 1).

ii. Characterization of all cell types present in the *D. rostriformis* trochophore larva by gene expression profiling using single cell RNA-seq (Manuscript 2).

iii. Characterization of genes expressed in the shell field of the trochophore larva of *D. rostriformis* using single cell RNA-seq (Manuscript 3).

1.5. References Introduction

Achim, K., Eling, N., Vergara, H. M., Bertucci, P. Y., Musser, J., Vopalensky, P., Brunet, T., Collier, P., Benes, V., Marioni, J. C., & Arendt, D. (2018). Whole-Body Single-Cell Sequencing Reveals Transcriptional Domains in the Annelid Larval Body. *Molecular Biology and Evolution*, *35*(5), 1047–1062. https://doi.org/10.1093/molbev/msx336

Achim, K., Pettit, J.-B., Saraiva, L. R., Gavriouchkina, D., Larsson, T., Arendt, D., & Marioni, J. C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nature Biotechnology*, *33*(5), 503–509.

Aldridge, D. C., Ho, S., & Froufe, E. (2014). The Ponto-Caspian quagga mussel, Dreissena rostriformis bugensis (Andrusov, 1897), invades Great Britain. *Aquatic Invasions*, *9*(4), 529–535.

Briggs, J. A., Weinreb, C., Wagner, D. E., Megason, S., Peshkin, L., Kirschner, M. W., & Klein, A. M. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science*, *360*(6392), eaar5780. https://doi.org/10.1126/science.aar5780

Carroll, S. B. (2008). Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell*, *134*(1), 25–36. https://doi.org/10.1016/j.cell.2008.06.030

De Oliveira, A. L., Wollesen, T., Kristof, A., Scherholz, M., Redl, E., Todt, C., Bleidorn, C., & Wanninger, A. (2016). Comparative transcriptomics enlarges the toolkit of known developmental genes in mollusks. *BMC Genomics*, *17*(1), 905. https://doi.org/10.1186/s12864-016-3080-9

de Oliveira, André L, Calcino, A. D., & Wanninger, A. (2019). Extensive conservation of

the proneuropeptide and peptide prohormone complement in mollusks. *Scientific Reports*, *9*. https://doi.org/10.1038/s41598-019-40949-0

Dictus, W. J. A. G., & Damen, P. (1997). Cell-lineage and clonal-contribution map of the trochophore larva of Patella vulgata (Mollusca)1Both authors contributed equally to this work.1. *Mechanisms of Development*, *62*(2), 213–226. https://doi.org/https://doi.org/10.1016/S0925-4773(97)00666-7

Duruz, J., Kaltenrieder, C., Ladurner, P., Bruggmann, R., Martìnez, P., & Sprecher, S. G. (2020). Acoel single-cell transcriptomics: cell-type analysis of a deep branching bilaterian. *BioRxiv*, 2020.07.10.196782. https://doi.org/10.1101/2020.07.10.196782

Farrell, J. A., Wang, Y., Riesenfeld, S. J., Shekhar, K., Regev, A., & Schier, A. F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science*, *360*(6392), eaar3131. https://doi.org/10.1126/science.aar3131

Fincher, C. T., Wurtzel, O., de Hoog, T., Kravarik, K. M., & Reddien, P. W. (2018). Cell type transcriptome atlas for the planarian Schmidtea mediterranea. *Science*, *360*(6391), eaaq1736. https://doi.org/10.1126/science.aaq1736

Foster, S., Oulhen, N., & Wessel, G. (2020). A single cell RNA sequencing resource for early sea urchin development. *Development*, *147*(17), dev191528. https://doi.org/10.1242/dev.191528

Foster, S., Teo, Y. V., Neretti, N., Oulhen, N., & Wessel, G. M. (2019). Single cell RNA-seq in the sea urchin embryo show marked cell-type specificity in the Delta/Notch pathway. *Molecular Reproduction and Development*, *86*(8), 931–934. https://doi.org/10.1002/mrd.23181

Fritsch, M., Wollesen, T., de Oliveira, A. L., & Wanninger, A. (2015). Unexpected co-linearity of Hox gene expression in an aculiferan mollusk. *BMC Evolutionary*

Biology, *15*(1), 151. https://doi.org/10.1186/s12862-015-0414-1

García-Castro, H., Kenny, N. J., Iglesias, M., Álvarez-Campos, P., Mason, V., Elek, A., Schönauer, A., Sleight, V. A., Neiro, J., Aboobaker, A., Permanyer, J., Irimia, M., Sebé-Pedrós, A., & Solana, J. (2021). ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics. *Genome Biology*, *22*(1), 89. https://doi.org/10.1186/s13059-021-02302-5

Heiler, K. C. M., Bij de Vaate, A., Ekschmitt, K., von Oheimb, P. V, Albrecht, C., & Wilke, T. (2013). Reconstruction of the early invasion history of the quagga mussel (Dreissena rostriformis bugensis) in Western Europe. *Aquatic Invasions*, *8*(1), 53–57.

Hejnol, A., Martindale, M. Q., & Henry, J. Q. (2007). High-resolution fate map of the snail Crepidula fornicata: The origins of ciliary bands, nervous system, and muscular elements. *Developmental Biology*, *305*(1), 63–76. https://doi.org/https://doi.org/10.1016/j.ydbio.2007.01.044

Henry, J. Q., Okusu, A., & Martindale, M. Q. (2004). The cell lineage of the polyplacophoran, Chaetopleura apiculata: variation in the spiralian program and implications for molluscan evolution. *Developmental Biology*, *272*(1), 145–160. https://doi.org/https://doi.org/10.1016/j.ydbio.2004.04.027

Hinman, V. F., O'Brien, E. K., Richards, G. S., & Degnan, B. M. (2003). Expression of anterior Hox genes during larval development of the gastropod Haliotis asinina. *Evolution & Development*, *5*(5), 508–521.

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., & Zinzen, R. P. (2017). The Drosophila embryo at single-cell transcriptome resolution. *Science*, *358*(6360), 194 LP – 199. https://doi.org/10.1126/science.aan3235

Karatayev, A., Burlakova, L., & Padilla, D. (2002). *Impacts of Zebra Mussels on Aquatic Communities and their Role as Ecosystem Engineers* (pp. 433–446). https://doi.org/10.1007/978-94-015-9956-6_43

Karatayev, A. Y., Burlakova, L. E., Padilla, D. K., & Johnson, L. E. (2003). Patterns of spread of the zebra mussel (Dreissena polymorpha (Pallas)): the continuing invasion of Belarussian lakes. *Biological Invasions*, *5*(3), 213–221.

Karatayev, A. Y., Burlakova, L. E., Pennuto, C., Ciborowski, J., Karatayev, V. A., Juette, P., & Clapsadl, M. (2014). Twenty five years of changes in Dreissena spp. populations in Lake Erie. *Journal of Great Lakes Research*, *40*(3), 550–559.

Karatayev, A. Y., Padilla, D. K., Minchin, D., Boltovskoy, D., & Burlakova, L. E. (2007). Changes in Global Economies and Trade: the Potential Spread of Exotic Freshwater Bivalves. *Biological Invasions*, *9*(2), 161–180. https://doi.org/10.1007/s10530-006-9013-9

Kocot, K. M., Cannon, J. T., Todt, C., Citarella, M. R., Kohn, A. B., Meyer, A., Santos, S. R., Schander, C., Moroz, L. L., Lieb, B., & Halanych, K. M. (2011). Phylogenomics reveals deep molluscan relationships. *Nature*, *477*(7365), 452–456. https://doi.org/10.1038/nature10382

Lee, P. N., Callaerts, P., De Couet, H. G., & Martindale, M. Q. (2003). Cephalopod Hox genes and the origin of morphological novelties. *Nature*, *424*(6952), 1061–1065. https://doi.org/10.1038/nature01872

Lillie, F. R. (1895). The embryology of the unionidae. A study in cell-lineage. *Journal of Morphology*, *10*(1), 1–100. https://doi.org/10.1002/jmor.1050100102

Luetjens, C. M., & Dorresteijn, A. W. C. (1995). Multiple, alternative cleavage patterns precede uniform larval morphology during normal development of Dreissena polymorpha (Mollusca, Lamellibranchia). *Roux's Archives of Developmental*

Biology, *205*(3), 138–149. https://doi.org/10.1007/BF00357760

Lyons, D. C., Perry, K. J., & Henry, J. Q. (2017). Morphogenesis along the animal-vegetal axis: fates of primary quartet micromere daughters in the gastropod Crepidula fornicata. *BMC Evolutionary Biology*, *17*(1), 217. https://doi.org/10.1186/s12862-017-1057-1

McCartney, M. A., Auch, B., Kono, T., Mallez, S., Zhang, Y., Obille, A., Becker, A., Abrahante, J. E., Garbe, J., Badalamenti, J. P., Herman, A., Mangelson, H., Liachko, I., Sullivan, S., Sone, E. D., Koren, S., Silverstein, K. A. T., Beckman, K. B., & Gohl, D. M. (2019). The Genome of the Zebra Mussel, Dreissena polymorpha: A Resource for Invasive Species Research. *BioRxiv*, 696732. https://doi.org/10.1101/696732

McNickle, G. G., Rennie, M. D., & Sprules, W. G. (2006). Changes in benthic invertebrate communities of South Bay, Lake Huron following invasion by zebra mussels (Dreissena polymorpha), and potential effects on lake whitefish (Coregonus clupeaformis) diet and growth. *Journal of Great Lakes Research*, *32*(1), 180–193.

Meisenheimer, J. (1900). *Entwicklungsgeschichte von Dreissensia polymorpha Pall.*

Mills, E. L., Dermott, R. M., Roseman, E. F., Dustin, D., Mellina, E., Conn, D. B., & Spidle, A. P. (1993). Colonization, ecology, and population structure of the" quagga’’mussel (Bivalvia: Dreissenidae) in the lower Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, *50*(11), 2305–2314.

Müller, G. B. (2007). Evo–devo: extending the evolutionary synthesis. *Nature Reviews Genetics*, *8*(12), 943–949. https://doi.org/10.1038/nrg2219

Nalepa, T. F., & Schloesser, D. W. (2019). *Quagga and zebra mussels: biology, impacts, and control*. CRC Press.

Paps, J., Xu, F., Zhang, G., & Holland, P. W. H. (2015). Reinforcing the egg-timer: recruitment of novel lophotrochozoa homeobox genes to early and late development in the pacific oyster. *Genome Biology and Evolution*, *7*(3), 677–688.

Parkhaev, P. Y. (2017). Origin and the early evolution of the phylum Mollusca. *Paleontological Journal*, *51*(6), 663–686.

Plass, M., Solana, J., Wolf, F. A., Ayoub, S., Misios, A., Glažar, P., Obermayer, B., Theis, F. J., Kocks, C., & Rajewsky, N. (2018). Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science*, *360*(6391), eaaq1723. https://doi.org/10.1126/science.aaq1723

Raikow, D. F. (2004). Food web interactions between larval bluegill (Lepomis macrochirus) and exotic zebra mussels (Dreissena polymorpha). *Canadian Journal of Fisheries and Aquatic Sciences*, *61*(3), 497–504.

Redl, E., Scherholz, M., Wollesen, T., Todt, C., & Wanninger, A. (2016). Cell Proliferation Pattern and Twist Expression in an Aplacophoran Mollusk Argue Against Segmented Ancestry of Mollusca. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *326*(7), 422–436. https://doi.org/10.1002/jez.b.22714

Render, J. (1997). Cell Fate Maps in theIlyanassa obsoletaEmbryo beyond the Third Division. *Developmental Biology*, *189*(2), 301–310. https://doi.org/https://doi.org/10.1006/dbio.1997.8654

Rosenberg, G. (2014). A New Critical Estimate of Named Species-Level Diversity of the Recent Mollusca. *American Malacological Bulletin*, *32*(2), 308–322. https://doi.org/10.4003/006.032.0204

Salamanca-Díaz, D. A., Calcino, A. D., de Oliveira, A. L., & Wanninger, A. (2021). Non-collinear Hox gene expression in bivalves and the evolution of morphological

novelties in mollusks. *Scientific Reports*, *11*(1), 3575. https://doi.org/10.1038/s41598-021-82122-6

Samadi, L., & Steiner, G. (2009). Involvement of Hox genes in shell morphogenesis in the encapsulated development of a top shell gastropod (Gibbula varia L.). *Development Genes and Evolution*, *219*(9–10), 523–530. https://doi.org/10.1007/s00427-009-0308-6

Samadi, L., & Steiner, G. (2010a). Conservation of ParaHox genes' function in patterning of the digestive tract of the marine gastropod Gibbula varia. *BMC Developmental Biology*, *10*(1), 74. https://doi.org/10.1186/1471-213X-10-74

Samadi, L., & Steiner, G. (2010b). Expression of Hox genes during the larval development of the snail, Gibbula varia (L.)-further evidence of non-colinearity in molluscs. *Development Genes and Evolution*, *220*(5–6), 161–172. https://doi.org/10.1007/s00427-010-0338-0

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*, *33*(5), 495–502. https://doi.org/10.1038/nbt.3192

Sebé-Pedrós, A., Saudemont, B., Chomsky, E., Plessier, F., Mailhé, M.-P., Renno, J., Loe-Mie, Y., Lifshitz, A., Mukamel, Z., Schmutz, S., Novault, S., Steinmetz, P. R. H., Spitz, F., Tanay, A., & Marlow, H. (2018). Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell*, *173*(6), 1520-1534.e20. https://doi.org/10.1016/j.cell.2018.05.019

Siebert, S., Farrell, J. A., Cazet, J. F., Abeykoon, Y., Primack, A. S., Schnitzler, C. E., & Juliano, C. E. (2019). Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science*, *365*(6451), eaav9314. https://doi.org/10.1126/science.aav9314

Smith, S. a., Wilson, N. G., Goetz, F. E., Feehery, C., Andrade, S. C. S., Rouse, G. W., Giribet, G., & Dunn, C. W. (2011). Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature*, *480*(7377), 364–367. https://doi.org/10.1038/nature10526

Strayer, D. L., Hattala, K. A., & Kahnle, A. W. (2004). Effects of an invasive bivalve (Dreissena polymorpha) on fish in the Hudson River estuary. *Canadian Journal of Fisheries and Aquatic Sciences*, *61*(6), 924–941.

Sur, A., & Meyer, N. P. (2021). Resolving Transcriptional States and Predicting Lineages in the Annelid Capitella teleta Using Single-Cell RNAseq . In *Frontiers in Ecology and Evolution* (Vol. 8, p. 523). https://www.frontiersin.org/article/10.3389/fevo.2020.618007

Svensson, V., Vento-Tormo, R., & Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nature Protocols*, *13*(4), 599–604.

Takeuchi, T., Koyanagi, R., Gyoja, F., Kanda, M., Hisata, K., Fujie, M., Goto, H., Yamasaki, S., Nagai, K., Morino, Y., Miyamoto, H., Endo, K., Endo, H., Nagasawa, H., Kinoshita, S., Asakawa, S., Watabe, S., Satoh, N., Kawashima, T., … Weisblat, D. (2016). Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zoological Letters*, *2*(1), 3. https://doi.org/10.1186/s40851-016-0039-2

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., & Surani, M. A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell*, *6*(5), 468–478.

Taylor, P. D., & Lewis, D. N. (2007). *Fossil invertebrates*. Harvard University Press.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. *Genome Research*, *25*(10), 1491–1498.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., & Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, *32*(4), 381.

Varney, R. M., Speiser, D. I., McDougall, C., Degnan, B. M., & Kocot, K. M. (2020). The iron-responsive genome of the chiton Acanthopleura granulata. *BioRxiv*, 2020.05.19.102897. https://doi.org/10.1101/2020.05.19.102897

Vergara, H. M., Bertucci, P. Y., Hantz, P., Tosches, M. A., Achim, K., Vopalensky, P., & Arendt, D. (2017). Whole-organism cellular gene-expression atlas reveals conserved cell types in the ventral nerve cord of Platynereis dumerilii. *Proceedings of the National Academy of Sciences*, *114*(23), 5878–5885.

Vinther, J. (2015). The origins of molluscs. *Palaeontology*, *58*(1), 19–34. https://doi.org/https://doi.org/10.1111/pala.12140

Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G., & Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science*, *360*(6392), 981 LP – 987. https://doi.org/10.1126/science.aar4362

Wakida-Kusunoki, A. T., Wakida, F. T., & De Leon-Sandoval, J. M. (2015). First record of quagga mussel Dreissena rostriformis bugensis (Andrusov, 1897)(Bivalvia, Dreissenidae) from Mexico. *BioInvasions Rec*, *4*, 31–36.

Wang, S., Zhang, J., Jiao, W., Li, J., Xun, X., Sun, Y., Guo, X., Huan, P., Dong, B., Zhang, L., Hu, X., Sun, X., Wang, J., Zhao, C., Wang, Y., Wang, D., Huang, X., Wang, R., Lv, J., … Bao, Z. (2017). Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature Ecology & Evolution*, *1*(April), 0120. https://doi.org/10.1038/s41559-017-0120

Wanninger, A., & Wollesen, T. (2015). Evolutionary developmental biology of invertebrates 2: Lophotrochozoa (Spiralia). In A. Wanninger (Ed.), *Evolutionary Developmental Biology of Invertebrates 2: Lophotrochozoa (Spiralia)* (pp. 103–154). Springer-Verlag Wien. https://doi.org/10.1007/978-3-7091-1871-9

Wanninger, A., & Wollesen, T. (2019). The evolution of molluscs. *Biological Reviews*, *94*(1), 102–115. https://doi.org/10.1111/brv.12439

Wendt, G., Zhao, L., Chen, R., Liu, C., O'Donoghue, A. J., Caffrey, C. R., Reese, M. L., & Collins, J. J. (2020). A single-cell RNA-seq atlas of Schistosoma mansoni identifies a key regulator of blood feeding. *Science*, *369*(6511), 1644 LP – 1649. https://doi.org/10.1126/science.abb7709

Wollesen, T., Rodriguez Monje, S. V., de Oliveira, A. L., & Wanninger, A. (2018). Staggered Hox expression is more widespread among molluscs than previously appreciated. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1888). https://doi.org/10.1098/rspb.2018.1513

Wollesen, T., Rodríguez Monje, S. V., McDougall, C., Degnan, B. M., & Wanninger, A. (2015). The ParaHox gene Gsx patterns the apical organ and central nervous system but not the foregut in scaphopod and cephalopod mollusks. *EvoDevo*, *6*(1), 41. https://doi.org/10.1186/s13227-015-0037-z

Wollesen, T., Rodríguez Monje, S. V., Todt, C., Degnan, B. M., & Wanninger, A. (2015). Ancestral role of Pax2/5/8 in molluscan brain and multimodal sensory system development. *BMC Evolutionary Biology*, *15*, 231. https://doi.org/10.1186/s12862-015-0505-z

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., Xiong, Z., Que, H., Xie, Y., Holland, P. W. H., Paps, J., Zhu, Y., Wu, F., Chen, Y., Wang, J., … Wang, J. (2012). The oyster genome reveals stress adaptation and

complexity of shell formation. *Nature*, *490*(7418), 49–54. https://doi.org/10.1038/nature11413

Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L., & Pan, N. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature*, *555*(7697), 524–528.

2. Results

Data resulting from experiments in this thesis are summarized in the form of scientific papers. One of these has been published in *Scientific Reports*, the second in *Frontiers of Ecology and Evolution*, and the third manuscript is under review in *Frontiers in Cell and Developmental Biology*. These three manuscripts form the major chapters and thus the core of this thesis. In order to avoid redundancy, the major outcomes and implications of the respective studies are only briefly summarized at the start of each chapter.

2.1. Chapter 1. Non-collinear Hox gene expression in bivalves and the evolution of morphological novelties in mollusks (published in *Scientific Reports*)

Salamanca-Díaz, D.A., Calcino, A.D., de Oliveira, A.L., Wanninger, A. Sci Rep 11, 3575 (2021).

For this chapter, we analyzed the expression across developmental stages and gene organization conformation of the Hox/ParaHox clusters in *Dreissena rostriformis*. Hox and ParaHox positions in the genome assembly revealed that these genes are distributed over five scaffolds, while two of the three ParaHox genes (*Gsx, Xlox*) were found on a single scaffold. Additionally, several genes with no apparent relationship were found between members of the Hox cluster (between *Hox5* and *Lox5, Lox2* and *Post2, Xlox* and *Cdx*). This denotates insertions all along the gene set and provides evidence in support of disorganized (but not broken) Hox/ ParaHox clusters.

With the level of organization assessed, the next step was to analyze the corresponding expression domains of each gene in both whole clusters. Among the expression domains analyzed, there are few genes (*Hox1, Post2, Lox4*) which might be interpreted as “partially spatially staggered” in the gastrula stage. However, during the transition to the trochophore stage, all *D. rostriformis* Hox genes display highly dynamic spatial expression profiles with changes in expression domains.

scientific reports

Check for updates

OPEN

Non-collinear Hox gene expression in bivalves and the evolution of morphological novelties in mollusks

David A. Salamanca-Díaz[1], Andrew D. Calcino[1], André L. de Oliveira[2] & Andreas Wanninger[1]✉

Hox genes are key developmental regulators that are involved in establishing morphological features during animal ontogeny. They are commonly expressed along the anterior–posterior axis in a staggered, or collinear, fashion. In mollusks, the repertoire of body plans is widely diverse and current data suggest their involvement during development of landmark morphological traits in Conchifera, one of the two major lineages that comprises those taxa that originated from a uni-shelled ancestor (Monoplacophora, Gastropoda, Cephalopoda, Scaphopoda, Bivalvia). For most clades, and bivalves in particular, data on Hox gene expression throughout ontogeny are scarce. We thus investigated Hox expression during development of the quagga mussel, *Dreissena rostriformis*, to elucidate to which degree they might contribute to specific phenotypic traits as in other conchiferans. The Hox/ParaHox complement of Mollusca typically comprises 14 genes, 13 of which are present in bivalve genomes including *Dreissena*. We describe here expression of 9 Hox genes and the ParaHox gene *Xlox* during *Dreissena* development. Hox expression in *Dreissena* is first detected in the gastrula stage with widely overlapping expression domains of most genes. In the trochophore stage, Hox gene expression shifts towards more compact, largely mesodermal domains. Only few of these domains can be assigned to specific developing morphological structures such as *Hox1* in the shell field and *Xlox* in the hindgut. We did not find traces of spatial or temporal staggered expression of Hox genes in *Dreissena*. Our data support the notion that Hox gene expression has been coopted independently, and to varying degrees, into lineage-specific structures in the respective conchiferan clades. The non-collinear mode of Hox expression in *Dreissena* might be a result of the low degree of body plan regionalization along the bivalve anterior–posterior axis as exemplified by the lack of key morphological traits such as a distinct head, cephalic tentacles, radula apparatus, and a simplified central nervous system.

Mollusca constitutes a metazoan phylum with unique morphological diversity. It is composed of two clades, Aculifera and Conchifera, that diverged from one another in the early Cambrian. The Aculifera includes the vermiform, spicule-bearing Solenogastres (Neomeniomorpha) and Caudofoveata (Chaetodermomorpha) as well as the dorso-ventrally flattened Polyplacophora with eight shell plates, while the primarily single-shelled Conchifera contains the Monoplacophora, Scaphopoda, Gastropoda, Bivalvia, and Cephalopoda[1]. Hox genes are key developmental regulators that are involved in specifying morphological regions of the body plan during animal ontogeny[2–6]. One of the central roles of Hox gene products is to define body regionalization along the anterior–posterior axis[2,7,8]. The Hox genes are usually arranged in close proximity to each other on the genome, and this micro-syntenic block is also referred to as the Hox cluster. There is often a correlation between a given Hox gene's relative chromosomal position, its relative spatial expression domain along the animal's anterior–posterior axis (spatial collinearity), and/or its temporal activation (temporal collinearity)[9–11]. A three-fold collinearity encompassing positional arrangement on the genome as well as tempo-spatial expression is commonly believed to be ancestral to bilaterian animals[12]. However, the recent increase in high quality genome assemblies as well as expression analyses on previously neglected clades have shown that numerous deviations from this model are found in almost all major bilaterian lineages[13–19].

[1]Department of Evolutionary Biology, Unit for Integrative Zoology, University of Vienna, Althanstraße 14, 1090 Vienna, Austria. [2]Department of Functional and Evolutionary Ecology, Unit for Bio-Oceanography and Marine Biology, University of Vienna, Althantraße 14, 1090 Vienna, Austria. ✉email: andreas.wanninger@univie.ac.at

In Mollusca, 11 Hox genes have been identified[14,16,18,20–38]. While no Hox expression data are available for the two aplacophoran clades, Neomeniomorpha and Chaetodermomorpha, staggered spatial expression of 10 Hox genes along the anterior–posterior body axis has been found in the polyplacophoran *Acanthochitona*[30,31]. This is different from the situation in the conchiferans, where Hox genes are expressed during development of distinct morphological features such as ganglia, the foot, or the shell field[39]. In the gastropod *Gibbula*, *Lox5, Hox7, Lox4*, and *Lox2* are expressed in the larval swimming device, the prototroch, and in the larval episphere, with little indication of spatial staggering other than in the nervous system[14,20,26,27,40]. Scaphopod Hox transcripts are primarily present in the foot (*Hox2, Hox4, Hox5, Lox5, Lox4*, *Post1*), the mantle (*Hox1*), and in the central nervous system, with a near-to-staggered expression only in the mid-trochophore stage[37]. Cephalopods express some of these genes in the arm crown (*Hox1*, *Hox3*, *Hox4*, *Lox5*, *Lox4*, *Hox7*, *Post2*, *Post1*), the funnel tube (*Hox3*, *Lox5*, *Lox4*), and in various ganglia as well as in other clade-specific structures[22,29]. For bivalves, rudimentary expression data from *Hox1*, *Hox4*, *Lox5*, and *Post2* are available for the gastrula stage of a scallop only, wherein the authors interpreted their results as evidence of staggered expression[35]. In order to fill this gap in knowledge and to contribute to the question as to what degree Hox genes might contribute to bivalve-specific features, we here provide the first comprehensive dataset of tempo-spatial expression of Hox and ParaHox genes in the invasive freshwater mussel *Dreissena rostriformis*, a bivalve with an ancestral life cycle that involves indirect development via a trochophore and a veliger larva.

Materials and methods

Animal collection and cultures. Sexually mature individuals of *Dreissena rostriformis* were collected in the Danube river in Vienna, Austria (N 48° 14′ 45.812″, O 16° 23′ 38.145″). Collection took place between April and September 2017 and 2018, respectively. Adults were gathered from underneath stones and transferred to the laboratory. Here, the animals were washed, cleaned, and maintained in aquaria with filtered river water (FRW) at 19 °C.

Spawning of animals was induced by exposing 10–20 sexually mature specimens for 15 min to a 10^{-3} M solution of serotonin (Sigma-Aldrich, Darmstadt, Germany), followed by one wash and subsequent maintenance in FRW. After approximately 30 min, up to 50% of the treated specimens started to spawn. Oocytes were collected from the water column, inseminated, and cultured in 50 ml glass beakers at 23 °C. After fertilization, water was changed every half an hour for the first three hours and then every six hours to remove excess sperm and to avoid bacterial or fungal infection.

RNA extraction and fixation of developmental stages. Several hundred individuals of each, cleavage, trochophore, and various veliger stages were stored in RNALater (Lifetechnologies, Vienna, Austria) at – 20 °C. RNA was extracted with an RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and was stored at – 80 °C.

Embryos and larvae from three developmental stages (gastrula, trochophore, and veliger) were fixed in 4% paraformaldehyde (PFA) in 0.1 M phosphate buffered saline (PBS) for 1 h at room temperature (23 °C). Veliger larvae aged 24 h post fertilization (hpf) or older were relaxed prior to fixation by adding cocaine crystals (Sigma-Aldrich, Darmstadt, Germany) to the FRW. All larvae were subsequently washed in 0.1 M PBS, transferred stepwise to 100% methanol, and stored at -20 °C.

Transcriptome assembly and expression profiling. RNA-seq transcription data collected from developmental stages of interest were assembled previously[41]. Gene expression levels of the 17 libraries were quantified with Kallisto (transcripts per kilobase million, TPM)[42]. Heatmaps showing relative quantitative expression of genes were plotted in R software with the heatmap.2 function from the gplots R package[43] and columns were normalized by z-score (Fig. 1, Supplementary Table S1).

Sequence alignment and phylogenetic analysis. Amino acid sequences of Hox orthologs were retrieved from published NCBI sequences of other molluscan, lophotrochozoan, and bilaterian gene orthologs. *Dreissena* Hox candidates were found in the previously assembled transcriptome by blasting against confirmed ortholog sequences. Gene names and GenBank accession numbers used for phylogenetic reconstruction are available in Supplementary Table S1. *Dreissena* Hox and ParaHox ortholog candidates were confirmed by searching for characteristic motifs in each gene from the quagga mussel transcriptome as previously described[33,41]. Amino acid sequences of candidate *Dreissena* Hox genes and metazoan orthologs were aligned using MAFFT v7.123b[44]. The trimmed multiple sequence alignment used for the phylogenetic analysis was obtained with TrimAl v1.2 with the following parameters: -cons 20 -st 0.8 -gt 0.7—w 1[45]. Model selection with Prottest3 v3.4.2[46] through the Akaike Information Criterion (AIC) determined the Jones-Taylor-Thornton (JTT) model of amino-acid substitution as most appropriate for the alignment and thus was selected for the phylogenetic analysis. The maximum likelihood phylogenetic analysis was carried out with RAxML v8.2.X[47] with gamma-distributed rates and 1,000 non-parametric bootstrap replicates. Bayesian analysis was performed using MrBayes v3.2.6[48] with 25% of sampled trees discarded as burn-in, six rate categories for the gamma distribution, and 30,000,000 generations. The Bayesian phylogenetic run output was analyzed with the R package RWTY[49]. Likelihoods for the model parameters and resulting tree topologies were highlighted as a function of the number of sampled generations during the phylogenetic inference after burn-in of 25% (Supplementary Data 1). The phylogenetic trees were manually rooted using FigTree v1.4.4[50].

Gene cloning. Roche first-strand cDNA Synthesis Kit for rt-PCR (Roche Diagnostics, Mannheim, Germany) was used for cDNA synthesis from a pooled RNA sample of several larval stages (i.e., gastrula, trocho-

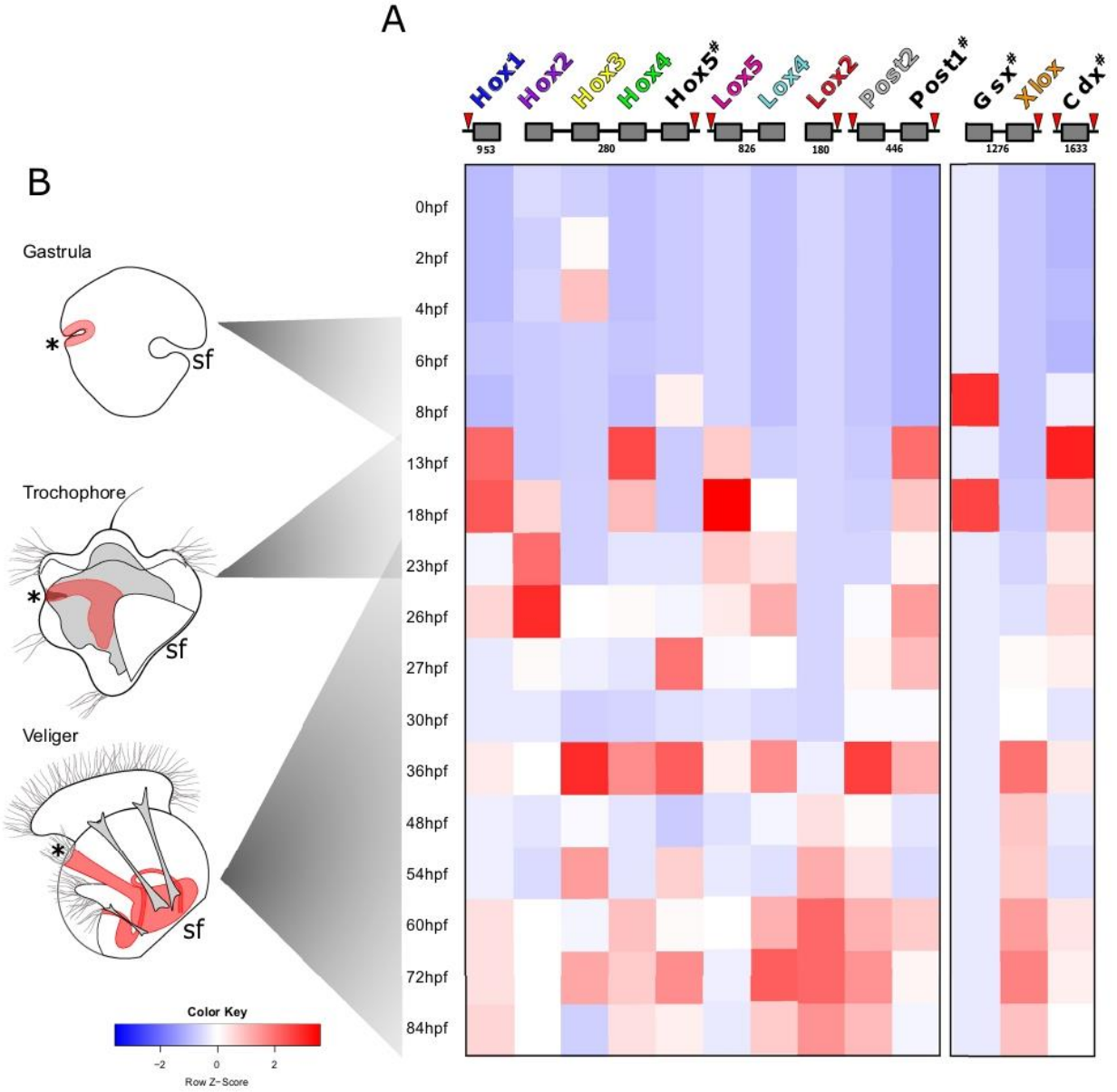


Figure 1. Relative quantitative analysis of Hox and ParaHox gene transcripts during development of *Dreissena rostriformis*. (**A**) Localization of Hox genes on scaffolds is shown in the upper part. Identification numbers of scaffolds where genes were found and location of non-Hox genes identified within the cluster are indicated by red arrowheads. Hashtags mark genes for which no expression data by in situ hybridization could be produced. (**B**) Heat map shows relative normalized expression levels for each gene (Z-score). Genes expressed above the normalization threshold are depicted in graded shades of red, those below the threshold in shades of blue. Time after fertilization and corresponding developmental stages at 23 °C are to the left. Germ layers and major derivatives in animal schemes are depicted in grey (mesoderm), red (endoderm), and white (ectoderm), respectively. Asterisks mark the blastopore/mouth, sf indicates the shell field.

phore and veliger). Identified Hox and ParaHox orthologs were used to design gene-specific primers. PCR amplified fragments were later separated by gel electrophoresis and purified with a QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). The products were cloned into pGEM-T Easy Vectors (Promega, Mannheim, Germany) and used to transform *E. coli* cells. Colonies were grown overnight, plasmids were later purified with a QIAprep Spin MiniprepKit (Qiagen) and sequenced for verification of insert orientation. For more descriptions of the whole experimental procedure see also[36,37,51–53]. Forward and reverse primers for each gene sequence can be found in Supplementary Table S2.

Figure 2. Hox gene expression in the late gastrula stage of *Dreissena rostriformis*. Graphical representation of the corresponding expression domains is to the right. Hox genes are not expressed in a staggered fashion. In the schemes, the future ectoderm is marked in white and the developing endoderm in red. Sites of gene expression are labeled blue and largely correspond to domains of future mesoderm. In the micrographs, the developing endoderm (site of gastrulation) is marked by dotted lines. (**A–C**) *Hox1* is expressed at the anterior margin of the shell field. (**D–F**) *Hox2* is expressed in the median region of the developing mesoderm. (**G–I**) *Hox3* is present in the antero-ventral region around the mouth. (**J–L**) *Hox4* is found in separate regions in the developing mesoderm. (**M–O**) *Lox5* is constrained to the ventro-median mesoderm. (**P–R**) *Lox4* is expressed in tissue underlying the posterior margin of the shell field. (**S–U**) *Lox2* is expressed in the mesoderm that extends towards the posterior region. (**V–X**) *Post2* is found in the posterior margin of the shell field. (**Y**,**Z**) *Xlox* is expressed in a group of cells in an anterior mesodermal domain. Arrowheads mark the blastopore. a, anterior; p, posterior; v, ventral; d, dorsal; l, Left; r, right. Scale bars equal 20 µm.

Probe synthesis and whole mount in situ hybridization. Plasmid inserts were amplified by PCR amplification using M13 primers as described previously[36,52]. Antisense riboprobes were synthesized using digoxigenin-UTP (DIG RNA Labeling Kit, Roche Diagnostics) and SP6/ T7 polymerase (Roche Diagnostics).

For whole mount in situ hybridization, animals stored in 100% methanol were rehydrated in 0.1 M PBS and decalcified with PPE (20% PFA + 10 × PBS + 0.5 M EGTA at pH 8 + diethylpyrocarbonate; DEPC-treated H_2O). Afterwards, samples were treated with proteinase-K at 37 °C for 10 min (10 µg/ml in PTw: 1 × PBS + 0.1% Tween-20). Samples were washed in PTw and post-fixed for 45 min in 4% PFA. Subsequently, samples were washed and transferred to hybridization buffer (50% formamide, 5 × SSC, 50 µg/ml heparin, 500 µg/ml yeast tRNA, 0.1% tween-20, pH 6.0) for 8–10 h at 56–60 °C. Probe hybridization was performed at the same temperature with probe concentrations ranging between 1 and 2 ng/µL for 30–48 h. Washing steps after hybridization were performed in wash buffer (75% hybridization buffer + 25% of SSC (3 M NaCl + 0.3 M saline-sodium citrate; SSC buffer + DEPC H_2O)) three times for 10 min each, then twice in 50% hybridization buffer + 50% SSC for 10 min each, and finally three times in 25% hybridization buffer + 75% SSC for 7 min (once) and 1xSSC + 0.1%Tween-20 for 5 min each. Next, three washes in 0.1 M MAB (maleic acid buffer) were performed. A digoxigenin(DIG)-labeled alkaline phosphatase (AP)-antibody raised in sheep (Roche Diagnostics) was used at a dilution of 1:5,000 in blocking solution (10 × blocking reagent, Roche Diagnostics, + 0.1 M MAB) at 4 °C overnight. Samples were then washed in PTw three times for 20 min each and twice for 10 min each. Development of the color reaction was done in a NBT/BCIP solution (Roche Diagnostics) diluted in 1 × alkaline phosphatase buffer (0.5 M NaCl + 0.5 M Tris at pH 9.5 + 50 mM $MgCl_2$ + 0.1%Tween-20 + DEPC H_2O) at 4 °C with periodic buffer replacement until signal was detected. Color reactions were stopped by washing five times for 10 min each in PTw.

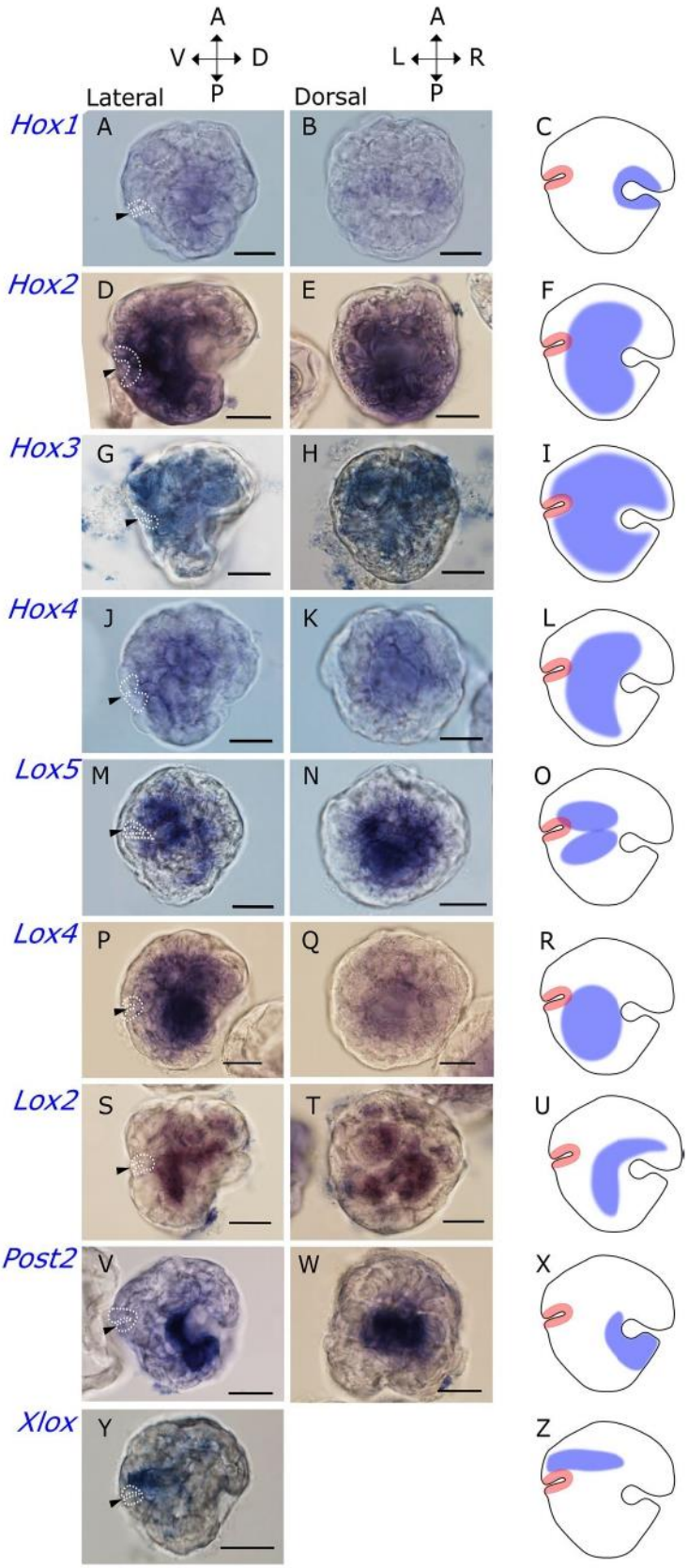
Image collection and figure processing. Larvae were mounted in 70% glycerol and examined and documented with an Olympus BX53 Microscope (Olympus, Hamburg, Germany). Light micrographs were processed in Inkscape (version 0.92.4–4) for contrast and brightness. All figures and graphical representations were prepared in Inkscape (version 0.92.4–4).

Results

Phylogenetic analysis of gene orthologs. Candidate orthologs of all ten Hox genes (*Hox7* appears to have been lost as in *Crassostrea gigas*) and three ParaHox genes were identified from the quagga mussel transcriptome. Multiple sequence alignment of the putative Hox proteins of *Dreissena rostriformis* and other bilaterians shows high sequence conservation within the homeodomain regions (Supplementary Data 2). Tree topologies obtained from Bayesian inference and maximum likelihood methods are highly consistent and reveal similar clusters among corresponding ortholog groups. All *Dreissena* gene candidates cluster with either a mollusk or another lophotrochozoan ortholog sequence (Supplementary Table S3, Supplementary Data 3, 4). *Dro-Hox1-5*, *Dro-Lox2, Dro-Lox4, Dro-Lox5*, *Dro-Post1*, and *Dro-Post2,* together with the ParaHox sequences *Dro-Gsx, Dro-Xlox*, and *Dro-Cdx*, were successfully cloned and the respective riboprobes were synthesized. All probes except those for *Dro-Hox5*, *Dro-Post1*, *Dro-Cdx*, and *Dro-Gsx* yielded results by in situ hybridization. The negative results of the latter were most likely due to low expression levels of these genes in the targeted developmental time points (Supplementary Table S1, Supplementary Data 5).

Hox and ParaHox gene arrangement. *Dreissena rostriformis* Hox and ParaHox genes are distributed along several scaffolds of the genome, which may either reflect the fragmented state of the assembly or disruption of the respective clusters (Supplementary Table S4, Supplementary Data 6). The Hox genes were distributed over five scaffolds, while two of the three ParaHox genes (*Gsx*, *Xlox*) were found on a single scaffold. Several non-Hox genes were found between members of the Hox cluster (between *Xlox* and *Cdx*, *Hox5* and *Lox5*, *Lox2* and *Post2*), demonstrating insertions all along the cluster (Fig. 1a, Supplementary Table S4, Supplementary Data 6). This indicates a somewhat disorganized (albeit not broken) Hox cluster[12].

Hox and ParaHox gene expression in *Dreissena rostriformis*. First detection of Hox and ParaHox gene expression by in situ hybridization started in the late gastrula stage (i.e. after 10hpf), in which *Dreissena* is heavily ciliated yet lacks a distinct mesodermal layer. The large invagination of the shell field marks the future dorsal side, and the smaller invagination of the blastopore lies on the opposite, ventral side (Fig. 2). Although this time point was not directly sampled for quantitative analysis, data from 8 and 13hpf old individuals suggest that significant transcription rates start between these stages. Both Hox and ParaHox gene expression is main-


A
V
D
P
Lateral
A
L
R
P
Dorsal
Hox1
A
B
C
Hox2
D
E
F
Hox3
G
H
I
Hox4
J
K
L
Lox5
M
N
O
Lox4
P
Q
R
Lox2
S
T
U
Post2
V
W
X
Xlox
Y
Z

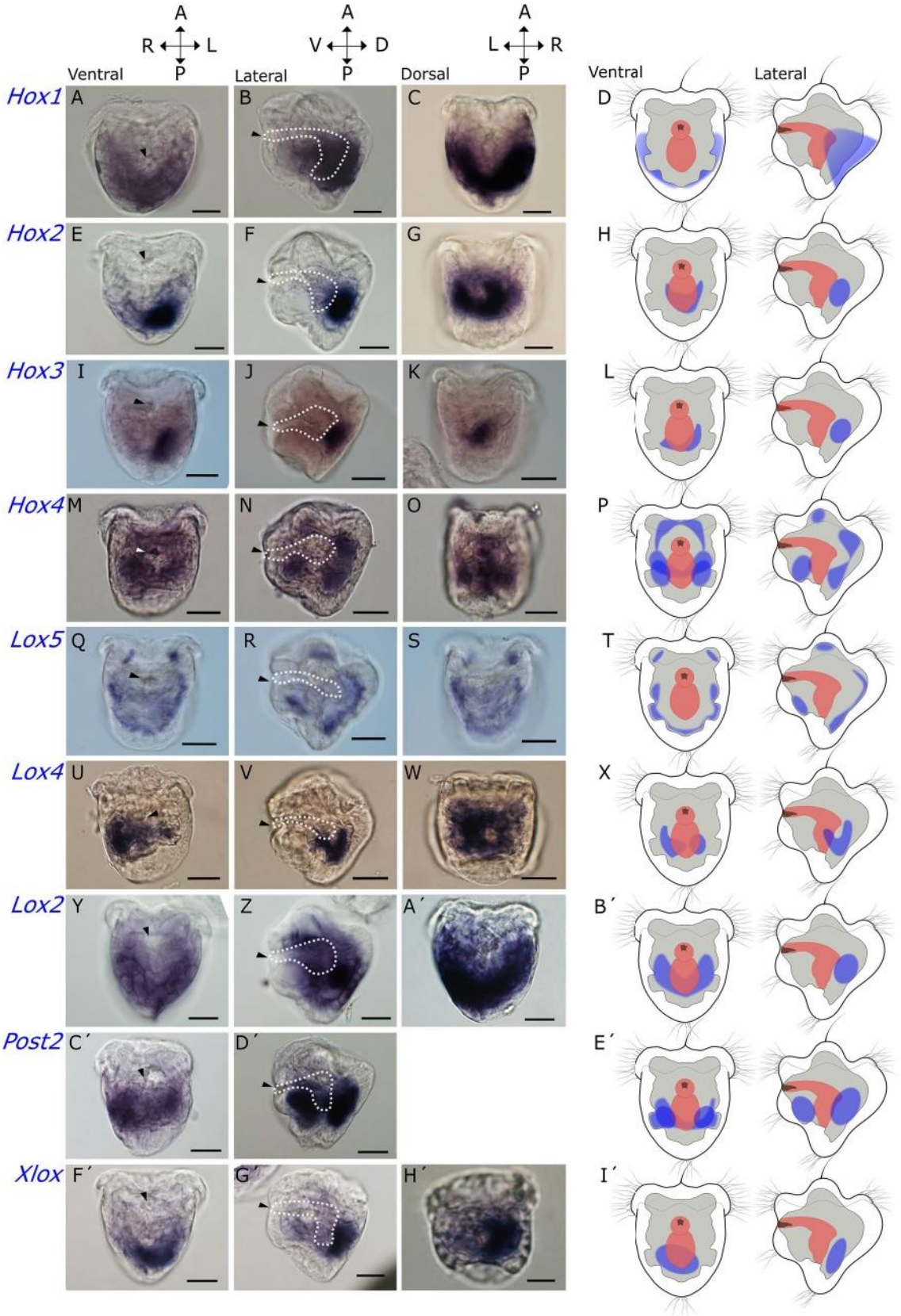

A
R
L
P
Ventral
A
V
D
P
Lateral
A
L
R
P
Dorsal
Ventral
Lateral
Hox1
A
B
C
D
Hox2
E
F
G
H
Hox3
I
J
K
L
Hox4
M
N
O
P
Lox5
Q
R
S
T
Lox4
U
V
W
X
Lox2
Y
Z
A´
B´
Post2
C´
D´
E´
Xlox
F´
G´
H´
I´

Figure 3. Hox gene expression in the trochophore stage of *Dreissena rostriformis*. Graphical representation of the corresponding expression domains is to the right with outline of the developing digestive tract in red, ectodermal domains in white, mesoderm in grey, and sites of gene expression in blue. In the micrographs, dotted lines mark the developing digestive tract. Hox genes are mostly co-expressed in regions of the developing mesoderm in a non-staggered fashion. (**A**–**D**) *Hox1* is expressed in all cells of the shell field, i.e. in domains dorso-lateral of the developing digestive tract. (**E**–**H**) *Hox2* is expressed in the mesodermal layer that underlies the shell field. (**I**–**L**) *Hox3* is expressed in the mesoderm that underlies the shell. (**M**–**P**) *Hox4* is expressed in separated cell masses belonging to the developing mesoderm. (**Q**–**T**) *Lox5* is expressed in individual mesodermal cell clusters throughout the larval body. (**U**–**X**) *Lox4* is expressed in the dorsal mesoderm that underlies the shell field. (**Y**–**B′**): *Lox2* is expressed in the dorsal mesoderm below the shell field and surrounds parts of the developing digestive tract. (**C′**–**E′**) *Post2* is found on the postero-ventral side. (**F′**–**I′**) *Xlox* is expressed in cells that lie adjacent to the developing hindgut. Arrowheads mark the mouth opening. a, anterior; p, posterior; v, ventral; d, dorsal; l, left; r, right. Scale bars equal 20 µm.

tained throughout the trochophore stage (i.e. around 13-16hpf) in which larvae become slightly elongated, the shell field has already evaginated, and the prototroch is distinct (Fig. 3). Hox and ParaHox gene expression was not detected in veliger stages.

Hox 1. Expression of *Dro-Hox1* is restricted to the shell field in the late gastrula and in the trochophore stage (Figs. 2a–c, 3a–d). Expression of *Hox1* is initially found at the anterior margin of the shell field in gastrula stage. In the trochophore larva the expression domain expands to all cells of the shell field.

Hox 2. *Dro-Hox2* shows strong expression in the median region of the developing mesoderm of the gastrula and is located towards the posterior half of the embryo (Fig. 2d–f). Subsequently, mesodermal expression becomes confined to the median region of the trochophore that corresponds to the mesodermal layer underlying the shell field (Fig. 3e–h).

Hox 3. *Dro-Hox3* is widely expressed in the mesoderm of the gastrula. Its expression domain partially overlaps with *Hox2*, although *Hox3* transcripts are distributed more widely all across the larval body (Fig. 2g–i). The unspecific stain around the periphery of the gastrula constitutes non-larval material. In the trochophore, *Hox3* expression overlaps with most genes except *Hox1*, *Post2*, and *Xlox*. Expression is confined to a mesodermal domain that underlies the shell field (Fig. 3i–l).

Hox 4. In the gastrula stage, low levels of *Dro-Hox4* expression are evident in the developing mesoderm, with expression overlapping with *Dro-Hox2* and *Dro-Hox3* expression domains. In the late trochophore, *Hox4* is expressed in separated cell clusters that belong to the developing mesoderm and are distributed all over the larval body (Figs. 2j–l, 3m–p).

Lox 5. In the gastrula stage, the expression domain of *Lox5* partially overlaps with *Dro-Hox2, Dro-Hox3*, and *Dro-Hox4*. Expression at this stage is restricted to the presumptive ventro-median mesoderm (Fig. 2s–u). In the late trochophore stage, gene expression occurs in individual mesodermal cell clusters throughout the larval body (Fig. 3q–t).

Lox 4. In the gastrula, expression of *Dro-Lox4* is located in mesodermal tissue that underlies the posterior margin of the shell field (Fig. 2p–r). In the trochophore stage, *Dro-Lox4* is expressed in the dorsal mesoderm that underlies the shell field (Fig. 3u–x).

Lox 2. *Dro-Lox2* shows a clear shift of expression between developmental stages. In the gastrula stage, the expression is in median mesodermal tissue that extends towards the posterior region of the embryo. In the trochophore, transcripts are located in the dorsal mesoderm below the shell field. The pattern overlaps with *Dro-Hox2*, *Dro-Hox3*, and, to some extent, with *Dro-Lox4* (Figs. 2m–o, 3y–b′).

Post 2. *Dro-Post2* expression is located in cells of the posterior margin of the shell field in the gastrula, opposite to *Dro-Hox1* (Fig. 2v–x). In the trochophore stage, transcripts are located more posteriorly on the ventral side of the larva, where expression co-localizes with the foot anlage (Fig. 3c′–e′).

Xlox. *Dro-Xlox* is first detected in the gastrula stage, in a group of cells corresponding to the anterior region of the developing digestive tract, i.e. in an endodermal domain, and in parts of the emerging anterior mesoderm (Fig. 2y, z). In the trochophore, *Dro-Xlox*-expressing cells lie adjacent to the developing hindgut (Fig. 3f′–i′).

Discussion

Non-staggered Hox gene expression in *Dreissena*. Hox genes in bilaterians are predominantly expressed along the anterior-posterior axis. They determine regionalization of morphological structures across the animal body. Since this staggered spatial expression is present in most clades, it is considered ancestral for Bilateria[9–12,54,55]. In bivalves, similar spatially staggered expression of *Hox1*, *Hox4*, *Lox5*, and *Post2* were proposed for a scallop, *Mizuhopecten yessoensis*. However, only the gastrula stage was investigated in this bivalve and data

are therefore largely incomplete. Transcriptomic data on this scallop seem to suggest a subcluster-level of temporal staggered expression whereby the first genes of each subcluster are expressed simultaneously, followed by the sequential expression of each subsequent gene within each subcluster[35]. By contrast, stage-specific quantitative transcriptome data for another bivalve, *Crassostrea gigas*, appears to more closely reflect the results from *Dreissena* insofar as no evidence of sequential or collinear activation of Hox genes was found[28,56].

In *Dreissena*, the relative expression domains of a few genes (*Hox1, Post2, Lox4*) might be interpreted as "partially spatially staggered" in the gastrula stage. *Dro-Lox4* maintains its pattern of expression in the dorso-posterior half of the gastrula and the trochophore, while *Dro-Hox1* is expressed anteriorly in the dorsal region of the gastrula. However, during the transition to the trochophore stage, *Dro-Hox1* expands throughout the entire shell field. Such a *Hox1* expression domain has also been found in gastropods, scaphopods, and the scallop, and might constitute an autapomorphy of Conchifera[26,35,37,40]. All other *Dreissena* Hox genes display highly dynamic spatial expression profiles with changes in expression domains frequently occurring between the gastrula and trochophore stages. As such, *Dro-Post2* expression is initially restricted to the posterior margin of the presumptive shell field, in close proximity to *Dro-Hox1*, but later extends ventro-posteriorly in the trochophore. This non-staggered expression of *Dro-Hox1, Dro-Hox2, Dro-Hox3, Dro-Hox4, Dro-Lox2, Dro-Lox5, Dro-Post2*, and *Dro-Xlox* is also reflected by their non-collinear but rather synchronous temporal activation (Fig. 1, Supplementary Data 5).

Cooption of Hox genes in mollusks. Conchiferan Hox gene expression is often located in specific developing morphological structures and does not strictly follow an anterior–posterior gradient, although rudiments of an ancestral staggered mode of expression are recognized to varying degrees in individual lineages and developmental stages[16,18,26,36,37,53] (Fig. 4b). *Hox2, Hox3, Hox4, Lox5, Lox4*, and *Lox2* show an overlapping expression in the developing mesoderm in the gastrula stage in *Dreissena* (Figs. 2, 4a). Subsequently, the expression domains still overlap, but are more confined to dorsal parts of the mesoderm (Figs. 3, 4a). In other conchiferans, Hox genes show more distinct spatial staggering along the anterior–posterior axis. In pre-torsional gastropod larvae, *Hox2, Hox3, Hox4, Hox5*, and *Hox7* expression domains appear somewhat staggered in the developing nervous system, while *Hox1, Lox5, Lox4, Lox2, Post2*, and *Post1* do not. Instead, these genes are expressed in distinct morphological features such as the shell field, foot, prototroch, and the larval apical organ[14,20,26,37,38,40]. A similar situation is found in cephalopods, where *Hox1, Hox3, Hox5, Lox4*, and *Post2* expression is located in the arm crown of the bobtail squid *Euprymna scolopes* in a somewhat staggered manner, but are also active in the pedal ganglia, the palliovisceral ganglia, and the light organ. *Hox3, Hox5, Lox4*, and *Lox5* are additionally expressed in the developing funnel[21,22]. Trochophore larvae of the scaphopod *Antalis entalis* likewise show spatially staggered expression of seven out of nine *Hox* genes, with *Hox2*, *Hox4*, *Hox5, Lox5, Lox4*, and *Post1* being expressed in the anlagen of the ganglia, the mantle margin, and the foot[37]. Taken together, it appears that Hox genes have been recruited independently into novel expression domains in various conchiferan lineages and have most likely also acquired novel roles at the cost of their original function in anterior–posterior axis specification.

Interestingly, and strikingly different from any conchiferan investigated to date, the aculiferan polyplacophoran *Acanthochitona fascicularis* shows almost textbook-like spatial (but not temporal) Hox gene expression along the anterior–posterior axis in the trochophore stage[30,31,38]. Here, Hox genes are not confined to distinct morphological features but instead to well-defined axial territories, thus probably resembling the conserved bilaterian condition[11]. Moreover, the genome of the polyplacophoran *Acanthopleura granulata* shows evidence of an intact cluster of the 11 Hox genes[57]. This supports the assumption that the last common ancestor of at least Aculifera likewise might have showed staggered and possibly collinear expression of Hox genes, although a final argument cannot be made without data from the aplacophoran clades (Fig. 4b). The deviation from this ancestral condition and deployment into novel roles in the conchiferan lineages might have facilitated the evolution of the immense diversity of body plans in the Conchifera[39].

Comparative aspects of ParaHox gene expression. ParaHox genes are a set of genes which originated from a duplication of the ancestral metazoan Proto-Hox cluster[58]. They comprise the *Cdx, Xlox*, and *Gsx* genes, which are commonly expressed in a staggered fashion in the hindgut, midgut, and foregut, respectively. Thus, it is assumed that ParaHox genes have an ancestral function in digestive tract patterning and regionalization[58,59]. This is supported by staggered expression of these genes along the developing gut in the polychaete annelids *Platynereis dumerilii*, *Alitta virens*, and *Capitella teleta*[60–62]. Data from the polyplacophoran *Acanthochitona* support this notion, with *Cdx* being expressed in the region of the developing hindgut, but data on other ParaHox genes are still lacking[31]. Once again, the conchiferans deviate from this condition. Although gastropod ParaHox genes appear spatially staggered expressed in both the gut and developing ganglia[27], *Gsx* is expressed in the developing nervous system of scaphopods and cephalopods[36]. Interestingly, *Dro-Xlox* expression in *Dreissena* resembles more the condition of other bilaterians including deuterostomes, where it is expressed in the midgut, rather than the one of its conchiferan relatives[60–65]. Thus, in contrast to the Hox gene condition, our results suggest that in *Dreissena*, *Xlox* has retained its ancestral expression domain reminiscent of the last common bilaterian ancestor.

Staggered Hox expression versus body plan regionalization in Mollusca. The Hox complement of Mollusca comprises 11 genes. These are expressed in a staggered fashion in polyplacophorans, a representative of the aculiferan lineage with a number of putative conserved ancestral traits such as a serially arranged dorsoventral musculature, a nervous system with non-ganglionated longitudinal nerve cords, an unpronounced cephalic region, and an elongated foot that extends more than three quarters along the longitudinal body axis[30]. The non-regionalized anterior–posterior distribution of these morphological traits and the conserved mode of

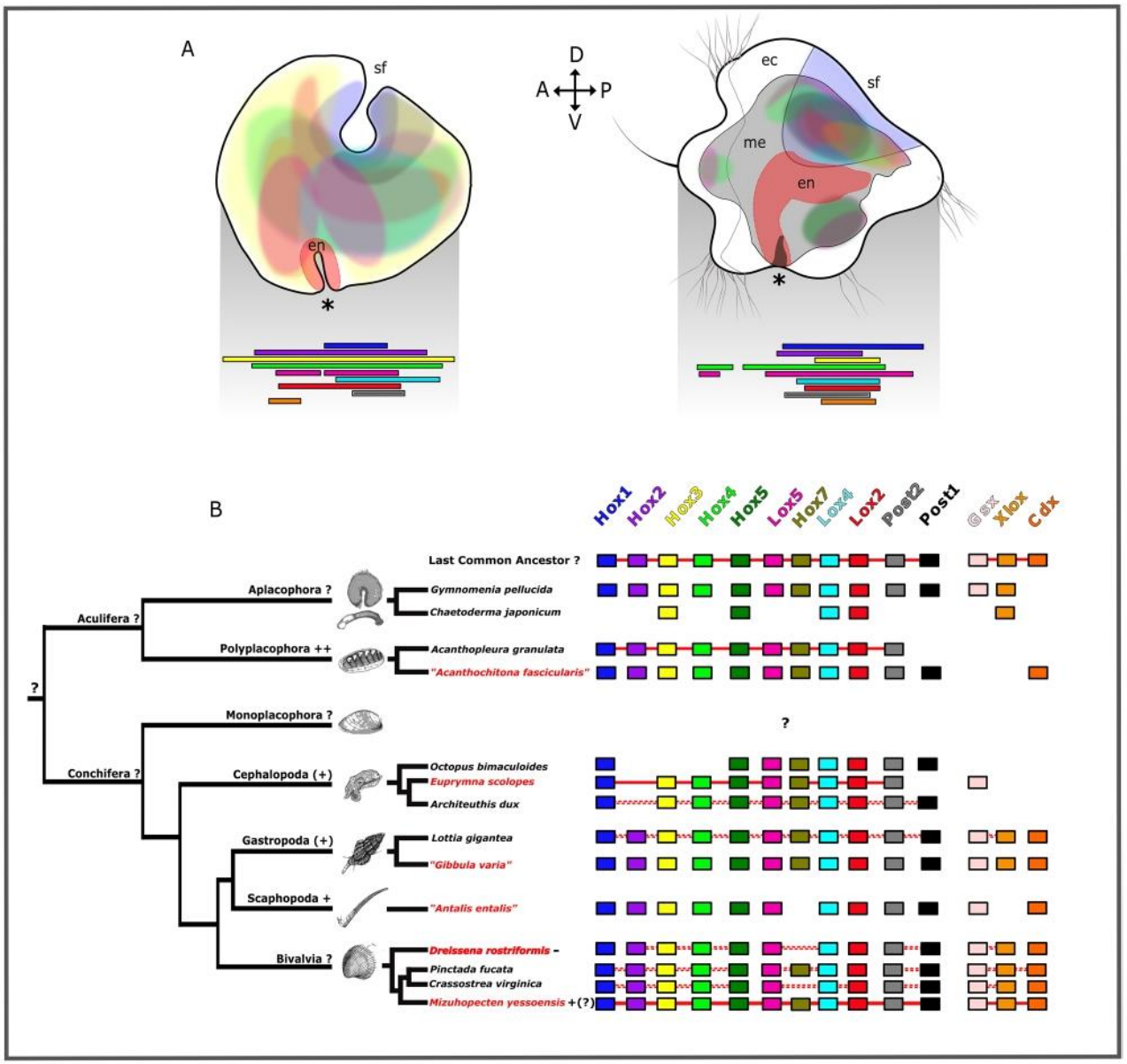


Figure 4. Schematic representation of Hox gene expression in *Dreissena rostriformis*. Summary of Hox expression patterns together with the localization of Hox genes in the genome of selected mollusk species. (**A**) Hox genes in the late gastrula (left) and the late trochophore larva (right) do not follow a temporally or spatially staggered mode of expression. (**B**) Hox gene arrangement in mollusks. Phylogenetic tree adapted from[1,66], animal schemes modified after[67]. Data from[16,18,19,35,39,57,68]. Straight red lines indicate absence of coding sequences between individual Hox genes, dotted red lines indicate presence of non-Hox genes between Hox genes, blank space between Hox genes means that presence or absence of non-Hox genes between Hox genes remains unknown. Species for which gene expression data by in situ hybridization are available are highlighted in red. ++ marks lineages with clear spatially staggered Hox expression, + indicates prominent hints of spatial Hox expression, (+) symbolizes weak hints of spatial expression, - indicates absence of staggered expression, ? indicates that no clear statement concerning staggered Hox expression can be made. Note that the situation in *Mizuhopecten yessoensis* is somewhat undecided since only few Hox genes were investigated in the gastrula stage only. Quotation marks indicate that species identity is not fully settled (*G. varia*, *A. entalis*) or that genomic and gene expression data for the identical species have been published under different species names (*A. fascicularis*, where gene expression data were erroneously assigned to *A. crinita*). a, anterior; d, dorsal; ec, ectoderm; en, endoderm; me, mesoderm; p, posterior; sf, shell field; v, ventral; asterisk, blastopore/mouth opening.

staggered Hox gene expression coincides with the situation in the polychaete annelid *Platynereis* which, despite significant differences with the polyplacophoran body plan such as segmentation and ganglionated nerve cords, likewise shows a largely evenly distributed set of organ systems along its longitudinal axis as well as staggered

expression of Hox genes[54]. From this common scheme, conchiferan molluscs have decoupled Hox gene expression from their original position along the anterior–posterior axis in various ways, depending on the degree of body regionalization. As such, scaphopod larvae with a pronounced longitudinal axis with relatively little structural concentration have largely retained the staggered mode of Hox expression, while gastropod veligers and cephalopod embryos with a high degree of organ system concentration, particularly in the anterior region, only show hidden traits of staggered Hox expression[37]. This independent loss (to varying degrees) of staggered Hox gene expression in the scaphopod-gastropod-cephalopod lineages has been replaced by coopted expression of these genes in (sometimes lineage-specific) distinct morphological structures such as the shell, the foot including funnel, the tentacles, or the larval prototroch. The situation in the bivalve *Dreissena* differs from that of their conchiferan (and polyplacophoran) relatives insofar as Hox genes are detected by in situ hybridization only until the trochophore stage, a larval type with little structural complexity. By the time the complex and regionalized veliger larva starts to establish, Hox gene expression is lost. Thus, the widely overlapping, non-staggered expression domains in the *Dreissena* trochophore and the low expression level during veliger patterning indicate that these genes might, at best, play a minor role in axis patterning in this bivalve, similar to the gastropods and cephalopods.

In conclusion, the data currently available corroborate the high plasticity of Hox gene expression in Mollusca. The use of functional assays and gene manipulation tools such as RNAi or the CRISPR/Cas9 system could help elucidate the function of individual Hox genes and the molecular basis of the developmental and evolutionary pathways that have resulted in the manifold morphological novelties in conchiferan mollusks.

Received: 17 August 2020; Accepted: 15 January 2021
Published online: 11 February 2021

References

1. Smith, S. A. *et al.* Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* **480**, 364–367 (2011).
2. McGinnis, W. & Krumlauf, R. Homeobox genes and axial patterning. *Cell* **68**, 283–302 (1992).
3. Pearson, J. C., Lemons, D. & McGinnis, W. Modulating Hox gene functions during animal body patterning. *Nat. Rev. Genet.* **6**, 893–904 (2005).
4. Carroll, S. B. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell* **134**, 25–36 (2008).
5. Alexander, T., Nolte, C. & Krumlauf, R. Hox genes and segmentation of the hindbrain and axial skeleton. *Annu. Rev. Cell Dev.* **25**, 431–456 (2009).
6. Wellik, D. M. Hox genes and vertebrate axial pattern. *Curr. Top. Dev. Biol.* **88**, 257–278 (2009).
7. McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. & Gehring, W. J. A homologous protein-coding sequence in drosophila homeotic genes and its conservation in other metazoans. *Cell* **37**, 403–408 (1984).
8. Halder, G., Callaerts, P. & Gehring, W. J. Induction of ectopic eyes by targeted expression of the *eyeless* gene in *Drosophila*. *Science (80-.).* **267**, 1788–1792 (1995).
9. Dollé, P., Izpisúa-Belmonte, J.-C., Falkenstein, H., Renucci, A. & Duboule, D. Coordinate expression of the murine Hox-5 complex homeobox-containing genes during limb pattern formation. *Nature* **342**, 767–772 (1989).
10. Graham, A., Papalopulu, N. & Krumlauf, R. The murine and Drosophila homeobox gene complexes have common features of organization and expression. *Cell* **57**, 367–378 (1989).
11. Gaunt, S. J. Hox cluster genes and collinearities throughout the tree of animal life. *Int. J. Dev. Biol.* **62**, 673–683 (2018).
12. Duboule, D. The rise and fall of Hox gene clusters. *Development* **134**, 2549–2560 (2007).
13. Seo, H.-C. *et al.* Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. *Nature* **431**, 67–71 (2004).
14. Samadi, L. & Steiner, G. Expression of Hox genes during the larval development of the snail, Gibbula varia (L.)-further evidence of non-colinearity in molluscs. *Dev. Genes Evol.* **220**, 161–172 (2010).
15. Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526 (2012).
16. Albertin, C. B. *et al.* The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**, 220 (2015).
17. Pace, R. M., Grbić, M. & Nagy, L. M. Composition and genomic organization of arthropod Hox clusters. *Evodevo* **7**, 11 (2016).
18. Belcaid, M. *et al.* Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. *Proc. Natl. Acad. Sci. USA* **116**, 3030–3035 (2019).
19. da Fonseca, R. R. *et al.* A draft genome sequence of the elusive giant squid, Architeuthis dux. *Gigascience* **9**, (2020).
20. Giusti, A. F., Hinman, V. F., Degnan, S. M., Degnan, B. M. & Morse, D. E. Expression of a Scr/Hox5 gene in the larval central nervous system of the gastropod Haliotis, a non-segmented spiralian lophotrochozoan. *Evol. Dev.* **2**, 294–302 (2000).
21. Callaerts, P. *et al.* HOX genes in the sepiolid squid Euprymna scolopes: Implications for the evolution of complex body plans. *Proc. Natl. Acad. Sci.* **99**, 2088–2093 (2002).
22. Lee, P. N., Callaerts, P., De Couet, H. G. & Martindale, M. Q. Cephalopod Hox genes and the origin of morphological novelties. *Nature* **424**, 1061–1065 (2003).
23. Canapa, A., Biscotti, M. A., Olmo, E. & Barucca, M. Isolation of Hox and ParaHox genes in the bivalve Pecten maximus. *Gene* **348**, 83–88 (2005).
24. Pérez-Parallé, M. L., Carpintero, P., Pazos, A. J., Abad, M. & Sánchez, J. L. The HOX Gene Cluster in the Bivalve Mollusc Mytilus galloprovincialis. *Biochem. Genet.* **43**, 417–424 (2005).
25. Iijima, M., Akiba, N., Sarashina, I., Kuratani, S. & Endo, K. Evolution of Hox genes in molluscs: a comparison among seven morphologically diverse classes. *J. Molluscan Stud.* **72**, 259–266 (2006).
26. Samadi, L. & Steiner, G. Involvement of Hox genes in shell morphogenesis in the encapsulated development of a top shell gastropod (Gibbula varia L.). *Dev. Genes Evol.* **219**, 523–530 (2009).
27. Samadi, L. & Steiner, G. Conservation of ParaHox genes' function in patterning of the digestive tract of the marine gastropod Gibbula varia. *BMC Dev. Biol.* **10**, 74 (2010).
28. Zhang, G. *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49 (2012).
29. Focareta, L., Sesso, S. & Cole, A. G. Characterization of homeobox genes reveals sophisticated regionalization of the central nervous system in the European cuttlefish sepia officinalis. *PLoS One* **9**, (2014).
30. Fritsch, M., Wollesen, T., de Oliveira, A. L. & Wanninger, A. Unexpected co-linearity of Hox gene expression in an aculiferan mollusk. *BMC Evol. Biol.* **15**, 151 (2015).

31. Fritsch, M., Wollesen, T. & Wanninger, A. Hox and ParaHox gene expression in early body plan patterning of polyplacophoran mollusks. *J. Exp. Zool. Part B Mol. Dev. Evol.* **326**, 89–104 (2016).
32. Barucca, M., Canapa, A. & Biscotti, M. An Overview of Hox Genes in Lophotrochozoa: Evolution and Functionality. *J. Dev. Biol.* **4**, 12 (2016).
33. De Oliveira, A. L. *et al.* Comparative transcriptomics enlarges the toolkit of known developmental genes in mollusks. *BMC Genomics* **17**, 905 (2016).
34. Takeuchi, T. *et al.* Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zool. Lett.* **2**, 3 (2016).
35. Wang, S. *et al.* Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat. Ecol. Evol.* **1**, 0120 (2017).
36. Wollesen, T., Rodríguez Monje, S. V., Todt, C., Degnan, B. M. & Wanninger, A. Ancestral role of Pax2/5/8 in molluscan brain and multimodal sensory system development. *BMC Evol. Biol.* **15**, 231 (2015).
37. Wollesen, T., Rodriguez Monje, S. V., de Oliveira, A. L. & Wanninger, A. Staggered Hox expression is more widespread among molluscs than previously appreciated. *Proc. R. Soc. B Biol. Sci.* **285**, (2018).
38. Huan, P., Wang, Q., Tan, S. & Liu, B. Dorsoventral decoupling of Hox gene expression underpins the diversification of molluscs. *Proc. Natl. Acad. Sci.* **117**, 503–512 (2020).
39. Wanninger, A. & Wollesen, T. The evolution of molluscs. *Biol. Rev.* **94**, 102–115 (2019).
40. Hinman, V. F., O'Brien, E. K., Richards, G. S. & Degnan, B. M. Expression of anterior Hox genes during larval development of the gastropod Haliotis asinina. *Evol. Dev.* **5**, 508–521 (2003).
41. Calcino, A. D. *et al.* The quagga mussel genome and the evolution of freshwater tolerance. *DNA Res.* **26**, 411–422 (2019).
42. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525 (2016).
43. Warnes, G. R. *et al.* gplots: Various R programming tools for plotting data. *R Packag. version* **2**, 1 (2009).
44. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
45. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
46. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. Europe PMC Funders Group ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2017).
47. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
48. Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.* **61**, 539–542 (2012).
49. Warren, D. L., Geneva, A. J. & Lanfear, R. RWTY (R We There Yet): an R package for examining convergence of bayesian phylogenetic analyses. *Mol. Biol. Evol.* **34**, 1016–1020 (2017).
50. Rambaut, A. FigTree 1.4. 2 software. *Inst. Evol. Biol. Univ. Edinburgh* (2014).
51. Redl, E., Scherholz, M., Wollesen, T., Todt, C. & Wanninger, A. Cell Proliferation Pattern and Twist Expression in an Aplacophoran Mollusk Argue Against Segmented Ancestry of Mollusca. *J. Exp. Zool. Part B Mol. Dev. Evol.* **326**, 422–436 (2016).
52. Wollesen, T., Rodríguez Monje, S. V., McDougall, C., Degnan, B. M. & Wanninger, A. The ParaHox gene Gsx patterns the apical organ and central nervous system but not the foregut in scaphopod and cephalopod mollusks. *Evodevo* **6**, 41 (2015).
53. Wollesen, T. *et al.* Brain regionalization genes are co-opted into shell field patterning in Mollusca. *Sci. Rep.* **7**, 5486 (2017).
54. Kulakova, M. *et al.* Hox gene expression in larval development of the polychaetes Nereis virens and Platynereis dumerilii (Annelida, Lophotrochozoa). *Dev. Genes Evol.* **217**, 39–54 (2006).
55. Duboule, D. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev. Suppl.* **42**, 135–142 (1994).
56. Paps, J., Xu, F., Zhang, G. & Holland, P. W. H. Reinforcing the egg-timer: recruitment of novel lophotrochozoa homeobox genes to early and late development in the pacific oyster. *Genome Biol. Evol.* **7**, 677–688 (2015).
57. Varney, R. M., Speiser, D. I., McDougall, C., Degnan, B. M. & Kocot, K. M. The iron-responsive genome of the chiton Acanthopleura granulata. *bioRxiv* 2020.05.19.102897 (2020) https://doi.org/10.1101/2020.05.19.102897.
58. Brooke, N. M., Garcia-Fernàndez, J. & Holland, P. W. H. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* **392**, 920–922 (1998).
59. Holland, P. W. H. Beyond the Hox: How widespread is homeobox gene clustering?. *J. Anat.* **199**, 13–23 (2001).
60. Fröbius, A. C. & Seaver, E. C. ParaHox gene expression in the polychaete annelid Capitella sp. I. *Dev. Genes Evol.* **216**, 81 (2006).
61. Kulakova, M. A., Cook, C. E. & Andreeva, T. F. ParaHox gene expression in larval and postlarval development of the polychaete Nereis virens (Annelida, Lophotrochozoa). *BMC Dev. Biol.* **8**, 61 (2008).
62. Hui, J. H. L. *et al.* Features of the ancestral bilaterian inferred from Platynereis dumerilii ParaHox genes. *BMC Biol.* **7**, 43 (2009).
63. Jonsson, J., Carlsson, L., Edlund, T. & Edlund, H. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* **371**, 606–609 (1994).
64. Osborne, P. W., Benoit, G., Laudet, V., Schubert, M. & Ferrier, D. E. K. Differential regulation of ParaHox genes by retinoic acid in the invertebrate chordate amphioxus (*Branchiostoma floridae*). *Dev. Biol.* **327**, 252–262 (2009).
65. Annunziata, R., Martinez, P. & Arnone, M. I. Intact cluster and chordate-like expression of ParaHox genes in a sea star. *BMC Biol.* **11**, 68 (2013).
66. Kocot, K. M., Poustka, A. J., Stöger, I., Halanych, K. M. & Schrödl, M. New data from Monoplacophora and a carefully-curated dataset resolve molluscan relationships. *Sci. Rep.* **10**, 101 (2020).
67. Giribet, G. *et al.* Evidence for a clade composed of molluscs with serially repeated structures: Monoplacophorans are related to chitons. *Proc. Natl. Acad. Sci.* **103**, 7723 LP–7728 (2006).
68. Li, Y. *et al.* Reconstruction of ancient homeobox gene linkages inferred from a new high-quality assembly of the Hong Kong oyster (Magallana hongkongensis) genome. *BMC Genomics* **21**, 713 (2020).

Acknowledgements

This study was supported by Grant P29455-B29 of the Austrian Science Fund (FWF) to AW.

Author contributions

A.W. designed the research. D.A.S.D. performed most parts of the research. A.D.C. contributed to the bioinforamtics analyses. A.L.D.O. contributed to the phylogenetic analyses. D.A.S.D. drafted the manuscript with input from A.W. All authors commented and provided input during manuscript writing. D.A.S.D. and A.W. finalized the manuscript. All authors approved the final version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-82122-6.

Correspondence and requests for materials should be addressed to A.W.

Reprints and permissions information is available at www.nature.com/reprints.

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

2.2. Chapter 2. Single-cell RNA sequencing atlas from a bivalve larva enhances classical cell lineage studies (published in *Frontiers of Ecology and Evolution*)

Salamanca-Díaz, D.A., Schulreich, S.M., Cole, A.G., Wanninger, A.

We analyzed the gene sets that are expressed during an ontogenetic stage of *D. rostriformis*. We utilized up to date methods for sequencing live, dissociated cells (Single Cell RNA-seq), followed by in silico pipelines which allowed to characterize transcriptomic profiles from seven distinct cell types (i.e., undifferentiated, epidermal, endoderm, muscle, neuronal, ciliary/prototroch and shell field).

Using this annotated dataset, we calculated and analyzed developmental trajectories of the extant cell populations. This resulted in observations of relationships among clusters which can be traced and compared to the cell lineage tree of *D. rostriformis* (i.e., ciliary/prototroch, epidermal, endoderm, muscle, shell field and neuronal). Furthermore, our results allow to gain a first glimpse into cluster structure within already defined cell lineages and demonstrate how transcriptomic signatures of cell populations can be used to annotate and characterize clusters present in early developmental stages of a bivalve mollusk.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 26 January 2022
doi: 10.3389/fevo.2021.783984

Single-Cell RNA Sequencing Atlas From a Bivalve Larva Enhances Classical Cell Lineage Studies

David A. Salamanca-Díaz[1†], Stephan M. Schulreich[1†], Alison G. Cole[2†] and Andreas Wanninger[1†]*

[1] Unit for Integrative Zoology, Department of Evolutionary Biology, University of Vienna, Vienna, Austria, [2] Division of Molecular Evolution and Development, Department of Neuroscience and Developmental Biology, University of Vienna, Vienna, Austria

Ciliated trochophore-type larvae are widespread among protostome animals with spiral cleavage. The respective phyla are often united into the superclade Spiralia or Lophotrochozoa that includes, for example, mollusks, annelids, and platyhelminths. Mollusks (bivalves, gastropods, cephalopods, polyplacophorans, and their kin) in particular are known for their morphological innovations and lineage-specific plasticity of homologous characters (e.g., radula, shell, foot, neuromuscular systems), raising questions concerning the cell types and the molecular toolkit that underlie this variation. Here, we report on the gene expression profile of individual cells of the trochophore larva of the invasive freshwater bivalve *Dreissena rostriformis* as inferred from single cell RNA sequencing. We generated transcriptomes of 632 individual cells and identified seven transcriptionally distinct cell populations. Developmental trajectory analyses identify cell populations that, for example, share an ectodermal origin such as the nervous system, the shell field, and the prototroch. To annotate these cell populations, we examined ontology terms from the gene sets that characterize each individual cluster. These were compared to gene expression data previously reported from other lophotrochozoans. Genes expected to be specific to certain tissues, such as *Hox1* (in the shell field), *Caveolin* (in prototrochal cells), or *FoxJ* (in other cillia-bearing cells) provide evidence that the recovered cell populations contribute to various distinct tissues and organs known from morphological studies. This dataset provides the first molecular atlas of gene expression underlying bivalve organogenesis and generates an important framework for future comparative studies into cell and tissue type development in Mollusca and Metazoa as a whole.

Keywords: Bivalvia, development, *Dreissena*, evo-devo, trochophore, evolution, cell type

OPEN ACCESS

Edited by:
Maria Ina Arnone,
Stazione Zoologica Anton Dohrn, Italy

Reviewed by:
Jose Maria Martin-Duran,
Queen Mary University of London, United Kingdom
Nathan Kenny,
University of Otago, New Zealand

***Correspondence:**
Andreas Wanninger
andreas.wanninger@univie.ac.at

†ORCID:
David A. Salamanca-Díaz
orcid.org/0000-0003-2082-3939
Stephan M. Schulreich
orcid.org/0000-0002-5309-0228
Alison G. Cole
orcid.org/0000-0002-7515-7489
Andreas Wanninger
orcid.org/0000-0002-3266-5838

Specialty section:
This article was submitted to Evolutionary Developmental Biology, a section of the journal Frontiers in Ecology and Evolution

Received: *27 September 2021*
Accepted: *16 December 2021*
Published: *26 January 2022*

Citation:
Salamanca-Díaz DA, Schulreich SM, Cole AG and Wanninger A (2022) Single-Cell RNA Sequencing Atlas From a Bivalve Larva Enhances Classical Cell Lineage Studies. Front. Ecol. Evol. 9:783984. doi: 10.3389/fevo.2021.783984

INTRODUCTION

The organization and diversity of cell types constitute key features during the ontogenetic establishment of animal body plans. Distinct transcriptional profiles of cell populations are an important prerequisite for the formation of cell types, and, ultimately, tissues, of multicellular animals (Shapiro et al., 2013; Stegle et al., 2015; Arendt et al., 2016). Accordingly, deciphering the transcriptomic code underlying the ontogeny of these cell types is crucial for sound inferences of the shared evolutionary history of cells and tissues across the animal kingdom. Among the multicellular animals, adult mollusks exhibit a particularly high morphological diversity that has resulted in a number of autapomorphic characters on various phylogenetic levels

[e.g., radula, shell(s), foot, neuromuscular systems] (Smith et al., 2011; Wanninger and Wollesen, 2019). However, despite their unique innovations in the adult body plan, numerous mollusks display a conserved, ciliated, trochophore-type larva that is also found in other lophotrochozoans including annelids and platyhelminths (Nielsen, 2018). The molecular pathways underlying the morphological diversification from this common larval type into clade-specific features remains largely unresolved.

Traditional studies have focused on the ontogeny of different tissues by tracing the mitotic history of a cell and generating so-called fate maps, resulting in the reconstruction of so-called cell lineages (Lillie, 1895; Meisenheimer, 1900; Luetjens and Dorresteijn, 1995; Dictus and Damen, 1997; Render, 1997; Henry et al., 2004; Hejnol et al., 2007; Lyons et al., 2017; Farrell et al., 2018). While cell lineages for certain putative homologous structures (e.g., the prototroch) have been identified, striking differences in mitotic history details are evident even between closely related taxa, particularly within Mollusca (Wanninger and Wollesen, 2015). Moreover, some lineage diagrams do not assign fates to all terminal ends of the cell lineage tree, hindering broad comparisons across taxa (Meisenheimer, 1900; Luetjens and Dorresteijn, 1995). For more detailed insights into differences and shared features on the ontogeny of tissues and organ systems, gene expression analyses using *in situ* hybridization has been widely used on a variety of molluscan taxa (e.g., Hinman et al., 2003; Lee et al., 2003; Wollesen et al., 2015a,b, 2018; Redl et al., 2016; Wang et al., 2017; Salamanca-Díaz et al., 2021). These studies represent a framework for comparative analyses of genes underlying the formation of features such as the shell field, nervous system, foot, sensory structures, and others (Wanninger and Wollesen, 2019). However, this approach is hampered by the fact that only a fraction of genes are known to be active in certain regions and developmental stages of the respective target species. Novel approaches of sequencing live, dissociated cells (single-cell RNA sequencing, scRNA-seq) followed by *in silico* methods using fine-grained computational pipelines have allowed for characterization of transcriptomic profiles related to cell types and cell states based on their distinct, relative gene expression levels. Such studies have resulted in comprehensive lists of genes that are expressed during ontogeny (Tang et al., 2010; Trapnell et al., 2014; Achim et al., 2015; Satija et al., 2015; Trapnell, 2015; Vergara et al., 2017; Svensson et al., 2018; Zhong et al., 2018). Accordingly, recent studies have uncovered the presence of previously uncharacterized cell types and an unexpected transcriptomic heterogeneity amongst cell populations and have revealed numerous genes with hitherto unknown function (Karaiskos et al., 2017; Achim et al., 2018; Briggs et al., 2018; Farrell et al., 2018; Fincher et al., 2018; Plass et al., 2018; Sebé-Pedrós et al., 2018; Wagner et al., 2018; Foster et al., 2019, 2020; Siebert et al., 2019; Duruz et al., 2020; Wendt et al., 2020; García-Castro et al., 2021; Paganos et al., 2021; Sur and Meyer, 2021).

The quagga mussel *Dreissena rostriformis*, native to the Dnieper drainage of the Ukrainian Black Sea region (Nalepa and Schloesser, 2019), has managed to rapidly spread across European and North American freshwater bodies (Mills et al., 1993; Heiler et al., 2013; Aldridge et al., 2014; Karatayev et al., 2014; Wakida-Kusunoki et al., 2015; Nalepa and Schloesser, 2019). Together with its congener, the zebra mussel *Dreissena polymorpha*, they build dense populations which consume large amounts of phytoplankton with often damaging impact on native invertebrate and fish populations (Raikow, 2004; Strayer et al., 2004; McNickle et al., 2006; Karatayev et al., 2014; Nalepa and Schloesser, 2019). This makes the quagga and zebra mussel one of the most prevalent and successful invasive mollusks (Karatayev et al., 2007). Efforts to understand and manipulate the rate of colonization and dispersal for these species have mainly focused on the adult stage when the population is already established, while data on the dispersive larval stages, the trochophore and the veliger, are still lacking (McCartney et al., 2019).

Here, we present the first comprehensive scRNA-seq study on the trochophore larva of any mollusk, the quagga mussel *Dreissena rostriformis*. We generated a scRNA-seq library from dissociated cells of entire larvae. Using this dataset, we define distinct cell types present in this larval type. We assign identities to identified cell populations based on characterization of distinct transcriptomic signatures and predict relationships between cell types by modeling developmental trajectories within the dataset. We analyze ciliary, neuronal, muscle, and other cell types with potential importance to environment sensing and larval spreading. Through this work, we also provide a comprehensive molecular atlas of gene expression underlying *Dreissena rostriformis* development, thereby laying the foundation for further research into understanding the features that allow quagga mussel larvae to rapidly conquer new freshwater habitats.

MATERIALS AND METHODS

Animal Collection and Cultures

Sexually mature individuals of *Dreissena rostriformis* were collected in the Danube River in Vienna, Austria (N 48°14′45.812″, O 16°23′38.145″) between April and September, 2019. Adults were gathered from underneath stones and transferred to the laboratory where they were cleaned and maintained in aquaria with filtered river water (FRW) at 19°C.

Spawning of animals was induced by exposing sexually mature specimens for 15 min to a 10^{-3} M solution of serotonin (#H9523, Sigma-Aldrich, Darmstadt, Germany), followed by one wash and subsequent maintenance in FRW. Individuals were kept isolated in FRW inside 50 mL glass beakers and after approximately 30 min, up to 50% of the treated specimens started to spawn. Oocytes and motile sperm were then collected separately from the water column. Fertilization rates were high when three to four drops of sperm-containing water were added to 50 ml glass beakers with oocytes. The cultures were maintained at 23°C. After fertilization, water was changed every half an hour for the first 3 h and then every 6 h to remove excess sperm and avoid bacterial or fungal infection.

10X Single-Cell 3′ RNA-Seq

Sample Preparation

Cell dissociations of *Dreissena* larvae were generated by first washing 13 h post fertilization (hpf) trochophore larvae over a 20 μm mesh with sterile autoclaved fresh river water (AFRW).

Larvae were concentrated into 1.5 mL tubes by centrifugation (1rcf) for 5 min and AFRW was replaced with a pronase enzyme in AFRW (8 mg/mL) (#10165921001, Roche Diagnostics, Mannheim, Germany) and the tubes were subsequently placed in gentle agitation for 30 min on a rocker at the slowest setting in order to keep the cells and reagents in suspension. Larvae were dissociated by first passing them through a syringe with a hypodermic needle with 0.4 mm diameter. This procedure was repeated and monitored until a uniform single cell solution was visible under the microscope. Cell viability was assessed by adding an acridine orange/propidium iodide (AO/PI) staining solution (#CS2-0106; Nexcelom Bioscience, Lawrence, MA, United States) to stain live and dead cells, respectively (Cellometer K2; Nexcelom Bioscience). A single cell suspension measuring 76.8% viability with a concentration of 1000 cells/μL was loaded into a 10x Chromium Controller using Chromium Single Cell 3′ Kit v2 reagents (#120237, 10x Genomics, United States). cDNA synthesis and library construction were made according to specifications from the manufacturer. Library quantification was performed on a Bioanalyzer (#5067-4626, High Sensitivity DNA reagents, Agilent Technology; Agilent 2100 Bioanalyzer) and sequenced on the Illumina platform.

Mapping Tool Preparation and Cell Clustering

The transcriptomes used for creating the mapping tool, together with the reference genome to map the reads against, were previously generated (Calcino et al., 2019). Gene models were elongated 2 kilobases in the 3′direction to account for poorly annotated three-prime ends (Levin et al., 2016). To annotate the transcriptome, we performed a BLASTX search in each individual gene sequence against the human and the Pacific giant oyster (*Crassostrea gigas*) genomes. For each transcript, the BLAST hit with the highest *E*-value was selected for annotation. We utilized InterProScan v5.46-81.0 (Jones et al., 2014) to search for gene ontology and allocate domains on the reference genome by surveying publicly available databases such as GO terms, Pfam, and PANTHER (**Supplementary Table 1**). The reference database was generated using CellRanger MakeRef v3.1.0 with default settings. Afterward, sample libraries were demultiplexed using CellRanger Makefastq v3.1.0 with default settings and filtered according to cell barcode and Unique Molecular Markers (UMIs) using CellRanger Count v3.1.0 with parameters of –force-cells = 1000 –chemistry = SC3Pv2. The resulting cell count gene expression matrix was analyzed in R v3.6.1 (R Developement Core Team, 2015) with the Seurat v4.0.1 package (Satija et al., 2015). Samples were considered to represent a single cell transcriptome if it contained at least 150 genes. Samples with more than 4,000 sequenced genes were removed as these represent multiplets. The count matrix was processed through a standard Seurat pipeline using default parameters as follows: counts were first normalized, and log scaled (NormalizeData and ScaleData functions) and principal component analysis was performed (RunPCA) using only highly variable features (FindVariableFeatures). We visually inspected an elbow plot of standard deviation of the principal components (**Supplementary Figure 1**) and selected the top 30 principal components for clustering. We then generated a KNN graph based on the Euclidean distance in PCA space following the Louvain algorithm [FindNeighbors (k.param = 10, nn.method = "annoy," annoy.metric = "cosine")] and clustered the data [FindClusters (random.seed = 0, res = 0.5)]. For data visualization, we projected all single cell transcriptomes into a uniform space of 2 dimensions [Uniform Manifold Approximation and Projection (UMAP): RunUMAP (n.neighbors = 10 L, spread = 0.45, min.dist = 0.1, local.connectivity = 100)]. Marker genes were identified with the FindAllMarkers function (random.seed = 0), which considers only genes that were enriched and expressed in at least 10% of the cells in each population (min.pct = 0.1) and with a log fold difference larger than 0.6 (logfc.threshold = 0.6). Later, a filter was performed for the top 10 genes with the higher average logarithmic fold change values (**Supplementary Tables 2, 3**). Additional cutoff thresholds were explored and the whole pipeline was applied to observe how they affect the remaining cells in the dataset (**Supplementary Figure 2** and **Supplementary Table 5**). We applied a graph imputation algorithm with the R package MAGIC (van Dijk et al., 2018) on single cells for de-noising the count matrix and fill in missing transcripts. The R package topGO (Alexa et al., 2006) was implemented to perform a GO terms enrichment analysis on genes from each cluster, determining statistically the molecular activity of a gene (molecular function), the place in the cell where the gene produces an effect (cellular component), and the physiological process influenced by a gene (biological process). Following this, expression of orthologous marker genes from distinct cell types were examined in this dataset to validate cluster identification assigned by GO term results.

Pseudotime Trajectory Analysis

In order to root the pseudotime trajectories in a single place and to calculate the number of transcriptomic changes of each population, we implemented the R package StemID to predict developmentally uncommitted cell populations from maximum transcriptome entropy calculations (Grün et al., 2016). This one cell population, which clustered together based upon the lack of differentially expressed genes compared with the other clusters in the dataset, also received the lowest entropy calculation. This suggests that these cells represent a pool of uncommitted cells within the embryo (Hemmrich et al., 2012; Fincher et al., 2018; Siebert et al., 2019). Cells were ordered along a calculated similarity trajectory with this population defined as the starting point using the Monocle3 R package (Cao et al., 2019), imported from Seurat with the package SeuratWrappers v0.3. Cells were re-clustered using the "cluster_cells" function (resolution = 0.02). The "learn_graph" function was implemented to model the cell differentiation paths throughout development. This principal graph was used to order cells in pseudotime using the "orderCells" function with undifferentiated cells defined as the root. Monocle interprets multiple ends to a trajectory as cellular decisions.

Gene Cloning

A first-strand cDNA Synthesis Kit for RT-PCR (#11483188001, Roche Diagnostics, Mannheim, Germany) was used for cDNA

synthesis from the pooled RNA. PCR products were size fractioned by gel electrophoresis and purified with a QIAquick Gel Extraction Kit (#28706, Qiagen, Hilden, Germany). PCR products were cloned in pGEM-T Easy Vectors (#A1380, Promega, Mannheim, Germany). Plasmid minipreps were grown overnight, purified with a QIAprep Spin Miniprep Kit (#27106, Qiagen), and sequenced for verification. Forward and reverse primers for each gene sequence can be found in **Supplementary Table 4**.

Probe Synthesis and Whole Mount *in situ* Hybridization

To provide spatial information associated with the transcriptomic signatures and further validate cell type annotations, whole mount *in situ* hybridization analysis of some of the highly expressed cell population-specific genes were performed on trochophore larvae aged 13–15 hpf. Plasmid products were linearized by PCR amplification using M13 primers as described previously (Wollesen et al., 2015b). Antisense riboprobes were synthesized using digoxigenin-UTP (#11175025910, DIG RNA Labeling Kit, Roche Diagnostics) and SP6/ T7 polymerase (#10810274001, #1088176001, Roche Diagnostics).

Larvae were fixed for 1 h in 4% paraformaldehyde (PFA) (20% PFA + 10x PBS; DEPC- treated H_2O), followed by three washes of 15 min each in Phosphate Buffer Solution (PBS). For long-term storage of samples, subsequent washing steps for at least 15 min each were performed in different concentrations of methanol (25, 50, and 75%) in 1x PBS. For whole mount *in situ* hybridization, larvae stored in 100% methanol were rehydrated in 0.1 M PBS and decalcified with PPE (20% PFA + 10x PBS + 0.5 M EGTA at pH 8.0 + diethyl pyrocarbonate; DEPC-treated H_2O). Afterward, samples were treated with proteinase-K at 37°C for 10 min (10 μg/mL in PTw: 1x PBS + 0.1% Tween-20). Samples were washed in PTw and post-fixed for 45 min in 4% PFA. Subsequently, samples were washed and transferred to hybridization buffer (50% formamide, 5X SSC, 50 μg/ml heparin, 500 μg/ml yeast tRNA, 0.1% Tween-20, pH 6.0) for 8–10 h at 56–60°C. Probe hybridization was performed at the same temperature with probe concentrations ranging between 1 and 2 ng/μL for 30–48 h. Washing steps after hybridization were done using the following solutions: 75% hybridization buffer + 25% of SSC (3 M NaCl + 0.3 M saline-sodium citrate; SSC buffer + DEPC H_2O) for three times 10 min each, then 50% hybridization buffer + 50% SSC twice for 10 min each, 25% hybridization buffer + 75% SSC for 7 min (once) and 1X SSC + 0.1% Tween-20 for three times for 5 min each. Next, three washes in 0.1 M maleic acid buffer (MAB) were performed. A digoxigenin (DIG)-labeled alkaline phosphatase (AP) antibody (#11093274910, Roche Diagnostics) was used at a dilution of 1:5000 in blocking solution (#11096176001, 10x blocking reagent; Roche Diagnostics + 0.1 M MAB) at 4°C overnight. Samples were then washed in PTw three times for 20 min each and twice for 10 min each. Development of the color reaction was done in a NBT/BCIP solution (#11383213001, #11383221001, Roche Diagnostics) diluted in 1x alkaline phosphatase buffer (0.5 M NaCl + 0.5 M Tris at pH 9.5 + 50 mM $MgCl_2$ + 0.1% Tween-20 + DEPC H_2O) at 4°C with constant buffer replacement until signal was detected. Color reactions were stopped by washing five times for 10 min each with PTw.

Immunocytochemistry

Trochophore larvae (13 hpf) were fixed in 4% PFA for 1 h, then washed and stored at 4°C in 1x PBS with 0.1% sodium azide added. Afterward, samples were permeabilized by three washing steps of 15 min in 1x PBS + 0.2% Triton X-100 + 0.1% NaN_3 (PTA). Blocking was performed in PTA + 6% normal goat serum (NGS) overnight at 4°C. Primary antibody incubation was performed over night at 4°C with anti-acetylated α-tubulin (raised in mouse, monoclonal, #T6793, Sigma-Aldrich) at a concentration of 1:400 diluted in PTA + 6% NGS. Two washes with PTA of 15 min each followed. For the secondary antibody, goat anti-mouse Alexa 568 (#A-11004, Thermo Fisher Scientific, Waltham, MA, United States) was added and the samples were incubated overnight at 4°C. This was followed by two washes with PBS for 15 min each. After that, a mix of CellMask Green Stain (#H32714, Invitrogen, Carlsbad, CA, United States; 1:100) and DAPI (#D1306, Sigma-Aldrich, 1:2000) in PBS was added for 1 h. Then, the samples were washed thrice with PBS for 15 min each and mounted on glass slides with Fluoromount-G (#0100-01, SouthernBiotech, Birmingham, AL, United States) and stored at 4°C for a few days prior to image collection.

Image Collection and Figure Processing

Trochophore larvae (13 hpf) used for *in situ* hybridization were mounted in 100% glycerol and documented with an Olympus BX53 Microscope (Olympus, Hamburg, Germany). Immunofluorescent antibody preparations were scanned with a Leica TCS SP5 II confocal microscope (Leica Microsystems GmbH, Wetzlar, Germany). FIJI (Schindelin et al., 2012) was used for image analysis and further processing to estimate cell number in a larva through nuclei counting, employing the 3D Object Counter v2 with a threshold of 75. All figures, light micrographs, and graphical representations were prepared, compiled, and adjusted for contrast and brightness using Inkscape version 0.92.4-4 (Inkscape Project, 2020).

RESULTS

De novo Cell Cluster Annotation and Marker Gene Identification in a Bivalve Trochophore Larva

Based on nuclei counts from immunostainings, we estimated that the trochophore larva of the quagga mussel *Dreissena rostriformis* is composed of approximately 110 cells. Therefore, we aimed to sequence ~1000 cells with the 10x Genomics platform. In total we sequenced 41325580 reads and 84.9% of these reads mapped to the reference genome, i.e., the remaining percentage could not be mapped to either exonic, intronic, intergenic regions or antisense genes. After filtering out cells with less than 150 molecules and with more than 4000 sequenced molecules, a total of 632

cells were captured, corresponding to around five individuals (**Figures 1A,B**). The number of genes detected in a cell (nFeatures) have a median of 1054.5 and a mean of 1141.8, while the number of molecules detected within a cell (nCount) have a median of 3253 and a mean value of 4172.8. After processing the cell count matrix with the Seurat pipeline, the heterogeneity of cell states in the trochophore larvae was assessed. Graph-based clustering revealed seven transcriptionally distinct cell populations (**Figures 2A–G**). Analysis of the genes underlying this clustering, as described below, identified cell populations corresponding to "ciliary/prototroch," "neuronal," "endoderm," "muscle," "epidermal," "shell field," and "undifferentiated" identities (**Figures 3A, 4A**).

Transcriptional Domains Confirm Cell Type Identity in the Larva

Ciliary cells in *Dreissena* start to develop early in the gastrula stage. GO terms analysis revealed one cluster of cells with a statistically significant enrichment in genes involved in processes related to cilia assembly, movement and cell motility, and microtubule-based processes (**Figures 2A–A″**), thus identifying a "ciliary/prototroch" cell population within the dataset. Moreover, further sub-structure within this group is exhibited when the clustering resolution parameter is increased (**Supplementary Figures 1I–K** and **Supplementary Table 3**). In the larva, there are several subtypes of ciliated cells such as prototroch, telotroch, and apical tuft cells (Meisenheimer, 1900; Pavlicek et al., 2018). The structure seen in this cluster could reflect the different subtypes of ciliary cells in the live animal.

Cells identified as "neuronal" present highly expressed genes where there is an enrichment of neurogenesis-related processes such as neuropeptide signaling pathway, proton transmembrane transport (iron ion transport), and acetylcholine receptor regulator activity (**Figures 2B–B″** and **Supplementary Tables 1, 2**). When increasing the resolution for finding clusters, results revealed subclusters within this population (**Supplementary Figures 1I–K** and **Supplementary Tables 2, 3**). Even though the two resulting subclusters have equal enrichment for neuronal processes, they have differential expression of markers setting them apart from each other (**Supplementary Tables 2, 3**). For example, while the red subcluster has predominant expression of *FMRFamide receptor,* the turquoise one on the right presents high expression of *neuronal acetylcholine receptor subunit alpha-6* (**Supplementary Figures 1I, 4C,D**). Annotations from transcriptomic signatures in the cluster "neuronal" indicate that neuropeptides are secreted from these cells. Accordingly, in-depth analysis of gene expression signatures of this cluster appears particularly promising for characterizing the gene regulatory networks responsible for the differentiation of neuronal cells.

The cluster "endoderm" shows all related cells share gene annotations from human and *Crassostrea* orthologs related to enzymatic reactions that are likely to occur in the developing digestive system, e.g., *arginase-1* (Gene.72022), *cathepsin K preproprotein* (Gene.17238), *neutral cholesterol ester hydrolase-1* (Gene.4751) (Kirschke et al., 1995; Duruz et al., 2020), but also included immune response-related genes, e.g., *galectin-8* (Gene.117090) (Vasta et al., 2015). Moreover, GO terms for processes related to ribosome biogenesis, RNA processing, translation, and ATP synthesis-coupled proton transport show enrichment in this cluster (**Figures 2C–C″**). Since all these activities are crucial for cell survival, one could assume the enriched processes are crucial for all cell states at this stage and some of the genes are expressed in other clusters. However, the annotations from the top differentially expressed genes identified in this cluster suggest that these cells are involved in endoderm development and immune responses (**Supplementary Figure 4**). Altogether, cells in this cluster are highly likely to have endodermal fates in the quagga mussel larva.

Bivalves have two predominant adductor muscles of which the anterior adductor and some larval retractor systems already start to form in the trochophore larva (unpublished observation). We identified cells with muscle identity in our dataset on the "muscle" cluster where there is gene ontology enrichment with processes associated with muscle system formation such as calcium ion binding, actin filament depolymerization, and protein peptidyl-prolyl isomerization (protein folding) (**Figures 2D–D″** and **Supplementary Table 1**). In addition, the commonly expressed marker during animal myogenesis, *myosin heavy chain (mhc*; Castellani and Cohen, 1987; Ladurner and Rieger, 2000; Dyachuk and Odintsova, 2009; Li et al., 2019), is expressed in this cluster. In the *Dreissena* trochophore, transcripts of *mhc* are found in the anterior mesoderm. They form spot-like expression domains in the anterior region between the developing digestive tract and the shell field (**Figures 3B–B″,E**). Hence, *in silico* annotations and *in situ* hybridization data support the identity of muscle cells for this cluster.

Cells with an "epidermal" identity formed a cluster with GO terms enrichment in processes that play a role as integral components of the membrane, such as epidermis morphogenesis and homophilic cell adhesion via plasma membrane (**Figures 2E–E″** and **Supplementary Table 1**). To visualize these cells in the larva, *in situ* hybridization of one of the top marker genes was performed. The gene *teashirt* (*tsh*) has an expression domain in cell populations adjacent to the anterior and posterior margins of the cells forming the shell gland (**Figures 3D–D″,G**). This gene expression domain and GO term annotations suggest that this cluster contributes to the formation of various types of epithelia and epidermal tissue in the trochophore larva of *Dreissena*.

Dreissena trochophores exhibit a developing shell field which grows and envelops the visceral region in later stages. We identified cells belonging to the cluster "shell field," which express several genes with unknown ontology term or ortholog match. Cells belonging to this cluster express numerous short-numbered-kilobases genes, which match hitherto uncharacterized molluscan genes (**Supplementary Table 1**). Nevertheless, the few GO terms collected show enrichment in processes that interact with the extracellular matrix, such as integrin-mediated signaling pathway and positive regulation of cell cycle G2/M phase progress (**Figures 2F–F″**

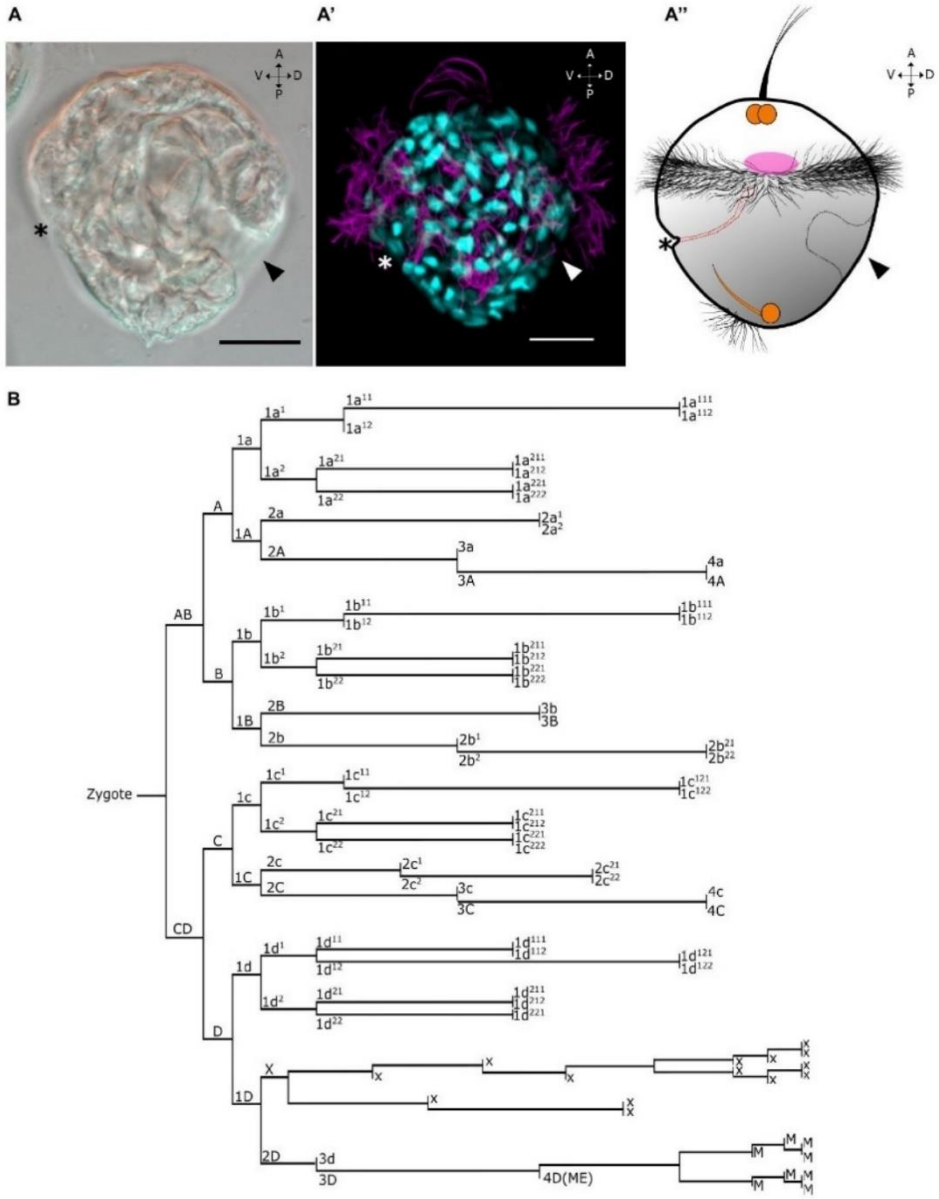


FIGURE 1 | Single-cell transcriptomics of *Dreissena rostriformis* larvae. Age of larvae is 13 hpf. **(A–A")** Lateral view of differential interference contrast (left), alpha-acetylated tubulin/ DAPI immunostained trochophores (center), and scheme illustrating the overall morphology of the trochophore larva (right). Asterisk on the larvae indicates the mouth opening. Arrowhead points to the shell field on the right side. Orange marked areas denote the developing nervous system based on data from Pavlicek et al. (2018). Pink marked area denotes the developing muscular system. Dotted black lines highlight shell field cells. Dotted red lines highlight the developing digestive system. Anterior (A), posterior (P), ventral (V), and dorsal (D). Scale bars equal 20 μm. **(B)** Cell lineage tree of *Dreissena* modified from Meisenheimer (1900) and Luetjens and Dorresteijn (1995).

and **Supplementary Table 1**). From the set of highly expressed genes forming the transcriptomic signature in this cluster, *hic31* was chosen to visualize the cluster in the *Dreissena* larva because it was previously shown to be involved in establishing the protein matrix prior to shell secretion (Liu et al., 2015, 2018). In *Dreissena* trochophores, *hic31* expression is present around the margin of the shell field (**Figures 3C–C″,F**). Thus, gene annotations and *in situ* hybridization data support the identity of shell field cells for this cluster.

We identified one cluster devoid of a unique transcriptional signature (**Figure 2G**). These cells were considered "undifferentiated" because their GO term annotations showed that they have significant levels of regulation of DNA replication, protein synthesis, and translational initiation in the *Dreissena* trochophore (**Figures 2G–G″**). The predominant expression of such "housekeeping genes" as well as the shared expression signatures of these genes with all clusters, in combination with a lack of distinct expression of marker genes

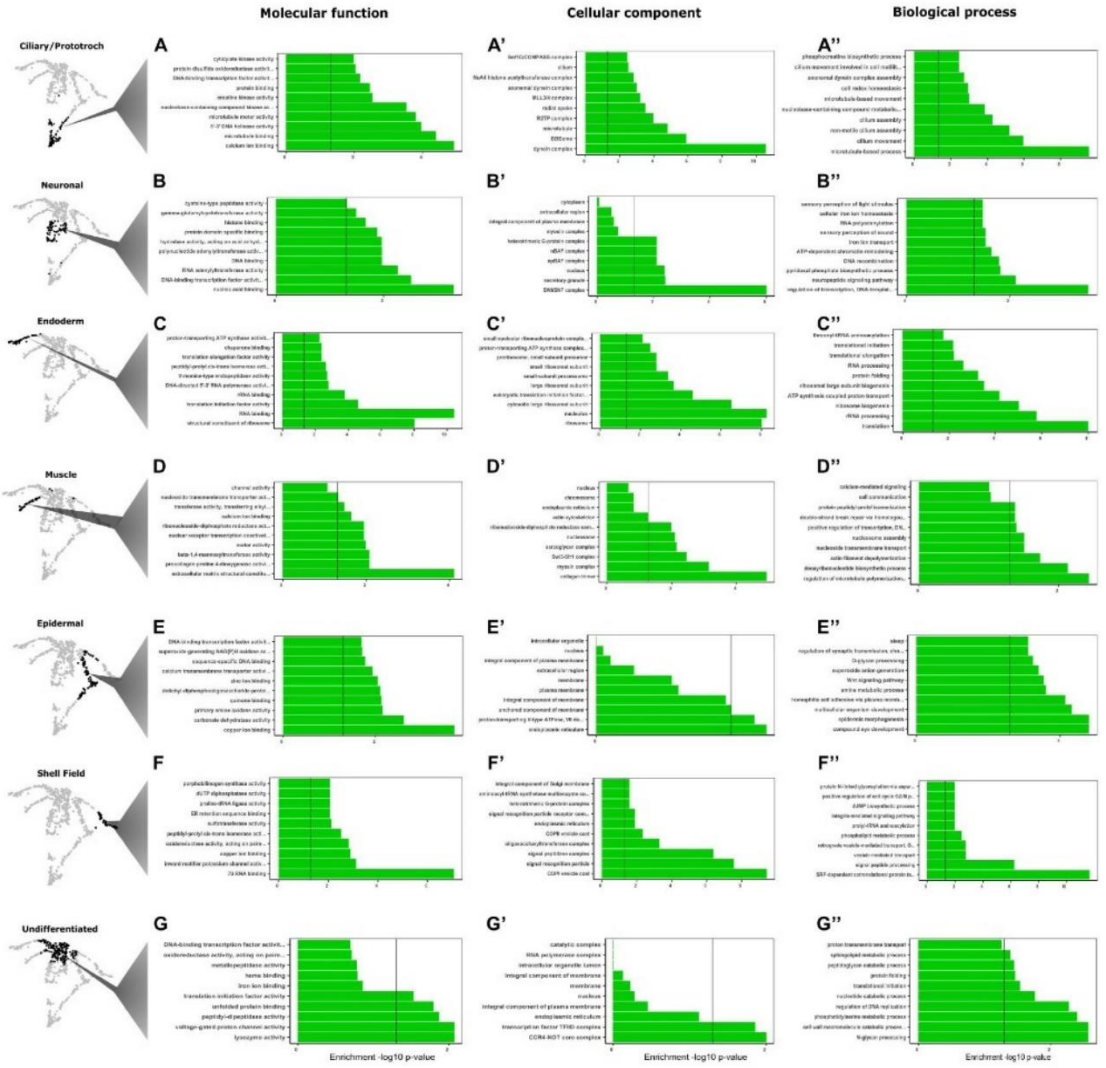


FIGURE 2 | Gene ontology (GO) analyses leads to identification and validation of clusters of differentially expressed genes. **(A–G")** GO terms associated with molecular function, cellular component, and biological processes. Classification of the corresponding GO terms form each highlighted cluster on the left side of the panel. Each row represents a cluster with significant transcriptomic signature and each column shows the categories of enrichment by molecular function, cellular component, and biological processes. Vertical black line (*p*-value = 0.05) in each graph represents statistical significance threshold set after performing a Fisher's exact test.

(**Supplementary Figure 4** and **Supplementary Tables 1–3**), is a typical feature of animal cells that have not yet undergone differentiation into developmental fates and is often observed in whole-organism single cell RNA-seq datasets (Hemmrich et al., 2012; Fincher et al., 2018; Siebert et al., 2019; Duruz et al., 2020). This cluster is composed of cells with low gene counts (151–1332), suggesting that these cells lack sufficient information to robustly cluster with similar cell populations and thus could represent a clustering artifact. In this case, this cell population would be expected to disappear with more stringent filtering (min. 500 detected genes), allowing for more robust clustering of the remaining cell populations. However, removing these low information cells does not improve clustering but rather shifts the cluster boundary, resulting in a cluster of cells with less information (**Supplementary Data 2**). Although this may represent a technical artifact, it could also be interpreted as a real biological signal at this stage of development wherein many of the embryonic cells are in a permissive cell state prior to activating cell differentiation pathways. In addition, results from StemID, an algorithm for predicting multipotent cell identities (Grün, 2020), suggest a statistically supported link of all clusters to these cells in the trochophore stage and indicate there is a starting point for differentiation trajectories in this cluster (**Supplementary Figure 3**). Altogether, *in silico* analyses support the notion that this cluster is composed of cells which have not yet developed a unique transcriptomic signature and thus the identified developmental trajectories are differentiated from this cluster.

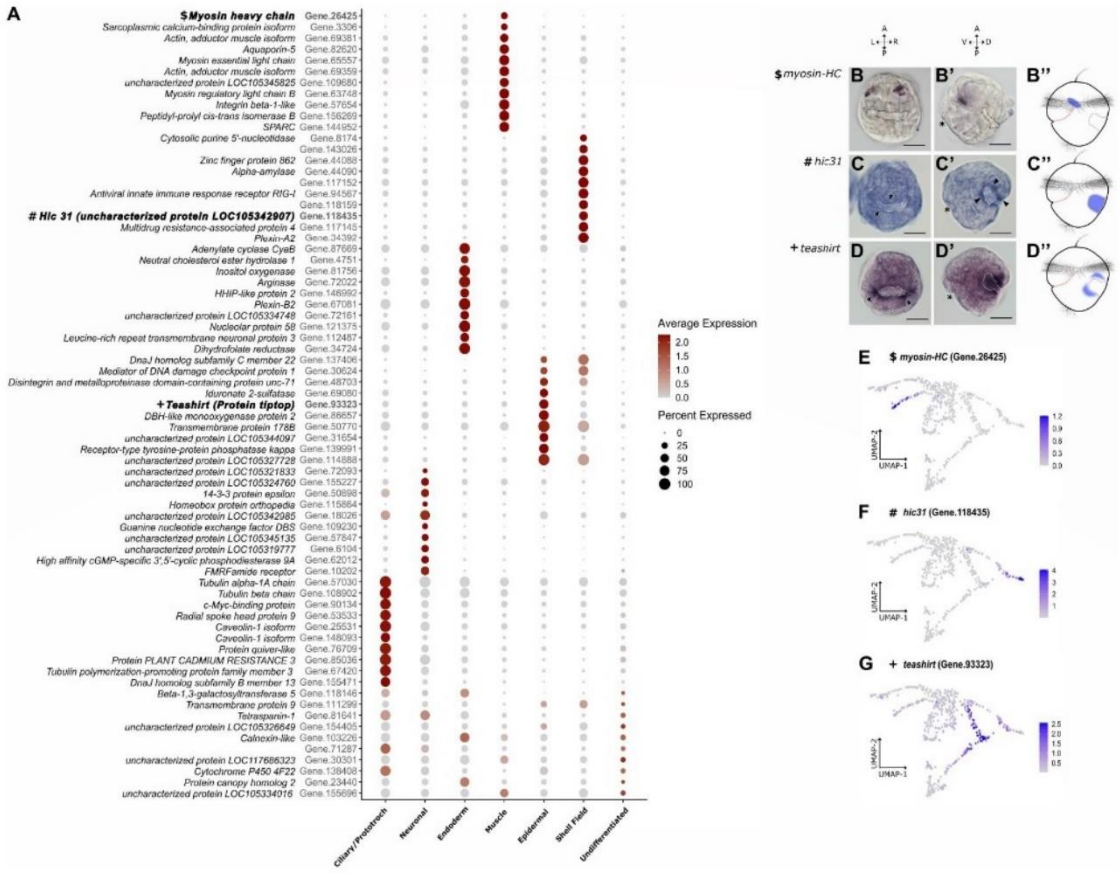


FIGURE 3 | Mapping cell state markers on the trochophore larva. **(A)** Top 10 highly expressed genes per cluster. On the left side of each gene is the putative ortholog match from *Crassostrea gigas*. Color intensity represents the average gene expression while circle size is the percentage of cells expressing a certain gene. In bold are the IDs of the genes used for cluster validation through *in situ* hybridization in **(B–D)**. **(B–D")** Expression domains of randomly sampled marker genes from a dorsal **(B–D)** and a lateral point of view **(B'–D')**. **(B–B")** *Mhc* expression located in cells from the anterior region between the developing digestive tract and the shell field. Dotted lines highlight the secreted shell field. **(C–C")** *Hic31* expression located in cells belonging to the outer margin of the shell field (arrowheads). **(D–D")** *Tsh* expression is found in the epidermal cells limiting the outline of the shell field. Asterisk indicates the mouth opening on the left side. Anterior (A), posterior (P), ventral (V), dorsal (D), left (L), and right (R). Scale bars are 20 μm. **(E–G)** Feature plots of each respective gene used for *in situ* hybridization and after running the gene imputation algorithm with MAGIC.

Pseudotime Trajectories Infer Relationships Between Cell Populations

We aimed to identify trajectories and reconstruct lineages from the cell clusters of *Dreissena rostriformis* trochophore larvae. While running the Monocle3 algorithm without any assumptions *a priori*, eight major trajectory ends could be recovered to model a differentiation tree with all cell populations and relate them to one single root, the "undifferentiated" cluster (**Figure 4B**). Two trajectory ends correspond to the clusters enriched with processes related to muscle development, endoderm development, and ribosomal activity (**Figures 2C–D′**, **3B**). "Muscle" and "endoderm" originate from a common cell lineage, the corresponding cell division can be observed on the node linking these two cell types (**Figure 4C**). The remaining trajectories show terminal ends in neuronal and ciliary clusters (**Figure 4B**). The presence of multiple differentiation pathways in these clusters supports further structure within the respective cell groups. There is one trajectory linking shell field and several epidermal cells (**Figure 4B**). This suggests a specification pathway that employs a subset of epithelial cells destined to contribute to the shell field. To assign putative terminal fates on the lineage tree of *Dreissena* and link them to our annotated clusters, we analyzed defined fate maps from other mollusks and made an estimation on the expected cell types per individual (**Figure 4C**). We then compared the number of expected clusters from the revised lineage tree of *Dreissena* to the actual number of cells obtained in our dataset (**Figure 4E**). Thereby, the "expected" cluster cells were counted from a developmental stage that was approximately one cell division cycle younger (67 cells) than that of the "outcome" cluster (110 cells). This was done to assess whether the same cell types described in classical cell lineage studies were also recovered by our single cell RNA-seq approach. While the proportions of expected versus outcome match well for the "neuronal," "ciliary," and "endoderm," the proportion of "shell field" and "muscle" cells show a higher proportion in the expected than in the

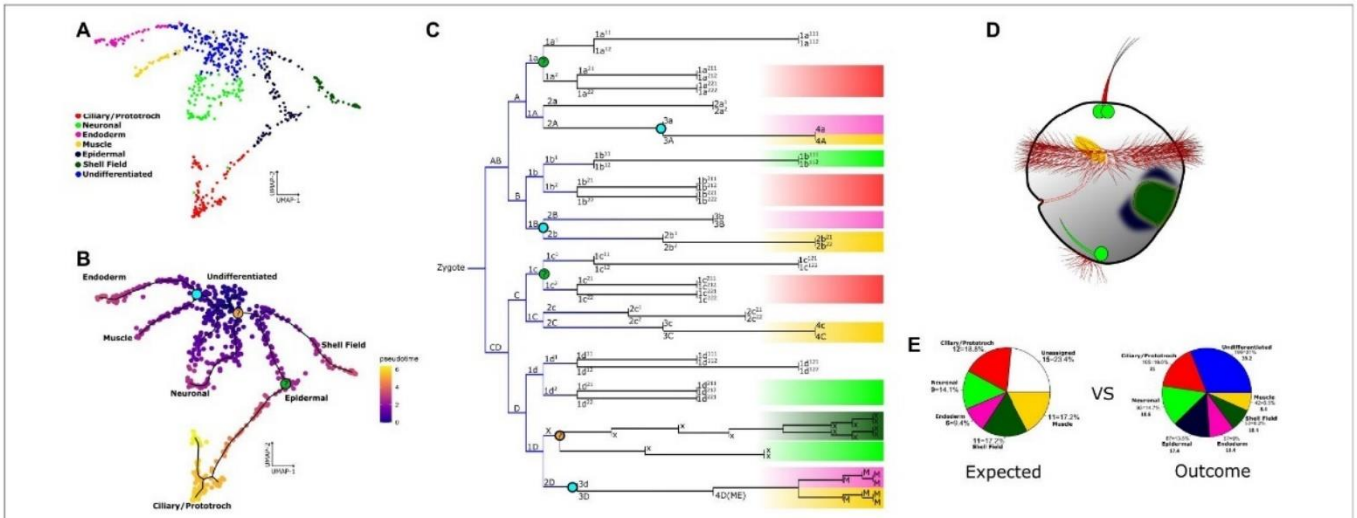


FIGURE 4 | Cell clusters mapped onto the *Dreissena* cell lineage tree and their location in the trochophore larva. **(A)** Resulting UMAP graph with cluster identification result of our annotation pipeline. Identified cell clusters are named and color-coded. **(B)** Calculated pseudotime showing recovered differentiation trajectories among extant clusters. Black lines represent the learned pseudotime trajectories. Colored nodes marked in the trajectory pathways and in the cell lineage tree represent the possible developmental decisions inferred from pseudotime and are drawn into the cell lineage map. The aquamarine node represents the developmental decision between endodermal and muscle cells. Light orange and light green represent the nodes that do not have a certain position on the fate map. **(C)** Cell lineage tree of *Dreissena* modified from Meisenheimer (1900) and Luetjens and Dorresteijn (1995). Highlighted are the putative terminal cell identities with the same color code from the UMAP projection. Blue gradient on the lines of early cell lineages represents the multipotency of the cells, which is lost in further cell divisions to acquire identity and could still be present in some lineages. Cell lineages not colored here mean the fate of these cell lineages was not identified in the fate map of *Dreissena*. **(D)** Putative location of clusters on the larva based on annotation methods described here. **(E)** Pie charts showing the percentage of cell types based on counts from cell lineage studies for one single larva (left, "expected") vs. the percentage of cell types based on counts of this study from five larvae (right, "outcome").

actual outcome analysis. This could be due to the annotation of previously unrecognized cell types on the lineage tree from *Dreissena* that might affect the cluster proportions of the "outcome" (**Figure 4E**). For example, concerning the shell field and epidermal cells, that both derive from the ectoderm, these might not yet have undergone fate specification in the analyzed developmental stage and thus might all fall in one category (here: shell field) in the "expected" analysis. The higher value for "muscle" cells in the "expected" chart relative to the "outcome" one might be because considerably fewer cells have undergone differentiation into muscle cells, a situation that is congruent with morphological analyses that showed a strikingly simple myoanatomy of *Dreissena* trochophores (Meisenheimer, 1900; Pavlicek et al., 2018, unpublished observations). Overall, the trajectories traced across the whole dataset position extant clusters uniformly along each respective path. This shows that developmental trajectories are well sampled in our dataset. In addition, it confirms that our cluster annotations recover the correct proportions of expected cell types from the lineage tree and that the pseudotime analysis is a good indicator of differentiation paths in the live larva.

DISCUSSION

Verification of Cluster Identity by Ortholog Comparison Across Metazoa

The trochophore larva of the quagga mussel is comprised of well-defined features such as a shell field, a ciliated prototroch, telotroch and apical tuft, a simply-built nervous system, and a developing digestive tract (Pavlicek et al., 2018; Salamanca-Díaz et al., 2021). Our single cell transcriptomic atlas of the *Dreissena* trochophore provides the first detailed analysis of the transcriptomic toolkit of a larval bivalve mollusk at single cell resolution, capturing cells corresponding to each of these morphological features (**Figures 4A,D**). Overall, our transcriptomic annotations in the quagga mussel *Dreissena rostriformis* confirm that key molecular factors expressed in all clusters are comparable to other species. As such, the cells identified as "ciliary/prototroch" are enriched in genes involved in ciliogenesis and cilia movement and the development of the dynein complex known to be a crucial part of locomotion in many metazoans (King, 2012). Ciliated cells found in the annelids *Platynereis dumerilii*, *Hydroides elegans*, and *Capitella teleta* similarly express orthologs of *Caveolin, Forkhead box J (foxJ)*, *rshp4*, *dynein heavy chain 9 (DNAH9)*, *tubulin polymerization-promoting family member 3*, and *tektin4* (**Supplementary Figure 4** and **Supplementary Tables 1–3**; Arenas-Mena et al., 2007; Achim et al., 2018; Sur and Meyer, 2021). Furthermore, genes expressed in ciliary cells of *Dreissena* are also present in motor cilia of more distantly related organisms such as the acoel *Isodiametra pulchra* (i.e., *Dynein heavy chain* and *tubulin* homologs) and in ciliary band cells of the sea urchin *Strongylocentrotus purpuratus* (e.g., *foxJ*) (Duruz et al., 2020; Paganos et al., 2021). This shows that the "ciliary/prototroch" cells in our data possess a distinct transcriptomic signature indicative of the ability to produce cilia, and that at least some of this molecular program was present in

the last common ancestor of bivalves, annelids, sea urchins, and xenacoelomorphs.

We identified cells that have ontology enrichment of neuroreceptors crucial for peptide signaling and factors related to mechanosensation and photosensation (**Figures 2B–B″**). Additionally, this cluster shows high expression of orthologs that play a role in the differentiation of the anlagen for the developing larval neurosecretory cells, e.g., *FMRFamide receptor, neuronal acetylcholine receptor subunit alpha-6* and *neuregulin 2*, in the bivalves *Macrocallista nimbosa*, *Mizuhopecten yessoensis*, and *Mytilus galloprovincialis* (**Supplementary Figure 4** and **Supplementary Tables 1–3**; Price and Greenberg, 1977; Wang et al., 2017; Gerdol et al., 2020). This indicates the presence of a neurosecretory center in this cluster that contains the key machinery for neuronal differentiation and suggests that the neuroanatomy of *Dreissena* trochophore larvae may be more complex than currently appreciated by morphological data using fluorescence-coupled antibody staining (Pavlicek et al., 2018).

Cells identified as "endoderm" expresses orthologs linked to protein breakdown processes and catalytic reactions which take place in the digestive system (**Figures 2C–C″**; Kirschke et al., 1995; Duruz et al., 2020). This is supported by the expression of crucial functional orthologs such as *hepatocyte nuclear factor 4 (hnf4), hematopoietically expressed homeobox gene* (*hhex*), and *galectin*. *Galectin* has been reported to promote phagocytosis of pathogens as a defense mechanism present in all cells and in hemocytes in adult stages, playing a role in innate immune responses and intracellular digestion in aquatic mollusks (Vasta et al., 2015). *Hnf4* is considered a marker of developing endoderm since its expression has been found in gut stem cells in several metazoans (Achim et al., 2018; Wendt et al., 2020; Paganos et al., 2021; Sur and Meyer, 2021) and was highly expressed exclusively in this cluster (in a limited number of cells only). *Hhex* is considered a regulator of hepatocyte development in deuterostomes (Wallace et al., 2001; Evseeva et al., 2021; Paganos et al., 2021) and was also found to be expressed in cells of this cluster (**Supplementary Figure 4** and **Supplementary Tables 1–3**). Although detection of these transcriptional markers is restricted to few cells, this is not uncommon for transcription factors, and overall the molecular fingerprint of this cluster is coherent with the developing gastrodermis.

Dreissena muscle progenitor cells consistently express common muscle and mesodermal markers. Among the gene orthologs identified in the "muscle" cluster, expression of *twist, Hand*, *Mox*, and *myosin heavy chain (mhc)* were detected (**Figures 3A,E** and **Supplementary Figure 4**). These genes are typically involved in mesoderm and/or muscle formation across metazoan clades (Castellani and Cohen, 1987; Ladurner and Rieger, 2000; Nederbragt et al., 2002; Dyachuk and Odintsova, 2009; Dobi et al., 2015; Achim et al., 2018; Li et al., 2019). In addition, there is gene ontology enrichment of several processes associated with muscle differentiation such as calcium ion binding and actin filament depolymerization (**Figures 2D–D″**).

The cluster "epidermal" shows enrichment of epidermis-related processes such as its morphogenesis and cell adhesion (**Figures 2E–E″**). Moreover, this cluster expresses orthologs of epidermal cells from the acoel *Isodiametra pulchra* and the annelid *Capitella teleta* that are associated with extracellular matrix factors and tight junctions such as *annexin*, cadherins, and claudin family members (**Supplementary Figure 4** and **Supplementary Tables 1–3**; Duruz et al., 2020; Sur and Meyer, 2021). Claudin is a transmembrane protein that is a crucial component of tight junctions in cells and cadherins are a type of cell adhesion molecules important for epithelium formation (Broders and Thiery, 1995; Krause et al., 2008; Piontek et al., 2008; Shapiro, 2016). For *annexin*, while playing a role in calcium-dependent membrane-binding and constituting an ectodermal marker in several metazoans (Meyer and Seaver, 2009; Duruz et al., 2020; Sur and Meyer, 2021), our results find this gene expressed also in ciliary and shell field cells (**Supplementary Figure 4** and **Supplementary Tables 1–3**). Hence, these shared transcriptomic signatures suggest that the precursor for both clusters have an ectodermal origin. Overall, the ortholog search supports the identity of this cluster by allowing one to one comparisons of cluster markers from different animals.

Orthologs of genes expressed in the shell field cluster that are known to play a role in early molluscan shell formation include *Hox1, chitin binding protein*, and *hic31* (**Supplementary Figure 4** and **Supplementary Tables 1–3**; Kakoi et al., 2008; Kin et al., 2009; Tan et al., 2017; Wollesen et al., 2018; Zhao et al., 2018; Liu et al., 2020; Salamanca-Díaz et al., 2021). Transcripts expressed here show ontology enrichment in processes such as peptide signaling and plasma membrane interactions, linking the activity of these cells to shell building mechanisms (**Figures 2F–F″**). However, most of the upregulated genes in this cluster do not have a known ontology term or an ortholog match with other metazoans (**Supplementary Tables 1–3**). Previous studies analyzing gene products expressed in shell secreting mantle tissues of various bivalves and gastropods found that, despite having a common biomineralization toolbox, the shell matrix expresses many novel, duplicated, and reorganized genes. These features have led to the conclusion that molluscan shells are "rapidly evolving" (Jackson et al., 2006; Aguilera et al., 2017; Zhao et al., 2018). This opens the possibility for further analyses in *Dreissena* shell field cells and to test whether these markers constitute hitherto unknown genes.

Enhancing the Cell Fate Map of *Dreissena*

Cell groups were linked together by tracing pseudotime trajectories across our dataset (**Figure 4B**). As these trajectories are represented in function of pseudotime, they can be used as a bioinformatical approximation of developmental decisions. By comparing terminal fates from other molluscan lineage trees to our results, we were able to link scRNA-seq data to the cellular lineage tree of *Dreissena* (Meisenheimer, 1900; **Figure 1B**). The "muscle" and "endoderm" clusters are separated here by a node in the scRNA-seq dataset (**Figure 4B**). According to cell fate maps from *Dreissena* and other mollusks, there is a significant portion of the endoderm that originates from mesodermal lineages

(Nielsen, 2005; Kurita et al., 2009; Chan and Lambert, 2011; Gharbiah et al., 2013; Lyons et al., 2017). Therefore, since these clusters have an origin in a common cell lineage, the node between these two groups may mark the corresponding cell division visualized in the cell fate map (**Figure 4C**). On the other hand, the clusters seen in the rest of the built trajectories, i.e., shell field and epidermal, neuronal and ciliary/prototroch (**Figures 4B,C**), have their origin in macromeres composed mainly of the ectodermal germ layer (Nielsen, 2005; Kurita et al., 2009; Chan and Lambert, 2011; Gharbiah et al., 2013; Lyons et al., 2017). Furthermore, after increasing cluster resolution and analyzing ortholog gene expression in both, ciliary and neuronal cells, there was occurrence of more than one subtype in each cluster (**Supplementary Figures 1I–K**). This was supported by the trajectory analyses that display multiple ends in both ciliary and neuronal cells (**Figure 4B**). We suggest that this diversity within a cluster could represent the different subtypes of ciliary and neuronal cells seen in the live larva, i.e., apical tuft, prototroch, telotroch, *FMFRamide receptor*-positive and *neuronal acetylcholine receptor subunit alpha-6*-positive cells, respectively (**Supplementary Figure 4** and **Supplementary Tables 1–3**). Similar cluster diversity in neuronal cell types is found in single cell datasets from other metazoans such as the annelid *Capitella teleta* and the sea urchin *Strongylocentrotus purpuratus*. All neuronal cell types of these species share a baseline transcriptomic signature, with only enough differences in expression to subdivide the overall cluster into different categories, e.g., 12 neuronal subtypes in the sea urchin and four neural subgroups in *C. teleta* (Achim et al., 2018; Foster et al., 2020; Paganos et al., 2021). Moreover, when analyzing the cell lineage tree of *Dreissena*, our data highlights that the underlying transcriptomic signature is similar across lineages that give rise to the same fate (**Figure 3C**).

CONCLUSION

Our analyses identified all expected major cell populations in the trochophore larva of the quagga mussel, *Dreissena rostriformis* (**Figure 4E**). When comparing our results to studies on other metazoans, we were able to support previous morphological annotations from cell groups such as "ciliary," "neuronal," "muscle," "endoderm," and "shell field." Our cluster validations, such as the BLASTX search, GO term annotations, and *in situ* hybridization identified marker genes for each of these cell types. Moreover, we were able to gain a first glimpse into developmental decisions, e.g., "endoderm" + "muscle," and cluster structure within already known cell lineages, e.g., "ciliary/prototroch" and "neuronal." Our data demonstrate that the levels of differentially expressed genes can be used as markers to identify clusters of cells in early developmental (larval) stages of (bivalve) mollusks, thus adding these animals to the suite of taxa that can now be accessed for comparative evolutionary studies on cell type level using scRNA-seq techniques. This should ultimately lead to a better understanding as to how animals form and which genes and cell types are involved in shaping larval and adult life cycle stages across Metazoa.

DATA AVAILABILITY STATEMENT

The data presented in this study are deposited in NCBI's Gene Expression Omnibus and are accessible through the GEO series accession number GSE192624 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE192624).

AUTHOR CONTRIBUTIONS

DS-D, AC, and AW designed the research and contributed to interpretation of data. DS-D performed the experiments, generated data with contribution of AC, performed the data analysis with assistance of AC, and drafted the manuscript with input from AW and AC. AC generated the sequencing library from cell suspensions provided by DS-D. SS contributed to gene annotation from the muscle cluster and performed *in situ* hybridization of *myosin heavy chain*. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by a completion grant of the Vienna Doctoral School of Ecology and Evolution (VDSEE) to DS-D and by the Austrian Science Fund (FWF) (grant P29455-B29 to AW and P31018-B29 to AC).

ACKNOWLEDGMENTS

The authors thank Julia Steger and Oliver Link (Vienna) for their support and valuable input when developing the dissociation suspension and assisting the first experimental trials, and Hannah Schmidbaur and Elena A. Ritschard (Vienna) for bioinformatics advice on the gene annotations. The authors also thank Uli Technau for hosting DS-D during this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.783984/full#supplementary-material

Supplementary Figure 1 | (A) Barcode plot showing the cumulative distribution of UMIs in function of barcodes. Violin plots show feature distribution for the overall data before **(B)** and after **(C)** setting thresholds to the data, and in every cluster **(D)** from the dataset. **(E)** Scatter plot showing the correlation between features, Pearson correlation is shown on top of the plot. **(F)** Top 2000 most variably expressed genes used for further calculation of significant principal components. Cells were filtered out based on less than 150 and more than 4000 features. **(G)** Elbow plot showing the standard deviations of the calculated principal components. **(H)** Statistics table on the single cell transcriptome libraries showing sequencing, mapping, and cell metrics summary for this study. **(I–K)**

UMAP graphs result of an increased cluster-finding algorithm resolution with 0.7, 1, and 2 as resolution threshold, respectively.

Supplementary Figure 2 | Dataset exploration on different filtering thresholds. **(A,C,D,G,H,K)** Represent the data after filtering out cells with less than 300 and more than 4000 genes, resulting in 490 cells. **(B,E,F,I,J,L)** Represent the data after filtering out cells with less than 500 and more than 4000 genes, resulting in 430 cells. **(A,B)** Violin plots showing feature and count distribution for the filtered data, with more than 300 and 500 genes, respectively. **(C,E)** Original clustering highlighted in the filtered data, with more than 300 and 500 genes, respectively. **(D–J)** UMAP graphs result of an increased cluster-finding algorithm resolution in filter settings of more than 300 and 500 genes, with 0.5, 1, and 5 as resolution threshold, respectively. **(K,L)** Pseudotime trajectory calculations for each respective filtered data threshold.

Supplementary Figure 3 | Prediction of a stem cell-like cluster at the root of the differentiation pathways based on the initial clustering. The entropy values are shown as color of the vertices. The color of the link represents the $-\log_{10}$ p-value. Thickness of the link indicates the score, reflecting cell density coverage on a link.

Supplementary Figure 4 | (A) Orthologs known to be stem cell markers identified in our dataset, expressed across all clusters. **(B)** Dot plot of ortholog gene expression per cluster. **(C)** Orthologs identified in the dataset used to annotate cluster identities. **(D)** Dot plot of ortholog gene expression per cluster. Color intensity represents the average gene expression while circle size is the percentage of cells expressing a certain gene.

Supplementary Figure 5 | Negative controls (sense riboprobes) of *in situ* hybridization experiments for each randomly sampled marker gene used for cluster annotation. **(A–C)** Lateral views of *mhc, hic31,* and *tsh,* respectively. Anterior (A), posterior (P), ventral (V), and dorsal (D). Scale bars are 20 μm. Dotted lines indicate the outline of the shell field. Asterisk indicates the mouth opening on the left side.

Supplementary Table 1 | Gene annotations of the transcriptome used for this dataset. Annotations from Pfam, PANTHER, GO terms, and BLASTX hit from human and *Crassostrea gigas* orthologs are indicated in each column accordingly.

Supplementary Table 2 | Differentially expressed markers per cluster result from the output FindAllMarkers function with a cluster threshold set up to 0.5.

Supplementary Table 3 | Top 10 differentially expressed markers per cluster result from the output FindAllMarkers function with a cluster threshold set up to 0.7.

Supplementary Table 4 | List of primers and nucleotide sequence for each gene used for cloning. Forward and reverse primers used for amplifying each gene by PCR are indicated.

Supplementary Table 5 | Explored cutoffs in the dataset and the number of remaining cells of each cluster after each filtering. Thresholds were set from more than 100 genes and less than 4000 to more than 500 genes and less than 2500, respectively. Dotted lines indicate the outline of the shell field and the symbol * indicates the mouth opening on the left side.

REFERENCES

Achim, K., Eling, N., Vergara, H. M., Bertucci, P. Y., Musser, J., Vopalensky, P., et al. (2018). Whole-body single-cell sequencing reveals transcriptional domains in the annelid larval body. *Mol. Biol. Evol.* 35, 1047–1062. doi: 10.1093/molbev/msx336

Achim, K., Pettit, J.-B., Saraiva, L. R., Gavriouchkina, D., Larsson, T., Arendt, D., et al. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* 33, 503–509.

Aguilera, F., McDougall, C., Degnan, B. M., and Irwin, D. (2017). Co-option and de novo gene evolution underlie molluscan shell diversity. *Mol. Biol. Evol.* 34, 779–792. doi: 10.1093/molbev/msw294

Aldridge, D. C., Ho, S., and Froufe, E. (2014). The ponto-caspian quagga mussel, *Dreissena rostriformis* bugensis (Andrusov, 1897), invades Great Britain. *Aquat. Invasions* 9, 529–535.

Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22, 1600–1607. doi: 10.1093/bioinformatics/btl140

Arenas-Mena, C., Wong, K. S.-Y., and Arandi-Forosani, N. (2007). Ciliary band gene expression patterns in the embryo and trochophore larva of an indirectly developing polychaete. *Gene Expr. Patterns* 7, 544–549. doi: 10.1016/j.modgep.2007.01.007

Arendt, D., Musser, J. M., Baker, C. V. H., Bergman, A., Cepko, C., Erwin, D. H., et al. (2016). The origin and evolution of cell types. *Nat. Rev. Genet.* 17, 744–757. doi: 10.1038/nrg.2016.127

Briggs, J. A., Weinreb, C., Wagner, D. E., Megason, S., Peshkin, L., Kirschner, M. W., et al. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* 360:eaar5780. doi: 10.1126/science.aar5780

Broders, F., and Thiery, J. P. (1995). "Structure and function of cadherins BT," in *Organization of the Early Vertebrate Embryo*, eds N. Zagris, A. M. Duprat, and A. Durston (Berlin: Springer), 183–208. doi: 10.1007/978-1-4899-1618-1_16

Calcino, A. D., de Oliveira, A. L., Simakov, O., Schwaha, T., Zieger, E., Wollesen, T., et al. (2019). The quagga mussel genome and the evolution of freshwater tolerance. *DNA Res.* 26, 411–422. doi: 10.1093/dnares/dsz019

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566, 496–502. doi: 10.1038/s41586-019-0969-x

Castellani, L., and Cohen, C. (1987). Myosin rod phosphorylation and the catch state of molluscan-muscles. *Science (New York, N.Y.)* 235, 334–337. doi: 10.1126/science.3026049

Chan, X. Y., and Lambert, J. (2011). Patterning a spiralian embryo: a segregated RNA for a Tis11 ortholog is required in the 3a and 3b cells of the *Ilyanassa embryo*. *Dev. Biol.* 349, 102–112. doi: 10.1016/j.ydbio.2010.10.001

Dictus, W. J. A. G., and Damen, P. (1997). Cell-lineage and clonal-contribution map of the trochophore larva of *Patella vulgata* (Mollusca)1Both authors contributed equally to this work.1. *Mech. Dev.* 62, 213–226. doi: 10.1016/S0925-4773(97)00666-7

Dobi, K. C., Schulman, V. K., and Baylies, M. K. (2015). Specification of the somatic musculature in *Drosophila*. *Wires Dev. Biol.* 4, 357–375. doi: 10.1002/wdev.182

Duruz, J., Kaltenrieder, C., Ladurner, P., Bruggmann, R., Martinez, P., and Sprecher, S. G. (2020). Acoel single-cell transcriptomics: cell-type analysis of a deep branching bilaterian. *BioRxiv* [Preprint]. doi: 10.1101/2020.07.10.196782 BioRxiv 2020.07.10.196782,

Dyachuk, V., and Odintsova, N. (2009). Development of the larval muscle system in the mussel *Mytilus trossulus* (Mollusca, Bivalvia): original article. *Dev. Growth Differ.* 51, 69–79. doi: 10.1111/j.1440-169X.2008.01081.x

Evseeva, M. N., Dyikanov, D. T., Karagyaur, M. N., Prikazchikova, T. A., Sheptulina, A. F., Balashova, M. S., et al. (2021). Hematopoietically-expressed homeobox protein HHEX regulates adipogenesis in preadipocytes. *Biochimie* 185, 68–77. doi: 10.1016/j.biochi.2021.02.011

Farrell, J. A., Wang, Y., Riesenfeld, S. J., Shekhar, K., Regev, A., and Schier, A. F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* 360:eaar3131. doi: 10.1126/science.aar3131

Fincher, C. T., Wurtzel, O., de Hoog, T., Kravarik, K. M., and Reddien, P. W. (2018). Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science* 360:eaaq1736. doi: 10.1126/science.aaq1736

Foster, S., Oulhen, N., and Wessel, G. (2020). A single cell RNA sequencing resource for early sea urchin development. *Development* 147:dev191528. doi: 10.1242/dev.191528

Foster, S., Teo, Y. V., Neretti, N., Oulhen, N., and Wessel, G. M. (2019). Single cell RNA-seq in the sea urchin embryo show marked cell-type specificity in the Delta/Notch pathway. *Mol. Reproduct. Dev.* 86, 931–934. doi: 10.1002/mrd.23181

García-Castro, H., Kenny, N. J., Iglesias, M., Álvarez-Campos, P., Mason, V., Elek, A., et al. (2021). ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics. *Genome Biol.* 22:89. doi: 10.1186/s13059-021-02302-5

Gerdol, M., Moreira, R., Cruz, F., Gómez-Garrido, J., Vlasova, A., Rosani, U., et al. (2020). Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biol.* 21:275. doi: 10.1186/s13059-020-02180-3

Gharbiah, M., Nakamoto, A., and Nagy, L. M. (2013). Analysis of ciliary band formation in the mollusc *Ilyanassa obsoleta*. *Dev. Genes Evol.* 223, 225–235. doi: 10.1007/s00427-013-0440-1

Grün, D. (2020). Revealing dynamics of gene expression variability in cell state space. *Nat. Methods* 17, 45–49. doi: 10.1038/s41592-019-0632-3

Grün, D., Muraro, M. J., Boisset, J.-C., Wiebrands, K., Lyubimova, A., Dharmadhikari, G., et al. (2016). De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* 19, 266–277. doi: 10.1016/j.stem.2016.05.010

Heiler, K. C. M., Bij de Vaate, A., Ekschmitt, K., von Oheimb, P. V., Albrecht, C., and Wilke, T. (2013). Reconstruction of the early invasion history of the quagga mussel (Dreissena rostriformis bugensis) in Western Europe. *Aquat. Invasions* 8, 53–57.

Hejnol, A., Martindale, M. Q., and Henry, J. Q. (2007). High-resolution fate map of the snail *Crepidula fornicata*: The origins of ciliary bands, nervous system, and muscular elements. *Dev. Biol.* 305, 63–76. doi: 10.1016/j.ydbio.2007.01.044

Hemmrich, G., Khalturin, K., Boehm, A.-M., Puchert, M., Anton-Erxleben, F., Wittlieb, J., et al. (2012). Molecular signatures of the three stem cell lineages in hydra and the emergence of stem cell function at the base of multicellularity. *Mol. Biol. Evol.* 29, 3267–3280. doi: 10.1093/molbev/mss134

Henry, J. Q., Okusu, A., and Martindale, M. Q. (2004). The cell lineage of the polyplacophoran, *Chaetopleura apiculata*: variation in the spiralian program and implications for molluscan evolution. *Deve. Biol.* 272, 145–160. doi: 10.1016/j.ydbio.2004.04.027

Hinman, V. F., O'Brien, E. K., Richards, G. S., and Degnan, B. M. (2003). Expression of anterior Hox genes during larval development of the gastropod *Haliotis asinina*. *Evol. Dev.* 5, 508–521.

Inkscape Project (2020). *Inkscape*. Available online at: https://inkscape.org (accessed March 3, 2021).

Jackson, D. J., McDougall, C., Green, K., Simpson, F., Wörheide, G., and Degnan, B. M. (2006). A rapidly evolving secretome builds and patterns a sea shell. *BMC Biol.* 4:40. doi: 10.1186/1741-7007-4-40

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kakoi, S., Kin, K., Miyazaki, K., and Wada, H. (2008). Early development of the japanese spiny oyster (*Saccostrea kegaki*): characterization of some genetic markers. *Zool. Sci.* 25, 455–464. doi: 10.2108/zsj.25.455

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., et al. (2017). The *Drosophila embryo* at single-cell transcriptome resolution. *Science* 358, 194–199. doi: 10.1126/science.aan3235

Karatayev, A. Y., Burlakova, L. E., Pennuto, C., Ciborowski, J., Karatayev, V. A., Juette, P., et al. (2014). Twenty five years of changes in *Dreissena* spp. populations in Lake Erie. *J. Great Lakes Res.* 40, 550–559.

Karatayev, A. Y., Padilla, D. K., Minchin, D., Boltovskoy, D., and Burlakova, L. E. (2007). Changes in global economies and trade: the potential spread of exotic freshwater bivalves. *Biol. Invasions* 9, 161–180. doi: 10.1007/s10530-006-9013-9

Kin, K., Kakoi, S., and Wada, H. (2009). A novel role for dpp in the shaping of bivalve shells revealed in a conserved molluscan developmental program. *Dev. Biol.* 329, 152–166. doi: 10.1016/j.ydbio.2009.01.021

King, S. M. (2012). Integrated control of axonemal dynein AAA+ motors. *J. Struct. Biol.* 179, 222–228. doi: 10.1016/j.jsb.2012.02.013

Kirschke, H., Barrett, A. J., and Rawlings, N. D. (1995). Proteinases 1: lysosomal cysteine proteinases. *Protein Profile* 2, 1581–1643.

Krause, G., Winkler, L., Mueller, S. L., Haseloff, R. F., Piontek, J., and Blasig, I. E. (2008). Structure and function of claudins. *Biochim. Biophys. Acta* 1778, 631–645. doi: 10.1016/j.bbamem.2007.10.018

Kurita, Y., Deguchi, R., and Wada, H. (2009). Early development and cleavage pattern of the Japanese Purple Mussel, *Septifer virgatus*. *Zool. Sci.* 26, 814–820. doi: 10.2108/zsj.26.814

Ladurner, P., and Rieger, R. (2000). Embryonic muscle development of *Convoluta pulchra* (Turbellaria–Acoelomorpha, Platyhelminthes). *Dev. Biol.* 222, 359–375.

Lee, P. N., Callaerts, P., De Couet, H. G., and Martindale, M. Q. (2003). Cephalopod Hox genes and the origin of morphological novelties. *Nature* 424, 1061–1065. doi: 10.1038/nature01872

Levin, M., Anavy, L., Cole, A. G., Winter, E., Mostov, N., Khair, S., et al. (2016). The mid-developmental transition and the evolution of animal body plans. *Nature* 531, 637–641. doi: 10.1038/nature16994

Li, H., Li, Q., Yu, H., and Du, S. (2019). Developmental dynamics of myogenesis in Pacific oyster *Crassostrea gigas*. *Comp. Biochem. Physiol. Part B* 227, 21–30.

Lillie, F. R. (1895). The embryology of the unionidae. A study in cell-lineage. *J. Morphol.* 10, 1–100. doi: 10.1002/jmor.1050100102

Liu, G., Huan, P., and Liu, B. (2020). Identification of three cell populations from the shell gland of a bivalve mollusc. *Dev. Genes Evol.* 230, 39–45. doi: 10.1007/s00427-020-00646-9

Liu, X., Jin, C., Li, H., Bai, Z., and Li, J. (2018). Morphological structure of shell and expression patterns of five matrix protein genes during the shell regeneration process in *Hyriopsis cumingii*. *Aquacult. Fish.* 3, 225–231. doi: 10.1016/j.aaf.2018.09.005

Liu, X., Zeng, S., Dong, S., Jin, C., and Li, J. (2015). A novel matrix protein hic31 from the prismatic layer of hyriopsis cumingii displays a collagen-like structure. *PLoS One* 10:e0135123. doi: 10.1371/journal.pone.0135123

Luetjens, C. M., and Dorresteijn, A. W. C. (1995). Multiple, alternative cleavage patterns precede uniform larval morphology during normal development of *Dreissena polymorpha* (Mollusca, Lamellibranchia). *Rouxs Arch. Dev. Biol.* 205, 138–149. doi: 10.1007/BF00357760

Lyons, D. C., Perry, K. J., and Henry, J. Q. (2017). Morphogenesis along the animal-vegetal axis: fates of primary quartet micromere daughters in the gastropod *Crepidula fornicata*. *BMC Evol. Biol.* 17:217. doi: 10.1186/s12862-017-1057-1

McCartney, M. A., Auch, B., Kono, T., Mallez, S., Zhang, Y., Obille, A., et al. (2019). The genome of the Zebra Mussel, *Dreissena polymorpha*: a resource for invasive species research. *BioRxiv* [Preprint]. doi: 10.1101/696732 BioRxiv 696732,

McNickle, G. G., Rennie, M. D., and Sprules, W. G. (2006). Changes in benthic invertebrate communities of South Bay, Lake Huron following invasion by zebra mussels (*Dreissena polymorpha*), and potential effects on lake whitefish (*Coregonus clupeaformis*) diet and growth. *J. Great Lakes Res.* 32, 180–193.

Meisenheimer, J. (1900). *Entwicklungsgeschichte von Dreissensia polymorpha Pall.* Leipzig: Wilhelm Engelmann.

Meyer, N. P., and Seaver, E. C. (2009). Neurogenesis in an annelid: Characterization of brain neural precursors in the polychaete *Capitella* sp. *Dev. Biol.* 335, 237–252. doi: 10.1016/j.ydbio.2009.06.017

Mills, E. L., Dermott, R. M., Roseman, E. F., Dustin, D., Mellina, E., Conn, D. B., et al. (1993). Colonization, ecology, and population structure of the "quagga"mussel (Bivalvia: Dreissenidae) in the lower Great Lakes. *Can. J. Fish. Aquat. Sci.* 50, 2305–2314.

Nalepa, T. F., and Schloesser, D. W. (2019). *Quagga and Zebra Mussels: Biology, Impacts, and Control*. Boca Raton, FL: CRC Press.

Nederbragt, A. J., Lespinet, O., Van Wageningen, S., Van Loon, A. E., Adoutte, A., and Dictus, W. J. A. G. (2002). A lophotrochozoan twist gene is expressed in the ectomesoderm of the gastropod mollusk *Patella vulgata*. *Evol. Dev.* 4, 334–343. doi: 10.1046/j.1525-142X.2002.02020.x

Nielsen, C. (2005). Trochophora larvae: cell-lineages, ciliary bands and body regions. 2. Other groups and general discussion. *J. Exp. Zool. Part B* 304, 401–447. doi: 10.1002/jez.b.21050

Nielsen, C. (2018). Origin of the trochophora larva. *R. Soc. Open Sci.* 5:180042. doi: 10.1098/rsos.180042

Paganos, P., Voronov, D., Musser, J., Arendt, D., and Arnone, M. I. (2021). Single cell RNA sequencing of the *Strongylocentrotus purpuratus* larva reveals the blueprint of major cell types and nervous system of a non-chordate deuterostome. *Elife* 10:e70416. doi: 10.7554/eLife.70416

Pavlicek, A., Schwaha, T., and Wanninger, A. (2018). Towards a ground pattern reconstruction of bivalve nervous systems: neurogenesis in the zebra mussel *Dreissena polymorpha*. *Organ. Divers. Evol.* 18, 101–114. doi: 10.1007/s13127-017-0356-0

Piontek, J., Winkler, L., Wolburg, H., Müller, S. L., Zuleger, N., Piehl, C., et al. (2008). Formation of tight junction: determinants of homophilic interaction between classic claudins. *FASEB J.* 22, 146–158. doi: 10.1096/fj.07-8319com

Plass, M., Solana, J., Wolf, F. A., Ayoub, S., Misios, A., Glažar, P., et al. (2018). Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* 360:eaaq1723. doi: 10.1126/science.aaq1723

Price, D. A., and Greenberg, M. J. (1977). Purification and characterization of a cardioexcitatory neuropeptide from the central ganglia of a bivalve mollusc. *Prep. Biochem.* 7, 261–281. doi: 10.1080/00327487708061643

R Developement Core Team (2015). *R: A Language and Environment for Statistical Computing*, Vol. 1. Vienna: R Foundation for Statistical Computing, 409. doi: 10.1007/978-3-540-74686-7

Raikow, D. F. (2004). Food web interactions between larval bluegill (*Lepomis macrochirus*) and exotic zebra mussels (*Dreissena polymorpha*). *Can. J. Fish. Aquat. Sci.* 61, 497–504.

Redl, E., Scherholz, M., Wollesen, T., Todt, C., and Wanninger, A. (2016). Cell proliferation pattern and twist expression in an aplacophoran mollusk argue against segmented ancestry of mollusca. *J. Exp. Zool. Part B* 326, 422–436. doi: 10.1002/jez.b.22714

Render, J. (1997). Cell Fate Maps in theIlyanassa obsoleta Embryo beyond the Third Division. *Dev. Biol.* 189, 301–310. doi: 10.1006/dbio.1997.8654

Salamanca-Díaz, D. A., Calcino, A. D., de Oliveira, A. L., and Wanninger, A. (2021). Non-collinear Hox gene expression in bivalves and the evolution of morphological novelties in mollusks. *Sci. Rep.* 11:3575. doi: 10.1038/s41598-021-82122-6

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502. doi: 10.1038/nbt.3192

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019

Sebé-Pedrós, A., Saudemont, B., Chomsky, E., Plessier, F., Mailhé, M.-P., Renno, J., et al. (2018). Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-seq. *Cell* 173, 1520–1534.e20. doi: 10.1016/j.cell.2018.05.019

Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* 14, 618–630. doi: 10.1038/nrg3542

Shapiro, L. (2016). "Structure and function of cadherin extracellular regions BT," in *The Cadherin Superfamily: Key Regulators of Animal Development and Physiology*, eds S. T. Suzuki and S. Hirano (Berlin: Springer), 71–91. doi: 10.1007/978-4-431-56033-3_4

Siebert, S., Farrell, J. A., Cazet, J. F., Abeykoon, Y., Primack, A. S., Schnitzler, C. E., et al. (2019). Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science* 365:eaav9314. doi: 10.1126/science.aav9314

Smith, S. A., Wilson, N. G., Goetz, F. E., Feehery, C., Andrade, S. C. S., Rouse, G. W., et al. (2011). Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* 480, 364–367. doi: 10.1038/nature10526

Stegle, O., Teichmann, S. A., and Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 16, 133–145. doi: 10.1038/nrg3833

Strayer, D. L., Hattala, K. A., and Kahnle, A. W. (2004). Effects of an invasive bivalve (*Dreissena polymorpha*) on fish in the Hudson River estuary. *Can. J. Fish. Aquat. Sci.* 61, 924–941.

Sur, A., and Meyer, N. P. (2021). Resolving transcriptional states and predicting lineages in the annelid capitella teleta using single-Cell RNAseq. *Front. Ecol. Evol.* 8:523.

Svensson, V., Vento-Tormo, R., and Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* 13, 599–604.

Tan, S., Huan, P., and Liu, B. (2017). Expression patterns indicate that BMP2/4 and Chordin, not BMP5-8 and Gremlin, mediate dorsal–ventral patterning in the mollusk *Crassostrea gigas*. *Dev. Genes Evol.* 227, 75–84. doi: 10.1007/s00427-016-0570-3

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., et al. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* 6, 468–478.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. *Genome Res.* 25, 1491–1498.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32:381.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. *Cell* 174, 716–729.e27. doi: 10.1016/j.cell.2018.05.061

Vasta, G. R., Feng, C., Bianchet, M. A., Bachvaroff, T. R., and Tasumi, S. (2015). Structural, functional, and evolutionary aspects of galectins in aquatic mollusks: From a sweet tooth to the Trojan horse. *Fish Shellfish Immunol.* 46, 94–106. doi: 10.1016/j.fsi.2015.05.012

Vergara, H. M., Bertucci, P. Y., Hantz, P., Tosches, M. A., Achim, K., Vopalensky, P., et al. (2017). Whole-organism cellular gene-expression atlas reveals conserved cell types in the ventral nerve cord of Platynereis dumerilii. *Proc. Natl. Acad. Sci. U.S.A.* 114, 5878–5885.

Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G., and Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* 360, 981–987. doi: 10.1126/science.aar4362

Wakida-Kusunoki, A. T., Wakida, F. T., and De Leon-Sandoval, J. M. (2015). First record of quagga mussel *Dreissena rostriformis* bugensis (Andrusov, 1897) (Bivalvia, Dreissenidae) from Mexico. *BioInvasions Rec.* 4, 31–36.

Wallace, K. N., Yusuff, S., Sonntag, J. M., Chin, A. J., and Pack, M. (2001). Zebrafish hhex regulates liver development and digestive organ chirality. *Genesis* 30, 141–143. doi: 10.1002/gene.1050

Wang, S., Zhang, J., Jiao, W., Li, J., Xun, X., Sun, Y., et al. (2017). Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat. Ecol. Evol.* 1:0120. doi: 10.1038/s41559-017-0120

Wanninger, A., and Wollesen, T. (2015). "Evolutionary developmental biology of invertebrates 2: lophotrochozoa (Spiralia)," in *Evolutionary Developmental Biology of Invertebrates 2: Lophotrochozoa (Spiralia)*, ed. A. Wanninger (Berlin: Springer-Verlag), 103–154. doi: 10.1007/978-3-7091-1871-9

Wanninger, A., and Wollesen, T. (2019). The evolution of molluscs. *Biol. Rev.* 94, 102–115. doi: 10.1111/brv.12439

Wendt, G., Zhao, L., Chen, R., Liu, C., O'Donoghue, A. J., Caffrey, C. R., et al. (2020). A single-cell RNA-seq atlas of andlt;emandgt;Schistosoma mansoniandlt;/emandgt; identifies a key regulator of blood feeding. *Science* 369, 1644–1649. doi: 10.1126/science.abb7709

Wollesen, T., Rodriguez Monje, S. V., de Oliveira, A. L., and Wanninger, A. (2018). Staggered Hox expression is more widespread among molluscs than previously appreciated. *Proc. R. Soc. B* 285:20181513. doi: 10.1098/rspb.2018.1513

Wollesen, T., Rodríguez Monje, S. V., McDougall, C., Degnan, B. M., and Wanninger, A. (2015a). The ParaHox gene Gsx patterns the apical organ and central nervous system but not the foregut in scaphopod and cephalopod mollusks. *EvoDevo* 6:41. doi: 10.1186/s13227-015-0037-z

Wollesen, T., Rodríguez Monje, S. V., Todt, C., Degnan, B. M., and Wanninger, A. (2015b). Ancestral role of Pax2/5/8 in molluscan brain and multimodal sensory system development. *BMC Evol. Biol.* 15:231. doi: 10.1186/s12862-015-0505-z

Zhao, R., Takeuchi, T., Luo, Y.-J., Ishikawa, A., Kobayashi, T., Koyanagi, R., et al. (2018). Dual gene repertoires for larval and adult shells reveal molecules essential for molluscan shell formation. *Mol. Biol. Evol.* 35, 2751–2761. doi: 10.1093/molbev/msy172

Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., et al. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* 555, 524–528.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

2.3. Chapter 3. Comparative single-cell transcriptomics reveals novel genes involved in bivalve embryonic shell formation and questions ontogenetic homology of molluscan shell types (submitted to *Frontiers in Cell and Developmental Biology*)

Salamanca-Díaz, D.A., Ritschard, E.A., Schmidbaur, H., Wanninger, A.

During the cluster analyses and annotations performed in Chapter 2, one of the clusters isolated resulted in very few matches for gene ontology. Additionally, cells belonging to this cluster express numerous short-numbered-kilobases genes, which match with previously uncharacterized molluscan genes. Accordingly, we analyzed the gene set that corresponds to protoconch I formation in the trochophore stage of the quagga mussel *Dreissena rostriformis*. We included annotations from shell field cells as identified in Chapter 2 and mixed them with ortholog analyzes applied on the previously assembled genome of *D. rostriformis*. Results showed that 24% (86) of the genes expressed in the shell field of the trochophore of *D. rostriformis* could not be assigned to any known orthogroup. At the same time, expression dynamics of these genes indicate fundamental differences in the gene repertoires present during protoconch I and protoconch II stages. This result points towards an independent evolutionary origin of these shell types and argues against ontogenetic homology. Moreover, there is a fraction of genes from the shell field of the *Dreissena* trochophore that are shared with the rest of the sampled metazoans (68 genes in 61 orthogroups, 19%). This set comprehends gene families previously described in secretomes from other bivalves and gastropods that originated before the emergence of the conchiferan clade (Aguilera et al., 2017; Kocot et al., 2016). This chapter demonstrates de existence in *Dreissena rostriformis* of both a mosaic of cooption

of known metazoan genes, and de novo recruitment of genes with hitherto unknown function or ortholog match. Altogether, these results confirm the high plasticity in the gene repertoire that underlie the ontogeny of adult and embryonic bivalve shells.

Comparative single-cell transcriptomics reveals novel genes involved in bivalve embryonic shell formation and questions ontogenetic homology of molluscan shell types

Authors: David A. Salamanca-Díaz[1], Elena A. Ritschard[2], Hannah Schmidbaur[2], Andreas Wanninger[1*]

[1]University of Vienna, Department of Evolutionary Biology, Unit for Integrative Zoology, Djerassiplatz 1, 1030 Vienna, Austria

[2]University of Vienna, Department of Neuroscience and Developmental Biology, Division of Molecular Evolution and Development, Djerassiplatz 1, 1030 Vienna, Austria

*__Author for correspondence__: Andreas Wanninger; email: andreas.wanninger@univie.ac.at

ORCID nr:

David A. Salamanca-Díaz: 0000-0003-2082-3939

Elena A. Ritschard: 0000-0002-4956-9703

Hannah Schmidbaur: 0000-0002-0099-7491

Andreas Wanninger: 0000-0002-3266-5838

Abstract

Mollusks are known for their highly diverse repertoire of body plans that often includes external armor in form of mineralized hardparts. Representatives of the Conchifera, one of the two major lineages that comprises taxa which originated from a uni-shelled ancestor (Monoplacophora, Gastropoda, Cephalopoda, Scaphopoda, Bivalvia), are particularly relevant regarding the evolution of mollusk shells. Previous studies have found that the shell matrix of the adult shell (teleoconch) is rapidly evolving, and that the gene set involved in shell formation is highly taxon-specific. However, detailed annotation of genes expressed in tissues involved in the formation of the embryonic shell (protoconch I) or the larval shell (protoconch II) are currently lacking. Here, we analyzed the genetic toolbox involved in embryonic and larval shell formation in the quagga mussel *Dreissena rostriformis* using single cell RNA sequencing. We found significant differences in genes expressed during embryonic and larval shell secretion, calling into question ontogenetic homology of these transitory bivalve shell types. Further ortholog comparisons throughout Metazoa indicates that a common genetic biomineralization toolbox, that was secondarily co-opted into molluscan shell formation, was already present in the last common metazoan ancestor. Genes included are *engrailed*, *carbonic anhydrase*, and *tyrosinase* homologs. However, we found that 25% of the genes expressed in the embryonic shell field of *D. rostriformis* lack an ortholog match with any other metazoan. This indicates that not only adult but also embryonic mollusk shells may be fast-evolving structures. We raise the question as to what degree, and on which taxonomic level, the gene complement involved in conchiferan protoconch formation may be lineage-specific or conserved across taxa.

Key words: Mollusca, Bivalvia, *Dreissena*, eco-evodevo, trochophore, larval shell, single-cell seq, proteomics

Introduction

Mollusca constitutes one of the most diverse metazoan phyla. It is composed of two major subclades, Aculifera and Conchifera, which diverged from one another in the Cambrian (Vinther, 2015; Wanninger & Wollesen, 2015; Parkhaev, 2017; Wanninger & Wollesen, 2019). The Aculifera includes the vermiform, spicule-bearing Solenogastres (Neomeniomorpha) and Caudofoveata (Chaetodermomorpha), as well as the dorso-ventrally flattened Polyplacophora with eight shell plates. The primarily single-shelled Conchifera contains the Monoplacophora, Scaphopoda, Gastropoda, Bivalvia, and Cephalopoda (Kocot et al., 2011; Smith et al., 2011; Vinther, 2015). One molluscan key characteristic is the presence of a mineralized exoskeleton that may come in form of spicules and scales, single or bipartite shells, or serially arranged shell plates. This external armor might have played a crucial role in the evolutionary success of the phylum (Marin et al., 2014).

Molluscan shells and spicules are highly versatile morphological innovations that provide protection and, together with an elaborated musculature, often aid in maintaining structural support (Lowenstam & Weiner, 1989; Simkiss & Wilbur, 2012). They are formed as mineralized secretions from epithelial cells of the mantle (Marin et al., 2007; Furuhashi et al., 2009; Kocot et al., 2016). Once mineralized, shells present a considerable amount of variation in form and shape up to the microstructural level across the different taxa (Chateigner et al., 2000; Furuhashi et al., 2009). In conchiferan mollusks, the shell matrix, *i.e.* the outer layer of the mantle, is primarily composed of polysaccharides, glycoproteins, chitin, and calcium carbonate (Addadi et al., 2006; Marin et al., 2007). Previous studies that analyzed gene expression in adult mantle tissues of various bivalves and gastropods found that, despite sharing a common set of genes, the expression profiles in the shell matrix differ considerably between taxa, irrespective of their phylogenetic position. This has been used to argue that conchiferan adult shells (teleoconchs) are rapidly evolving features, thus providing an explanation for their high degree of morphological variation

across lineages (Jackson et al., 2006; Aguilera et al., 2017; Clark et al., 2020; Yarra et al., 2021).

While conchiferan teleoconchs are continuously secreted from the mantle margin and are highly variable in shape and color, the first-formed embryonic shell (protoconch I) emerges in the gastrula or in the early trochophore larva from the dorsally situated embryonic shell gland (or shell field) in a short time window. It is typically of smooth, non-sculptured appearance (see Wanninger and Wollesen, 2015 for review). Some gastropods with long-lived veliger stages as well as most bivalves form an additional, intermediate shell type, the larval shell (protoconch II) that – similar to its developmental successor, the teleoconch – is secreted from the mantle edge. Only very few studies have focused on the cell lineage, morphological, biochemical, and molecular aspects of the formation of these elusive and microscopic protoconch types (Henry et al., 2004; Kakoi et al., 2008; Lyons et al., 2017; X. Liu et al., 2018; G. Liu et al., 2020). While embryonic and larval shell-forming cells have shown to express a common toolbox of markers such as chitin-binding proteins, *von Willebrand factor type A*, and carbonic anhydrases, they display numerous shell matrix proteins (SMPs) that are likely species-specific and also differ from those involved in teleoconch formation (Zhao et al., 2018, 2020). However, detailed analyses to assess the number and type of genes that are expressed during protoconch I and protoconch II formation are currently lacking. To fill this gap in knowledge, we reconstructed the shell formation toolbox during protoconch I development in the trochophore larva of the quagga mussel, *Dreissena rostriformis*, using a previously generated single-cell RNA-Seq dataset (Salamanca-Díaz et al. 2022). We also analyzed previously annotated genes which were shown to be expressed in the developing embryonic shell field across conchiferan mollusks for insights into the putative involvement of conserved versus hitherto unknown genes in this key developmental process in the bivalve life cycle.

Materials and Methods

Animal collection and cultures

Sexually mature individuals of *Dreissena rostriformis* were collected in the Danube River in Vienna, Austria (N 48°14´45.812“, O 16°23´38.145“). Collection took place between April and September, 2019. Adults were gathered from underneath stones and transferred to the laboratory where they were cleaned and maintained in aquaria with filtered river water (FRW) at 19°C.

Spawning of animals was induced by incubating sexually mature specimens in a 10^{-3}M solution of serotonin for 15 minutes (Sigma-Aldrich, Darmstadt, Germany) in FRW, followed by one wash and subsequent maintenance in FRW. Individuals were kept isolated in FRW in 50mL glass beakers and after approximately 30 minutes, up to 50% of the treated specimens started to spawn. Fertilization occurred when three to four drops of sperm-containing water were added to 50ml glass beakers with oocytes. After fertilization, water was changed every half an hour for the first three hours and then every six hours to remove excess sperm and avoid bacterial and fungal growth. The embryos were cultured at 23°C.

10X single-cell 3´RNAseq

Sample preparation

Cell dissociations of *Dreissena* larvae were generated by first washing 13 hours post fertilization (hpf) old trochophore larvae over a 20µm mesh with sterile media (autoclaved fresh river water; AFRW). Larvae were concentrated and dissociated by first passing them through a syringe with a hypodermic needle with 0.4mm diameter. A single-cell suspension was loaded into a 10x Chromium Controller using Chromium Single Cell 3’ Kit v2 reagents (Cat #120237, 10xGenomics, USA). cDNA synthesis and library construction were made according to specifications from the manufacturer. Library quantification was

performed on a bioanalyzer (High Sensitivity DNA reagents, Agilent Technology #5067-4626; Agilent 2100 Bioanalyzer) and sequenced on the Illumina platform as previously described (Salamanca-Díaz et al., 2022).

Mapping tool preparation and cell clustering

The transcriptomes used for creating the mapping tool, together with the reference genome to map the reads against, were previously generated (Calcino et al., 2019). In our study, gene models were elongated by 2 kilobases in the 3´ direction to account for poorly annotated three-prime ends in the gene models (Levin et al., 2016). To annotate the transcriptome, we performed a BLASTX search against both human and the Pacific giant oyster (*Crassostrea gigas*) genome for each individual gene sequence. For each transcript, the BLAST hit with the highest E-value was selected for annotation. We utilized InterProScan v5.46-81.0 (Jones et al., 2014) to search for gene ontology and to allocate domains on the reference genome by surveying publicly available databases such as GO terms, Pfam, and PANTHER (Supplementary Table 3). The reference database was generated using CellRanger Makeref v3.1.0 and demultiplexed using CellRanger Makefastq v3.1.0 with default settings and filtered according to cell barcode and Unique Molecular Markers (UMIs) using Cell Ranger Count v3.1.0 with parameters of --force-cells=1000 --chemistry=SC3Pv2. The resulting cell count gene expression matrix was analyzed in R v3.6.1 (R Development Core Team, 2015) with the Seurat v4.0.1 package (Satija et al., 2015). The count matrix was processed through a standard Seurat pipeline using default parameters. We then generated a KNN graph based on the euclidean distance in PCA space following the Louvain algorithm (FindNeighbors (k.param=10, nn.method='annoy', annoy.metric= 'cosine')) and clustered the data (FindClusters (random.seed=0, res = 0.5)). Marker genes were identified with the FindAllMarkers function (random.seed= 0), which considers only genes that were enriched and expressed in 1% of the cells in each

population (min.pct = 0.1) and with a log fold difference larger than 0.6 (logfc.threshold = 0.6). Following this, expression of orthologous marker genes from distinct cell types were examined in this dataset to support homology of the previously assigned cluster identities. After this, we selected the differentially expressed genes from the cluster annotated as “shell field” for in-depth homology assessments with respective sequences from other metazoan taxa.

Assessment of unknown genes and gene architecture annotations

To assess the orthology relationships of shell field-specific genes in the trochophore stage of *D. rostriformis* with genes of other metazoan species, a comparative analysis was performed using OrthoFinder2 (Emms & Kelly, 2019). The transcriptomes and genome assembly of *Dreissena* used for this analysis were previously generated (Calcino *et al*., 2019). The genomes, transcriptomes, and gene models for 30 additional species representing major sub-phylum-level metazoan lineages were obtained from publicly available data (Supplementary Table 1). At first, proteins were filtered for the longest transcript per gene and used as an input to OrthoFinder. After this, all-versus-all similarity search was obtained using DIAMOND v0.9.15 (Buchfink *et al*., 2015) and used by OrthoFinder2 to identify orthogroups, which are groups of proteins that are likely homologous. Subsequently, proteins belonging to each orthogroup were aligned using MAFFT v7.221 (Katoh & Standley, 2013) for multiple sequence alignments to generate gene trees using FastTree (Price *et al*., 2009) (-a 16 -b WorkingDirectory -M msa -A mafft -T fasttree). The resulting trees were parsed with the OrthoFinder2 pipeline to discriminate between orthologs and paralogs within each orthogroup. Afterwards, we overlapped these results with the gene sets previously characterized through differentially expressed genes in the

single-cell RNA sequencing of the shell field. This resulted in identification of the orthogroups which contain differentially expressed genes in the shell field of the trochophore larva.

For insights into the architecture of genes that are differentially expressed in the shell field, we implemented the webserver of SignalP v5.0 with default parameters (Almagro Armenteros et al., 2019) to search for signal peptides in each corresponding sequence. Additionally, TMHMM v2.0 webserver (Krogh et al., 2001) was also used to screen transmembrane domains and predict which amino acid sequences have domains on the outer side of the plasma membrane. Further gene annotations, corresponding to Pfam, PANTHER, GO term, human, and *Crassostrea gigantea* ortholog similarity, were implemented from a previous study (Salamanca et al., 2022). Gene expression levels of the 17 existing transcriptome libraries (Calcino et al., 2019) were quantified with Kallisto (transcripts per kilobase million, TPM) (Bray et al., 2016). Expression data from *Crassostrea gigas* were collected from public databases (Supplementary Table 1) and TPM values were calculated following the pipeline of a previous study (Zieger et al., 2021). Heatmaps showing normalized quantitative expression of genes were plotted with R (R Developement Core Team, 2015) with the heatmap function from the ComplexHeatmap R package (Gu *et al*., 2016) (Figure 1, Supplementary Table 3).

Results

Overall orthogroup statistics of the *Dreissena rostriformis* genome

Around one third of all predicted orthogroups (31.4%) predicted from 30 different metazoan species contain *Dreissena rostriformis* (“DRERO”) genes (cf. Supplementary Table 3). In addition, *D. rostriformis* possesses species-specific orthogroups with non-annotated genes, meaning there is a noteworthy number of genes that have no known match with any other animal sampled. However, all other species used in our analysis show similar

low percentages of genes that can be assigned to known orthogroups (Supplementary Table 3), corroborating the common notion of the vital role of lineage specific genes or families during animal genome evolution (Fernández & Gabaldón, 2020). In *D. rostriformis*, such genes identified from the genome mount up to 19.8% (7469 genes; see Supplementary Tables 2, 3).

Orthogroups containing genes from the trochophore shell field

Using the outputs from the OrthoFinder and Single-cell seq pipelines, we characterized the shell field-specific genes according to their orthogroup correspondence (Supplementary Table 2). In addition, expression dynamics of shell matrix genes during development of the trochophore of *D. rostriformis* were analyzed and compared with shell field-specific genes from the oyster *Crassostrea gigas* using previously published RNA-seq data (Figure 2, Supplementary Tables 7, 8) (Zhao et al., 2018; Calcino et al., 2019). Expression of most of these genes starts early in development, *i.e.,* shortly after fertilization, likely by maternal transcripts. High normalized peaks of transcription are seen throughout the late gastrula and trochophore stages (during which the protoconch I is established) and continue in the veliger stages (continuous protoconch II formation) *i.e.*, between 13 and 48hpf. *In situ* hybridization experiments of some of these genes have previously shown a high level of expression in stages of embryonic shell (protoconch I) formation (e.g., *Hox 1*, *hic31*) (Salamanca-Díaz et al., 2021, 2022). Furthermore, numerous orthogroups that contain genes that are specific to the embryonic shell field are shared across Metazoa (Supplementary Tables 2, 4). This demonstrates multiple cooption events of these genes into various functions in the respective metazoan lineages (Supplementary Table 1). However, we also found a group of genes (68 genes in 61 orthogroups; e.g., *engrailed*, *cyclin-A2, carbonic anhydrase*, *tyrosinase* homologs) active in *Dreissena* shell field formation that are also involved in shell formation of other mollusks (Nederbragt *et al*., 2002;

Iijima *et al.*, 2008; Kin *et al*., 2009; Samadi and Steiner, 2009; Huan *et al.*, 2013; Zhang *et al.*, 2020). These genes are also present in all other metazoans screened for herein and are commonly known to be related to body plan specification, cell cycle, and metalloenzyme activity (Figure 1 and Supplementary Table 2).

A closer analysis of specific taxonomic orthogroups (e.g., Protostomia, Lophotrochozoa) shows that the percentage of shell field genes from *Dreissena* is notably low (Figure 1). Only 2.24% of these genes are shared only in the sampled protostomes species (8 genes in 6 orthogroups) and 1.1% are restricted even more to lophotrochozoans (4 genes in 4 orthogroups). Most genes in these orthogroups do not have a match in the InterPro database but show low level similarities with human orthologs (e-values higher than 1). Among the few genes in these groups, there is a member of the Claudin protein family, a *keratin* ortholog, epidermal growth factor domains, *heat shock 70kDa protein*, and an *endonuclease 2* ortholog. Mollusca itself shows a similar situation, with few unique orthogroups (4 genes in 4 orthogroups, 1.1%) containing shell field-specific genes. This subset of genes expressed in the *Dreissena* shell field lack confident annotations (Supplementary Table 2). Human blast hits show only few domain commonalities to genes with a role in protein modification, DNA repair, Nuclear envelope component and ion exchange, i.e., *DDB1 and CUL4 associated factor 4*, *DNA repair protein XRCC1, nuclear envelope integral membrane protein 2*, and *sodium-driven bicarbonate exchanger*.

The number of hitherto non-annotated genes seems to decrease when analyzing the different lineages inside Mollusca. This suddenly changes in the branch leading to *Dreissena rostriformis*, where the number of shell field-specific genes notably increases (Figure 1). Within Conchifera, a putative Bivalvia + Gastropoda clade shows 2 hitherto undescribed genes in 2 separate orthogroups and Bivalvia alone presents a unique set of 9 genes in 9 orthogroups (Figure 1, Supplementary Table 2). In this gene set, there are low e-values and little similarity to human as well as *Crassostrea gigas* orthologs, with blast

hits to genes associated to antioxidant reactions, cell migration, cell attachment, and cellular proliferation, i.e., *superoxide dismutase*, *myomegalin*, *laminin subunit beta-4*, and *ETS domain-containing transcription factor ERF*. Additionally, from the genes expressed in the *D. rostriformis* embryonic shell field which were not assigned to any orthogroup or have a specific identity, and thus are considered here to be species-specific (86 genes in total, 24% of all shell field genes), 39 have transmembrane domains and 41 have signal peptides. This suggests that almost half of this gene subset is probably crucial for cell signaling since it has domains that interact directly with the outside of the cell membrane (Supplementary Tables 2, 5 and 6). Altogether, our results show that, while there is a core gene set expressed in the embryonic shell field which is present throughout Metazoa, there is also strong indication of novel gene emergence that is specific to the embryonic shell field of the *Dreissena* trochophore.

Discussion

High number of putative novel *Dreissena*-specific genes involved in embryonic shell formation

Previous studies have characterized the shell secretomes from larval and adult stages of two marine bivalves, *Pinctada fucata* and *Crassostrea gigas*. They found that, despite having some common gene expression signatures (e.g., *carbonic anhydrase*, *chitin binding protein*, and *von Willebrand factor type A*), they also show distinct expression patterns of larval shell matrix proteins depending on species and developmental stages. One significant subset of the genes (around 90 out of 156 genes) involved in shell secretion is expressed in trochophore stages, while the other genes are expressed during the later D-shape veliger stages, suggesting different molecular signatures underlying embryonic versus larval shell formation (Zhao et al., 2018). This calls into question the homology of embryonic and larval shells in Bivalvia. Since solid data on the genes involved in bivalve

teleoconch formation are still lacking, evolutionary relationships between the adult and the two transitory protoconch shell types currently remain unknown. This underlines that more in-depth comparative studies are needed to assess the decades-old question of (ontogenetic) homology of conchiferan embryonic, larval, and adult shells within the respective sublineages (particularly bivalves, gastropods, and scaphopods).

Our study shows that 24% (86) of the genes differentially expressed in the shell field of the trochophore of *D. rostriformis* could not be assigned to any known orthogroup (Figure 1A). From these, 13 unassigned genes have known annotations, 41 have low e-value similarity with human or *Crassostrea* orthologs, and 32 of these genes have no known annotation or ortholog match. Similar trends are also known from other mollusks, where unassigned and undescribed genes expressed in shell- and plate-forming cells appear to be highly taxon-specific. For example, secretomes from adult gastropods, bivalves, polyplacophorans, and a nautiloid cephalopod show considerable level of lineage-specific orphan genes (Jackson et al., 2006, 2009; Immel et al., 2016; Kocot et al., 2016; Marin, 2020; Setiamarga et al., 2021). It has been argued previously that the rapid evolutionary rate of these genes may be a possible reason for the lack of orthology detection of these shell matrix toolbox genes (Aguilera et al., 2017). This, in turn, could be the result of evolutionary responses to the widely varying ecological conditions shell-bearing mollusks are exposed to, since most of the gene products in the shell field are in direct contact with the environment. Interestingly, almost half of these lineage-specific orphan genes have transmembrane domains and/or signaling peptides (Supplementary Tables 2, 5 and 6). This suggests that genes expressed in the shell field at the trochophore stage might be significantly influenced by the ecology of the larva. Previous studies have found that molluscan shell proteomes drastically change when ecological factors such as the pH or the temperature are altered, but combined experimental and transcriptomic studies are currently too scarce for robust conclusions on an evolutionary level (Timmins-Schiffman et

al., 2014; Wei et al., 2015). However, since the environmental conditions during protoconch I and protoconch II formation are identical in *D. rostriformis*, this might hint towards an independent evolutionary origin (and thus argue against ontogenetic homology) of these shell types. This is further supported by the fact that, after shell field formation, there is a fluctuation of gene expression throughout development, *i.e.*, *Chitin binding domain* ortholog, (Gene.49769) and *voltage-dependent calcium channel subunit alpha-2/delta-4* ortholog (Gene.25093) (Figure 2A), demonstrating putatively different expression dynamics during protoconch I and protoconch II formation, respectively. A similar tendency emerges when comparing temporal expression dynamics of shell-specific genes of *D. rostriformis* with *C. gigas* (Figure 2B). The Pacific giant oyster seems to have different sets of genes with alternate expression throughout developmental stages where shell field formation is active, just as in *Dreissena,* thus calling into question the homology of bivalve ontogenetic shell types (cf. Zhao et al., 2018). However, further comparative studies employing different developmental stages of the same as well as similar developmental stages of different species are needed to further assess this assumption.

Metazoan biomineralization gene repertoires

Our single-cell RNAseq and OrthoFinder analyses grouped 271 (76%) out of 357 identified genes that are differentially expressed in shell field cells from the trochophore stage of *Dreissena rostriformis* into orthogroups shared with different taxa. From these resulting orthogroups, there is a fraction of *Dreissena* trochophore shell field-specific genes that are shared with the rest of the sampled metazoans (68 genes in 61 orthogroups, 19%) (Figure 1, Supplementary Table 2). It has been shown previously that mantle secretomes in other bivalves and gastropods also possess a wide range of gene families that originated prior to the emergence of the conchiferan clade (Kocot et al., 2016; Aguilera et al., 2017). Among this, a set of genes from the shell field of *Dreissena*, which are present in

other metazoans, is known to be involved in extracellular matrix formation, such as orthologs of *laminin*, *C-type lectin* domains, and immunoglobulins (Supplementary Table 2). Additionally, among this set there are genes containing leucine-rich repeat domains and semaphoring, which are also found throughout Metazoa (Supplementary Table 2). Moreover, genes from this subset of orthogroups were coopted into embryonic shell formation in conchiferan mollusks such as *Dreissena*, e.g., *Hox 1* (Gene.152834)*, Hox 4* (Gene.66474)*, Lox 4* (Gene.142102)*,* and *engrailed* (Gene.126286) (Figure 1, Supplementary Table 2) (Samadi & Steiner, 2009; Fritsch et al., 2015; Wollesen et al., 2018; Huan et al., 2020; Liu et al., 2020; Salamanca-Díaz et al., 2021). Furthermore, in these orthogroups there are genes that have been found to be also involved in biomineralization processes in echinoderms and vertebrates, e.g., *Cyclophilin-type* (Gene.103270) and *Carbonic anhydrase* (Gene.82229) (Livingston *et al.*, 2006; Mann *et al*., 2008; 2010). Such an organic matrix is formed prior to secretion of the mineralized part of the shell and is thus of crucial importance for conchiferan mollusks, but the respective factors involved are also present in other metazoans that lack a shell (Supplementary Table 2) (Marie *et al.*, 2010; 2011a; 2011b; 2012; 2014). Altogether, our data points towards a shared "molecular biomineralization toolbox" across Metazoa, but a broader taxon sampling especially from key invertebrate phyla are required for deeper evolutionary insights. Given the fact that numerous animal phyla outside the mentioned deuterostomes and mollusks contain taxa with mineralized hard parts, including accessible representatives such as annelids, brachiopods, other lophotrochozoans, as well as numerous arthropods, this hypothesis can be tested by comparative studies using single-cell RNA transcriptomic approaches.

Taken together, the quagga mussel *Dreissena rostriformis* shows a mosaic of co-option of known metazoan genes and de novo recruitment of genes with hitherto unknown function or ortholog match into embryonic (protoconch I) shell formation. Our data suggest

that not only adult but also embryonic bivalve shells are highly plastic in the gene repertoire that underlie their ontogeny, which may be indicative of non-homology of bivalve – and possibly conchiferan - ontogenetic shell types.

Author contributions

DASD and AW designed the research. DASD performed the experiments and generated data with contribution of EAR and HS. DASD performed the data analysis with assistance of EAR. DASD drafted the manuscript with input from AW, EAR and HS. DASD and AW interpreted and discussed the findings and finalized the manuscript. All authors contributed to interpretation of data and approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors thank Nicolas S. M. Robert (Vienna) for his support and valuable input when developing the OrthoFinder pipeline and for advice during data acquisition.

Funding

This work was supported by the completion grant of the Vienna Doctoral School Ecology and Evolution (VDSEE) to DASD and by the Austrian Science Fund (FWF) (grant P29455-B29 to AW).

References

Addadi, L., Joester, D., Nudelman, F., & Weiner, S. (2006). Mollusk shell formation: a source of new concepts for understanding biomineralization processes. *Chemistry–A European Journal*, *12*(4), 980–987.

Aguilera, F., McDougall, C., Degnan, B. M., & Irwin, D. (2017). Co-option and de novo gene evolution underlie molluscan shell diversity. *Molecular Biology and Evolution*, *34*(4), 779–792. https://doi.org/10.1093/molbev/msw294

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*(4), 420–423. https://doi.org/10.1038/s41587-019-0036-z

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*, 525. https://doi.org/10.1038/nbt.3519

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60.

Calcino, A. D., de Oliveira, A. L., Simakov, O., Schwaha, T., Zieger, E., Wollesen, T., & Wanninger, A. (2019). The quagga mussel genome and the evolution of freshwater tolerance. *DNA Research*, *26*(5), 411–422. https://doi.org/10.1093/dnares/dsz019

Chateigner, D., Hedegaard, C., & Wenk, H.-R. (2000). Mollusc shell microstructures and crystallographic textures. *Journal of Structural Geology*, *22*(11), 1723–1735. https://doi.org/https://doi.org/10.1016/S0191-8141(00)00088-2

Clark, M. S., Peck, L. S., Arivalagan, J., Backeljau, T., Berland, S., Cardoso, J. C. R., Caurcel, C., Chapelle, G., De Noia, M., Dupont, S., Gharbi, K., Hoffman, J. I., Last, K. S., Marie, A., Melzner, F., Michalek, K., Morris, J., Power, D. M., Ramesh, K., … Harper, E. M. (2020). Deciphering mollusc shell production: the roles of genetic

mechanisms through to ecology, aquaculture and biomimetics. *Biological Reviews*, *n/a*(n/a). https://doi.org/10.1111/brv.12640

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 1–14.

Fernández, R., & Gabaldón, T. (2020). Gene gain and loss across the metazoan tree of life. *Nature Ecology & Evolution*, *4*(4), 524–533. https://doi.org/10.1038/s41559-019-1069-x

Fritsch, M., Wollesen, T., de Oliveira, A. L., & Wanninger, A. (2015). Unexpected co-linearity of Hox gene expression in an aculiferan mollusk. *BMC Evolutionary Biology*, *15*(1), 151. https://doi.org/10.1186/s12862-015-0414-1

Furuhashi, T., Schwarzinger, C., Miksik, I., Smrz, M., & Beran, A. (2009). Molluscan shell evolution with review of shell calcification hypothesis. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *154*(3), 351–371. https://doi.org/https://doi.org/10.1016/j.cbpb.2009.07.011

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, *32*(18), 2847–2849. https://doi.org/10.1093/bioinformatics/btw313

Henry, J. Q., Okusu, A., & Martindale, M. Q. (2004). The cell lineage of the polyplacophoran, Chaetopleura apiculata: variation in the spiralian program and implications for molluscan evolution. *Developmental Biology*, *272*(1), 145–160. https://doi.org/https://doi.org/10.1016/j.ydbio.2004.04.027

Huan, P., Wang, Q., Tan, S., & Liu, B. (2020). Dorsoventral decoupling of Hox gene expression underpins the diversification of molluscs. *Proceedings of the National Academy of Sciences*, *117*(1), 503–512. https://doi.org/10.1073/PNAS.1907328117

Immel, F., Broussard, C., Catherinet, B., Plasseraud, L., Alcaraz, G., Bundeleva, I., & Marin, F. (2016). The Shell of the Invasive Bivalve Species Dreissena polymorpha:

Biochemical, Elemental and Textural Investigations. *PLOS ONE*, *11*(5), e0154264. https://doi.org/10.1371/journal.pone.0154264

Jackson, D. J., McDougall, C., Green, K., Simpson, F., Wörheide, G., & Degnan, B. M. (2006). A rapidly evolving secretome builds and patterns a sea shell. *BMC Biology*, *4*(1), 40. https://doi.org/10.1186/1741-7007-4-40

Jackson, D. J., McDougall, C., Woodcroft, B., Moase, P., Rose, R. A., Kube, M., Reinhardt, R., Rokhsar, D. S., Montagnani, C., Joubert, C., Piquemal, D., & Degnan, B. M. (2009). Parallel Evolution of Nacre Building Gene Sets in Molluscs. *Molecular Biology and Evolution*, *27*(3), 591–608. https://doi.org/10.1093/molbev/msp278

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kakoi, S., Kin, K., Miyazaki, K., & Wada, H. (2008). Early Development of the Japanese Spiny Oyster (Saccostrea kegaki): Characterization of Some Genetic Markers. *Zoological Science*, *25*(5), 455–464. https://doi.org/10.2108/zsj.25.455

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kocot, K. M., Aguilera, F., McDougall, C., Jackson, D. J., & Degnan, B. M. (2016). Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. *Frontiers in Zoology*, *13*(1), 23. https://doi.org/10.1186/s12983-016-0155-z

Kocot, K. M., Cannon, J. T., Todt, C., Citarella, M. R., Kohn, A. B., Meyer, A., Santos, S.

R., Schander, C., Moroz, L. L., Lieb, B., & Halanych, K. M. (2011). Phylogenomics reveals deep molluscan relationships. *Nature*, *477*(7365), 452–456. https://doi.org/10.1038/nature10382

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580.

Laumer, C. E., Fernández, R., Lemer, S., Combosch, D., Kocot, K. M., Riesgo, A., Andrade, S. C. S., Sterrer, W., Sørensen, M. V, & Giribet, G. (2019). Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1906), 20190831. https://doi.org/10.1098/rspb.2019.0831

Lemer, S., Bieler, R., & Giribet, G. (2019). Resolving the relationships of clams and cockles: dense transcriptome sampling drastically improves the bivalve tree of life. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1896), 20182684. https://doi.org/10.1098/rspb.2018.2684

Levin, M., Anavy, L., Cole, A. G., Winter, E., Mostov, N., Khair, S., Senderovich, N., Kovalev, E., Silver, D. H., Feder, M., Fernandez-Valverde, S. L., Nakanishi, N., Simmons, D., Simakov, O., Larsson, T., Liu, S.-Y., Jerafi-Vider, A., Yaniv, K., Ryan, J. F., … Yanai, I. (2016). The mid-developmental transition and the evolution of animal body plans. *Nature*, *531*(7596), 637–641. https://doi.org/10.1038/nature16994

Li, Y., Shen, X.-X., Evans, B., Dunn, C. W., & Rokas, A. (2021). Rooting the Animal Tree of Life. *Molecular Biology and Evolution*, *38*(10), 4322–4333. https://doi.org/10.1093/molbev/msab170

Liu, G., Huan, P., & Liu, B. (2020). Identification of three cell populations from the shell gland of a bivalve mollusc. *Development Genes and Evolution*.

https://doi.org/10.1007/s00427-020-00646-9

Liu, X., Jin, C., Li, H., Bai, Z., & Li, J. (2018). Morphological structure of shell and expression patterns of five matrix protein genes during the shell regeneration process in Hyriopsis cumingii. *Aquaculture and Fisheries*, *3*(6), 225–231. https://doi.org/https://doi.org/10.1016/j.aaf.2018.09.005

Livingston, B. T., Killian, C. E., Wilt, F., Cameron, A., Landrum, M. J., Ermolaeva, O., Sapojnikov, V., Maglott, D. R., Buchanan, A. M., & Ettensohn, C. A. (2006). A genome-wide analysis of biomineralization-related proteins in the sea urchin Strongylocentrotus purpuratus. *Developmental Biology*, *300*(1), 335–348.

Lowenstam, H. A., & Weiner, S. (1989). *On biomineralization*. Oxford University Press on Demand.

Lyons, D. C., Perry, K. J., & Henry, J. Q. (2017). Morphogenesis along the animal-vegetal axis: fates of primary quartet micromere daughters in the gastropod Crepidula fornicata. *BMC Evolutionary Biology*, *17*(1), 217. https://doi.org/10.1186/s12862-017-1057-1

Mann, K., & Edsinger, E. (2014). The Lottia gigantea shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. *Proteome Science*, *12*(1), 1–12.

Mann, K., Poustka, A. J., & Mann, M. (2008). The sea urchin (Strongylocentrotus purpuratus) test and spine proteomes. *Proteome Science*, *6*(1), 1–10.

Mann, K., Wilt, F. H., & Poustka, A. J. (2010). Proteomic analysis of sea urchin (Strongylocentrotus purpuratus) spicule matrix. *Proteome Science*, *8*(1), 1–12.

Marie, B., Joubert, C., Tayalé, A., Zanella-Cléon, I., Belliard, C., Piquemal, D., Cochennec-Laureau, N., Marin, F., Gueguen, Y., & Montagnani, C. (2012). Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. *Proceedings of the National Academy of*

Sciences, *109*(51), 20986–20991.

Marie, B., Marie, A., Jackson, D. J., Dubost, L., Degnan, B. M., Milet, C., & Marin, F. (2010). Proteomic analysis of the organic matrix of the abalone Haliotis asinina calcified shell. *Proteome Science*, *8*(1), 1–11.

Marie, B., Trinkler, N., Zanella-Cleon, I., Guichard, N., Becchi, M., Paillard, C., & Marin, F. (2011). Proteomic identification of novel proteins from the calcifying shell matrix of the Manila clam Venerupis philippinarum. *Marine Biotechnology*, *13*(5), 955–962.

Marie, B., Zanella-Cléon, I., Guichard, N., Becchi, M., & Marin, F. (2011). Novel proteins from the calcifying shell matrix of the Pacific oyster Crassostrea gigas. *Marine Biotechnology*, *13*(6), 1159–1168.

Marin, F. (2020). Mollusc shellomes: Past, present and future. *Journal of Structural Biology*, *212*(1), 107583. https://doi.org/https://doi.org/10.1016/j.jsb.2020.107583

Marin, F., Le Roy, N., Marie, B., Ramos-Silva, P., Bundeleva, I., Guichard, N., & Immel, F. (2014). Metazoan calcium carbonate biomineralizations: macroevolutionary trends – challenges for the coming decade. *Bulletin de La Société Géologique de France*, *185*(4), 217–232. https://doi.org/10.2113/gssgfbull.185.4.217

Marin, F., Luquet, G., Marie, B., & Medakovic, D. B. T.-C. T. in D. B. (2007). Molluscan Shell Proteins: Primary Structure, Origin, and Evolution. In *Current Topics in Developmental Biology* (Vol. 80, pp. 209–276). Academic Press. https://doi.org/https://doi.org/10.1016/S0070-2153(07)80006-8

Parkhaev, P. Y. (2017). Origin and the early evolution of the phylum Mollusca. *Paleontological Journal*, *51*(6), 663–686.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, *26*(7), 1641–1650. https://doi.org/10.1093/molbev/msp077

R Developement Core Team. (2015). R: A Language and Environment for Statistical

Computing. *R Foundation for Statistical Computing*, *1*, 409. https://doi.org/10.1007/978-3-540-74686-7

Salamanca-Díaz, D. A., Calcino, A. D., de Oliveira, A. L., & Wanninger, A. (2021). Non-collinear Hox gene expression in bivalves and the evolution of morphological novelties in mollusks. *Scientific Reports*, *11*(1), 3575. https://doi.org/10.1038/s41598-021-82122-6

Salamanca-Díaz, D. A., Schulreich, S. M., Cole, A. G., & Wanninger, A. (2022). Single-Cell RNA sequencing atlas from a vivalve larva enhances classical cell lineage studies. *Frontiers in Ecology and Evolution* 9: 783984. https://doi.org/10.3389/fevo.2021.783984

Samadi, L., & Steiner, G. (2009). Involvement of Hox genes in shell morphogenesis in the encapsulated development of a top shell gastropod (Gibbula varia L.). *Development Genes and Evolution*, *219*(9–10), 523–530. https://doi.org/10.1007/s00427-009-0308-6

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*, *33*(5), 495–502. https://doi.org/10.1038/nbt.3192

Setiamarga, D. H. E., Hirota, K., Yoshida, M., Takeda, Y., Kito, K., Ishikawa, M., Shimizu, K., Isowa, Y., Ikeo, K., Sasaki, T., & Endo, K. (2021). Hydrophilic Shell Matrix Proteins of Nautilus pompilius and the Identification of a Core Set of Conchiferan Domains. In *Genes* (Vol. 12, Issue 12). https://doi.org/10.3390/genes12121925

Simkiss, K., & Wilbur, K. M. (2012). *Biomineralization*. Elsevier.

Smith, S. a., Wilson, N. G., Goetz, F. E., Feehery, C., Andrade, S. C. S., Rouse, G. W., Giribet, G., & Dunn, C. W. (2011). Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature*, *480*(7377), 364–367.

https://doi.org/10.1038/nature10526

Timmins-Schiffman, E., Coffey, W. D., Hua, W., Nunn, B. L., Dickinson, G. H., & Roberts, S. B. (2014). Shotgun proteomics reveals physiological response to ocean acidification in Crassostrea gigas. *BMC Genomics*, *15*(1), 1–18.

Vinther, J. (2015). The origins of molluscs. *Palaeontology*, *58*(1), 19–34. https://doi.org/https://doi.org/10.1111/pala.12140

Wanninger, A., & Wollesen, T. (2015). Evolutionary developmental biology of invertebrates 2: Lophotrochozoa (Spiralia). In A. Wanninger (Ed.), *Evolutionary Developmental Biology of Invertebrates 2: Lophotrochozoa (Spiralia)* (pp. 103–154). Springer-Verlag Wien. https://doi.org/10.1007/978-3-7091-1871-9

Wanninger, A., & Wollesen, T. (2019). The evolution of molluscs. *Biological Reviews*, *94*(1), 102–115. https://doi.org/10.1111/brv.12439

Wei, L., Wang, Q., Ning, X., Mu, C., Wang, C., Cao, R., Wu, H., Cong, M., Li, F., & Ji, C. (2015). Combined metabolome and proteome analysis of the mantle tissue from Pacific oyster Crassostrea gigas exposed to elevated pCO2. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, *13*, 16–23.

Wollesen, T., Rodriguez Monje, S. V., de Oliveira, A. L., & Wanninger, A. (2018). Staggered Hox expression is more widespread among molluscs than previously appreciated. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1888). https://doi.org/10.1098/rspb.2018.1513

Yarra, T., Blaxter, M., & Clark, M. S. (2021). A Bivalve Biomineralization Toolbox. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msab153

Zhao, R., Takeuchi, T., Koyanagi, R., Villar-Briones, A., Yamada, L., Sawada, H., Ishikawa, A., Iwanaga, S., Nagai, K., Che, Y., Satoh, N., & Endo, K. (2020). Phylogenetic comparisons reveal mosaic histories of larval and adult shell matrix protein deployment in pteriomorph bivalves. *Scientific Reports*, *10*(1), 22140.

https://doi.org/10.1038/s41598-020-79330-x

Zhao, R., Takeuchi, T., Luo, Y.-J., Ishikawa, A., Kobayashi, T., Koyanagi, R., Villar-Briones, A., Yamada, L., Sawada, H., Iwanaga, S., Nagai, K., Satoh, N., & Endo, K. (2018). Dual Gene Repertoires for Larval and Adult Shells Reveal Molecules Essential for Molluscan Shell Formation. *Molecular Biology and Evolution*, *35*(11), 2751–2761. https://doi.org/10.1093/molbev/msy172

Zieger, E., Robert, N. S. M., Calcino, A., & Wanninger, A. (2021). Ancestral Role of Ecdysis-Related Neuropeptides in Animal Life Cycle Transitions. *Current Biology*, *31*(1), 207-213.e4. https://doi.org/https://doi.org/10.1016/j.cub.2020.10.004

Figure Legends

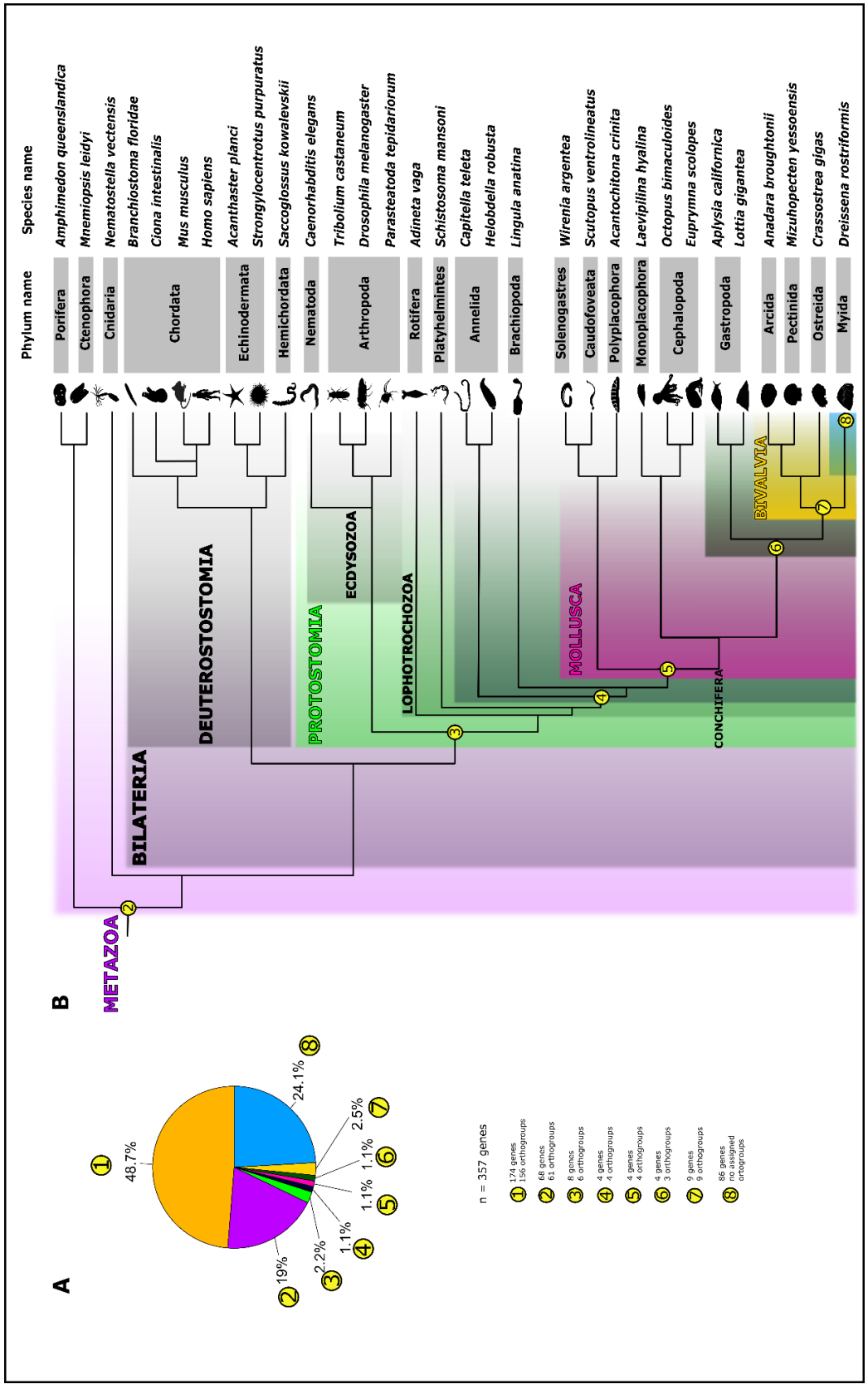


Figure 1. Distribution of orthogroups containing shell field-specific genes from *Dreissena rostriformis* across Metazoa. A: Pie chart representing the number of orthogroups and genes found in the respective taxa. Each subset of orthogroups is numbered (1-8), indicating how many shell field-specific genes are contained in each taxon, together with the total amount of shell field-specific genes analyzed. B: Dendrogram representing phylogenetic relationships of the sampled species and the presence of orthogroups and genes on each node. Phylogenetic relationships of the sampled species are plotted on a class-level tree based on previous studies (Smith et al., 2011; Laumer et al., 2019; Lemer et al., 2019; Fernández & Gabaldón, 2020; Li et al., 2021). Numbers correspond to those in (A). Number 1 refers to all shell field orthogroups that are randomly distributed (i.e., without distinct pattern) among the sampled metazoans (48.7%). Numbers 2-8 depict the shell field genes/orthogroups identified for the respective nodes in the phylogeny. Node (2) refers to the 19% of all shell field orthogroups present in all sampled metazoan genomes in this study. Node (3) is equivalent to 2.2% of all shell field orthogroups present in sampled protostome organisms. Node (4) corresponds to the 1.1% of orthogroups present in the sampled organisms classified as Lophotrochozoa. Node (5) refers to all shell field orthogroups present exclusively in the sampled mollusks (1.1%). Node (6) represents all shell field orthogroups (1.1%) present in the gastropod and bivalve genomes sampled. Node (7) depicts all shell field orthogroups (2.5%) present in bivalve genomes analyzed here. Node (8) represents all *D. rostriformis*-specific shell field genes that could not be assigned to any orthogroup. Species silhouettes were obtained from www.phylopic.org and are either licensed under Creative Commons Attribution 3.0 Unported or are available under public domain.

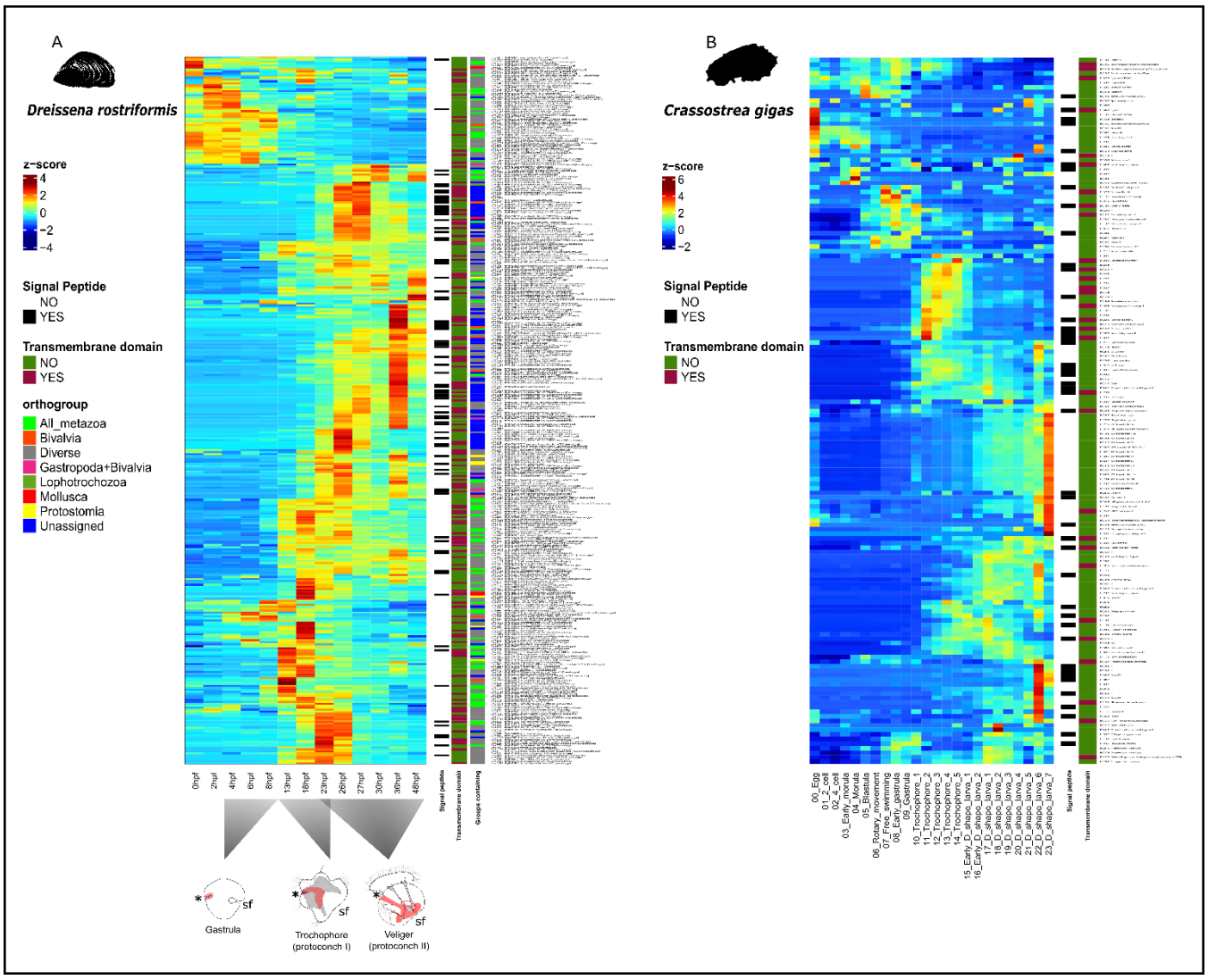


Figure 2. Relative quantitative expression of shell field-specific genes during development of two bivalve species. A: Heat map showing relative normalized expression levels for each isolated gene from the shell field of the trochophore of *Dreissena rostriformis*. Normalized gene expression (transcripts per kilobase million; TPM) is depicted in graded shades of red when values are above the median, those below this threshold and with a value close to zero are in shades of blue. Details on gene annotations, orthogroup assignments, presence of signaling peptides, and transmembrane domains are provided in Supplementary Table 2. Blast top hit to the the Pacific giant oyster is next to each gene name. Time after fertilization (hpf) and corresponding developmental stages at 23°C are

ordered chronologically at the bottom of the x axis. Germ layers and their major derivatives in animal schemes are depicted in grey (mesoderm), red (endoderm), and white (ectoderm), respectively. Asterisks mark the blastopore/mouth, sf indicates the shell field. B: Heat map showing relative normalized expression levels of genes isolated from larval and adult shells of *Crassostrea gigas* as described in Zhao et al., 2018. Normalized gene expression is depicted in graded shades of red when values are above the median, those below this threshold and with a value close to zero are in shades of blue. Each developmental stage is organized chronologically from left to right on the x axis. Details on gene annotations, orthogroup assignments, presence of signaling peptides, and transmembrane domains are provided in Supplementary Table 9 (cf. Zhao et al., 2018).

Supplementary Table 1. Genome and proteomes of publicly available data. Columns show species scientific names, common names, their respective abbreviation used in this study, databases where the data were obtained from, as well as the molecular nature (either genome or transcriptome assembly) of each sample.

Supplementary Table 2. Annotations of shell field genes from *Dreissena rostriformis*. Results show the orthofinder, InterProScan, TMHMM, SignalP, and BLAST searches against the human genome. Each column corresponds to the gene code from the quagga mussel, the analysis performed on the aminoacid sequence, the signature accession result of the analysis, signature description, accession code on the InterPro database, InterPro description, associated gene ontology terms, similarity index of the analysis (e-value), blast top hit against the human genome and the Pacific giant oyster with the respective e-value, presence of transmembrane domains and signal peptides, orthogroup containing the corresponding gene and the taxonomic clades containing that orthogroup, and the length of the protein sequence of each gene.

Supplementary Table 3. Statistics result of the orthofinder analysis showing the number of genes assigned to each orthogroup for all species used in the analysis.

Supplementary Table 4. Resulting orthogroups that contain at least one shell field gene obtained from the single cell RNA seq analysis from *D. rostriformis*.

Supplementary Table 5. Transmembrane helices domains predicted for trochophore shell field genes from *D. rostriformis* with the TMHMM 2.0 webserver.

Supplementary Table 6. Signal peptide-positive proteins prediction for the trochophore shell field genes from *D. rostriformis* with the SignalP 5.0 software.

Supplementary Table 7. Raw data of relative expression levels (transcripts per kilobase million, TPM) of *Dreissena rostriformis* shell field genes throughout development.

Supplementary Table 8. Raw data of relative expression levels (transcripts per kilobase million, TPM) of *Crassostrea gigas* shell field genes throughout development (based on Zhao et al., 2018).

Supplementary Table 9. Annotations of shell-related genes from *Crassostrea gigas.* Results from annotations of shell protein matrix genes taken from Zhao et al., 2018. Each column corresponds to annotations assigned in Zhao et al., 2018. Indicated is the gene code from the NCBI database, the result of the analysis performed on the aminoacid sequence showing the presence of transmembrane domains and signal peptides, and the assigned signature description.

Supplementary material available online at:

https://drive.google.com/drive/folders/1Fad1AQDDwgf5irU-XFicPUe8k-Aa2zJp?usp=sharing

3. General Discussion

The following headers are result of the discussion from experiments made in this thesis. As mentioned above these chapters are being compiled in the form of scientific papers. Therefore, the key points product of each chapter is only briefly summarized in this section.

3.1. Hox/ParaHox genes and their implications for bivalve development (cf. Manuscript 1)

Hox and ParaHox positions in the genome assembly revealed that these genes were distributed over five scaffolds, while two of the three ParaHox genes (*Gsx, Xlox*) were found on a single scaffold. Additionally, several genes with no apparent relationship were found between members of the Hox cluster (between *Hox5* and *Lox5, Lox2* and *Post2, Xlox* and *Cdx*). This denotates insertions all along the gene set and provides evidence in support of disorganized (but not broken) Hox/ ParaHox clusters.

The non-staggered expression of *Dro-Hox1, Dro-Hox2, Dro-Hox3, Dro-Hox4, Dro-Lox2, Dro-Lox5, Dro-Post2,* and *Dro-Xlox* is also reflected by their non-collinear but rather synchronous temporal activation. Moreover, *D. rostriformis* shows few similarities of Hox/ParaHox expression domains when compared to other conchiferan with different degrees of Hox spatial staggering along the anterior–posterior axis (Cephalopods, Gastropods, Scaphopods, Bivalves) (Lee et al., 2003; Samadi & Steiner, 2009, 2010b, 2010a; Wang et al., 2017; Wollesen et al., 2018) and aplacophoran mollusks with an ancestral-like Hox spatial staggered expression (Polyplacophorans) (Fritsch et al., 2015). The

quagga mussel shows overall overlapping mesodermal expression of Hox genes which are also highly dynamic throughout development. Hence, results of Chapter 1 could suggest Hox genes have been recruited independently into novel expression domains in various conchiferan lineages and have most likely also acquired novel roles at the cost of their original function in anterior–posterior axis specification. Additionally, when comparing with other mollusks the dynamic non-staggered expression domains of *Dreissena* demonstrate to have a minor role on whole body regionalization during later stages of bivalves.

3.2. First single cell transcriptomic expression atlas from a larval stage of a freshwater bivalve (cf. Manuscript 2)

Results from Chapter 2 manage to compile and annotate all expected major cell populations on the trochophore larva of the quagga mussel. Together with the bioinformatic validations of top marker genes of each cluster, we provide a first glimpse into developmental decisions of a mollusk larva. Cell groups were linked together by tracing pseudotime trajectories across our dataset. As these trajectories are represented in function of pseudotime, they are used as a bioinformatical approximation of developmental decisions. By comparing terminal fates from other molluscan lineage trees to our results, we were able to link scRNA-seq data to the cellular lineage tree of *D. rostriformis* (Meisenheimer 1900). For example, the linkage between ‘muscle’ and ‘endoderm’ clusters most likely represents the origin of these cell groups from mesodermal lineages. Additionally, the grouping of other clusters i.e., ‘shell field’ and ‘epidermal’, ‘neuronal’ and ‘ciliary/prototroch’, could represent the origin of macromeres arising from the ectodermal germ layer (Nielsen, 2005; Kurita et al., 2009; Chan and Lambert, 2011; Gharbiah et al., 2013; Lyons

et al., 2017). Our data demonstrate that the levels of differentially expressed genes can be used as markers to identify clusters of cells in early developmental (larval) stages of (bivalve) mollusks, thus adding these animals to the suite of taxa that can now be accessed for comparative evolutionary studies on cell type level using scRNA-seq techniques.

3.3. Transcriptomic signatures of shell field cells reveal gene emergence (cf. Manuscript 3)

Previous studies analyzing gene products expressed in mantles of bivalves and gastropods have stated that, despite having a common biomineralization toolbox, the shell matrix has a rapidly evolving nature in mollusks. Since it is a morphological novelty in the mollusks, it deploys domain shuffling and many gene family expansions in genes that play a role in shell architecture (i.e., carbonic anhydrase, tyrosinase, SPARC and chitin-binding protein) (Aguilera et al., 2017). Although part of our data points to a shared biomineralization genetic makeup across Metazoa, each one of the species lineages follows a unique specification pathway, exemplified here in the trochophore larva of *Dreissena*. Results of manuscript 2 show that 24% (86) of the genes expressed in the shell field of the trochophore of Dreissena could not be assigned to any known orthogroup. Given the physical location in the cell and properties of the translated proteins, i.e., transmembrane domains and/or signal peptides, there could be an indication of influence of the ecology of the larva in these unassigned genes. However, since the environmental conditions along across developmental stages are indistinguishable, this rather hints towards a mo-

saic of co-option of known metazoan genes and de novo recruitment of genes with hitherto unknown function or ortholog match into biomineralization processes during embryonic (protoconch I) shell formation. Of those genes that are differentially expressed in the shell field, a majority are active during extracellular matrix formation and may point towards non-homology of bivalve ontogenetic shell types. Altogether, results of this manuscript corroborate the high plasticity and fast evolution in the gene repertoire involved in development of bivalve shells.

3.4. References General Discussion

Achim, K., Eling, N., Vergara, H. M., Bertucci, P. Y., Musser, J., Vopalensky, P., Brunet, T., Collier, P., Benes, V., Marioni, J. C., & Arendt, D. (2018). Whole-Body Single-Cell Sequencing Reveals Transcriptional Domains in the Annelid Larval Body. Molecular Biology and Evolution, 35(5), 1047–1062. https://doi.org/10.1093/molbev/msx336

Achim, K., Pettit, J.-B., Saraiva, L. R., Gavriouchkina, D., Larsson, T., Arendt, D., & Marioni, J. C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. Nature Biotechnology, 33(5), 503–509.

Aguilera, F., McDougall, C., Degnan, B. M., & Irwin, D. (2017). Co-option and de novo gene evolution underlie molluscan shell diversity. Molecular Biology and Evolution, 34(4), 779–792. https://doi.org/10.1093/molbev/msw294

Briggs, J. A., Weinreb, C., Wagner, D. E., Megason, S., Peshkin, L., Kirschner, M. W., & Klein, A. M. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science, 360(6392), eaar5780.

https://doi.org/10.1126/science.aar5780

Chan, X. Y., & Lambert, J. (2011). Patterning a spiralian embryo: A segregated RNA for a Tis11 ortholog is required in the 3a and 3b cells of the Ilyanassa embryo. Developmental Biology, 349, 102–112. https://doi.org/10.1016/j.ydbio.2010.10.001

Duruz, J., Kaltenrieder, C., Ladurner, P., Bruggmann, R., Martìnez, P., & Sprecher, S. G. (2020). Acoel single-cell transcriptomics: cell-type analysis of a deep branching bilaterian. BioRxiv, 2020.07.10.196782. https://doi.org/10.1101/2020.07.10.196782

Farrell, J. A., Wang, Y., Riesenfeld, S. J., Shekhar, K., Regev, A., & Schier, A. F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science, 360(6392), eaar3131. https://doi.org/10.1126/science.aar3131

Fincher, C. T., Wurtzel, O., de Hoog, T., Kravarik, K. M., & Reddien, P. W. (2018). Cell type transcriptome atlas for the planarian Schmidtea mediterranea. Science, 360(6391), eaaq1736. https://doi.org/10.1126/science.aaq1736

Foster, S., Oulhen, N., & Wessel, G. (2020). A single cell RNA sequencing resource for early sea urchin development. Development, 147(17), dev191528. https://doi.org/10.1242/dev.191528

Foster, S., Teo, Y. V., Neretti, N., Oulhen, N., & Wessel, G. M. (2019). Single cell RNA-seq in the sea urchin embryo show marked cell-type specificity in the Delta/Notch pathway. Molecular Reproduction and Development, 86(8), 931–934. https://doi.org/10.1002/mrd.23181

Fritsch, M., Wollesen, T., de Oliveira, A. L., & Wanninger, A. (2015). Unexpected co-linearity of Hox gene expression in an aculiferan mollusk. BMC Evolutionary Biology, 15(1), 151. https://doi.org/10.1186/s12862-015-0414-1

García-Castro, H., Kenny, N. J., Iglesias, M., Álvarez-Campos, P., Mason, V., Elek, A.,

Schönauer, A., Sleight, V. A., Neiro, J., Aboobaker, A., Permanyer, J., Irimia, M., Sebé-Pedrós, A., & Solana, J. (2021). ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics. Genome Biology, 22(1), 89. https://doi.org/10.1186/s13059-021-02302-5

Gharbiah, M., Nakamoto, A., & Nagy, L. M. (2013). Analysis of ciliary band formation in the mollusc Ilyanassa obsoleta. Development Genes and Evolution, 223(4), 225–235. https://doi.org/10.1007/s00427-013-0440-1

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., & Zinzen, R. P. (2017). The Drosophila embryo at single-cell transcriptome resolution. Science, 358(6360), 194 LP – 199. https://doi.org/10.1126/science.aan3235

Kocot, K. M., Aguilera, F., McDougall, C., Jackson, D. J., & Degnan, B. M. (2016). Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. Frontiers in Zoology, 13(1), 23. https://doi.org/10.1186/s12983-016-0155-z

Kurita, Y., Deguchi, R., & Wada, H. (2009). Early Development and Cleavage Pattern of the Japanese Purple Mussel, Septifer virgatus. Zoological Science, 26(12), 814–820. https://doi.org/10.2108/zsj.26.814

Lee, P. N., Callaerts, P., De Couet, H. G., & Martindale, M. Q. (2003). Cephalopod Hox genes and the origin of morphological novelties. Nature, 424(6952), 1061–1065. https://doi.org/10.1038/nature01872

Lyons, D. C., Perry, K. J., & Henry, J. Q. (2017). Morphogenesis along the animal-vegetal axis: fates of primary quartet micromere daughters in the gastropod Crepidula fornicata. BMC Evolutionary Biology, 17(1), 217. https://doi.org/10.1186/s12862-017-1057-1

Nielsen, C. (2005). Trochophora larvae: Cell-lineages, ciliary bands and body regions. 2. Other groups and general discussion. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution, 304(5), 401–447. https://doi.org/10.1002/jez.b.21050

Plass, M., Solana, J., Wolf, F. A., Ayoub, S., Misios, A., Glažar, P., Obermayer, B., Theis, F. J., Kocks, C., & Rajewsky, N. (2018). Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. Science, 360(6391), eaaq1723. https://doi.org/10.1126/science.aaq1723

Samadi, L., & Steiner, G. (2009). Involvement of Hox genes in shell morphogenesis in the encapsulated development of a top shell gastropod (Gibbula varia L.). Development Genes and Evolution, 219(9–10), 523–530. https://doi.org/10.1007/s00427-009-0308-6

Samadi, L., & Steiner, G. (2010a). Conservation of ParaHox genes' function in patterning of the digestive tract of the marine gastropod Gibbula varia. BMC Developmental Biology, 10(1), 74. https://doi.org/10.1186/1471-213X-10-74

Samadi, L., & Steiner, G. (2010b). Expression of Hox genes during the larval development of the snail, Gibbula varia (L.)-further evidence of non-colinearity in molluscs. Development Genes and Evolution, 220(5–6), 161–172. https://doi.org/10.1007/s00427-010-0338-0

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature Biotechnology, 33(5), 495–502. https://doi.org/10.1038/nbt.3192

Sebé-Pedrós, A., Saudemont, B., Chomsky, E., Plessier, F., Mailhé, M.-P., Renno, J., Loe-Mie, Y., Lifshitz, A., Mukamel, Z., Schmutz, S., Novault, S., Steinmetz, P. R. H., Spitz, F., Tanay, A., & Marlow, H. (2018). Cnidarian Cell Type Diversity and

Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. Cell, 173(6), 1520-1534.e20. https://doi.org/10.1016/j.cell.2018.05.019

Siebert, S., Farrell, J. A., Cazet, J. F., Abeykoon, Y., Primack, A. S., Schnitzler, C. E., & Juliano, C. E. (2019). Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. Science, 365(6451), eaav9314. https://doi.org/10.1126/science.aav9314

Sur, A., & Meyer, N. P. (2021). Resolving Transcriptional States and Predicting Lineages in the Annelid Capitella teleta Using Single-Cell RNAseq . In Frontiers in Ecology and Evolution (Vol. 8, p. 523). https://www.frontiersin.org/article/10.3389/fevo.2020.618007

Svensson, V., Vento-Tormo, R., & Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. Nature Protocols, 13(4), 599–604.

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., & Surani, M. A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. Cell Stem Cell, 6(5), 468–478.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. Genome Research, 25(10), 1491–1498.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., & Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature Biotechnology, 32(4), 381.

Vergara, H. M., Bertucci, P. Y., Hantz, P., Tosches, M. A., Achim, K., Vopalensky, P., & Arendt, D. (2017). Whole-organism cellular gene-expression atlas reveals conserved cell types in the ventral nerve cord of Platynereis dumerilii. Proceedings of the National Academy of Sciences, 114(23), 5878–5885.

Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G., & Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science, 360(6392), 981 LP – 987. https://doi.org/10.1126/science.aar4362

Wang, S., Zhang, J., Jiao, W., Li, J., Xun, X., Sun, Y., Guo, X., Huan, P., Dong, B., Zhang, L., Hu, X., Sun, X., Wang, J., Zhao, C., Wang, Y., Wang, D., Huang, X., Wang, R., Lv, J., … Bao, Z. (2017). Scallop genome provides insights into evolution of bilaterian karyotype and development. Nature Ecology & Evolution, 1(April), 0120. https://doi.org/10.1038/s41559-017-0120

Wendt, G., Zhao, L., Chen, R., Liu, C., O'Donoghue, A. J., Caffrey, C. R., Reese, M. L., & Collins, J. J. (2020). A single-cell RNA-seq atlas of Schistosoma mansoni identifies a key regulator of blood feeding. Science, 369(6511), 1644 LP – 1649. https://doi.org/10.1126/science.abb7709

Wollesen, T., Rodriguez Monje, S. V., de Oliveira, A. L., & Wanninger, A. (2018). Staggered Hox expression is more widespread among molluscs than previously appreciated. Proceedings of the Royal Society B: Biological Sciences, 285(1888). https://doi.org/10.1098/rspb.2018.1513

Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L., & Pan, N. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. Nature, 555(7697), 524–528.

4. Conclusion and key findings

The contributions of this thesis constitute the first source of gene developmental data for the entire clade of Bivalvia. This was made with the purpose of using these results for comparative evolutionary studies across all animals, while making evo-devo conclusions. The key findings presented in this thesis are the following:

- The characterization of the Hox cluster in the quagga mussel revealed non colinear way of expression in bivalves, contrasted with other conchiferans which have at least rudimentary hints of collinearity at some point in development. This suggests a lineage specific way of coopting developmental genes in conchiferan mollusks.
- Trochophore larvae of *D. rostriformis* present characteristic cell types (i.e., ciliary, muscle, neuronal, shell field and endodermal cells) with a repertoire of conserved transcriptomic signatures that can be compared across metazoans.
- All the extant cell clusters in the trochophore of *D. rostriformis* were successfully annotated, showing a distinct transcriptomic signature with further subcluster structure.
- Shell field cells on the trochophore whilst expressing a significant number of unknown genes, present the existence of an expression toolbox likely to be present in the last common ancestor of all metazoans.
- The gene novelty and expression patterns described in shell field cells of *D. rostriformis* trochophore stage indicates that not only adult, but also embryonic mollusk shells may be fast-evolving structures.

- Temporal dynamics of shell field-specific genes present a highly variable expression throughout development after shell field formation. This hints towards an independent evolutionary origin of bivalve – and potentially other conchiferan - protoconch I, protoconch II, and teleoconchs.

5. Appendix

5.1. Relevant coauthorships

5.1.1. In preparation or under review

Schulreich, S.M; Salamanca-Diaz, D. A; Zieger, E; Calcino, A.D; Wanninger, A. A mosaic of conserved and novel modes of gene expression and morphogenesis in mesoderm and muscle formation of a larval bivalve. Submitted to Organisms Diversity & Evolution.