Developmental Small RNA Transcriptomics of *Pristionchus pacificus*: Insights on a Transcriptional Regulatory Mechanism from a Comparative Developmental Perspective

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Devansh Raj Sharma
aus Udaipur, Indien

Tübingen
2024

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	13.02.2025
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Ralf J. Sommer
2. Berichterstatter:	Prof. Dr. Heinz-R. Köhler

Acknowledgements

The last six years of my life have been a tremendous learning curve, with each day having presented me with a new stepping stone to reach the stage where I find myself today. None of this would have been possible without the presence of the following people in my life.

First and foremost, as the creator of this doctoral thesis, I am indebted to my creators, **Dr. Vinita Sharma** and **Dr. Rajendra Sharma**. It would be silly of me to even attempt to elucidate their role in my life, let alone my PhD, so I will leave it at that by just mentioning that if I could ever raise a kid as they raised me, that would certainly be the biggest achievement of my life.

Next, I would like to thank my supervisor, **Prof. Dr. Ralf J. Sommer**, for singularly changing my world upside down when he decided to have me as a PhD student in Sommerlab. From showing me the fascinating world of nematodes to letting me explore the world of scientific biological research in his lab to sharing his invaluable tips on debating and public presentations, the list of learnings that I have received from him is too long to mention in its entirety here. But certainly, the two most important life lessons that I have received from him are his toiling energy, and his love and enthusiasm for research. If, as I proceed through the turns and twists of my own professional career in the days to come, I am able to incorporate and imbibe even a fraction of his dedication and enthusiasm in myself, then I shall consider myself a success.

My time in Sommerlab went by smoothly due to the ever-competent and helpful technical staff. I would like to sincerely thank **Tobias Loschko** and **Waltraud Röseler** for coping up with me on a daily basis in the same room, to **Metta Riebesell** and **Hanh Witte** for their unending patience while teaching me micro-injections in nematodes, to **Heike Haussmann** for freezing and storing numerous worm strains for me (and for giving me wonderful ideas for road trips in Europe), and to **Sandra Mäck**, **Gabi Eberhardt**, **Christian Weiler**, **Dorothee Harbecke,** and **Raphaela Zurheide** for bailing me out whenever I peppered them with my troubles. A special shoutout to **Maria Gölz**, **Kostadinka Krause** and **Karin Klein** for helping me with all the administrative assistance, paperwork and bureaucratic nitty-gritties, and to **Aida Kalac**, **Janaine Miranda**, and **Alexander Richter**, for their help and support in creating a great

atmosphere, quite literally, in my doctoral lab.

I have been fortunate to have shared my time in Sommerlab with some great postdoctoral scientists and staff scientists. I extend my gratitude to **Dr. Adrian Streit**, **Dr. Matthias Herrmann**, **Dr. Catia Igreja**, **Dr. Eduardo Moreno**, **Dr. Chuanfu Dong**, **Dr. Kohta Yoshida**, and **Dr. Ata Kalirad** for lending me their ears and their thoughts whenever I felt stuck and presented them with my conundrums. A heartfelt thanks also to **Dr. James Lightfoot**, **Dr. Ziduan Han**, and **Dr. Wen-Sui Lo**, who not only provided me with professional guidance whenever I knocked on their doors, but also shared their wisdom about science and useful tips on the code of conduct of a scientist.

I would not be the scientist I am today had it not been for the following two mentors that I was fortunate to have: **Dr. Michael S. Werner** and **Dr. Christian Rödelsperger**. "Captain" Mike was the safety net that I have fallen back on multiple times in my formative years, in whose shadow I learned not only scientific but also life's lessons. I am indebted to him for his professional as well as personal contribution in my grooming. While Mike was the driving force behind my back to motivate me to learn diverse skills such as bioinformatics, Christian took me under his wing and made sure that I reached home with those tasks. Christian has been a sentinel for an entire generation of students such as myself, and is someone who has entertained my doubts and naivety (bioinformatics or otherwise) without a frown, and has on numerous occasions shepherded me back from the wilderness of my thoughts, always with a smile on his face.

I owe a good chunk of my personal and professional development to my seniors: **Dr. Suryesh Namdeo**, **Dr. Praveen Baskaran**, **Dr. Bogdan Sieriebriennikov**, **Dr. Neel Duti Prabh**, **Dr. Anja Holz**, **Dr. Alex Dulovic**, **Dr. Martin Wilecki** and **Dr. Nermin Akduman**. Each one of them has been an exemplary predecessor during my time, helping me right from day one in settling down in Tübingen and in the lab, and sharing with me their ideas whenever I pestered them with my questions. Special mention to Neel, Praveen, their wives **Dr. Nidhi Singh** and **Arthi Praveen**, Suryesh, and **Dr. Gauri Tendulkar**, for they were my flatmates and helped me settle down and helped me find a home away from home.

My time in the lab and outside of it was made memorable by my wonderful

colleagues **Marina Athanasouli**, **Dr. Mohannad Dardiry**, **Neha Agyal**, **Penghieng Theam**, **Linxuan Li**, **Veroni de Ree**, **Yinan Wang**, **Dr. Sara Wighard**, **Dr. Shuai Sun**, **Dr. Veysi Piskobulu**, **Zeeshan Ayaz Mirza**, **Shiela Pearl Quiobe**, **Elbin Aloshy**, **Xiaoxiao Yin**, **Tuğba Sehitogullari**, and **Merve Gül Turan**. I am happy to have shared this part of my life with these wonderful souls. Along with my own, I have had the opportunity to witness their career development, and I can confidently say that I have had the fortune to know some of the brightest stars of the future.

Of special importance to me has been the company of my colleague and friend **Radhika Sharma**. Ever since she joined Sommerlab, Radhika has been a constant support every day, a pillar on whom I have leaned on innumerable occasions, and continue to do so even after my time in the lab.

Seldom do contemporaries shape and mould one's professional and personal life as **Dr. Tess Renahan** and **Dr. Tobias Renahan** (née Theska) have mine. A flat-mate, lab-mate, bench-mate, and drinking partner, Theska has been the perfect 'brother from another mother' that I could have asked for during my doctorate. His sheer presence has, on occasion, given me a sense of relief at times when my going has gone tough. To say that I cherish his friendship and company is an understatement; I live it. Tess is the most cheerful and kind-hearted person that I have had the good fortune of working with, who has lent me her shoulders to lean on, on multiple occasions, during our time together in lab 1 and also after it. In both of them, I am lucky to have found friends who have been there for me at all times, in the lab and outside, cheering me the loudest amidst crowds during my moments of triumph and encouraging me with their words and their company during my tough times.

Outside of the Sommerlab, I am in deep awe of the company of my compatriots **Dr. Anasuya Moitra**, **Dr. Minakshi Singh**, **Dr. Amit Landge** and **Anupam Gautam**. They became the family that I had the fortune of having in Tübingen. Be it our 'chai pe charchaa' (translation: discussion over tea) or the non-stop discussions by the Neckar, the Lake of Constance or in each others' kitchens, I shall cherish these moments for the rest of my life.

As a former representative of the PhD students, I have had a great time with fellow representatives and friends **Dr. Thanvi Srikanth**, **Dr. Alba Gonzalez**, **Dr.**

Joe Joiner, **Dr. Liam Fitzstevens**, **Dr. Marco Podobnik** and **Dr. Felix Räsch**. In addition, a sincere vote of thanks to the former PhD program coordinators **Dr. Dagmar Sigurdardottir**, **Dr. Sarah Danes**, and to the former and current members of the Research Support Team: **Dr. Jeanette Müller**, **George Deffner**, **Sibylle Patheiger**, **Ines Müller**, **Susan Jones** and **Dr. Alejandra Duque**, all of whom have played an important part in my time here at the Max Planck Institute for Biology Tübingen.

I would like to take this opportunity to also thank our collaborators **Prof. Hendrik Korswagen** and **Dr. Annabel Ebbing**, from Hubrecht Institute, Utrecht for hosting me in the winter of 2018 in their institute and extending their hospitality and their research facilities to me, and to **Dr. Claudia Martinho** for bestowing upon me her experienced opinions and ideas in the field of small RNA biology.

Outside of the professional circle, I am deeply thankful to my Kava Conspiracy band-mates: **Dr. Richard Schulz**, **Dr. Andre Weiß**, **Dr. Tobias Ensslen**, and **Magdalena Broll**. My weekly getaways with them have been a big personal achievement and my personal let-out corner, which has helped me come back to my work with vigour on multiple occasions. Additionally, in them and their families I found new friends who made me feel welcome and loved.

Lastly, I would also like to thank my partner-in-life-and-everything, **Sadhana Golani**, for standing by me like a rock and believing in me even on occasions when I had lost the ability to do so in myself. Every achievement of mine is an achievement of hers.

Dedication

I would like to dedicate this work to the Bordia family:
Firstly, to Indira and Ashok, chief architects,
Next, to Aabha and Abeer, the sentinels,
and lastly, to Amit, a timeless inspiration, who taught me to stick to my lane.

Table of contents

A. Abstract 9

B. Zusammenfassung 11

C. List of publications 13

D. Introduction
1. Origins of evolutionary developmental biology 14
2. Developmental system drift, phenotypic novelty, and developmental plasticity: key concepts of evo-devo 19
3. The study system: *Pristionchus pacificus* 21
4. Environmental sensing, genetic architecture, and their mediators: epigenetic mechanisms 23

E. Aims of this thesis 28

F. Results
1. Crowdsourcing and the feasibility of manual gene annotation: a pilot study in the nematode *Pristionchus pacificus* 30
2. Spatial transcriptomics of nematodes identifies sperm cells as a source of genomic novelty and rapid evolution 31
3. Developmental small RNA transcriptomics reveals divergent evolution of the conserved microRNA miR-100 and the *let-7-complex* in nematodes 32
4. Environmental influence on *Pristionchus pacificus* mouth form through different culture methods 33

G. Discussion and conclusions 34

H. References 38

I. Appendix 53

A. Abstract

Embryology, or developmental biology, has been widely studied since centuries, as a field that contains answers to several fundamental questions in biology. One such persistent biological question has been that of evolution, with many classical evolutionists turning to comparative embryology in their quest for answers and explanations. However, with the rediscovery of Mendel's laws and subsequent advent of genetics in the early and mid-20th century, followed by the identification of DNA as the primary carrier of genetic information across generations, DNA and genes have taken centre stage in explaining how species evolve from pre-existing species, at the neglect of other aspects, such as development and ecology. Over the last years, though, a combination of factors - such as (re)merging of evolutionary and developmental biology, advent of a wide array of model organisms for comparative developmental studies, and technological advances in the field of molecular and computational biology - have led to a growing need to redress and reassess the fundamentals of evolution and origins of species diversity. In this context, the nematode *Pristionchus pacificus* has emerged as a key model system. *P. pacificus*, along with its famous distant relative *Caenorhabditis elegans*, provides us with a great opportunity to understand the finer aspects of contemporary evolutionary developmental biology (evo-devo) through comparative analyses between the two species. Specifically, *P. pacificus* has emerged has a keystone species to study developmental plasticity, a phenomenon in which organisms have the ability to give rise to alternative phenotypes on the basis of their environmental experiences during development. In this context, a wealth of knowledge has been accumulated with regards to the effects of environmental factors on developmentally plastic traits of *P. pacificus*, as well as on the underlying genetic machinery manifesting and regulating said plastic traits. However, the mediating link connecting the environmental reception to the morphological manifestation is only starting to be peered into in *P. pacificus*.

In this doctoral thesis, I, with support from my colleagues, improve the genomic reference of the orthologues of *C. elegans* genes in *P. pacificus* genome, using it to subsequently identify, using comparative spatial tomography, a dissociation between genomic and morphological homology between the two nematode species, and further explore the spatial expression of evolutionarily young genes, one of the keystone evolutionary facilitators of morphological novelty and

evolutionary diversity, in the masculine part of the gonads, providing the first example in support of the 'out-of-testis' hypothesis from nematodes. Next, I present the first developmentally resolved small RNA transcriptome of *P. pacificus*. Small non-coding RNAs have emerged as important post-transcriptional regulators, and implicated in wide-ranging biological phenomena, including development, genome protection, and environmental response, among others. Temporal resolution of the *P. pacificus* small RNAs gives us the opportunity to look at the role of conserved microRNAs in the worm's development, and explore their evolutionary divergence between *P. pacificus* and other nematodes, including *C. elegans*. In addition to being the first of their kind for *P. pacificus*, the developmental small RNA transcriptome also sets up the stage for elucidating the potential roles of small RNAs in mediating the crosstalk between environmental influence and the genetic machinery controlling developmental plasticity within the species. Lastly, I contribute in exploring how the environment itself may regulate developmental decision-making without any genetic perturbation in perfectly isogenic worm populations.

B. Zusammenfassung

Die Embryologie oder Entwicklungsbiologie wird seit Jahrhunderten eingehend untersucht, da sie Antworten auf mehrere grundlegende Fragen der Biologie bietet. Eine dieser hartnäckigen biologischen Fragen ist die der Evolution, und viele Vertreter der klassischen Evolutionstheorie haben sich auf der Suche nach Antworten und Erklärungen der vergleichenden Embryologie zugewandt. Mit der Wiederentdeckung der Mendelschen Gesetze und dem darauffolgenden Aufkommen der Genetik im frühen und mittleren 20. Jahrhundert sowie der anschließenden Identifizierung der DNA als primärer Träger genetischer Informationen über Generationen hinweg sind DNA und Gene jedoch in den Mittelpunkt der Erklärung gerückt, wie sich neue Arten aus bereits existierenden Arten entwickeln, wobei andere Bereiche wie Entwicklung und Ökologie vernachlässigt wurden. In den letzten Jahren hat jedoch eine Kombination von Faktoren – wie die (Wieder-)Verschmelzung von Evolutions- und Entwicklungsbiologie, das Aufkommen einer breiten Palette von Modellorganismen für vergleichende Entwicklungsstudien und technologische Fortschritte auf dem Gebiet der Molekularbiologie und Computerbiologie – zu einem wachsenden Bedarf geführt, die Grundlagen der Evolution und die Ursprünge der Artenvielfalt zu überdenken und neu zu bewerten. In diesem Zusammenhang hat sich der Fadenwurm *Pristionchus pacificus* als wichtiges Modellsystem herauskristallisiert. *P. pacificus* bietet uns zusammen mit seinem berühmten entfernten Verwandten *Caenorhabditis elegans* eine ausgezeichnete Gelegenheit, die feineren Aspekte der modernen evolutionären Entwicklungsbiologie (Evo-Devo) durch vergleichende Analysen zwischen den beiden Arten zu verstehen. Insbesondere hat sich *P. pacificus* als Schlüsselart für die Untersuchung der Entwicklungsplastizität herausgestellt, ein Phänomen, bei dem Organismen aufgrund ihrer Umwelterfahrungen während der Entwicklung alternative Phänotypen hervorbringen können. In diesem Zusammenhang wurde ein umfangreiches Wissen über die Auswirkungen von Umweltfaktoren auf entwicklungsplastische Merkmale von *P. pacificus* sowie über die zugrunde liegende genetische Maschinerie, die diese plastischen Merkmale manifestiert und reguliert, angesammelt. Das vermittelnde Bindeglied zwischen der Umweltaufnahme und der morphologischen Manifestation wird bei *P. pacificus* jedoch erst allmählich erforscht.

In dieser Doktorarbeit verbessere ich mit Unterstützung meiner Kollegen die

genomische Referenz der Orthologen der *C. elegans*-Gene im Genom von *P. pacificus* und nutze sie, um anschließend mithilfe der vergleichenden räumlichen Tomographie eine Dissoziation zwischen genomischer und morphologischer Homologie zwischen den beiden Fadenwurmarten festzustellen. Außerdem erforsche ich die räumliche Expression evolutionär junger Gene, einen der wichtigsten evolutionären Förderer morphologischer Neuheit und evolutionärer Vielfalt, im männlichen Teil der Gonaden und liefere damit das erste Beispiel zur Unterstützung der „Out-of-testis"-Hypothese von Fadenwürmern. Als nächstes präsentiere das erste entwicklungsmäßig aufgelöste kleine RNA-Transkriptom von *P. pacificus*. Kleine nicht-kodierende RNAs haben sich als wichtige posttranskriptionelle Regulatoren herausgestellt und sind an weitreichenden biologischen Phänomenen beteiligt, darunter Entwicklung, Schutz des Genoms und Reaktion auf Umweltbedingungen. Die zeitliche Auflösung der Expression kleiner RNAs von *P. pacificus* gibt uns die Möglichkeit, die Rolle konservierter microRNAs bei der Entwicklung des Wurms zu untersuchen und ihre evolutionäre Divergenz zwischen *P. pacificus* und anderen Fadenwürmern, einschließlich *C. elegans*, zu erforschen. Das Transkriptom der kleinen RNAs während der Entwicklung ist nicht nur das erste seiner Art für *P. pacificus*, sondern schafft auch die Grundlage für die Aufklärung der möglichen Rolle kleiner RNAs bei der Vermittlung der Wechselwirkung zwischen Umwelteinflüssen und dem genetischen Mechanismus, der die Entwicklungsplastizität innerhalb der Art steuert. Schließlich trage ich dazu bei, zu erforschen, wie die Umwelt selbst die Entwicklungsentscheidungen ohne genetische Störungen in perfekt isogenen Wurmpopulationen regulieren kann.

C. List of publications

Werner, M. S., Sieriebriennikov, B., Loschko, T., Namdeo, S., Lenuzzi, M., Dardiry, M., Renahan, T., **Sharma, D. R.**, Sommer, R. J. (2017). Environmental influence on *Pristionchus pacificus* mouth form through different culture methods. *Scientific Reports* 7, 7207. https://doi.org/10.1038/s41598-017-07455-7

Rödelpserger, C., Athanasouli, M., Lenuzzi, M., Theska, T., Sun, S., Dardiry, M., Wighard, S., Hu, W., **Sharma, D. R.**, Han, Z. (2019). Crowdsourcing and the feasibility of manual gene annotation: A pilot study in the nematode *Pristionchus pacificus*. *Scientific Reports* 9, 18789. https://doi.org/10.1038/s41598-019-55359-5

Rödelsperger, C., Ebbing, A., **Sharma, D. R.**, Okumura, M., Sommer, R. J., Korswagen, H. C. (2021). Spatial transcriptomics of nematodes identifies sperm cells as a source of genomic novelty and rapid evolution. *Molecular Biology and Evolution*, Volume 38, Issue 1, Pages 229–243, https://doi.org/10.1093/molbev/msaa207

Sharma, D. R., Röseler, W., Witte, H., Werner, M. S., Sommer, R. J (2024). Developmental small RNA transcriptomics reveals divergent evolution of the conserved microRNA miR-100 and the *let-7-complex* in nematodes. bioRxiv 2024.07.19.604269; doi: https://doi.org/10.1101/2024.07.19.604269

D. Introduction

D.1. Origins of evolutionary developmental biology

Developmental biology - the study of progressive changes in individual living beings, starting from a single cell, embryonic progression leading up to birth, and even beyond - has been a field of interest since ancient times [1, 2], with many considering the ancient Greek philosopher Aristotle as the first noted embryologist in history [1, 3]. Modern developmental biology was prefaced by the field of embryology [4], which arose as a field of study in the 18th and 19th centuries to settle the preformation *vs.* epigenesis debate. Study of embryos, along with the advent of microscopy, chemistry and histological staining methods, helped conclude that embryos are not fully formed individuals in minuscule (the preformationist tenet), but rather undergo changes starting from the morphologically simple zygotes, or in other words, undergo 'epigenesis' or 'development' [5].

In 1828, Karl Ernst von Baer, based on his observations while studying embryonic development in different species, formulated some of the earliest laws of embryology [6], which later came to be known as 'von Baer's laws of embryology' [7], and set up the stage for comparative embryology. von Baer suggested that at their earliest stages, embryos of different animals are very similar. As time progresses, embryos undergo similar developmental changes in the beginning, and only at a later time do they start undergoing developmental changes that set them apart from other animals (species-specific changes) [8].

In the second half of 19th century, contemporary evolutionary thought started looking towards embryology to solve the question of genealogy. Building up on von Baer's laws, Charles Darwin suggested that the degree of embryonic and larval similarities between different animals could explain their evolutionary closeness [9]. On the basis of their work with barnacle larvae, both Darwin and John Vaughan Thompson classified them as crustaceans (a subphylum within arthropods) and not as molluscs as was believed until then [1, 10, 11]. Similarly, Alexander Kowalevsky classified tunicates (which were also classified as molluscs until then) as chordates, on the basis of the presence of a larval notochord [12-14]. Darwin, while outlining 'descent with modification' in his theory of evolution by natural selection, explained genealogical descent on the

basis of embryonic similarities (while alterations in embryonic development would give rise to morphological modifications between homologous features, adapted for survival in different environments), thus bringing together comparative embryology and evolutionary morphology to amalgamate into evolutionary embryology [1, 9]. The field was furthered by Fritz Müller and Francis Balfour, who, based on their studies with crustaceans and vertebrates respectively, suggested that similar to their adult counterparts, embryos and larvae of different species were also subject to natural selection in response to their corresponding environments (in contrast to the views held by von Baer and Darwin, who believed early developmental stages to be unchanging) [15-17].

At around the same time, Ernst Haeckel, in his attempt to unite embryonic development with evolution, proposed the biogenetic law (also known as 'Haeckel's recapitulation theory') [18], an oppugnant theory to that of von Baer's laws of embryology. In it, Haeckel suggested that only the adults were subject to evolution by natural selection, and the different embryonic stages of a 'more evolved' species would resemble the adults of its 'primitive' ancestors, adding new features only towards its end, causing it to be more 'advanced' than its predecessors [19]. With its simplistic catchphrase, that "ontogeny recapitulates phylogeny", the biogenetic law thus rendered evolution as a linear progression with some species more advanced than others (for example, it suggested that humans are more advanced than other primates or vertebrates) [20], in contrast to the branched topography of evolution as a conduit for biodiversity, as envisioned by Darwin and von Baer [9]. Although once widely popular, and having introduced concepts that still hold their relevance today (for example, heterochrony), Haeckel's biogenetic law eventually fell out of favour among biologists, largely because of lack of evidence in favour of the biogenetic law, its pertinacity that a 'more evolved' implied its superiority over others, and in particular because of his embryonic sketches that have since been considered as fraudulent [9, 20, 21]. One of the earliest dissenters of the biogenetic law was Walter Garstang, who (in line with Müller and Balfour) argued that larval stages too were adaptable and amenable to natural selection. Furthermore, it was these larval adaptations that would set the stage for evolution of novelties in adults [17]. As such, he argued, "ontogeny does not recapitulate phylogeny, but creates it" [17, 22]. Garstang, based on his work with chordates, proposed the concept of neoteny/paedomorphosis, a mechanism by which larvae of early chordates (with sessile adult stages) with accelerated sexual maturity would evolve to give

rise to mobile vertebrates [17].

Towards the end of 19th and beginning of 20th centuries, evolutionary embryology paved way for experimental embryology, pioneered by Wilhelm Roux and later carried forward by Hans Driesch and Hans Spemann, among others, experimentalists who studied embryology analytically rather than just theoretically through observations, by performing experiments with them under controlled lab conditions [23]. Paradoxically, contrasting results of experiments with embryos rendered understanding the role of embryonic development in evolution murkier [23]. Contemporarily, publication of August Weismann's germ plasm theory (which proposed that specific 'germ' cells carried the necessary information for inheritance and subsequent development in their nuclei), rediscovery of Greogor Mendel's experiments with pea plants, Thomas Hunt Morgan's experiments with the fruit fly to study inheritance of mutations, and William Bateson's work on biological variation [9, 24-26], lead to the establishment of the field of genetics. Genetics supplanted embryology in understanding evolution, leading to the paradigm that evolution could be understood by studying the inheritance of genetic material alone. Consequently, when the Modern Synthesis of evolutionary biology (also referred to as 'neo-Darwinism') was first published in 1942 [27], it did not include embryology as a relevant field for explaining evolution [28]. Interestingly, and notably, another vertical that got dissociated in the Modern Synthesis due to a combination of genetics-centric approach, lab-oriented experimentation regime, and its closer association with embryonic development than evolution, was ecology, despite both Darwin and Weismann having acknowledged and emphasised on the importance of environmental factors in inheritance and evolution [29, 30].

The reintroduction of the importance of development and ecology in evolution is largely credited to the works of Oscar Hertwig, Richard Woltereck, Richard Goldschmidt, Gavin de Beer, Ivan Ivanovich Schmalhausen, Conrad Hal Waddington, and Stephen Jay Gould [17, 19, 31-35], among others. Hertwig, best known for his observation of sexual reproduction using sea urchin gametes, experimentally verified his observations from nature that external environmental factors, such as diet and temperature, and not just genetic determinants, were important criteria in determining the gender of embryos in many species [9, 31]. Similar observation of environmentally influenced phenotypic variations in the helmet size of the water flea *Daphnia* were made by Richard Woltereck [36, 37].

He further defined - building up on the then recently furnished terms ‘phenotype’ and ‘genotype’ by Wilhelm L. Johannsen [38] - the observed variation of a phenotype upon changing the environmental factor(s) over a range as the former’s ‘norm of reaction’ [37]. These observations, in our current times, can be categorised under the broader phenomenon known as ‘developmental plasticity’, which has become one of the central topics in understanding evolution [39]. Richard Goldschmidt, a staunch promulgator of the interplay between genetics and development, was often perceived controversially by his peers for the same [40]. He argued - based on his work with the moth *Lymantria dispar* and the fruit fly *Drosophila melanogaster* - that while local adaptations within a species had a strong genetic input, such ‘microevolutionary’ changes (he termed the gradual intraspecific changes, accumulated over time, as ‘microevolution’) were insufficient for evolution of new species. Instead, large jumps in evolution could occur by ‘macromutations’, which would either involve mutations in developmentally important loci, or by a systemic rearrangement within chromosomes. While an overwhelming majority of such mutations would result in deleterious ‘monsters’, every once in a while, by pure chance, such mutations would give rise to ‘hopeful monsters’ that might be preferred by natural selection, leading to speciation events [35, 40-42]. His understanding of a gene too was different from his contemporaries, and he considered chromosomes as the fundamental heritable unit. The relative repositioning of genes due to chromosomal rearrangements would result in changes in gene function and in quantitative differences, in contrast to the then pervasive Mendelian connotation of genes causing qualitatively dominant or recessive phenotypes [40, 42, 43]. Sir Gavin de Beer, although originally trained as a comparative evolutionary embryologist, was also well versed in experimental embryology and genetics, which he used to his advantage to amalgamate the different fields together in the first half of 20th century [17, 44]. de Beer shed light on the importance of timing in developmental processes, expanding on the concept of heterochrony first introduced by Ernst Haeckel [44]. Using Goldschmidt’s concept of developmental genes, de Beer suggested that changes in the activity of such genes could result in an ontogenic shift, resulting in heterochronic delays (such as paedomorphosis, as described by Walter Garstang) or accelerations [17, 34, 40, 44]. He further argued that such changes during development result in phylogenetic differences among species (as suggested by Garstang), rather than recapitulate them (as suggested by Haeckel) [18, 22, 34, 44]. de Beer also sought to address the long-standing question of homology in embryology and

development. Francis Balfour, who had spent the majority of his career hypothesising homology of structures across species on the basis of their origin from the same germ layer during development, was not fully certain of the hypothesis, as shown in his last works [16, 17, 45]. de Beer, based on examples from vertebrates and the fruit fly, and expanding the original hypothesis to include the genetic component, concluded that homologous structures need not necessarily arise from the same germ layer during embryogenesis, nor do they necessitate homology of the underlying genes responsible for their formation (and vice versa) [17, 46]. The lattermost aspect of de Beer's conclusion, specifically the dissociation of homology between the phenotypes and their underlying genetic architecture, forms the basis of 'developmental system drift', an important concept in modern evolutionary developmental biology [47].

Ivan I. Schmalhausen and Conrad H. Waddington, independently but thematically similarly, blended genetics, environment, development, and evolution together to expand on the Modern Synthesis of evolutionary biology [9]. Schmalhausen is best known for his theory of stabilising selection [33, 48]. According to him, adaptive traits in any species would necessitate them to undergo stabilising selection, that is, remain unperturbed by environmental variation. Genetic variation found in any given species at any point of time would permit the extremes of environmental fluctuation to unravel new traits, which can then be stabilised subsequently by natural selection, and ultimately lead to their hereditary fixation [48]. What Schmalhausen defined as 'stabilising selection' and 'hereditary fixation', Waddington called them as 'canalisation' and 'genetic assimilation' respectively, a conclusion that the latter reached through his inheritance experiments with the fruit fly [32, 48-51]. A geneticist-turned-embryologist (unlike most other scientists discussed here), Waddington suggested that a would-be adaptive trait (which would itself be genetically regulated) that was initially triggered by environmental stimulation, would in due course of evolutionary time be incorporated into the genetic machinery during development. This would, in turn, remove the need for the original environmental trigger for the manifestation of the new adaptive phenotype. He went further and suggested that once the trait has undergone 'genetic assimilation', it will be developmentally constrained, or 'canalised'. As such, subsequent variations in environmental conditions will not lead to morphological/phenotypic variations despite allowing for accumulation of changes in the genotypes within a population, as long as the aforementioned

environmental changes are not drastic [32, 51, 52]. However, despite its growing popularity, both Schmalhausen and Waddington faced resistance in their attempts to extend the Modern Synthesis [9, 48, 51].

Stephen Jay Gould, a palaeontologist and geologist by training, built up on and brought together the different themes discussed above, culminating in 'Ontogeny and Phylogeny', published in 1977, widely credited as establishing the field of evolutionary developmental biology in its current form [9, 53]. In it, along with his other works, Gould threw light on varied topics, such as von Baer's laws, Haeckel's theory of recapitulation, Goldschmidt's macromutations and hopeful monsters *vs.* Darwin's gradualism, and de Beer's developmental timing and heterochrony, among others [19, 53, 54]. In the years since the publication of 'Ontogeny and Phylogeny', a number of treatises exploring and expanding on the field further have been published, such as 'Embryos, Genes, and Evolution' by Rudolf Raff and Thomas Kaufman, 'The Shape of Life' by Rudolf Raff, 'Evolutionary Developmental Biology' by Brian Hall, and 'Developmental Plasticity and Evolution' by Mary Jane West-Eberhard, among others [55], cementing the union between the two fields in modern research pursuits. Furthermore, Gould was a regular author for popular magazines and newspapers, intertwining scientific theories with everyday examples for lay persons, making his ideas reach a wider audience [53].

D.2. Developmental system drift, phenotypic novelty, and developmental plasticity: key concepts of evo-devo

Modern evolutionary developmental biology incorporates several key concepts. Two such key concepts are developmental system drift (DSD) and phenotypic novelty [47, 56, 57]. DSD refers to the phenomenon when homologous morphological traits in closely related species have undergone divergence - over the course of evolution - in the underlying genetic network responsible for their development in those species [47]. In other words, DSD is when non-homologous genes and proteins result in the formation of homologous features/traits in different species. At first, DSD appears to be very counter-intuitive, since parsimonious logic would imply that a particular gene network, once adopted and fine-tuned to give rise to a particular morphological trait, would be kept intact over evolutionary timescales. However, DSD implies that the molecular genetic underpinnings of developmental processes can continuously undergo all

sorts of modifications, as long as they continue to interact in a phenotypically viable manner. These changes - such as gene duplication, neofunctionalization, subfunctionalization, functional redundancy, and co-option - over evolutionary timescales can give rise to new species via functional repurposing of genes and gene networks (even creating new ones) in distinct lineages [47, 56].

One of the possible outcomes of such a genetic repurposing as described above can be the appearance of phenotypic novelties [58, 59]. While divergence of existing traits in species is readily understood to be adaptive under favourable environmental conditions, and thus, understood in great detail [27, 60], how such traits come by in the first place is a central question of evolutionary biology that is relatively poorly understood [57-59]. One of the suggested mechanisms is by perturbation of the developmental capacitance of a system. Developmental capacitance refers to the robustness of developmental processes, resulting in functionally and morphologically stable phenotypes and individuals across a range of environmental and genotypic variation [61]. Developmental capacitance can facilitate buildup of genetic variation in a population, termed 'cryptic genetic variation' [58, 62, 63]. However, the definition of developmental capacitance also implies that if pushed to its threshold limits, whether genetic or environmental, the robustness of developmental processes may yield, potentially giving rise to new trait phenotypes as a result of exposure to novel environments, or strong genetic mutations (what Richard Goldschmidt would call 'macromutations'). Such alternative phenotypic traits, in the off chance that they might be adaptive in the altered environmental or genetic conditions, can then be selected for, and give rise to phenotypic novelty. Some of the earliest experimental evidence for occurrence of new traits under extreme developmental conditions come from Waddington and his colleagues [49, 50, 63, 64].

Another mechanism suggested as a facilitator of origin of phenotypic novelty is developmental plasticity, also called phenotypic plasticity. It is defined as the ability of a single genotype to give rise to different phenotypes as a response to the received environmental cues during development [39, 58, 65, 66]. Some of the earliest examples of developmental plasticity were observed by Oscar Hertwig and Richard Woltereck [31, 36, 37]. Developmental plasticity, by its definition, permits venturing beyond the widely accepted one genotype-one phenotype paradigm, and identifies environment not as a passive agent of natural selection, but an active contributor to the rise of morphological, biochemical, or even

behavioural variations, via its interactions with the developmental processes [39, 67]. Similarly, under the tenets of plasticity, development itself is understood not only as a stabilising or canalising phenomenon [32, 33]. Rather, emphasis is put on the early, undifferentiated developmental stages, wherein environment can influence the plastic genetic architecture of the organism (what Waddington would call as 'epigenetic' modifications) to result in alternative phenotypic variations [39, 51, 52, 67, 68]. Such variations can then be further refined by the combination of natural selection (through environmental conditions) and the underlying genetic architecture of the organism, leading to novelty within the phenotype [67, 69]. At the population level, with each individual member contributing with its own extent of plasticity of the genotype, there would be a wide range of available phenotypic variations, some of which might undergo genetic accommodation over evolutionary timescales, and result in speciation and evolutionary diversity [59, 67, 69]. Nature abounds with examples of phenotypic plasticity, indicating at the pervasiveness of the above-mentioned hypothesis in the wild. However, its systematic analysis has only started in the last few decades [39, 65-67, 70-74].

D.3. The study system: *Pristionchus pacificus*

Over the last three decades, the nematode *Pristionchus pacificus* has emerged as a conducive model system to study DSD and developmental plasticity. *P. pacificus*, a free-living hermaphroditic nematode, is a member of the family Diplogastridae [75]. Its evolutionary distance from the model nematode *Caenorhabditis elegans* is estimated between 130 and 200 million years [76], making it a convenient choice for comparative studies with *C. elegans*. Since the first morphological, genetic, and molecular description of the species [77], the repertoire of molecular, genomic, ecological, phylogenetic, morphological, and biochemical resources for *P. pacificus* has expanded rapidly [78-84]. From a practical standpoint, *P. pacificus*, as a self-fertilising hermaphrodite with a low propensity of spontaneous males, has the same advantages as *C. elegans* (easy and low-cost maintenance, rapid generation time, small size, and ease of biological material access for molecular analyses) [77, 85], making it amenable to biological research.

Nematode vulval development is one of the best studied and tractable examples to study DSD [47, 86]. In *P. pacificus*, adult hermaphrodites have a vulva in the mid-body region, homologous to the *C. elegans* vulva. However, developmental

analyses of the organ showed that the number of precursor cells that result in the formation of vulva in *P. pacificus* are different from those in *C. elegans* [86, 87]. Subsequent genetic analyses showed that the developmental fate of the cells and the corresponding underlying genetic network has undergone evolutionary divergence [88-95]. Another example of DSD between the two nematode species is the alternative, starvation-resistant dauer developmental stage, which, interestingly, is also an example of developmental plasticity (see below) [96, 97]. While conserved transcriptional regulators determine the decision to enter dauer or not in both species [98, 99], the underlying genetic, as well as biochemical, machinery is different [100-103].

Holistic to the study of genetic divergence, as exemplified by DSD, is the evolution of genomes. Diplogastrid nematodes, including the genus *Pristionchus*, are found in close association with beetles and have been sampled globally [78, 83, 104-109], providing a valuable phylogenetic resource of closely related species. This, in turn, has facilitated phylogenomic studies to evaluate evolutionary age and divergence of genes [110-112], allowing their categorisation on the basis of their evolutionary age as conserved, taxonomically restricted, or orphan genes [113]. Of these, orphan, *de novo* and young genes are of particular research interest in an evolutionary context, and work in *Pristionchus* and other nematodes has revealed the importance of young genes in the context of genome evolution, evolution of modes of reproduction, and architecture of neighbourhood of young genes [110-116]. Young genes have also been shown to be conduits of functional, morphological, and behavioural novelty, including examples from *P. pacificus* [102, 117-119].

P. pacificus is also a conducive model to study developmental plasticity (Fig. 1). In addition to dauer formation, *P. pacificus*, along with other species of family Diplogastridae, shows another developmentally plastic trait: a dimorphism in its buccal cavity which contains chitinous teeth-like denticles [120-122]. Adult *P. pacificus* worms have two alternative mouth morphs: a wide but shallow buccal cavity with two large and claw-like denticles, called the 'eurystomatous' morph, or a narrow and deep cavity with one small flint-like denticle, called the 'stenostomatous' morph (Fig. 1). This developmental plasticity also has a behavioural aspect, as stenostomatous worms are strictly bacterivorous, whereas eurystomatous worms can predate on other smaller nematodes to draw nutrition in addition to feeding on bacteria [123]. Although the phenotype is manifested

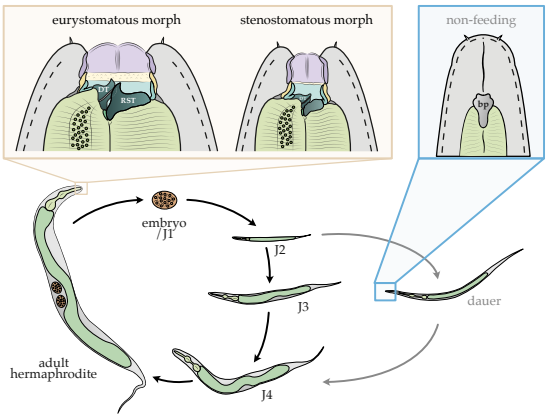


Fig. 1: Two instances of developmental plasticity - alternative development as dauers, and mouth dimorphism - from life history of *P. pacificus*. Normal ambient lab conditions result in a direct life cycle, wherein *P. pacificus* embryos progress through four juvenile stages (J1-J4) before reaching adulthood. Contrastingly, under poor environmental conditions, such as high population density and/or low food abundance, worms develop into the arrested, stress-resistant, non-feeding (shown here with a non-feeding buccal cavity; 'bp': buccal plug), and dispersal-oriented 'dauer' stage, which, on encountering favourable conditions (abundant food availability, low population density), progresses through development into the J4 stage. Additionally, *P. pacificus* shows a polyphenism in its buccal cavity, wherein adult worms can have a wide 'eurystomatous' mouth with two large teeth-like denticles (shown here with 'DT' and 'RST', for dorsal tooth and right subventral tooth, respectively) or a narrow 'stenostomatous' mouth with a single small dorsal tooth (Note that the J1 stage of *P. pacificus* is still within the egg-shell, in contrast to its homologous L1 stage in *C. elegans*, which occurs post hatching. This is an example of heterochrony, a key concept of evo-devo, for which *P. pacificus*, with *C. elegans*, can serve as a potential study system in the future; figure adapted from Theska *et. al.*, *Zoological Letters*, 2024)

during adulthood, the decision for it is made during the worm's development through juvenile stages [124, 125]. Recently, Susoy and colleagues showed that developmental plasticity facilitates phenotypic divergence using a broad range of nematode species that include the Diplogastridae family [126]. Here, on an evolutionary timescale, the authors were able to show a strong correlation between the onset of developmental plasticity in the form of mouth dimorphism, and an increase in both morphological complexity and rate of evolution (Fig. 2). Interestingly, a subsequent loss of dimorphism in secondarily monomorphic Diplogastridae, standing in as proxy for assimilated phenotypes, simultaneously showed an even further increase in rate of evolution, although the morphological complexity was reduced, indicating that developmental plasticity facilitates evolutionary divergence and promotes phenotypic innovation, even after the original trigger causing the morphological divergence has been secondarily fixed.

D.4. Environmental sensing, genetic architecture, and their mediators: epigenetic mechanisms

In order to study developmental plasticity, a model system needs three components: 1) a trait with alternative phenotypes responsive to environment, 2) fluctuating environmental conditions which are capable of inducing a change in the plastic trait, and 3) molecular machinery capable of carrying the environmentally-triggered information to the site of manifestation of plastic phenotypes of the trait. In recent years, several ongoing and completed studies have tapped into a plethora of environmental conditions, such as starvation, overcrowding/high population density, secreted metabolites, diet, and



Fig. 2: Mouthparts of diplogastrid nematodes as an example of emergence of morphological novelty. Compared to the outgroup species (in green), nematodes of the Diplogastridae family (in red and blue) show developmental plasticity in their feeding structures, resulting in individuals with alternative mouth morphologies, which correlates with an increase in structural complexity in their mouthparts and emergence of novel structures in their buccal cavities (blue: polymorphic species, red: secondarily monomorphic species; figure adapted from Susoy *et. al.*, *Elife*, 2015)

temperature, to see their effect on the mouth morphs of adult *P. pacificus* worms [100, 120, 127-130].

Similarly, a wealth of knowledge has been accumulated steadily over the years, unraveling the developmental genetic machinery underlying the worm's feeding-structure plasticity (Fig. 3) [99, 120-122, 124, 125, 131-142], centred around a genetic switch [67]. In brief: forward genetic screens identified that a gene coding for a sulfatase, subsequently named '*eud-1*' (for 'eurystomatous deficient'), led to entire worm populations being absolutely monomorphic in either direction depending on the gene's absence or over-expression, regardless of the genetic background and environmental conditions [121]. Interestingly, a sulfotransferase (a biochemical antagonist of a sulfatase), *sult-1*, showed opposite phenotypic effects to that of *eud-1* (although not as absolute), indicating that the switch-making decision is biochemically centred on sulfation and desulfation of core ligands [136, 143]. Subsequent analysis of the genetic architecture around *eud-1* found existence of a multi-gene locus that evolutionarily originated alongside mouth dimorphism in diplogastrid phylogeny, containing the sulfatase, its paralogue, and two genes coding for N-acetylglucosaminidases, all contributing to the mouth morph decision [137]. N-acetylglucosaminidases have been implicated in proper extracellular matrix formation in humans [144].

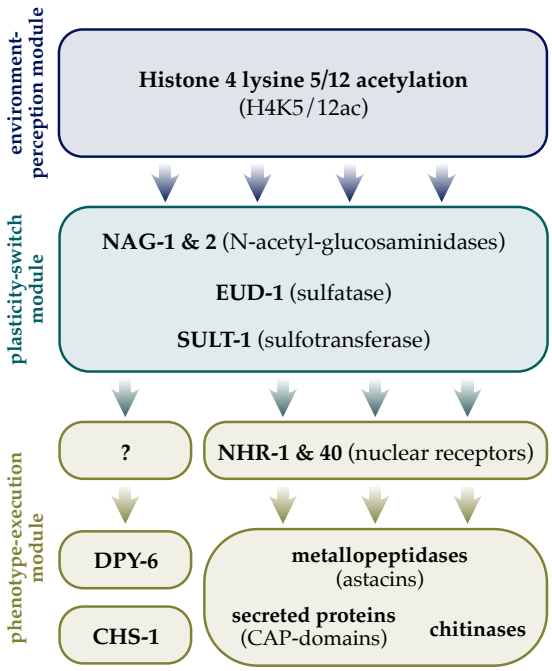


Fig. 3: Modular gene regulatory network the development of mouth morphs in *P. pacificus*. The bulk of studies pertaining to unravelling this network have focussed on genetic players involved in the phenotypic switch between the two morphs, as well as their downstream regulators and effectors involved in the manifestation of phenotypes. Recently, histone modifications have been implicated in sensing the environment and its effect on the phenotypic decision, revealing potential epigenetic regulation of the *P. pacificus* in general, and of mouth dimorphism in particular (Figure adapted from Theska *et. al.*, *Zoological Letters*, 2024)

Further studies focussing on downstream targets of *eud-1* identified two transcription factor-coding genes that are conserved between rhabditids and diplogastrids - *nhr-1* and *nhr-40* - as key regulators of downstream structural genes, such as chitinases and astacins, that likely participate in building up the mouthparts [131, 138]. Interestingly, a recent study has shown that despite their conservation across *P. pacificus* and *C. elegans*, *nhr-1* and *nhr-40* do not participate in the development of mouth parts in the latter, indicating that development of feeding structures, as in the case of vulva, has undergone developmental system drift between the two [142].

While chitin has recently been shown to contribute to the formation of denticles in *P. pacificus* [122], the role of astacins in formation of mouth parts is less clear, in part owing to the massive expansion of the family in nematodes, which renders their phenotypic analysis by genetic perturbation difficult [138, 145]. Astacins are a family of genes that encode for metalloproteases (extracellular matrix-degrading enzymes), and are known to be conserved across the animal kingdom [145]. In nematodes, astacins have mostly been implicated in cuticle shedding, a periodic phenomenon displayed by nematodes as they progress through developmental stages [146].

While environmental conditions and the gene regulatory network have been and continue to be studied in considerable depth in the context of *P. pacificus* development, currently there is a paucity of insights into the molecular machinery linking the two aspects, providing us with an understanding of how environmental signals manifest themselves at the phenotypic level via the gene regulatory network. Over the last few years, histone modifications, DNA methylation, and small RNA-mediated regulation, commonly grouped together as 'epigenetic modifications' [68], have been widely reported as molecular

mechanisms responsible for perception of environmental information, such as temperature changes, chemical and toxicological exposure, diet and nutrition, pathogenic infections, and other biotic and abiotic stresses, across the kingdom of life [147-163]. Of the three, histone modifications have recently been shown to mediate between environmental sensing and developmental decision-making in feeding plasticity of *P. pacificus* (Fig. 3) [124, 125], and enzymes regulating histone modifications have been shown to determine the sensitivity to environmental conditions and the degree of the phenotypic response, effects that can permeate multiple generations [141]. Furthermore, comparative analysis of epigenetic machinery with *C. elegans* has identified notable differences between the two species, but of which some are conserved across multiple diplogastrid species [164].

In the context of nematodes, the most well-studied DNA methylation mark, 5-methylcytosine (5mC), widely prevalent in animals as well as other kingdoms of life, is selectively absent in some nematodes [165, 166], including *P. pacificus* [167]. In recent years, another DNA methylation mark, N6-methyladenine (N-6mA) has garnered interest as a novel epigenetic modification within the scientific community. While N-6mA is an established mark in prokaryotes and protists [168-170], its presence and relevance is contestable in higher animals and plants [171, 172]. Despite the absence of methylated DNA, many organisms, such as *P. pacificus,* seem to have preserved the machinery that is responsible for 'reading', 'writing', and 'erasing' these modifications (using DNA methyl-binding proteins, methyltransferases, and methylases, respectively) [167].

Small non-coding RNAs, originally identified as regulators of developmental processes in *C. elegans*, followed by discovery of their orthologues in other species [173-175], have emerged as an epigenetic mechanism of great interest, with implications in not just development and environmental sensing, but also in ageing, RNA interference, tissue differentiation, nutrition, stress response, immunity against selfish and viral genetic elements, and gonadal fertility, among others [176-185]. Notwithstanding lineage-, phylum-, and kingdom-specific variations across organisms, at its centre, the small RNA machinery consists of two conserved components. The first component are small RNA molecules. With lengths generally between 20-30 nucleotides, small RNAs can be further classified into microRNAs, small-interfering RNAs, and PIWI-interacting RNAs, on the basis of their biogenesis and processing mechanisms. The second

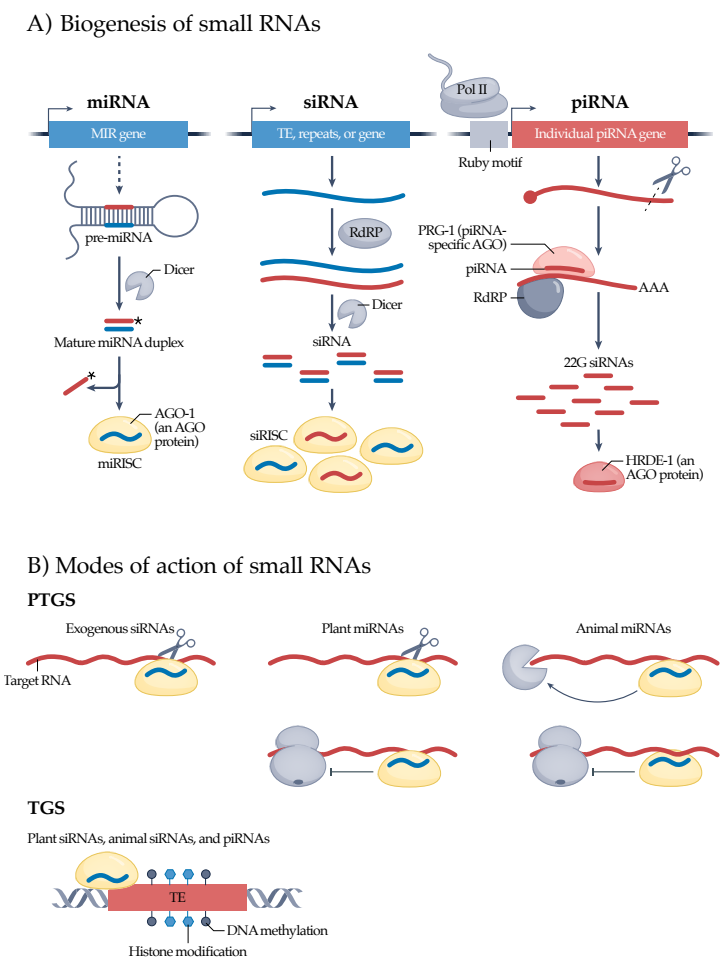


Fig. 4: Small RNAs can be classified on the basis of their biogenesis mechanisms and modes of action. A) miRNA-coding genes are transcribed to form primary miRNAs (pri-miRNAs; not shown here), which are then further processed into hairpin-shaped precursor miRNAs (pre-miRNAs) by the conserved Microprocessor complex, which are then subsequently cleaved by the conserved enzyme Dicer to form mature miRNA duplexes. One strand of the duplex (sometimes called the 'star' strand, shown here with *) is degraded, while the other is loaded on to the Argonaute (AGO) protein-containing complex, called the RNA-induced silencing complex (RISC), to form miRISC. siRNAs usually are products of cleavage of long double-stranded RNA (dsRNA) molecules, originating from transposable elements or repeat-rich regions, and subsequently formed into dsRNA by the transcriptional activity of RNA-dependent RNA polymerases (RdRPs). dsRNA is then cleaved by Dicer into small duplexes, which are then loaded on to RISC to form siRISC. Given the long dsRNA precursors, siRNAs are generally more diverse in their sequences than miRNAs. The piRNA pathway shown here is nematode-specific (other organisms generate piRNAs through a pathway known as the 'ping-pong cycle', not shown here), wherein piRNA genes are transcribed to form long piRNA precursors, which are then cleaved to form mature piRNAs. These piRNAs then associate with a piRNA-specific AGO protein, and the resultant ribonucleoprotein complex then recruits RdRPs to transcribe secondary siRNAs (also called 22Gs, given their preferential length of 22 nucleotides and triphosphated guanosines as the first nucleotide) downstream of the piRNA-binding site. B) Most small RNAs cause post-transcriptional gene silencing (PTGS) by guiding the RISC to the target RNA molecules, leading to transcript degradation or translational repression. In some cases, small RNA-guided RISC activity recruits proteins responsible for DNA methylation and/or repressive histone modification, thereby leading to transcriptional gene silencing (TGS) (Figure adapted from Chen and Rechavi, *Nature Reviews Molecular Cell Biology*, 2022).

component of the machinery are Argonaute proteins (Fig. 4). Broadly, small RNAs bind to the Argonaute proteins within the RNA-induced silencing complex (RISC). This complex, with the small RNA loaded into it, can then target messenger RNA (mRNA) molecules, based on sequence-complementarity with the small RNA molecule, resulting in post-transcriptional and transcriptional silencing of the target mRNA molecules (Fig. 4) [186-191]. A plethora of studies on multiple nematodes, particularly the model *C. elegans*, have contributed to our understanding of the diversity within small RNAs and their modes of action [183, 185, 190, 192-197]. Notably, small RNA research in *P. pacificus* has largely focussed on phylogenetic analyses across nematode species [198-201], leveraging the availability of well-annotated genomes. So far, however, the roles of small RNAs in the development and physiology of *P. pacificus* remain to be unveiled.

E. Aims of this thesis

Broadening our current understanding of the environment-phenotype interplay during developmental progression and building up on the existing pedestal of resources on which *P. pacificus* stands, the primary aim of this doctoral thesis is to further our understanding of the small RNA machinery - the only unexplored of the three molecular mechanisms that regulate the epigenetic/transcriptional landscape - in *P. pacificus*. As a resource, our knowledge of *P. pacificus* small RNAs during development would allow us to look at its ontogeny from a perspective that has been shown to be a key regulator of development in other species and was not explored in *P. pacificus* until now. Furthermore, it would set the stage for future studies wherein we can elucidate putative role/s of small RNAs as mediators of environmental information received during early development to determine the phenotypic outcomes in the context of the worm's plasticity. Lastly, it would allow for comparative developmental analysis with *C. elegans* and other nematodes, to help unveil putative conservation and/or divergence of the small RNA machinery across different species within the phylum.

To this end, I:

- Participated in improving the genome annotation of *P. pacificus*, to improve our genomic reference for more robust bioinformatic analyses, assist in reliable prediction of *P. pacificus* genes for their subsequent functional characterisation, and also help throw light on the genomic evolution
- Contributed to evaluating the transcriptome of *P. pacificus* on a spatial resolution, and its comparison with its *C. elegans* counterpart, unveiling the relationship between spatial expression and orthologous genes between the species, and a tissue-specific presence for the expression of young genes in nematodes
- Created the first developmental small RNA transcriptome of *P. pacificus*, profiling and classifying small RNAs across the worm's ontogeny, and providing a previously unavailable resource dataset for the community
- Explored the developmental dynamics of deeply conserved microRNAs, and their potential roles in regulation of ontogenic transitions by phenotypic characterisation
- Explored the evolutionary similarities and differences of deeply conserved microRNAs in other nematode species, including *C. elegans*

- Participated in exploring the interaction between phenotype and environment, by gauging the effect of the latter on the former in light of developmental plasticity of mouth morphs in *P. pacificus* worms

F. Results

The observations accumulated in the process of fulfilment of above-stated aims have resulted in the following publications:

F.1. Crowdsourcing and the feasibility of manual gene annotation: a pilot study in the nematode *Pristionchus pacificus*

C. Rödelsperger, M. Athanasouli, M. Lenuzzi, T. Theska, S. Sun, M. Dardiry, S. Wighard, W. Hu, **D. R. Sharma**, Z. Han

Scientific Reports 9, 18789 (2019). https://doi.org/10.1038/s41598-019-55359-5

Synopsis:

Organismal research has expanded our fundamental understanding of biology as a scientific field, and the use of different model systems across phylogenies has helped us to evaluate the generality or specificity of biological phenomena. Amongst animals in general, and specifically within nematodes, *C. elegans* is a prime model system to study varied frontiers of biology. Given the expansive knowledge of *C. elegans* genetics that has accumulated over the last decades of community research, we were able to leverage the existing high-quality genome annotation of *C. elegans*, along with existing *P. pacificus* transcriptome datasets, to manually improve the genome annotation of *P. pacificus*. This undertaking improved the annotation of *P. pacificus* orthologues of *C. elegans* genes by about 10%, and improved the completeness of the orthologues to 97%, while simultaneously demonstrating the feasibility of manual curation of the genome. Future studies with a focus on characterisation of *C. elegans* orthologues in *P. pacificus* will benefit from the improved genome annotation, and the current study would assist in future comparative evo-devo studies between the two species.

Own contribution:

I was involved in the manual curation of the candidate genes, as well as in the final review of the manuscript. My contribution to the publication is 10%.

F.2. Spatial transcriptomics of nematodes identifies sperm cells as a source of genomic novelty and rapid evolution

C. Rödelsperger, A. Ebbing, **D. R. Sharma**, M. Okumura, R. J. Sommer, H. C. Korswagen

Molecular Biology and Evolution, Volume 38, Issue 1, January 2021, Pages 229–243, https://doi.org/10.1093/molbev/msaa207

Synopsis:

Evolutionary divergence between phylogenetically related species can be understood in terms of genetic and transcriptomic differences between homologous anatomical structures. Furthermore, the study of evolution of gene families in deeply sampled taxons casts light on the conservation of ancient families as well as on the emergence of orphan genes. In this study, we used RNA tomography to prepare a spatially resolved transcriptome of *P. pacificus*, and used this dataset to study the evolution of genes by comparing to a similarly prepared dataset for *C. elegans*. Interestingly, our analysis revealed that one-to-one orthologous genes in both species did not show high levels of conservation in their expression-pattern topology. Additionally, phylostratigraphic analysis of regional expression patterns, leveraging the availability of sequenced genomes in closely related outgroup species, revealed an enrichment of young genes in the sperm-cell producing regions of the gonad, giving the first strong support to the 'out of testis' hypothesis - which hypothesises that sperm cells provide a hotbed for genomic innovation, supported by the testes-biased expression of young and novel genes in different model systems - within nematodes. Lastly, the dataset we created constitutes the first genome-wide transcriptional profile with spatial resolution and, thus, forms a rich resource for future studies within the nematode community.

Own contribution:

I cultured the *P. pacificus* worms and travelled to the Hubrecht Institute in Utrecht, the Netherlands, to prepare individual nematodes for sectioning and sequencing. I was also involved in the discussions during bioinformatics analyses, and in the final review of the manuscript. My contribution is 10%.

F.3. Developmental small RNA transcriptomics reveals divergent evolution of the conserved microRNA miR-100 and the *let-7-complex* in nematodes

D. R. Sharma, W. Röseler, H. Witte, M. S. Werner, R. J. Sommer

bioRxiv 2024.07.19.604269; doi: https://doi.org/10.1101/2024.07.19.604269

Synopsis:

Since their discovery as mediators of normal development in *C. elegans*, subsequently followed by the discovery of their ubiquity in all kingdoms of life, small RNAs have rapidly emerged as a field of research interest to understand post-transcriptional regulation. In addition, small RNAs have also been implicated in mediating environmental perceptions on a molecular level, as well as in transmitting these environmental perceptions across generations, thus rendering them as interesting candidates to study in the context of developmental plasticity and evolutionary developmental biology. While the genetics of developmental plasticity in *P. pacificus* has been elucidated in great detail, there still exists only a limited understanding of how the upstream environmental signals are channelled into the phenotype-determining gene regulatory network. To elucidate the role of small RNAs in its development, we developed the first developmental-stage-specific small RNA transcriptome of *P. pacificus*, throwing light on the developmental dynamics of this molecular machinery in this species for the first time, and simultaneously creating a bioinformatics resource for *P. pacificus* that was unavailable until now. Furthermore, we identified that an ancient, highly conserved microRNA - miR-100, along with another microRNA let-7 - regulate normal development in *P. pacificus* by regulating the two large protein families of collagens and metalloproteases. Finally, we also discovered that the genes coding for the two above-mentioned microRNAs constitute a genomic locus - known as the *let-7-complex* - that is conserved across bilaterians, but shows secondarily acquired alternative genomic arrangements in nematodes.

Own contribution:

I designed all and performed most of the experiments, carried out all analyses, and wrote the manuscript. My contribution is 90%.

F.4. Environmental influence on *Pristionchus pacificus* mouth form through different culture methods

M. S. Werner, B. Sieriebriennikov, T. Loschko, S. Namdeo, M. Lenuzzi, M. Dardiry, T. Renahan, **D. R. Sharma**, R. J. Sommer

Scientific Reports 7, 7207 (2017). https://doi.org/10.1038/s41598-017-07455-7

Synopsis:

Developmental plasticity, the ability of an organism to give rise to alternative phenotypes in the face of changing environmental conditions during development, is a well-documented phenomenon that unfortunately suffers from a lack of systematic analysis under lab conditions, either because of a dearth of study systems, or environmental conditions, or both. In this study, we found that, in contrast to the standard rearing on agar plates which gives rise to the predatory eurystomatous mouth morph, *P. pacificus* wild-type strain switches to highly stenostomatous populations when reared in liquid-culture conditions. Further experimentation and analyses revealed that the percentage of stenostomatous individuals in liquid-culture populations depend on various factors, such as composition of the culture as well as rotational speed of the culture flasks during incubation. Thus, we were able to conclude that the causal factor/s for alteration of mouth-form ratios in these culture conditions lie upstream of any known genetic players, and that an inexpensive and amenable array of environmental conditions can be used to procure *P. pacificus* populations with desired mouth-form ratios, without a need to alter the worm's genetic background.

Own contribution:

I participated in the qPCR experiments to analyse the expression of known mouth-form genes in the two culture conditions. My contribution is 5%.

G. Discussion and conclusions

The neo-Darwinian evolutionary theory of the mid-20th century placed genetics at the centre-stage of evolution. It posited that mutation-based changes in DNA sequences that can be inherited are the sole mediators of variation within existing species, and subsequent fodder for natural selection [202-204]. However, in recent years, there has been a strong acknowledgement of the role of development and developmental plasticity in determining the evolutionary trajectory and emergence of biological diversity/novelty [39, 56, 65-67, 71, 74, 205-207]. The roundworm *P. pacificus* has emerged as an excellent model system in this regard [73, 75]. In *P. pacificus*, availability of large-scale molecular resource datasets has led to rapid advances in elucidating the physiology, ontogeny, and evolutionary history of the species [79, 81, 116, 208]. Additionally, comparative analysis of the species alongside the model nematode *C. elegans* has highlighted important anatomical, developmental, and evolutionary differences between the two species [94, 98, 101, 114, 120, 138, 140, 142, 164, 201, 209].

Novel/young genes form one of the widely debated topics in mainstream evolutionary theory, as technical limitations of most genome-assembly and annotation pipelines render differentiation of real and artificial novel genes difficult [113, 114]. To address this issue, I, along with my colleagues, first manually improved the gene annotation of orthologues of *C. elegans* genes in *P. pacificus* genome across its entirety, taking their completeness to near saturation, and providing an important resource for future comparative genetic studies. Subsequently, using comparative spatial RNA tomography between the two species, we have shown, for the first time, a tissue-specific enrichment of expression of young genes in sperm-producing tissues of nematode gonads. Leveraging the availability of high-quality genomes in its phylogenetic proximity, recent research in *P. pacificus* has shown that not only are most novel genes no artefacts, but that they also tend to have distinct evolutionary signatures across the genome [110-112, 116]. Furthermore, novel genes are also capable of facilitating co-option of existing morphological structures for new functions, or, in an extreme scenario, facilitate morphological and behavioural novelty, thus contributing to phenotypic diversity [102, 118]. It remains to be tested whether the relative enrichment of activity of young genes in sperm-producing tissues in our work is indicative of their role/s in spermatogenesis itself, or whether it is a by-product of the overall relaxed nature of transcriptional

constraints generally found in this tissue [210]. In addition, through comparative spatial transcriptomics, we also show topological heterogeneity in the expression of orthologous genes between the two species, reminiscent of Gavin de Beer's suggestion that morphological and genomic homology need not go hand in hand [17, 44, 46]. Lastly, through this study, we also provide the first genome-wide spatial description of *P. pacificus* transcription, which will come in handy to the researcher community, especially as the number of available tissue-specific reporter constructs and modern single-cell sequencing technologies for *P. pacificus* are still limited.

Another cause for our improved understanding of environmental conditions as agents of biological variation is the advancement of molecular and computational biology, especially with respect to high-throughput sequencing [211, 212]. This, in turn, has allowed us to go from the traditional map-based cloning for characterisation of single genes to whole genome-wide analyses, and further beyond to the level of biochemical and transcriptional regulation of genomes, in a very short time window. Characterisation of epigenetic modifications has thrown light on their role/s in molecular mediation of environmental information [68, 147, 160, 213]. In this context, I present the first developmental small RNA transcriptome for *P. pacificus*, which will serve as a resource for researchers within the nematode community, particularly in the fields of comparative, developmental, and RNA biology. Using this dataset, I show that a highly conserved, arguably the most ancient microRNA [214], miR-100, is the most abundant microRNA in the juvenile stages of *P. pacificus*. Furthermore, I found that mir-100, along with another well-conserved microRNA, let-7, regulates normal developmental progression to adulthood, indicating a conservation of the joint regulatory pathways between the two microRNAs [215-219]. With the first draft of the developmental small RNA transcriptome now available, ongoing and future research will be able to shed light on the extent to which such regulatory epigenetic mechanisms are involved in the environmental perception to bring about morphological, physiological, and/or behavioural variations in *P. pacificus* worms [124, 125, 141].

Comparison of the microRNAs transcribed by the conserved locus *let-7-complex* between *P. pacificus* and *C. elegans* revealed some interesting insights with respect to their sequence similarity (or lack thereof) along with their functional divergence. While ppc-miR-100 shares sequence similarity with miR-100 from

other species, the *C. elegans* orthologues (the miR-51 family) show a strong divergence in their sequences, along with an expansion of the family into six functionally redundant microRNAs [220]. Functionally, too, miR-100 orthologues between the two species show differences, as miR-51 microRNAs in *C. elegans* are transcribed at embryonic stages, and necessary for the worm's progression to post-embryonic stages [219, 220], in contrast to the relatively mild phenotypic perturbation seen in *Ppa-miR-100* mutants in *P. pacificus*. In this light, it will be interesting to see to what extent do knockouts of the entire miR-100 family in *P. pacificus* stall its developmental progression, and at what stage/s. However, even more interesting are the comparisons with respect to the other two microRNAs of the *let-7-complex* - let-7 and lin-4 - for these two microRNAs, in addition to being the most well-known developmental regulators in *C. elegans* [173, 175, 221-223], are identical between the two species. However, their sequence identity does not directly translate to their functional similarity, as neither do the *Ppa-let-7* mutants show a fully penetrant lethal phenotype like their *C. elegans* counterparts, nor do the *Ppa-lin-4* mutants show a vulvaless phenotype incapable of egg-laying. In fact, as far as we could phenotypically quantify and qualitatively observe, *Ppa-lin-4* mutants were hardly discernible from their wild-type conspecifics. While it is too early to conclusively say if the heterochronic pathway between the two species has also undergone DSD, the functional divergence of these microRNAs between these two distantly-related nematode species reported in this thesis provides us with new prospects to study their molecular evolution.

Lastly, exploring the role of environment itself in bringing about the morphological variations, we unravel the previously unknown role of liquid culture in switching the mouth morphs of isogenic adult *P. pacificus* siblings to stenostomatous from the wild-type eurystomatous morph on lab-standard NGM-agar plates. One of the overarching aims of this thesis has been to explore the role of environment in manifesting morphological diversity through its interaction with development. With liquid culture conditions in *P. pacificus*, we showed one of the strongest examples of how environmental manipulation alone can determine alternative phenotypic outcomes without any perturbations to the genotype, as, notwithstanding genome-wide spontaneous mutations (which are relatively lower in comparison to other species [224]), *P. pacificus* worms of a given strain are genetic clones of each other, given their hermaphroditic mode of reproduction.

Taken together, the results, observations, and developments made in this thesis, along with contributions from colleagues, set the stage for future research in exploring and unravelling the interactions between environment, development, its robustness and flexibility, and morphological innovations, and will hopefully contribute to expanding our understanding of evolutionary mechanisms.

H. References

1. Gilbert, S.F., *Developmental Biology*. 2014: Sinauer Associates, Incorporated Publishers.
2. Stern, C.D., *Reflections on the past, present and future of developmental biology.* Dev Biol, 2022. **488**: p. 30-34.
3. Aristotle and A.L. Peck, *Generation of animals, with an English translation by A.L. Peck*. 1943, Cambridge: Harvard University Press.
4. Needham, J. and A. Hughes, *A History of Embryology*. 2015: Cambridge University Press.
5. Van Speybroeck, L., D. De Waele, and G. Van de Vijver, *Theories in early embryology: close connections between epigenesis, preformationism, and self-organization.* Ann N Y Acad Sci, 2002. **981**: p. 7-49.
6. von Baer, K.E., *Über Entwickelungsgeschichte der Thiere. Beobachtung und Reflexion*. Vol. Th.1. 1828, Königsberg: Bei den Gebrüdern Borntträger.
7. Henfrey, A. and T.H. Huxley, *Scientific memoirs, selected from the transactions of foreign academies of science, and from foreign journals. Natural history*. 1853, London: Taylor and Francis.
8. Barnes, M.E. *Karl Ernst von Baer's Laws of Embryology*. Embryo Project Encyclopedia, 2014.
9. Gilbert, S.F., *The morphogenesis of evolutionary developmental biology.* Int J Dev Biol, 2003. **47**(7-8): p. 467-77.
10. Thompson, J.V. and J. McGrigor, *XX. Discovery of the metamorphosis in the second type of the cirripedes, viz. the lepades, completing the natural history of these singular animals, and confirming their affinity with the crustacea.* Philosophical Transactions of the Royal Society of London, 1835. **125**: p. 355-358.
11. Darwin, C., *On the Origin of the Species by Means of Natural Selection: Or, The Preservation of Favoured Races in the Struggle for Life*. 1859: J. Murray.
12. Kovalevskij, A.O., *Entwickelungsgeschichte der einfachen Ascidien*. 1866: Kaiserliche Akad. der Wiss.
13. Darwin, C., *The Descent of man*. 1871: D. Appleton and Company.
14. Ereskovsky, A.V., *Alexander Onufrievich Kowalevsky (1840–1901)*, in *Evolutionary Developmental Biology: A Reference Guide*, L. Nuño de la Rosa and G.B. Müller, Editors. 2021, Springer International Publishing: Cham. p. 217-233.
15. Müller, F., *Für Darwin*. 1864: Wilhelm Engelmann.
16. Balfour, F.M., *A Treatise on Comparative Embryology*. 1880: Macmillan and Company.

17. Hall, B.K., *Balfour, Garstang and de Beer: The First Century of Evolutionary Embryology.* American Zoologist, 2000. **40**(5): p. 718-728.
18. Haeckel, E., *Generelle Morphologie der Organismen: Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformierte Descendenz-Theorie*. 1866, Berlin: G. Reimer.
19. Gould, S.J., *Ontogeny and phylogeny*. Ontogeny and phylogeny. 1977, Oxford, England: Harvard U Press. ix, 501-ix, 501.
20. Haeckel, E., *Anthropogenie; oder, Entwickelungsgeschichte des menschen, Keimes- und stammesgeschichte*. 3. umgearb. aufl. ed. 1877, Leipzig: W. Engelmann.
21. Richardson, M.K., et al., *There is no highly conserved embryonic stage in the vertebrates: implications for current theories of evolution and development.* Anatomy and Embryology, 1997. **196**(2): p. 91-106.
22. Garstang, W., *The Theory of Recapitulation: A Critical Re-statement of the Biogenetic Law.* Zoological Journal of the Linnean Society, 1922. **35**(232): p. 81-101.
23. Hamburger, V., *Wilhelm Roux: Visionary with a Blind Spot.* Journal of the History of Biology, 1997. **30**(2): p. 229-238.
24. Weismann, A., *Das Keimplasma: eine Theorie der Vererbung*. 1892: G. Fischer.
25. Weismann, A., *The Germ-plasm: a theory of heredity. Translated by W. Newton Parker and Harriet Rönnfeldt*. 1893, New York: Scribner.
26. Bateson, W., C.J. Clay, and F.H. Pierpont, *Materials for the study of variation treated with especial regard to discontinuity in the origin of species*. 1894, London: Macmillan.
27. Huxley, J., *Evolution: The Modern Synthesis*. 1942, Cambridge, MA: MIT Press.
28. Hamburger, V., *Embryology and the Modern Synthesis in Evolutionary Theory*, in *The Evolutionary Synthesis*, E. Mayr and W.B. Provine, Editors. 1980, Harvard University Press: Cambridge, MA and London, England. p. 97-112.
29. Darwin, C., *The Variation of Animals and Plants Under Domestication*. 1868: J. Murray.
30. Weismann, A., *Studien zur Descendenz-Theorie*. 1875, Leipzig: Engelmann.
31. Hertwig, O., *Zeit- und Streitfragen der Biologie I: Präformation oder Epigenese? Grundzüge einer Entwicklungstheorie der Organismen*. Vol. 1. 1894, Jena: G. Fischer.
32. Waddington, C.H., *CANALIZATION OF DEVELOPMENT AND THE INHERITANCE OF ACQUIRED CHARACTERS.* Nature, 1942. **150**(3811): p. 563-565.
33. Schmalhausen, I.I., *Factors of evolution: the theory of stabilizing selection*. Factors of evolution: the theory of stabilizing selection. 1949, Oxford, England:

Blakiston. xiv, 327-xiv, 327.
34. de Beer, G.R., *Embryos and Ancestors*. Third ed. 1958, Oxford: Oxford University Press.
35. Goldschmidt, R., *The Material Basis of Evolution*. 1st ed. 1982, New Haven and London: Yale University Press.
36. Woltereck, R., *Weitere experimentelle untersuchungen über Artänderung, speziell über das Wesen quantitativer Artunterschiede bei Daphniden.* Zeitschrift für induktive Abstammungs- und Vererbungslehre, 1913. **9**: p. 146.
37. Harwood, J., *Weimar Culture and Biological Theory: A Study of Richard Woltereck (1877–1944).* History of Science, 1996. **34**(3): p. 347-377.
38. Johannsen, W., *Elemente der exakten erblichkeitslehre. Deutsche wesentlich erweiterte ausgabe in fünfundzwanzig vorlesungen*. 1909, Jena: G. Fischer.
39. West-Eberhard, M.J., *Developmental Plasticity and Evolution*. 2003: Oxford University Press.
40. Zou, Y., *Richard Goldschmidt (1878–1958)*, in *Evolutionary Developmental Biology: A Reference Guide*, L. Nuño de la Rosa and G.B. Müller, Editors. 2021, Springer International Publishing: Cham. p. 249-261.
41. Carson, H.L., *A Provocative View of the Evolutionary Process*, in *Controversial Geneticist and Creative Biologist: A Critical Review of His Contributions with an Introduction by Karl von Frisch*, L.K. Piternick, Editor. 1980, Birkhäuser Basel: Basel. p. 24-26.
42. Dietrich, M.R., *From Hopeful Monsters to Homeotic Effects: Richard Goldschmidt's Integration of Development, Evolution, and Genetics.* American Zoologist, 2000. **40**(5): p. 738-747.
43. Bakken, A.H., *Views on the Nature of the Gene, the Structure and Function of the Chromosome, and Heterochromatic Heredity*, in *Controversial Geneticist and Creative Biologist: A Critical Review of His Contributions with an Introduction by Karl von Frisch*, L.K. Piternick, Editor. 1980, Birkhäuser Basel: Basel. p. 32-40.
44. Zou, Y., *Gavin de Beer (1899–1972)*, in *Evolutionary Developmental Biology: A Reference Guide*, L. Nuno de la Rosa and G. Müller, Editors. 2018, Springer International Publishing: Cham. p. 289-297.
45. Balfour, F.M., *On the structure and homologies of the germinal layers of the embryo.* Journal of Cell Science, 1880. **s2-20**(79): p. 247-273.
46. de Beer, G.R., *Homology, an Unsolved Problem*. 1971: Oxford University Press.
47. Haag, E.S. and J.R. True, *Developmental System Drift*, in *Evolutionary Developmental Biology: A Reference Guide*, L. Nuño de la Rosa and G.B. Müller,

Editors. 2021, Springer International Publishing: Cham. p. 99-110.
48. Rispoli, G. and F. D'Abramo, *Ivan I. Schmalhausen (1884–1963)*, in *Evolutionary Developmental Biology: A Reference Guide*, L. Nuno de la Rosa and G. Müller, Editors. 2019, Springer International Publishing: Cham. p. 1-13.
49. Waddington, C.H., *Genetic Assimilation of an Acquired Character.* Evolution, 1953. **7**(2): p. 118-126.
50. Waddington, C.H., *Genetic Assimilation of the Bithorax Phenotype.* Evolution, 1956. **10**(1): p. 1-13.
51. Fabris, F., *Conrad Hal Waddington (1905–1975)*, in *Evolutionary Developmental Biology: A Reference Guide*, L. Nuno de la Rosa and G. Müller, Editors. 2019, Springer International Publishing: Cham. p. 1-15.
52. Gilbert, S.F., *Diachronic Biology Meets Evo-Devo: C. H. Waddington's Approach to Evolutionary Developmental Biology.* American Zoologist, 2000. **40**(5): p. 729-737.
53. Turriziani Colonna, F., *Stephen Jay Gould (1941–2002)*, in *Evolutionary Developmental Biology: A Reference Guide*, L. Nuño de la Rosa and G.B. Müller, Editors. 2021, Springer International Publishing: Cham. p. 329-337.
54. Gould, S.J., *The return of hopeful monsters.* Natural History, 1977. **86**(6): p. 24-30.
55. Nuño de la Rosa, L. and G.B. Müller, *A Reference Guide to Evo-Devo*, in *Evolutionary Developmental Biology: A Reference Guide*, L. Nuno de la Rosa and G. Müller, Editors. 2020, Springer International Publishing: Cham. p. 1-10.
56. True, J.R. and E.S. Haag, *Developmental system drift and flexibility in evolutionary trajectories.* Evol Dev, 2001. **3**(2): p. 109-19.
57. Müller, G.B., *Developmental Innovation and Phenotypic Novelty*, in *Evolutionary Developmental Biology: A Reference Guide*, L. Nuno de la Rosa and G. Müller, Editors. 2020, Springer International Publishing: Cham. p. 1-16.
58. Moczek, A.P., *On the origins of novelty in development and evolution.* Bioessays, 2008. **30**(5): p. 432-47.
59. West-Eberhard, M.J., *Toward a Modern Revival of Darwin's Theory of Evolutionary Novelty.* Philosophy of Science, 2008. **75**(5): p. 899-908.
60. Dobzhansky, T., *Genetics and the Origin of Species*. 1982: Columbia University Press.
61. Moczek, A.P., *Developmental capacitance, genetic accommodation, and adaptive evolution.* Evol Dev, 2007. **9**(3): p. 299-305.
62. Gibson, G. and I. Dworkin, *Uncovering cryptic genetic variation.* Nat Rev Genet, 2004. **5**(9): p. 681-90.

63. Dworkin, I., *Towards a genetic architecture of cryptic genetic variation and genetic assimilation: the contribution of K. G. Bateman.* J Genet, 2005. **84**(3): p. 223-6.
64. Bateman, K.G., *The genetic assimilation of four venation phenocopies. (Received Oct. 1, 1957).* J Genet, 2005. **84**(3): p. 227-57.
65. Pigliucci, M., *Phenotypic Plasticity: Beyond Nature and Nurture*. 2001: Johns Hopkins University Press.
66. Paaby, A.B. and N.D. Testa, *Developmental Plasticity and Evolution*, in *Evolutionary Developmental Biology: A Reference Guide*, L. Nuno de la Rosa and G. Müller, Editors. 2018, Springer International Publishing: Cham. p. 1-14.
67. West-Eberhard, M.J., *Phenotypic Plasticity and the Origins of Diversity.* Annual review of ecology and systematics, 1989. **20**(70863836): p. 249-278.
68. Deans, C. and K.A. Maggert, *What do you mean, "epigenetic"?* Genetics, 2015. **199**(4): p. 887-96.
69. West-Eberhard, M.J., *Phenotypic accommodation: adaptive innovation due to developmental plasticity.* J Exp Zool B Mol Dev Evol, 2005. **304**(6): p. 610-8.
70. Wang, X. and L. Kang, *Molecular mechanisms of phase change in locusts.* Annu Rev Entomol, 2014. **59**: p. 225-44.
71. Levis, N.A., A.J. Isdaner, and D.W. Pfennig, *Morphological novelty emerges from pre-existing phenotypic plasticity.* Nat Ecol Evol, 2018. **2**(8): p. 1289-1297.
72. Weiss, L.C., *Sensory Ecology of Predator-Induced Phenotypic Plasticity.* Front Behav Neurosci, 2018. **12**: p. 330.
73. Sommer, R.J., *Phenotypic Plasticity: From Theory and Genetics to Current and Future Challenges.* Genetics, 2020. **215**(1): p. 1-13.
74. Uller, T., et al., *Twenty years on from Developmental Plasticity and Evolution: middle-range theories and how to test them.* J Exp Biol, 2024. **227**(Suppl_1).
75. Sommer, R.J., *Evolution of development in nematodes related to C. elegans.* WormBook, 2005: p. 1-17.
76. Howard, R.J., et al., *The Ediacaran origin of Ecdysozoa: integrating fossil and phylogenomic data.* Journal of the Geological Society, 2022. **179**(4).
77. Sommer, R.J., et al., *Morphological, genetic and molecular description of Pristionchus pacificus sp. n. (Nematoda : Neodiplogastridae).* Fundamental and applied nematology, 1996. **196**: p. 511-521.
78. Herrmann, M., et al., *The nematode Pristionchus pacificus (Nematoda: Diplogastridae) is associated with the oriental beetle Exomala orientalis (Coleoptera: Scarabaeidae) in Japan.* Zoolog Sci, 2007. **24**(9): p. 883-9.
79. Dieterich, C., et al., *The Pristionchus pacificus genome provides a unique perspective on nematode lifestyle and parasitism.* Nat Genet, 2008. **40**(10): p. 1193-8.

80. Witte, H., et al., *Gene inactivation using the CRISPR/Cas9 system in the nematode Pristionchus pacificus.* Dev Genes Evol, 2015. **225**(1): p. 55-62.
81. Rodelsperger, C., et al., *Single-Molecule Sequencing Reveals the Chromosome-Scale Genomic Architecture of the Nematode Model Organism Pristionchus pacificus.* Cell Rep, 2017. **21**(3): p. 834-844.
82. Theska, T., et al., *Geometric morphometrics of microscopic animals as exemplified by model nematodes.* Nat Protoc, 2020. **15**(8): p. 2611-2644.
83. Kanzaki, N., et al., *Nine new Pristionchus (Nematoda: Diplogastridae) species from China.* Zootaxa, 2021. **4943**(1): p. zootaxa 4943 1 1.
84. Igreja, C., et al., *Application of ALFA-Tagging in the Nematode Model Organisms Caenorhabditis elegans and Pristionchus pacificus.* Cells, 2022. **11**(23).
85. Stiernagle, T., *Maintenance of C. elegans.* WormBook, 2006: p. 1-11.
86. Sommer, R.J., *Evolution and development--the nematode vulva as a case study.* Bioessays, 1997. **19**(3): p. 225-31.
87. Sommer, R.J. and P.W. Sternberg, *Apoptosis and change of competence limit the size of the vulva equivalence group in Pristionchus pacificus: a genetic analysis.* Curr Biol, 1996. **6**(1): p. 52-9.
88. Clark, S.G., A.D. Chisholm, and H.R. Horvitz, *Control of cell fates in the central body region of C. elegans by the homeobox gene lin-39.* Cell, 1993. **74**(1): p. 43-55.
89. Eizinger, A. and R.J. Sommer, *The homeotic gene lin-39 and the evolution of nematode epidermal cell fates.* Science, 1997. **278**(5337): p. 452-5.
90. Sigrist, C.B. and R.J. Sommer, *Vulva formation in Pristionchus pacificus relies on continuous gonadal induction.* Dev Genes Evol, 1999. **209**(8): p. 451-9.
91. Jungblut, B. and R.J. Sommer, *Novel cell-cell interactions during vulva development in Pristionchus pacificus.* Development, 2000. **127**(15): p. 3295-303.
92. Zheng, M., et al., *Conservation and diversification of Wnt signaling function during the evolution of nematode vulva development.* Nat Genet, 2005. **37**(3): p. 300-4.
93. Tian, H., et al., *Wnt signaling induces vulva development in the nematode Pristionchus pacificus.* Curr Biol, 2008. **18**(2): p. 142-6.
94. Wang, X. and R.J. Sommer, *Antagonism of LIN-17/Frizzled and LIN-18/Ryk in nematode vulva induction reveals evolutionary alterations in core developmental pathways.* PLoS Biol, 2011. **9**(7): p. e1001110.
95. Sommer, R.J., *Evolution of regulatory networks: nematode vulva induction as an example of developmental systems drift.* Adv Exp Med Biol, 2012. **751**: p. 79-91.
96. Sommer, R.J. and A. Ogawa, *Hormone signaling and phenotypic plasticity in nematode development and evolution.* Curr Biol, 2011. **21**(18): p. R758-66.

97. Theska, T., T. Renahan, and R.J. Sommer, *Starvation resistance in the nematode Pristionchus pacificus requires a conserved supplementary nuclear receptor.* Zoological Lett, 2024. **10**(1): p. 7.
98. Ogawa, A., et al., *A conserved endocrine mechanism controls the formation of dauer and infective larvae in nematodes.* Curr Biol, 2009. **19**(1): p. 67-71.
99. Ogawa, A., et al., *Pristionchus pacificus daf-16 is essential for dauer formation but dispensable for mouth form dimorphism.* Development, 2011. **138**(7): p. 1281-4.
100. Bose, N., et al., *Complex small-molecule architectures regulate phenotypic plasticity in a nematode.* Angew Chem Int Ed Engl, 2012. **51**(50): p. 12438-43.
101. Sinha, A., R.J. Sommer, and C. Dieterich, *Divergent gene expression in the conserved dauer stage of the nematodes Pristionchus pacificus and Caenorhabditis elegans.* BMC Genomics, 2012. **13**: p. 254.
102. Mayer, M.G., et al., *The Orphan Gene dauerless Regulates Dauer Development and Intraspecific Competition in Nematodes by Copy Number Variation.* PLoS Genet, 2015. **11**(6): p. e1005146.
103. Markov, G.V., et al., *Functional Conservation and Divergence of daf-22 Paralogs in Pristionchus pacificus Dauer Development.* Mol Biol Evol, 2016. **33**(10): p. 2506-14.
104. Herrmann, M., W.E. Mayer, and R.J. Sommer, *Nematodes of the genus Pristionchus are closely associated with scarab beetles and the Colorado potato beetle in Western Europe.* Zoology (Jena), 2006. **109**(2): p. 96-108.
105. Mayer, W.E., M. Herrmann, and R.J. Sommer, *Molecular phylogeny of beetle associated diplogastrid nematodes suggests host switching rather than nematode-beetle coevolution.* BMC Evol Biol, 2009. **9**: p. 212.
106. Kanzaki, N., et al., *Pristionchus bucculentus n. sp. (Rhabditida: Diplogastridae) Isolated from a Shining Mushroom Beetle (Coleoptera: Scaphidiidae) in Hokkaido, Japan.* J Nematol, 2013. **45**(1): p. 78-86.
107. Susoy, V. and M. Herrmann, *Preferential host switching and codivergence shaped radiation of bark beetle symbionts, nematodes of Micoletzkya (Nematoda: Diplogastridae).* J Evol Biol, 2014. **27**(5): p. 889-98.
108. Kanzaki, N., et al., *Samplings of Millipedes in Japan and Scarab Beetles in Hong Kong result in five new Species of Pristionchus (Nematoda: Diplogastridae).* J Nematol, 2018. **50**(4): p. 587-610.
109. Yoshida, K., et al., *Two New Species of Pristionchus (Nematoda: Diplogastridae) from Taiwan and the Definition of the pacificus Species-Complex Sensu Stricto.* J Nematol, 2018. **50**(3): p. 355-368.
110. Prabh, N. and C. Rodelsperger, *Are orphan genes protein-coding, prediction*

artifacts, or non-coding RNAs? BMC Bioinformatics, 2016. **17**(1): p. 226.
111. Prabh, N., et al., *Deep taxon sampling reveals the evolutionary dynamics of novel gene families in Pristionchus nematodes.* Genome Res, 2018. **28**(11): p. 1664-1674.
112. Prabh, N. and C. Rodelsperger, *Multiple Pristionchus pacificus genomes reveal distinct evolutionary dynamics between de novo candidates and duplicated genes.* Genome Res, 2022. **32**(7): p. 1315-1327.
113. Rodelsperger, C., N. Prabh, and R.J. Sommer, *New Gene Origin and Deep Taxon Phylogenomics: Opportunities and Challenges.* Trends Genet, 2019. **35**(12): p. 914-922.
114. Rodelsperger, C., *Comparative Genomics of Gene Loss and Gain in Caenorhabditis and Other Nematodes.* Methods Mol Biol, 2018. **1704**: p. 419-432.
115. Rodelsperger, C., et al., *Phylotranscriptomics of Pristionchus Nematodes Reveals Parallel Gene Loss in Six Hermaphroditic Lineages.* Curr Biol, 2018. **28**(19): p. 3123-3127 e5.
116. Werner, M.S., et al., *Young genes have distinct gene structure, epigenetic profiles, and transcriptional regulation.* Genome Res, 2018. **28**(11): p. 1675-1687.
117. Santos, M.E., et al., *Taxon-restricted genes at the origin of a novel trait allowing access to a new environment.* Science, 2017. **358**(6361): p. 386-390.
118. Lightfoot, J.W., et al., *Small peptide-mediated self-recognition prevents cannibalism in predatory nematodes.* Science, 2019. **364**(6435): p. 86-89.
119. Parker, B.J. and J.A. Brisson, *A Laterally Transferred Viral Gene Modifies Aphid Wing Plasticity.* Curr Biol, 2019. **29**(12): p. 2098-2103 e5.
120. Bento, G., A. Ogawa, and R.J. Sommer, *Co-option of the hormone-signalling module dafachronic acid-DAF-12 in nematode evolution.* Nature, 2010. **466**(7305): p. 494-7.
121. Ragsdale, E.J., et al., *A developmental switch coupled to the evolution of plasticity acts through a sulfatase.* Cell, 2013. **155**(4): p. 922-33.
122. Sun, S., H. Witte, and R.J. Sommer, *Chitin contributes to the formation of a feeding structure in a predatory nematode.* Curr Biol, 2023. **33**(1): p. 15-27 e6.
123. Wilecki, M., et al., *Predatory feeding behaviour in Pristionchus nematodes is dependent on phenotypic plasticity and induced by serotonin.* J Exp Biol, 2015. **218**(Pt 9): p. 1306-13.
124. Werner, M.S., et al., *Author Correction: Histone 4 lysine 5/12 acetylation enables developmental plasticity of Pristionchus mouth form.* Nat Commun, 2023. **14**(1): p. 2837.
125. Werner, M.S., et al., *Histone 4 lysine 5/12 acetylation enables developmental*

plasticity of Pristionchus mouth form. Nat Commun, 2023. **14**(1): p. 2095.
126. Susoy, V., et al., *Rapid diversification associated with a macroevolutionary pulse of developmental plasticity.* Elife, 2015. **4**.
127. Werner, M.S., et al., *Adult Influence on Juvenile Phenotypes by Stage-Specific Pheromone Production.* iScience, 2018. **10**: p. 123-134.
128. Akduman, N., et al., *Bacterial vitamin B(12) production enhances nematode predatory behavior.* ISME J, 2020. **14**(6): p. 1494-1507.
129. Akduman, N., et al., *Correction: Bacterial vitamin B(12) production enhances nematode predatory behavior.* ISME J, 2020. **14**(7): p. 1911.
130. Lenuzzi, M., et al., *Influence of environmental temperature on mouth-form plasticity in Pristionchus pacificus acts through daf-11-dependent cGMP signaling.* J Exp Zool B Mol Dev Evol, 2023. **340**(2): p. 214-224.
131. Kieninger, M.R., et al., *The Nuclear Hormone Receptor NHR-40 Acts Downstream of the Sulfatase EUD-1 as Part of a Developmental Plasticity Switch in Pristionchus.* Curr Biol, 2016. **26**(16): p. 2174-9.
132. Serobyan, V., et al., *Chromatin remodelling and antisense-mediated up-regulation of the developmental switch gene eud-1 control predatory feeding plasticity.* Nat Commun, 2016. **7**: p. 12337.
133. Sieriebriennikov, B., et al., *The Role of DAF-21/Hsp90 in Mouth-Form Plasticity in Pristionchus pacificus.* Mol Biol Evol, 2017. **34**(7): p. 1644-1653.
134. Bui, L.T., N.A. Ivers, and E.J. Ragsdale, *A sulfotransferase dosage-dependently regulates mouthpart polyphenism in the nematode Pristionchus pacificus.* Nat Commun, 2018. **9**(1): p. 4119.
135. Bui, L.T., N.A. Ivers, and E.J. Ragsdale, *Author Correction: A sulfotransferase dosage-dependently regulates mouthpart polyphenism in the nematode Pristionchus pacificus.* Nat Commun, 2018. **9**(1): p. 4835.
136. Namdeo, S., et al., *Two independent sulfation processes regulate mouth-form plasticity in the nematode Pristionchus pacificus.* Development, 2018. **145**(13).
137. Sieriebriennikov, B., et al., *A Developmental Switch Generating Phenotypic Plasticity Is Part of a Conserved Multi-gene Locus.* Cell Rep, 2018. **23**(10): p. 2835-2843 e4.
138. Sieriebriennikov, B., et al., *Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell.* PLoS Genet, 2020. **16**(4): p. e1008687.
139. Sun, S., et al., *The oscillating Mucin-type protein DPY-6 has a conserved role in nematode mouth and cuticle formation.* Genetics, 2022. **220**(3).
140. Ishita, Y., et al., *Co-option of an Astacin Metalloprotease Is Associated with an*

Evolutionarily Novel Feeding Morphology in a Predatory Nematode. Mol Biol Evol, 2023. **40**(12).
141. Levis, N.A. and E.J. Ragsdale, *A histone demethylase links the loss of plasticity to nongenetic inheritance and morphological change.* Nat Commun, 2023. **14**(1): p. 8439.
142. Theska, T. and R.J. Sommer, *Feeding-structure morphogenesis in "rhabditid" and diplogastrid nematodes is not controlled by a conserved genetic module.* Evol Dev, 2024. **26**(2): p. e12471.
143. Igreja, C. and R.J. Sommer, *The Role of Sulfation in Nematode Development and Phenotypic Plasticity.* Front Mol Biosci, 2022. **9**: p. 838148.
144. Beesley, C.E., et al., *Molecular defects in Sanfilippo syndrome type B (mucopolysaccharidosis IIIB).* J Inherit Metab Dis, 2005. **28**(5): p. 759-67.
145. Gomis-Ruth, F.X., S. Trillo-Muyo, and W. Stocker, *Functional and structural insights into astacin metallopeptidases.* Biol Chem, 2012. **393**(10): p. 1027-41.
146. Suzuki, M., et al., *Metalloproteases with EGF, CUB, and thrombospondin-1 domains function in molting of Caenorhabditis elegans.* Biol Chem, 2004. **385**(6): p. 565-8.
147. Alvarado, S., et al., *The dynamic nature of DNA methylation: a role in response to social and seasonal variation.* Integr Comp Biol, 2014. **54**(1): p. 68-76.
148. Nielsen, C.H., A. Larsen, and A.L. Nielsen, *DNA methylation alterations in response to prenatal exposure of maternal cigarette smoking: A persistent epigenetic impact on health from maternal lifestyle?* Arch Toxicol, 2016. **90**(2): p. 231-45.
149. Villalobos-Escobedo, J.M., A. Herrera-Estrella, and N. Carreras-Villasenor, *The interaction of fungi with the environment orchestrated by RNAi.* Mycologia, 2016. **108**(3): p. 556-71.
150. Huang, X., et al., *Rapid response to changing environments during biological invasions: DNA methylation perspectives.* Mol Ecol, 2017. **26**(23): p. 6621-6633.
151. Frohlich, K.S. and S. Gottesman, *Small Regulatory RNAs in the Enterobacterial Response to Envelope Damage and Oxidative Stress.* Microbiol Spectr, 2018. **6**(4).
152. Godwin, J. and S. Farrona, *Plant Epigenetic Stress Memory Induced by Drought: A Physiological and Molecular Perspective.* Methods Mol Biol, 2020. **2093**: p. 243-259.
153. McCaw, B.A., T.J. Stevenson, and L.T. Lancaster, *Epigenetic Responses to Temperature and Climate.* Integr Comp Biol, 2020. **60**(6): p. 1469-1480.
154. Zhao, J., et al., *Plant Responses to Heat Stress: Physiology, Transcription, Noncoding RNAs, and Epigenetics.* Int J Mol Sci, 2020. **22**(1).

155. Li, S., et al., *Histone Acetylation Changes in Plant Response to Drought Stress.* Genes (Basel), 2021. **12**(9).
156. Qin, W., B.P. Scicluna, and T. van der Poll, *The Role of Host Cell DNA Methylation in the Immune Response to Bacterial Infection.* Front Immunol, 2021. **12**: p. 696280.
157. Kang, H., et al., *Histone modification and chromatin remodeling in plant response to pathogens.* Front Plant Sci, 2022. **13**: p. 986940.
158. Lai, Y., et al., *Regulatory Roles of Histone Modifications in Filamentous Fungal Pathogens.* J Fungi (Basel), 2022. **8**(6).
159. Li, D., et al., *Diet-gut microbiota-epigenetics in metabolic diseases: From mechanisms to therapeutics.* Biomed Pharmacother, 2022. **153**: p. 113290.
160. Liu, Y., et al., *Dynamic regulation of DNA methylation and histone modifications in response to abiotic stresses in plants.* J Integr Plant Biol, 2022. **64**(12): p. 2252-2274.
161. Nwanade, C.F., et al., *DNA Methylation Variation Is a Possible Mechanism in the Response of Haemaphysalis longicornis to Low-Temperature Stress.* Int J Mol Sci, 2022. **23**(23).
162. Perrella, G., I. Baurle, and M. van Zanten, *Epigenetic regulation of thermomorphogenesis and heat stress tolerance.* New Phytol, 2022. **234**(4): p. 1144-1160.
163. Singh, R.K., et al., *Regulation of small RNA-mediated high temperature stress responses in crop plants.* Plant Cell Rep, 2022. **41**(3): p. 765-773.
164. Brown, A.L., et al., *Characterization of the Pristionchus pacificus "epigenetic toolkit" reveals the evolutionary loss of the histone methyltransferase complex PRC2.* Genetics, 2024. **227**(1).
165. Simpson, V.J., T.E. Johnson, and R.F. Hammen, *Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging.* Nucleic Acids Res, 1986. **14**(16): p. 6711-9.
166. Gao, F., et al., *Differential DNA methylation in discrete developmental stages of the parasitic nematode Trichinella spiralis.* Genome Biol, 2012. **13**(10): p. R100.
167. Gutierrez, A. and R.J. Sommer, *Evolution of dnmt-2 and mbd-2-like genes in the free-living nematodes Pristionchus pacificus, Caenorhabditis elegans and Caenorhabditis briggsae.* Nucleic Acids Res, 2004. **32**(21): p. 6388-96.
168. Ehrlich, M., et al., *DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine.* Nucleic Acids Res, 1985. **13**(4): p. 1399-412.
169. Beh, L.Y., et al., *Identification of a DNA N6-Adenine Methyltransferase*

Complex and Its Impact on Chromatin Organization. Cell, 2019. **177**(7): p. 1781-1796 e25.
170. Sheng, Y., et al., *Semiconservative transmission of DNA N (6)-adenine methylation in a unicellular eukaryote.* Genome Res, 2024. **34**(5): p. 740-756.
171. Douvlataniotis, K., et al., *No evidence for DNA N (6)-methyladenine in mammals.* Sci Adv, 2020. **6**(12): p. eaay3335.
172. Boulias, K. and E.L. Greer, *Means, mechanisms and consequences of adenine methylation in DNA.* Nat Rev Genet, 2022. **23**(7): p. 411-428.
173. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. **75**(5): p. 843-54.
174. Pasquinelli, A.E., et al., *Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA.* Nature, 2000. **408**(6808): p. 86-9.
175. Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.* Nature, 2000. **403**(6772): p. 901-6.
176. Giraldez, A.J., et al., *MicroRNAs regulate brain morphogenesis in zebrafish.* Science, 2005. **308**(5723): p. 833-8.
177. Giraldez, A.J., et al., *Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs.* Science, 2006. **312**(5770): p. 75-9.
178. Brennecke, J., et al., *Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila.* Cell, 2007. **128**(6): p. 1089-103.
179. Batista, P.J., et al., *PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans.* Mol Cell, 2008. **31**(1): p. 67-78.
180. Han, T., et al., *26G endo-siRNAs regulate spermatogenic and zygotic gene expression in Caenorhabditis elegans.* Proc Natl Acad Sci U S A, 2009. **106**(44): p. 18674-9.
181. Petrelli, A., et al., *Sequential analysis of multistage hepatocarcinogenesis reveals that miR-100 and PLK1 dysregulation is an early event maintained along tumor progression.* Oncogene, 2012. **31**(42): p. 4517-26.
182. Zheng, Y.S., et al., *MiR-100 regulates cell differentiation and survival by targeting RBSP3, a phosphatase-like tumor suppressor in acute myeloid leukemia.* Oncogene, 2012. **31**(1): p. 80-92.
183. Buck, A.H., et al., *Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity.* Nat Commun, 2014. **5**: p. 5488.
184. Huang, J., et al., *MicroRNA-100 inhibits osteosarcoma cell proliferation by targeting Cyr61.* Tumour Biol, 2014. **35**(2): p. 1095-100.

185. Buck, A.H., et al., *Erratum: Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity.* Nat Commun, 2015. **6**: p. 8772.
186. Moazed, D., *Small RNAs in transcriptional gene silencing and genome defence.* Nature, 2009. **457**(7228): p. 413-20.
187. Sarkies, P. and E.A. Miska, *Small RNAs break out: the molecular cell biology of mobile small RNAs.* Nat Rev Mol Cell Biol, 2014. **15**(8): p. 525-35.
188. Swarts, D.C., et al., *The evolutionary journey of Argonaute proteins.* Nat Struct Mol Biol, 2014. **21**(9): p. 743-53.
189. Czech, B., et al., *piRNA-Guided Genome Defense: From Biogenesis to Silencing*. Annu Rev Genet, 2018. **52**: p. 131-157.
190. Ketting, R.F. and L. Cochella, *Concepts and functions of small RNA pathways in C. elegans.* Curr Top Dev Biol, 2021. **144**: p. 45-89.
191. Chen, X. and O. Rechavi, *Plant and animal small RNA communications between cells and organisms.* Nat Rev Mol Cell Biol, 2022. **23**(3): p. 185-203.
192. Wang, J., et al., *Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles.* Genome Res, 2011. **21**(9): p. 1462-77.
193. Srinivasan, J., et al., *The draft genome and transcriptome of Panagrellus redivivus are shaped by the harsh demands of a free-living lifestyle.* Genetics, 2013. **193**(4): p. 1279-95.
194. Sarkies, P., et al., *Ancient and novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages.* PLoS Biol, 2015. **13**(2): p. e1002061.
195. Zamanian, M., et al., *Release of Small RNA-containing Exosome-like Vesicles from the Human Filarial Parasite Brugia malayi.* PLoS Negl Trop Dis, 2015. **9**(9): p. e0004069.
196. Ambros, V. and G. Ruvkun, *Recent Molecular Genetic Explorations of Caenorhabditis elegans MicroRNAs.* Genetics, 2018. **209**(3): p. 651-673.
197. Marks, N.D., et al., *Profiling microRNAs through development of the parasitic nematode Haemonchus identifies nematode-specific miRNAs that suppress larval development.* Sci Rep, 2019. **9**(1): p. 17594.
198. de Wit, E., et al., *Repertoire and evolution of miRNA genes in four divergent nematode species.* Genome Res, 2009. **19**(11): p. 2064-74.
199. Ahmed, R., et al., *Conserved miRNAs are candidate post-transcriptional regulators of developmental arrest in free-living and parasitic nematodes.* Genome Biol Evol, 2013. **5**(7): p. 1246-60.

200. Holz, A. and A. Streit, *Gain and Loss of Small RNA Classes-Characterization of Small RNAs in the Parasitic Nematode Family Strongyloididae.* Genome Biol Evol, 2017. **9**(10): p. 2826-2843.
201. Beltran, T., et al., *Comparative Epigenomics Reveals that RNA Polymerase II Pausing and Chromatin Domain Organization Control Nematode piRNA Biogenesis.* Dev Cell, 2019. **48**(6): p. 793-810 e6.
202. Dawkins, R., *The Selfish Gene*. 1976: Oxford University Press.
203. Mayr, E., *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. 1982: Belknap Press.
204. Mayr, E. and W.B. Provine, *The Evolutionary Synthesis: Perspectives on the Unification of Biology*. 1998: Harvard University Press.
205. Muller, G.B., *Evo-devo: extending the evolutionary synthesis.* Nat Rev Genet, 2007. **8**(12): p. 943-9.
206. Laland, K.N., et al., *The extended evolutionary synthesis: its structure, assumptions and predictions.* Proc Biol Sci, 2015. **282**(1813): p. 20151019.
207. Skinner, M.K. and E.E. Nilsson, *Role of environmentally induced epigenetic transgenerational inheritance in evolutionary biology: Unified Evolution Theory.* Environ Epigenet, 2021. **7**(1): p. dvab012.
208. Baskaran, P., et al., *Ancient gene duplications have shaped developmental stage-specific expression in Pristionchus pacificus.* BMC Evol Biol, 2015. **15**: p. 185.
209. Sinha, A., et al., *Genome-wide analysis of trans-splicing in the nematode Pristionchus pacificus unravels conserved gene functions for germline and dauer development in divergent operons.* RNA, 2014. **20**(9): p. 1386-97.
210. Kaessmann, H., *Origins, evolution, and phenotypic impact of new genes.* Genome Res, 2010. **20**(10): p. 1313-26.
211. Reuter, J.A., D.V. Spacek, and M.P. Snyder, *High-throughput sequencing technologies.* Mol Cell, 2015. **58**(4): p. 586-97.
212. Hu, T., et al., *Next-generation sequencing technologies: An overview.* Hum Immunol, 2021. **82**(11): p. 801-811.
213. Ashe, A., et al., *piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans.* Cell, 2012. **150**(1): p. 88-99.
214. Grimson, A., et al., *Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals.* Nature, 2008. **455**(7217): p. 1193-7.
215. Christodoulou, F., et al., *Ancient animal microRNAs and the evolution of tissue identity.* Nature, 2010. **463**(7284): p. 1084-8.
216. Rubio, M. and X. Belles, *Subtle roles of microRNAs let-7, miR-100 and miR-125 on wing morphogenesis in hemimetabolan metamorphosis.* J Insect Physiol, 2013.

59(11): p. 1089-94.
217. Peng, T., et al., *Reduced abundance of the CYP6CY3-targeting let-7 and miR-100 miRNAs accounts for host adaptation of Myzus persicae nicotianae.* Insect Biochem Mol Biol, 2016. **75**: p. 89-97.
218. Gendron, C.M. and S.D. Pletcher, *MicroRNAs mir-184 and let-7 alter Drosophila metabolism and longevity.* Aging Cell, 2017. **16**(6): p. 1434-1438.
219. Dexheimer, P.J., J. Wang, and L. Cochella, *Two MicroRNAs Are Sufficient for Embryonic Patterning in C. elegans.* Curr Biol, 2020. **30**(24): p. 5058-5065 e5.
220. Shaw, W.R., et al., *The conserved miR-51 microRNA family is redundantly required for embryonic development and pharynx attachment in Caenorhabditis elegans.* Genetics, 2010. **185**(3): p. 897-905.
221. Ambros, V. and H.R. Horvitz, *The lin-14 locus of Caenorhabditis elegans controls the time of expression of specific postembryonic developmental events.* Genes Dev, 1987. **1**(4): p. 398-414.
222. Arasu, P., B. Wightman, and G. Ruvkun, *Temporal regulation of lin-14 by the antagonistic action of two other heterochronic genes, lin-4 and lin-28.* Genes Dev, 1991. **5**(10): p. 1825-33.
223. Wightman, B., I. Ha, and G. Ruvkun, *Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans.* Cell, 1993. **75**(5): p. 855-62.
224. Molnar, R.I., et al., *Mutation rates and intraspecific divergence of the mitochondrial genome of Pristionchus pacificus.* Mol Biol Evol, 2011. **28**(8): p. 2317-26.

I. Appendix

SCIENTIFIC REPORTS
nature research

OPEN

Crowdsourcing and the feasibility of manual gene annotation: A pilot study in the nematode *Pristionchus pacificus*

Christian Rödelsperger*, Marina Athanasouli, Maša Lenuzzi, Tobias Theska, Shuai Sun, Mohannad Dardiry, Sara Wighard, Wen Hu, Devansh Raj Sharma & Ziduan Han

Nematodes such as *Caenorhabditis elegans* are powerful systems to study basically all aspects of biology. Their species richness together with tremendous genetic knowledge from *C. elegans* facilitate the evolutionary study of biological functions using reverse genetics. However, the ability to identify orthologs of candidate genes in other species can be hampered by erroneous gene annotations. To improve gene annotation in the nematode model organism *Pristionchus pacificus*, we performed a genome-wide screen for *C. elegans* genes with potentially incorrectly annotated *P. pacificus* orthologs. We initiated a community-based project to manually inspect more than two thousand candidate loci and to propose new gene models based on recently generated Iso-seq and RNA-seq data. In most cases, misannotation of *C. elegans* orthologs was due to artificially fused gene predictions and completely missing gene models. The community-based curation raised the gene count from 25,517 to 28,036 and increased the single copy ortholog completeness level from 86% to 97%. This pilot study demonstrates how even small-scale crowdsourcing can drastically improve gene annotations. In future, similar approaches can be used for other species, gene sets, and even larger communities thus making manual annotation of large parts of the genome feasible.

How well can biological knowledge be transferred across species? Are biological functions carried out by the same genes in different organisms? How fast do regulatory networks diverge? In order to address these fundamental questions, more than 20 years ago, the nematode *Pristionchus pacificus* has been introduced as a so-called "satellite" model organism to one of the most successful animal model systems, *Caenorhabditis elegans*[1,2]. Since then, several comparative studies in developmental and ecological contexts have highlighted the importance of developmental system drift as a concept in evolution[3] and have demonstrated that the divergence between *Pristionchus* and *Caenorhabditis* was accompanied by extensive chemical[4–6], genic[7–9], and morphological[10–12] innovations. The establishment of multiple genetic[13,14] and genomic tools and resources[15,16] by Sommer and colleagues motivated an increasing number of independent groups to adapt *P. pacificus* as a model system for comparative studies at a mechanistic level[17–21]. However, reverse genetic approaches based on candidate genes with known functions in *C. elegans*[22,23] have been hampered not only by the huge amount of lineage-specific duplications[23–26], but also by missing and incorrect gene annotations. Traditionally, protein-coding genes are annotated by gene prediction algorithms that model general sequence features of transcription and translation start and end sites, as well as splicing signals[27–29]. This can be complemented with evidence based approaches using transcriptomic and protein homology data[30,31]. While automated annotation pipelines perform reasonably well to be useful for genetic screens[32–34] and evolutionary genomic analyses[35–37], their outcomes by far do not meet the standards of the gene annotations from classical model organisms such as *C. elegans*, *Drosophila melanogaster*, and *Mus musculus* that have been curated over decades by a large research community[38]. In order to make the *P. pacificus* system more tractable for researchers without extensive genomic and phylogenetic expertise, we need to minimize the discrepancy in gene annotation quality between *C. elegans* and *P. pacificus*. To this end, we employed an integrative approach using comparative genomic and transcriptomic data combined with crowdsourcing to improve the *P. pacificus* annotations of *C. elegans* homologs and orthologs. First, we carry out a comparative assessment of 22

Max Planck Institute for Developmental Biology, Department for Integrative Evolutionary Biology, Max-Planck-Ring 9, 72076, Tübingen, Germany. *email: christian.roedelsperger@tuebingen.mpg.de

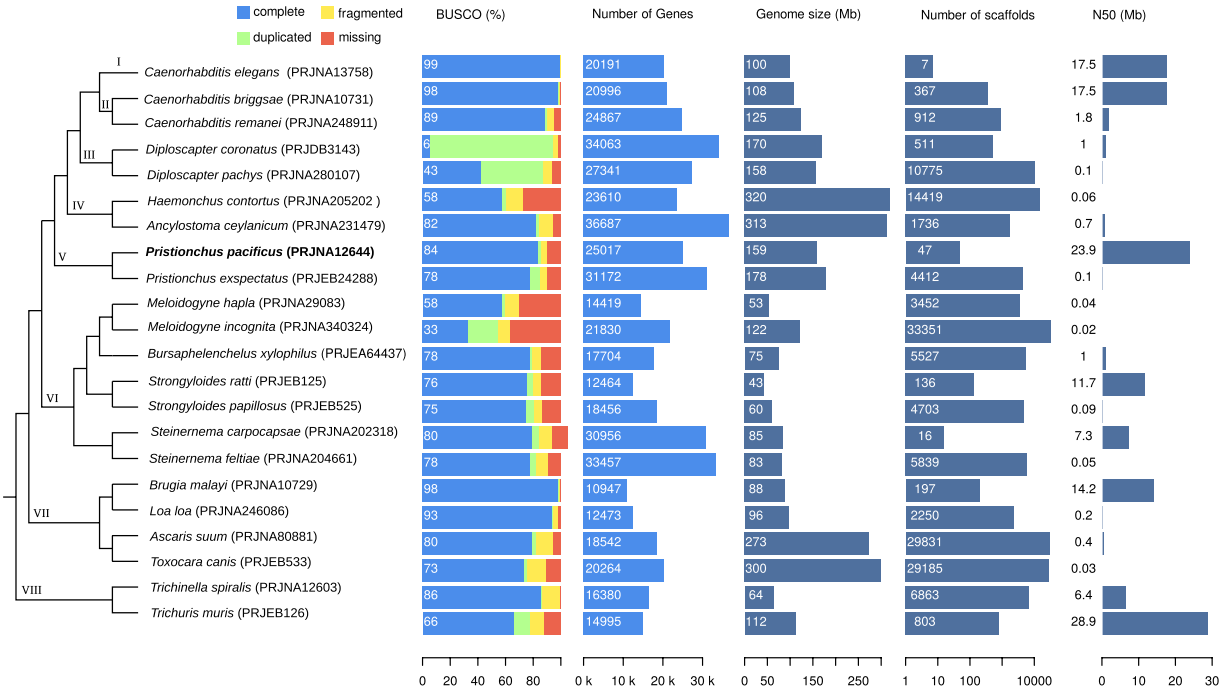


Figure 1. Comparative assessment of nematode genome quality. Genomic data for 22 nematode species was obtained from WormBase ParaSite (release WBPS13) and evaluated based on completeness level of gene annotations and genome assembly contiguity. The barplots show the results of a benchmarking of single copy orthologs (BUSCO[40]) analysis, the number of genes, genome sizes, number of scaffolds, and the N50 measure of assembly contiguity. The genome and annotations of *P. pacificus* exhibit an overall comparatively high quality. The schematic phylogeny is based on phylogenomic analysis of 108 nematodes[39], Roman numerals indicate phylostrata that are used for further analysis.

nematode genomes and demonstrate that *P. pacificus* has one of the best available nematode genomes. Second, we perform a genomewide screen for *C. elegans* genes where homologs and orthologs are not or incorrectly annotated in the *P. pacificus* genome. Third, a community-based manual curation of suspicious gene models reveals thousands of hidden orthologs and missing homologs. This pilot study can be extended to even larger gene sets and communities possibly employing citizen scientists, which would raise the quality of gene annotations to the next level[38].

Results

The quality of nematode draft genomes is highly heterogeneous. To obtain a general overview of the current status of nematode genome quality, we analyzed assemblies and gene annotations of 22 species (Fig. 1). The species were arbitrarily selected to span the diversity of the nematode phylum[39]. We will further use this taxon sampling to perform an analysis of gene age, i.e. phylostratigraphic analysis where each phylostratum is defined by at least two outgroup species to minimize the effect of species-specific gene loss. Nematode genomes range in size between 43 and 320 Mb and contain between 11 and 37 thousand annotated protein-coding genes (Fig. 1). Analyses of assembly features and gene annotations indicate a wide range of qualitative variability. Some genomes are assembled and scaffolded to the level of chromosomes with high degrees of contiguity (the N50 value which is a measure of genome assembly contiguity is up to 29 Mb) whereas others are largely fragmented into up to 33 thousand scaffolds with N50 values below 0.1 Mb (Fig. 1). Similarly, analyses of completeness levels based on benchmarking univeral single copy orthologs (BUSCO[40]) reveal substantial amount of either missing or duplicated genes and it is not totally clear to what extent these differences are of biological or technical nature[41]. In the case of *Diploscapter coronatus*, the apparent high fraction of duplicated genes could either be explained by hybridization of two divergent lineages or a whole genome duplication[42]. The genome of *P. pacificus*, which was generated by assembly from single-molecule, long-read sequencing data and scaffolding with the help of a genetic linkage map[15], shows one of the highest levels of contiguity (47 scaffolds, N50 = 24 Mb). Gene annotations were generated by the MAKER2 pipeline[30,31] which combined gene prediction algorithms, transcriptome data, and protein homology data from other *Pristionchus* species[11,15,43]. The completeness level of gene annotations (BUSCO completeness: 84%) is in the upper range when compared to most other nematode genomes (median 78%, interquartile range (IQR): 68–85%, Fig. 1). This demonstrates the relatively high quality of the current *P. pacificus* assembly and gene annotations.

Complementary genome and transcriptomes reveal potentially missing gene models. The completeness analysis as implemented in the software BUSCO[40] can also be applied to the raw genome assembly of *P. pacificus*. This yielded a combined completeness value of 93% (complete single copy and duplicates) as

Data set	BUSCO (%)				Ref
	Complete Single Copy (+Duplicates)	Duplicate	Fragmented	Missing	
Genome assembly (El Paco assembly)	91.6 (92.9)	1.3	3.1	4.0	[15]
El Paco annotation v1/WS268	84.0 (85.8)	1.8	4.3	9.9	[15]
de novo transcriptome assembly	59.1 (97.1)	38.0	2.6	0.3	[16]
Iso-Seq assembly	48.0 (73.3)	25.3	10.9	15.8	[44]
El Paco annotation v2	95.4 (97.1)	1.7	2.0	0.9	this study

Table 1. Completeness analysis of different *P. pacificus* data set. The high level of duplicates in the two transcriptomic data sets is due to the presence of isoforms.

compared to 86% for the *P. pacificus* gene annotations and indicates towards the presence of incorrectly annotated or missing *C. elegans* orthologs in the genome of *P. pacificus*. Moreover, the fact that a recent *de novo* transcriptome assembly that was based on a strand-specific RNA-seq data set exhibited an even higher combined completeness level of 97% (Table 1) demonstrates even further room for improvement[16]. Finally, single-molecule, long-read transcriptome sequencing data were recently generated for *P. pacificus* which allows a much more accurate definition of gene structures from reference alignments of single reads[44]. However, neither transcriptomic data set was available when the existing gene annotations (version: El Paco annotation v1/WormBase release: WS268) were generated and they could still be used for further improvement.

To systematically identify potentially missing genes in the *P. pacificus* genome, we searched for *C. elegans* genes lacking homologs in the current *P. pacificus* gene annotations (BLASTP e-value $< 10^{-5}$) but having a matching open reading frame in the *de novo* transcriptome assembly (Fig. 2a). While 12,504 (62%) *C. elegans* genes had BLASTP hits in both data sets, 634 (3%) *C. elegans* genes showed only BLASTP hits against the current gene annotations suggesting that these genes are properly annotated but are expressed so weakly that they were not captured in the transcriptome assembly of mixed-stage cultures[45]. Similarly, we identified 526 (3%) *C. elegans* genes that were only found in the transcriptome assembly and therefore represent candidates for missing gene annotations.

Community-based curation identifies missing genes in the *P. pacificus* genome. In order to improve the existing gene annotations, we chose to manually inspect and classify all 526 missing gene candidates in the *P. pacificus* genome browser (http://www.pristionchus.org). Thereby, we recruited and trained colleagues as community annotators, who would be capable to classify a genomic locus and to propose a correction to the existing gene models (see *Methods*). Lists of missing gene candidates were shared in online spreadsheets and documents, which allowed multiple annotators to inspect and correct candidate loci in parallel. 119 (25%) of the 486 non-redundant *P. pacificus* loci were classified as missing genes in predicted UTRs of annotated genes (Fig. 2b). We would speculate that this is caused by the fact that nematode genomes are compact and UTR regions can frequently overlap[45]. This can cause artificial fusion of transcripts during the assembly of RNA-seq data. Consequently, only the largest ORF of such a gene is annotated as protein-coding and the rest is classified as 3′ and 5′ UTR. Alternatively, this problem could arise when a fused gene prediction from the sister species is used as homology information but MAKER2 fails to generate a complete gene model out of it. The *C. elegans* gene C29H12.2 is one example of a missing gene model residing in the UTR of a *P. pacificus rars-2* homolog (Fig. 2c). The corresponding *P. pacificus* locus is spanned by two assembled transcripts that are homologous two C29H12.2 and *rars-2*, respectively. Both transcripts are also well supported by Iso-seq data and exhibit different expression levels[44,46]. In such a case, we would propose a replacement of the old *P. pacificus* gene model by the two distinct transcripts.

After manual inspection of all 526 missing gene candidates, 201 (41%) of the 486 non-redundant *P. pacificus* loci were classified as missing genes (Fig. 2b). Presumably this kind of error could arise when the gene annotation pipeline is mostly dependent on gene prediction algorithms which fail to predict all genes in gene dense regions (e.g. operon structures) as the intergenic distances might span only a few hundred nucleotides, which could be too small for triggering the initiation of a new gene model. The *C. elegans* gene *apn-1* is one example of a missed gene model in a gene dense region (Fig. 2d). Given that the *P. pacificus* homolog of *apn-1* has good transcriptomic support, the correction in this case would simply add the transcript to the existing gene models. Other instances of missing homologs are due to borderline cases in the BLASTP searches where one search resulted in an e-value slightly below the e-value threshold (10^{-5}) and the result of the other BLASTP search was slightly above the threshold. In total, we encountered 87 of such cases which we termed 'weak similarity' (Fig. 2b). For such cases no correction was proposed. In summary, we compiled corrections for 280 *P. pacificus* genes which were replaced by 714 new gene models. All these changes were submitted to WormBase and were incorporated in the release WS272.

Artificial gene fusions mask thousands of hidden orthologs. A small number of *C. elegans* genes with missing homologs in the current gene annotations (version: El Paco v1/WS268) of *P. pacificus* were classified as located in fused gene models (Fig. 2b). One potential explanation could be that an artificially fused gene prediction from the sister species is taken as homology data to annotate the orthologous locus in *P. pacificus*, but small errors cause parts of the gene model to be either incompletely or incorrectly annotated in *P. pacificus* resulting in a loss of detectable homology (Fig. 2c). Even if the homolog of a *C. elegans* gene is incorporated in the correct ORF within an artificially fused gene model, this could still cause a loss of one-to-one orthology as the

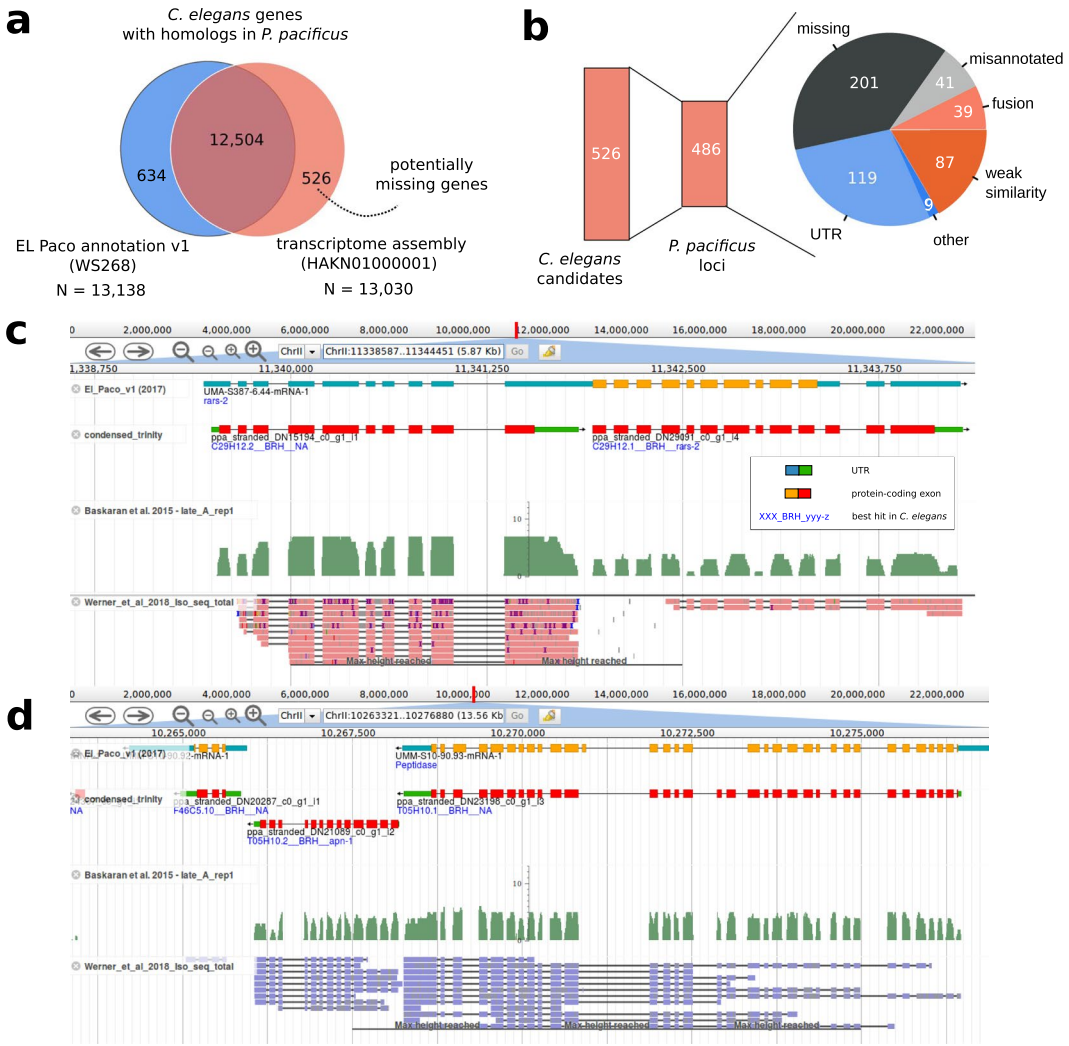


Figure 2. Identification of missing genes. (**a**) 526 potentially missing genes were identified based on *C. elegans* genes with homologs in the transcriptome assembly but not in current gene annotations. (**b**) The 526 missing gene candidates were located in 486 *P. pacificus* loci that were classified based on community annotators. (**c**) The genome browser screenshot shows a homolog of *C. elegans* C29H12.2 which is located in the annotated 5′UTR of a *P. pacificus* gene. This locus harbors two *P. pacificus* transcripts with different expression levels and well supported as non-overlapping transcripts based on RNA-seq and Iso-seq data. (**d**) A homolog of apn-1 is completely missing from current gene annotations.

corresponding *P. pacificus* gene can only be identified as one-to-one ortholog of a single *C. elegans* gene. Thus, we performed a second screen for *C. elegans* genes that had a predicted one-to-one ortholog (best-reciprocal hit) in the transcriptome assembly but not in current gene annotations (Fig. 3a). In total, 6075 (93%) of *C. elegans* genes with a predicted one-to-one ortholog (based on best-reciprocal hits) in current gene annotations, also had a predicted one-to-one ortholog against the *de novo* transcriptome assembly (Fig. 3a). Nevertheless, we found 2075 *C. elegans* genes that only had predicted one-to-one orthologs in the *de novo* transcriptome assembly. Excluding *C. elegans* genes that were identified already in the previous screen for missing homologs, this resulted in 1692 *C. elegans* genes with predicted one-to-one orthologs in the *de novo* transcriptome assembly but not in the current set of gene annotations (version: El Paco v1/WS268). Community-based classification and curation of the 1281 corresponding *P. pacificus* loci classified 912 (71%) cases as artificial gene fusions (Fig. 3b). One such an example is the *C. elegans* gene D1053.3. Its putative ortholog is fused with the *P. pacificus mvb-12* ortholog (Fig. 3c). Apart from being orthologous to two different *C. elegans* genes, both *P. pacificus* genes are supported as non-overlapping transcripts by RNA-seq and Iso-seq, and are expressed at different levels. This confirmed the interpretation of an artificially fused annotation. The proposed correction in this case would be a replacement of the old gene model by the two non-overlapping transcripts. In total, we updated 1241 *P. pacificus* gene models and replaced them with 3305 new models. These updates were submitted to WormBase and will be released following curation. The new *P. pacifcus* gene annotation (version: El Paco v2) with 28,036 gene models is also available on http://www.pristionchus.org/download. The results of the BUSCO analysis (Complete and Single Copy: 95.4%, Duplicated:

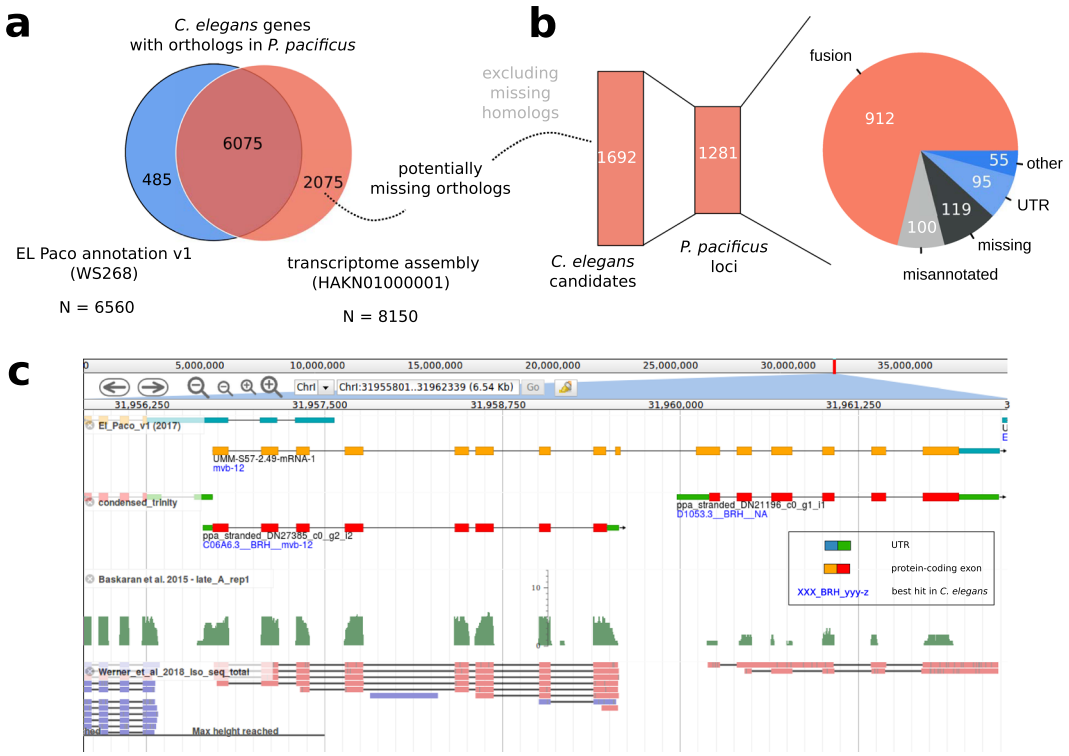


Figure 3. Community-based curation of hidden orthologs. (**a**) We identified 2075 putative *C. elegans* one-to-one orthologs that were specific to the *P. pacificus* transcriptome assembly. (**b**) Community-based curation classified most of the corresponding gene loci as artificial gene fusions. (**c**) Non-overlapping transcripts corresponding to *P. pacificus* orthologs of *mvb-12* and D1053.3 are artificially fused in a current gene model. This prohibits the detection of a one-to-one ortholog of D1053.3 based on a genome-wide approach such as best reciprocal hits.

1.7%, Fragmented: 2.0%, Missing: 0.9%) indicate that the new annotation represents a substantial improvement over the previous annotations[15] (Table 1).

Improved gene annotations facilitate the establishment of a catalog of *C. elegans* homologs and orthologs in the *P. pacificus* genome. Since our primary focus was to improve the annotation of *C. elegans* orthologs in the *P. pacificus* genome, we wanted to use the updated gene annotation to generate a catalog of predicted orthologs between *C. elegans* and *P. pacificus*. As the identification of orthologs typically requires sufficient genomic and phylogenetic knowledge to retrieve relevant protein data sets and to perform reconstruction of gene trees[24,45,46], a genome-wide catalog of orthologs would be highly useful as a starting point for researchers without sufficient expertise. Previous comparisons between *C. elegans* and *P. pacificus* identified putative one-to-one orthologs for roughly 6000–8000 genes[44,46]. To further characterize *C. elegans* genes without orthologs in *P. pacificus*, we additionally carried out a phylostratigraphic analysis[47] to estimate their relative age. Basically, phylostratigraphy uses absence-presence patterns of a gene to map its origin to an internal branch in a species tree[47]. Our analysis revealed that 5258 (26%) of *C. elegans* genes do not have BLAST hits in *Pristionchus* or more distantly related species (Phylostrata I–IV, Supplemental Table 1). This strongly suggests that they are younger than the common ancestor between *C. elegans* and *P. pacificus* and consequently have no orthologs. Next, we applied two different approaches to predict orthologs between *C. elegans* and *P. pacificus*: best reciprocal hits and Markov clustering as implemented in the software orthAgogue[48]. Computation of best reciprocal hits is a standard approach for predicting one-to-one orthologs across species[49,50]. In order to capture more complex orthology relationships (e.g. many-to-many), more general approaches such as Markov clustering have been widely applied[48,51]. Based on best reciprocal hits, we identified 8348 predicted one-to-one orthologs between both species (Supplemental Table 1) whereas the orthAgogue pipeline identified 7643 orthologous clusters, of which only 3345 corresponded to one-to-one orthologs. The large majority (98%) of these predicted one-to-one orthologs was also identified as best reciprocal hits and in 3260 (99%) cases, the same *P. pacificus* gene was predicted as one-to-one ortholog. The large discrepancy between the total number of best reciprocal hits and one-to-one orthologs defined by orthAgogue could be explained by the fact that best reciprocal hits do not take inparalogs into account[49]. However, only 1049 (21%) of *C. elegans* genes that were not identified as one-to-one orthologs by orthAgogue could be explained by the presence of lineage-specific inparalogs, suggesting that orthAgogue with default settings might be too conservative for this analysis. This is further supported by the reanalysis of 57 one-to-one orthologous pairs that were previously confirmed by phylogenetic analysis[46]. While 53 of the previously confirmed one-to-one orthologs were captured as best reciprocal hits, only 33 were also identified

by orthAgogue. Taken together, the improved gene annotation facilitated the prediction of substantially more one-to-one orthologs (Fig. 3a, Supplemental Table 1). This resource can be taken as a starting point to identify candidate genes in *P. pacificus*.

Discussion

With *C. elegans*, *C. briggsae*, and *P. pacificus*, three genetically tractable and free living nematode model organisms have been well established and can be used to study the evolution of gene function at various time-scales[2,3,52]. For example, recent reverse genetic approaches in *P. pacificus* have revealed functional divergence of genes with known roles in *C. elegans* dauer formation[22,23,53]. In addition, mutant screens in *P. pacificus* for social behaviours have uncovered multiple orthologous *C. elegans* genes for which a behavioral phenotype had been overlooked previously[33,54]. Together with complementary studies of the functional importance of novel genes[7,32,55], this makes nematodes an extremely powerful system to study genome evolution and gene function at a mechanistic level.

In order to facilitate fruitful functional studies across multiple model organisms, it is crucial to generate genomic resources (e.g. assemblies, annotations) and experimental genetic toolkits (e.g. forward and reverse genetics) of comparable quality. The chromosome-scale assembly of the *P. pacificus* genome[15] was a major step towards making this species more tractable for other groups. In our study, we aimed to minimize the discrepancy between automatically generated gene annotations for *P. pacificus* and heavily curated annotations for *C. elegans*. To this end, we incorporated recently generated Iso-seq and RNA-seq data into current gene annotations by manual curation of suspicious candidate loci that were identified by comparative genomic analysis. While application of alternative annotation pipelines can generate overall better gene annotations[29,41], they cannot guarantee that gene annotations will only improve. In certain cases, new annotation pipelines will also cause new errors. In contrast, during manual inspection, each community curator has the choice to not propose any change of gene models in case of uncertainty. Thus, manual inspection should only lead to removal of errors and thus improve annotation quality without introducing biases elsewhere. While manual annotation is an incredibly tedious task that is probably not scalable to complete genomes[38], we minimized the workload by focusing on a small gene set of *C. elegans* orthologs, recruiting colleagues as community curators, and restricting the task just to the selection of alternative gene models that were generated from transcriptomic data[16,44] or previous rounds of gene prediction[56,57]. In our opinion, the most crucial aspect of this community project is a good training of new annotators. We achieved this by personal training sessions between experienced and new annotators and the possibility to always discuss cases of uncertainty with other curators. For larger projects, initial training could be achieved by comprehensive online tutorials and communication via email, but this will likely be less efficient. In the case of the *P. pacificus* gene annotations, our study raised the gene count from 25,517 to 28,036 and increased the single copy ortholog completeness level from 86% to 97%. In the *P. pacificus* genome, the greatest source of error was the artificial fusion of neighboring genes. This type of error might be more prevalent in nematodes where genomes are compact[9] and genes frequently overlap[37,45]. Consequently, manual annotation of restricted gene sets has been proposed and applied previously to circumvent this problem[58]. Given that nematode genomes tend to be pretty compact (Fig. 1), we anticipate that misannotation due to overlapping gene models should be much less pronounced in large vertebrate or plant genomes. Nevertheless, it would be interesting to apply similar screens for gene annotation artifacts to other systems and eventually this could reveal some incorrect annotations in the genomes of classical model organisms.

While this study was restricted to *P. pacificus* genes with putative orthologs in *C. elegans*, we cannot reliably estimate the fraction of erroneous gene models across the whole genome. Our results would suggest that the fraction of missing genes is around one percent (Fig. 2a,b) and the amount of gene models affected by artificial fusions may be up to 15% (Fig. 3a). However, as the *P. pacificus* genome has a higher gene density and a higher concentration of old genes at the chromosome centers[8,15], we hypothesize that errors due to artificial gene fusions should be much less pronounced at chromosome arms. To test this, an unbiased quantification of error rates across genomic segments would be needed. In future, we also plan to focus on large gene families and lineage-specific orphan genes[55] that were not explicit subjects of this study. Artificial fusions in these classes of genes could be identified by screens for unexpectedly long gene models or unusual protein domain content. For orphan genes abundant RNA-seq studies of different developmental stages[22,46], tissues[10,46], environmental conditions[59], sexes[16], and genetic backgrounds[60,61] could be used to detect non-overlapping transcripts that exhibit anticorrelated expression within a single locus. Thus, while our study has demonstrated that community-based curation of gene annotations is feasible and can lead to substantial improvements, continued effort is needed to lift its quality to a level that would be similar to classical model organisms.

Methods

Comparative assessment of nematode genomes. We downloaded 22 nematode genomes and corresponding protein sequences from WormBase ParaSite (release WBPS13). For *Steinernema carpocapsae*, the latest version at WBPS14 was used. In case of multiple isoforms, we selected the longest isoform for further analysis. We ran BUSCO (version 3.0.1) in protein mode (option: -m prot) against the nematode_odb9 data set (N = 982 genes) to evaluate the completeness level of available protein sequences.

Genome browser integration of transcriptomic resources. To allow community annotators to propose alternative gene models, we integrated recent raw read alignments and reference guided transcript assemblies of Iso-seq data[44] and a *de novo* assembly of strand-specific RNA-seq data[16] into our genome browser (implemented in jbrowse[62]) on our webserver (http://www.pristionchus.org). Genomic coordinates for the *de novo* transcriptome assembly were generated by alignment to the *P. pacificus* reference genome (version El Paco) with the program exonerate[63] (version: 2.2.0, options: -m est2genome – dnawordlen 20 – maxintron 20000). To reduce the complexity of this data set, a condensed version of the *de novo* transcriptome assembly (selection of the isoform with the longest ORF as single representative isoform per gene, minimum peptide length of 60 amino acids, removal of single exon transcripts) with annotated best-reciprocal hits and best hits (BLASTP, e-value $< 10^{-5}$) in

C. elegans was also incorporated into our jbrowse instance. In addition, predicted protein sequences of previous versions of *P. pacificus* annotations (Hybrid1[56] and TAU2011[57]) were mapped against the *P. pacificus* assembly by exonerate (version: 2.2.0, options: -m protein2genome – dnawordlen 20 – maxintron 20000). All data sets are available under the gene annotation track of the El Paco reference assembly in our genome browser. To evaluate the quality of the two recent transcriptome assemblies, we ran BUSCO (version 3.0.1, options -m trans) against the nematode_odb9 data set (N = 982 genes) for completeness assessment (Table 1).

Identification of missing and fused gene models in current gene annotations. We ran bidirectional BLASTP (e-value $< 10^{-5}$) searches between *C. elegans* (version: WS260, longest isoform per gene) and two different *P. pacificus* data sets: the annotated proteins (version: El Paco v1, WS268) and the *de novo* transcriptome assembly[16]. For the *de novo* transcriptome, we reduced the redundancy resulting from different isoforms by selecting the longest ORFs per gene. Based on the different BLASTP searches, we first screened for *C. elegans* proteins with BLASTP hits against ORFs in the *de novo* transcriptome assembly but not against the currently annotated proteins. This yielded 526 candidate genes. In a second phase, we screened for *C. elegans* proteins with putative orthologs, defined by best-reciprocal BLASTP relationships, in the *de novo* transcriptome assembly but not in the annotated proteins, resulting in 2075 candidate genes.

Community-based manual curation of candidate loci. All *C. elegans* genes together with their candidate homologs and orthologs in the *P. pacificus de novo* transcriptome assembly were stored in a shared online spreadsheet. Community annotators were trained to find the corresponding locus in the genome browser by entering the transcript identifier and to manually inspect the surrounding regions that were defined by the encompassing *P. pacificus* gene model. The candidate locus was then classified as untranslated region (UTR) (the query transcript overlapped exons that were annotated as UTR), missing gene (the query transcript did not overlap any annotated exon), gene fusion (the query transcript did overlap protein-coding exons and homology was detected by BLASTP), misannotation (the query transcript did overlap protein-coding exons but no BLASTP hit was found due incorrect reading frame annotation or minimal overlap) or inconclusive. After classification, a correction was proposed that either added new genes (identifiers could be selected from the *de novo* assembled transcripts, Iso-seq assemblies, or previous versions of gene annotations) or replaced an existing gene model by one or more new genes. In such a case the objective was to lose as little annotated coding sequence as possible. Thus, new genes were selected from the above mentioned data set in order to cover as much coding sequence of the initial gene model as possible. If parts of the old gene model were not covered, BLAST searches against *C. elegans* and other *Pristionchus* species were used to split the old gene model into several parts with sequence matches to distinct *C. elegans* genes, or to extract partial protein sequences of the old gene model that were not covered. Such protein sequence stretches were given a pseudo identifier and were stored in a shared online document. All these sequences were later automatically reannotated by mapping them against the reference genome with the help of exonerate. In case that an existing gene model was replaced by multiple new gene models, we additionally selected one of the new gene models to inherit the WormBase identifier of the old gene model to allow WormBase to record the history of a given gene model. Usually, either the most conserved or the longest new gene model was chosen. Due to the fact that a single artificially fused gene could cause missing homologs and orthologs for multiple *C. elegans* genes, some loci were curated multiple times. We randomly picked some of these cases to compare the classifications and the corresponding corrections from multiple curators, which turned out to be largely consistent. In case of redundant curations, one out of many possible curations for a given locus was chosen based on the following criteria: preference towards higher number of new models, experience of the curator (number of curated loci), and transcriptional evidence over gene prediction.

Phylostratigraphy and orthology predictions. Outgroup data sets were defined by concatenating all protein sequence data from different species in the ladder-like phylogeny leading to *C. elegans* (Fig. 1). More precisely, we pooled all data from species in an induced subtree defined by branches with roman numbers in Fig. 1. We then ran a BLASTP search (e-value < 0.001) of *C. elegans* proteins (longest isoform per gene) against each of the outgroup data sets. Starting from the *C. elegans* genes with homologs in the most distant outgroup set (VIII), we iteratively defined phylostrata by comparison with the next, more closely related outgroup set. The results of this analysis are summarized in Supplemental Table 1. *C. elegans* specific genes are assigned to phylostratum I, whereas genes that are present in the most divergent outgroups are assigned to phylostratum VIII. Orthologs were defined after performing all pairwise BLASTP searches including self-searches (e-value $< 10^{-5}$) between *C. elegans* and *P. pacificus* and extracting best reciprocal hits from the BLAST output files. Simultaneously, the program orthAgogue was run with default setting on the same input files[48].

Data availability

The strand-specific *de novo* transcriptome was submitted to the European Nucleotide Archive under the accession HAKN01000001[16] and the Iso-seq data was submitted to the European Nucleotide Archive under the accessions ERX2315712 and ERX2315713[44]. All data sets are also available at http://www.pristionchus.org/download. The initial set of *P. pacificus* gene annotations corresponds to WormBase WS268. Corrections from this study were submitted to WormBase and will be released following curation.

Received: 1 October 2019; Accepted: 20 November 2019;
Published online: 11 December 2019

References

1. Sommer, R. J., Carta, L. K., Kim, S.-Y. & Sternberg, P. W. Morphological, genetic and molecular description of *Pristionchus pacificus* sp. n. (Nematoda: Neodiplogastridae). *Fundam. Appl. Nematol.* **19**, 511–521 (1996).
2. Sternberg, P. W. Why *Caenorhabditis elegans* is great and *Pristionchus pacificus* might be better. In *Pristionchus pacificus* (ed. Sommer, R. J.) **11**, 1–17 (BRILL).
3. Sommer, R. J. Evolution of regulatory networks: nematode vulva induction as an example of developmental systems drift. *Adv. Exp. Med. Biol.* **751**, 79–91 (2012).
4. Bose, N. *et al*. Complex small-molecule architectures regulate phenotypic plasticity in a nematode. *Angew. Chem. Int. Ed Engl.* **51**, 12438–12443 (2012).
5. Yim, J. J., Bose, N., Meyer, J. M., Sommer, R. J. & Schroeder, F. C. Nematode signaling molecules derived from multimodular assembly of primary metabolic building blocks. *Org. Lett.* **17**, 1648–1651 (2015).
6. Falcke, J. M. *et al*. Linking genomic and metabolomic natural variation uncovers nematode pheromone biosynthesis. *Cell Chem Biol* **25**, 787–796.e12 (2018).
7. Mayer, M. G., Rödelsperger, C., Witte, H., Riebesell, M. & Sommer, R. J. The orphan gene *dauerless* regulates dauer development and intraspecific competition in nematodes by copy number variation. *PLoS Genet.* **11**, e1005146 (2015).
8. Prabh, N. *et al*. Deep taxon sampling reveals the evolutionary dynamics of novel gene families in *Pristionchus* nematodes. *Genome Res.* **28**, 1664–1674 (2018).
9. Rödelsperger, C., Streit, A. & Sommer, R. J. Structure, function and evolution of the nematode genome. In: *eLS*. John Wiley & Sons, Ltd: Chichester, https://doi.org/10.1002/9780470015902.a0024603 (2013).
10. Lightfoot, J. W., Chauhan, V. M., Aylott, J. W. & Rödelsperger, C. Comparative transcriptomics of the nematode gut identifies global shifts in feeding mode and pathogen susceptibility. *BMC Res. Notes* **9**, 142 (2016).
11. Susoy, V. *et al*. Large-scale diversification without genetic isolation in nematode symbionts of figs. *Sci Adv* **2**, e1501031 (2016).
12. Bumbarger, D. J., Riebesell, M., Rödelsperger, C. & Sommer, R. J. System-wide rewiring underlies behavioral differences in predatory and bacterial-feeding nematodes. *Cell* **152**, 109–119 (2013).
13. Witte, H. *et al*. Gene inactivation using the CRISPR/Cas9 system in the nematode *Pristionchus pacificus*. *Dev. Genes Evol.* **225**, 55–62 (2015).
14. Srinivasan, J. *et al*. A bacterial artificial chromosome-based genetic linkage map of the nematode *Pristionchus pacificus*. *Genetics* **162**, 129–134 (2002).
15. Rödelsperger, C. *et al*. Single-molecule sequencing reveals the chromosome-scale genomic architecture of the nematode model organism *Pristionchus pacificus*. *Cell Rep.* **21**, 834–844 (2017).
16. Rödelsperger, C. *et al*. Phylotranscriptomics of *Pristionchus* nematodes reveals parallel gene loss in six hermaphroditic lineages. *Curr. Biol.* **28**, 3123–3127.e5 (2018).
17. Namai, S. & Sugimoto, A. Transgenesis by microparticle bombardment for live imaging of fluorescent proteins in *Pristionchus pacificus* germline and early embryos. *Dev. Genes Evol.* **228**, 75–82 (2018).
18. Lo, T.-W. *et al*. Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. *Genetics* **195**, 331–348 (2013).
19. Bui, L. T. & Ragsdale, E. J. Multiple plasticity regulators reveal targets specifying an induced predatory form in nematodes. *Mol. Biol. Evol*, https://doi.org/10.1093/molbev/msz171 (2019).
20. Ishita, Y., Chihara, T. & Okumura, M. Serotonergic modulation of feeding behavior in *Caenorhabditis elegans* and other related nematodes. *Neurosci. Res*, https://doi.org/10.1016/j.neures.2019.04.006 (2019).
21. Liu, Z. *et al*. Predator-secreted sulfolipids induce defensive responses in *C. elegans*. *Nature Communications* **9** (2018).
22. Moreno, E. *et al*. DAF-19/RFX controls ciliogenesis and influences oxygen-induced social behaviors in *Pristionchus pacificus*. *Evol. Dev.* **20**, 233–243 (2018).
23. Markov, G. V. *et al*. Functional conservation and divergence of *daf-22* paralogs in *Pristionchus pacificus* dauer development. *Mol. Biol. Evol.* **33**, 2506–2514 (2016).
24. Markov, G. V., Baskaran, P. & Sommer, R. J. The same or not the same: lineage-specific gene expansions and homology relationships in multigene families in nematodes. *J. Mol. Evol.* **80**, 18–36 (2015).
25. Namdeo, S. *et al*. Two independent sulfation processes regulate mouth-form plasticity in the nematode. *Development* **145** (2018).
26. Rödelsperger, C. Comparative genomics of gene loss and gain in *Caenorhabditis* and Other Nematodes. In *Methods in Molecular Biology* 419–432 (2018).
27. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
28. Stanke, M. *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–9 (2006).
29. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **1962**, 65–95 (2019).
30. Cantarel, B. L. *et al*. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
31. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
32. Lightfoot, J. W. *et al*. Small peptide-mediated self-recognition prevents cannibalism in predatory nematodes. *Science* **364**, 86–89 (2019).
33. Moreno, E. *et al*. Regulation of hyperoxia-induced social behaviour in *Pristionchus pacificus* nematodes requires a novel cilia-mediated environmental input. *Sci. Rep.* **7**, 17550 (2017).
34. Kieninger, M. R. *et al*. The nuclear hormone receptor NHR-40 acts downstream of the sulfatase EUD-1 as part of a developmental plasticity switch in *Pristionchus*. *Curr. Biol.* **26**, 2174–2179 (2016).
35. Baskaran, P. & Rödelsperger, C. Microevolution of duplications and deletions and their impact on gene expression in the Nematode *Pristionchus pacificus*. *PLoS One* **10**, e0131136 (2015).
36. Weller, A. M., Rödelsperger, C., Eberhardt, G., Molnar, R. I. & Sommer, R. J. Opposing forces of A/T-biased mutations and G/C-biased gene conversions shape the genome of the nematode *Pristionchus pacificus*. *Genetics* **196**, 1145–1152 (2014).
37. Prabh, N. & Rödelsperger, C. Divergence, and mixed origin contribute to the emergence of orphan genes in nematodes. *G3* **9**, 2277–2286 (2019).
38. Salzberg, S. L. Next-generation genome annotation: we still struggle to get it right. *Genome Biol.* **20**, 92 (2019).
39. Smythe, A. B., Holovachov, O. & Kocot, K. M. Improved phylogenomic sampling of free-living nematodes enhances resolution of higher-level nematode phylogeny. *BMC Evolutionary Biology* **19** (2019).
40. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
41. McLean, F., Berger, D., Laetsch, D. R., Schwartz, H. T. & Blaxter, M. Improving the annotation of the *Heterorhabditis bacteriophora* genome. *Gigascience* **7** (2018).
42. Hiraki, H. *et al*. Genome analysis of Diploscapter coronatus: insights into molecular peculiarities of a nematode with parthenogenetic reproduction. *BMC Genomics* **18**, 478 (2017).
43. Rödelsperger, C. *et al*. Characterization of genetic diversity in the nematode *Pristionchus pacificus* from population-scale resequencing data. *Genetics* **196**, 1153–1165 (2014).

44. Werner, M. S. *et al*. Young genes have distinct gene structure, epigenetic profiles, and transcriptional regulation. *Genome Res.* **28**, 1675–1687 (2018).
45. Rödelsperger, C., Menden, K., Serobyan, V., Witte, H. & Baskaran, P. First insights into the nature and evolution of antisense transcription in nematodes. *BMC Evol. Biol.* **16**, 165 (2016).
46. Baskaran, P. *et al*. Ancient gene duplications have shaped developmental stage-specific expression in *Pristionchus pacificus*. *BMC Evol. Biol.* **15**, 185 (2015).
47. Domazet-Loso, T., Brajković, J. & Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* **23**, 533–539 (2007).
48. Ekseth, O. K., Kuiper, M. & Mironov, V. orthAgogue: an agile tool for the rapid prediction of orthology relations. *Bioinformatics* **30**, 734–736 (2014).
49. Remm, M., Storm, C. E. & Sonnhammer, E. L. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052 (2001).
50. Tatusov, R. L. A Genomic Perspective on Protein Families. *Science* **278**, 631–637 (1997).
51. Li, L., Stoeckert, C. J. Jr & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
52. Verster, A. J., Ramani, A. K., McKay, S. J. & Fraser, A. G. Comparative RNAi Screens in *C. elegans* and *C. briggsae* Reveal the Impact of Developmental System Drift on Gene Function. *PLoS Genetics* **10**, e1004077 (2014).
53. Sieriebriennikov, B., Markov, G. V., Witte, H. & Sommer, R. J. The Role of DAF-21/Hsp90 in Mouth-Form Plasticity in *Pristionchus pacificus*. *Mol. Biol. Evol.* **34**, 1644–1653 (2017).
54. Moreno, E. & Sommer, R. J. A cilia-mediated environmental input induces solitary behaviour in *Caenorhabditis elegans* and *Pristionchus pacificus* nematodes. *Nematology* **20**, 201–209 (2018).
55. Prabh, N. & Rödelsperger, C. Are orphan genes protein-coding, prediction artifacts, or non-coding RNAs? *BMC Bioinformatics* **17**, 226 (2016).
56. Borchert, N. *et al*. Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. *Genome Res.* **20**, 837–846 (2010).
57. Sinha, A., Sommer, R. J. & Dieterich, C. Divergent gene expression in the conserved dauer stage of the nematodes *Pristionchus pacificus* and *Caenorhabditis elegans*. *BMC Genomics* **13**, 254 (2012).
58. Stoltzfus, J. D., Minot, S., Berriman, M., Nolan, T. J. & Lok, J. B. RNAseq analysis of the parasitic nematode *Strongyloides stercoralis* reveals divergent regulation of canonical dauer pathways. *PLoS Negl. Trop. Dis.* **6**, e1854 (2012).
59. Sanghvi, G. V. *et al*. Life history responses and gene expression profiles of the nematode *Pristionchus pacificus* cultured on *Cryptococcus* yeasts. *PLoS One* **11**, e0164881 (2016).
60. Serobyan, V. *et al*. Chromatin remodelling and antisense-mediated up-regulation of the developmental switch gene *eud-1* control predatory feeding plasticity. *Nat. Commun.* **7**, 12337 (2016).
61. Moreno, E., McGaughran, A., Rödelsperger, C., Zimmer, M. & Sommer, R. J. Oxygen-induced social behaviours in *Pristionchus pacificus* have a distinct evolutionary history and genetic regulation from *Caenorhabditis elegans*. *Proc. Biol. Sci.* **283**, 20152263 (2016).
62. Buels, R. *et al*. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* **17**, 66 (2016).
63. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).

Acknowledgements

The authors would like to thank the complete *Pristionchus* community for their long-term interest in studying *P. pacificus* and thus motivating this work. Further thanks to Bogdan Sieriebriennikov for providing additional manual curations and to all members of the Sommer lab for general discussions. Finally, special thanks to Michael Paulini for incorporating the updated gene models into WormBase. This work was funded by the Max Planck Society.

Author contributions

Conceptualization, C.R.; Investigation, C.R., M.A., M.L., T.T., S.S., M.D., S.W., W.H., D.R.S. and Z.H.; Writing – Original Draft, C.R.; Writing – Review & Editing, C.R., M.A., M.L., T.T., S.S., M.D., S.W., W.H., D.R.S. and Z.H.; Supervision, C.R.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-55359-5.

Correspondence and requests for materials should be addressed to C.R.

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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019

Spatial Transcriptomics of Nematodes Identifies Sperm Cells as a Source of Genomic Novelty and Rapid Evolution

Christian Rödelsperger,*[†,1] Annabel Ebbing,[†,2] Devansh Raj Sharma,[1] Misako Okumura,[3] Ralf J. Sommer ,[1] and Hendrik C. Korswagen*[,2,4]

[1]Department for Integrative Evolutionary Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany

[2]Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences and University Medical Center Utrecht, Utrecht, The Netherlands

[3]Program of Biomedical Science, Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan

[4]Developmental Biology, Department of Biology, Institute of Biodynamics and Biocomplexity, Utrecht University, Utrecht, The Netherlands

[†]These authors contributed equally to this work.

***Corresponding authors**: E-mails: christian.roedelsperger@tuebingen.mpg.de; r.korswagen@hubrecht.eu.

Associate editor: Ilya Ruvinsky

Sequencing data have been submitted to the European Nucleotide Archive under the study accession PRJEB37439.

Abstract

Divergence of gene function and expression during development can give rise to phenotypic differences at the level of cells, tissues, organs, and ultimately whole organisms. To gain insights into the evolution of gene expression and novel genes at spatial resolution, we compared the spatially resolved transcriptomes of two distantly related nematodes, *Caenorhabditis elegans* and *Pristionchus pacificus*, that diverged 60–90 Ma. The spatial transcriptomes of adult worms show little evidence for strong conservation at the level of single genes. Instead, regional expression is largely driven by recent duplication and emergence of novel genes. Estimation of gene ages across anatomical structures revealed an enrichment of novel genes in sperm-related regions. This provides first evidence in nematodes for the "out of testis" hypothesis that has been previously postulated based on studies in *Drosophila* and mammals. "Out of testis" genes represent a mix of products of pervasive transcription as well as fast evolving members of ancient gene families. Strikingly, numerous novel genes have known functions during meiosis in *Caenorhabditis elegans* indicating that even universal processes such as meiosis may be targets of rapid evolution. Our study highlights the importance of novel genes in generating phenotypic diversity and explicitly characterizes gene origination in sperm-related regions. Furthermore, it proposes new functions for previously uncharacterized genes and establishes the spatial transcriptome of *Pristionchus pacificus* as a catalog for future studies on the evolution of gene expression and function.

***Key words:* comparative genomics, orphan genes, de novo gene, duplication, *Pristionchus pacificus*, *Caenorhabditis elegans*, meiosis.**

Introduction

The immense morphological diversity across all living organisms has raised the fundamental question, how novelty arises in the first place. Genetic screens in multiple model systems have dissected the underlying architecture of many traits and have highlighted the role of *cis*-regulatory evolution, as well as the emergence of novel genes in shaping phenotypic diversity (Wray 2007; Khalturin et al. 2009). Novel genes can arise from a spectrum of processes ranging from duplication of existing genes to de novo formation from noncoding sequences (McLysaght and Hurst 2016; Rödelsperger, Prabh, et al. 2019; Van Oss and Carvunis 2019). In addition, sequence divergence can contribute to the formation of orphan genes without detectable homologs in other lineages. Such orphan genes make up to one-third of all genes in a given genome (Tautz and Domazet-Lošo 2011; Rödelsperger, Prabh, et al. 2019). Their high abundance combined with the paucity of homologs raises the fundamental question, where orphan genes are coming from and what their biological function might be. Although the origin of orphan genes has been extensively studied in yeast (Carvunis et al. 2012; Vakirlis et al. 2018), vertebrates (Toll-Riera et al. 2008; Knowles and McLysaght 2009), and insects (Levine et al. 2006; Wissler et al. 2013), recent studies in plants and other animal phyla such as nematodes have broadened our knowledge of novel gene formation (Prabh et al. 2018; Prabh and Rödelsperger 2019; Zhang et al. 2019). In particular, nematodes with their small genome sizes and species richness are ideal model systems to study the evolution of novel genes by means of

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administration - MPS user on 08 August 2023

Article

phylogenomic approaches relying on deep taxon sampling (Prabh et al. 2018; Rödelsperger 2018; Rödelsperger, Prabh, et al. 2019). Specifically, each of the two nematode genera *Caenorhabditis* and *Pristionchus* has around 40 described species and continuous sampling efforts discover new species almost every year (Rödelsperger et al. 2018; Herrmann et al. 2019; Stevens et al. 2019). The combination of deep taxon sampling with genome sequencing allowed us to characterize the evolutionary dynamics and regulation of novel genes in *Pristionchus* nematodes as well as to elucidate several mechanisms including de novo formation that give rise to orphan genes (Prabh et al. 2018; Werner et al. 2018; Prabh and Rödelsperger 2019). Moreover, the two nematode species *Caenorhabditis elegans* and *Pristionchus pacificus* have the additional advantage to be established model systems with genetic toolkits including transgenes and genome editing that allow for comparative studies of gene function (Witte et al. 2015; Markov et al. 2016; Moreno et al. 2017, 2018).

In this study, we focus on the spatial expression of novel genes, which can provide further insights into their birth processes and enables to generate hypotheses about their functions. The *P. pacificus* genome has around 8,000 orphan genes without homologs in other nematode families (Rödelsperger 2018; Prabh and Rödelsperger 2019). However, only two of these orphan genes have been functionally characterized in unbiased genetic investigations. First, the orphan gene *dauerless* was found to control the entry into the dauer stage, which represents a highly conserved dispersal and long-term survival strategy in nematodes (Mayer et al. 2015). The second orphan gene *self-1* encodes a small peptide responsible for self-recognition in natural populations of *P. pacificus*. Nematodes of the *P. pacificus* lineage developed tooth-like structures that allow them to predate on other nematodes (Bento et al. 2010; Wilecki et al. 2015). Simultaneously, they evolved a self-recognition system to prevent cannibalism. Notably, a change of a single amino acid in the SELF-1 peptide was sufficient for a mutant line to be killed by its parent (Lightfoot et al. 2019). Although these examples indicate that orphan genes can be identified in unbiased genetic studies, developing rationales for large-scale targeting of novel genes remains difficult.

Therefore, in this study, we focus on the spatial expression of novel genes using RNA tomography (tomo-seq), which can provide insight into their birth processes and allows to generate hypotheses about their functions. Previously, the combination of spatial transcriptomics with knockdown experiments in *C. elegans* provided strong evidence that regional expression was indeed highly indicative for function (Ebbing et al. 2018). To acquire information about regionalized gene expression in *P. pacificus* on a genome-wide scale, we performed spatial transcriptomics of young adult hermaphrodites. First, this allows us to investigate in which anatomical structure a particular gene is expressed. This is of particular importance for the *Pristionchus* research community as spatial expression based on reporter lines is only known for a few *P. pacificus* genes (Ragsdale et al. 2013; Kieninger et al. 2016; Sieriebriennikov et al. 2020) and only a single tissue-specific RNA-seq data set is currently available (Lightfoot et al. 2016). Second, this data set can be used for a comparative analysis of gene expression between *P. pacificus* and *C. elegans*, which have been estimated to have diverged 60–90 Ma (Prabh et al. 2018; Werner et al. 2018). Finally, we can compare the distribution and evolutionary dynamics of genes of different age classes including novel genes across different anatomical regions. This analysis revealed a striking enrichment of novel genes in sperm-related regions, providing first evidence for the "out of testis" hypothesis in nematodes (Levine et al. 2006; Kaessmann 2010; Soumillon et al. 2013; Witt et al. 2019). "Out of testis" genes represent a mix of products of pervasive transcription as well as fast evolving members of ancient gene families. Together, the spatially resolved transcriptome of *P. pacificus* extends our understanding of novel gene formation in nematodes and will be an important resource for future studies of gene function.

Results

Spatial Transcriptomics Identifies Thousands of Genes with Regional Expression

In order to acquire genome-wide gene expression data with spatial resolution, we performed RNA tomography (Junker et al. 2014; Ebbing et al. 2018) which combines cryosectioning of individual *P. pacificus* specimens with RNA sequencing. This approach has been applied previously to compare the spatial transcriptomes of *C. elegans* hermaphrodites and males (Ebbing et al. 2018). In total, four young adult *P. pacificus* hermaphrodites were frozen and cut into around 40 sections per animal (20 µm per section) along the anterior–posterior (A–P) axis. RNA-seq library preparation and sequencing resulted in the detection of 21,046–32,930 (median of all sections per animal) transcripts per section. In total, we detected 12,265 *P. pacificus* genes with expression in all four specimens (supplementary table S1, Supplementary Material online). Reanalysis of the spatial transcriptomics data from *C. elegans* hermaphrodites with the same filtering criteria yielded 6,890 robustly expressed genes (Ebbing et al. 2018). We used orthologs of previously established *C. elegans* marker genes with punctuated spatial expression to define distinct regions (P1–P11 in *P. pacificus* and C1–C12 in *C. elegans*) that would correspond to different anatomical structures (fig. 1*a* and *b*) (Ebbing et al. 2018). For example, the neuropeptide encoding gene *flp-1* (P3, C3), which in *C. elegans* is expressed in the nerve ring (Nelson 1998), defines regions P3, C3, and the anterior regions (P1, P2, C1, and C2) as different parts of the head (fig. 1*a* and *b*). The gonads including the germline are defined by expression of the helicase *glh-1* (P5–P9 and C5–C11) (Gruidl et al. 1996). Within the gonads, the expression of major sperm protein (MSP) genes marks sperm-related regions (P5, P9 and C6, C10) (Ebbing et al. 2018; Tzur et al. 2018). In contrast to males, spermatogenesis in hermaphrodites occurs in a finite duration during development. In *P. pacificus* hermaphrodites, spermatogenesis is delayed relative to *C. elegans* hermaphrodites and starts only at the last larval molt, with the transition to oogenesis occurring 4–6 h into adulthood (Rudel et al. 2005). As mature sperm, which are stored in the spermatheca (fig. 1*a* and *b*),

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

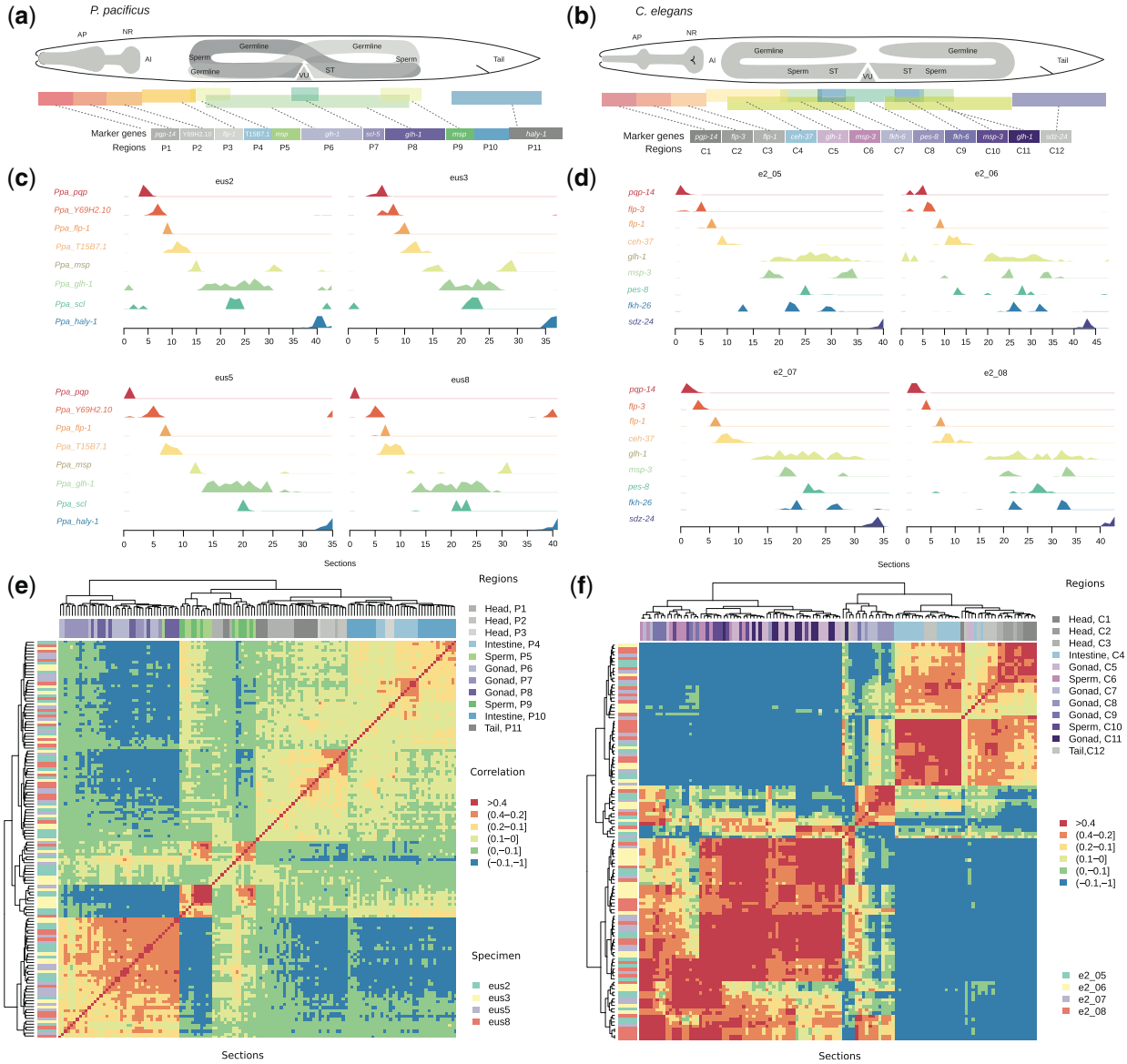


Fig. 1. Comparative spatial transcriptomics of nematodes. (*a*, *b*) The schematics represent anatomical overviews of *Pristionchus pacificus* (*a*) and *Caenorhabditis elegans* (*b*) hermaphrodites together with expression domains of selected marker genes which were used to define regions across the A–P axis. For example, P5 was defined as the first peak of *Ppa_msp*, P9 as the second peak, and P10 was defined indirectly as sections between P9 and P11. AP, anterior pharynx region; NR, nerve ring; AI, anterior intestine region; ST, spermatheca; VU, vulva and uterus region. (*c*, *d*) The plots show the normalized expression of marker genes across all sections for four specimens of *P. pacificus* (*c*) and *C. elegans* (*d*). (*e*, *f*) Hierarchical clustering of *P. pacificus* (*e*) and *C. elegans* (*f*) sections based on Pearson correlation indicates that sections within a region are generally highly similar on the whole transcriptome level.

are transcriptionally and translationally silent, the detected MSP expression rather represents residual mRNA levels than de novo transcription. Regions P4 and C4 are enriched for signals from the intestine that are not intermingled with either neuronal or germline expression. The comparison of the spatial transcriptome with previously generated *P. pacificus* reporter lines showed a high level of agreement (supplementary fig. S1, Supplementary Material online) (Ragsdale et al. 2013; Kieninger et al. 2016; Sieriebriennikov et al. 2020). For example, *eud-1*, the master regulator of the mouth form dimorphism in *P. pacificus* (Ragsdale et al. 2013; Kieninger et al. 2016), which is expressed in a few head neurons, is exclusively enriched in region P2 (supplementary fig. S1, Supplementary Material online). Importantly, the expression profiles of many of these marker genes are highly similar across species and individuals (fig. 1*c* and *d*). Taken together, these first analyses indicate that our data set represents a high-quality, spatially resolved transcriptome of *P. pacificus* that is suited for comparative analysis with *C. elegans*.

Based on the partitioning into regions, we summarized expression values per gene by taking the median *z*-score for all sections in a given region. We further defined regional genes based on a relative enrichment of expression in a given region (median *z*-score >1). This identified 3,502 regional genes in *P. pacificus* (supplementary table S3, Supplementary Material online) and 3,656 for the reanalyzed

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

C. elegans data (supplementary table S4, Supplementary Material online) (Ebbing et al. 2018). Next, we performed a clustering approach to assess whether regions that reflect similar anatomical structures (as defined using only a few marker genes) show similar transcriptomic profiles on a global scale. Despite the variable degrees of cellular heterogeneity across sections, hierarchical clustering based on Pearson correlation grouped individual sections by anatomical region and not by specimen (fig. 1*e* and *f*). Thus, anatomical regions identified in individual animals are reproducible across different animals, and they are not affected by batch effects or individual-to-individual variation. Together, our data set of 3,502 genes with regional expression profiles represents an enormous extension to the knowledge about spatial expression in *P. pacificus*.

Regional Expression Is Not Highly Conserved between *P. pacificus* and *C. elegans*

The availability of spatially resolved and highly comparable transcriptomes of *C. elegans* and *P. pacificus* provides the opportunity to study the evolution of regional expression in nematodes. Despite the high conservation of expression patterns of individual marker genes such as *flp-3*, *glh-1*, and MSP genes and their orthologs in *P. pacificus* (fig. 1*c* and *d*), it remains elusive to what extent regional expression is generally conserved between species. In addition, since partially non-orthologous marker genes were selected to segment the spatial transcriptomes of *P. pacificus* and *C. elegans* (fig. 1*a*–*d*), it is unclear how the different regions of both species correspond to each other. To assess the overall degree of similarity between different regions across species, we compared expression profiles of regional one-to-one orthologs. We combined *z*-score normalized expression data of 1,559 one-to-one orthologous genes with regional expression in at least one of the species and performed a clustering analysis (fig. 2*a*). The most dominant signal is the separation between the gonadal regions (P6–P8, C5–C11) and the rest of the worms. Specifically, the majority of genes is highly expressed in the gonadal regions of both species resulting in a joint clustering of data from both species (fig. 2*a*). In contrast, outside the gonadal regions, the clustering appears to be predominantly driven by species-specific signals, such as varying levels of intestinal or cuticular expression and similarities between head and tail regions (fig. 2*a*). Interestingly, the most anterior regions of *C. elegans* (C1–C4) show hundreds of genes with highly specific expression that are absent in *P. pacificus* suggesting substantial divergence between the two species. These genes are mostly involved in muscle-related functions, such as mitochondrion ($P < 10^{-15}$), striated muscle ($P < 10^{-11}$), and locomotion ($P < 10^{-3}$) (Huang et al. 2009). This difference is consistent with a recent morphological study showing a pronounced shift in the *P. pacificus* lineage from muscular to glandular tissues in the mouth regions (Riebesell and Sommer 2017). To quantify the amounts of regional genes with highly conserved expression more explicitly, we counted the number of shared one-to-one orthologs in every pairwise comparison (fig. 2*b* and supplementary table S5, Supplementary Material online). Surprisingly, only a few dozens of one-to-one orthologs are shared between regions. This finding suggests substantial expression divergence, because overall, 25–30% of the *P. pacificus* genes have one-to-one orthologs in *C. elegans* (Rödelsperger, Athanasouli, et al. 2019). Thus, regionally expressed genes seem to be largely composed of one-to-one orthologs with divergent expression and novel genes.

Conserved and Symmetric Expression Patterns Support the Homology of Anatomical Structures between *P. pacificus* and *C. elegans*

Although the overall level of conserved regional expression is low, the comparison of regional expression based on shared orthologs also provides some indication how the different regions in both nematodes correspond to each other. In particular, the head and tail regions exhibit a high similarity across the data sets of both species. For example, the *P. pacificus* region P2 shares 40 regional one-to-one orthologs with *C. elegans* region C2, but also 35 with *C. elegans* region C1 (fig. 2*b*). This shows that there is no unambiguous mapping between individual regions in the anterior part of *P. pacificus* to their counterparts in *C. elegans*. However, as a single unit, the three most anterior *P. pacificus* regions together (P1–P3) share dozens of one-to-one orthologs specifically with the most anterior regions in *C. elegans* (C1–C3) and thus, represent distinct segments of the heads (fig. 2*b*). Also, the intestinal regions P4 and C4 from both species seem to correspond to one another, and the most posterior regions (P11 and C12) show highly significant sharing of one-to-one orthologs (fig. 2*b*). Within regions representing the gonads (consisting of both the somatic and germline gonad), the signal is obscured partially by the symmetry of gonad structures and the anatomical differences between the species (fig. 1*a* and *b*). To illustrate this, we tested for gene sharing between regions of the same species. This perfectly recapitulated the symmetry of the gonads, as many of these regions exhibit highly significant gene sharing with themselves and one additional region. For example, both MSP expressing regions (P5 and P9 in *P. pacificus* and C6 and C10 in *C. elegans*) have hundreds of genes in common (fig. 2*c* and *d*). In the case of *C. elegans*, sperm-related signals seem to be stronger mixed with general gonadal expression than in *P. pacificus* because the global transcriptomic profiles in *C. elegans* are more similar to other gonadal regions and the amount of shared regional genes is higher (figs. 1*e*, 1*f*, and 2*d*). Thus, even though the overall level of conserved regional expression is rather low, it is still sufficient to identify homologous anatomical regions across species as follows: head, intestine, general gonad, sperm-related, and tail regions.

Spatial Expression Patterns Reveal Regional Clustering of Large Gene Families

The limited degree of conservation at the level of one-to-one orthologous genes (fig. 2*b*) is indirect evidence that certain novel gene families are enriched among regional genes. This would be consistent with a previous studies showing that large portions of developmentally regulated genes in

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

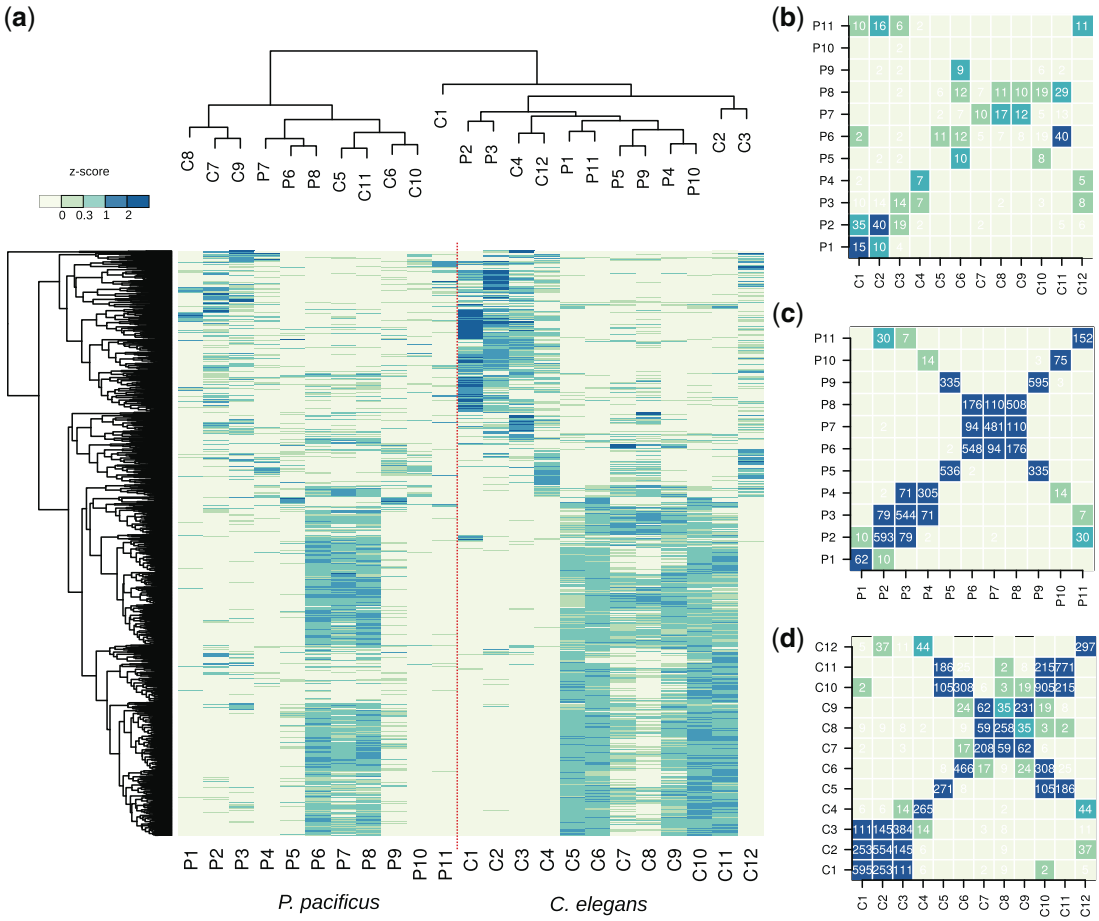


FIG. 2. Limited conservation of highly spatially restricted expression. (*a*) The heatmap shows the median *z*-score-normalized expression values for one-to-one orthologs of regional genes in both species. Joint clustering of regions separates the germline from the rest of the worms. (*b*–*d*) The heatmaps show the numbers of shared one-to-one orthologs (*b*) or shared genes in *Pristionchus pacificus* (*c*) and *Caenorhabditis elegans* (*d*) between different regions. The color code scales with the significance of the overlap (Fisher's exact test).

C. elegans and *P. pacificus* originated from lineage-specific duplication events and therefore, do not have one-to-one orthologs (Castillo-Davis and Hartl 2002; Baskaran et al. 2015). Thus, we tested regional genes for an overrepresentation of gene families as defined by the presence of certain protein domains (fig. 3*a* and *b*). Consistent with previous analysis based on reporter lines, we found a significant enrichment (false discovery rate-corrected *P* value <0.05) of astacin genes in the head region P2 of *P. pacificus* (Sieriebriennikov et al. 2020), which is not observed in the *C. elegans* data (fig. 3*a* and *b*). This example of expression divergence might be an indirect effect caused by cellular heterogeneity across species as the corresponding gland cell expressing these astacins is greatly expanded in *P. pacificus* (Riebesell and Sommer 2017; Sieriebriennikov et al. 2020). Overall, this analysis revealed that most regions are strongly biased toward the expression of members of a particular gene family. The numbers of such lineage-specific duplicates can by far exceed the amount of one-to-one orthologs in a given region (fig. 3*a* and *b*). For example, although regions P4 and C4 only share seven one-to-one orthologs (fig. 2*b*), P4 and C4 contain 26 and 9 regionally expressed C-type lectins. C-type lectins have been implicated in the innate immunity of *C. elegans* and were previously found to be expressed in the intestines of both nematodes (Schulenburg et al. 2008; Lightfoot et al. 2016). Similarly, MSP expressing regions share almost no one-to-one orthologs, yet both nematode species express between 19 and 54 different MSPs in a regional manner (fig. 3*a* and *b*). Thus, gene families that have undergone recent duplications show strong regional clustering in the spatial transcriptomes of *P. pacificus* and *C. elegans*.

Previously Uncharacterized Gene Families Are Expressed in Sperm-Related Regions

Although many of the enriched gene families such as C-type lectins, MSPs, vitellogenins, and kinases have well-characterized molecular or biological functions, we observed multiple domains of unknown function (DUF) with regional expression as well (fig. 3*a* and *b*). DUFs mostly represent evolutionary constrained sequence stretches in lineage-specific clusters of orthologous and paralogous genes (El-Gebali et al. 2019). On a functional level, these novel gene families are still poorly characterized, probably because different paralogs are functionally redundant or because large-scale RNAi screens did not yield obvious developmental or morphological phenotypes (Kamath et al. 2003). For example, the

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

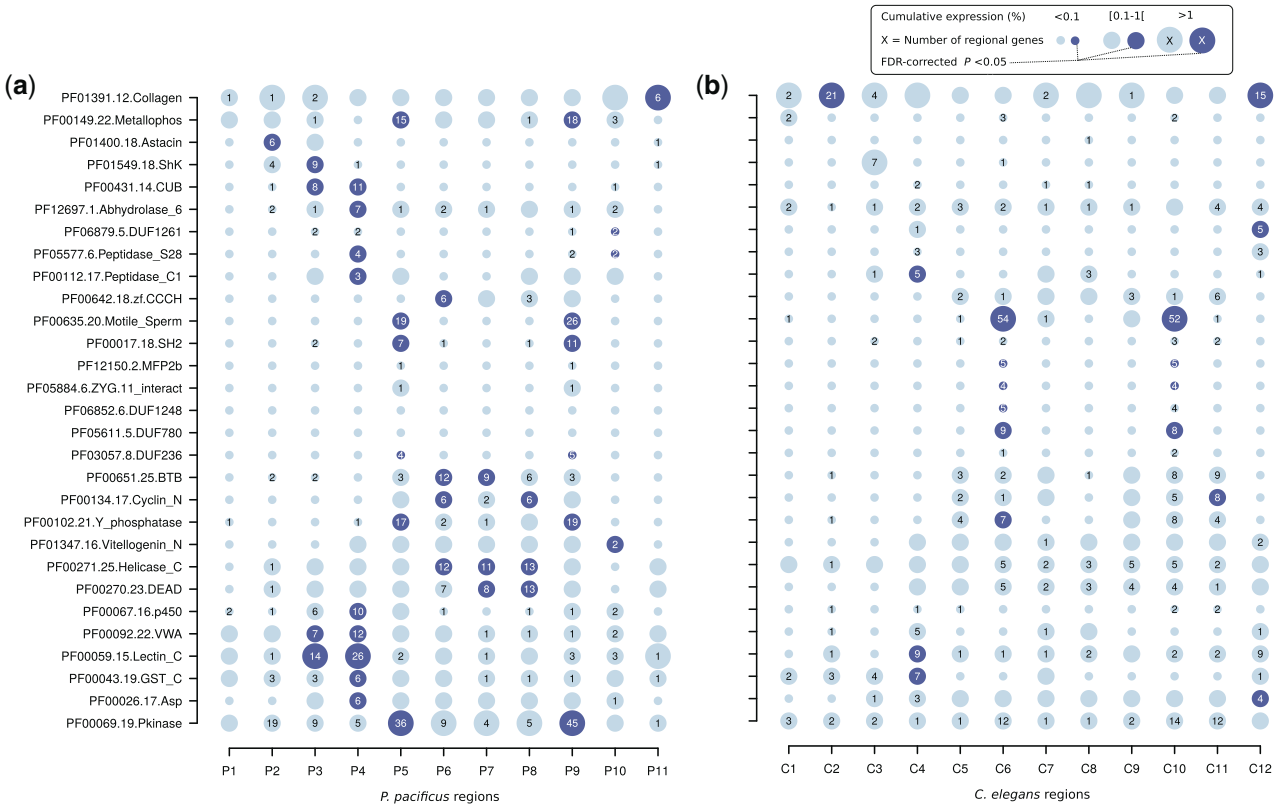


Fig. 3. Overrepresentation of protein domains among regional genes. (*a*, *b*) We identified 29 protein domains that are significantly enriched in regional genes in at least one of the species. The plots show the number of regional genes, the cumulative expression, and an indication of significance for a given protein domain in *Pristionchus pacificus* (*a*) and *Caenorhabditis elegans* (*b*).

spatial expression profile of DUF1261-containing genes is suggestive of intestinal expression (fig. 3*a* and *b*) and indeed five of them were previously detected in the intestinal transcriptome of *C. elegans* (Lightfoot et al. 2016). Interestingly, DUF780, DUF236, and DUF1248 show the highest signal in MSP expressing sperm-related regions in at least one of the species (fig. 3*a* and *b*). It is important to note that all these gene families are largely restricted to nematodes and thus, homologous sequences are unknown from vertebrates and most insects (El-Gebali et al. 2019). Following the idea that regional expression could be an indicator of potential biological function (Ebbing et al. 2018), we would hypothesize that the coexpression of the three gene families DUF780, DUF236, and DUF1248 together with MSPs possibly indicates a role in sperm cells (fig. 3*a* and *b*).

The Phylogenies of Sperm-Related Gene Families Are Consistent with Recent Expansions

Given that reproductive genes are frequently associated with rapid evolution (Swanson and Vacquier 2002; Kasimatis and Phillips 2018), we tested whether the phylogenies of the corresponding gene families are indeed consistent with rapid evolution and exhibit signatures of recent lineage-specific gene expansions. To this end, we reconstructed phylogenies of members of all three families and compared these with the phylogeny of DUF1261-containing proteins (as an example for a DUF family without sperm-related expression profile, fig. 4*a*). In addition, we generated phylogenies for two other gene families that are also coexpressed with MSPs, metallophosphatases including *C. elegans gsp-3* and homologs of MSP fiber protein 2 (MFP2), which are involved in sperm motility (Buttery et al. 2003; Wu et al. 2012) (fig. 4*b* and *c*). This analysis revealed that individual members of the DUF1261 gene family are perfectly preserved as one-to-one orthologs between *P. pacificus* and *C. elegans* (fig. 4*a*). In contrast, the two sperm-related gene families are consistent with recent gene expansions (fig. 4*b* and *c*). For example, four out of five of the sperm-related MFP2 genes from *C. elegans* arose very recently from duplication after the split from its sister species *Caenorhabditis inopinata* (C. sp34, fig. 4*b*) (Kanzaki et al. 2018). Similarly, sperm-related metallophoshophatases in *P. pacificus* tend to cluster in lineage-specific groups of paralogs which is a common sign of recent duplications. Also, in the case of the three uncharacterized but putatively sperm-related gene families DUF780, DUF236, and DUF1248, clusters of paralogous genes most-likely point toward recent lineage-specific duplication (fig. 4*d*–*f*). We conclude from these observations that the phylogenies of all five sperm-related gene families show signatures of recent expansions that are consistent with rapid evolution.

Spatially Resolved Phylostratigraphy Supports the *Out of Testis* Hypothesis in Nematodes

Genomic studies in *Drosophila* and mammals revealed that novel genes frequently exhibit testis-biased expression, resulting in the "out of testis" hypothesis, which postulates that

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

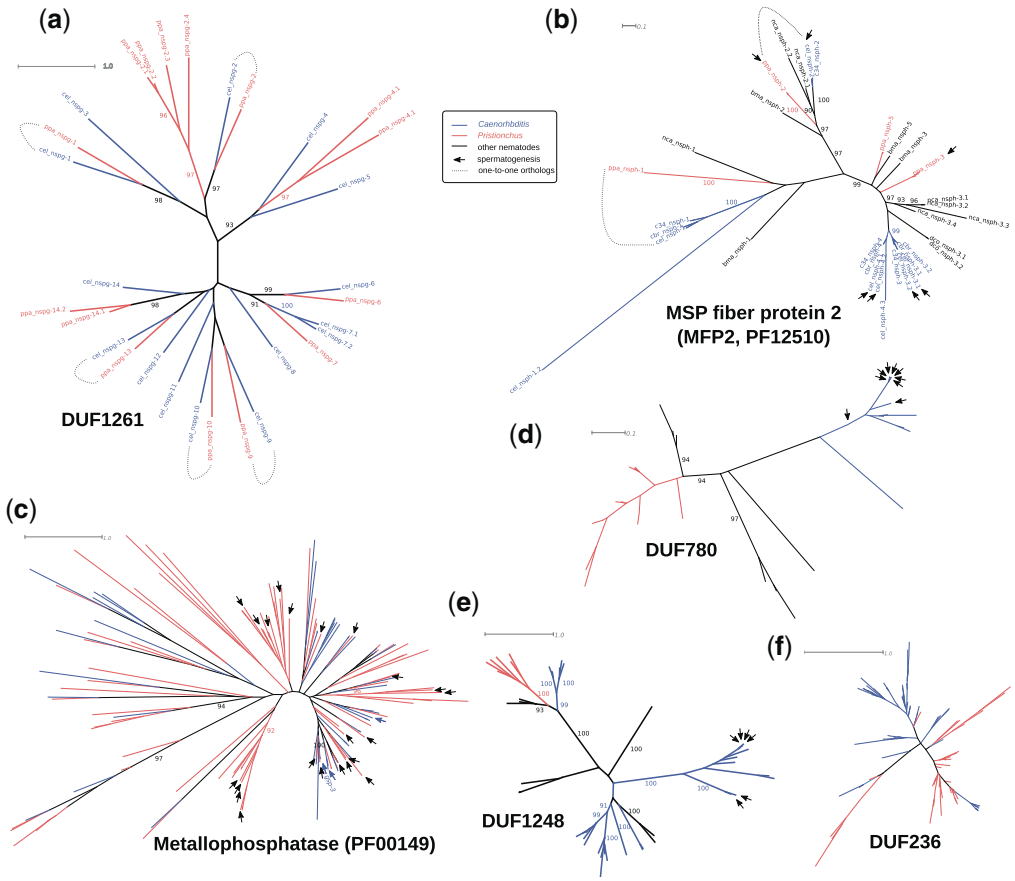


Fig. 4. Rapid gene expansion of sperm-related gene families. (*a*) DUF1261 represents an uncharacterized gene family that is expressed in the nematode intestine. A number of genes within this gene family are preserved as one-to-one orthologs between *Pristionchus pacificus* and *Caenorhabditis elegans*. (*b*, *c*) Members of the MFP2 (*b*) and metallophosphatase (*c*) gene families are known to be involved in spermatogenesis and sperm motility. Multiple sperm-related genes in both families are found in lineage-specific clusters and likely arose by recent gene duplication. (*d*–*f*) The phylogenies of other uncharacterized gene families also exhibit strong signatures of recent gene expansions and sperm-related genes are found in *Caenorhabditis*-specific subtrees of DUF780 (*d*) and DUF1248 (*e*). DUF236 containing genes usually have multiple repeated DUF236 domains; the phylogenetic tree (*f*) represents the relationship of individual repeat units. Thus, individual sperm-related genes cannot be indicated.

testis might represent a special environment facilitating the formation of novel genes (Levine et al. 2006; Soumillon et al. 2013; Witt et al. 2019). The above-mentioned examples of the novel gene families (DUF780, DUF236, and DUF1248) that are expressed in sperm-related regions provide first evidence that the "out of testis" hypothesis may also hold true in nematodes. To provide additional support for the "out of testis" hypothesis in nematodes, we systematically assessed the amount of novel genes that are expressed in different anatomical structures, by performing a phylostratigraphic analysis of all regionally expressed genes (Domazet-Loso et al. 2007). To this end, we combined comparative genomic data from multiple non-*Pristionchus* species with the genomes of seven additional *Pristionchus* species that were recently sequenced on the same platform with similar protocols to study the evolutionary dynamics of novel gene families (Prabh et al. 2018) (fig. 5*a*). The underlying ladder-like phylogeny enables us to date the emergence of a gene based on the presence/absence patterns of homologous sequences in other taxa. This defined phylostrata (age classes), which can be quantified across anatomical regions (fig. 5*a*). For example, phylostratum I denotes genes that are specific to *P. pacificus*, whereas genes of other phylostrata are shared with more distantly related species and are therefore to be considered older. The *P. pacificus* data set supports the "out of testis" hypothesis as regions P5 and P9 show the overall highest ratio of young phylostrata (I–XIII, fig. 5*a*). We repeated the phylostratigraphic analysis for *C. elegans* using a different phylogenomic context, which included the high-quality *Caenorhabditis briggsae* genome and the most basal *Caenorhabditis* species *C. monodelphis* (Stein et al. 2003; Slos et al. 2017). Also, in this data set, the sperm-related regions (C6 and C10) exhibit the highest fraction of young phylostrata (I–VIII, fig. 5*b*). To further support this observation, we reanalyzed tomo-seq data of *C. elegans* males (Ebbing et al. 2018). Consistent with the fact that *C. elegans* males have only one gonadal arm, we observe a single peak of young phylostrata in region M7 which also expresses MSP genes (fig. 5*c*). Interestingly, region M9 that overlaps the seminal vesicle and vas deferens (Ebbing et al. 2018) also exhibits high levels of young phylostrata suggesting a role of novel genes in sperm maturation and seminal fluid function (fig. 5*c*). To further narrow down the location of highest expression of novel genes, we repeated this analysis with a previously

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

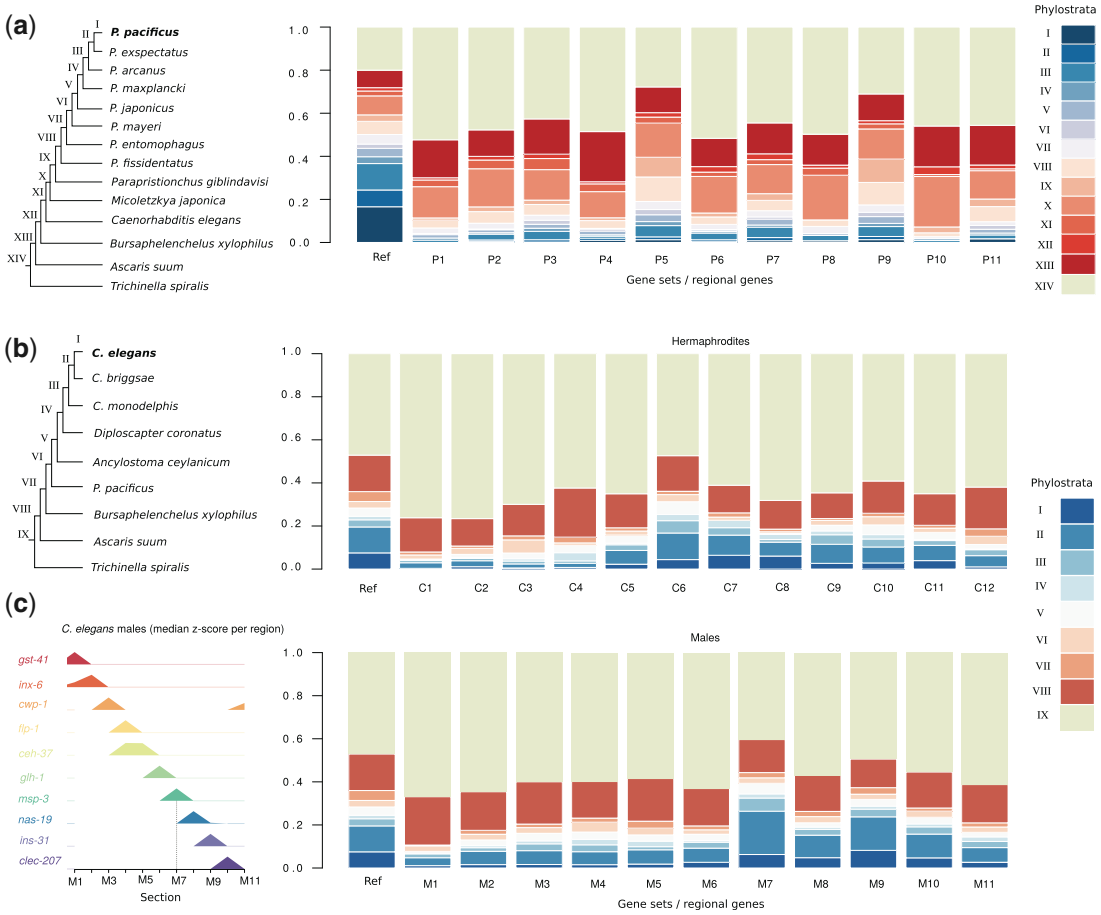


Fig. 5. Expression of novel genes in sperm-related regions. (*a*, *b*) We classified *Pristionchus pacificus* genes into gene ages by means of phylostratigraphy and plotted the fraction of phylostrata across the different regional gene sets. The stacked bars labeled as "Ref" show the genome-wide, phylostratigraphic distribution which is dependent on the phylogenetic resolution around the focal species. Sperm-related regions in *P. pacificus* (*a*, P5, P9) and *Caenorhabditis elegans* (*b*, C6, C10) showed the highest fraction of young age classes which is consistent with the "out of testis" hypothesis. (*c*) Reanalysis of tomo-seq data of *C. elegans* males identifies 11 regions (M1–M11) and confirms that the MSP expressing region (M7) indeed shows the highest fraction of young phylostrata.

published *C. elegans* tomo-seq data set consisting of ten gonadal regions (Tzur et al. 2018). Although expression of MSP was not detectable in hermaphrodite data of the investigated developmental stage (L4 + 24 h), we observed a pronounced peak of young phylostrata in regions of crossover formation and desynapsis (supplementary fig. S2, Supplementary Material online). These findings indicate that meiotic spermatocytes exhibit the highest levels of young phylostrata and they are consistent across data sets generated from different laboratories, nematode sexes, and species. Together, our observations provide strong support for the "out of testis" hypothesis in nematodes.

Young *Out of Testis* Genes May Result from Pervasive Transcription and Evolve Rapidly

Despite widespread support for the "out of testis" hypothesis in insects and mammals, it is still not entirely clear to what extent novel gene formation is driven by sexual selection or by an overall permissive chromatin state. A comprehensive, phylotranscriptomic analysis in mammals supported the idea that the observed pattern is largely a product of the permissive chromatin state resulting in pervasive transcription (Soumillon et al. 2013). In agreement with this, we found that not only are more novel genes expressed in sperm-related regions, but that sperm-related regions also tend to express the overall highest numbers of genes irrespective of the gene's age (supplementary fig. S3, Supplementary Material online). This observation suggests that the expression of novel genes in nematodes is also, at least partially, a byproduct of the higher transcriptomic complexity in sperm-related regions. To test if sperm-related genes in *C. elegans* are actually functional, we used signatures of purifying selection as indirect evidence of function (Prabh and Rödelsperger 2016). Comparison of nonsynonymous versus synonymous substitutions (dN/dS) of orthologous pairs between *C. elegans* and *C. briggsae* indicated that basically all *C. elegans* genes with *C. briggsae* orthologs exhibit strong purifying selection, which is indirect evidence of biological function (Prabh and Rödelsperger 2016) (fig. 6). Contrasting different phylostrata, we found that younger genes are generally less constrained than genes of more ancient phylostrata, which is consistent with previous phylogenomic analysis of *P. pacificus and Pristionchus exspectatus* orthologs (Prabh et al. 2018) (fig. 6*a*). When estimating selective constraint across the

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

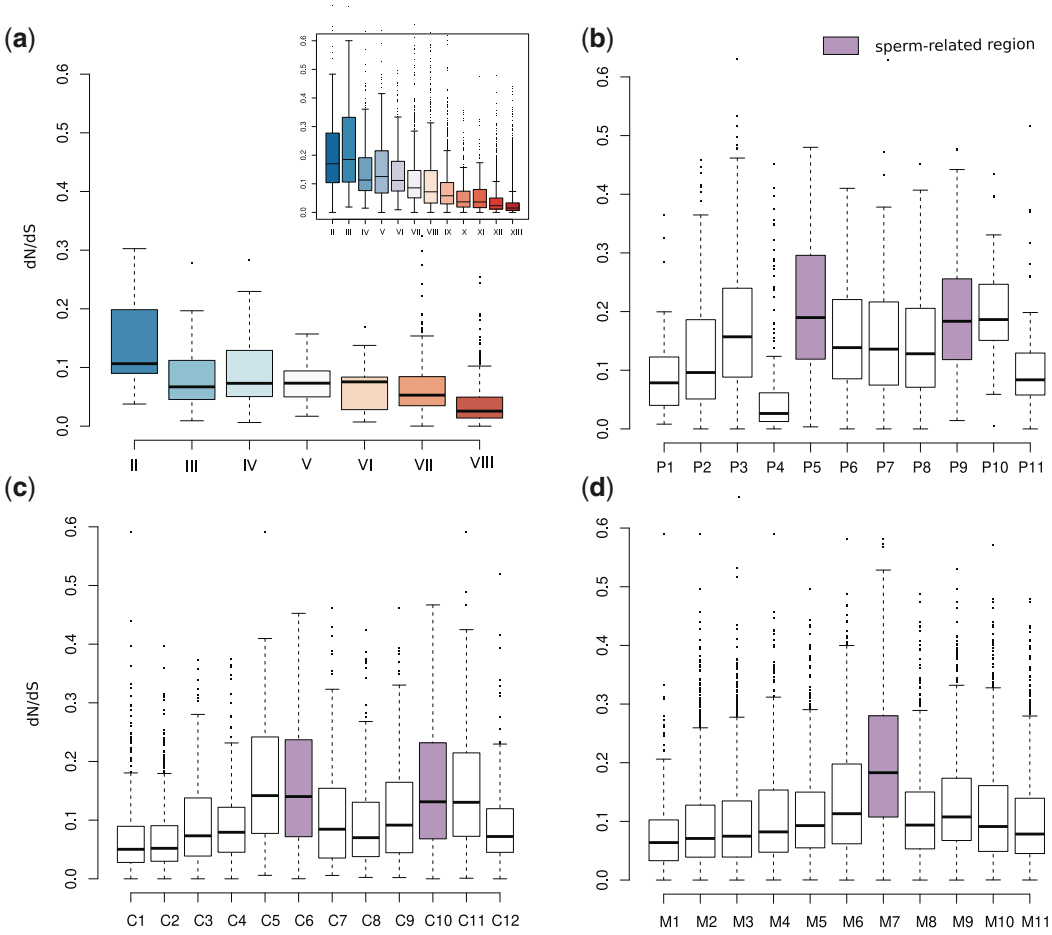


Fig. 6. Relaxed evolutionary constraint in sperm-related regions. (*a*) The boxplots show median and interquartile ranges of dN/dS ratios inferred from orthologs between *Caenorhabditis elegans* and *C. briggsae* for different phylostrata. The inset shows equivalent data for *Pristionchus pacificus* and *P. exspectatus* orthologs. Generally, older genes underlie stronger purifying selection than more recent gene classes. (*b–d*) The distribution of dN/dS values across sets of regional genes in *P. pacificus* hermaphrodites (*b*), *C. elegans* hermaphrodites (*c*), and *C. elegans* males (*d*) indicates that sperm-related regions have high ratios of genes underlying relaxed evolutionary constraint.

sets of regional genes, we found in all three data sets that sperm-related regions exhibit high ratios of genes under relaxed evolutionary constraint (fig. 6*b–d*). Thus, pervasive transcription but also rapid evolution, seem to be inherent features of sperm-related regions.

Young *Out of Testis* Genes Affect Spermatogenesis and Meiosis

Finally, to further support the idea that recently evolved "out of testis" genes are not just a byproduct of pervasive transcription but actually got integrated into the biology of their host (Rödelsperger, Prabh, et al. 2019), we exploited the enormous genetic knowledge of *C. elegans* to screen for novel genes with well-characterized biological functions. Although the largest portion (~90%) of novel genes (Phylostrata I–III) has not even been characterized in *C. elegans*, we found a number of recently evolved genes with functions in spermatogenesis and meiosis (table 1). Notably, we found regional expression of several meiosis-related genes such as *him-5* (Meneely et al. 2012), *lab-1* (Tzur et al. 2012), *syp-2* and *syp-3* (Schild-Prüfert et al. 2011), *kca-1* (McNally et al. 2012), *mesp-1* (Wolff et al. 2016), *brc-2* (Checchi et al. 2014), and *ify-1* (Kitagawa et al. 2002) in sperm-related regions C6 and C10 of *C. elegans* hermaphrodites. The same regions also express other spermatogenesis-related genes such as *spe-11* (L'Hernault et al. 1988) and *fog-2* (Nayak et al. 2005). In region CM7 of *C. elegans* males, we found expression of multiple novel spermatogenesis and meiosis-related genes such as *spe-12* (Nance et al. 1999), *spe-27* (Minniti et al. 1996), *spe-11* (L'Hernault and Singson 2000), *spe-19* (Geldziler et al. 2005), *spe-38* (Chatterjee 2005), *lab-1* (Tzur et al. 2012), and *ify-1* (Kitagawa et al. 2002) as well as nine members of another uncharacterized nematode-specific protein family (NSP group A, table 1). These results demonstrate that multiple seemingly newly arisen genes have been integrated into the biology of their host and modulate universal processes including spermatogenesis and meiosis.

Discussion

Nematodes are one of the most species-rich animal phyla and with *C. elegans*, they harbor one of the best studied model organisms (Lambshead 1993; Blaxter 1998). Many of the free-living nematodes, including *C. elegans*, are characterized by their small body, which was one the key features for the

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

Table 1. Candidate List of Novel, Sperm-Related *Caenorhabditis elegans* Genes.

Category	*C. elegans* Genes
Meiosis	*him-5*, *cdc-26*, *lab-1*, *kca-1*, *syp-2*, *syp-3*, *mesp-1*, *brc-2*, *ify-1*, *hcp-2*, and *rsa-2*
Spermatogenesis	*fog-2*, *spe-27*, *spe-19*, *spe-11*, *spe-38*, and *sss-1*
Other F-box proteins	*fbxb-38*, *fbxb-36*, *fbxc-20*, *fbxa-201*, *fbxa-196*, *fbxa-218*, *fbxa-125*, *fbxc-44*, and *fbxb-52*
Nematode-specific protein group A	*nspa-1*, *nspa-2*, *nspa-3*, *nspa-4*, *nspa-5*, *nspa-6*, *nspa-7*, *nspa-8*, and *nspa-10*
Germ granules	*pgl-1*, *pgl-2*, and *pgl-3*
Others	*sdz-27*, *hil-8*, *cec-2*, *numr-2 egl-1*, *dlc-6*, *kbp-1*, *mut-16*, *hmg-12*, *hpo-40*, *gpr-1*, *gpr-2*, *clec-260*, and *clec-142*

NOTE.—*Caenorhabditis elegans* genes of phylostrata I–III that have been characterized previously.

selection and success of *C. elegans* as a model system. However, for a long time, their small body size has also hindered the generation of large-scale tissue-specific expression catalogs that are comparable to mammals (Dupuy et al. 2007; Spencer et al. 2011; GTEx Consortium 2013; Smith et al. 2019). With the advent of spatial transcriptomics and single-cell sequencing, the nematode research community can now overcome this hindrance (Cao et al. 2017; Ebbing et al. 2018). Although single-cell RNA-seq still has obvious advantages in terms of resolution, spatial transcriptomics provides additional positional information that is very useful to disentangle signals from symmetric anatomical structures such as the hermaphroditic gonads. In this study, we combined cryosectioning of individual *P. pacificus* worms with RNA-seq in order to study the evolution of novel genes at spatial resolution and to establish a resource for comparative studies of gene function between *C. elegans* and *P. pacificus* (Markov et al. 2016; Moreno et al. 2017, 2018). Since individual sections represent a mix of heterogeneous tissues and cells (e.g., cuticle, neurons, intestine, somatic gonad, and germline), we focused on the analysis of regional genes with enriched expression in a given region relative to the whole worm in order to minimize contributions from other tissues. However, further experimental analysis would be needed to confirm the expression of these regional transcripts in the proposed tissue. By comparison with the spatial transcriptome of *C. elegans* hermaphrodites, we found that only a small number of one-to-one orthologs show conserved regional expression. This result resembles recent findings of expression divergence between the developmental transcriptomes of *C. elegans* and *C. briggsae* (Lu et al. 2020) and together, these results suggest that nematode transcriptomes harbor considerable species-specific signals at a spatial and temporal level. However, further analysis will be needed to disentangle to what extent the limited spatial conservation is caused by expression divergence, heterogeneous composition of cells and tissues across sections, lineage-specific duplications, or low expression levels that are below the current limits of detection (individual sections of our spatial transcriptome are sequenced at a substantially lower depth than samples in previous RNA-seq studies [Baskaran et al. 2015; Lightfoot et al. 2016]). At the same time, the limited level of conserved regional expression highlighted important differences between species, including the gain of spatially restricted expression of complete gene families. Focusing on sperm-related regions, we identified some previously uncharacterized gene families that might play roles in spermatogenesis or sperm cell biology (fig. 4) and we found broad support for the "out of testis" hypothesis in nematodes (fig. 5). This is particularly interesting as unlike *Drosophila* and most mammals, *C. elegans* and *P. pacificus* are androdioecious species where most individuals are hermaphrodites that are capable of self-fertilization and only a small percentage of a population is composed of males. Hermaphrodites of both species produce spermatids during larval development which are activated upon contact with oocytes that are formed after transitioning to the adult stage (L'Hernault 2006). Although these spermatids are transcriptionally and translationally quiescent (no de novo transcription and translation), transcripts are still detectable which allowed us to investigate the "out of testis" hypothesis in young adult hermaphrodites of both species. Note however, that the enrichment of novel genes in sperm-related regions was strongest in the spatial transcriptomic data of *C. elegans* males with active spermatogenesis (fig. 5*c* and supplementary fig. S2, Supplementary Material online). Interestingly, the evolution of selfing in nematodes led to genome contractions and gene losses that preferentially affect male-biased and sperm-related genes (Fierst et al. 2015; Rödelsperger et al. 2018; Yin et al. 2018). Thus, the analysis of gonochoristic sister species could reveal sperm-related expression of further novel genes that have been lost in the *P. pacificus* and *C. elegans* lineages.

Gene duplication is a major source for generating novel genes and functions (Ohno 1970), but other processes such as divergence and de novo formation can also generate novel orphan genes (Schlötterer 2015; McLysaght and Hurst 2016; Van Oss and Carvunis 2019) and first cases of these processes have been recently characterized in nematodes (Prabh and Rödelsperger 2019; Rödelsperger, Prabh, et al. 2019; Zhang et al. 2019). The spermatogenesis-related gene *fog-2* is a good example of sequence divergence. Mutations in *fog-2* transform hermaphrodites into females that are not capable of self-fertilization (Schedl and Kimble 1988). *fog-2* arose recently via gene duplication in *C. elegans* and has no one-to-one ortholog in *C. briggsae* (Nayak et al. 2005). According to our phylostratigraphic analysis, it is classified as *Caenorhabditis* specific. However, the FOG-2 protein and also other sperm-related genes (table 1) contain a widely distributed *F-box* motif that is also recognized by searches using profile hidden Markov models (El-Gebali et al. 2019), supporting the notion that the absence of sequence similarity is due to divergence. Similarly, BRC-2 has been shown to be homologous to the mammalian tumor suppressor BRCA2 (Bork et al. 1996), suggesting that the actual gene age of highly

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

divergent members of ancient gene families can be underestimated in phylostratigraphic analysis (Moyers and Zhang 2015). However, even if more sensitive homology searches can still detect some level of sequence similarity, we would argue that such a sequence has diverged enough from its ancestor to be called novel (Rödelsperger, Prabh, et al. 2019). On top of the lack of strong evolutionary constraint, positive selection can additionally drive sequence divergence, especially in the context of reproductive genes (Swanson and Vacquier 2002; Kasimatis and Phillips 2018). Consistent with this, previous analysis has found evidence for positive selection acting on *fog-2* (Nayak et al. 2005). Interestingly, *fog-2* is not the only F-box gene and many of its paralogs are also expressed in sperm-related regions (table 1). Thus, following the argument that spatial expression is indicative for biological function (Ebbing et al. 2018), we would also hypothesize that these F-box genes might play some previously undescribed roles during spermatogenesis. Nevertheless, not every gene that is found to be coexpressed with MSPs is likely to perform biological functions in sperm cells. The high transcriptional complexity in sperm-related regions as a result of a permissive chromatin state may lead to the expression of putative de novo genes without any specific function. Thus, the transcriptome of sperm-related regions might harbor a mix of many classes of novel genes some of which are simply byproducts of pervasive transcription. Even if such sequences do not have any biological function, they form the raw material out of which new genes may evolve. Some of these potential de novo genes may start to interact with other cellular components and gain biological functions (Keeling et al. 2019). At the same time, recent gene duplicates and rapidly evolving members of ancient gene families also contribute to this heterogenous pool of novel genes that is highly expressed in sperm-related regions. This is further supported by the recent finding in mammals that even though error correction processes such as transcriptional scanning exist, male reproductive genes exhibit an unusually high level of sequence divergence (Xia et al. 2020). In future, more detailed analysis and functional studies of individual genes and gene families are needed to comprehensively characterize the roles of novel genes in sperm cell biology.

Conclusion

By comparison of the spatial transcriptomes between *P. pacificus* and *C. elegans*, we have shed light on the evolution of novel genes in a tissue-specific context. We present highly spatially restricted expression profiles for many previously uncharacterized gene families in *C. elegans* and *P. pacificus*, which might help future work to study their function in a more targeted context. Focusing on sperm-related genes, we provide first evidence in nematodes supporting the previously postulated "out of testis" hypothesis and demonstrate that the sequences expressed during spermatogenesis represent a mix of biologically relevant genes as well as potential byproducts of pervasive transcription. Furthermore, we find evidence for rapid evolution of sperm-related genes at the level of sequence and copy number, which reflects typical patterns of the evolution of reproductive genes. Exploiting the wealth of genetic knowledge in *C. elegans*, we could show that even universal cellular processes such as meiosis are affected by the birth of novel genes. Finally, our data set may be taken as a starting point to investigate how gene expression changes and genomic novelty coincide with morphological and behavioral diversity.

Materials and Methods

Worm Culturing and Sample Preparation

Pristionchus pacificus strain PS312 was grown on OP50 *Escherichia coli* strain on S-medium-agar petri plates (Stiernagle 2006; Werner et al. 2017). Bleached egg populations were seeded on to plates to obtain synchronized worm cultures. Worms were screened ~64 h post bleaching, to determine their developmental stage, under a high-magnification stereomicroscope. Young adults were sought out for, by looking out for the invaginated vulval shape and absence of embryos in the germline (Sommer 1997; Mok et al. 2015), and were selectively picked from the plates for subsequent cryosectioning and sample preparation steps.

Cryosectioning and Preparation of Sequencing Libraries

Young adult worms were straightened in cryosolution (OCT). The head and the tail were marked by a single blue bead (Affigel Blue Gel agarose beads). The worms were then frozen, within 10 min of preparation, on dry ice and stored at −80 °C prior to cryosectioning. The embedded worms were sectioned using a cooled histology cryotome (−20 °C). Sections of 20 μm width were made starting from the first blue bead until the last blue bead. Typically, a worm, in-between two blue beads, would comprise about 40 sections. The sections of one individual worm were placed in a cooled pre-prepared 96-well plate, previously stored at −80 °C. The 96-well plates contain barcoded primers used for the CELseq protocol; consisting of a PolyT stretch, a unique 8-nt section-specific barcode, a 4-nt random barcode, the 5′ Illumina adaptor, and a T7 promotor for in vitro transcription. 96-Well plates containing sectioned worms were sealed, spinned down at 300 rpm at 4 °C, and stored at −80 °C. mRNA processing, reverse transcription, in vitro transcription, and illumina sequencing library preparation were performed according to the robotized version of the CELseq2 protocol (Hashimshony et al. 2012, 2016; Muraro et al. 2016), by the Single Cell Facility at the Hubrecht Institute. Sample processing was performed within a week of cryosectioning.

Preparation of a Reference Transcriptome

As current gene annotations for *P. pacificus* do not include 3′UTRs (Rödelsperger, Athanasouli, et al. 2019), we generated a customized *P. pacificus* reference annotation based on a previously established strand-specific transcriptome assembly (Rödelsperger et al. 2016, 2018). To reduce isoform redundancy, we applied the following set of rules to pick one representative isoform per gene: 1) pick the isoform that had a poly-A tail which was defined as the last six nucleotides or

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

more being adenine, 2) in case of multiple isoforms with a poly-A tail, the isoform that had the most 3′ alignment to the best protein hit in *C. elegans* was chosen, 3) if neither rule 1 nor rule 2 could resolve the ambiguity, we picked the longest isoform. Open reading frames were then extracted from these sequences by taking the longest partial or complete translatable peptide sequence that was not interfered by a stop codon (Rödelsperger et al. 2018). We then applied an Inparanoid-like approach of detection of one-to-one orthologs with *C. elegans* (Remm et al. 2001). In summary, BlastP searches against combined protein data from both species were performed (*e*-value <0.001) and only best-reciprocal pairs with lowest cross-species *e*-value (self-hits were excluded) were reported to avoid calling one-to-one orthologs in the presence of inparalogs. Finally, to annotate protein domains in the reference annotation, we ran the hmmsearch program of the HMMER3.0 package (-E 0.001) (Mistry et al. 2013).

Analysis of Tomo-Seq Data Expression Data

Raw reads were demultiplexed, aligned to the reference annotation, and transcript counts were quantified as described in Ebbing et al. (2018). Transcript counts were normalized by the total number of transcripts per section. Subsequently, for each gene in a single worm, z-scores were calculated using the scale function as implemented in R on the transcript normalized expression values across all sections. To make sure that the *P. pacificus* and *C. elegans* data sets are maximally comparable, we reanalyzed the previously generated spatial transcriptomics data of *C. elegans* males and hermaphrodites (Ebbing et al. 2018) using the same normalization procedures and rules for defining the anatomical regions. We then used the expression pattern of selected marker genes to classify the sections into regions. Data from multiple individuals were combined by pooling all sections corresponding to a certain region and taking the median z-score from all these sections. Clustering of sections based on Pearson correlation and clustering of regions based on z-scores of one-to-one orthologous genes were done using complete linkage clustering as implemented in the hclust and heatmap functions of R. The tomo-seq data of *C. elegans* hermaphrodites and males were analyzed in the same way. Similarly, spatial transcriptomic data from the study of Tzur et al. (2018) were extracted from supplementary tables S2 and S10, Supplementary Material online, and were z-score normalized. For all data sets, expressed genes were defined as genes with a positive median normalized transcript count of all sections corresponding to a region and regional genes were defined as genes with a z-score >1. Overrepresentation analysis of protein domains (results of hmmsearch) was done by calculating the intersection between a set of regional genes and a given protein domain and quantifying the significance by performing a Fisher's exact test against a background of all expressed genes. To correct for multiple comparisons, *P* values were adjusted by the false discovery rate method.

Phylostratigraphic, Phylogenetic, and Selection Analysis

Phylostratigraphic analysis for *C. elegans* was performed by obtaining annotated protein sets for *C. elegans* (WS260), *C. briggsae* (WS260), *C. monodelphis* (JU1667 v1), *Diploscapter coronatus* (WBPS10), *Ancylostoma ceylanicum* (WS248), *P. pacificus* (WS260), *Bursaphelenchelus xylophilus* (WS248), *Ascaris suum* (WS248), and *Trichinella spiralis* (WS248). The set of *C. elegans* proteins (longest isoform per gene) were taken as query sequences and phylostrata were defined from the pairwise BlastP (version 2.6) searches against all other species (*e*-value <0.001) (Domazet-Loso et al. 2007). Phylostratigraphy for *P. pacificus* was done analogously using the predicted proteins from recently sequenced diplogastrid genomes (Prabh et al. 2018) and proteins from *C. elegans*, *Bursaphelenchelus xylophilus*, *Ascaris suum*, and *Trichinella spiralis* as target BlastP databases (*e*-value <0.001). To account for potentially incorrectly annotated open reading frames (ORFs) in the *P. pacificus* reference annotation, additional BlastN searches against the corresponding diplogastrid transcriptome assemblies (Rödelsperger et al. 2018) were performed (*e*-value $<10^{-5}$). Phylogenetic analysis of selected gene families was done by extracting *C. elegans* and *P. pacificus* protein sequences based on the protein domain annotation created by hmmsearch. Further sequences were included after manual Blast searches on pristionchus.org, wormbase.org, and parasite.wormbase.org. A multiple sequence alignment was then generated by the MUSCLE program (version 3.8.31) and maximum likelihood trees were reconstructed using the phangorn R package (pml and optim.pml function with the following parameters, model = "LG," optNni = TRUE, optBf = TRUE, and optInv = TRUE). One hundred bootstrap pseudoreplicates were calculated and edges with support values above 90 were labeled in the resulting trees using the Dendroscope and Inkscape softwares. In order to quantify the strength of selection, we first computed orthologous clusters between *C. elegans* and *C. briggsae* proteins, *P. pacificus* and *P. exspectatus* proteins, and *P. pacificus* and *Pristionchus fissidentatus* proteins with the help of orthAgogue (Ekseth et al. 2014). Subsequently, pairs of orthologous proteins were aligned by MUSCLE (version 3.8.31) (Edgar 2004) and translated into a codon alignment by the program pal2nal (version 14) (Suyama et al. 2006). We then ran the codeml program of the paml package (version 4) to estimate one dN/dS value for each orthologous pair (codeml model 0) (Yang 2007). For each query gene for *C. elegans* and *P. pacificus*, we selected the ortholog with lowest estimated dS as one-to-one ortholog for further selection analysis. dN/dS values for subsets of genes, for example, different phylostrata or sperm-related genes, were then visualized as boxplots in R. For better comparison of dN/dS values across anatomical regions and nematode genera, we chose a pair of *Pristionchus* species that is separated by similar evolutionary distance (measured as median dS) as *C. elegans* and *C. briggsae*. As median dS between *C. elegans* and *C. briggsae* was 2.1, the matching *Pristionchus* species pair corresponded to *P. pacificus* and *P. fissidentatus* (median dS = 1.9) which was

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

considerably more diverged than *P. pacificus* and *P. exspectatus* (median dS = 0.3).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank all members of the Sommer lab for helpful discussions. Moreover, we are grateful to Marco C. Betist for help with freezing the samples and to Judith Vivié and Mauro Muraro from Single Cell Discoveries for processing of the samples and library preparation. This work was funded by the research program (14NOISE01) of the Foundation for Fundamental Research on Matter (FOM) of the Netherlands Organization for Scientific Research (NWO) (A.E. and H.C.K.) and the Max Planck Society (C.R., D.R.S., M.O., and R.J.S).

Author Contributions

H.C.K., C.R., and R.J.S. performed conceptualization; A.E., D.R.S., M.O., R.J.S., and H.C.K. provided resources; C.R., A.E., D.R.S., and M.O. contributed to investigation; C.R. and A.E. performed data curation, formal analysis, and visualization; C.R. was responsible for the writing original draft; C.R., A.E., R.J.S., H.C.K., D.R.S., and M.O. were responsible for the writing review and editing; C.R. was responsible for project administration; C.R., H.C.K., and R.J.S. were responsible for supervision; and H.C.K. and R.J.S. were responsible for funding acquisition.

References

Baskaran P, Rödelsperger C, Prabh N, Serobyan V, Markov GV, Hirsekorn A, Dieterich C. 2015. Ancient gene duplications have shaped developmental stage-specific expression in *Pristionchus pacificus*. *BMC Evol Biol*. 15:185.

Bento G, Ogawa A, Sommer RJ. 2010. Co-option of the hormone-signalling module dafachronic acid-DAF-12 in nematode evolution. *Nature* 466(7305):494–497.

Blaxter M. 1998. *Caenorhabditis elegans* is a nematode. *Science* 282(5396):2041–2046.

Bork P, Blomberg N, Nilges M. 1996. Internal repeats in the BRCA2 protein sequence. *Nat Genet*. 13(1):22–23.

Buttery SM, Ekman GC, Seavy M, Stewart M, Roberts TM. 2003. Dissection of the *Ascaris* sperm motility machinery identifies key proteins involved in major sperm protein-based amoeboid locomotion. *Mol Biol Cell* 14(12):5082–5088.

Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, et al. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357(6352):661–667.

Carvunis A-R, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, Charloteaux B, Hidalgo CA, Barbette J, Santhanam B, et al. 2012. Proto-genes and de novo gene birth. *Nature* 487(7407):370–374.

Castillo-Davis CI, Hartl DL. 2002. Genome evolution and developmental constraint in *Caenorhabditis elegans*. *Mol Biol Evol*. 19(5):728–735.

Chatterjee I. 2005. The *Caenorhabditis elegans* spe-38 gene encodes a novel four-pass integral membrane protein required for sperm function at fertilization. *Development* 132(12):2795–2808.

Checchi PM, Lawrence KS, Van MV, Larson BJ, Engebrecht J. 2014. Pseudosynapsis and decreased stringency of meiotic repair pathway choice on the hemizygous sex chromosome of *Caenorhabditis elegans* males. *Genetics* 197(2):543–560.

Domazet-Loso T, Brajković J, Tautz D. 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet*. 23(11):533–539.

Dupuy D, Bertin N, Hidalgo CA, Venkatesan K, Tu D, Lee D, Rosenberg J, Svrzikapa N, Blanc A, Carnec A, et al. 2007. Genome-scale analysis of in vivo spatiotemporal promoter activity in *Caenorhabditis elegans*. *Nat Biotechnol*. 25(6):663–668.

Ebbing A, Vértesy Á, Betist MC, Spanjaard B, Junker JP, Berezikov E, van Oudenaarden A, Korswagen HC. 2018. Spatial transcriptomics of *C. elegans* males and hermaphrodites identifies sex-specific differences in gene expression patterns. *Dev Cell* 47(6):801–813.e6.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5):1792–1797.

Ekseth OK, Kuiper M, Mironov V. 2014. orthAgogue: an agile tool for the rapid prediction of orthology relations. *Bioinformatics* 30(5):734–736.

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res*. 47(D1):D427–D432.

Fierst JL, Willis JH, Thomas CG, Wang W, Reynolds RM, Ahearne TE, Cutter AD, Phillips PC. 2015. Reproductive mode and the evolution of genome size and structure in *Caenorhabditis* nematodes. *PLoS Genet*. 11(6):e1005323.

Geldziler B, Chatterjee I, Singson A. 2005. The genetic and molecular analysis of spe-19, a gene required for sperm activation in *Caenorhabditis elegans*. *Dev Biol*. 283(2):424–436.

Gruidl ME, Smith PA, Kuznicki KA, McCrone JS, Kirchner J, Roussell DL, Strome S, Bennett KL. 1996. Multiple potential germ-line helicases are components of the germ-line-specific P granules of *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A*. 93(24):13837–13842.

GTEx Consortium. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 45:580–585.

Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, Gennert D, Li S, Livak KJ, Rozenblatt-Rosen O, et al. 2016. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol*. 17(1):77.

Hashimshony T, Wagner F, Sher N, Yanai I. 2012. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep*. 2(3):666–673.

Herrmann M, Kanzaki N, Weiler C, Yoshida K, Rödelsperger C, Sommer RJ. 2019. Two new species of *Pristionchus* (Nematoda: Diplogastridae) include the gonochoristic sister species of *P. fissidentatus*. *J Nematol*. 51:1–14.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 4(1):44–57.

Junker JP, Noël ES, Guryev V, Peterson KA, Shah G, Huisken J, McMahon AP, Berezikov E, Bakkers J, van Oudenaarden A. 2014. Genome-wide RNA Tomography in the zebrafish embryo. *Cell* 159(3):662–675.

Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res*. 20(10):1313–1326.

Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, et al. 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421(6920):231–237.

Kanzaki N, Tsai IJ, Tanaka R, Hunt VL, Liu D, Tsuyama K, Maeda Y, Namai S, Kumagai R, Tracey A, et al. 2018. Biology and genome of a newly discovered sibling species of *Caenorhabditis elegans*. *Nat Commun*. 9(1):3216.

Kasimatis KR, Phillips PC. 2018. Rapid gene family evolution of a nematode sperm protein despite sequence hyper-conservation. *G3 (Bethesda)* 8:353–362.

Keeling DM, Garza P, Nartey CM, Carvunis A-R. 2019. The meanings of "function" in biology and the problematic case of de novo gene emergence. *eLife* 8:e47014.

Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. 2009. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet*. 25(9):404–413.

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

Kieninger MR, Ivers NA, Rödelsperger C, Markov GV, Sommer RJ, Ragsdale EJ. 2016. The nuclear hormone receptor NHR-40 acts downstream of the sulfatase EUD-1 as part of a developmental plasticity switch in *Pristionchus*. *Curr Biol*. 26(16):2174–2179.

Kitagawa R, Law E, Tang L, Rose AM. 2002. The Cdc20 homolog, FZY-1, and its interacting protein, IFY-1, are required for proper chromosome segregation in *Caenorhabditis elegans*. *Curr Biol*. 12(24):2118–2123.

Knowles DG, McLysaght A. 2009. Recent de novo origin of human protein-coding genes. *Genome Res*. 19(10):1752–1759.

Lambshead PJD. 1993. Recent developments in marine benthic biodiversity research. Oceanis 19:5–24.

Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc Natl Acad Sci U S A*. 103(26):9935–9939.

L'Hernault SW. 2006. Spermatogenesis. WormBook, ed. The C. elegans Research Community, WormBook. Available from: https://doi.org/10.1895/wormbook.1.85.1.

L'Hernault SW, Shakes DC, Ward S. 1988. Developmental genetics of chromosome I spermatogenesis-defective mutants in the nematode *Caenorhabditis elegans*. *Genetics* 120:435–452.

L'Hernault SW, Singson AW. 2000. Developmental genetics of spermatogenesis in the nematode *Caenorhabditis elegans*. In: The testis. New York: Springer. p. 109–119.

Lightfoot JW, Chauhan VM, Aylott JW, Rödelsperger C. 2016. Comparative transcriptomics of the nematode gut identifies global shifts in feeding mode and pathogen susceptibility. *BMC Res Notes* 9:142.

Lightfoot JW, Wilecki M, Rödelsperger C, Moreno E, Susoy V, Witte H, Sommer RJ. 2019. Small peptide–mediated self-recognition prevents cannibalism in predatory nematodes. *Science* 364(6435):86–89.

Lu MR, Lai C-K, Liao B-Y, Tsai IJ. 2020. Comparative transcriptomics across nematode life cycles reveal gene expression conservation and correlated evolution in adjacent developmental stages. *Genome Biol Evol*. 12(7):1019–1030.

Markov GV, Meyer JM, Panda O, Artyukhin AB, Claaßen M, Witte H, Schroeder FC, Sommer RJ. 2016. Functional conservation and divergence of daf-22 paralogs in *Pristionchus pacificus* dauer development. *Mol Biol Evol*. 33(10):2506–2514.

Mayer MG, Rödelsperger C, Witte H, Riebesell M, Sommer RJ. 2015. The orphan gene dauerless regulates dauer development and intraspecific competition in nematodes by copy number variation. *PLoS Genet*. 11(6):e1005146.

McLysaght A, Hurst LD. 2016. Open questions in the study of de novo genes: what, how and why. *Nat Rev Genet*. 17(9):567–578.

McNally KLP, Fabritius AS, Ellefson ML, Flynn JR, Milan JA, McNally FJ. 2012. Kinesin-1 prevents capture of the oocyte meiotic spindle by the sperm aster. *Dev Cell* 22(4):788–798.

Meneely PM, McGovern OL, Heinis FI, Yanowitz JL. 2012. Crossover distribution and frequency are regulated by him-5 in *Caenorhabditis elegans*. *Genetics* 190(4):1251–1266.

Minniti AN, Sadler C, Ward S. 1996. Genetic and molecular analysis of spe-27, a gene required for spermiogenesis in *Caenorhabditis elegans* hermaphrodites. *Genetics* 143(1):213–223.

Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res*. 41(12):e121.

Mok DZL, Sternberg PW, Inoue T. 2015. Morphologically defined substages of *C. elegans* vulval development in the fourth larval stage. *BMC Dev Biol*. 15:26.

Moreno E, Lenuzzi M, Rödelsperger C, Prabh N, Witte H, Roeseler W, Riebesell M, Sommer RJ. 2018. DAF-19/RFX controls ciliogenesis and influences oxygen-induced social behaviors in *Pristionchus pacificus*. *Evol Dev*. 20(6):233–243.

Moreno E, Sieriebriennikov B, Witte H, Rödelsperger C, Lightfoot JW, Sommer RJ. 2017. Regulation of hyperoxia-induced social behaviour in *Pristionchus pacificus* nematodes requires a novel cilia-mediated environmental input. *Sci Rep*. 7(1):17550.

Moyers BA, Zhang J. 2015. Phylostratigraphic bias creates spurious patterns of genome evolution. *Mol Biol Evol*. 32(1):258–267.

Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, van Gurp L, Engelse MA, Carlotti F, D, Koning EJP, et al. 2016. A single-cell transcriptome atlas of the human pancreas. *Cell Syst*. 3(4):385–394.e3.

Nance J, Minniti AN, Sadler C, Ward S. 1999. spe-12 encodes a sperm cell surface protein that promotes spermiogenesis in *Caenorhabditis elegans*. *Genetics* 152(1):209–220.

Nayak S, Goree J, Schedl T. 2005. fog-2 and the evolution of self-fertile hermaphroditism in *Caenorhabditis*. *PLoS Biol*. 3(1):e6.

Nelson LS. 1998. Disruption of a neuropeptide gene, flp-1, causes multiple behavioral defects in *Caenorhabditis elegans*. *Science* 281(5383):1686–1690.

Ohno S. 1970. *Evolution by gene duplication*. New York: Springer.

Prabh N, Rödelsperger C. 2016. Are orphan genes protein-coding, prediction artifacts, or non-coding RNAs? *BMC Bioinf*. 17(1):226.

Prabh N, Rödelsperger C. 2019. De novo, divergence, and mixed origin contribute to the emergence of orphan genes in Pristionchus nematodes. *G3 (Bethesda)* 9:2277–2286.

Prabh N, Roeseler W, Witte H, Eberhardt G, Sommer RJ, Rödelsperger C. 2018. Deep taxon sampling reveals the evolutionary dynamics of novel gene families in *Pristionchus* nematodes. *Genome Res*. 28(11):1664–1674.

Ragsdale EJ, Müller MR, Rödelsperger C, Sommer RJ. 2013. A developmental switch coupled to the evolution of plasticity acts through a sulfatase. *Cell* 155(4):922–933.

Remm M, Storm CE, Sonnhammer EL. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol*. 314(5):1041–1052.

Riebesell M, Sommer RJ. 2017. Three-dimensional reconstruction of the pharyngeal gland cells in the predatory nematode *Pristionchus pacificus*. *J Morphol*. 278(12):1656–1666.

Rödelsperger C. 2018. Comparative genomics of gene loss and gain in *Caenorhabditis* and other nematodes. *Comp Genomics*. 419–432.

Rödelsperger C, Athanasouli M, Lenuzzi M, Theska T, Sun S, Dardiry M, Wighard S, Hu W, Sharma DR, Han Z. 2019. Crowdsourcing and the feasibility of manual gene annotation: a pilot study in the nematode *Pristionchus pacificus*. *Sci Rep*. 9(1):18789.

Rödelsperger C, Menden K, Serobyan V, Witte H, Baskaran P. 2016. First insights into the nature and evolution of antisense transcription in nematodes. *BMC Evol Biol*. 16(1):165.

Rödelsperger C, Prabh N, Sommer RJ. 2019. New gene origin and deep taxon phylogenomics: opportunities and challenges. *Trends Genet*. 35(12):914–922.

Rödelsperger C, Röseler W, Prabh N, Yoshida K, Weiler C, Herrmann M, Sommer RJ. 2018. Phylotranscriptomics of *Pristionchus* nematodes reveals parallel gene loss in six hermaphroditic lineages. *Curr Biol*. 28(19):3123–3127.e5.

Rudel D, Riebesell M, Sommer RJ. 2005. Gonadogenesis in *Pristionchus pacificus* and organ evolution: development, adult morphology and cell–cell interactions in the hermaphrodite gonad. *Dev Biol*. 277(1):200–221.

Schedl T, Kimble J. 1988. fog-2, a germ-line-specific sex determination gene required for hermaphrodite spermatogenesis in *Caenorhabditis elegans*. *Genetics* 119(1):43–61.

Schild-Prüfert K, Saito TT, Smolikov S, Gu Y, Hincapie M, Hill DE, Vidal M, McDonald K, Colaiácovo MP. 2011. Organization of the synaptonemal complex during meiosis in *Caenorhabditis elegans*. *Genetics* 189(2):411–421.

Schlötterer C. 2015. Genes from scratch—the evolutionary fate of de novo genes. *Trends Genet*. 31(4):215–219.

Schulenburg H, Hoeppner MP, Weiner J, Bornberg-Bauer E. 2008. Specificity of the innate immune system and diversity of C-type lectin domain (CTLD) proteins in the nematode *Caenorhabditis elegans*. *Immunobiology* 213(3–4):237–250.

Sieriebriennikov B, Sun S, Lightfoot JW, Witte H, Moreno E, Rödelsperger C, Sommer RJ. 2020. Conserved nuclear hormone receptors

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

controlling a novel plastic trait target fast-evolving genes expressed in a single cell. *PLoS Genet*. 16(4):e1008687.
Slos D, Sudhaus W, Stevens L, Bert W, Blaxter M. 2017. *Caenorhabditis monodelphis* sp. n.: defining the stem morphology and genomics of the genus *Caenorhabditis*. *BMC Zool*. 2:4.
Smith CM, Hayamizu TF, Finger JH, Bello SM, McCright IJ, Xu J, Baldarelli RM, Beal JS, Campbell J, Corbani LE, et al. 2019. The mouse Gene Expression Database (GXD): 2019 update. *Nucleic Acids Res*. 47(D1):D774–D779.
Sommer RJ. 1997. Evolution and development–the nematode vulva as a case study. *BioEssays* 19(3):225–231.
Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, Barthès P, Kokkinaki M, Nef S, Gnirke A, et al. 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep*. 3(6):2179–2190.
Spencer WC, Zeller G, Watson JD, Henz SR, Watkins KL, McWhirter RD, Petersen S, Sreedharan VT, Widmer C, Jo J, et al. 2011. A spatial and temporal map of *C. elegans* gene expression. *Genome Res*. 21(2):325–341.
Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, et al. 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol*. 1(2):E45.
Stevens L, Félix M-A, Beltran T, Braendle C, Caurcel C, Fausett S, Fitch D, Frézal L, Gosse C, Kaur T, et al. 2019. Comparative genomics of 10 new *Caenorhabditis* species. *Evol Lett*. 3(2):217–236.
Stiernagle T. 2006. Maintenance of *C. elegans*. WormBook, ed. The C. elegans Research Community, WormBook. Available from: https://doi.org/10.1895/wormbook.1.101.1.
Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 34(Web Server):W609–W612.
Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nat Rev Genet*. 3(2):137–144.
Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. *Nat Rev Genet*. 12(10):692–702.
Toll-Riera M, Bosch N, Bellora N, Castelo R, Armengol L, Estivill X, Mar Alba M. 2008. Origin of primate orphan genes: a comparative genomics approach. *Mol Biol Evol*. 26(3):603–612.
Tzur YB, Egydio de Carvalho C, Nadarajan S, Van Bostelen I, Gu Y, Chu DS, Cheeseman IM, Colaiácovo MP. 2012. LAB-1 targets PP1 and restricts Aurora B kinase upon entrance into meiosis to promote sister chromatid cohesion. *PLoS Biol*. 10(8):e1001378.
Tzur YB, Winter E, Gao J, Hashimshony T, Yanai I, Colaiácovo MP. 2018. Spatiotemporal gene expression analysis of the germline uncovers a syncytial expression switch. *Genetics* 210(2):587–605.
Vakirlis N, Hebert AS, Opulente DA, Achaz G, Hittinger CT, Fischer G, Coon JJ, Lafontaine I. 2018. A molecular portrait of de novo genes in yeasts. *Mol Biol Evol*. 35(3):631–645.
Van Oss SB, Carvunis A-R. 2019. De novo gene birth. *PLoS Genet*. 15(5):e1008160.
Werner MS, Sieriebriennikov B, Loschko T, Namdeo S, Lenuzzi M, Dardiry M, Renahan T, Sharma DR, Sommer RJ. 2017. Environmental influence on *Pristionchus pacificus* mouth form through different culture methods. *Sci Rep*. 7(1):7207.
Werner MS, Sieriebriennikov B, Prabh N, Loschko T, Lanz C, Sommer RJ. 2018. Young genes have distinct gene structure, epigenetic profiles, and transcriptional regulation. *Genome Res*. 28(11):1675–1687.
Wilecki M, Lightfoot JW, Susoy V, Sommer RJ. 2015. Predatory feeding behaviour in *Pristionchus* nematodes is dependent on phenotypic plasticity and induced by serotonin. *J Exp Biol*. 218(9):1306–1313.
Wissler L, Gadau J, Simola DF, Helmkampf M, Bornberg-Bauer E. 2013. Mechanisms and dynamics of orphan gene emergence in insect genomes. *Genome Biol Evol*. 5(2):439–455.
Witt E, Benjamin S, Svetec N, Zhao L. 2019. Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in *Drosophila*. *eLife* 8:e47138.
Witte H, Moreno E, Rödelsperger C, Kim J, Kim J-S, Streit A, Sommer RJ. 2015. Gene inactivation using the CRISPR/Cas9 system in the nematode *Pristionchus pacificus*. *Dev Genes Evol*. 225(1):55–62.
Wolff ID, Tran MV, Mullen TJ, Villeneuve AM, Wignall SM. 2016. Assembly of *Caenorhabditis elegans* acentrosomal spindles occurs without evident microtubule-organizing centers and requires microtubule sorting by KLP-18/kinesin-12 and MESP-1. *Mol Biol Cell* 27(20):3122–3131.
Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet*. 8(3):206–216.
Wu J-C, Go AC, Samson M, Cintra T, Mirsoian S, Wu TF, Jow MM, Routman EJ, Chu DS. 2012. Sperm development and motility are regulated by PP1 phosphatases in *Caenorhabditis elegans*. *Genetics* 190(1):143–157.
Xia B, Yan Y, Baron M, Wagner F, Barkley D, Chiodin M, Kim SY, Keefe DL, Alukal JP, Boeke JD, et al. 2020. Widespread transcriptional scanning in the testis modulates gene evolution rates. *Cell* 180(2):248–262.e21.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24(8):1586–1591.
Yin D, Schwarz EM, Thomas CG, Felde RL, Korf IF, Cutter AD, Schartner CM, Ralston EJ, Meyer BJ, Haag ES. 2018. Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. *Science* 359(6371):55–61.
Zhang W, Gao Y, Long M, Shen B. 2019. Origination and evolution of orphan genes and de novo genes in the genome of *Caenorhabditis elegans*. *Sci China Life Sci*. 62(4):579–593.

Downloaded from https://academic.oup.com/mbe/article/38/1/229/5891668 by Administrative Headquarters - MPS user on 08 August 2023

Developmental small RNA transcriptomics reveals divergent evolution of the conserved microRNA *miR-100* and the *let-7-complex* in nematodes

Devansh Raj Sharma[1,a], Waltraud Röseler[1], Hanh Witte[1], Michael S. Werner[1,2] and Ralf J. Sommer[1,*]

[1]Max Planck Institute for Biology, Tübingen, Germany
[2]School of Biological Sciences, University of Utah, Salt Lake City, Utah, USA

[a]Current address: Institute of Biological and Chemical Systems - Biological Information Processing, Karlsruhe Institute of Technology - Campus North, Eggenstein-Leopoldshafen, Germany

[*]Corresponding author: Ralf J. Sommer (ralf.sommer@tuebingen.mpg.de)

Abstract

Small RNAs have emerged as one of the key post-transcriptional regulators and can be broadly categorised into microRNAs (miRNAs), small-interfering RNAs (siRNAs), and PIWI-interacting RNAs (piRNAs). Since their discovery as regulators of developmental timing in the nematode *Caenorhabditis elegans*, some miRNAs were found to be highly conserved throughout animal evolution. miR-100 is a highly conserved miRNA, possibly predating the origin of bilaterians. Furthermore, *miR-100* is a member of the conserved *let-7-complex*, a locus containing three ancient miRNA-coding genes (*miR-100*, *let-7*, *miR-125*/*lin-4*) that has been secondarily lost in *C. elegans.* Here, we performed small RNA-sequencing across development in the satellite nematode *Pristionchus pacificus*, identifying miR-100 as the most abundant miRNA in post-embryonic juvenile stages. Additionally, *P. pacificus* exhibited a novel constellation of the *let-7-complex*, wherein *mir-100* is closely linked to *let-7*, but without *lin-4* in the same locus. Knockouts of *mir-100*, *let-7*, and *mir-100 let-7* double mutants in *P. pacificus* resulted in viable worms with developmental defects related to dysregulation of collagen. Surprisingly, *lin-4* mutants showed nearly normal development. This study provides the first developmental small RNA transcriptome in *P. pacificus*, and indicates that the *let-7-complex* and small RNA pathways undergo rapid evolutionary divergence in nematodes.

Keywords: *Pristionchus pacificus*, *C. elegans*, miRNAs, *let-7*, *miR-100*

Introduction

Development is a complex process, requiring a robust temporal and spatial regulation of gene expression (Nüsslein-Volhard and Wieschaus, 1980; Ambros, 1989; Han and Sternberg, 1990; Lawrence and Struhl, 1996). Over the last three decades, small RNAs have emerged as one of the key post-transcriptional regulators. Small RNAs can be broadly categorised into microRNAs (miRNAs), small-interfering RNAs (siRNAs), and PIWI-interacting RNAs (piRNAs). These RNAs bind to and guide Argonaute proteins for post-transcriptional repression of target genes. Since their discovery as regulators of developmental timing, first in the free-living nematode *Caenorhabditis elegans* and later in other animals (Lee, Feinbaum and Ambros, 1993; Pasquinelli *et al.*, 2000; Reinhart *et al.*, 2000), small RNAs have been implicated in a wide range of biological pathways, including development and ageing, protection against transposable elements, RNAi, cellular and tissue differentiation, relaying of environmental information, stress response, regulation of access to chromatin, metabolism, antiviral response, and transgenerational inheritance (Moazed, 2009; Ambros, 2011; Sokol, 2012; Sarkies and Miska, 2014; Rechavi and Lev, 2017; Ambros and Ruvkun, 2018; Berkhout, 2018; Kabekkodu *et al.*, 2018; Ketting and Cochella, 2021; Chen and Rechavi, 2022). Availability of high-quality genome assemblies and high-throughput sequencing has made studying small RNAs easier in different biological study systems.

Nematodes are arguably the most abundant phylum on earth (van den Hoogen *et al.*, 2020) and widely studied in life sciences. While the availability of molecular and genetic tools and resources in *C. elegans* has provided an opportunity to investigate the small RNA molecular machinery in detail (Fire *et al.*, 1998; Sijen *et al.*, 2001; Lee, Hammell and Ambros, 2006; Ruby *et al.*, 2006; Pak and Fire, 2007; Kato *et al.*, 2009; Beltran *et al.*, 2020; Houri-Zeevi *et al.*, 2020; Shukla *et al.*, 2020), an increasing number of small RNA studies in other nematodes has widened our understanding of the role of small RNAs in evolutionary and developmental biology, physiology, and in the case of parasitic nematodes, pathology (Wang *et al.*, 2011; Srinivasan *et al.*, 2013; Buck *et al.*, 2014; Zamanian *et al.*, 2015; Marks *et al.*, 2019). Despite the overarching similarities across the animal kingdom, nematodes show several unique features related to the molecular composition of small RNA-mediated silencing machinery. The most notable anomalies occur in the piRNAs, including their shorter length in comparison to other animals, clustered *vs.* dispersed arrangement in different species, replacement of ping-pong cycle with RdRP-mediated secondary siRNA generation to mediate gene silencing, and the absence of piRNAs altogether in several species (Brennecke *et al.*, 2007; Wang *et al.*, 2011; Bagijn *et al.*, 2012; Sarkies *et al.*, 2015; Holz and Streit, 2017; Beltran *et al.*, 2019, 2021). Another stark difference is the expansion of the Argonaute protein family along with the apparent loss of DNA methylation machinery in many, but not all, nematodes (Simpson, Johnson and Hammen, 1986; Gutierrez and Sommer, 2004; Yigit *et al.*, 2006; Gao *et al.*, 2012; Greer *et al.*, 2015; O'Brown *et al.*, 2019; Charlesworth *et al.*, 2021). In this context, the free-living, hermaphroditic diplogastrid nematode *Pristionchus pacificus*, originally developed as a satellite study system for comparative evolutionary developmental studies alongside *C. elegans* (Jungblut and Sommer, 2000; Sommer, 2005), is now a model organism in its own right, given the establishment of genetic, genomic, molecular, ecological, behavioural, phylogenetic, morphometric, and biochemical resources (Herrmann *et al.*, 2007; Dieterich *et al.*, 2008; Ogawa *et al.*, 2009; Witte *et al.*, 2015; Rödelsperger *et al.*, 2017; Han *et al.*, 2020; Theska *et al.*, 2020; Igreja *et al.*, 2022). Specifically, *P. pacificus* has emerged as a prime model system to investigate developmental plasticity due to its unique feeding polyphenism with two alternative mouth forms (Bento, Ogawa and Sommer, 2010; Ragsdale *et al.*, 2013; Levis and Ragsdale, 2023). Given that one of the two morphs allows predation on other nematodes, *P. pacificus* is also a new model system to investigate cannibalism, self-recognition and collective behaviour

(Lightfoot *et al.*, 2019; Hiramatsu and Lightfoot, 2023). The repertoire of small RNAs in *P. pacificus* has been examined in a comparative context alongside other nematodes (de Wit *et al.*, 2009; Ahmed *et al.*, 2013; Sarkies *et al.*, 2015; Holz and Streit, 2017). However, the role of small RNAs in *P. pacificus* development has not been investigated.

Here, we report the small RNA transcriptome of *P. pacificus* across multiple developmental timepoints, and identify miR-100 as a highly abundant, developmentally-restricted miRNA. Subsequent analyses identified miR-100 as an ancient miRNA, present in other free-living and parasitic nematodes, but secondarily diversified into a family of 6 miRNAs in *C. elegans*, all of which show limited sequence conservation with miR-100. Furthermore, miR-100 in *P. pacificus* is syntenically conserved in a locus with another miRNA, let-7, partially reminiscent of the conserved locus *let-7-complex* that is, however, secondarily lost in *C. elegans*. Genetic knockout mutants of the two miRNAs showed pleiotropic phenotypes pertaining to developmental defects and reduction in brood size. RNA sequencing showed that the two miRNAs ensure normal development by regulating transcription of collagens and astacins, two large protein classes found in nematodes that regulate various aspects of development, such as fertilisation, hatching, and periodic cuticle building and shedding, among others.

Results and discussion

Stage-specific small RNA sequencing reveals developmental dynamics in *P. pacificus*

To investigate the small RNA transcriptome of *P. pacificus*, worms from four synchronised time points were collected, covering five developmental stages: embryos, J2s, J3s, and late J4s with young adults (YA)(3 biological replicates per time point)(Fig. 1A). Nearly all small RNAs were between 19 and 26 nucleotides in length, and began with either a uracil, guanine, or adenine (Fig. 1B-E). This observation is in line with the previous analyses of *P. pacificus* small RNAs, and with studies showing a 5' bias towards adenine or uracil in miRNAs and piRNAs, and towards guanine in primary and secondary siRNAs in *C. elegans* (Elbashir, Lendeckel and Tuschl, 2001; Sijen et al., 2001; Ruby et al., 2006; Pak and Fire, 2007; Gu et al., 2009; Han et al., 2009; Frank, Sonenberg and Nagar, 2010; Suzuki et al., 2015; Holz and Streit, 2017).

To classify small RNAs as miRNAs, piRNAs, and siRNAs, a second set of sequencing experiments was performed without pyrophosphatase treatment, to exclude 5' triphosphate and capped-small RNAs, in particular, RdRP-generated secondary siRNAs (a peculiarity of nematodes), while retaining 5'-monophosphate bearing miRNAs, piRNAs, and primary siRNAs (Sijen *et al.*, 2001; Gu *et al.*, 2009; Han *et al.*, 2009; Sarkies *et al.*, 2015; Holz and Streit, 2017). Comparing the corresponding stages between the two experiments (Fig. 1B-E and Fig. S1A-D) showed depletion in small RNAs with a 5' guanine, except for reads with length of 26 nucleotides, corresponding to primary siRNAs. This finding is in accordance with previous knowledge of secondary siRNAs having a bias towards 5' guanine and length 22-23 nucleotides (Pak and Fire, 2007; Gu *et al.*, 2009; Holz and Streit, 2017). Further categorisation of pyrophosphatase-treated small RNAs (Fig. 1B-E) as miRNAs or piRNAs was limited to reads between 19 and 26 nucleotides (Fig. 1F-I). Sequences classified as piRNAs were identified by the presence of the conserved Ruby motif upstream of the sequences (Ruby *et al.*, 2006; Beltran *et al.*, 2019). Similar to previous results (Sarkies *et al.*, 2015; Holz and Streit, 2017), piRNAs in *P. pacificus* have a bias for a length of 21 nucleotides. Within our analysis, relative abundance of piRNAs was lower in J2 and J3 stages than in J4/YA and embryos, correlating with the proliferation of the germline in *P. pacificus* and parental inheritance of piRNAs in *C. elegans* embryos (Rudel, Riebesell and Sommer, 2005; Batista *et al.*, 2008). Furthermore, those small RNA sequences that had a perfect sequence match to existing mature sequences on miRBase (release 22.1)(Griffiths-Jones, 2004; Griffiths-Jones *et al.*, 2006; Kozomara, Birgaoanu and Griffiths-

Jones, 2019) were annotated as miRNAs. The observed increase in miRNAs of lengths 23 and 24 nucleotides during J2 and J3 was notably different from adults (Holz and Streit, 2017). Lastly, the remaining reads that did not categorise as either piRNAs or miRNAs were annotated as 'others', which would include siRNAs. Taken together, these results present the first comprehensive dataset of developmental small RNA transcriptome for *P. pacificus*.

A conserved miRNA is the most abundant juvenile-specific miRNA in *P. pacificus*

Two developmental patterns in the small RNA profiles stood out: the large relative abundance of category 22U in the embryo with progressive reduction, and the dominance of category 24A in juveniles and young adults in contrast to embryos. To explore these further, unique reads constituting more than 1% of the total abundance of 22U at embryonic stage were identified, showing that a single sequence made up for nearly half of 22U reads (Fig. S2A), corresponding to ~10% of the entire embryonic small RNA transcriptome. Cross-referencing with miRBase revealed that the sequence corresponded to ppc-miR-2235a-3p, a previously annotated *P. pacificus* miRNA. ppc-miR-2235a-3p shows a seed sequence conservation (Bartel, 2009, 2018) with the miR-35 family of *C. elegans*, a miRNA family with embryonic expression (Martinez *et al.*, 2008; Dexheimer, Wang and Cochella, 2020). The miR-35 family in *C. elegans* comprises of eight miRNAs (miR-35-42), and the genes encoding seven of them (miR-35-41) are arranged in a cluster on chromosome 2 (Alvarez-Saavedra and Horvitz, 2010; Liu *et al.*, 2011). The miR-35 family has been implicated in cell cycle regulation, RNAi sensitivity, and more recently, in maternally-influenced sex determination (Alvarez-Saavedra and Horvitz, 2010; Liu *et al.*, 2011; Massirer *et al.*, 2012; McJunkin and Ambros, 2017; Benner, Prothro and McJunkin, 2019). Given the clustered arrangement of miR-35 family genes in *C. elegans*, we searched for the gene encoding the miRNA ppc-miR-2235a-3p, *Ppa-miR-2235a-3p*, in *P. pacificus* genome (note that the gene names for miRNA-coding genes in *P. pacificus* carry the prefix '*Ppa-*', same as the gene nomenclature of protein-coding genes and in accordance with the 3-letter species-specific abbreviation designated to *P. pacificus* within the nematode community, whereas the corresponding miRNAs carry the prefix 'ppc-', in accordance with the 3-letter species-specific abbreviation designated to *P. pacificus* on miRBase). Surprisingly, *Ppa-miR-2235a-3p* was found to exist in 35 copies on chromosome 2 in an ~8000 base-pair window, with any two copies of the 22-nucleotide miRNA-coding gene separated by a precise 202 base-pairs of consensus genomic sequence (except in between copies #18 and #19) (Fig. S2B). This genomic arrangement is reminiscent of *miR-430* family in zebrafish, wherein a triplet of three miRNA-coding genes - *miR-430a*, *miR-430b*, and *miR-430c* - forms a cluster of nearly 90 copies in a 120-kb genomic locus (Giraldez *et al.*, 2005). *miR-430* too is expressed early in zebrafish development, and is responsible for the deadenylation and subsequent degradation of maternally-provided mRNAs at the onset of zygotic transcription (Giraldez *et al.*, 2006).

Similar to 22U, a single unique read made up the bulk of the 24A category, constituting between 60% and 70% of 24A in the three post-embryonic time points (Fig. 2A-C), and thus amounting between 15% and 20% of the entire small RNA transcriptome. This corresponds to approximately 1.5 times the abundance of ppc-miR-2235a-3p at embryonic stage. Interestingly, we could not find this small RNA read in the list of known *P. pacificus* miRNAs on miRBase. To test for the possibility of sequencing artefacts, a putative biogenesis site for the read was searched for by perfect sequence complementarity within the *P. pacificus* genome. The 24 base-pair region between the genomic coordinates 2633305 and 2633328 on chromosome X of *P. pacificus* (Rödelsperger et al., 2017) was a perfect and unique match against our query sequence. Next, leveraging the availability of sequenced genomes of other species in the family Diplogastridae (Prabh *et al.*, 2018; Röseler *et al.*, 2022; Wighard *et al.*, 2022; Yoshida *et al.*, 2023), the 24-nucleotide read was queried against and gave perfect matches in eight of the 10 genomes. The two exceptions were *Pristionchus fissidentatus*, where the match was limited to

the first 23 nucleotides, and *Pristionchus mayeri*, where no match was found. Furthermore, in *Pristionchus exspectatus*, the gonochoristic sister species to *P. pacificus* (Yoshida et al., 2023), the query sequence matched on the same chromosome as in *P. pacificus*, and in *Allodiplogaster sudhausi*, a diplogastrid nematode with a whole genome duplication (Wighard *et al.*, 2022), two perfect matches against the genome were identified. These results, coupled with the small RNA sequencing results from *P. pacificus* across multiple replicates, and the abundance of the read, strongly suggest that the 24A sequence element is not an artefact, but a bona fide and abundant small RNA with putative conservation of expression in other diplogastrid nematodes.

There was no Ruby motif upstream of this small RNA, suggesting it is not a piRNA (Ruby *et al.*, 2006; Beltran *et al.*, 2019). We did, however, discover complete sequence identity (taking into account, the first 22 nucleotides of the read or more) in other organisms on miRBase, corresponding to miRNA miR-100. Therefore, we named this highly abundant 24A small RNA read of *P. pacificus* as ppc-miR-100. mir-100 has been implicated in differentiation, cell signalling, and multiple types of cancer (Grundmann *et al.*, 2011; Petrelli *et al.*, 2012; Zeng *et al.*, 2012; Zheng *et al.*, 2012; Li *et al.*, 2015; Qin, Huang and Wang, 2015; Ottaviani *et al.*, 2018; Chen *et al.*, 2023). However, its biology in invertebrates remains largely enigmatic. Notably, although miR-100 is a highly conserved regulatory RNA with a strong sequence conservation that may predate bilaterian evolution (Grimson *et al.*, 2008), the *C. elegans* miRNA repertoire does not contain a cel-miR-100. Instead, it transcribes six miRNAs, cel-miR-51, -52, -53, -54, -55, and -56, all of which share the seed sequence with miR-100 from other organisms, but show divergence from an otherwise deeply conserved sequence (Fig. 2D)(Grimson *et al.*, 2008; Alvarez-Saavedra and Horvitz, 2010; Shaw *et al.*, 2010; Kozomara, Birgaoanu and Griffiths-Jones, 2019; Dexheimer, Wang and Cochella, 2020). The transcription of miR-51 family in *C. elegans* is switched on from early embryonic stages. Furthermore, miR-51 family has recently been shown to be one of the two miRNA families (the other being miR-35 family) whose transcription is necessary and sufficient for embryogenesis in *C. elegans* (Dexheimer, Wang and Cochella, 2020), and is responsible for pharyngeal attachment in the embryo (Shaw *et al.*, 2010). However, the sequence similarity between the cel-miR-51 family members and ppc-miR-100 ceases to exist beyond the seed region. Thus, the presence of miR-100 in *Pristionchus* provides a new and potentially powerful avenue to study its biology.

The *let-7-complex* shows two alternative arrangements in nematodes

In *D. melanogaster* and other animals, including vertebrates, *miR-100* occurs in a clustered arrangement with two other miRNA-coding genes, *let-7* and *miR-125* (Sempere *et al.*, 2003)(Fig. 3A). The *D. melanogaster* cluster is polycistronic, giving rise to the precursors of the three miRNAs from the same primary transcript, and is known as the *let-7-complex* (Sokol et al., 2008; Sokol, 2012). *C. elegans* is an exception to this arrangement. The orthologues of both *let-7* and *miR-125* (the latter called *lin-4*) are located on different chromosomes (Fig. 3B)(Sempere *et al.*, 2003). Both *lin-4* and *let-7* are important for the temporal control of developmental decision-making in *C. elegans* (Lee, Feinbaum and Ambros, 1993; Reinhart et al., 2000), albeit at different stages of the worm's ontogeny. This correlates with the genomically separate arrangement of the two miRNA-coding genes in *C. elegans* on different chromosomes (*Cel-let-7* is on chromosome X and *Cel-lin-4* is on chromosome 2), thus untying the polycistronic, simultaneous expression of the *let-7-complex* observed in *D. melanogaster* (Sempere et al., 2003). As for *miR-100* orthologues, while three out of six genes (*Cel-miR-54, -55, -56*) coding for miR-51 family in *C. elegans* are found on chromosome X, same as *Cel-let-7*, and form a cluster of their own (Shaw *et al.*, 2010), the chromosomal distance between the two is approximately 1.5 megabases, and thus not tightly linked.

To determine if the three miRNA-coding gene constituents of *let-7-complex* in *P. pacificus* constitute a locus similar to the one in *D. melanogaster* and other animals (ppc-let-7 and ppc-lin-4 were already annotated in miRBase), we next studied their genomic arrangement. We found that both *Ppa-lin-4* and *Ppa-let-7* reside on the same chromosomes as their *C. elegans* counterparts. Interestingly, *Ppa-miR-100* and *Ppa-let-7* are tightly linked, with *Ppa-let-7* being only 200 base-pairs downstream and in the same orientation as *Ppa-miR-100* (Fig. 3C), reminiscent of the arrangement in the *let-7-complex*. Exploring the genomic arrangements of these three miRNAs in other nematodes, we found that *C. briggsae*, *C. remanei* and *C. brenneri* all follow the same pattern as *C. elegans* in not having a 'classical' *miR-100* within their genome, and having *lin-4* and *let-7* genes dispersed over different chromosomes or genome contigs (Fig. 3B). Conversely, four nematode species with *miR-100* within their genomes - *Ascaris suum*, *Panagrellus redivivus*, *Heligmosomoides polygyrus* and *Brugia malayi* - all have a *P. pacificus*-like arrangement in which *miR-100* and *let-7* are tightly linked in a ~200 base-pair region, whereas *lin-4* is found on a different chromosome or contig (Fig. 3C). These four species and *P. pacificus* are spread over three nematode clades (two are free-living, three are parasitic)(Blaxter *et al.*, 1998), suggesting a progressive disintegration of the *let-7-complex* in nematodes. We speculate that the most downstream of the three miRNAs, miR-125/lin-4, was the first to dissociate from the complex, as it became relocated to a different chromosome in the last common ancestor of all nematodes investigated in this study. This was followed by the loss of the most upstream miRNA, seen here by the absence of miR-100 orthologues from the locus in all investigated *Caenorhabditis* species. In a final step, let-7 itself may get lost, as has been reported recently in multiple species within the *japonica* group of *Caenorhabditis* (Nelson and Ambros, 2021).

Next, we analysed the expression levels of the three miRNAs (Fig. 3D-F). As reported above, miR-100 was found to be highly abundant during J2 and J3 stages, accounting for ~20% of the entire small RNA transcriptome at its peak. Similar to miR-100, let-7 also peaked at J3, indicating the possibility that their origins might be from a polycistronic locus as in *D. melanogaster* despite the differences in their relative abundance levels (Vilimova and Pfeffer, 2023). Finally, lin-4 was the least abundant of the three, peaking at nearly 0.2% of the total during J3. The maxima of abundance attained by the three miRNAs at the same developmental stage led us to explore their putative roles in *P. pacificus*.

Phenotypic characterisation of *P. pacificus* miRNA mutants

We used CRISPR/Cas9-mediated genome editing (Witte *et al.*, 2015; Han *et al.*, 2020) to knock out *Ppa-miR-100*, isolating seven independent alleles from two micro-injection experiments (Table 1A). Homozygous *Ppa-miR-100* mutants were viable. However, in an allele-specific manner, *Ppa-miR-100* mutants were sick, showed egg retention, moulting defects, slower locomotion, and a mild him (high incidence of males) phenotype. To quantify these differences further, we used synchronised egg clutches to measure egg-laying rates, progeny survival, and overall brood sizes, eliminating potential effects of maternal age (Fig. S4; see also 'phenotypic assays' in the methods section for details). We found that *Ppa-miR-100* mutants showed an overall reduction in egg-laying rate, albeit with varying severity (Fig. S3A). Frequently, miRNAs show redundancy, wherein multiple members of the same family cater to the same regulatory pathway, resulting in single mutants having no phenotypes, or showing mild phenotypes, sometimes in combination with stressful environmental conditions (Abbott *et al.*, 2005; Miska *et al.*, 2007; Alvarez-Saavedra and Horvitz, 2010; Shaw *et al.*, 2010; Massirer *et al.*, 2012). Using the phenotypically strongest allele, *Ppa-miR-100(tu1684)*, we scored embryonic lethality, survival to adulthood, and egg-laying rate. *Ppa-miR-100(tu1684)* mutants performed worse than wild-type, *i.e.*, had higher embryonic lethality, lower survival to adulthood, and lower egg-laying rate, than wild-type (Fig. 4C-E). Overall, brood sizes of *Ppa-miR-100(tu1684)* worms were significantly lower than those of wild-type over the course of five days (Fig. 4A). Resolving

these data temporally showed that *Ppa-miR-100(tu1684)* mutants laid fewer eggs at the peak of reproductive activity, and maintained their production for a longer window than wild-type (Fig. 4B).

Next, we generated *Ppa-let-7* mutants. We obtained 11 mutations (Table 1B), which were also not lethal under lab conditions, but showed stronger pleiotropic phenotypes than those observed in *Ppa-miR-100* mutants (sick worms, bagging of eggs in the uterus, moulting defects). Additionally, we observed larval arrest, mid-body rupture, internal hatching, dumpy, and blistered worms. Phenotypic characterisation showed that disparities in egg-laying rates were stronger and more consistent, with statistically significant differences across all 11 alleles (Fig. S3B). We observed that *Ppa-let-7(tu1719)* mutants had lower egg-laying and survival-to-adulthood rates than *Ppa-miR-100(tu1684)* mutants, but showed marginally lower embryonic lethality, although still higher than wild-type (Fig. 4C-E). *Ppa-let-7(tu1719)* mutants also had smaller overall brood sizes than *Ppa-miR-100(tu1684)* mutants (Fig. 4A-B).

To test if a double mutant of both miRNAs would have a stronger phenotype, *Ppa-miR-100* was knocked out in *Ppa-let-7(tu1724)* background, yielding four *Ppa-miR-100 Ppa-let-7(tu1724)* double mutants (Table 1C). Double mutants had significantly lower egg-laying rates than wild-type worms (Fig. S3C), but did not show a strong enhancement of the pleiotropic phenotype of the two single mutants. Using *Ppa-miR-100(tu1898) Ppa-let-7(tu1724)* for subsequent phenotypic analyses showed that the double mutant had a stronger phenotype than *Ppa-miR-100(tu1684),* but weaker than *Ppa-let-7(tu1719)* mutants (Fig. 4A-E).

Lastly, we knocked out *Ppa-lin-4* and obtained two independent alleles (Table 1D). Surprisingly, *Ppa-lin-4(tu1942)* mutants had near-normal growth and showed no differences when compared to wild-type worms (Fig. 4C-E). Thus, in *P. pacificus*, miR-125/lin-4 is not required in the same regulatory machinery as let-7 and miR-100. The original *lin-4* mutant strain (e912) in *C. elegans* was investigated by the Ruvkun and Ambros labs in the 1990s because of its striking developmental timing defects. *lin-4* mutants exhibit heterochronic defects wherein juvenile cells and moulting phenotypes are extended into adults (Arasu, Wightman and Ruvkun, 1991; Lee, Feinbaum and Ambros, 1993). Yet, in *Pristionchus,* and potentially several other nematodes with similar loci, it appears to have no essential role in development.

Small RNA sequencing confirms absence of miRNAs and incomplete phenotypic rescue by miR-100 family members

To study the small RNA profiles in miRNA mutants, J3 stage-specific small RNA sequencing of wild-type, *Ppa-miR-100(tu1684)*, *Ppa-let-7(tu1719)*, and *Ppa-miR-100(tu1898) Ppa-let-7(tu1724)* was performed (3 biological replicates for *Ppa-miR-100(tu1684)*, 2 for the rest). The J3 stage was selected since both miR-100 and let-7 were the most abundant at this stage (Fig. 3D-E). Small RNA profiles showed that 24A and 22U, housing miR-100 and let-7, respectively, had lower relative abundances in *Ppa-miR-100(tu1684)* and *Ppa-let-7(tu1719)* mutant worms, respectively, in comparison to wild-type, and both were less abundant in the *Ppa-miR-100(tu1898) Ppa-let-7(tu1724)* double mutant (Fig. 5A-D). Indeed, we found that *ppc-miR-100* and *ppc-let-7* reads were completely absent in the corresponding mutants (Fig. 5M-N).

As referred to earlier, frequently, individual miRNA mutants show weak to no phenotypes, due to functional redundancy provided by other miRNAs that belong to the same family (Abbott *et al.*, 2005; Miska *et al.*, 2007; Alvarez-Saavedra and Horvitz, 2010; Shaw *et al.*, 2010; Massirer *et al.*, 2012). To test the possibility that the phenotypic inconsistency in *Ppa-miR-100* mutants could be due to redundancy with other miR-100 family members potentially transcribed in the *P. pacificus* genome, we queried for the expression levels of entire miR-100 family in wild-type

and mutant worms. A majority of the miR-100 family in wild-type worms was constituted by miR-100 itself, which was abrogated in the *Ppa-miR-100(tu1684)* and *Ppa-miR-100(tu1898) Ppa-let-7(tu1724)* double mutants, but not in *Ppa-let-7(tu1719)* mutants (Fig. 5E-H). However, a small number of 22 nucleotide-length reads belonging to the miR-100 family were also found in wild-type worms (Fig. 5E), which were still present in *Ppa-miR-100(tu1684)* and the *Ppa-miR-100(tu1898) Ppa-let-7(tu1724)* double mutants (Fig. 5F, H), indicating that these are miR-100 family members which could potentially partially rescue the loss of miR-100, thus resulting in an incomplete and inconsistent mutant phenotype (Fig. S3A)(Abbott *et al.*, 2005; Miska *et al.*, 2007; Shaw *et al.*, 2010). Such a pattern was not seen for let-7; the let-7 family in *P. pacificus* is primarily composed of let-7 itself, which, when knocked out, has nearly no family members to rescue its loss (Fig. 5I-L), resulting in a stronger and more consistent phenotype (Fig. S3B).

Lastly, to rule out the possibility that the miRNA miR-100 itself might be partially present in the strains with weaker phenotypes, four more mutant alleles across the three mutants - *Ppa-miR-100(tu1638)*, *Ppa-miR-100(tu1685)*, *Ppa-let-7(tu1724)*, and *Ppa-miR-100(tu1900) Ppa-let-7(tu1724)* double mutant - were sequenced (3 biological replicates for *Ppa-miR-100(tu1638)* and *Ppa-miR-100(tu1685)*, 2 biological replicates for *Ppa-let-7(tu1724)* and *Ppa-miR-100(tu1900) Ppa-let-7(tu1724)* double mutant), and then searched for presence of miR-100 *and* let-7. Regardless of their phenotypes, or lack thereof, no miR-100 or let-7 was found in corresponding mutants (Fig. S5). Thus, the presence of other miRNA family members, rather than the partial presence of miR-100 in the alleles, is presumably responsible for incomplete phenotypic penetration.

Extracellular matrix-regulating developmental genes are downstream targets of miRNAs *miR-100* and *let-7*

To identify putative regulatory targets of the *P. pacificus* miRNAs, we performed RNA sequencing of mutant strains (3 biological replicates for *Ppa-miR-100(tu1684)*, 2 for the rest). *Ppa-miR-100(tu1684)* mutant worms had the most differentially expressed genes (DEGs), with a total of 548, a majority of which (449 genes) were upregulated in comparison to wild-type, while the remaining 99 genes were downregulated (Supplementary table S1). *Ppa-let-7(tu1719)* mutants had fewer DEGs, with a total of 224, of which approximately one half was upregulated (116 genes) and the other (108 genes) downregulated (Supplementary table S2). Lastly, the *Ppa-miR-100(tu1898) Ppa-let-7(tu1724)* double mutant had a total of 214 DEGs, with 31 downregulated and 183 upregulated genes (Supplementary table S3).

The skewed distribution of upregulated versus downregulated DEGs in the *Ppa-miR-100(tu1684)* and double mutants was intriguing (Fig. 6A-C). Therefore, we looked at the overlap of DEGs between the three mutants in a pairwise manner, and then across all three together. A total of 67 genes were commonly differentially expressed between the two single mutants, whereas *Ppa-miR-100(tu1684)* mutant and the double mutant shared 85 DEGs, and *Ppa-let-7(tu1719)* mutant and the double mutant shared 71 DEGs. Also, there were 35 DEGs shared among all three mutants (Fig. 6D). Looking at the direction of differential expression in each pairwise comparison, the number of common genes oppositely regulated between the double mutant and the single mutants was 2/85 and 4/71 for *Ppa-miR-100(tu1684)* and *Ppa-let-7(tu1719),* respectively. The common genes between the two single mutants stood apart from the rest: 14/67 common genes between *Ppa-miR-100(tu1684)* and *Ppa-let-7(tu1719)* mutants were oppositely regulated, none of which were present in the common genes among all three mutants (Supplementary tables S4-S7). In *Ppa-let-7(tu1719)* mutants, 11 of these 14 were downregulated, with 10 out of those 11 strongly downregulated ($\log_2$(fold change) < -2). Thus, let-7, in parts, oppositely regulates miR-100 targets, an effect that is abrogated in the double mutant, which we hypothesise explains the more modest effect in phenotype.

To explore the DEGs further, we performed an overrepresentation analysis of their protein domains (Fig. 6E-J). For each mutant, upregulated genes had stronger signals for overrepresentation (*i.e.*, smaller *P*-values) and higher relative abundances of domains (*i.e.*, higher Rich Factors) than their downregulated counterparts. *Ppa-miR-100(tu1684)* mutant worms showed an overrepresentation for collagen and nematode cuticle domains in both upregulated and downregulated genes, indicating a general perturbation of these structural proteins. For *Ppa-let-7(tu1719)* mutants, the overrepresentation of collagen and nematode cuticle-containing proteins in the upregulated genes was accompanied by a strong overrepresentation of astacins in the downregulated genes (Fig. 6G). Astacins are a class of developmental genes highly conserved across animals, and encode for metalloproteases, a conserved protein family known for their zinc-dependent extracellular proteolytic activity during development in multiple organisms (Gomis-Rüth, Trillo-Muyo and Stöcker, 2012). Humans have six genes encoding for astacins, which includes BMP1 (bone morphogenetic protein 1), whereas invertebrates such as *C. elegans* have nearly 40 (Wozney *et al.*, 1988; Möhrlen, Hutter and Zwilling, 2003; Gomis-Rüth, Trillo-Muyo and Stöcker, 2012). In mice and sea urchins, astacins are responsible for the cleavage of a protein-rich layer of vitelline envelope (zona pellucida in higher animals) to facilitate hatching and to prevent polyspermy post fertilisation (Nomura *et al.*, 1991; Burkart *et al.*, 2012). In the context of nematodes, astacins are known to participate in multiple developmental pathways. The first astacin identified and characterised in *C. elegans* was *hch-1*. HCH-1 is an embryonic protein, whose absence leads to a delayed albeit viable hatching in *C. elegans* (Hishida et al., 1996). However, the most common pathway that nematode astacins are known to play a role in is moulting of its cuticle. *nas-37*, *nas-36* and *dpy-31*/*nas-35* were found to be expressed in the hypodermal cells, and mutants for these genes were found to have moulting defects (Davis *et al.*, 2004; Novelli, Ahmed and Hodgkin, 2004; Suzuki *et al.*, 2004) due to incomplete degradation of the older cuticle, an extracellular matrix layer composed of collagen (Cox, Kusch and Edgar, 1981; Cox, Staprans and Edgar, 1981). Astacins are capable of regulating collagen levels by processing of procollagen (Kronenberg *et al.*, 2010), and subsequent research has shown that moulting-related functions of astacins are evolutionarily conserved between free-living and parasitic nematodes (Stepek, McCormack and Page, 2010; Stepek *et al.*, 2011, 2015; Huang *et al.*, 2021).

Overrepresentation analysis for the double mutant broadly recapitulated the results of *Ppa-let-7* worms; astacin was the most statistically significant domain in downregulated genes, whereas collagen and cuticle were the most statistically significant within the upregulated genes (Fig. 6I, J). There was also an overall increase in the total number of overrepresented domains in the double mutant, and a strong increase in ribosomal protein domains in the upregulated genes (Fig. 6J).

In *P. pacificus*, the interest in astacins goes beyond their presumed functional conservation in moulting, and into their potential roles in development of plastic adult mouth morphs (Sieriebriennikov *et al.*, 2020; Ishita *et al.*, 2023). Recently, Sieriebriennikov and colleagues identified that the common regulatory targets of two nuclear hormone receptors which control feeding plasticity in *P. pacificus*, *Ppa-nhr-40* and *Ppa-nhr-1*, are enriched in astacins (Sieriebriennikov *et al.*, 2020). We found 15 out of 24 common transcriptional targets of *Ppa-nhr-1* and *Ppa-nhr-40* to be overlapping with the DEGs in *Ppa-let-7* mutants, eight of which were also overlapping with the DEGs in the double mutants (Fig. S7). Out of the 12 astacins that were found among the 24 NHR targets, eight were differentially expressed in *Ppa-let-7* worms, five of which were also differentially expressed in the double mutant. Interestingly, all of these genes were downregulated in our miRNA mutants (Fig. S7). The unexpected regulatory convergence on astacins and collagens may suggest that small RNAs have a potential role in

regulation of feeding plasticity too. Small RNAs are known for their ability to respond to environmental perturbations and even convey the information across generations (Ashe *et al.*, 2012; Rechavi *et al.*, 2014). Although we did not see any mouth-form phenotypes in either of the three mutants (Eu morph > 95% in wild-type and miRNA mutants; n = 90-100 over three biological replicates) in line with the duodecuple astacin mutant (Sieriebriennikov *et al.*, 2020), future analyses will reveal the extent to which small RNAs contribute to such phenotypic plasticity.

Finally, to gain biological insights about the DEGs, we performed enrichment analyses of our DEGs in two recently published genome-wide datasets, categorising *P. pacificus* genes on the basis of co-expression modules (Athanasouli *et al.*, 2023) and identifying oscillating genes in the genome (Sun, Rödelsperger and Sommer, 2021). Gene module enrichment analysis showed an overrepresentation of genes within module #11, which includes genes involved in cuticle development and maintenance, across all three mutants (Fig. S6A-C). Other modules with an overrepresentation in at least one mutant were associated with developmental pathways such as oogenesis (module #7)(Rudel, Riebesell and Sommer, 2005), or expression in tissues undergoing rapid developmental changes, such as the intestine (module #3), nervous system (module #8) and regions associated with mouth dimorphism (module #24) (Fig. S6A-C). Simultaneously, enrichment analysis within the oscillating genes showed that all three mutants had a higher representation of oscillating genes than the genome-wide baseline of ~10% (Fig. S6D)(Sun, Rödelsperger and Sommer, 2021). Most of the oscillations coincide with the moulting cycles and coordinate systematic progression through developmental stages (Supplementary tables S8-S10). Taken together, miRNAs miR-100 and let-7 are important for overall developmental progression in *P. pacificus*, and their perturbation leads to pleiotropic defects - characteristic of developmental genes - as shown above (Fig. 4).

Conclusion

In this study, we provide the first description of the developmental repertoire of small RNAs in *P. pacificus*, which can serve as a molecular resource for the nematode community to study the roles of small RNAs in an evo-devo context. Furthermore, we identify that an ancient, highly conserved miRNA, miR-100, is present as the most abundant miRNA in juvenile stages of *P. pacificus*. Despite its transcriptional abundance, miR-100 is non-essential for viability of the worm, likely due to functional redundancy and partial rescue by other miR-100 miRNA family members (Fig. 5E-H)(Miska *et al.*, 2007; Alvarez-Saavedra and Horvitz, 2010). We hypothesise that a complete knockout of the miR-100 family in *P. pacificus* would result in stronger phenotypes, possibly even lethality.

The classical *let-7-complex* is non-existent in nematodes, instead presenting its constituent miRNA-coding genes in two alternative genomic arrangements (Fig. 3B-C). This dissociation indicates a (potentially ongoing) disintegration of the *let-7-complex* in the lineages leading to *Pristionchus* and *Caenorhabditis*, where the individual components of the locus are progressively rearranged within the respective genomes. While *P. pacificus* deviates from the norm by having relocated its copy of *lin-4* (*Ppa-lin-4*) on a different chromosome, *C. elegans* shows a dissociation of *miR-100* from the locus along with an expansion into genes transcribing a family of six miRNAs: the miR-51 family (Shaw *et al.*, 2010), in addition to the relocation of *Cel-lin-4* as in *P. pacificus.*

Lastly, we show that the two miRNAs - ppc-miR-100 and ppc-let-7 - as the two constituents of the remnants of *let-7-complex* in *P. pacificus*, ensure normal development by regulating transcription of collagens and astacins (along with other protein families, such as heat-shock

proteins (HSPs), ribosomal protein-subunits, and the von Willebrand factor (VWF) domains, among others)(Fig. 6E-J), two developmental gene classes vastly expanded in nematodes, and important for developmental progression to adulthood.

Methods

Stage-specific worm pellet collection

10-cm NGM-agar plates spotted with 800 µl *E. coli* OP50 were seeded with *P. pacificus* wild-type strain (PS312) and incubated at 20°C (Stiernagle, 2006). Fully-populated plates with eggs and gravid adults were washed down 6 days post seeding using M9, then filtered through a 5 µm nylon filter to separate bacteria (in the flow-through) and worms (on the filter). Filtered worms were collected and bleach-treated to rid the mixture of all post-hatched worms and any remaining bacteria (Stiernagle, 2006), followed by a filtration through a 120 µm nylon filter to separate eggs (in the flow-through) from any undissolved adult worm carcasses (on the filter). The flow-through was gently centrifuged (500g for 2 minutes) to collect the egg pellet and remove the supernatant, followed by washing and resuspension in M9. Approximately 30% of the egg resuspension volume was transferred to microcentrifuge tubes to collect the embryo developmental stage (Fig. 1A) by centrifugation followed by flash-freezing in liquid nitrogen and storage at -80°C. Remaining egg resuspension was spotted on to empty 10-cm NGM-agar plates with OP50 for collection of later developmental stages. Three biological replicates were collected per developmental time point.

For J2 pellet collection, plates spotted with bleached eggs were washed off 20-22 hours post seeding, using M9. Washed worms were first filtered through a 20 µm nylon filter to separate J2 and bacteria (in the flow-through) from unhatched and dead embryos (on the filter), followed by filtering the flow-through using a 5 µm nylon filter to separate bacteria (in the flow-through) and worms (on the filter). For J3 pellet collection, bleached eggs were washed off 42-44 hours post seeding. Washed worms were filtered through a 20 µm nylon filter to separate J2 and bacteria (in the flow-through) from J3 worms (on the filter). For J4/YA pellet collection, bleached eggs were washed off 66-70 hours post seeding. Washed worms were filtered through a 5 µm nylon filter to separate J4/YA worms (on the filter) from bacteria (in the flow-through). In all cases, worms were washed off the filter patch and resuspended using M9, transferred and centrifuged in microcentrifuge tubes to collect into a pellet and flash-frozen in liquid nitrogen before being stored at -80°C.

For analysis of wild-type *vs.* mutants, J3 pellets were collected post bleaching as described above. Three (for *Ppa-miR-100*) or two (for the rest) biological replicates were collected per strain.

Total RNA extraction, small RNA size selection, NGS library preparation and sequencing

Worm pellets as collected above were thawed on ice and added with 500 µl of TRIzol (Invitrogen), followed by a quick vortex treatment for resuspension of pellets. Next, three cycles of flash-freezing the worm-TRIzol mix, followed by thawing at 37°C, were performed. Following three freeze-thaw cycles, 100 µl of chloroform was added to the worm-TRIzol mix, and the contents of tubes mixed by shaking and inverting. Microcentrifuge tubes containing the worm-TRIzol-chloroform mix were incubated at room temperature for 2-3 minutes, followed by a centrifugation at 12000g and 4°C for 15 minutes. Post centrifugation, the aqueous phase on top was transferred to a new sterile microcentrifuge tube for each sample, equal volume (approximately 300 µl) of 100% ethanol was added, and the contents of tubes mixed by vortexing. Total RNA from the aqueous phase-ethanol mix was subsequently extracted using RNA Clean & Concentrator-25 kit (Zymo Research) following the manufacturer's protocol.

Extracted RNA was then subjected to CapClip acid pyrophosphatase (CCAP)(Biozym) treatment for removal of 5' triphosphate cap (exception: 5' monophosphate-biased library, where this step was skipped; Fig. S1). For CCAP treatment, total RNA was sequentially added with sterile-grade water, 5 µl of 10x reaction buffer (Biozym), and lastly with 5 µl of CCAP enzyme (Biozym) in a microcentrifuge tube over ice (total reaction volume = 50 µl). Contents of the tubes were then mixed by gentle pipetting, followed by incubation at 37°C for 2 hours, followed by a cleanup step. RNA concentration was determined using the Qubit fluorometer (Invitrogen) before and after CCAP treatment for assessment of RNA losses.

Following RNA clean-up, CCAP-treated RNA was used for preparation of small RNA next-generation sequencing libraries with NEBNext Multiplex Small RNA Library Prep Kit (E7560S, New England BioLabs Inc.) using the manufacturer's protocol, including a negative (no RNA) control. RNA libraries thus prepared were subjected to size-selection using Blue Pippin and compliant 3% agarose gel cassettes (Sage Science Inc.) following the manufacturer's protocol. RNA samples before library preparation, and NGS libraries after size selection and a DNA cleanup step (using Monarch PCR & DNA cleanup kit; T1030S, New England BioLabs Inc.) were analysed with Bioanalyzer (Agilent Technologies, Inc.) using the manufacturer's protocol to check for integrity of RNA, library size, and quality. Once satisfactory profiles were observed on Bioanalyzer, libraries were pooled and run on a single lane on an Illumina HiSeq 3000 machine.

For wild-type *vs.* mutants, total RNA was extracted as described above. One part of total RNA was kept for RNA sequencing, while the other part was treated with CCAP enzyme for small RNA sequencing. Both types of samples were then submitted to Novogene (UK) Company Ltd. for RNA and small RNA sequencing respectively.

Bioinformatic analysis

Small RNA sequencing raw reads were first subjected to adapter trimming using Cutadapt (version 4.1)(Martin, 2011), then filtered for reads mapping to *P. pacificus* tRNA and rRNA genes identified through tRNAscan-SE (version 2.0.5) and RNAmmer (version 1.2) respectively (Lowe and Eddy, 1997; Lagesen *et al.*, 2007; Chan and Lowe, 2019). Remaining reads were then mapped to the *P. pacificus* genome using Bowtie2 (version 2.3.5.1)(Langmead *et al.*, 2009; Langmead and Salzberg, 2012; Rödelsperger *et al.*, 2017), procuring the FASTA sequences from BAM alignments using SAMtools (version 1.10)(Li *et al.*, 2009; Danecek *et al.*, 2021). FastQC (Babraham Bioinformatics) was used for ensuring quality of sequences and presence or absence of adapters.

Custom shell scripts were written for sorting reads on the basis of length and first 5' nucleotide, the output of which was fed to R (version 4.2.3) to generate small RNA profiles (Fig. 1B-E, Fig. S1, Fig. 5A-D). For identifying the putative Ruby motif in *P. pacificus*, using existing knowledge about *P. pacificus* piRNAs (Sarkies *et al.*, 2015), we queried for top 5 motifs 70-100 base-pairs upstream of 21U reads using MEME (Bailey, 1994; Bailey *et al.*, 2006) from the MEME Suite (version 5.2.0). For all predicted motifs for each of the 4 developmental timepoints, the best motifs predicted had a much better E-value (E-value < 10^{-4}) than the second best (E-value ~ 0.1) and closely resembled the Ruby motif. Using the best predicted motifs for each developmental stage as query, their occurrence was searched for 100 base-pairs upstream of all small RNA reads between lengths 19 and 26 nucleotides using FIMO (Grant, Bailey and Noble, 2011) from the MEME Suite, followed by looking at the length and first 5' nucleotide of positive sequences. For miRNAs, we looked for sequences with perfect matches to the mature miRNAs FASTA sequence from miRBase (Kozomara, Birgaoanu and Griffiths-Jones, 2019). Remaining sequences were categorised as 'others' (Fig. 1F-I).

For identification of the most abundant reads, for each category (combination of read length and first 5' nucleotide) of small RNAs, the number of reads pertaining to each unique genome-mapped small RNA sequence were counted between lengths 19 and 26 using custom Shell scripts. The resulting tables were fed into R to identify reads accounting for 1% or more of the corresponding category (Fig. 2A-C, Fig. S2A). For the expression levels of the three miRNAs in *P. pacificus*, the number of reads were searched for the perfect matches to the three sequences (ppc-miR-100: AACCCGTAGATCCGAACTTGTGTT, ppc-let-7: TGAGGTAGTAGGTTGTATAGTT, ppc-lin-4: TCCCTGAGACCTCAATTGTG) across the corresponding developmental stages or mutant strains (Fig. 3D-F, Fig. 5M-N, Fig. S5). For miRNA family expression levels, custom Shell scripts were written to identify the number of sequences matching the seed sequence of corresponding miRNAs (nucleotides 2-8 (Bartel, 2018)) across the mutant strains (Fig. 5E-L).

Paired-end RNA sequencing reads were aligned against the *P. pacificus* genome using STAR (version 2.7.9a)(Dobin *et al.*, 2013; Rödelsperger *et al.*, 2017), and mapped sequences were counted against the El Paco V3 annotation using featureCounts from the Subread package (version 2.0.0)(Liao, Smyth and Shi, 2014; Athanasouli *et al.*, 2020). Differential expression analysis was done between wild-type and mutant strains in a pairwise manner in R using DESeq2 (version 1.38.3)(Love, Huber and Anders, 2014), using genes with a mean count ≥ 10 in the wild-type strain. We always included a $\log_2$-fold change cut-off of 0.5849625 (absolute value; corresponding to a fold change of 1.5, or 50% change in expression levels) into our null hypothesis. Differentially expressed genes (DEGs) with an adjusted p-value < 0.1 were selected for subsequent overrepresentation analyses (Supplementary tables S1-S3). The latter were based on two-tailed Fisher's Exact tests with FDR-correction of p-values.

Data wrangling for graphical representation was done in R using ggplot2 (version 3.4.3)(Wickham, 2009). The Venn diagram (Fig. 6D) was prepared using the eulerr package (version 7.0.0) (Larsson, 2022). Post plotting, aesthetic modifications were performed in Adobe Illustrator CS6 (version 16.0.4) and Affinity Designer 2 (version 2.5.0).

CRISPR/Cas9-mediated mutagenesis

CRISPR-mediated mutagenesis was performed based on previously published protocols with some modifications (Witte *et al.*, 2015; Han *et al.*, 2020; Theska and Sommer, 2024; Theska, Renahan and Sommer, 2024). In short, crRNAs were designed with high sequence specificity 20 base-pairs upstream of the protospacer adjacent motif (PAM) sites (Integrated DNA technologies, Inc.). Cas9 protein (Catalog no. 1081058), tracrRNA (Catalog no. 1072534) and all three CRISPR RNAs were purchased from Integrated DNA Technologies. 5 µl of tracrRNA (0.4 µg/µl) and 1.12 µl of crRNA (0.4 µg/µl) were mixed and incubated at 37°C for 10 minutes followed by a 5-minute incubation at room temperature. To the above mix, 0.5 µl of Cas9 protein (10 µg/µl), 1 µl (100 mM) of 100-nucleotide long DNA repair template carrying a deletion for the miRNA sequence and 3 µl (170 ng/µl) of plasmid PZH009 (Han *et al.*, 2020) carrying the promoter of *Ppa-eft-3p* coupled with TurboRFP (Sun *et al.*, 2022) as co-injection marker were added, and the total reaction volume made up to 20 µl using Tris-EDTA buffer. Injected worms (P_0) were put onto NGM-plates with 50 µl OP50 lawns and allowed to lay eggs at 20°C for 24-32 hours before being sacrificed. Three to four days later, F1 J4s/young adults were screened for RFP co-injection marker fluorescence, and those containing RFP+ F1s were segregated onto individual plates, and genotyped after ~24 hours. F2 progeny of heterozygous F1s were genotyped (typically 10-16 F2s per F1) for potential homozygous mutants. Upon identification of such homozygous F2s, their progeny (F3) were subjected to the above procedure for a confirming Sanger sequencing (typically 4-6 F3s per homozygous F2), following which mutant lines were established. Sanger sequencing was done by Azenta Life

Sciences. Allele names and molecular lesions of all mutants generated in this study are mentioned in Table 1.

Phenotypic assays

A schematic layout is provided in Fig. S4. Briefly, day 1 adult hermaphrodites, with no more than 1-2 eggs in the uterus, from non-starved stock maintenance plates for 5+ generations were seeded to empty NGM-agar plates with 300 µl OP50 lawns, and allowed to lay eggs for 2 hours, after which they were removed, leaving the plates with 2-hour synchronised P_0 worms (maternal synchronisation; Fig. S4). For the egg-laying and survival assay, these synchronised P_0 young adult worms, after 90-96 hours, were passed onto NGM-agar plates with fresh (<24 hours old) 50 µl OP50 lawns (5 worms per plate) and allowed to lay eggs for 2 hours at room temperature before being removed. The number of synchronised F1 eggs were scored for egg-laying rate (Fig. 4C, Fig. S3A-C), before being incubated at 20°C. Approximately 28-30 hours later, the incubated F1 embryos were scored for embryos still unhatched (Fig. 4D), before being incubated again at 20°C. 90-96 hours post F1 synchronisation, the number of surviving F1 adults were scored (Fig. 4E). A total of five biological replicates (20 P_0 worms per replicate) were performed per strain.

For brood size assay, starting from 2-hour synchronised P_0 embryos, these embryos were incubated at 20°C for ~70 hours, by which time they would have developed into the J4 stage. At this stage, P_0 worms were singly segregated onto individual "day 1" plates and incubated at 20°C. After 24 hours, these worms were seeded to new "day 2" plates, and both sets of plates were incubated at 20°C. The experiment continued till accumulation of "day 5" plates. On day 6, P_0 worms on "day 5" plates were sacrificed and the plates incubated at 20°C, and the "day 1" plates were scored for the number of F1 adult worms (Fig. 4B). 24 hours later (day 7), "day 2" plates were scored, and so on until day 10, when "day 5" plates were scored for F1 adults. Total brood size was calculated by summing up the day-wise scores (Fig. 4A). Two replicates (25 P_0 worms per replicate) were performed per strain.

To test for mutant phenotypes in our assay data, we applied a permutation-based version of MANOVA in which null model residuals were randomised (in 10,000 iterations) using the RRPP package (version 1.4.0). *P*-values for the pairwise comparisons of individual strains are reported in the corresponding figure panels (*: p-adj < 0.05, **: p-adj < 0.01, ***: p-adj < 0.001). We considered our data as incompatible with the null hypothesis (i.e., "statistically significant") if the derived (FDR-corrected) *P*-values were < 0.05.

Data availability

The raw sequencing data generated in this study can be accessed at the European Nucleotide Archive (ENA) repository under the accession number PRJEB76785.

Funding

This research was funded by the Max Planck Society.

Acknowledgements

The authors are grateful to Heike Haussmann for freezing all the mutant alleles generated in this study, Heike Budde of the Genome Centre facility for her assistance in performing small RNA sequencing of the developmental stages, and Gabi Eberhardt for her assistance in shipping RNA samples to Novogene Ltd. for sequencing of mutants. Furthermore, the authors thank

Tobias Loschko, who, along with WR, mentored DRS in the initial stages of the project in RNA biology and quality control techniques, Metta Riebesell, who, along with HW, mentored DRS in nematode microinjections, Dr. Christian Rödelsperger for his expertise in bioinformatics and engaging discussions, Dr. Tobias Theska, Dr. Tess Renahan, and Dr. Adrian Streit for critically reading the manuscript and providing their insights, and members of Sommerlab, in particular Dr. Catia Igreja, for their thoughtful questions and valuable suggestions.

Author contributions

DRS, MSW and RJS conceptualised the study. DRS and RJS designed the experiments. DRS performed staging and sample collection experiments, RNA extraction and library preparation for developmental analysis, generated CRISPR mutants (except *lin-4* mutants), performed mutant phenotypic assays, analysed the sequencing data, and created the visualisations. WR performed RNA extraction for mutant analysis, and mentored DRS in library preparation and quality control. HW generated the *lin-4* CRISPR mutants, and mentored DRS in nematode microinjections and creating CRISPR mutants. DRS wrote the manuscript, with inputs from MSW and RJS. RJS provided the resources and supervised the study.

References

1. Abbott, A.L. *et al.* (2005) 'The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans', *Developmental cell*, 9(3), pp. 403–414.

2. Ahmed, R. *et al.* (2013) 'Conserved miRNAs are candidate post-transcriptional regulators of developmental arrest in free-living and parasitic nematodes', *Genome biology and evolution*, 5(7), pp. 1246–1260.

3. Alvarez-Saavedra, E. and Horvitz, H.R. (2010) 'Many families of C. elegans microRNAs are not essential for development or viability', *Current biology: CB*, 20(4), pp. 367–373.

4. Ambros, V. (1989) 'A hierarchy of regulatory genes controls a larva-to-adult developmental switch in C. elegans', *Cell*, 57(1), pp. 49–57.

5. Ambros, V. (2011) 'MicroRNAs and developmental timing', *Current opinion in genetics & development*, 21(4), pp. 511–517.

6. Ambros, V. and Ruvkun, G. (2018) 'Recent Molecular Genetic Explorations of MicroRNAs', *Genetics*, 209(3), pp. 651–673.

7. Arasu, P., Wightman, B. and Ruvkun, G. (1991) 'Temporal regulation of lin-14 by the antagonistic action of two other heterochronic genes, lin-4 and lin-28', *Genes & development*, 5(10), pp. 1825–1833.

8. Ashe, A. *et al.* (2012) 'piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans', *Cell*, 150(1), pp. 88–99.

9. Athanasouli, M. *et al.* (2020) 'Comparative genomics and community curation further improve gene annotations in the nematode Pristionchus pacificus', *BMC genomics*, 21(1), p. 708.

10. Athanasouli, M. *et al.* (2023) 'Thousands of Pristionchus pacificus orphan genes were integrated into developmental networks that respond to diverse environmental microbiota', *PLoS genetics*, 19(7), p. e1010832.

11. Bagijn, M.P. *et al.* (2012) 'Function, targets, and evolution of Caenorhabditis elegans piRNAs', *Science*, 337(6094), pp. 574–578.

12. Bailey, T.L. (1994) *Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Bipolymers*.

13. Bailey, T.L. *et al.* (2006) 'MEME: discovering and analyzing DNA and protein sequence motifs', *Nucleic acids research*, 34(Web Server issue), pp. W369–73.

14. Bartel, D.P. (2009) 'MicroRNAs: target recognition and regulatory functions', *Cell*, 136(2), pp. 215–233.

15. Bartel, D.P. (2018) 'Metazoan MicroRNAs', *Cell*, 173(1), pp. 20–51.

16. Batista, P.J. *et al.* (2008) 'PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans', *Molecular cell*, 31(1), pp. 67–78.

17. Beltran, T. *et al.* (2019) 'Comparative Epigenomics Reveals that RNA Polymerase II Pausing and Chromatin Domain Organization Control Nematode piRNA Biogenesis', *Developmental cell*, 48(6), pp. 793–810.e6.

18. Beltran, T. *et al.* (2020) 'Epimutations driven by small RNAs arise frequently but most have limited duration in Caenorhabditis elegans', *Nature ecology & evolution*, 4(11), pp. 1539–1548.

19. Beltran, T. *et al.* (2021) 'Integrator is recruited to promoter-proximally paused RNA Pol II to generate Caenorhabditis elegans piRNA precursors', *The EMBO journal*, 40(5), p. e105564.

20. Benner, L.K., Prothro, K.P. and McJunkin, K. (2019) 'The Family Links Maternal Germline Sex to Embryonic Viability in', *G3* , 9(3), pp. 901–909.

21. Bento, G., Ogawa, A. and Sommer, R.J. (2010) 'Co-option of the hormone-signalling module dafachronic acid-DAF-12 in nematode evolution', *Nature*, 466(7305), pp. 494–497.

22. Berkhout, B. (2018) 'RNAi-mediated antiviral immunity in mammals', *Current opinion in virology*, 32, pp. 9–14.

23. Blaxter, M.L. *et al.* (1998) 'A molecular evolutionary framework for the phylum Nematoda', *Nature*, 392(6671), pp. 71–75.

24. Brennecke, J. *et al.* (2007) 'Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila', *Cell*, 128(6), pp. 1089–1103.

25. Buck, A.H. *et al.* (2014) 'Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity', *Nature communications*, 5, p. 5488.

26. Burkart, A.D. *et al.* (2012) 'Ovastacin, a cortical granule protease, cleaves ZP2 in the zona pellucida to prevent polyspermy', *The Journal of cell biology*, 197(1), pp. 37–44.

27. Chan, P.P. and Lowe, T.M. (2019) 'tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences', *Methods in molecular biology* , 1962, pp. 1–14.

28. Charlesworth, A.G. *et al.* (2021) 'Two isoforms of the essential C. elegans Argonaute CSR-1 differentially regulate sperm and oocyte fertility', *Nucleic acids research*, 49(15), pp. 8836–8865.

29. Chen, C. *et al.* (2023) 'Downregulation of miR-100-5p in cancer-associated fibroblast-derived exosomes facilitates lymphangiogenesis in esophageal squamous cell carcinoma', *Cancer medicine*, 12(13), pp. 14468–14483.

30. Chen, X. and Rechavi, O. (2022) 'Plant and animal small RNA communications between cells and organisms', *Nature reviews. Molecular cell biology*, 23(3), pp. 185–203.

31. Cox, G.N., Kusch, M. and Edgar, R.S. (1981) 'Cuticle of Caenorhabditis elegans: its isolation and partial characterization', *The Journal of cell biology*, 90(1), pp. 7–17.

32. Cox, G.N., Staprans, S. and Edgar, R.S. (1981) 'The cuticle of Caenorhabditis elegans. II. Stage-specific changes in ultrastructure and protein composition during postembryonic development', *Developmental biology*, 86(2), pp. 456–470.

33. Danecek, P. *et al.* (2021) 'Twelve years of SAMtools and BCFtools', *GigaScience*, 10(2). Available at: https://doi.org/10.1093/gigascience/giab008.

34. Davis, M.W. *et al.* (2004) 'A conserved metalloprotease mediates ecdysis in Caenorhabditis elegans', *Development* , 131(23), pp. 6001–6008.

35. Dexheimer, P.J., Wang, J. and Cochella, L. (2020) 'Two MicroRNAs Are Sufficient for Embryonic Patterning in C. elegans', *Current biology: CB*, 30(24), pp. 5058–5065.e5.

36. Dieterich, C. *et al.* (2008) 'The Pristionchus pacificus genome provides a unique perspective on nematode lifestyle and parasitism', *Nature genetics*, 40(10), pp. 1193–1198.

37. Dobin, A. *et al.* (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics* , 29(1), pp. 15–21.

38. Elbashir, S.M., Lendeckel, W. and Tuschl, T. (2001) 'RNA interference is mediated by 21- and 22-nucleotide RNAs', *Genes & development*, 15(2), pp. 188–200.

39. Fire, A. *et al.* (1998) 'Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans', *Nature*, 391(6669), pp. 806–811.

872 40. Frank, F., Sonenberg, N. and Nagar, B. (2010) 'Structural basis for 5'-nucleotide base-
873 specific recognition of guide RNA by human AGO2', *Nature*, 465(7299), pp. 818–822.

874 41. Gao, F. *et al.* (2012) 'Differential DNA methylation in discrete developmental stages of
875 the parasitic nematode Trichinella spiralis', *Genome biology*, 13(10), p. R100.

876 42. Giraldez, A.J. *et al.* (2005) 'MicroRNAs regulate brain morphogenesis in zebrafish',
877 *Science*, 308(5723), pp. 833–838.

878 43. Giraldez, A.J. *et al.* (2006) 'Zebrafish MiR-430 promotes deadenylation and clearance of
879 maternal mRNAs', *Science*, 312(5770), pp. 75–79.

880 44. Gomis-Rüth, F.X., Trillo-Muyo, S. and Stöcker, W. (2012) 'Functional and structural
881 insights into astacin metallopeptidases', *Biological chemistry*, 393(10), pp. 1027–1041.

882 45. Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) 'FIMO: scanning for occurrences of a
883 given motif', *Bioinformatics* , 27(7), pp. 1017–1018.

884 46. Greer, E.L. *et al.* (2015) 'DNA Methylation on N6-Adenine in C. elegans', *Cell*, 161(4), pp.
885 868–878.

886 47. Griffiths-Jones, S. (2004) 'The microRNA Registry', *Nucleic acids research*, 32(Database
887 issue), pp. D109–11.

888 48. Griffiths-Jones, S. *et al.* (2006) 'miRBase: microRNA sequences, targets and gene
889 nomenclature', *Nucleic acids research*, 34(Database issue), pp. D140–4.

890 49. Grimson, A. *et al.* (2008) 'Early origins and evolution of microRNAs and Piwi-
891 interacting RNAs in animals', *Nature*, 455(7217), pp. 1193–1197.

892 50. Grundmann, S. *et al.* (2011) 'MicroRNA-100 regulates neovascularization by
893 suppression of mammalian target of rapamycin in endothelial and vascular smooth muscle
894 cells', *Circulation*, 123(9), pp. 999–1009.

895 51. Gutierrez, A. and Sommer, R.J. (2004) 'Evolution of dnmt-2 and mbd-2-like genes in the
896 free-living nematodes Pristionchus pacificus, Caenorhabditis elegans and Caenorhabditis
897 briggsae', *Nucleic acids research*, 32(21), pp. 6388–6396.

898 52. Gu, W. *et al.* (2009) 'Distinct argonaute-mediated 22G-RNA pathways direct genome
899 surveillance in the C. elegans germline', *Molecular cell*, 36(2), pp. 231–244.

900 53. Han, M. and Sternberg, P.W. (1990) 'let-60, a gene that specifies cell fates during C.
901 elegans vulval induction, encodes a ras protein', *Cell*, 63(5), pp. 921–931.

902 54. Han, T. *et al.* (2009) '26G endo-siRNAs regulate spermatogenic and zygotic gene
903 expression in Caenorhabditis elegans', *Proceedings of the National Academy of Sciences of*
904 *the United States of America*, 106(44), pp. 18674–18679.

905 55. Han, Z. *et al.* (2020) 'Improving Transgenesis Efficiency and CRISPR-Associated Tools
906 Through Codon Optimization and Native Intron Addition in Nematodes', *Genetics*, 216(4),
907 pp. 947–956.

908 56. Herrmann, M. *et al.* (2007) 'The nematode Pristionchus pacificus (Nematoda:
909 Diplogastridae) is associated with the oriental beetle Exomala orientalis (Coleoptera:
910 Scarabaeidae) in Japan', *Zoological science*, 24(9), pp. 883–889.

911 57. Hiramatsu, F. and Lightfoot, J.W. (2023) 'Kin-recognition and predation shape
912 collective behaviors in the cannibalistic nematode Pristionchus pacificus', *PLoS genetics*,
913 19(12), p. e1011056.

914 58. Hishida, R. *et al.* (1996) 'hch-1, a gene required for normal hatching and normal
915 migration of a neuroblast in C. elegans, encodes a protein related to TOLLOID and BMP-1',
916 *The EMBO journal*, 15(16), pp. 4111–4122.

917 59. Holz, A. and Streit, A. (2017) 'Gain and Loss of Small RNA Classes-Characterization of
918 Small RNAs in the Parasitic Nematode Family Strongyloididae', *Genome biology and*
919 *evolution*, 9(10), pp. 2826–2843.

920 60. van den Hoogen, J. *et al.* (2020) 'A global database of soil nematode abundance and
921 functional group composition', *Scientific data*, 7(1), p. 103.

922 61. Houri-Zeevi, L. *et al.* (2020) 'Three Rules Explain Transgenerational Small RNA
923 Inheritance in C. elegans', *Cell*, 182(5), pp. 1186–1197.e12.

924 62. Huang, Y. *et al.* (2021) 'A Zinc Metalloprotease Is Required for Molting and Survival in
925 Parasitic Nematode', *Frontiers in cell and developmental biology*, 9, p. 695003.

926 63. Igreja, C. *et al.* (2022) 'Application of ALFA-Tagging in the Nematode Model Organisms
927 and', *Cells* , 11(23). Available at: https://doi.org/10.3390/cells11233875.

928 64. Ishita, Y. *et al.* (2023) 'Co-option of an Astacin Metalloprotease Is Associated with an
929 Evolutionarily Novel Feeding Morphology in a Predatory Nematode', *Molecular biology and*
930 *evolution*, 40(12). Available at: https://doi.org/10.1093/molbev/msad266.

931 65. Jungblut, B. and Sommer, R.J. (2000) 'Novel cell-cell interactions during vulva
932 development in Pristionchus pacificus', *Development* , 127(15), pp. 3295–3303.

933 66. Kabekkodu, S.P. *et al.* (2018) 'Clustered miRNAs and their role in biological functions
934 and diseases', *Biological reviews of the Cambridge Philosophical Society*, 93(4), pp. 1955–
935 1986.

936 67. Kato, M. *et al.* (2009) 'Dynamic expression of small non-coding RNAs, including novel
937 microRNAs and piRNAs/21U-RNAs, during Caenorhabditis elegans development', *Genome*
938 *biology*, 10(5), p. R54.

939 68. Ketting, R.F. and Cochella, L. (2021) 'Concepts and functions of small RNA pathways in
940 C. elegans', *Current topics in developmental biology*, 144, pp. 45–89.

69. Kozomara, A., Birgaoanu, M. and Griffiths-Jones, S. (2019) 'miRBase: from microRNA sequences to function', *Nucleic acids research*, 47(D1), pp. D155–D162.

70. Kronenberg, D. *et al.* (2010) 'Processing of procollagen III by meprins: new players in extracellular matrix assembly?', *The Journal of investigative dermatology*, 130(12), pp. 2727–2735.

71. Lagesen, K. *et al.* (2007) 'RNAmmer: consistent and rapid annotation of ribosomal RNA genes', *Nucleic acids research*, 35(9), pp. 3100–3108.

72. Langmead, B. *et al.* (2009) 'Ultrafast and memory-efficient alignment of short DNA sequences to the human genome', *Genome biology*, 10(3), p. R25.

73. Langmead, B. and Salzberg, S.L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature methods*, 9(4), pp. 357–359.

74. Lawrence, P.A. and Struhl, G. (1996) 'Morphogens, compartments, and pattern: lessons from drosophila?', *Cell*, 85(7), pp. 951–961.

75. Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) 'The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14', *Cell*, 75(5), pp. 843–854.

76. Lee, R.C., Hammell, C.M. and Ambros, V. (2006) 'Interacting endogenous and exogenous RNAi pathways in Caenorhabditis elegans', *RNA* , 12(4), pp. 589–597.

77. Levis, N.A. and Ragsdale, E.J. (2023) 'A histone demethylase links the loss of plasticity to nongenetic inheritance and morphological change', *Nature communications*, 14(1), p. 8439.

78. Liao, Y., Smyth, G.K. and Shi, W. (2014) 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics* , 30(7), pp. 923–930.

79. Li, C. *et al.* (2015) 'Multiple Roles of MicroRNA-100 in Human Cancer and its Therapeutic Potential', *Cellular physiology and biochemistry: international journal of experimental cellular physiology, biochemistry, and pharmacology*, 37(6), pp. 2143–2159.

80. Lightfoot, J.W. *et al.* (2019) 'Small peptide-mediated self-recognition prevents cannibalism in predatory nematodes', *Science*, 364(6435), pp. 86–89.

81. Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics* , 25(16), pp. 2078–2079.

82. Liu, M. *et al.* (2011) 'mir-35 is involved in intestine cell G1/S transition and germ cell proliferation in C. elegans', *Cell research*, 21(11), pp. 1605–1618.

83. Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome biology*, 15(12), p. 550.

84. Lowe, T.M. and Eddy, S.R. (1997) 'tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence', *Nucleic acids research*, 25(5), pp. 955–964.

85. Marks, N.D. *et al.* (2019) 'Profiling microRNAs through development of the parasitic nematode Haemonchus identifies nematode-specific miRNAs that suppress larval development', *Scientific reports*, 9(1), p. 17594.

86. Martinez, N.J. *et al.* (2008) 'Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity', *Genome research*, 18(12), pp. 2005–2015.

87. Massirer, K.B. *et al.* (2012) 'The miR-35-41 family of microRNAs regulates RNAi sensitivity in Caenorhabditis elegans', *PLoS genetics*, 8(3), p. e1002536.

88. McJunkin, K. and Ambros, V. (2017) 'A microRNA family exerts maternal control on sex determination in', *Genes & development*, 31(4), pp. 422–437.

89. Miska, E.A. *et al.* (2007) 'Most Caenorhabditis elegans microRNAs are individually not essential for development or viability', *PLoS genetics*, 3(12), p. e215.

90. Moazed, D. (2009) 'Small RNAs in transcriptional gene silencing and genome defence', *Nature*, 457(7228), pp. 413–420.

91. Möhrlen, F., Hutter, H. and Zwilling, R. (2003) 'The astacin protein family in Caenorhabditis elegans', *European journal of biochemistry / FEBS*, 270(24), pp. 4909–4920.

92. Nelson, C. and Ambros, V. (2021) 'A cohort of Caenorhabditis species lacking the highly conserved let-7 microRNA', *G3* , 11(3). Available at: https://doi.org/10.1093/g3journal/jkab022.

93. Nomura, K. *et al.* (1991) 'The specificity of sea urchin hatching enzyme (envelysin) places it in the mammalian matrix metalloproteinase family', *Biochemistry*, 30(25), pp. 6115–6123.

94. Novelli, J., Ahmed, S. and Hodgkin, J. (2004) 'Gene interactions in Caenorhabditis elegans define DPY-31 as a candidate procollagen C-proteinase and SQT-3/ROL-4 as its predicted major target', *Genetics*, 168(3), pp. 1259–1273.

95. Nüsslein-Volhard, C. and Wieschaus, E. (1980) 'Mutations affecting segment number and polarity in Drosophila', *Nature*, 287(5785), pp. 795–801.

96. O'Brown, Z.K. *et al.* (2019) 'Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA', *BMC genomics*, 20(1), p. 445.

97. Ogawa, A. *et al.* (2009) 'A conserved endocrine mechanism controls the formation of dauer and infective larvae in nematodes', *Current biology: CB*, 19(1), pp. 67–71.

98. Ottaviani, S. *et al.* (2018) 'TGF-β induces miR-100 and miR-125b but blocks let-7a through LIN28B controlling PDAC progression', *Nature communications*, 9(1), p. 1845.

99. Pak, J. and Fire, A. (2007) 'Distinct populations of primary and secondary effectors during RNAi in C. elegans', *Science*, 315(5809), pp. 241–244.

100. Pasquinelli, A.E. *et al.* (2000) 'Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA', *Nature*, 408(6808), pp. 86–89.

101. Petrelli, A. *et al.* (2012) 'Sequential analysis of multistage hepatocarcinogenesis reveals that miR-100 and PLK1 dysregulation is an early event maintained along tumor progression', *Oncogene*, 31(42), pp. 4517–4526.

102. Prabh, N. *et al.* (2018) 'Deep taxon sampling reveals the evolutionary dynamics of novel gene families in nematodes', *Genome research*, 28(11), pp. 1664–1674.

103. Qin, C., Huang, R.-Y. and Wang, Z.-X. (2015) 'Potential role of miR-100 in cancer diagnosis, prognosis, and therapy', *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine*, 36(3), pp. 1403–1409.

104. Ragsdale, E.J. *et al.* (2013) 'A developmental switch coupled to the evolution of plasticity acts through a sulfatase', *Cell*, 155(4), pp. 922–933.

105. Rechavi, O. *et al.* (2014) 'Starvation-induced transgenerational inheritance of small RNAs in C. elegans', *Cell*, 158(2), pp. 277–287.

106. Rechavi, O. and Lev, I. (2017) 'Principles of Transgenerational Small RNA Inheritance in Caenorhabditis elegans', *Current biology: CB*, 27(14), pp. R720–R730.

107. Reinhart, B.J. *et al.* (2000) 'The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans', *Nature*, 403(6772), pp. 901–906.

108. Rödelsperger, C. *et al.* (2017) 'Single-Molecule Sequencing Reveals the Chromosome-Scale Genomic Architecture of the Nematode Model Organism Pristionchus pacificus', *Cell reports*, 21(3), pp. 834–844.

109. Röseler, W. *et al.* (2022) 'The improved genome of the nematode Parapristionchus giblindavisi provides insights into lineage-specific gene family evolution', *G3* , 12(10). Available at: https://doi.org/10.1093/g3journal/jkac215.

110. Ruby, J.G. *et al.* (2006) 'Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans', *Cell*, 127(6), pp. 1193–1207.

111. Rudel, D., Riebesell, M. and Sommer, R.J. (2005) 'Gonadogenesis in Pristionchus pacificus and organ evolution: development, adult morphology and cell-cell interactions in the hermaphrodite gonad', *Developmental biology*, 277(1), pp. 200–221.

112. Sarkies, P. *et al.* (2015) 'Ancient and novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages', *PLoS biology*, 13(2), p. e1002061.

113. Sarkies, P. and Miska, E.A. (2014) 'Small RNAs break out: the molecular cell biology of mobile small RNAs', *Nature reviews. Molecular cell biology*, 15(8), pp. 525–535.

114. Sempere, L.F. *et al.* (2003) 'Temporal regulation of microRNA expression in Drosophila melanogaster mediated by hormonal signals and broad-Complex gene activity', *Developmental biology*, 259(1), pp. 9–18.

115. Shaw, W.R. *et al.* (2010) 'The conserved miR-51 microRNA family is redundantly required for embryonic development and pharynx attachment in Caenorhabditis elegans', *Genetics*, 185(3), pp. 897–905.

116. Shukla, A. *et al.* (2020) 'poly(UG)-tailed RNAs in genome protection and epigenetic inheritance', *Nature*, 582(7811), pp. 283–288.

117. Sieriebriennikov, B. *et al.* (2020) 'Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell', *PLoS genetics*, 16(4), p. e1008687.

118. Sijen, T. *et al.* (2001) 'On the role of RNA amplification in dsRNA-triggered gene silencing', *Cell*, 107(4), pp. 465–476.

119. Simpson, V.J., Johnson, T.E. and Hammen, R.F. (1986) 'Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging', *Nucleic acids research*, 14(16), pp. 6711–6719.

120. Sokol, N.S. *et al.* (2008) 'Drosophila let-7 microRNA is required for remodeling of the neuromusculature during metamorphosis', *Genes & development*, 22(12), pp. 1591–1596.

121. Sokol, N.S. (2012) 'Small temporal RNAs in animal development', *Current opinion in genetics & development*, 22(4), pp. 368–373.

122. Sommer, R.J. (2005) 'Evolution of development in nematodes related to C. elegans', *WormBook: the online review of C. elegans biology*, pp. 1–17.

123. Srinivasan, J. *et al.* (2013) 'The draft genome and transcriptome of Panagrellus redivivus are shaped by the harsh demands of a free-living lifestyle', *Genetics*, 193(4), pp. 1279–1295.

124. Stepek, G. *et al.* (2011) 'The astacin metalloprotease moulting enzyme NAS-36 is required for normal cuticle ecdysis in free-living and parasitic nematodes', *Parasitology*, 138(2), pp. 237–248.

125. Stepek, G. *et al.* (2015) 'A highly conserved, inhibitable astacin metalloprotease from Teladorsagia circumcincta is required for cuticle formation and nematode development', *International journal for parasitology*, 45(5), pp. 345–355.

126. Stepek, G., McCormack, G. and Page, A.P. (2010) 'Collagen processing and cuticle formation is catalysed by the astacin metalloprotease DPY-31 in free-living and parasitic nematodes', *International journal for parasitology*, 40(5), pp. 533–542.

127. Stiernagle, T. (2006) 'Maintenance of C. elegans', *WormBook: the online review of C. elegans biology*, pp. 1–11.

128. Sun, S. *et al.* (2022) 'The oscillating Mucin-type protein DPY-6 has a conserved role in nematode mouth and cuticle formation', *Genetics*, 220(3). Available at: https://doi.org/10.1093/genetics/iyab233.

129. Sun, S., Rödelsperger, C. and Sommer, R.J. (2021) 'Single worm transcriptomics identifies a developmental core network of oscillating genes with deep conservation across nematodes', *Genome research*, 31(9), pp. 1590–1601.

130. Suzuki, H.I. *et al.* (2015) 'Small-RNA asymmetry is directly driven by mammalian Argonautes', *Nature structural & molecular biology*, 22(7), pp. 512–521.

131. Suzuki, M. *et al.* (2004) 'Metalloproteases with EGF, CUB, and thrombospondin-1 domains function in molting of Caenorhabditis elegans', *Biological chemistry*, 385(6), pp. 565–568.

132. Theska, T. *et al.* (2020) 'Geometric morphometrics of microscopic animals as exemplified by model nematodes', *Nature protocols*, 15(8), pp. 2611–2644.

133. Theska, T., Renahan, T. and Sommer, R.J. (2024) 'Starvation resistance in the nematode Pristionchus pacificus requires a conserved supplementary nuclear receptor', *Zoological letters*, 10(1), p. 7.

134. Theska, T. and Sommer, R.J. (2024) 'Feeding-structure morphogenesis in "rhabditid" and diplogastrid nematodes is not controlled by a conserved genetic module', *Evolution & development*, 26(2), p. e12471.

135. Vilimova, M. and Pfeffer, S. (2023) 'Post-transcriptional regulation of polycistronic microRNAs', *Wiley interdisciplinary reviews. RNA*, 14(2), p. e1749.

136. Wang, J. *et al.* (2011) 'Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles', *Genome research*, 21(9), pp. 1462–1477.

137. Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer Science & Business Media.

138. Wighard, S.S. *et al.* (2022) 'A New Hope: A Hermaphroditic Nematode Enables Analysis of a Recent Whole Genome Duplication Event', *Genome biology and evolution*, 14(12). Available at: https://doi.org/10.1093/gbe/evac169.

139. de Wit, E. *et al.* (2009) 'Repertoire and evolution of miRNA genes in four divergent nematode species', *Genome research*, 19(11), pp. 2064–2074.

140. Witte, H. *et al.* (2015) 'Gene inactivation using the CRISPR/Cas9 system in the nematode Pristionchus pacificus', *Development genes and evolution*, 225(1), pp. 55–62.

141. Wozney, J.M. *et al.* (1988) 'Novel regulators of bone formation: molecular clones and activities', *Science*, 242(4885), pp. 1528–1534.

142. Yigit, E. *et al.* (2006) 'Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi', *Cell*, 127(4), pp. 747–757.

143. Yoshida, K. *et al.* (2023) 'Chromosome fusions repatterned recombination rate and facilitated reproductive isolation during Pristionchus nematode speciation', *Nature ecology & evolution*, 7(3), pp. 424–439.

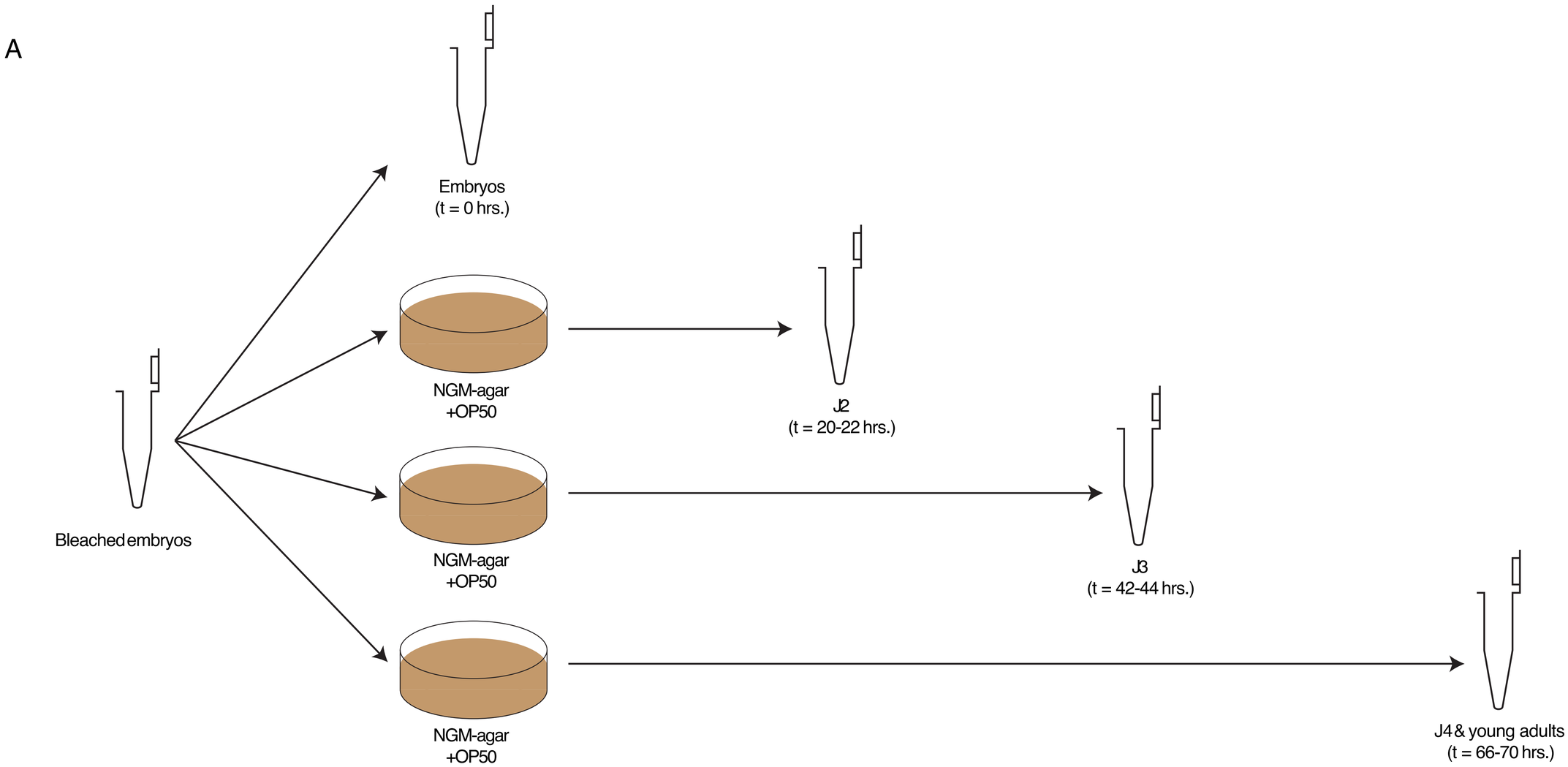
144. Zamanian, M. *et al.* (2015) 'Release of Small RNA-containing Exosome-like Vesicles from the Human Filarial Parasite Brugia malayi', *PLoS neglected tropical diseases*, 9(9), p. e0004069.

145. Zeng, Y. *et al.* (2012) 'MicroRNA-100 regulates osteogenic differentiation of human adipose-derived mesenchymal stem cells by targeting BMPR2', *FEBS letters*, 586(16), pp. 2375–2381.

146. Zheng, Y.-S. *et al.* (2012) 'MiR-100 regulates cell differentiation and survival by targeting RBSP3, a phosphatase-like tumor suppressor in acute myeloid leukemia', *Oncogene*, 31(1), pp. 80–92.

147. Larsson J (2022). *eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses*. R package version 7.0.0, https://CRAN.R-project.org/package=eulerr

148. Martin, M (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, [S.l.], v. 17, n. 1, p. pp. 10-12, may 2011. ISSN 2226-6089. doi:https://doi.org/10.14806/ej.17.1.200.



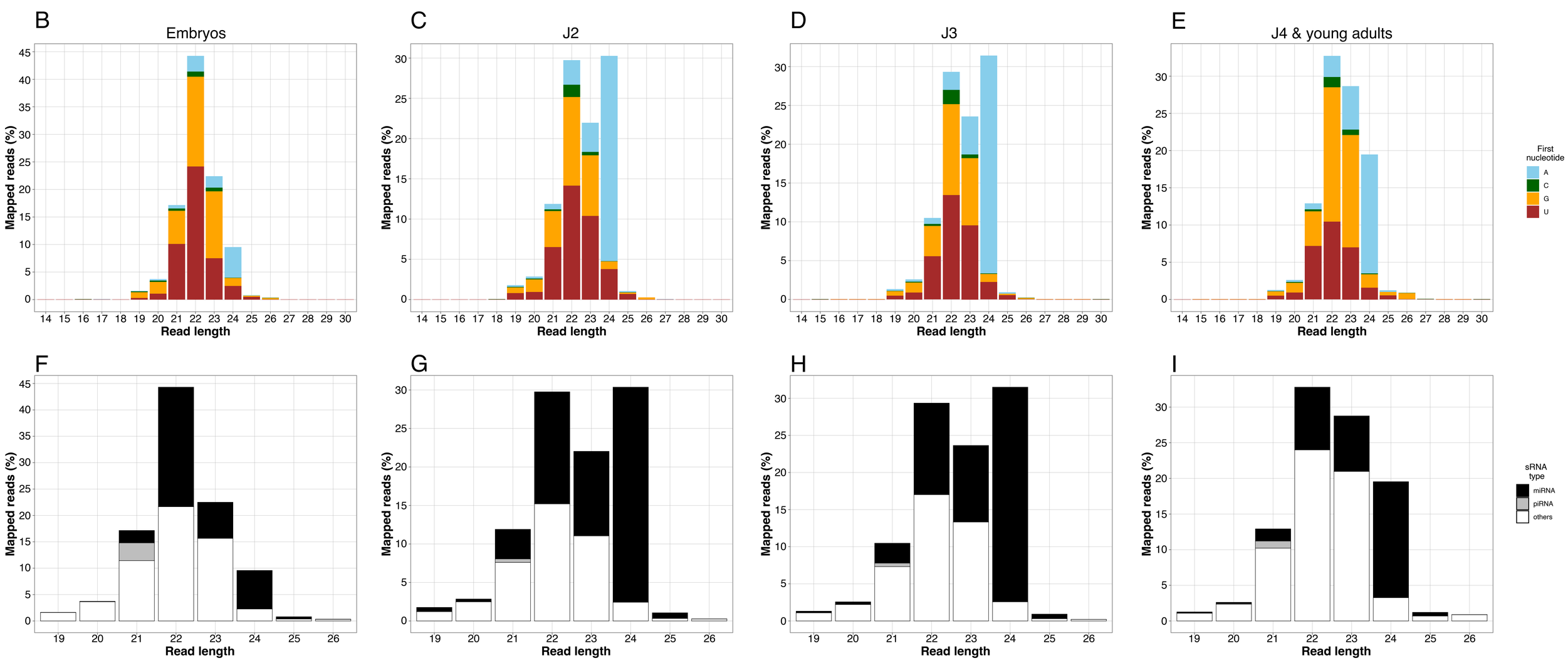


Fig. 1: Developmental small RNA profile of *P. pacificus*. (A) Schematic representation of experimental design. (B-E) Small RNAs in different developmental stages are categorised on the basis of their length (x-axis) and first 5' nucleotide (colour legend 1). (F-I) Corresponding developmental stages showing small RNAs on basis of their length and small RNA type (colour legend 2). Y-axis denotes abundance of each category as fraction of whole small RNA transcriptome for that corresponding stage. Note that the y-axis is different for different stages.

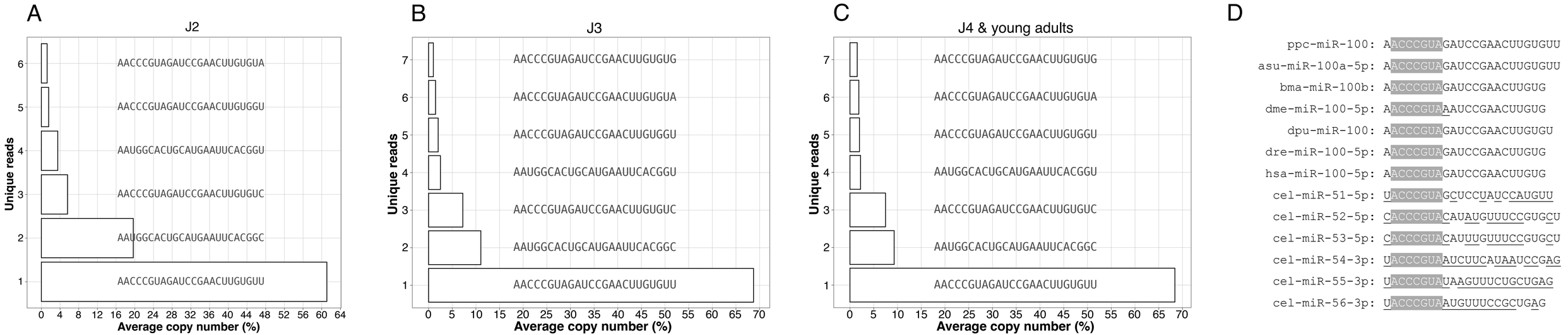


Fig. 2: ppc-miR-100, a highly conserved miRNA, is abundant in juvenile developmental stages of *P. pacificus*. (A-C) Most abundant 24A reads in corresponding developmental stages represented as horizontal bar plots as percentage of total 24A reads. Read#1 corresponds to a miRNA miR-100 on miRBase. (D) Comparison of miR-100 sequences from different species w.r.t. ppc-miR-100. Note the sequence divergence outside of the seed region in *C. elegans* in comparison to other species (seed sequence is highlighted, and sequence alterations are underlined) (species abbreviations: ppc; *P. pacificus*, asu; *A. suum*, bma; *B. malayi*, dme; *D. melanogaster*, dpu; *D. pulex*, dre; *D. rerio*, hsa; *H. sapiens*, cel; *C. elegans*)

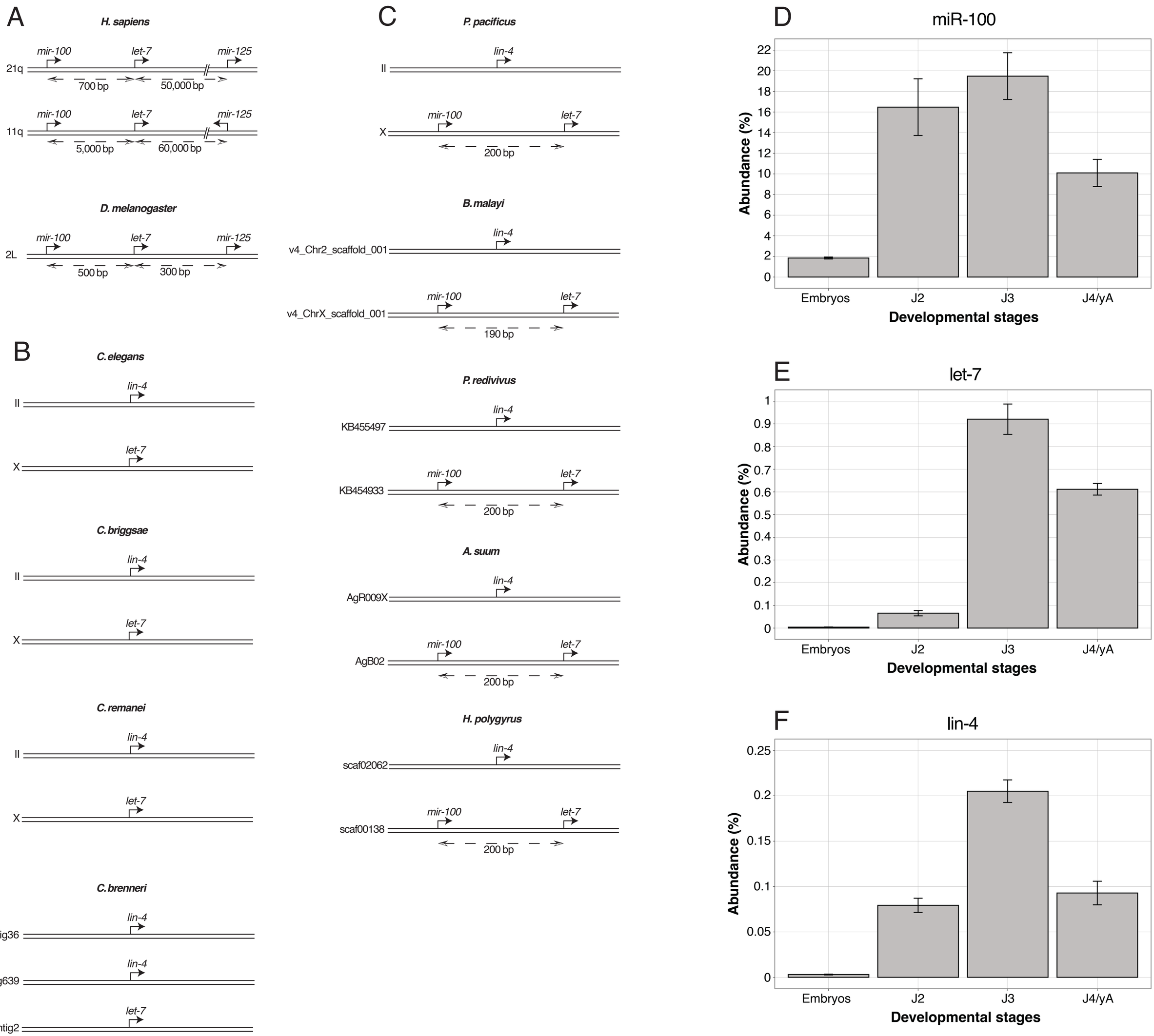
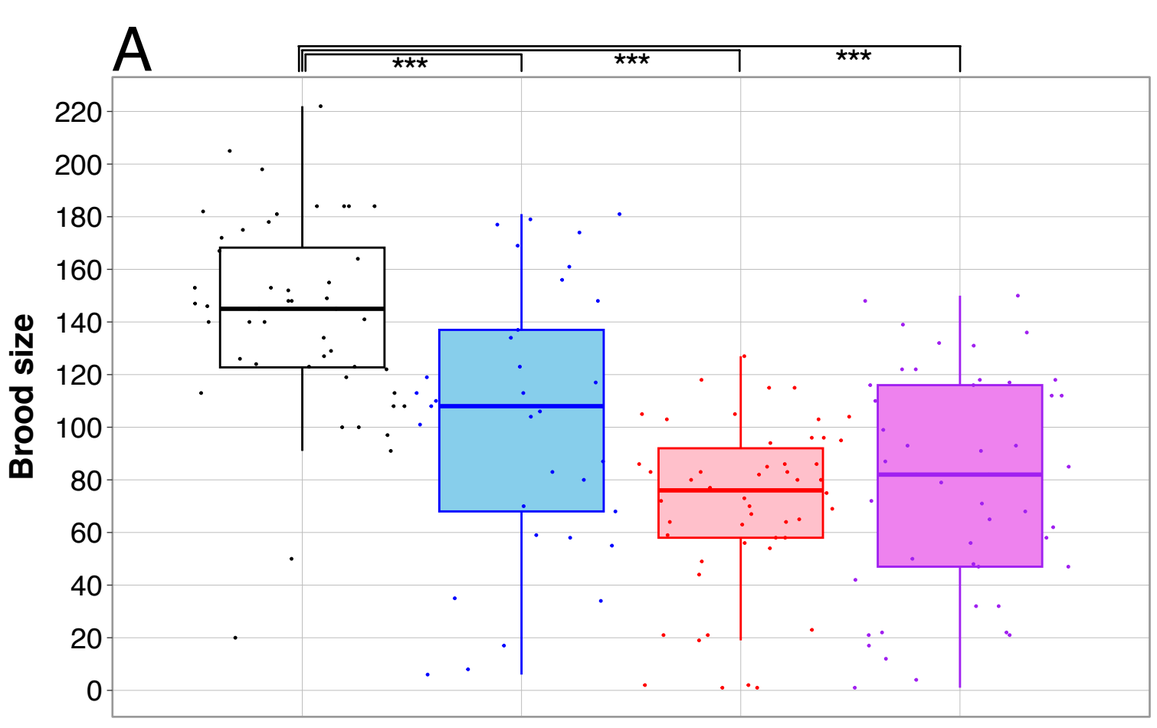
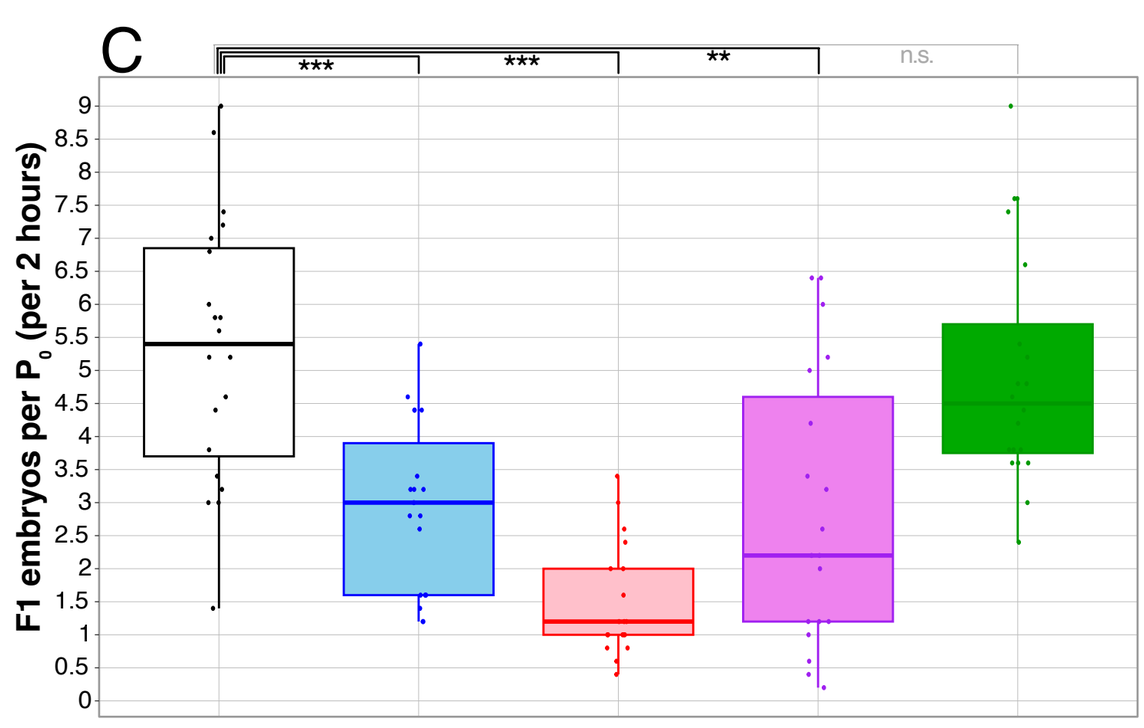
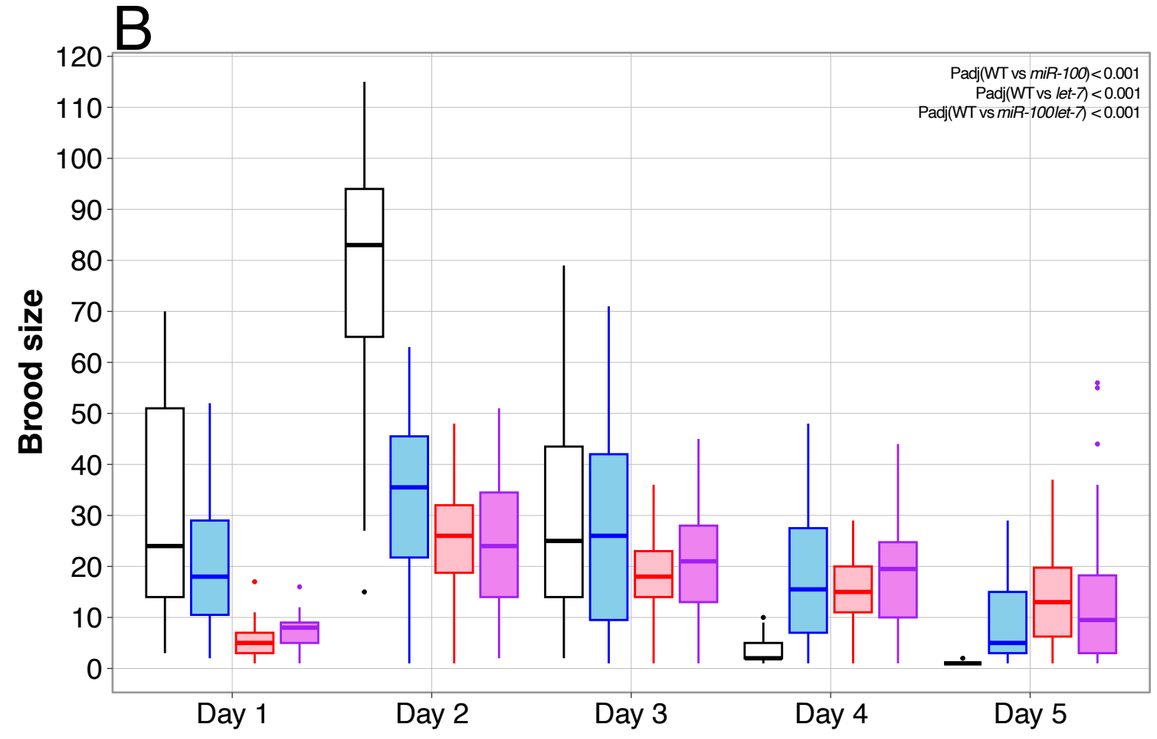
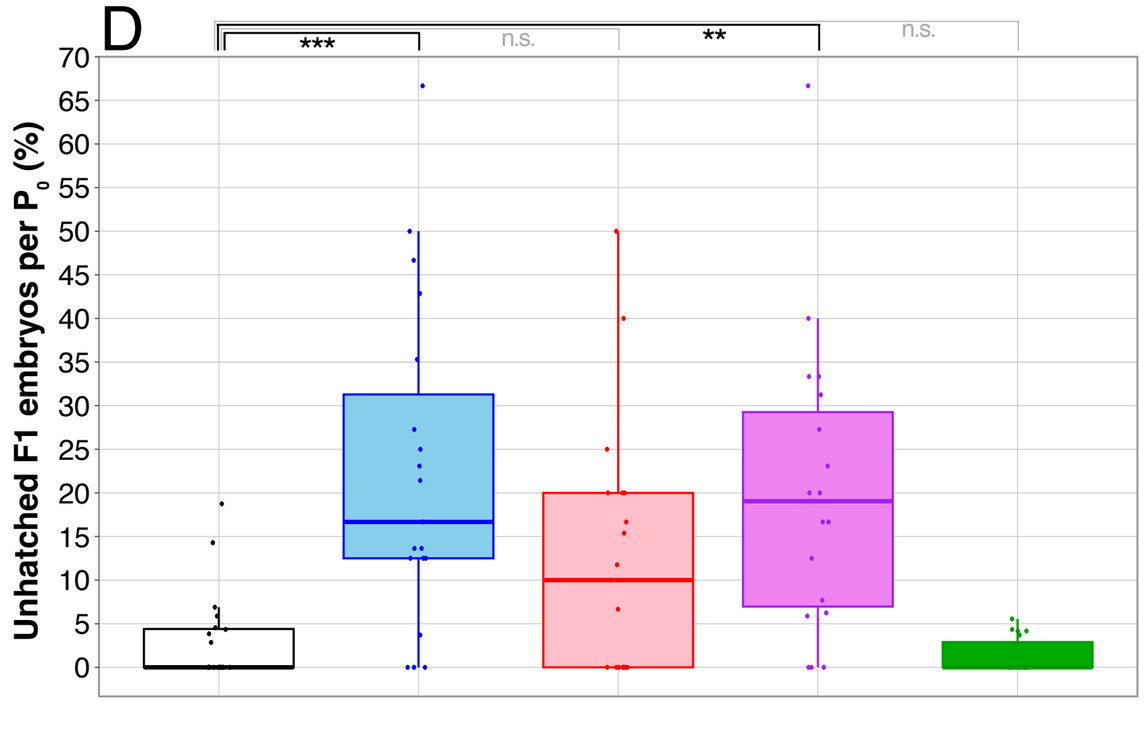


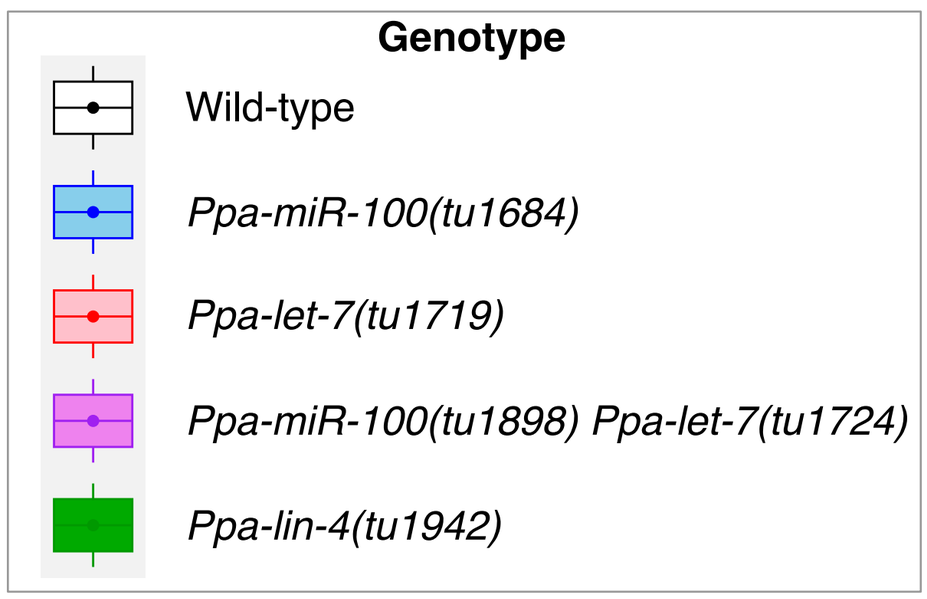
Fig. 3: Two alternative arrangements of *let-7-complex* in nematodes. (A) In humans, fruit flies, and other animals, miRNA-coding genes *miR-100*, *let-7* and *miR-125*/*lin-4* are found on the same chromosome in a genomic locus called the *let-7-complex*. (B) *Caenorhabditis* nematodes lack *miR-100* altogether, and *lin-4* & *let-7* are dispersed on different chromosomes. (C) *P. pacificus* and other non-*Caenorhabditis* nematodes show an alternative arrangement, in which *miR-100* is retained in the genome, and is found upstream of *let-7* on the same chromosome in close proximity, but *lin-4* is found on a different chromosome. (D-F) Developmental expression profiles of miRNAs miR-100, let-7 and lin-4 in *P. pacificus*. All three miRNAs are at their developmental peak in J3. Note that the y-axis is different for the 3 miRNAs.











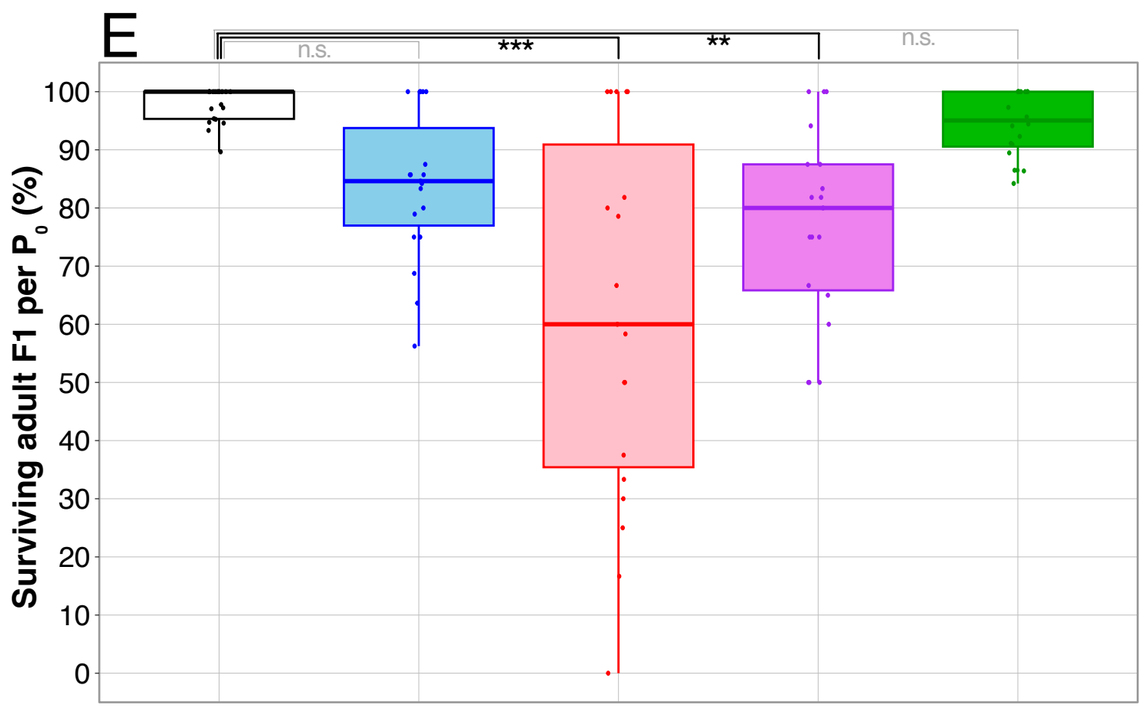


Fig. 4: *Ppa-miR-100*, *Ppa-let-7*, *Ppa-miR-100 Ppa-let-7* double mutant and *Ppa-lin-4* phenotypes. (A-B) Brood sizes of different genotypes taken all together (A) or on a temporal basis (B). n = 50 mothers per genotype for panels A and B. (C) Rate of egg-laying over 2 hours per synchronised mother for each genotype. (D) Percentage of embryonic lethality per synchronised mother per genotype. (E) Percentage of surviving embryos that reach adulthood per synchronised mother per genotype. n = 100 mothers per genotype for panels C-E.

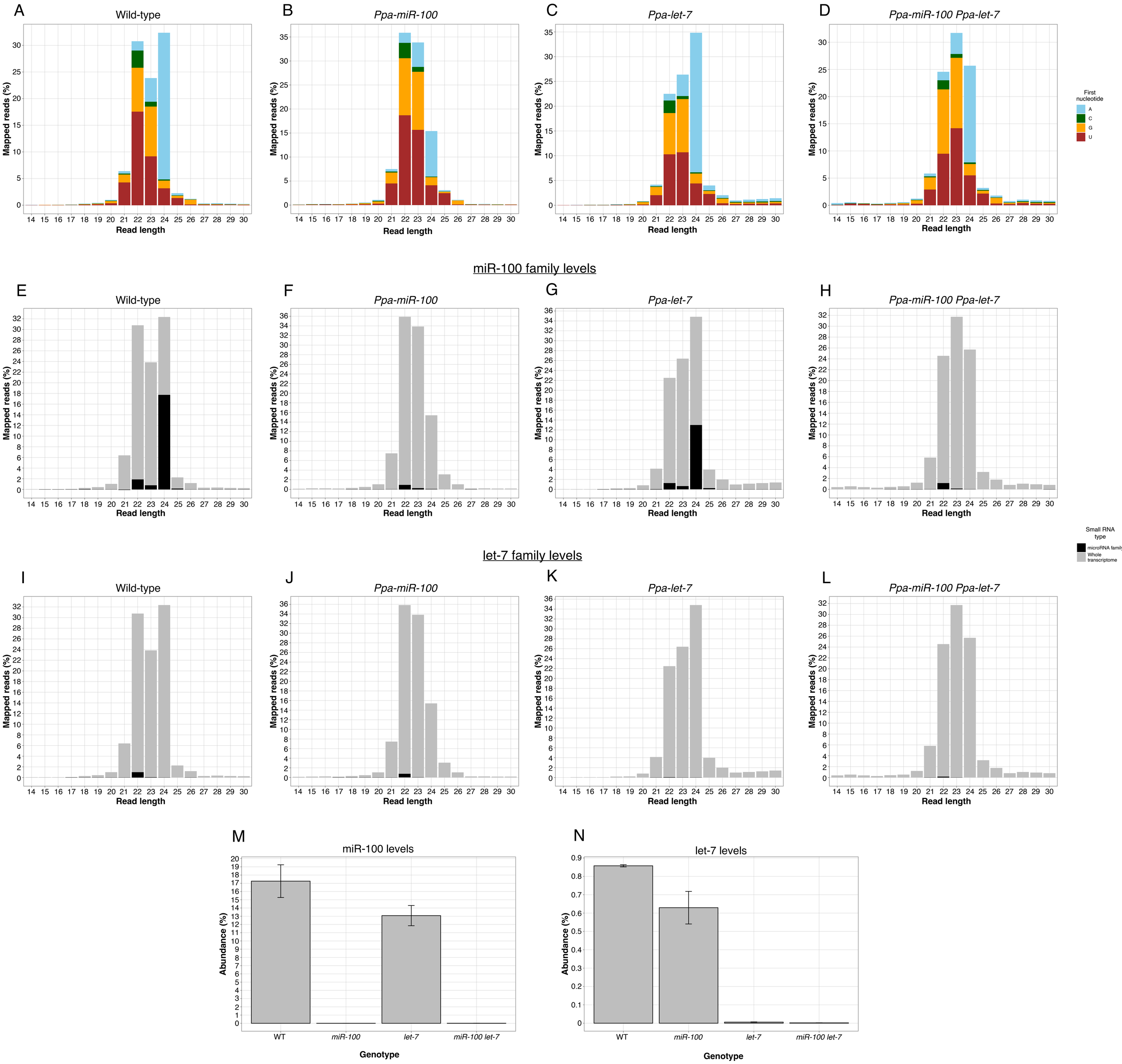


Fig. 5: Small RNA sequencing of wild-type, *Ppa-miR-100*, *Ppa-let-7* and *Ppa-miR-100 Ppa-let-7* double mutants of *P. pacificus*. (A-D) Small RNA profile at J3 stage categorised on the basis of their length (x-axis) and first 5' nucleotide (colour legend 1). (E-L) Expression levels of miR-100 family of miRNAs (E-H) or of let-7 family of miRNAs (I-L) w.r.t. the complete small RNA transcriptome. (M-N) Expression levels of miRNAs miR-100 (M) and let-7 (N) across different genotypes. Note that the y-axis is different between panels A-D, and between N and O.

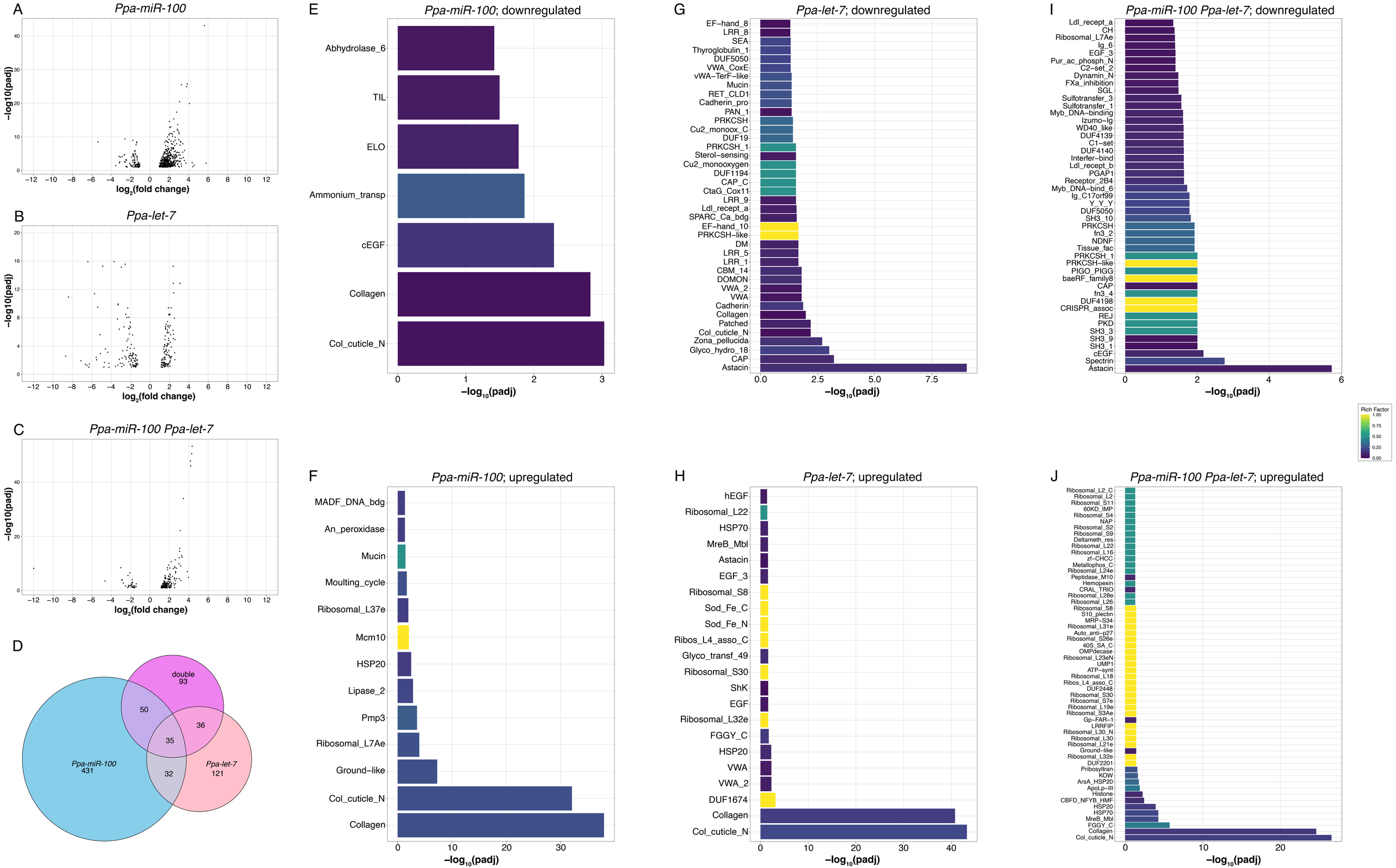


Fig. 6: RNA sequencing of wild-type, *Ppa-miR-100*, *Ppa-let-7*, and *Ppa-miR-100 Ppa-let-7* double mutants of *P. pacificus*. (A-C) Volcano plots at J3 stage resolved along fold-change (x-axis) and adjusted p-value (y-axis). (D) Venn diagram showing the number of overlapping genes between the three mutants. (E-J) Domain enrichment analysis of differentially expressed genes in the mutants, showing all significantly enriched (padj < 0.05) domains on y-axis, and their corresponding adjusted p-values on x-axis. Note that the y-axis is different between panels A-C, and x-axis between E-J.

A: *Ppa-miR-100* alleles

Chromosomal coordinates of wild-type sequence: ChrX: 2633305-2633328 (negative strand)

```
     Wild-type: AACCCGTAGATCCGAACTTGTGTT
RS3883(tu1637): AACCCGTA--------CTTGTGTT
RS3884(tu1638): AACCCGTAGACCAACATCCGAACTTGTGTT
RS3885(tu1639): AACCCG--------AACTTGTGTT
RS3988(tu1684): AACCCGTAGTTACGCTCAACCCGT-----AACTTGTGTT
RS3989(tu1685): AACCCGTA--------CTTGTGTT
RS3990(tu1686): AACCCGTCCTTCTCCTTCCTTCCAATGGCACAGAGCTCAGTGCGCCGATGCAAACGCGCTGGCATTTGCAGCGAATTCTCCGAGCGGATCCGAACACAAGTTCGGATCCGCTC
RS3991(tu1687): AACCCGTAG-------CTTGTGTT
```

B: *Ppa-let-7* alleles

Chromosomal coordinates of wild-type sequence: ChrX: 2633507-2633528 (negative strand)

```
     Wild-type: TGAGGTAGTAGGTTGTATAGTT
RS4025(tu1716): TAGTAGGTTAGTAGGTTGTATAGTT
RS4026(tu1717): AGAGGTAGTAGGTTGTATAGTT
RS4033(tu1718): ---GGTAGTAGGTTGTATAGTT
RS4034(tu1719): AATTCCAACTGCGAGCGAGCGGAGTAGTAGGTTGTATAGTT
RS4035(tu1720): -----TAGTAGGTTGTATAGTT
RS4036(tu1721): AATACCAACTATAGAACTAACGTTAGTTAGCTCGTTAGTATAGTT
RS4037(tu1722): ----------------------
RS4038(tu1723): T-AGGTAGTAGGTTGTATAGTT
RS4039(tu1724): ----------------------
RS4040(tu1725): ----------------------
RS4041(tu1726): -GA-GTAGTAGGTTGTATAGTT
```

C: *Ppa-miR-100 Ppa-let-7* double mutant alleles

Chromosomal coordinates of wild-type sequence: ChrX: 2633305-2633328 (negative strand)

```
            Wild-type: AACCCGTAGATCCGAACTTGTGTT
RS4261(tu1897 tu1724): AACCCGTAGAACAACAACAGACAACAGAG-CCGAACTTGTGTT
RS4262(tu1898 tu1724): AACCCGTAGCTCAACCCCTAGCTCAACCCGTAGCTCAACCGAACTTGTGTT
RS4263(tu1899 tu1724): AACCCGTA-------ACTTGTGTT
RS4264(tu1900 tu1724): AACCCGTTTTTCATTTTTTATT------GAACTTGTGTT
```

D: *Ppa-lin-4* alleles

Chromosomal coordinates of wild-type sequence: ChrII: 381658-381677 (negative strand)

```
     Wild-type: TCCCTGAGACCTCAATTGTG
RS4290(tu1942): TC-------CCTCAATTGTG
RS4291(tu1943): TCCCT--GACCTCAATTGTG
```

Table 1: molecular lesions of all alleles generated in this study. All chromosomal coordinates of the corresponding wild-type sequences of miRNAs are according to El Paco V3 assembly of *P. pacificus* genome. Deletions are shown by a dash (-), insertions by underlined letters, and SNPs by bold letters. (A) *Ppa-miR-100* alleles. Allele *tu1684* has a 15-nucleotide insertion followed by a 5-nucleotide deletion w.r.t. wild-type genomic sequence, and allele *tu1686* has a truncated gene after 7 nucleotides followed by 147-nucleotide insertion w.r.t. wild-type genomic sequence. Also note that the alleles *tu1637* and *tu1685* have the same net molecular lesion, despite being acquired from independent experiments. (B) *Ppa-let-7* alleles. 5 alleles (*tu1718*, *tu1719*, *tu1720*, *tu1721*, and *tu1726*) have gene truncations at their 5' end, and 3 alleles (*tu1722*, *tu1724*, and *tu1725*) have complete deletions, *i.e.*, null mutants. (C) *Ppa-miR-100 Ppa-let-7* double mutant alleles. All double mutant alleles were generated by knocking out *Ppa-miR-100* in *Ppa-let-7(tu1724)* null mutant background. As such, only molecular lesions pertaining to *Ppa-miR-100* gene are shown. Null mutation of *Ppa-let-7* in all double mutant alleles was confirmed using Sanger sequencing. (D) *Ppa-lin-4* alleles

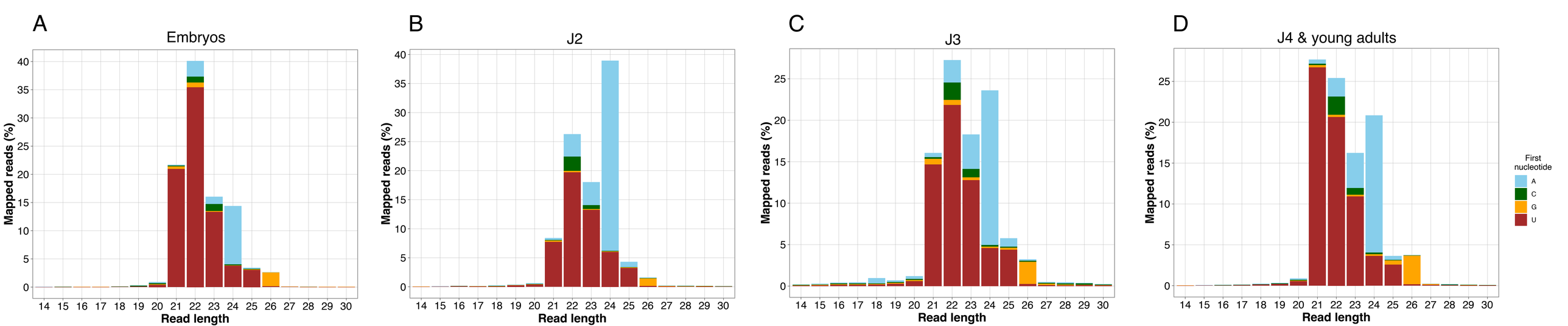
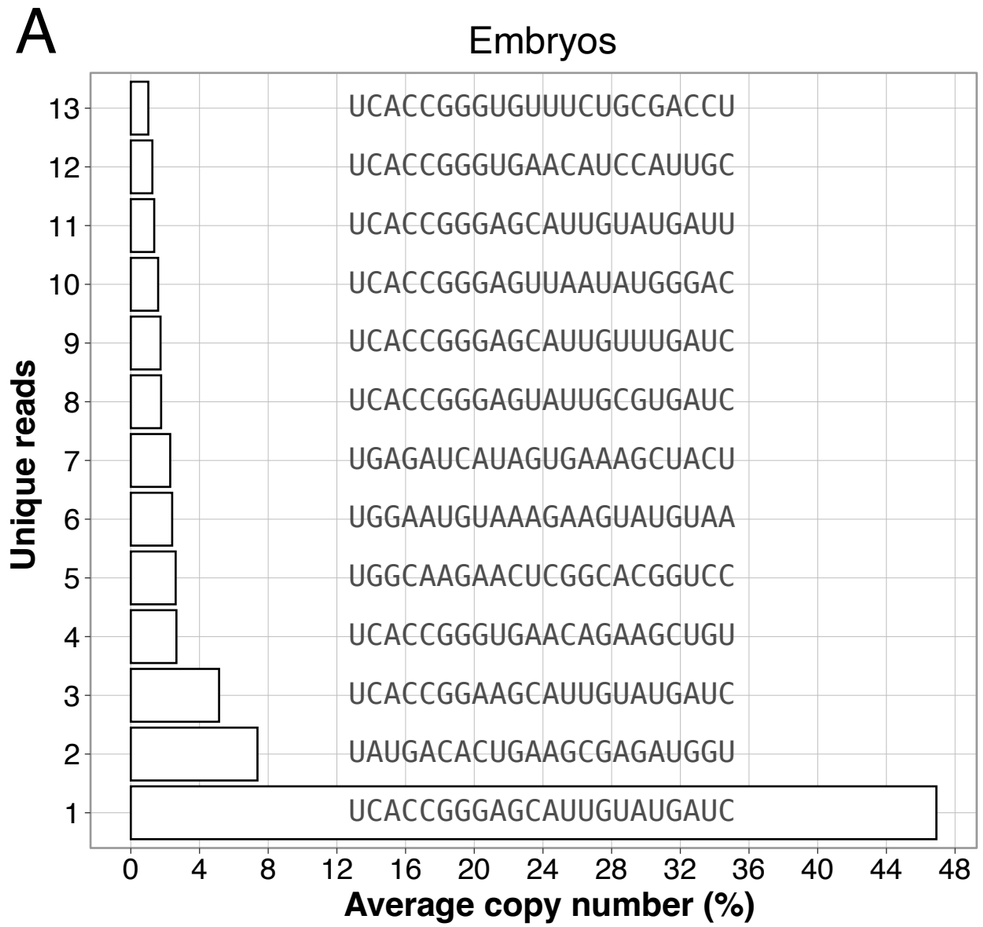


Fig. S1: Developmental small RNA profile of *P. pacificus*, without CCAP treatment. See Fig. 1 for comparison.



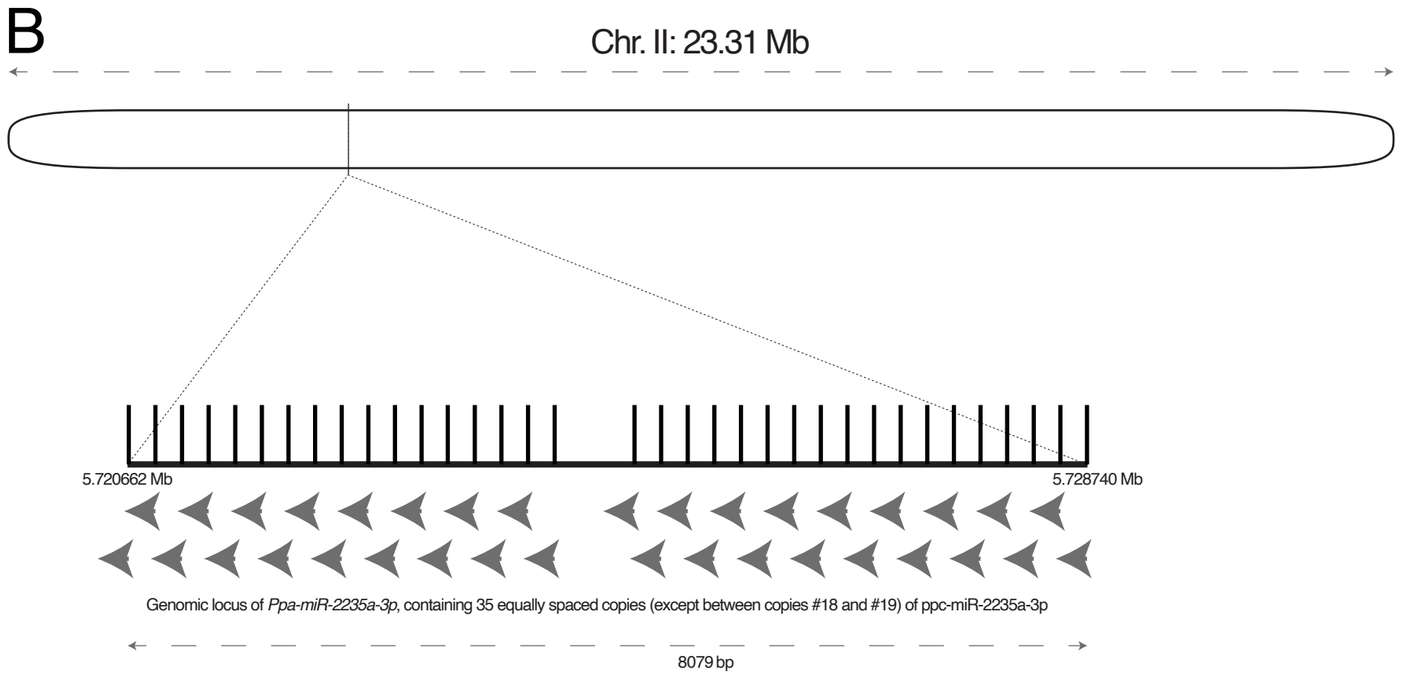
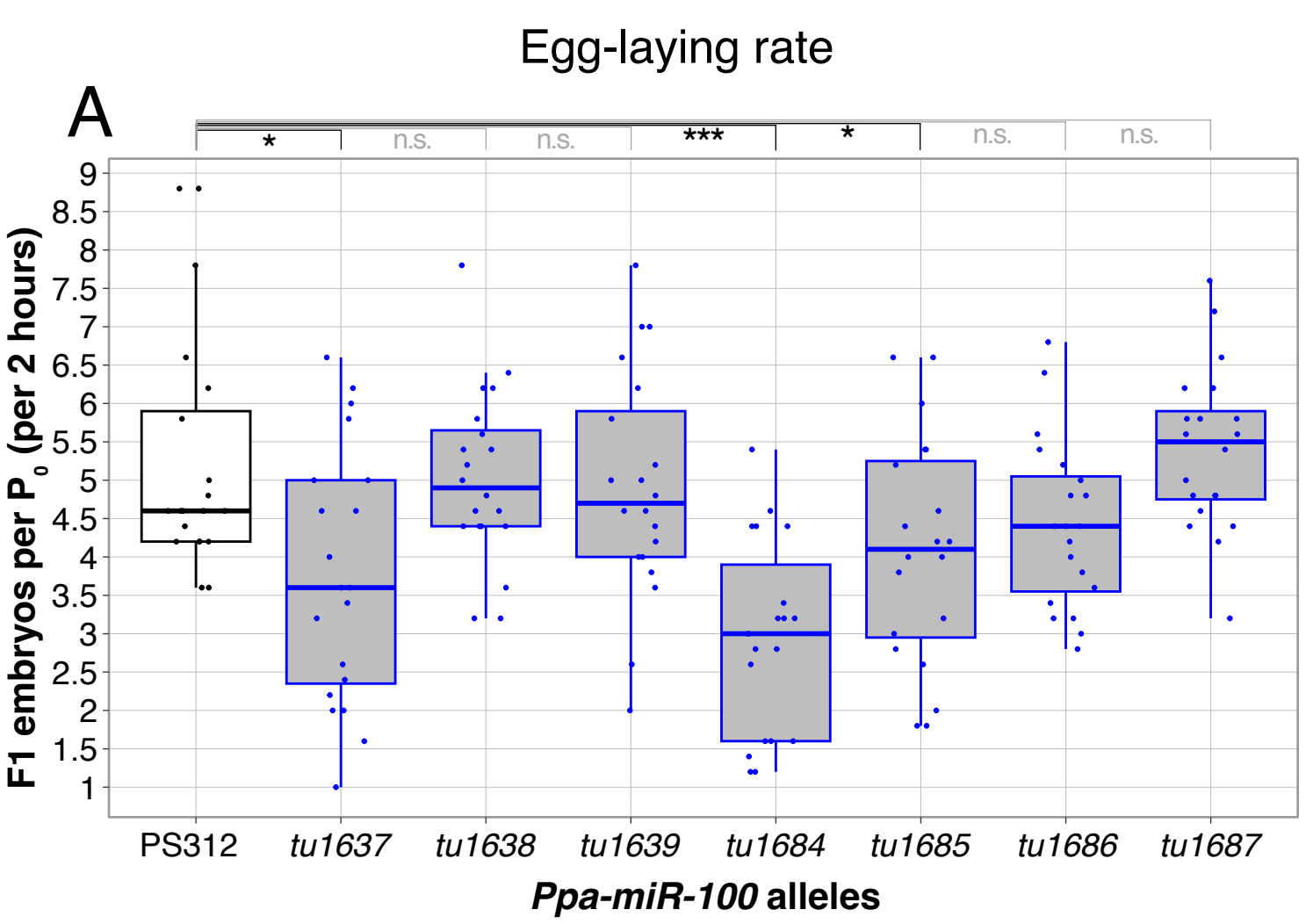
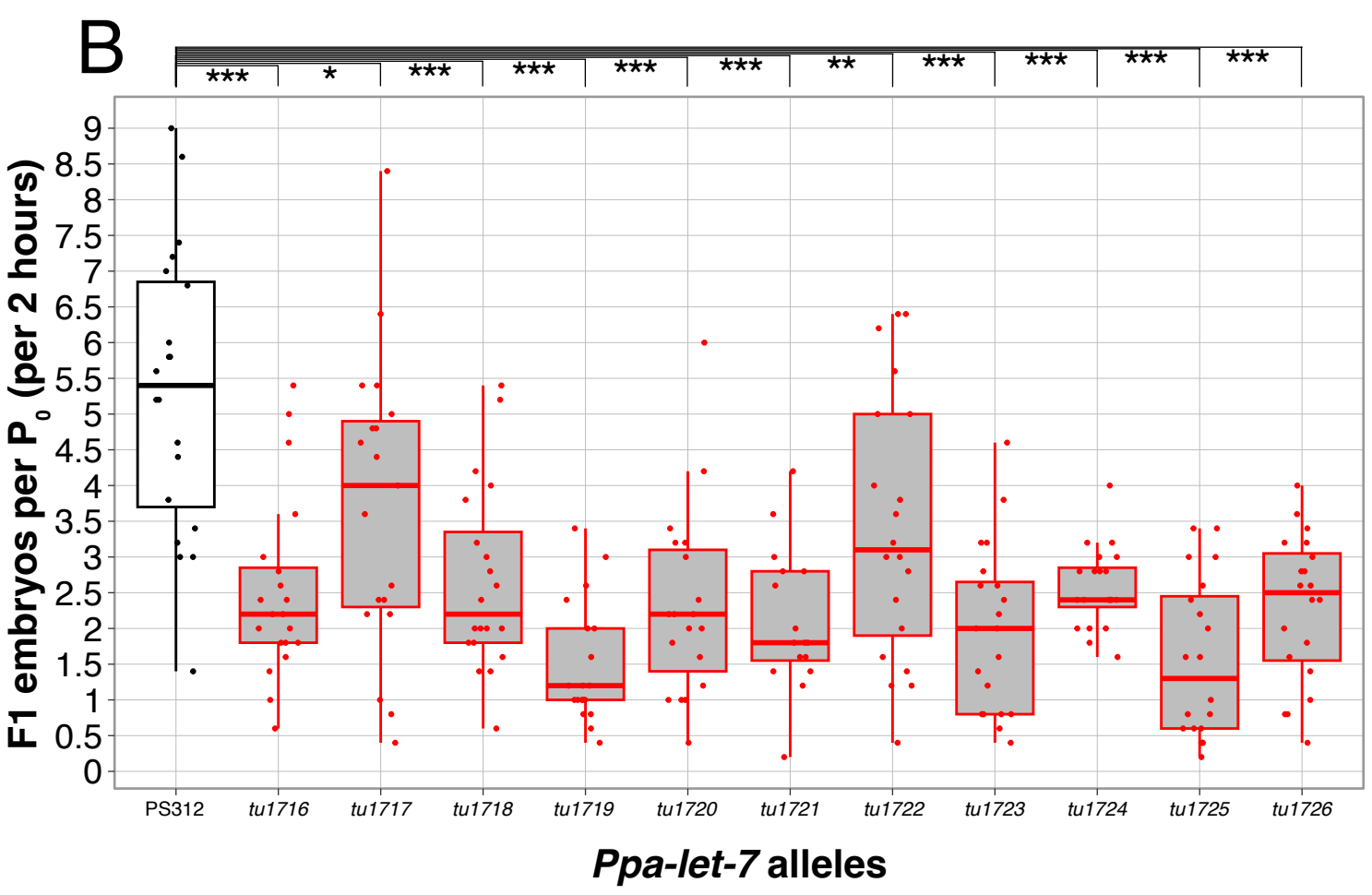


Fig. S2: ppc-miR-2235a-3p, a highly abundant embryonic miRNA, maps to a big locus on chromosome 2 of *P. pacificus*. (A) Most abundant 22U reads in embryos represented as horizontal bar plots as percentage of total 22U reads. Read#1 corresponds to the sequence of ppc-miR-2235a-3p. (B) Schematic representation of *Ppa-miR-2235a-3p* locus on chromosome 2.





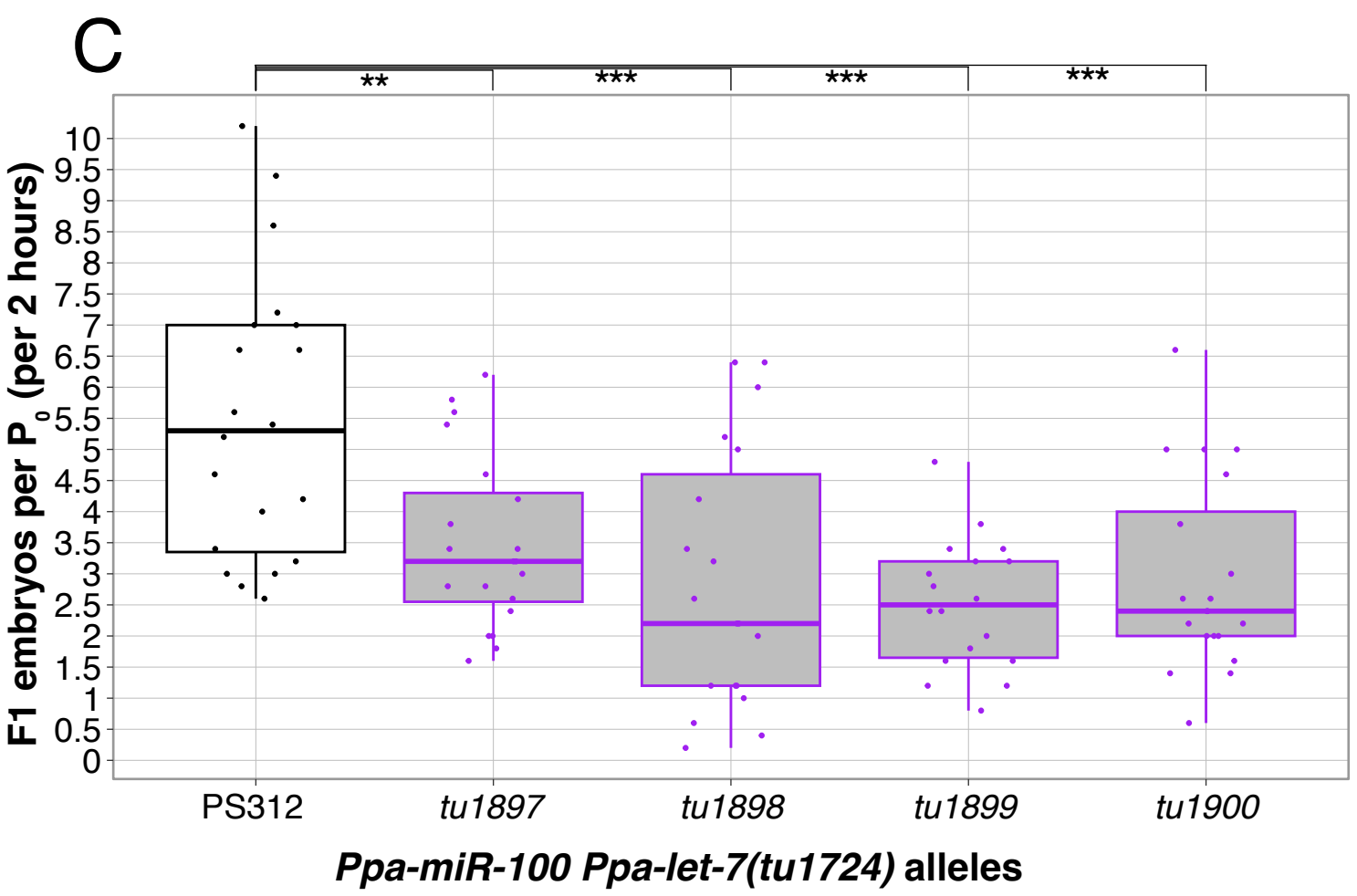


Fig. S3: *Ppa-miR-100*, *Ppa-let-7*, and *Ppa-miR-100 Ppa-let-7* double mutant alleles have varying phenotypic intensities. (A-C) Rate of egg-laying over 2 hours per synchronised mother for each allele. n = 100 mothers per allele per mutant.

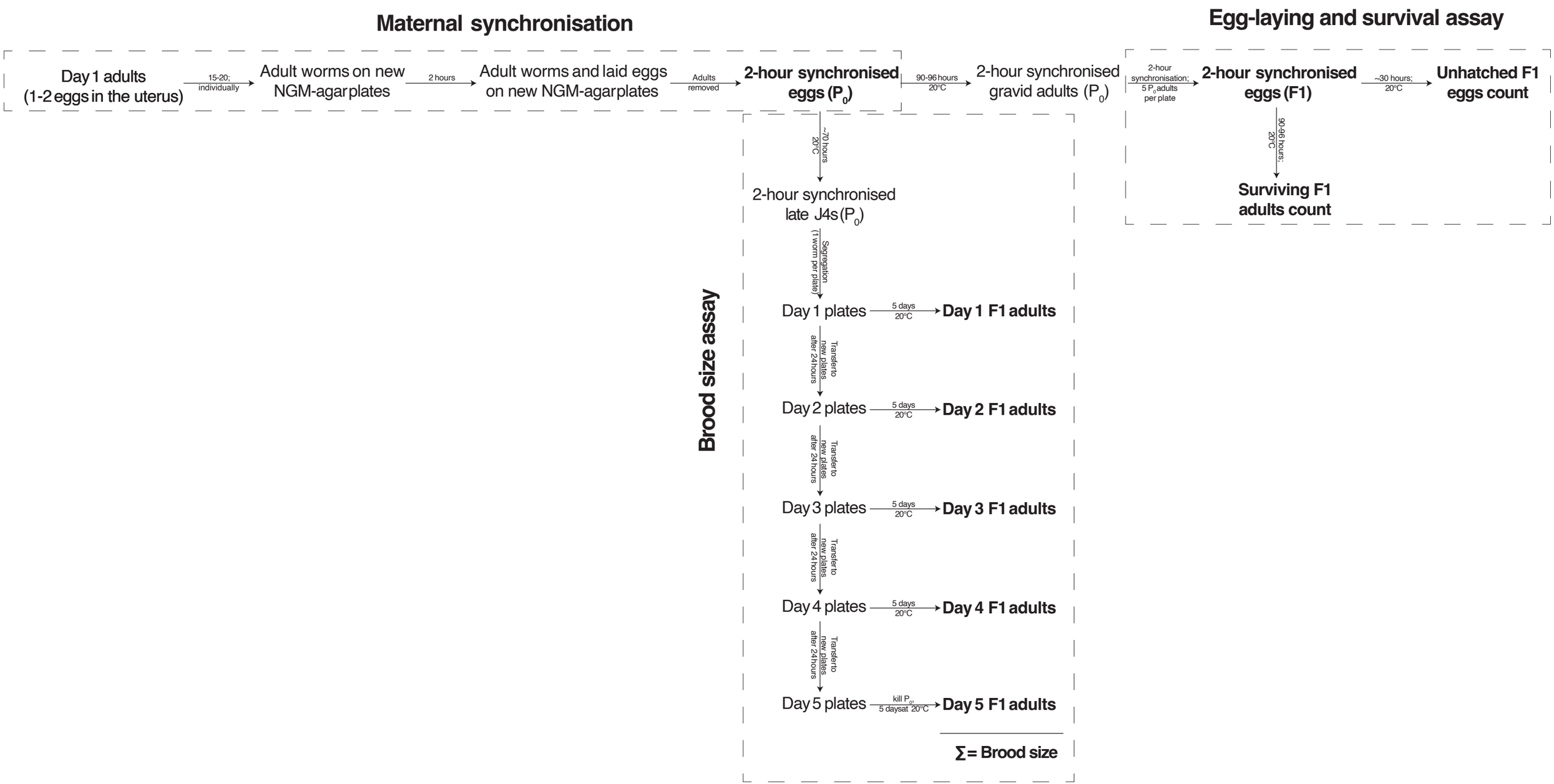


Fig. S4: Schematic representation of phenotypic assays for *Ppa-miR-100*, *Ppa-let-7*, *Ppa-miR-100 Ppa-let-7*, and *Ppa-lin-4* mutants. Synchronised mothers were used to initiate brood size or egg-laying and survival assay, to remove potential effects of maternal age for phenotypic scoring.

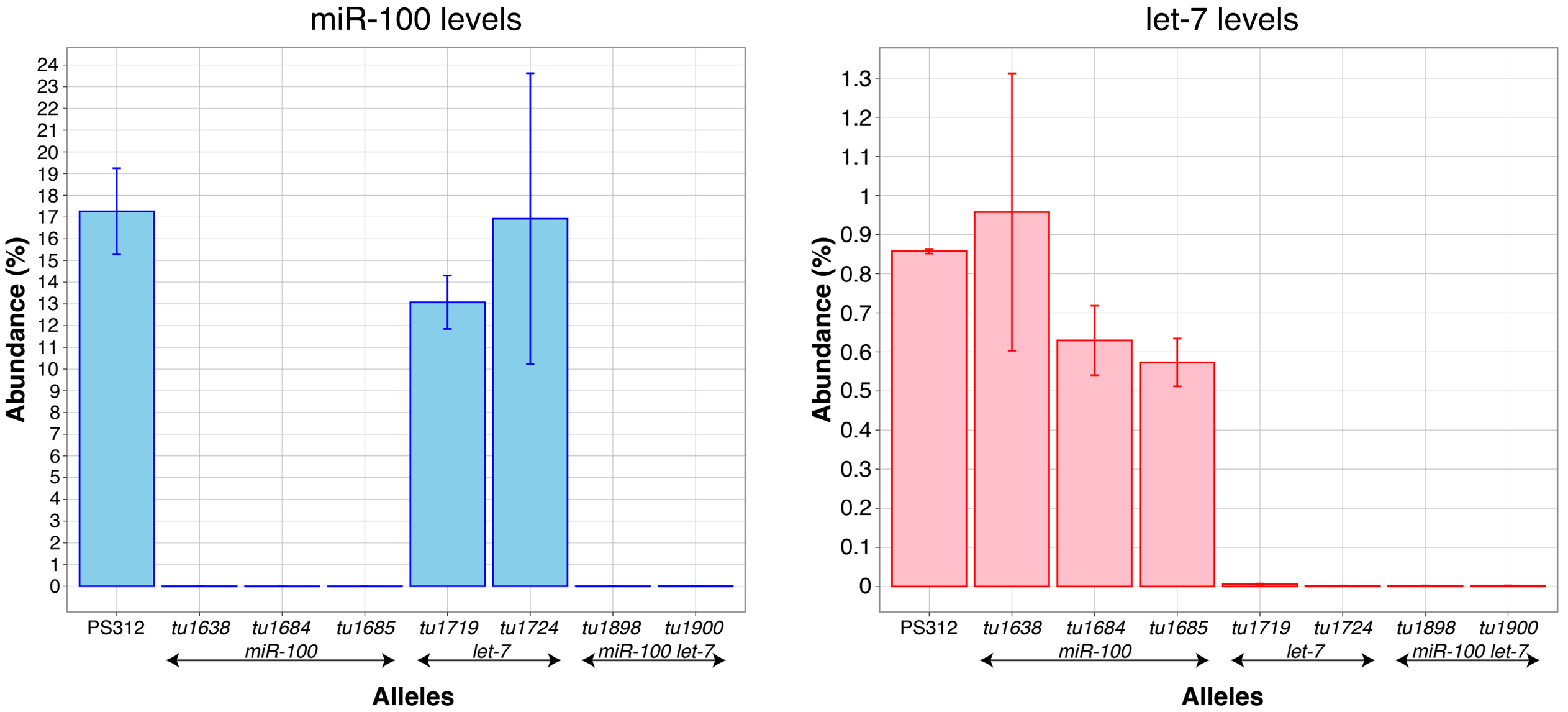
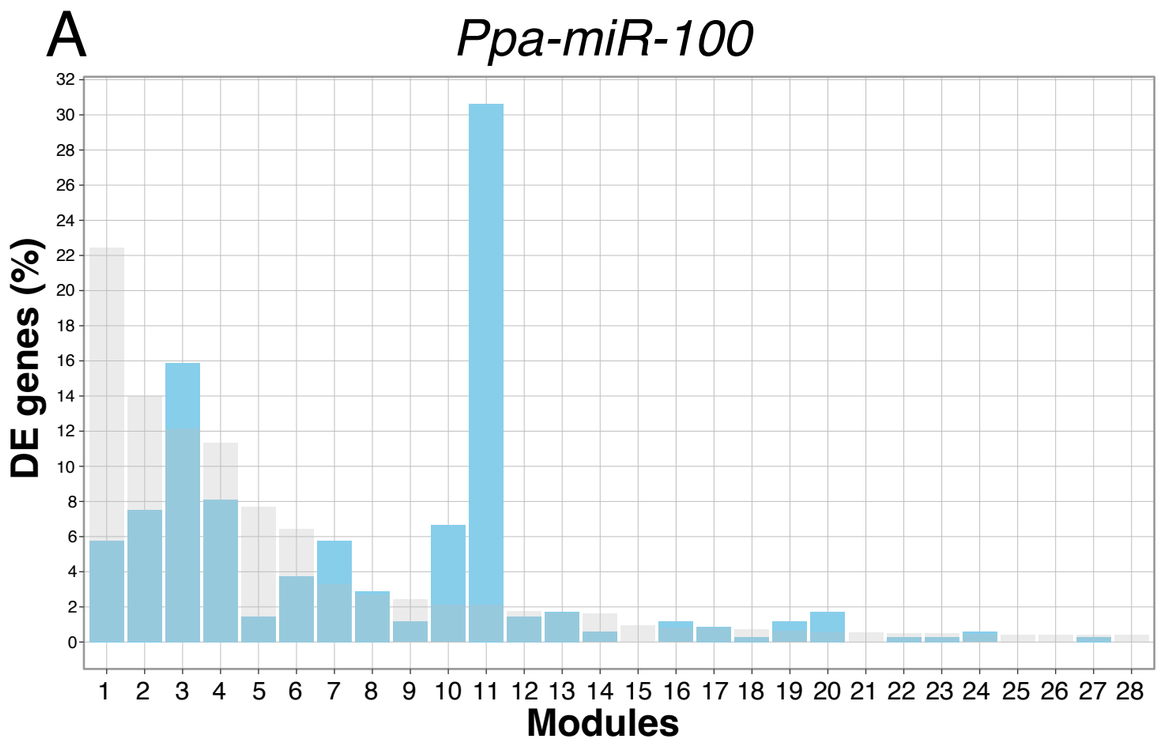
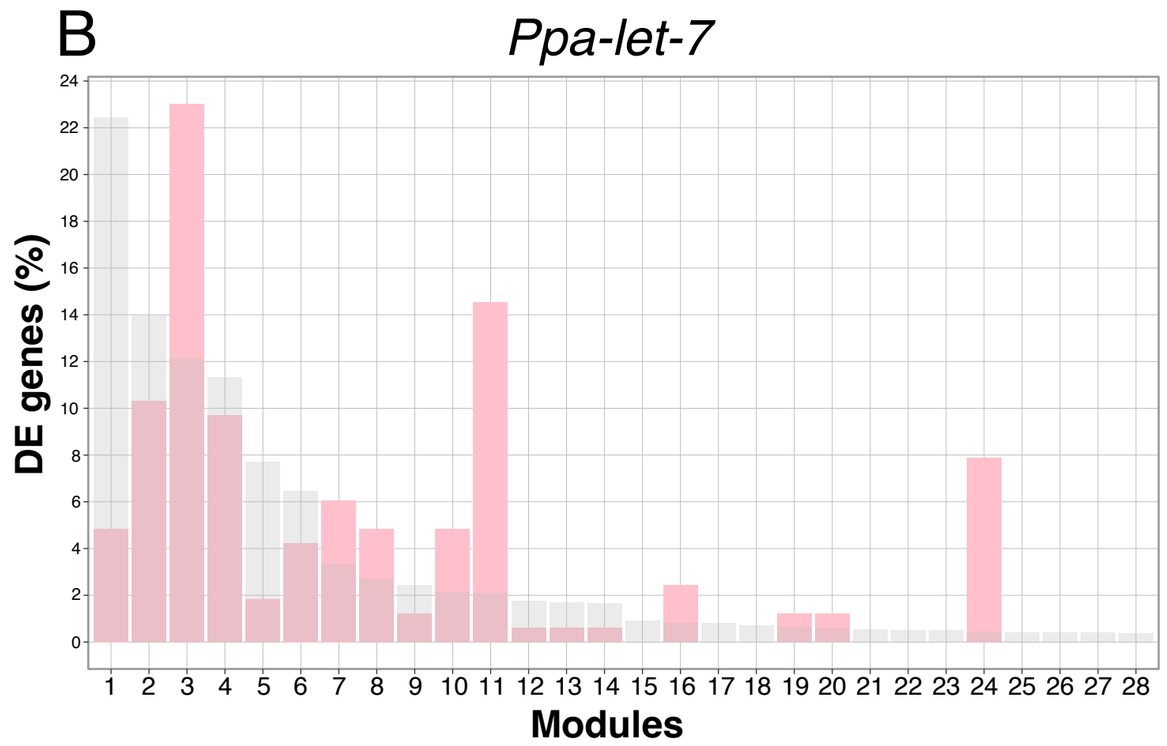
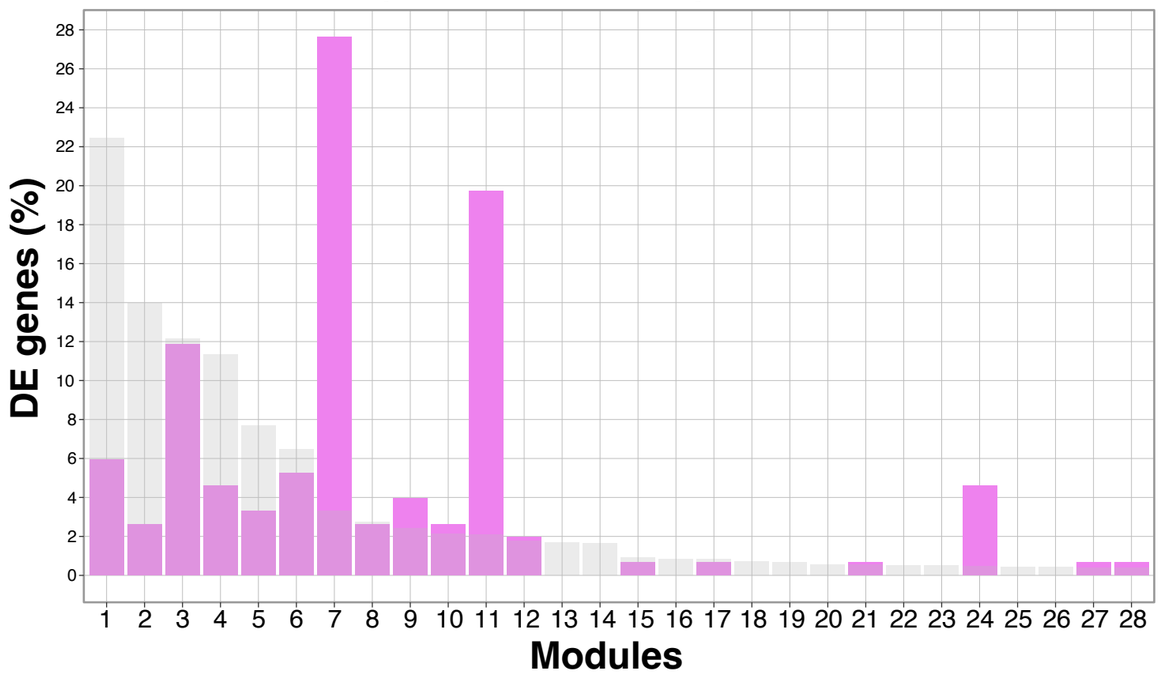
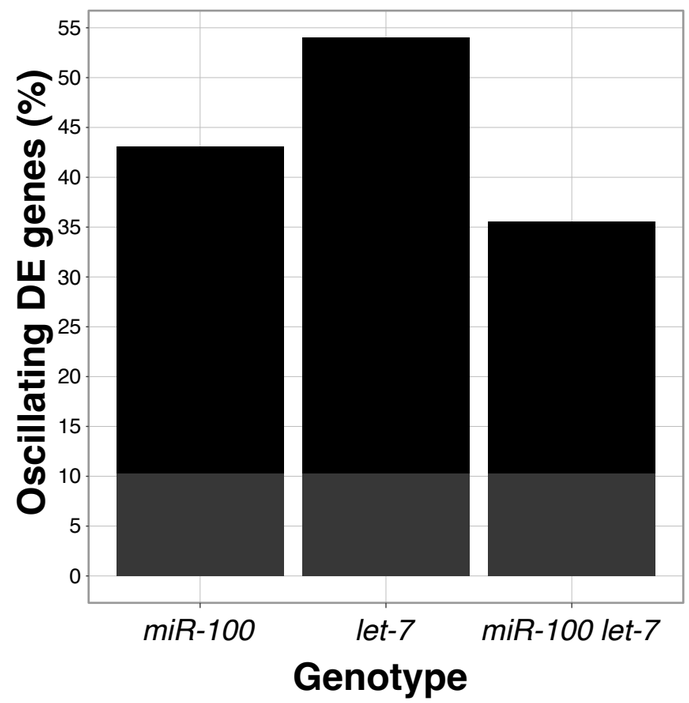


Fig. S5: Expression levels of miRNAs miR-100 and let-7 across different alleles. Note that the y-axis is different between panels.









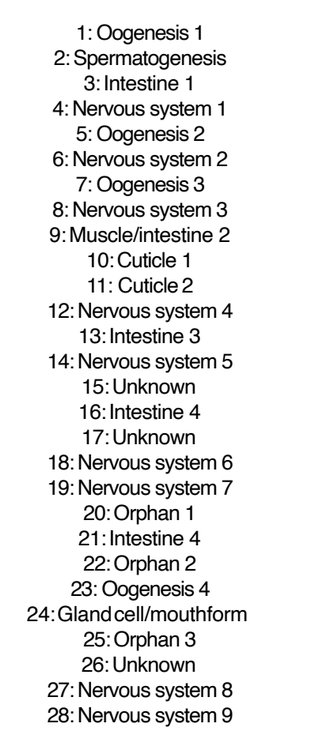
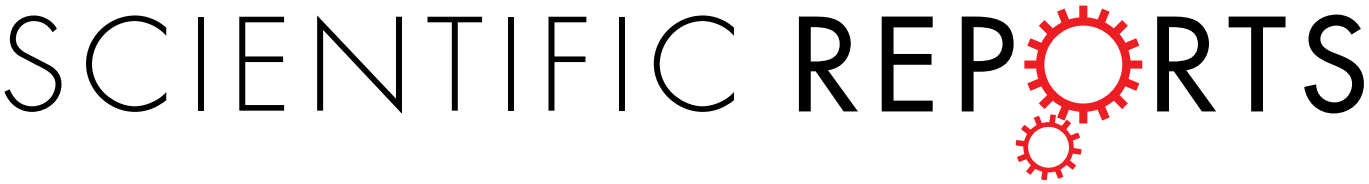


Fig. S6: (A-C) Gene module enrichment analysis of *Ppa-miR-100*, *Ppa-let-7*, and *Ppa-miR-100 Ppa-let-7* double mutants of *P. pacificus*. Grey bars in all three panels represent reference distribution (Athanasouli *et. al.*, 2023). (D) Oscillating genes enrichment analysis of *Ppa-miR-100*, *Ppa-let-7*, and *Ppa-miR-100 Ppa-let-7* double mutants of *P. pacificus*. Grey bars represent the baseline percentage of oscillating genes in the genome (Sun *et. al.*, 2021). (E) Names of modules in panels A-C as given in Athanasouli *et. al.*, 2023. Note that the y-axis is different between panels A-C.

El Paco V1 annotation	El Paco V3 annotation	PFAM	Chromosome	log_2(FC) *miR-100*	log_2(FC) *let-7*	log_2(FC) double
UMM-S328-9.28-mRNA-1	PPA05669	Astacin	X	n.a.	-4.685878953	n.a.
UMM-S2847-7.46-mRNA-1	PPA42525	Astacin	IV	n.a.	n.a.	n.a.
UMM-S2847-6.45-mRNA-1	PPA05955	Astacin	IV	n.a.	-2.375104017	-1.937456096
UMM-S328-7.47-mRNA-1	PPA05618	Astacin	X	n.a.	-3.704532949	-3.00210115
UMA-S293-8.46-mRNA-1	PPA16331	Astacin	X	n.a.	-3.350225515	-2.867998997
UMM-S328-10.33-mRNA-1	PPA39735	CAP	X	n.a.	-1.776499128	-1.624100142
UMM-S57-4.91-mRNA-1	PPA32730	Astacin	I	n.a.	-2.944111154	-2.263647157
UMM-S328-10.78-mRNA-1	PPA13058	CAP	X	n.a.	-1.517549385	-1.588453213
UMM-S283-11.38-mRNA-1	PPA39293	Glyco_hydro_18	IV	n.a.	-1.6499921	n.a.
UMM-S322-3.5-mRNA-1	PPA29522	CAP	X	n.a.	n.a.	n.a.
UMM-S293-11.30-mRNA-1	PPA39470	CAP	X	n.a.	-2.041405047	n.a.
UMA-S322-3.38-mRNA-1	PPA21910	CAP	X	n.a.	-1.475708175	n.a.
UMM-S283-11.45-mRNA-1	Contig1590-snapTAU.1	Glyco_hydro_18; MFS_1	IV	n.a.	-1.686063894	n.a.
UMA-S322-7.39-mRNA-1	PPA21987	Astacin	X	n.a.	n.a.	n.a.
UMS-S2861-1.50-mRNA-1	PPA27985	Astacin	X	n.a.	-2.307702765	n.a.
UMS-S328-0.4-mRNA-1	PPA30108	none	X	n.a.	n.a.	n.a.
UMS-S10-46.25-mRNA-1	PPA27560	none	II	n.a.	-2.378455761	-2.181107418
UMM-S57-36.5-mRNA-1	PPA30435	Lectin_C	I	n.a.	n.a.	n.a.
UMA-S2861-1.27-mRNA-1	PPA34430	Astacin	X	n.a.	n.a.	n.a.
UMM-S250-3.76-mRNA-1	ppa_stranded_DN18096_c0_g1_i1	ShK	X	n.a.	n.a.	n.a.
UMM-S2857-0.30-mRNA-1	PPA20266	Astacin	X	n.a.	-3.158198324	-2.295116979
UMM-S2857-0.41-mRNA-1	PPA42924	Astacin	X	n.a.	-2.810086147	n.a.
UMM-S7-5.16-mRNA-1	PPA03932	Astacin	I	n.a.	n.a.	n.a.
UMA-S2838-46.74-mRNA-1	PPA06264	adh_short; KR; THF_DHG_CYH_C	IV	n.a.	n.a.	n.a.

Fig. S7: Overlap between common targets of *Ppa-nhr-40* & *Ppa-nhr-1*, and miRNA mutants



OPEN

Environmental influence on *Pristionchus pacificus* mouth form through different culture methods

Received: 8 May 2017
Accepted: 28 June 2017
Published online: 03 August 2017

Michael S. Werner, Bogdan Sieriebriennikov, Tobias Loschko, Suryesh Namdeo, Masa Lenuzzi, Mohannad Dardiry, Tess Renahan, Devansh Raj Sharma & Ralf J. Sommer

Environmental cues can impact development to elicit distinct phenotypes in the adult. The consequences of phenotypic plasticity can have profound effects on morphology, life cycle, and behavior to increase the fitness of the organism. The molecular mechanisms governing these interactions are beginning to be elucidated in a few cases, such as social insects. Nevertheless, there is a paucity of systems that are amenable to rigorous experimentation, preventing both detailed mechanistic insight and the establishment of a generalizable conceptual framework. The mouth dimorphism of the model nematode *Pristionchus pacificus* offers the rare opportunity to examine the genetics, genomics, and epigenetics of environmental influence on developmental plasticity. Yet there are currently no easily tunable environmental factors that affect mouth-form ratios and are scalable to large cultures required for molecular biology. Here we present a suite of culture conditions to toggle the mouth-form phenotype of *P. pacificus*. The effects are reversible, do not require the costly or labor-intensive synthesis of chemicals, and proceed through the same pathways previously examined from forward genetic screens. Different species of *Pristionchus* exhibit different responses to culture conditions, demonstrating unique gene-environment interactions, and providing an opportunity to study environmental influence on a macroevolutionary scale.

Phenotypes can be dramatically influenced by environmental conditions experienced during development, a phenomenon referred to as developmental plasticity[1–3]. Examples of plastic phenotypes have been studied for nearly a century, including differences in morphology[4], sex and caste determination[5–7], and innate immunity[8]. However, despite long-held interest in the field, and decade's worth of progress linking genotype to phenotype, relatively little is known about the mechanisms connecting environment to phenotype. To study the mechanisms of environmental influence on phenotype, easily tunable methods to induce phenotypic changes and model organisms amenable to molecular biology techniques are required. For example, temperature and diet have been utilized to explore plasticity in insects and nematodes[9–14], some of which have revealed fundamental principles of dynamic gene regulation. In particular, investigating life cycle plasticity in *C. elegans* contributed to our understanding of nutrition and endocrine signaling[15–18], and the discovery of regulatory RNAs[19]. However, the number of case studies remains small, and heuristic insight of ecologically relevant phenotypes within an evolutionary framework is still lacking.

The model organism *P. pacificus* exhibits an environmentally sensitive developmental switch of its feeding structures[20]. In the wild *P. pacificus* exists in a dormant state (dauer) on beetles. When beetles die *Pristionchus* exits the dauer state to feed on decomposition bacteria, and proceeds to reproductive maturity[21,22] (Fig. 1A). While developing under crowded conditions a "wide-mouthed" eurystomatous (Eu) morph with two teeth is built, which allows adults to prey on other nematodes (Fig. 1B). Alternatively, a "narrow-mouthed" stenostomatous (St) morph with one tooth relegates diet exclusively to microorganisms (Fig. 1C). While Eu animals can exploit additional food sources[23] and attack and kill competitors[24], St animals mature slightly faster[25], creating a tradeoff of strategies depending on the environment perceived during development. Under monoxenic growth conditions in the laboratory using *Escherichia coli* OP50 bacteria as a food source on NGM-agar plates, 70–90% of the reference *P. pacificus* strain PS312 develop the Eu morph. Metabolic studies have elucidated compounds that affect this mouth-form decision. For example, the steroid hormone dafachronic acid shifts mouth-form frequencies to St[20]. Conversely, the pheromone dasc#1 shifts the frequency to Eu[26]. Recent mutant screens

Department of Evolutionary Biology, Max Planck Institute for Developmental Biology, 72076, Tübingen, Germany. Correspondence and requests for materials should be addressed to R.J.S. (email: ralf.sommer@tuebingen.mpg.de)

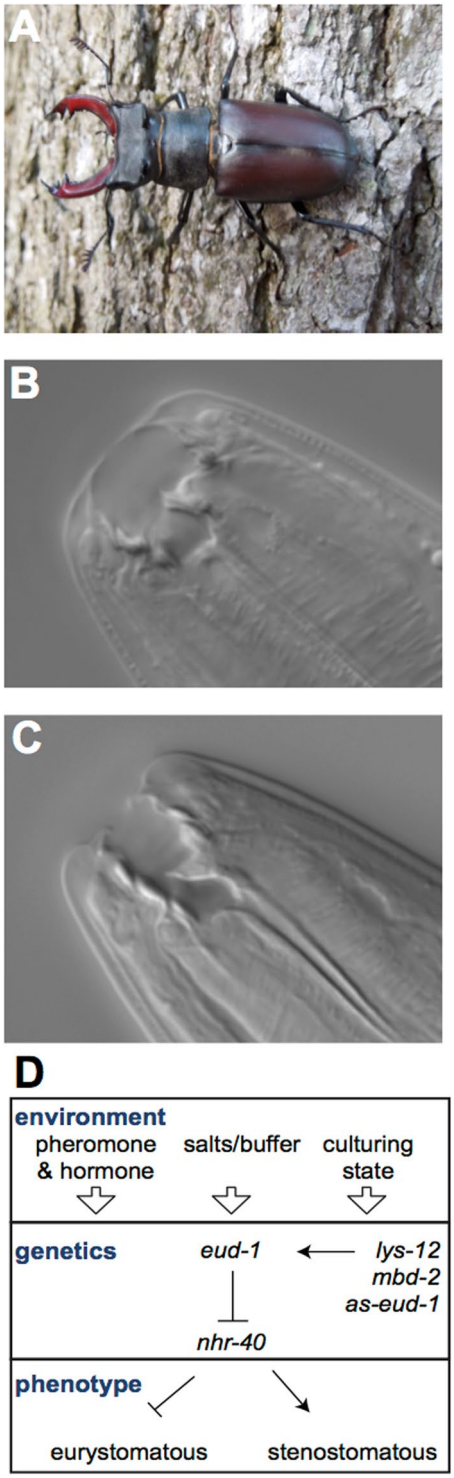


Figure 1. Life cycle and phenotypic plasticity of *Pristionchus pacificus*. (**A**) *P. pacificus* exist in a necromenic relationship with host beetles (i.e. shown here *Lucanus cervus*), and upon decomposition of the beetles the worms exit the dormant (dauer) state. Photo taken by M Herrmann and R Sommer. Depending on environmental conditions experienced during this period, adults develop either (**B**) a wide-mouth "eurystomatous" (Eu) morph with an additional tooth allowing them to prey on other nematodes, or (**C**) a microbivorous narrow mouth "stenostomatous" (St) morph. (**D**) Diagram integrating the environment into known gene-phenotype interactions of the *P. pacificus* mouth-form pathway.

have established several genes in the mouth-form regulatory pathway[27–29]. The sulfatase *eud-1* (*eu*rystomatous *d*efective) is a dosage-dependent "switch" gene[30]: *eud-1* mutants are 100% St, while overexpression of a *eud-1*

transgene confers 100% Eu[27]. The nuclear-hormone-receptor *Ppa-nhr-40* was identified as a suppressor of *eud-1*, and regulates downstream genes[28]. *C. elegans* homologs of the epigenetic enzymes acetyltransferase *lsy-12* and methyl-binding protein *mbd-2* have also been identified to control mouth-form plasticity, and are attractive factors for channeling environmental cues to changes in gene regulation. Both mutants led to global losses of activating-histone modifications, and decreased expression of *eud-1*[29].

Identification of these switch genes affords the opportunity to track regulatory mechanisms that respond to environmental cues[31,32]. Unfortunately, the application of small molecules to affect mouth-form ratios in large enough quantities for biochemical fractionation or epigenetic profiling (e.g. ChIP) is impractical given the labor and expense of chemical synthesis or purification. Moreover, it is difficult to obtain consistent mouth-form ratios with pharmacological compounds as they are in constant competition with endogenous hormones and pheromones[20]. Finally, while crowding/starvation can also induce the Eu morph, it is technically challenging to compare different population densities, or to synchronize starved vs. un-starved larvae. To adequately study environmental effects on phenotypic plasticity, cheap, consistent, and simple methods are needed that can tune mouth-form ratios in synchronized populations. Here, we establish a set of culture conditions to affect environmental influence on mouth form. These methods are fast, reproducible, and only require the differential application of buffer, and culturing state (solid vs. liquid). Intriguingly, different species of *Pristionchus* exhibit different response regimes, suggesting evolutionary divergence of gene-environment interactions.

Results

Liquid culture affects *Pristionchus pacificus* mouth-form. In order to accumulate large amounts of biological material for molecular and biochemical experiments we grew the laboratory California strain (PS312) of *P. pacificus* in liquid culture. To our surprise, this culture condition reversed the mouth-form phenotype from preferentially Eu to preferentially St. To better examine this observation we screened mouth-forms of adults representing a parental generation (P), and obtained[33] and split eggs evenly to either agar plates or liquid culture, and screened adults of the next generation (G1) (Fig. 2A). Reproducibly, this simple difference in culturing method led to a dramatic shift in mouth-form ratio ($>95\%$ Eu on agar compared to ~10% Eu in liquid culture, $p < 0.001$, *paired t-test*) (Fig. 2B). Importantly, *P. pacificus* developed at similar rates in agar and liquid culture, allowing facile comparisons between conditions (Fig. 2C), and arguing against nutritional deprivation inducing the mouth-form shift. St animals have a slightly faster development than Eu animals when grown on agar[25], however we found developmental speed to be indistinguishable between morphs in liquid culture (Supplementary Fig. 1). The different environmental conditions present distinct energy requirements (eg. swimming and feeding on motile bacteria in 3-dimensional liquid culture) that might offset potential tradeoffs in resource allocation.

Next, we investigated whether the change in mouth-form ratio induced by liquid culture was capable of being inherited. The mouth-from ratio of adults was consistent with the culture method they developed in regardless of the culture method of the parental generation, suggesting the effect is not transgenerational (Fig. 2D). These results also demonstrate the immediate and robust nature of this plasticity, and similar experiments coupled to mutagenesis may be useful for identifying genes involved in the ability to sense and respond to changing environments.

Buffer components and culture state affect mouth form. To investigate the potential influence of culture conditions on mouth form we examined differences in buffer composition, and solid vs. liquid culturing state. In our previous experiments we had used standard liquid culture protocols for *C. elegans*[33], which utilize S media (S), whereas we normally grow *P. pacificus* on Nematode Growth Media (NGM) agar plates[33]. To assess the contribution of the chemical composition of the medium, as opposed to solid vs. liquid environments (hereafter referred to as 'culture state'), we performed reciprocal culture experiments. Nematodes that were grown on either S-agar or NGM-liquid exhibited intermediate mouth-form ratios ($51 \pm 5\%$ Eu and $38 \pm 13\%$ Eu, respectively, $p < 0.001$ relative to solid or liquid states of the same medium, *paired t-test*) (Table 1d,h,i,p), revealing a growth-medium composition effect. However, as these mouth-form ratios were in-between the extremes of NGM-agar and S-liquid, it also suggests other environmental factors are operating.

S medium contains phosphate (50 mM) and sulfate (14 mM) - both of which have previously been shown to affect mouth-form ratios at 120 mM[27]. To test whether this concentration of phosphate was causing the S-medium effect we made alternative formulations by replacing phosphate with 50 mM Tris ("T-Medium") or Hepes ("H-Medium"), pH 7.5. Liquid culture in T- and H-medium yielded reproducibly higher Eu ratios ($35 \pm 8\%$ and $28 \pm 10\%$, respectively, $p < 0.05$, *paired t-test*) (Table 1d–f), demonstrating a specific, albeit subtle contribution from phosphate. Furthermore, *P. pacificus* grown in axenic (without bacteria)[34], M9[33], or PBS (which does not contain sulfate) -based liquid cultures were all highly St (Table 1a–c). Although nematode survival rate was poor in PBS, and development was slowed in axenic culture (9–10 days for sexual maturation, rather than 3–4).

Rotation speed of liquid culture affects mouth form. Further exploration of liquid culture methodology revealed that decreasing the rotation per minute (rpm) also affected mouth-form ratios. Previous experiments that led to high St ratios had been performed at 180 rpm, but when shifted to "slow" speeds of 70 or 50 rpm, the mouth-form ratio shifted to an intermediate Eu bias ($55 \pm 11\%$ and $66 \pm 9\%$, respectively, $p < 0.05$, *t-test*) (Table 1j,l). The simplicity of changing rpm shaking-speed to affect mouth-form ratios is an intriguing environmental perturbation as other factors like food source, buffer, and culturing state are identical. When examined without bacteria, it became evident that at slow speeds (<90 rpm) nematodes aggregated in the center of the liquid column, whereas at higher speeds they were dispersed. When combined with conditions that exhibited intermediate St ratios the effects were additive, yielding up to $87 \pm 3\%$ Eu with NGM-liquid culture (Table 1k,m,n). The higher density of nematodes at slow speeds suggests that pheromones may be responsible. Consistent with this hypothesis, we passed multiple *P. pacificus* generations from one liquid culture to another,

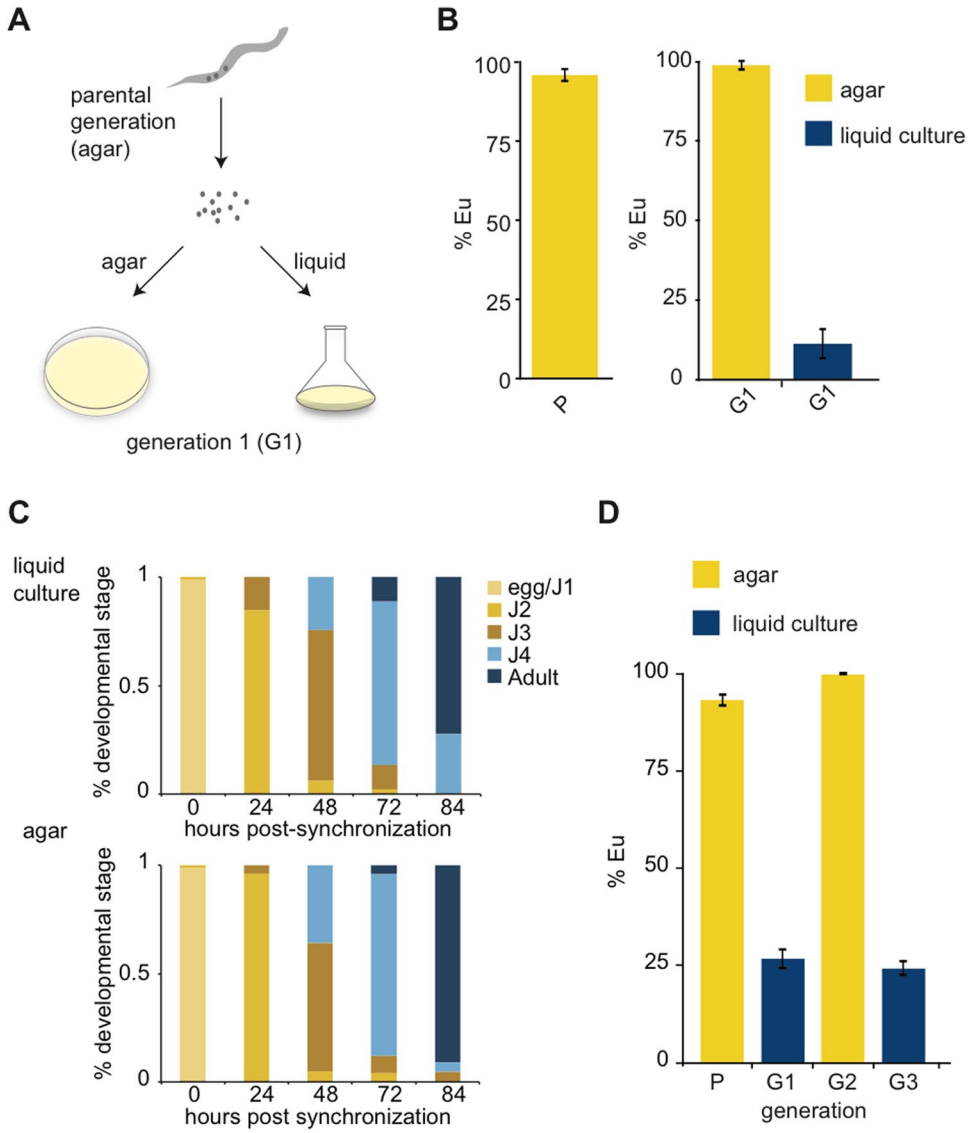


Figure 2. Different culture methods affect mouth-form phenotypic plasticity. (**A**) Diagram of experimental design to compare culture conditions from the same population after bleach synchronization. (**B**) Mouth-form ratios presented as percent eurystomatous (% Eu) from the parental generation (P) and the next generation (G1) grown in either liquid culture or agar plates, $n = 18$ biological replicates, $p < 0.05$, *students two-tailed t-test*, error bars represent SEM. (**C**) Developmental stages of bleach-synchronized *P. pacificus* in either agar plates or liquid culture. Bar graphs represent a typical experiment measuring >30 animals at the indicated time-points. (**D**) Mouth-form ratios of switching experiments between agar and liquid cultures. Nematodes were bleached between generations (P, G1, G2), and eggs-J1 larvae were passed to the next condition, $n = 3$, error bars represent SEM.

either by a 1:10 dilution, or by bleaching and washing. When passed by bleaching the next generation remained highly St ($8 \pm 4\%$). However when passed by dilution the next generation of worms exhibited intermediate Eu ratios ($51 \pm 16\%$, $p < 0.05$, *unpaired t-test*), perhaps because pheromones from the first generation were passed on to the second.

Liquid culture affects body morphology. We also observed morphological differences of body length and width between agar and liquid culture, demonstrating an additional plastic response (Supplementary Figure 2). Worms that develop in liquid culture exhibit longer, narrower bodies compared to worms that develop in agar, a phenomenon that has also been observed in *C. elegans*[33]. To disentangle whether the effect on mouth form is discrete or connected to the change in body shape we grew worms in NGG culture, which is intermediate between liquid and solid states[35]. Similar to liquid culture, adult worms grown in NGG exhibited a more slender body morphology than on agar plates ($p < 0.05$, Mann-Whitney), but they exhibited the highly Eu mouth-form ratio of worms grown in agar culture (Supplementary Fig. 2, Table 1d,o,p). While it is difficult to completely exclude the possibility that they are connected, there is no obvious correlation between the St mouth form and

	Condition	% Eu	S.E.M.
A	LC PBS, 180 rpm	7	3.2
B	LC Axenic Culture, 180 rpm	8.8	8.8
C	LC M9, 180 rpm	11.5	6.2
D	LC S-medium, 180 rpm	12.8	3.2
E	LC H-medium, 180 rpm	28	10.1
F	LC T-medium, 180 rpm	35	7.6
G	LC S-medium, 100 rpm	35.2	3
H	LC NGM, 180 rpm	37.8	12.9
I	AG S-medium	51.4	5.4
J	LC S-medium, 70 rpm	55.1	10.9
K	LC T-medium, 50 rpm	61.5	16.9
L	LC S-medium, 50 rpm	65.9	9
M	LC H-medium, 50 rpm	70.6	15
N	LC NGM, 50 rpm	87.3	3.3
O	NGG	97.1	2.5
P	AG NGM	98.7	0.7

Table 1. Buffer/ions and physical culture state affect mouth-form phenotype. A panel of culturing methods covers phenotypic ratios from ~10–99% Eu. LC = liquid culture, AG = agar, T and H medium = S-medium with phosphate replaced with 50 mM Tris or HEPES, pH 7.5, respectively, NGG = NGM with agar replaced with Gelrite/Gelzan CM (Sigma)[35]. $N \geq 3$ biological replicates per condition, and standard error mean (SEM) is presented in the last column. Mouth-form phenotypes were assessed 4–5 days after bleach-synchronization (see Methods).

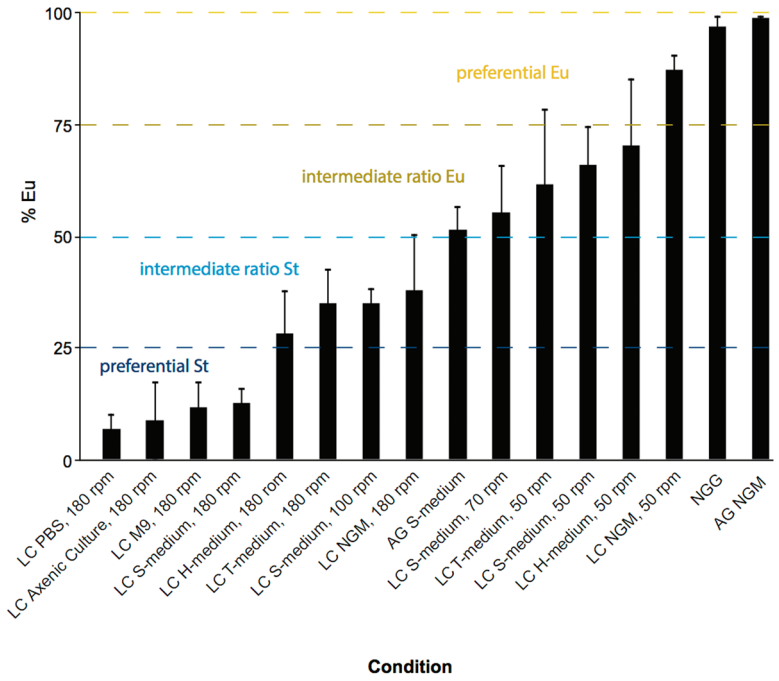


Figure 3. Comprehensive evaluation of culture method on mouth-form ratio in *P. pacificus*. Same data as in Table 1, but presented according to gradation of effect on mouth-form phenotype, from low to high % eurystomatous. LC = liquid culture, AG = agar, T and H medium = S-medium with phosphate replaced with 50 mM Tris or HEPES, pH 7.5, respectively, NGG = NGM with agar replaced with Gelrite/Gelzan CM (Sigma)[35]. Error bars represent standard error mean (SEM) for different biological replicates ($n \geq 3$, Methods).

slender morphology observed in liquid culture. Therefore, it seems these two instances of phenotypic plasticity are under distinct regulation.

Collectively, we have established a broad range of culturing methods that allow the acquisition of almost any mouth-form ratio from an isogenic strain (Fig. 3). A variety of liquid culture conditions, including buffers without phosphates or sulfates, exhibited an effect on mouth form, suggesting an unknown environmental effect that is perhaps specific to solid or liquid states.

Liquid culture acts upstream of known switch genes. Next, we sought to place the environmental effects of liquid culture relative to known genetic and environmental factors. First, we examined whether liquid culture had an effect on mutants that are 100% Eu on agar plates[27,28]. Animals from a *eud-1* overexpression line and *Ppa-nhr-40* mutant line remained 100% Eu in liquid culture, arguing that these genes act downstream of the environmental effect of liquid culture (Fig. 4A). Next, we assessed whether the dasc#1 pheromone was capable of inducing the Eu mouth-form in liquid culture, as it does on agar. dasc#1 experiments demonstrate a large variability in phenotypic ratio (Fig. 4B), however they typically exhibited a higher Eu proportion than control worms without dasc#1 treatment ($p = 0.068$, *paired t-test*). This intermediate and variable effect suggests that liquid culture and the dasc#1 pheromone act in parallel and antagonistically to each other. Finally, we also compared the expression of four genes in different culturing conditions that are up-or down-regulated in *eud-1* mutants (100% St) vs. wild-type (70–100% Eu)[27]. There was a strong correspondence between *eud-1* vs. wild-type RNA-seq data, and liquid vs. agar culture RT-qPCR (Fig. 4C,D). These results provide further evidence that the environmental effect of liquid culture is upstream of *eud-1*, and that this method is suitable for studying genetic pathways that have been determined through mutational experiments[27–29].

Liquid culture effect is dependent on genetic background. Finally, we explored whether there was a macro-evolutionary difference in responses to culture conditions. We chose four *Pristionchus* species that flank *P. pacificus* phylogenetically; three are highly Eu on agar (>95%), and one is highly St (>95%) (Fig. 5A,B). Remarkably, each species exhibited distinct phenotypic responses to liquid culture. For example, *P. maupasi* was highly Eu in both conditions, while *P. entomophagus* shifted to almost 100% St (Fig. 5C) in liquid culture. Meanwhile *P. mayeri* was St in both culture conditions. Taken together, these data show a genetic basis to environmental effects on phenotypic plasticity, which can be exploited for evolutionary, genetic, and molecular exploration of plasticity mechanisms. Whether these differences in response reflect adaptive changes to different environments, or are a result of drift remains to be seen in future investigations.

Discussion

We describe multiple methods for the culture of preferentially St (<25% Eu), intermediate St (25–50% Eu), intermediate Eu (50–75 Eu%), and preferentially Eu (>75% Eu) *P. pacificus* (Fig. 3, Methods). Growth rates are similar between conditions, allowing the generation of developmentally synchronized populations. The effects are immediate, and immediately reversible when switching between liquid and agar, suggesting they are not transgenerational. Importantly, the genetic pathways towards building each respective mouth form are consistent with pathways established from prior forward genetics[27,28]. Finally, the environmental response is unique in four species of *Pristionchus* tested, arguing that evolution has acted, passively or actively, on gene-environment interactions. The ability to toggle between mouth forms with simple culturing conditions provides powerful new tools to study the genetic and molecular mechanisms of phenotypic plasticity.

Perturbation of environmental factors such as salt concentration[15,36–38], pathogen[8,39–42], temperature[7,10,13,43–45], and diet[46,47] have been exploited for decades to study adaptive responses. More recent genome-wide profiling of epigenetic information carriers has revealed potential mechanisms for communicating stimuli to changes in gene expression. So called 'poised' or 'permissive' chromatin states can respond to external signals, leading to changes in transcription that ultimately affect tissue differentiation[48–55]. The time is now ripe to test whether similar processes affect phenotypic plasticity, a critical link between ecology and molecular mechanism that has just begun to be explored[56–60].

Our panel of *P. pacificus* culture conditions saturates the mouth-form frequency space (Fig. 3). The ability to shift ratios by rpm shaking-speed provides perhaps the cleanest method because of its simplicity. In shaking speeds greater than 90 rpm nematodes are dispersed, while below 90 rpm they are concentrated in the center of the liquid vortex. Since different buffer formulations also affected mouth-form ratios, and the combination with slow rpm yielded an additive effect, it seems that alterations in the abundance, diffusion, and local concentration of pheromones and ions (i.e. phosphate and sulfate) contribute to the observed differences between liquid and agar culture conditions. However, we note that densely packed nematodes at slow rpm (much denser than on a plate) in NGM-liquid media are still insufficient to recapitulate the >95% Eu phenotype seen on NGM-agar plates. While it remains possible that these are the only contributing factors, we speculate an additional unknown factor is extant related to bacterial density, metabolism, or the liquid environment itself.

Whether liquid culture is a direct stimulator of the St mouth form is currently unknown. Field observations and competition experiments are required to (1) assess if *Pristionchus* experiences wet-enough conditions in the wild to mimic liquid culture conditions as with other lotic, lentic or marine nematodes[61–63], and (2) determine whether the St mouth form provides an advantage in this environment. Both *C. elegans* and *P. pacificus* exhibit a slender morphology in liquid culture, suggesting a conserved plastic response to this environment. It is conceivable that a liquid culture-dependent signaling pathway related to mouth form also exists, although it could be mediated indirectly through other factors. Seemingly unrelated stimuli are capable of inducing the same developmental pathway by eventually descending on a downstream switch or "evocator"[64–66]. Regardless of the ultimate environmental factor, our analysis of gene expression in liquid culture reflects patterns observed in constitutive St mutants, suggesting that similar downstream pathways are utilized (Fig. 1D). Importantly however, we did not observe faster St development in liquid culture as has been observed on agar, and which is predicted to be the tradeoff advantage of the St morph[25]. It is formally possible that we did not have enough temporal resolution to identify the small but significant differences previously observed (55 hours for St and 61 hours for Eu). It is also worth noting that laboratory culture conditions are highly artificial, and it is perhaps not surprising that they could affect ecological strategies. Nevertheless, our results suggest that caution should be taken when studying *P. pacificus* ecology across different environments, as it may be context dependent. Going forward, it will be

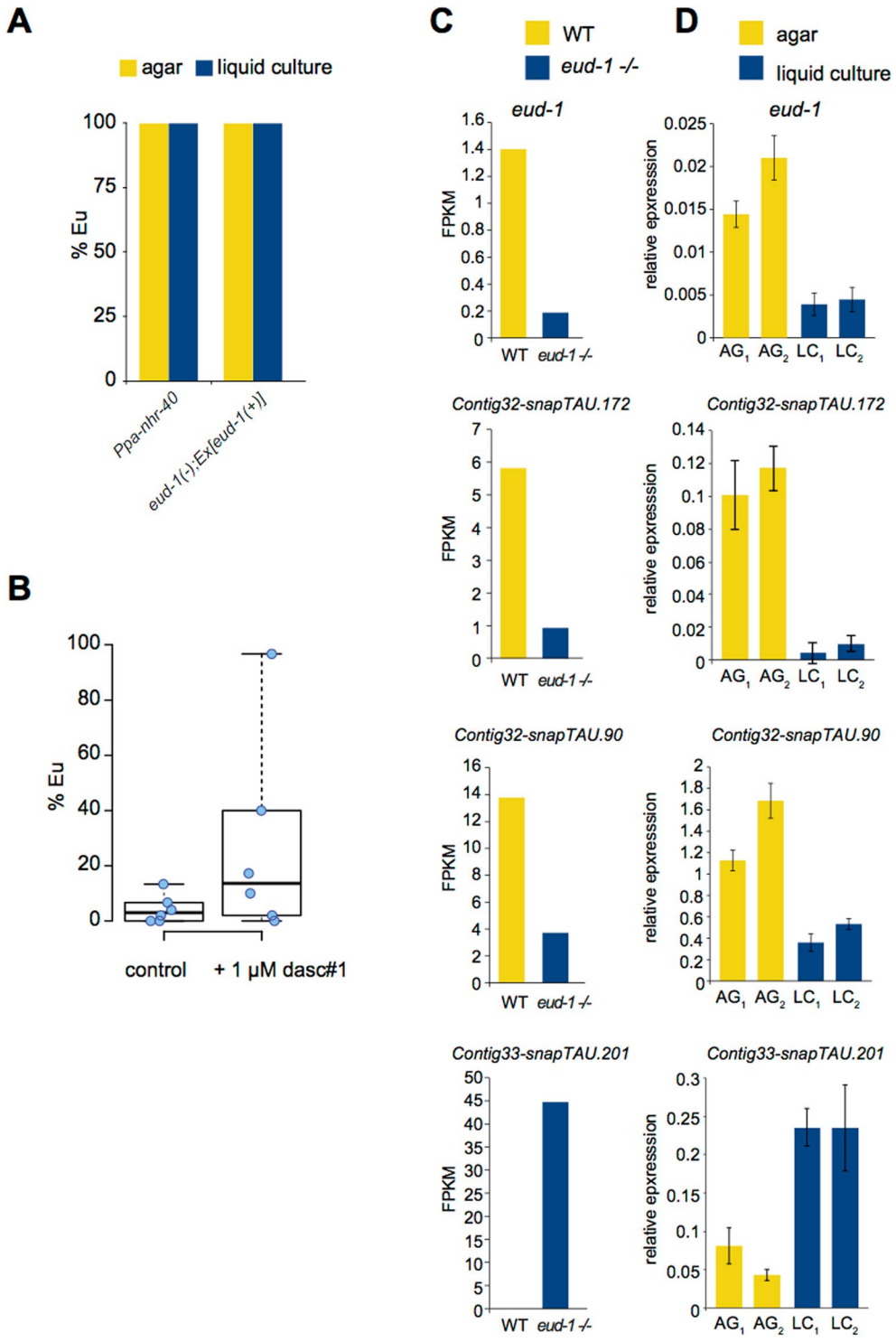


Figure 4. The environmental effect of liquid culture is upstream of known genetic components and induces similar pathways. (**A**) Mouth-form ratios of *eud-1* overexpression[27] and a *Ppa-nhr-40*[28] mutant in liquid culture reveals no effects, suggesting these genes are downstream, $n = 3$ biological replicates. (**B**) Addition of 1 µM dasc#1 exhibits a variable response that appears to induce Eu, although it is not statistically significant ($p = 0.068$). (**C**) Expression analysis of four genes by RNA-seq from *eud-1* mutants (the average of 4 homozygous mutant alleles is represented)[27] (100% St) compared to the RS2333 California strain (70–100% Eu), y-axis = fpkm (relative expression). (**D**) Reverse transcription-quantitative PCR (RT-qPCR) of *P. pacificus* PS312 grown in liquid culture/S-medium (LC) vs. NGM-agar plates (AG) for the two biological replicates displayed, with four technical replicates each. The y-axis represents $2^{\Delta Ct}$ (relative expression) compared to the housekeeping gene *Ppa-Y45F10D.4* (iron binding protein)[69], error bars represent standard deviation of n = 4 technical replicates.

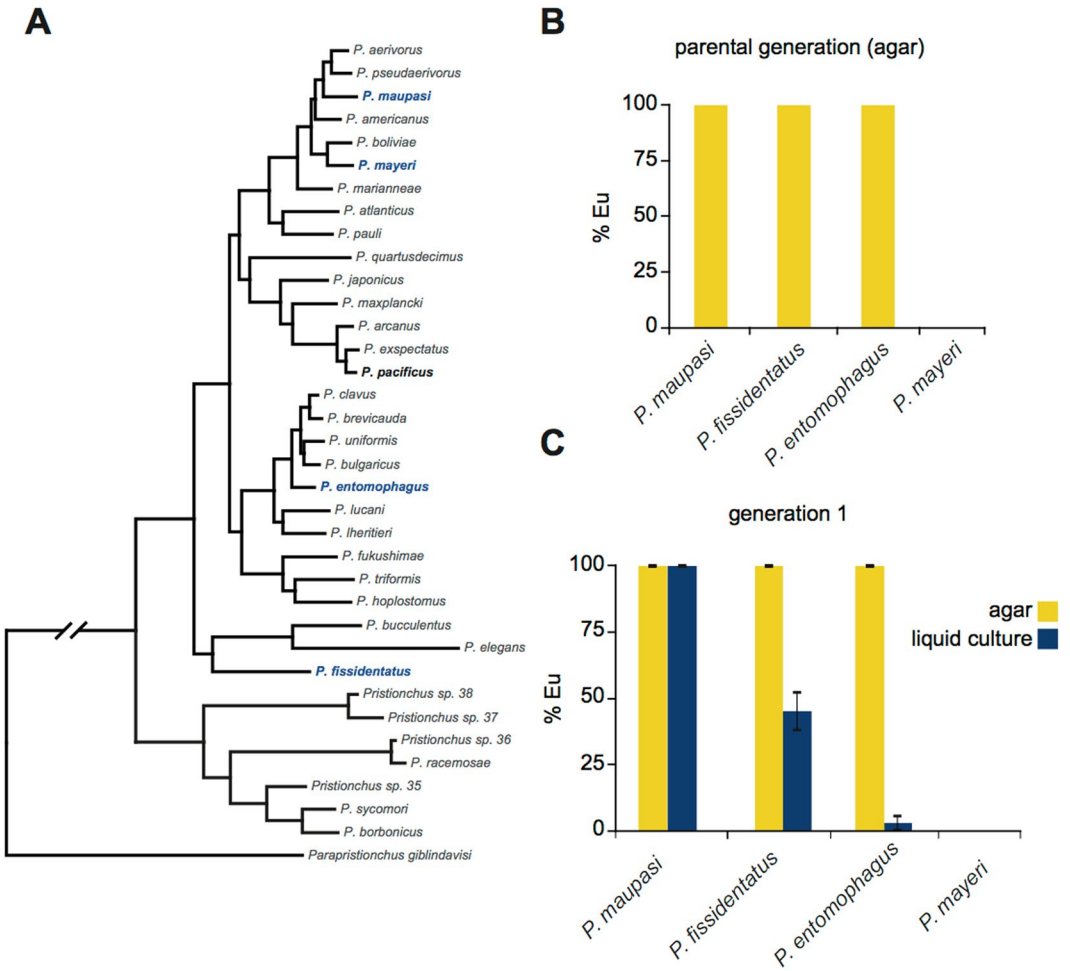


Figure 5. Macro-evolutionary view of liquid culture environmental influence. (**A**) Phylogeny of *Pristionchus* species[70] highlighting *P. pacificus* (bold), *P. fissidentatus*, *P. mayeri*, *and P. entomophagus* (blue). (**B**) Mouth-form ratio of parental generations ($n = 3$) of indicated species on NGM-agar after three consecutive healthy generations on OP50. (**C**) Mouth-form ratios of indicated species in either NGM-agar or liquid culture/S-medium ($n = 3$), error bars represent SEM.

informative to assess if the developmental speed of different species correlates with their response to liquid culture, and the aqueous content in which they are found in nature.

Which culture method is utilized will depend on the purpose of the experiment. Exploiting intermediate ratio conditions may be useful to study genes or other environmental factors predicated to effect mouth form but in an unknown direction (Eu or St). For experiments that require the greatest separation in mouth-form frequencies we recommend S-medium at 180 rpm (St) vs. NGM agar plates (Eu). We also frequently observed a modest degree of variation, which is expected for a stochastic phenotypic trait[67]. As such, every measurement utilizing these culturing methods should be performed side-by-side with control samples. It is our hope that these methods will be a contribution to the study of environmental effects on *P. pacificus* mouth form, and phenotypic plasticity in general.

Methods

Strains and species. For all *P. pacificus* experiments the California strain PS312 was used, except comparisons to RNA-seq data, which used a more grown-out version of the same strain (RS2333). For experiments with different species (Fig. 5) *P. maupasi*, *P. fissidenatus*, and *P. mayeri* were compared to *P. pacificus*. Epistasis experiments (Fig. 4A) were performed with *Ppa-nhr-40*(*tu505*) and *eud-1*(*tu445*);*tuEx[eud-1*(+)*]*.

Culture methods. Five young adult *Pristionchus* nematodes were passed every 4–6 days on 10 ml NGM-agar, 60 mm plates at 20 °C seeded with 300 µl of overnight cultures of *Escherichia coli* OP50 (grown in LB at 37 °C) and covered with parafilm to avoid experiencing starvation for three consecutive generations[33]. The mouth-form phenotype of 4th generation adults represents the parental (P) generation (Fig. 2A,C, and D). Prior to all subsequent phenotyping experiments adults were synchronized by washing off of plates with M9 using plastic Pasteur pipettes into 15 ml conical tubes, and adding 30% final volume NaOH/bleach (0.5 ml NaOH, 1 ml bleach/3.5 ml washed worms) for 9 minutes with gentle agitation every few minutes. Carcasses were filtered through a 120 µm

nylon net (Millipore) fixed between two rubber gaskets in a plastic funnel, washed by applying 3 ml M9 drop-wise on the filter, then pelleted 500 × g, 1 minute, room temperature. Eggs-J1 were washed by gentle re-suspension in 3 ml M9, and re-centrifuged 500 × g, 1 minute, room temperature. It is important not to wash worms with S-medium before or directly after bleach because it will start to precipitate. M9 wash was removed by pipette, and then eggs-J1s were ready for re-suspension in the appropriate buffer depending on the experiment.

For the majority of experiments, eggs-J1 larvae were re-suspended in 100 µl M9 × the number of test conditions (i.e. 200 µl for comparing one agar vs. one liquid culture condition). For re-culturing on agar, eggs-J1 were pipetted in the center of the OP50 lawn on 60 mm agar plates (NGM or S-medium), then the plate was tilted in 360° to spread and dry the eggs. Afterwards the plates were stored at 20° and adults were phenotyped 4–5 days later (see below for details of phenotyping). For culturing in liquid formats, 100 µl of eggs-J1 were pipetted into 10 ml of medium in 50 ml-volume autoclaved Erlenmeyer flasks. To prepare monoxenic liquid cultures the amount of OP50 *E. coli* was empirically determined. For all liquid cultures described (except axenic culture) 100 ml of overnight OP50 *E. coli* (grown in LB) to an optical density (OD_{600}) of 0.5, was pelleted 30 minutes, 4 °C at 3,000 × g in an SLA-3000 rotor and re-suspended in 10 ml filter-sterilized (0.22 µm, Millipore) S-medium[33] unless otherwise noted (e.g. M9 or PBS, Fig. 2). The concentration of bacteria is a critical parameter. The procedure described above led to healthy cultures of *P. pacificus* at the normal developmental rate observed on agar plates (3–4 days[21]), while adding less (50 ml or 10 ml) OP50 led to slower rates, or even the inability to develop beyond the J2 larval stage when significantly less was added. Liquid cultures were incubated 180 rpm, 20–22 °C unless otherwise noted for "slow" rpm experiments (50 and 70 rpm).

For experiments with "H" or "T" medium, S-medium was prepared as before[33] except that phosphates were replaced with 50 mM of HEPES or Tris, pH 7.5, respectively. Axenic culture was prepared according to Samuel *et al.*[34] with the exception that flavin-mononucleotide was replaced with riboflavin (Sigma) at the same amount, and cultures were shaken at 180 rpm instead of 70. As previously noted[34] with *C. elegans*, *P. pacificus* also develops slower in axenic culture, reaching maturity (adults) at 9–10 days after adding eggs. Culture in NGG was performed similar to Muschiol and Traunspurger 2007[35]. In short, 3 ml of NGM was prepared with agar replaced with Gelrite/Gelzan CM (Sigma) at 0.75 g/L and seeded with 300 µl of OP50 and bleached eggs, then incubated at 20 °C.

To collect nematodes from liquid cultures for tracking developmental stages or mouth-form phenotyping we developed a filtering method using removable 5 µm filters (Millipore) combined with the Sterifil aseptic system (47 mm, Millipore). Filters are applied to the Sterifil apparatus and a small amount of M9 is added and vacuumed through to ensure a tight and continuous seal. Then liquid cultures are decanted into the funnel and slowly vacuumed. All *P. pacificus* developmental stages are large enough to be blocked by the 5 µm filter, while bacteria pass through. However when attempting to isolate J2s we recommend applying 2 × 5 µm filters. After all liquid has passed through the filter, nematodes were washed with ~25 ml of M9 by decanting directly on to the filter and applying vacuum pressure. Then the funnel was removed, and forceps were used to transfer the filter to an open 50 ml conical tube in a curved shape to fit into the opening. Nematodes were then washed from the filter by repeatedly applying the same 1 ml of M9 over the filter. Then this 1 ml was transferred to 1.5 ml microcentrifuge tube, and incubated at room temperature for 5 minutes to allow adults to swim to the bottom. Adults were pelleted by a quick (2–3 seconds) centrifugation, and the supernatant was removed. If juveniles are desired, the tube, now free of bacteria after filtering, can also be centrifuged at max speed >5 minutes to pellet. Nematode pellets were then phenotyped, or flash-frozen in liquid N2 and stored −80 °C for subsequent processing (e.g. RT-qPCR).

Developmental rate determination. Worms were grown in liquid culture after bleach synchronization then filtered through a 20 µM filter 2 hours post bleach to isolate synchronous J2 animals, and then returned to liquid culture. Individual aliquots from the same flasks were monitored at regular intervals, and mouth-forms of adults were recorded at the J4-adult transition ($n = 2$). Flasks were rotated at 50 rpm to obtain large quantities of both St and Eu animals. Although not shown, several J4 were present at the earlier time points of 59 and 62 hours, which verified that we were observing the J4-adult transition.

Mouth-form phenotyping. For phenotyping nematodes grown on agar plates or NGG[35], adults were selected with a wire pick and transferred to 3–5 µl of M9 spotted on 4% agar pads (containing 10 mM sodium azide) on a standard microscope slide, then covered with a cover slip. For nematodes grown in liquid culture, after gently pelleting adults, they were re-suspended in the remaining M9 and 3–5 µl were directly pipetted onto the agar pad. When comparing mouth-forms of different conditions, we often performed 'blind' comparisons by writing the identity of the sample (i.e. "agar" or "liquid") on the slide, and then using laboratory tape to cover the identity, and blindly selecting a slide before placing it in the microscope holder. After counting, the identity of the sample was revealed by removing the tape. Phenotyping was performed at 40–100x/1.4 oil objective on a Differential Interference Contrast (DIC) microscope (Zeiss) according to buccal landmarks previously described[20]. In short, Eu were determined by the presence of a wide-mouth, a hooked dorsal tooth, and an additional subventral tooth. Conversely St animals were determined by a narrow-mouth, flint-like dorsal tooth, and absence of a subventral tooth (Fig. 1B,C). The number of biological replicates (n) was ≥3 for all conditions, and as high as 18 for liquid culture/S-medium, with each replicate including ≥50 animals with the exception that PBS and NGM-liquid cultures yielded significantly fewer animals, and included ≥20 animals per replicate. Mouth-forms were assessed 4–5 days after bleach-synchronization. Error bars represent standard error means (SEM), and statistical significance was assessed by *paired 2-tailed t-tests* unless otherwise indicated in the text.

dasc#1 experiments. dasc#1 was added at 1 µM final concentration according to previous methods[26] to eggs-J1 larvae in liquid culture. Mouth-forms were phenotyped as described above after 4 days and compared to

control liquid cultures without dasc#1. The p-value was determined by a 1-tailed, paired *t-test* for $n = 6$ biological replicates.

Morphology measurements. Length and width measurements were performed on synchronized adult animals four days after bleaching. Measurements were made of 12 animals grown on agar, 13 grown on NGG, and 10 in liquid culture using the ImageJ plug-in WormSizer[68]. Box plots in Supplementary Figure 2 show quartile edges (25% and 75%) of the distribution and medians (black bars), made in R {boxplot(shape~Condition, data = worm_sizes, horizontal = TRUE, notch = FALSE)}.

Expression analysis. RNA-seq data was obtained from Ragsdale, Müller *et al.*[27], and average fpkms from 4 mutant alleles of *eud-1* vs. one wild-type California RS2333 were plotted. For RT-qPCR, RNA was first extracted from either 1 agar plate or 1 liquid culture of synchronized young adults (4 days post-bleaching) of the California strain PS312 (same as RS2333 but an earlier frozen stock) by Trizole extraction followed by purification with Zymo RNA-Clean & Concentrator-25 columns following manufacturers instructions from Zymo. 500–1,000 ng of purified RNA was converted to cDNA using SuperScript II (Invitrogen) for 1 hour with Oligo$(dT)_{18}$ primer in 20 µl reactions, and then heat-inactivated with 40 µl of 150 mM KOH/20 mM Tris-base for 10 minutes at 99 °C followed by 40 µl of 150 mM HCl, and 100 µl of TE. 4 µl of cDNA was used for each technical replicate in 10 µl qPCR reactions with 1x LightCycler® 480 SYBR Green I Master Mix (Roche) and 0.25 µM of each primer on a Light-Cycler 480, 384 well format. All primer sets were validated for single amplicon production with Tm melt-curve analysis, and efficiency with a 5-log titration of cDNA. Relative expression ($2^{\Delta Ct}$) was measured relative to *Ppa-Y45F10D.4* (iron binding protein)[69] for each gene.

Data availability. All data generated or analyzed during this study are included in this article and its Supplementary Information files.

References

1. West-Eberhard, M. J. Developmental Plasticity and Evolution. (2003).
2. Bradshaw, A. D. Evolutionary Significance of Phenotypic Plasticity in Plants. *Adv. in Genetics* **13**, 115–155 (1965).
3. Gauze, G. F. *Problems of evolution*. **37**, part 2 (1947).
4. Huxley, J. Problems of relative growth. (1932).
5. Weaver, N. Rearing Honeybee Larvae on Royal Jelly in the Laboratory. *Bee World* (1955).
6. Charnov, E. L. & Bull, J. When is sex environmentally determined? *Nature* **266**, 828–830 (1977).
7. Ferguson, M. W. & Joanen, T. Temperature of egg incubation determines sex in Alligator mississippiensis. *Nature* **296**, 850–853 (1982).
8. Palacios, M. G., Sparkman, A. M. & Bronikowski, A. M. Developmental plasticity of immune defence in two life-history ecotypes of the garter snake, Thamnophis elegans - a common-environment experiment. *J Anim Ecol* **80**, 431–437 (2011).
9. Waddington, C. H. Genetic Assimilation of an Acquired Character. *Evolution* **7**, 118 (1953).
10. Gibert, J.-M., Peronnet, F. & Schlötterer, C. Phenotypic Plasticity in Drosophila Pigmentation Caused by Temperature Sensitivity of a Chromatin Regulator Network. *PLoS Genet* **3**, e30 (2007).
11. Golden, J. W. & Riddle, D. L. The Caenorhabditis elegans dauer larva: Developmental effects of pheromone, food, and temperature. *Developmental Biology* **102**, 368–378 (1984).
12. Sikkink, K. L., Reynolds, R. M. & Ituarte, C. M. Rapid evolution of phenotypic plasticity and shifting thresholds of genetic assimilation in the nematode Caenorhabditis remanei. *G3:* (2014).
13. Powsner, L. The effects of temperature on the durations of the developmental stages of Drosophila melanogaster. *Physiological Zoology* (1935).
14. Brakefield, P. M., Gates, J., Keys, D. & Kesbeke, F. Development, plasticity and evolution of butterfly eyespot patterns. *Nature* (1996).
15. Fielenbach, N. & Antebi, A. C. elegans dauer formation and the molecular basis of plasticity. *Genes & Development* **22**, 2149–2165 (2008).
16. Riddle, D. L., Swanson, M. M. & Albert, P. S. Interacting genes in nematode dauer larva formation. *Nature* **290**, 668–671 (1981).
17. Albert, P. S. & Riddle, D. L. Mutants of Caenorhabditis elegans that form dauer-like larvae. *Developmental Biology* **126**, 270–293 (1988).
18. Gottlieb, S. & Ruvkun, G. daf-2, daf-16 and daf-23: genetically interacting genes controlling Dauer formation in Caenorhabditis elegans. *Genetics* **137**, 107–120 (1994).
19. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. **75**, 843–854 (1993).
20. Bento, G., Ogawa, A. & Sommer, R. J. Co-option of the hormone-signalling module dafachronic acid–DAF-12 in nematode evolution. *Nature* **466**, 494–497 (2010).
21. Sommer, R. J. & McGaughran, A. The nematode Pristionchus pacificus as a model system for integrative studies in evolutionary biology. *Molecular Ecology* **22**, 2380–2393 (2013).
22. Meyer, J. M. *et al*. Succession and dynamics of Pristionchus nematodes and their microbiome during decomposition of Oryctes borbonicus on La Réunion Island. *Environmental Microbiology* **19**, 1476–1489 (2017).
23. Serobyan, V., Ragsdale, E. J. & Sommer, R. J. Adaptive value of a predatory mouth-form in a dimorphic nematode. *Proceedings of the Royal Society of London B: Biological Sciences* **281**, 20141334–989 (2014).
24. Wilecki, M., Lightfoot, J. W., Susoy, V. & Sommer, R. J. Predatory feeding behaviour in Pristionchus nematodes is dependent on phenotypic plasticity and induced by serotonin. *J. Exp. Biol.* **218**, 1306–1313 (2015).
25. Serobyan, V., Ragsdale, E. J., Müller, M. R. & Sommer, R. J. Feeding plasticity in the nematode Pristionchus pacificus is influenced by sex and social context and is linked to developmental speed. *Evolution & Development* **15**, 161–170 (2013).
26. Bose, N. *et al*. Complex Small-Molecule Architectures Regulate Phenotypic Plasticity in a Nematode. *Angewandte Chemie International Edition* **51**, 12438–12443 (2012).
27. Ragsdale, E. J., Müller, M. R., Rödelsperger, C. & Sommer, R. J. A Developmental Switch Coupled to the Evolution of Plasticity Acts through a Sulfatase. **155**, 922–933 (2013).
28. Kieninger, M. R. *et al*. The Nuclear Hormone Receptor NHR-40 Acts Downstream of the Sulfatase EUD-1 as Part of a Developmental Plasticity Switch in Pristionchus. *Curr. Biol.* (2016).
29. Serobyan, V. *et al*. Chromatin remodelling and antisense-mediated up-regulation of the developmental switch gene eud-1 control predatory feeding plasticity. *Nat Commun* **7**, 12337 (2016).
30. Mather, K. & De Winton, D. Adaptation and counter-adaptation of the breeding system in Primula. *Annals of Botany* (1941).
31. Sommer, R. J. *et al*. The genetics of phenotypic plasticity in nematode feeding structures. *Open Biology* **7**, 160332–118 (2017).

32. Serobyan, V. & Sommer, R. J. Developmental systems of plasticity and trans-generational epigenetic inheritance in nematodes. *Current Opinion in Genetics & Development* **45**, 51–57 (2017).
33. Stiernagle, T. *Maintenance of C. elegans*, *WormBook*, *ed. The C. elegans Research Community*. (WormBook, 2006).
34. Samuel, T. K., Sinclair, J. W., Pinter, K. L. & Hamza, I. Culturing Caenorhabditis elegans in Axenic Liquid Media and Creation of Transgenic Worms by Microparticle Bombardment. *Journal of Visualized Experiments: JoVE* e51796 (2014).
35. Muschiol, D. & Traunspurger, W. Life cycle and calculation of the intrinsic rate of natural increase of two bacterivorous nematodes from chemoautotrophic Movile Cave, Romania. *Nematology* **9**, 271–284 (2007).
36. Barberon, M. *et al*. Adaptation of Root Function by Nutrient-Induced Plasticity of Endodermal Differentiation. *Cell* **164**, 447–459 (2016).
37. Abbruzzese, G. *et al*. Leaf morphological plasticity and stomatal conductance in three Populus alba L. genotypes subjected to salt stress. *Environmental and Experimental Botany* **66**, 381–388 (2009).
38. Wang, Y. *et al*. Salt-induced plasticity of root hair development is caused by ion disequilibrium in Arabidopsis thaliana. *J Plant Res* **121**, 87–96 (2008).
39. Touchon, J. C. T. C., Gomez-Mestre, I. G.-M. & Warkentin, K. M. W. M. Hatching plasticity in two temperate anurans: responses to a pathogen and predation cues. *Canadian Journal of Zoology* **84**, 556–563 (2006).
40. Hong, J. K. & Hwang, B. K. Induction by pathogen, salt and drought of a basic class II chitinase mRNA and its *in situ* localization in pepper (Capsicum annuum). *Physiologia Plantarum* **114**, 549–558 (2002).
41. Pulendran, B., Palucka, K. & Banchereau, J. Sensing Pathogens and Tuning Immune Responses. *Science* **293**, 253–256 (2001).
42. Huang, Q. *et al*. The Plasticity of Dendritic Cell Responses to Pathogens and Their Components. *Science* **294**, 870–875 (2001).
43. Plunkett, C. R. Temperature as a tool of research in phenogenetics: methods and results (1932).
44. Woodward, D. E. & Murray, J. D. On the Effect of Temperature-Dependent Sex Determination on Sex Ratio and Survivorship in Crocodilians. *Proceedings of the Royal Society of London B: Biological Sciences* **252**, 149–155 (1993).
45. Manenti, T., Loeschcke, V., Moghadam, N. N. & Sørensen, J. G. Phenotypic plasticity is not affected by experimental evolution in constant, predictable or unpredictable fluctuating thermal environments. *Journal of Evolutionary Biology* **28**, 2078–2087 (2015).
46. Brzek, P., Kohl, K., Caviedes-Vidal, E. & Karasov, W. H. Developmental adjustments of house sparrow (Passer domesticus) nestlings to diet composition. *J. Exp. Biol.* **212**, 1284–1293 (2009).
47. Watson, E., MacNeil, L. T., Arda, H. E., Zhu, L. J. & Walhout, A. J. M. Integration of Metabolic and Gene Regulatory Networks Modulates the C. elegans Dietary Response. *Cell* **153**, 253–266 (2013).
48. Bernstein, B. E. *et al*. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic. *Stem Cells.* **125**, 315–326 (2006).
49. Rougvie, A. E. & Lis, J. T. Postinitiation transcriptional control in Drosophila melanogaster. *Mol. Cell. Biol.* **10**, 6041–6045 (1990).
50. Rada-Iglesias, A. *et al*. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
51. Zentner, G. E., Tesar, P. J. & Scacheri, P. C. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* **21**, 1273–1283 (2011).
52. Maxwell, C. S. *et al*. Pol II Docking and Pausing at Growth and Stress Genes in C. elegans. *Cell Rep* **6**, 455–466 (2014).
53. Gaertner, B. *et al*. Poised RNA Polymerase II Changes over Developmental Time and Prepares Genes for Future Expression. *Cell Rep* **2**, 1670–1683 (2012).
54. Ernst, J. *et al*. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
55. Hsu, H. T. *et al*. Recruitment of RNA polymerase II by the pioneer transcription factor PHA-4. *Science* **384**, 1372–1376 (2015).
56. Ozawa, T. *et al*. Histone deacetylases control module-specific phenotypic plasticity in beetle weapons. *Proc. Natl. Acad. Sci. USA* **113**, 15042–15047 (2016).
57. Kucharski, R., Maleszka, J., Foret, S. & Maleszka, R. Nutritional Control of Reproductive Status in Honeybees via DNA Methylation. *Science* **319**, 1827–1830 (2008).
58. Simola, D. F. *et al*. Epigenetic (re)programming of caste-specific behavior in the ant Camponotus floridanus. *Science* **351**, aac6633–aac6633 (2016).
59. Kooke, R. *et al*. Epigenetic basis of morphological variation and phenotypic plasticity in Arabidopsis thaliana. *Plant Cell* **27**, 337–348 (2015).
60. Zhang, Y. Y., Fischer, M., Colot, V. & Bossdorf, O. Epigenetic variation creates potential for evolution of plant phenotypic plasticity. *New Phytologist* **197**, 314–322 (2013).
61. Traunspurger, W. The biology and ecology of lotic nematodes. *Freshwater Biology* **44**, 29–45 (2000).
62. Tietjen, J. H. & Lee, J. J. Life history and feeding habits of the marine nematode, Chromadora macrolaimoides steiner. *Oecologia* **12**, 303–314 (1973).
63. Herman, P. M. J. & Vranken, G. Studies of the life-history and energetics of marine and brackish-water nematodes. *Oecologia* **77**, 457–463 (1988).
64. Waddington, C. H. *Organisers and Genes by C. H. Waddington*. (The University Press, 1940).
65. Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* (1942).
66. Baeuerle, P. A. & Baltimore, D. IkappaB: a specific inhibitor of the NF-kappaB transcription factor. *Science* (1988).
67. Susoy, V. & Sommer, R. J. Stochastic and Conditional Regulation of Nematode Mouth-Form Dimorphisms. *Front. Ecol. Evol.* **4**, 6706 (2016).
68. Moore, B. T., Jordan, J. M. & Baugh, L. R. WormSizer: high-throughput analysis of nematode size and shape. *PLoS ONE* **8**, e57142 (2013).
69. Schuster, L. N. & Sommer, R. J. Expressional and functional variation of horizontally acquired cellulases in the nematode Pristionchus pacificus. *Gene* **506**, 274–282 (2012).
70. Susoy, V. *et al*. Large-scale diversification without genetic isolation in nematode symbionts of figs. *Science Advances* **2**, e1501031–e1501031 (2016).

Acknowledgements

We are thankful to current members of the Sommer laboratory for thoughtful critique of experiments, results, and interpretations, and Dr. Matthias Herrmann for assistance with the photo in Figure 1A. This study was funded by the Max Planck Society.

Author Contributions

M.S.W., B.S. and R.J.S. conceived and designed the experiments. M.S.W. and T.L. performed mouth-form experiments with help from B.S. in making axenic culture, RNA-seq analysis (also with assistance from S.N.), and rpm experiments. M.S.W., B.S., S.N., M.L., M.D., T.R. and D.R.S. all contributed to RT-qPCR experiments. M.S.W. wrote the manuscript with edits and assistance from R.J.S., and with contribution and approval from all other authors.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-07455-7

Competing Interests: The authors declare that they have no competing interests.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017