

**High-resolution single-worm transcriptomics and the
function of oscillating genes in mouth-form development
in *Pristionchus pacificus***

Dissertation

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

vorgelegt von
Shuai Sun
aus Shan Dong, China

Tübingen
2022

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

Tag der mündlichen Qualifikation:

21.11.2022

Dekan:

Prof. Dr. Thilo Stehle

1. Berichterstatter/-in:

Prof. Dr. Ralf J. Sommer

2. Berichterstatter/-in:

Prof. Dr. Gerd Jürgens

Table of Contents

I Acknowledgements	6
II Summary	7
III Zusammenfassung	8
IV Abbreviations	10
V List of publications	11
1 Introduction	12
1.1 The nematode model system <i>Pristionchus pacificus</i>	12
1.1.1 Model system for studying phenotypic plasticity	12
1.1.2 Mouth-form dimorphism in <i>P. pacificus</i>	13
1.1.3 Molecular basis of mouth-form plasticity in <i>P. pacificus</i>	15
1.2 The cuticular extracellular matrix of nematode	16
1.2.1 Collagenous cuticle.....	17
1.2.1.1 The collagens and their mutations in nematode.....	18
1.2.1.2 The roles of astacin metalloproteases in collagen formation.....	18
1.2.2 Chitinous cuticle	19
1.2.2.1 Chitin formation and catalytic mechanism of chitin synthase	19
1.2.2.2 The roles of chitin synthase in the development of worms.....	21
1.2.2.3 Inhibitors affecting chitin synthesis	22
1.2.3 Remodeling of cuticular structure during molts in nematodes	22
1.3 High-resolution temporal of transcriptome sequencing	23
1.3.1 Development of transcriptome sequencing.....	24
1.3.2 Adaptation of low input RNAs-seq method for nematodes	25
1.3.3 Gene expression oscillations	26
2 Aims of this thesis	27
3 Results	28
3.1 Single worm transcriptomics identifies a developmental core network of oscillating genes with deep conservation across nematodes	28
3.2 The oscillating Mucin-type protein DPY-6 has a conserved role in nematode mouth and cuticle formation.....	29

3.3 Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell	30
3.4 Improving transgenesis efficiency and CRISPR-associated tools through codon optimization and native intron addition in <i>Pristionchus</i> nematodes	31
3.5 Crowdsourcing and the feasibility of manual gene annotation: A pilot study in the nematode <i>Pristionchus pacificus</i>	32
3.6 Teethless chitin synthase-2 mutants in a predatory nematode reveal a role of chitin for the formation of complex feeding structures.....	33
3.7 Protocol for single-worm transcriptomics in <i>Pristionchus pacificus</i> , step by step.....	34
4 Discussion.....	35
4.1 Future use of the single-worm transcriptome sequencing	35
4.2 Developmental oscillations in <i>P. pacificus</i>	36
4.3 The oscillating <i>dpy-6</i> is involved in nematode cuticle and mouth formation ...	37
4.4 The role of chitin synthase 2 in the mouth-form plasticity	38
5 Bibliography	40
6 Appendix.....	54

I Acknowledgements

I would like to thank my advisor, Prof. Dr. Ralf Sommer for giving me the opportunity to pursue various projects as well as for his expertise, support, encouragement, and freedom he gave me. Next, I would like to thank all members of my Thesis Advisory Committee (TAC), Prof. Dr. Gerd Juergens, Dr. Frank Chan, and Dr. Oliver Weichenrieder for their insightful scientific advice.

I would like to thank the members of my examination committee: Prof. Dr. Thilo Stehle and Prof. Dr. Gerd Jürgens for evaluating this dissertation.

I would like to thank present and past members of the Sommer lab. In particular, I want to thank Dr. Bogdan Sieriebriennikov for providing the first hand-on training in the beginning of my Ph.D. I want to thank Dr. Christian Rödelsperger, Dr. Wen-sui Lo and Dr. Michael Werner for their bioinformatics support, and Dr. Kohta Yoshida, Dr. Catia Igreja, Dr. Adrian Streit, Dr. Mohannad Dardiry, Tobias Theska, and Devansh Sharma for giving their constructive criticism. I'd like to thank Prof. Dr. Ralf Sommer, Dr. Christian Rödelsperger and Dr. Ata Kalirad for providing corrections and modifications for this thesis. Thanks to all the technicians, especially Hanh Witte and Waltraud Röseler, who made an important contribution to this thesis. I am also grateful to all the current and past members of lab 17, namely Dr. Eduardo Moreno, Dr. Bogdan Sieriebriennikov, Dr. Nermin Akduman, Dr. Mohannad Dardiry, Sara Wighard, Veysi Piskobulu, Penghieng Theam, Hanh Witte, Metta Riebesell, and Sandra Mäck for creating a fun and productive working environment.

Cordial thanks go to my parents, Yuxian Sun and Xiaoling Ma, and my sisters, Xue Sun and Xiao Sun, who tolerated my absence during these years and gave me their love and support. Finally, I am appreciative of the Max Planck Society and the China Scholarship Council for their generous support.

II Summary

Development is largely under the control of genes. Specific spatiotemporal gene expression can modify development and the final phenotype of an organism. To obtain a high-resolution catalog of the developmental transcriptome in *Pristionchus pacificus*, I developed and implemented a single worm transcriptomic approach for the nematode model organism *P. pacificus*, and performed temporal transcriptome analysis over the entire postembryonic development with 38 time points. I focused on investigating oscillating gene expression patterns and found that i) nearly 3000 oscillating genes are periodically expressed during postembryonic development, ii) there is an overrepresentation of ancient gene classes among oscillatory genes, and iii) the developmental switch gene *eud-1* mediates numerous oscillatory genes including collagens, indicating the potential roles of these oscillating collagens in regulating mouth-form plasticity.

Mouth-form dimorphism in *P. pacificus* provides an ideal example to study the mechanisms of phenotype plasticity. Many previous studies focused on the regulation of mouth-form development, and identified environmental influences, developmental switches, the gene regulatory network involved in mouth-form plasticity and the associated evolutionary processes. However, the molecular and structural basis of the teeth in *P. pacificus* remains poorly understood. To address this fundamental question, I used two complementary approaches. First, I performed a large-scale genetic screen and obtained six mutants displaying morphological changes in both stomatal structures and body shape. Using whole genome sequencing and genome editing (CRISPR/Cas9 system) technologies, I identified *Ppa-dpy-6*, which encodes a mucin-type protein, as the first structural component of the nematode stoma involving in the specification of the cheilostom and cuticle. Second, I investigated the function of two chitin synthase genes (*chs*) in *P. pacificus*. Phylogenetic analysis revealed that two chitin synthase genes are highly conserved across nematodes. Mutations in the C-terminus of *chs-2* in *P. pacificus* result in a viable but teethless phenotype. Moreover, animals with this teethless phenotype were observed after injection of the chitin-synthase inhibitor Nikkomycin Z. These results suggest that the conserved *Ppa-chs-2* is essential for *P. pacificus* teeth formation. In addition, such teethless mutants can feed on various bacterial food sources, yet they are incapable of predation.

III Zusammenfassung

Die Entwicklung von der befruchteten Eizelle zu einem multizellulären Organismus besteht aus Zellteilungs- und Differenzierungsprozessen, die sich auf der Ebene der Genexpression widerspiegeln. Diese Prozesse können plastisch auf Umwelteinflüsse reagieren. So können im Fadenwurm *Pristionchus pacificus* externe Faktoren wie Pheromone, Futter und Temperatur unterschiedliche Entwicklungsprogramme aktivieren, die sich in zwei verschiedenen Mundformen manifestieren. Während bisherige Studien sich auf das genregulatorische Netzwerk um das Schlüsselgen *eud-1* konzentriert haben, sind bisher noch keine strukturellen Komponenten des Mundformdimorphismus bekannt. Um solche Komponenten zu identifizieren, habe ich zunächst eine genaue Charakterisierung der Entwicklungsprozesse von *P. pacificus* auf transkriptioneller Ebene durchgeführt. Zu diesem Ziel habe ich ein Protokoll für die Erstellung von Genexpressionsprofilen von einzelnen Würmern etabliert. Mit dieser Methode habe ich die Expression aller Gene zu 38 Zeitpunkten der postembryonalen Entwicklung gemessen. Dies hat zu der Identifizierung von ungefähr 3000 oszillierenden Genen geführt. Viele dieser Gene sind hochkonserviert und werden durch *eud-1* reguliert. Diesen hochauflösenden Katalog der Genexpression habe ich mit Mutagenese-Experimenten komplementiert. Dadurch isolierte ich sechs Mutanten mit gleichzeitigen Defekten in der Mundöffnung, als auch in der allgemeinen Körperform. Mit Hilfe von Genomsequenzierung und -editierung (CRISPR/Cas9 System) habe ich das Muzin-ähnliche Gen *Ppa-dpy-6* als erste strukturelle Komponente der Mundform in Nematoden identifiziert. Darüber hinaus habe ich die Funktion zweier hochkonservierter Chitin-Synthetasen in *P. pacificus* untersucht. Würmer mit Mutation am C-terminalen Ende von *Ppa-chs-2* sind lebensfähig, aber zeigen einen „Zahnlos“ Phänotyp. Diese Mutanten können sich von Bakterien ernähren, sind aber unfähig andere Würmer durch Bisse zu verletzen. Der „Zahnlos“ Phänotyp konnte auch durch Injektion des Chitin-Synthetase Inhibitors Nikkomyzin Z bestätigt werden. Diese Ergebnisse zeigen, dass *Ppa-chs-2* essentiell für die Bildung der zahnähnlichen Strukturen in *P. pacificus* ist und erbringt damit den Nachweis, dass Chitin an der Ausbildung der Mundstrukturen in Fadenwürmern benötigt wird. Damit beleuchtet meine Arbeit die Entstehung der Mundform in *P.*

pacificus auf transkriptioneller und molekularer Ebene und hat die ersten strukturellen Komponenten des Mundformdimorphismus identifiziert.

IV Abbreviations

BSA	B ulk s egregant a nalysis
<i>C. elegans</i> (<i>C.el.</i>)	<i>Caenorhabditis elegans</i>
CRISPR-Cas-9	C lustered r egularly interspaced s hort p alindromic r epeats C RISPR a ssociated protein 9
cDNA	C omplementary d eoxyribonucleic a cid
DNA	D eoxyribonucleic a cid
EMS	E thyl m ethanesulfonate
Eu	E urystomatous mouth form
FPKM	F ragments p er kilobase m illion
GRN	G ene regulatory n etwork
mRNA	M essenger ribonucleic a cid
NHR	N uclear h ormone r eceptor
NZ	N ikkomycin Z
<i>P. pacificus</i> (<i>P. pa.</i>)	<i>Pristionchus pacificus</i>
PCR	P olymerase c hain r eaction
qPCR	Q uantitative p olymerase c hain r eaction
RACE	R apid a mplification of c DNA- e nds
RNA	R ibonucleic a cid
RNAi	R ibonucleic a cid i nterference
TI	T eethless mouth form
TPM	T ranscripts p er kilobase m illion
St	S tenoumatous mouth form
STD-seq	S tandard transcriptomic s equencing
SWT-seq	S ingle-worm transcriptomic s equencing

V List of publications

- 1) **Sun, S.**, Hanh W. and Sommer, R.J. Chitin contributes to the formation of a feeding structure in a predatory nematode. ***Current Biology*** 33, 1-13.
- 2) **Sun, S.**, Rödelsperger, C. and Sommer, R.J. Single worm transcriptomics identifies a developmental core network of oscillating genes with deep conservation across nematodes. ***Genome Research***, 31(9), pp.1590-1601.
- 3) **Sun, S.** †, Tobias T. †, Hanh W., Erik R.J. and Sommer, R.J. The oscillating Mucin-type protein DPY-6 has a conserved role in nematode mouth and cuticle formation. ***Genetics***, iyab233.
- 4) Sieriebriennikov, B., **Sun, S.**, Lightfoot, J.W., Witte, H., Moreno, E., Rödelsperger, C. and Sommer, R.J. Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell. ***PLoS Genetics***, 16(4), p.e1008687.
- 5) Han, Z. †, Lo, W.S. †, Lightfoot, J.W., Witte, H., **Sun, S.** and Sommer, R.J. Improving transgenesis efficiency and CRISPR-associated tools through codon optimization and native intron addition in *Pristionchus* nematodes. ***Genetics***, 216(4), pp.947-956.
- 6) Rödelsperger, C., Athanasouli, M., Lenuzzi, M., Theska, T., **Sun, S.**, Dardiry, M., Wighard, S., Hu, W., Sharma, D.R. and Han, Z. Crowdsourcing and the feasibility of manual gene annotation: A pilot study in the nematode *Pristionchus pacificus*. ***Scientific Reports***, 9(1), pp.1-9.
- 7) **Sun, S.**, and Sommer, R.J. Single-worm transcriptomics protocol in *P. pacificus*, step by step. ***Ready for submission***.

1 Introduction

1.1 The nematode model system *Pristionchus pacificus*

Pristionchus pacificus is a well-established model system for comparative and evolutionary studies (Sommer et al. 1996; Sommer and Streit 2011; Sommer and McGaughran 2013). The primary advantages of *P. pacificus* include, i) the rapid growth, with a four-day generation time, and a large brood size, making it easy to culture and maintain in the laboratory, ii) being a self-fertilizing hermaphrodite that usually propagates by selfing, with the rare occurrence of males, allowing for genetic crosses that enables complex genetic analysis, iii) a transparent body, which accommodates direct observation under the microscope. Unlike the majority of free-living nematodes, the life cycle of *P. pacificus* and other nematodes of the same family called Diplogastridae, includes only three free-living juvenile stages, separated by molts. Specially, *P. pacificus* hatches into the J2 juvenile stage after the first molt in the egg shell (Fürst von Lieven 2005). Thus, there are only three free-living larval stages, the J2, J3, and J4 stage that correspond to the *C. elegans* L2, L3 and L4 stages, respectively. Consequently, there are only three molting cycles to adulthood after hatching, in contrast to four molts after hatching in *C. elegans*. Nematodes of the Diplogastridae family can either go through the direct cycle, or enter the non-feeding, stress-resistant juvenile dauer stage to evade unsuitable environmental conditions (Grant et al. 2011; Ogawa et al. 2009).

1.1.1 Model system for studying phenotypic plasticity

P. pacificus has been established as a model organism with forward and reverse genetic methods, including CRISPR-associated engineering (Witte et al. 2015; Han et al. 2020; Nakayama et al. 2020), molecular tools containing transgenesis (Schlager et al. 2009), in situ hybridization (Tian et al. 2008), and integrative genomic technologies (Dieterich et al. 2008; Borchert et al. 2010; Sinha et al. 2012; Werner et al. 2018). *P. pacificus* and the common nematode model organism *Caenorhabditis elegans* evidently shared a last common ancestor ~200 mya (Howard et al. 2022).

The genome of *P. pacificus* was originally sequenced in 2008, supplemented by single molecule re-sequencing in 2019 (Dieterich et al. 2008; Rödelsperger et al. 2017), resulting in 28,896 annotated genes in the 168-Mb genome (Athanasouli et al. 2020). In comparison, the genome of *C. elegans* is 100 Mb in size and contains 20,040 coding genes. Approximately 30% of the *P. pacificus* genes have 1:1 orthologs in *C. elegans* (Rödelsperger et al. 2019).

P. pacificus is a free-living nematode and is distributed worldwide. The majority of *Pristionchus* species, including *P. pacificus*, have been isolated from soil and/or from scarab beetles (Herrmann et al. 2007, 2010). In total, around 50 species in the *Pristionchus* genus are available as living cultures for molecular investigations (Kanzaki et al. 2021). The large number of culturable species allows for conducting deep taxon sampling with high phylogenetic resolution (Prabh et al. 2018), and provides an opportunity to perform comparative studies of gene expression and gene regulation (Rödelsperger et al. 2018; Werner et al. 2018). In the last decade, molecular mechanistic studies on the development and evolution of mouth-form plasticity, predation and associated self-recognition processes provided insight into life history strategies and the evolution of novelty.

1.1.2 Mouth-form dimorphism in *P. pacificus*

The members of the nematode family Diplogastridae are characterized by the absence of the grinder in the terminal bulb of the pharynx. This grinder, which is present in the majority of free-living nematodes, is the primary structure to break open bacteria and thus, is essential for the feeding of *C. elegans* and other nematodes. The absence of a grinder in Diplogastridae and *P. pacificus* resulted in a diversity of feeding structures and strategies (Sudhaus and Fürst von Lieven, 2003). *P. pacificus* and many other species, form teeth-like denticles in their mouth, which makes them omnivorous feeders of bacteria, fungi, and other nematodes (Lightfoot and Sommer, 2022). From the cellular structural perspective, the mouth (stoma) consists of the 'buccal cavity', i.e., the lumen that extends from the mouth opening to the anterior end of the pharynx and the 'buccal capsule', i.e., its cuticular lining. De Ley and co-workers distinguished the buccal capsule into six different regions, including 'cheilostom', 'gymnostom', and 'stegostom', which is further subdivided into 'pro-', 'meso-', 'meta'- and 'telo'-

stegostom (Fig. 1C) (De Ley et al. 1995). Specifically, the most anterior part of the buccal cavity was defined as the 'cheilostom', which is surrounded by labial cells, thought to be the continuation of the cuticle of the worm, and is separated from the remaining parts of the buccal cavity. The 'gymnostom' is surrounded by arcade epidermis and is located at the middle of the buccal capsule. The posterior-most part of the buccal cavity is named 'stegostom', which is formed by interradiial muscle cells. Strikingly, diplogastrid nematodes display a huge evolutionary diversification in the stegostomal region and exhibit remarkably morphological diversity in their teeth-like denticle (von Lieven and Sudhaus 2000; Susoy et al. 2015). Most recently, Ragsdale and co-workers have reconstructed the epithelial and myoepithelial cells that form the mouth and pharynx of *P. pacificus* from transmission electron micrographs (Harry et al. 2021)

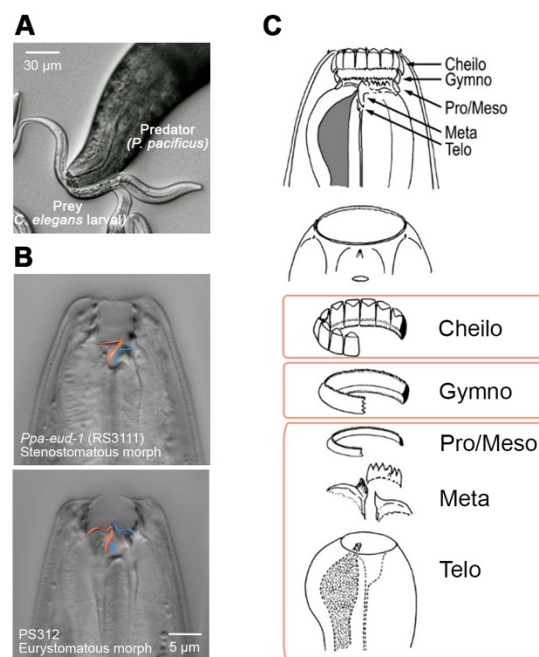


Figure 1. Mouth-form plasticity of predatory morphology in *Pristionchus pacificus*

(A) Shown here is a eurystomatous *Pristionchus pacificus* preying on a larva of *Caenorhabditis elegans*. (B) Mouth dimorphism in stenostomatous (St) and eurystomatous (Eu) morphs in median plane. The microbivorous St morph (RS3111) (Upper image) has a narrow mouth, a dorsal tooth (in orange), and a minute denticle on the right ventrosublateral part of stegostom (in blue), whereas the omnivorous Eu morph (PS312) (Lower image) has a wide mouth, a claw-shaped dorsal tooth (in orange), and a hooked ventrosublateral tooth (in blue). (C) Schematic representation of six distinguished regions in the buccal capsule, containing the cheilostom (Cheilo), gymnostom (Gymno), pro- and mesostegostom (Pro/Meso), metastegostom (Meta), and telostegostom (Telo). The schematic in C was modified from previous work (Ragsdale et al. 2013a)

P. pacificus forms two alternative mouth-forms with different teeth-like denticles, enabling different behaviors (Bento et al. 2010). Based on the morphology of the six regions described above, the dimorphic mouth-form of *P. pacificus* has been described as the stenostomatous (St) morph, expressing a narrow buccal cavity and only one single dorsal tooth, and the eurytomatous (Eu) morph, showing a broader buccal cavity and two large teeth (Fig. 1A, B). The morphological differences in these two morphs elicit diversity in predation and feeding strategy. Specifically, St animals follow a strict bacterial diet and are incapable of killing and successfully attacking potential prey under laboratory conditions (Wilecki et al. 2015; Seroby et al. 2014). In contrast, Eu individuals are more aggressive and prey upon both fungi and other nematodes, in addition to feeding on Bacteria (Susoy et al. 2016; Sudhaus 2010; Ragsdale et al. 2013b). Interestingly, the development of teeth-like denticles, mouth-form plasticity, and predation are characters that are shared across several taxa of the Diplogastridae but are absent in *C. elegans* and other nematodes of the Rhabditidae (Susoy et al. 2015).

1.1.3 Molecular basis of mouth-form plasticity in *P. pacificus*

Mouth-form plasticity in *P. pacificus* represents a model system for studying the mechanisms of phenotypic plasticity (Sommer 2020). Ragsdale and co-workers identified a developmental switch gene through a large forward genetic screen for morph-defective mutants. (Ragsdale et al. 2013b). The first mutant identified was named *eud-1*, for *eurytomatous-form-defective*, and encodes a sulfatase. Mutations in *eud-1* are dominant so that hermaphrodites with a single mutant copy are already preferentially St, and overexpression of *eud-1* in a *eud-1* mutant background can completely revert the all-St mutant mouth form into an all-Eu morph. These results strongly support a role of *eud-1* as a genetic switch gene in the development of mouth form. Furthermore, *eud-1* is located in a multigene locus that contains two pairs of duplicated genes in a tandem inverted configuration (Sieriebriennikov et al. 2018). Unlike the all-St phenotype of mutant in *eud-1*, mutants in the other genes in the multiple gene locus express an all-Eu mouth form. Apart from *eud-1*, the nuclear hormone receptor, *nhr-40* gene, was identified as another switch gene based on the

fact that gain-of-function mutants express all-Eu morph, whereas the loss-of-function mutants exhibiting all-St phenotype (Sieriebriennikov et al. 2020).

The irreversible decision of alternative mouth form is made by the switch network at an early juvenile stage. A “phenotypic execution network” is required for the eventual mouth-form differentiation (Bui and Ragsdale 2019; Sieriebriennikov et al. 2020). Sieriebriennikov and co-workers performed a large-scale unbiased genetic screen and identified the *nhr-1* gene as the downstream genetic factor of the switch network. By analyzing the stage-specific transcriptomes of mutants in *nhr-40* and *nhr-1*, they discovered the common targets of the nuclear hormone receptors NHR-40 and NHR-1. The phylogenetic analysis revealed that all of these common targets are fast-evolving genes, which suggests that rapidly evolving genes and the associated proteins have an important role in the evolution of a novel morphological structure (Sieriebriennikov et al. 2020). Altogether, research on mouth-form dimorphism in the past decade was mainly focused on (a) the environmental influences, (b) the developmental switches, and (c) the gene regulatory network associated with mouth-form plasticity and the associated evolutionary processes (Werner et al. 2017; Ragsdale et al. 2013b; Sieriebriennikov et al. 2018, 2020; Bui and Ragsdale 2019; Casasa et al. 2021). However, little know about the structure genes that directly build the mouth form of nematode.

1.2 The cuticular extracellular matrix of nematode

In nematodes, a cuticular extracellular matrix (ECM) forms the first barrier against infection and other environmental insults. This flexible and resilient exoskeleton is essential to maintain body morphology, and fulfils important roles in locomotion via attachment to muscle (Kramer et al. 1988; von Mende et al. 1988; Johnstone 1994). This complex ECM is synthesized repeatedly, during the molting process, to form larval- and adult- specific cuticles that permit growth. In the ECM experimental model system, the cuticle of *C. elegans* has been described as consisting of two types of cuticle: (a) collagenous cuticle, which lines in the main body of the animal and is flexible for exoskeleton growth in a continuous fashion, and (b) chitinous cuticle, which locates at the buccal cavity containing the entrance to the pharynx and enables growth in a saltatory fashion (Knight et al. 2002; Veronico et al. 2001). Unlike the body wall

cuticle, the cuticular linings of the pharynx is thought to contain the hardy biomaterial chitin (Zhang et al. 2005), as well as possibly functional amyloids in *C. elegans* (George-Raizen et al. 2014). However, this line of reasoning is still considered controversial, and much as there has been scant research on the structural basis of the mouth (stoma) and the potential role of chitin than on the body wall cuticle.

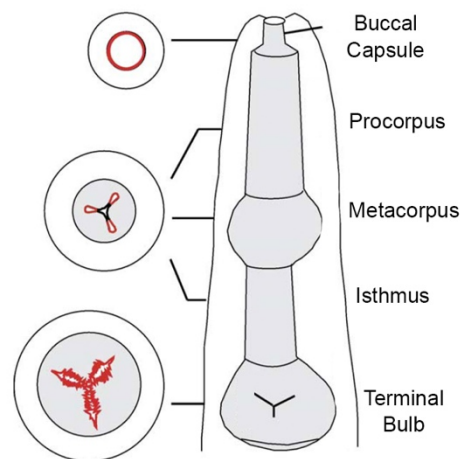


Figure 2. The schematic illustrates the proposed localization of chitin (red) in transverse sections of the pharynx (shaded gray) in *C. elegans*. the schematic was modified from the previous study (Zhang et al. 2005).

1.2.1 Collagenous cuticle

The collagenous cuticle in the Nematoda phylum has a well-ordered cytoarchitecture, which, at the electron-microscopic level, is characterized by distinct layers and transverse structures (Bird and Bird, 1991). In general, the nematode cuticle is composed of highly cross-linked, soluble, and insoluble structural proteins, which have not been thoroughly characterized. Collagens and cuticulins are some of the best studied components of the cuticle, but other minor proteins and lipids and carbohydrates are found in its structure. The presence of different layers of varying thickness is dependent on the stages and the species of nematode, e.g., a medial layer with struts is present only in adult stage of *C. elegans*. The collagenous cuticle is synthesized by the underlying epithelial tissue, called epidermal or hypodermal (Wright and Thomson 1981; Lee 2002), and is shed at each molting stage.

1.2.1.1 The collagens and their mutations in nematode

The cuticle collagens are encoded by the collagen superfamily with more than 150 members in *C. elegans* (Cohen and Sundaram 2020). Individual collagen genes display distinct temporal gene expression patterns during the process of cuticle collagen biosynthesis. Cuticle collagen gene expression in *C. elegans* was originally detected to be developmentally regulated (Cox and Hirsh 1985), revealing that collagens are synthesized at high rates during molts and at lower rates during the intermolt. With the development of RNA-sequencing, the high-resolution temporal gene expression catalogs were established in *C. elegans* (Hendriks et al. 2014). 91 oscillating collagens synchronized with molt cycles as well as 35 adult-specific collagens were identified precisely. Nonetheless, noticeably distinct expression patterns were detected in different classes of collagens, e.g., the *col* genes highly expressed mostly during the molt, which is in contrast to the gene expression of the *rol* and *dpy* families, which suggest that individual collagen genes with highly specific temporal expression patterns may facilitate assembly of a complex structure.

Collagen gene mutants aid in the comprehensive understanding of the function of individual collagen in the structure of the cuticle. In *C. elegans*, mutating some of the collagen-encoding genes or collagen-modifying enzymes result in abnormal morphology (Brenner 1974; Kusch and Edgar 1986). To date, 21 cuticle collagen genes have been identified that result in the body morphology defects, including phenotypes described as DumPY (Dpy), ROLLer (Rol), BLIster (Bli), SQuaT (Sqt), Ray AbnorMal (Ram), and LONg (Lon) (Page 2007).

1.2.1.2 The roles of astacin metalloproteases in collagen formation

Collagen biosynthesis is a complex multi-step process (Prockop and Kivirikko 1995; Myllyharju and Kivirikko 2004), including numerous modifying, folding, and processing enzymes. Similar processes are expected to be involved in nematode collagen biosynthesis, and mutations in some of these enzymes display profound changes in body morphology (Page and Winter 2003). Astacin (NAS) metalloproteases belong to the M12A group of metalloproteases family, which is one of the important proteases in nematodes and is demonstrated to be involved in hatching, cuticle collagen processing and cuticle molting in *C. elegans* (Hishida et al. 1996; Davis et al. 2004; Novelli et al. 2004). Interestingly, the procollagen C-proteinase NAS-35 (also named DPY-31 since

mutating this gene results in a dumpy phenotype) has been proven to be an essential procollagen C-peptidase, involved in SQT-3 cuticle collagen processing and proper cuticle formation (Novelli et al. 2004, 2006). Mutations in *nas-35* affect the cuticle structure and nematode morphogenesis. Furthermore, the orthologues of NAS-35 from the parasitic nematodes both *H. contortus* and *Brugia malayi* were demonstrated to efficiently process the *C. elegans* cuticle collagen SQT-3 at the correct C-terminal procollagen processing site, suggesting that similar specific astacin substrates are shared between these diverse nematode species (Davis et al. 2004; Stepek et al. 2010). In addition, the astacin metalloprotease NAS-6 is required for digestion of the L4 grinder but not the formation of the adult teeth. In *nas-6* mutant animals, the old L4 cuticle cannot be digested as well, whereas the adult grinder appeared normal (Sparacio et al. 2020).

1.2.2 Chitinous cuticle

Chitin is the second most abundant amino polysaccharide in nature, and is commonly found in lower eukaryotes and invertebrates. Chitin is a crystalline polymer of β -1,4 linked N-acetyl-D-glucosamine, and is thought to fulfill its role through the association with different types of cuticular proteins and catecholamines (Wigglesworth 1957; Andersen 1979; Wright 1987; Hopkins and Kramer 1992; Andersen et al. 1995; Sugumaran 1998). Chitin contributes to the cell wall and a specialized cell wall structure called the septum, which is found in the yeast *Saccharomyces cerevisiae* (Cabib et al. 2001). Arthropod chitin is a major part of both the exoskeleton and the insect peritrophic matrices, which serves to separate and protect internal tissue from the digestive process (Merzendorfer and Zimoch 2003).

1.2.2.1 Chitin formation and catalytic mechanism of chitin synthase

Synthesis, transport, and deposition of chitin are complex processes in which single chitin chains are produced at the cytoplasmic site of the chitin synthase. After translocation to the extracellular space, they are assembled as microfibrils and cross-linked with components of the extracellular matrix. In insects, chitin formation has been described as a three-step process. The first step involves the catalytic domain of chitin synthases facing the cytoplasmic site to form the polymer. In the second step, the

nascent polymer translocate across the membrane and is released into the extracellular space. In the last step, single polymers spontaneously assemble to form crystalline microfibrils and the process is completed. In subsequent reactions, the microfibrils combine with other sugars, proteins, glycoproteins, and proteoglycans to form fungal septa and cell walls, as well as arthropod cuticles and peritrophic matrices (Merzendorfer 2006).

Chitin formation is catalyzed by the enzyme chitin synthase, a widespread glycosyltransferase found in every chitin-synthesizing organism, such as fungi, nematodes, and arthropods (Roncero 2002; Coutinho et al. 2003; Merzendorfer and Zimoch 2003). The sequences within the catalytic domain are remarkably conserved in fungal and insect. The acceptor saccharide binding site and the product-binding site of fungal chitin synthases have been particularly investigated. The crucial functions of the signature sequences EDR and QRRRW were identified via studying the activity of the enzyme by site directed mutagenesis in yeast (Fig. 3) (Nagahashi et al. 1995). Specifically, the conservative substitutions in the putative acceptor saccharide binding site [E(D562E)R] and product-binding site [(Q601)RR(R604L)(W605Y)] dramatically decrease the activity of the yeast chitin synthase (Nagahashi et al. 1995; Cos et al. 1998), hinting at the essential function of the conserved motifs in catalysis. Additionally, the highly conserved motif (S/T)WGTKG, which is located within the carboxy-terminus, is required for *in vitro* activity (Cos et al. 1998), indicating that the carboxy-terminal part of the chitin synthase has a potential role in the translocation process of nascent chitin polymers across the membrane (Fig. 3).

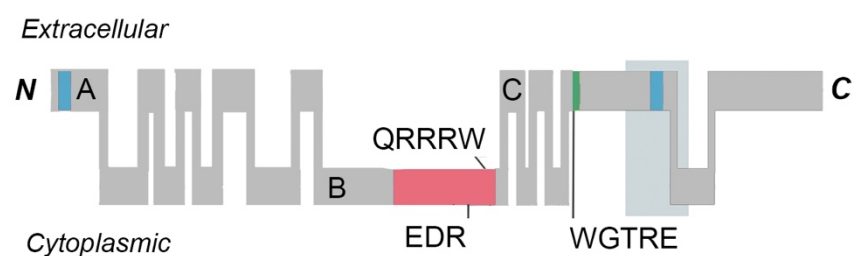


Figure 3. Schematic representation of the domain structures and membrane topology of the *Manduca sexta* chitin synthase 1. Transmembrane helices are shown as vertical bars, extracellular or cytoplasmic regions are depicted as horizontal bars. The N-terminal, catalytic and C-terminal domains are labeled with A, B and C, respectively. The three motifs, EDR, QRRRW and WGTRE are highly conserved chitin synthase sequences. The blue, red, and green boxes show supposed sites of *N*-glycosylation, catalysis and coiled-coils, respectively. This schematic was modified from the previous study (Merzendorfer and Zimoch 2003).

1.2.2.2 The roles of chitin synthase in the development of worms

Chitin synthases have been described in a wide variety of phyletic groups. In *Drosophila*, it has been shown that mutations in the *Chitin Synthase-1 (CS1)* gene are allelic with *krotzkopf verkehrt (kkv)*, a gene that was originally identified in various screens for lethal mutations (Jürgens et al. 1984; Ostrowski et al. 2002). A detailed morphological study of *CS-1/kkv* mutant flies demonstrated that chitin synthesis has essential roles in (a) the integrity of the procuticle, (b) the stability of the epicuticle, and (c) the maintenance of epidermal morphology and activity of underlying enzymes involved in sclerotization and iv) pigmentation (Moussian et al. 2005).

In nematodes, chitin is known to be present in the eggshells of many free living (Wharton 1980) and parasitic species (Brydon et al. 1987). Initial molecular studies in the parasitic nematodes *Brugia malayi* (Harris et al. 2000), *Dirofilaria immitis* (Harris and Fuhrman 2002) and *Meloidogyne artiellia* (Veronico et al. 2001) revealed the presence of one chitin synthase gene. After the whole genome sequencing of *C. elegans* was completed, two putative chitin synthases, *Cel-chs-1* (T25G3.2) and *Cel-chs-2* (F48A11.1), have been identified (Harris et al. 2000; Veronico et al. 2001), indicating the likely presence of orthologs of these two genes in other nematodes. The studies of the spatiotemporal expression patterns of these two chitin synthase genes revealed their essential and non-overlapping functions in development of nematodes. *Cel-chs-1* is specific for forming the eggshell and highly expressed in late larvae and adult hermaphrodites, while *Cel-chs-2* was shown to be detected in the pharynx and its RNA transcription correlates with the larval molting (Veronico et al. 2001; Zhang et al. 2005). Knocking down *Cel-chs-1* by RNA interference leads to eggs that are fragile and permeable to small molecular and the absence of embryonic cell division, revealing that *Cel-chs-1* has essential role in embryo development. Reducing *Cel-chs-2* by RNAi leads to a defect in the pharynx and causes L1 larval arrest, and the expression of cytoplasmic *Cel-chs-2::GFP* reporter in the pharynx is coupled with the presence of chitin, indicating that chitin has potential roles in the functions of the pharynx, e.g. providing rigidity to the pumping pharynx and/or strength for the mechanical breakdown of food in the grinder.

1.2.2.3 Inhibitors affecting chitin synthesis

As chitin is synthesized by insects and fungi but not vertebrates, chitin synthase represents a target for combating insect pests or disease-causing fungi. Inhibitors of chitin synthesis have been historically classified into two major clusters: peptidyl nucleosides and acylureas. Peptidyl nucleosides isolated from various *Streptomyces* bacteria act as substrate analogs to directly inhibit chitin synthase at its catalytic site. In contrast, acylureas may function in several processes of chitin formation including (a) altering either vesicle transport or fusion, (b) inhibiting translocation of chitin polymers across the membrane, and (c) interfering in the hormonal regulation of chitin synthesis by influencing the ecdysteroid production (Nakagawa and Matsumura 1994; Fournet et al. 1995; Cohen 2001). Peptidyl nucleosides contain polyoxins (Hori et al. 1971) and nikkomycins (Dähn et al. 1976).

Nikkomycin Z (NZ) was originally discovered from *Streptomyces* as an agent with antifungal activity (Möhrle et al. 1995; Georgopapadakou and Walsh 1996). The subsequent studies in *Saccharomyces cerevisiae* revealed that NZ is a competitive inhibitor of chitin synthase (ScChs) as a substrate analogue, and it was found that NZ blocks the activities of two of three ScChs isozymes, (ScChs) 1 and 3, but not 2 (Cabib 1991; Choi et al. 1994; Gaughran et al. 1994). Studying the susceptibility of CaChs isozymes to NZ *in vitro* revealed that NZ affects the activities of all three isozymes of chitin synthase in *Candida albicans* (CaChs) (Kim et al. 2002). In *Tegillarca granosa*, the inhibition of TgCHS via NZ significantly increases the percentage of polyspermy, indicating that TgCHS plays an important role in polyspermy blocking (Han et al. 2016). It is assumed that the inhibition is due to the reduction of nascent chitin by NZ, which subsequently slows the exocytosis of cortical vesicles (Schönitzer et al. 2011; Cabib and Arroyo 2013; Walker et al. 2015). Collectively, NZ has been demonstrated as a competitive inhibitor of chitin synthase in several organisms, however, thus far, no investigation of NZ has been performed in nematodes.

1.2.3 Remodeling of cuticular structure during molts in nematodes

Ecdysozoa, the clade of animals defined by the presence of an exoskeleton, exchange their cuticle during the episodic molts (Aguinaldo et al. 1997; Dunn et al. 2008). Nematodes can increase in size during the intermolt period and after the final molt,

growth arrest occurring at molting (Wilson 1976). Most nematodes undergo four molts during their post-embryonic development, except for *P. pacificus* which has only three molts during the external larval stages (Félix et al. 1999). The cuticle covering the body walls, and the cuticular linings of the buccal cavity, pharynx, and other opening tissues are shed and replaced during molts.

The morphological changes of molting in *C. elegans* have been investigated by high resolution Normarski interference contrast microscopy. Molting has been described to consist of three major steps: (a) lethargus occurs at the onset of the molt, and corresponds to a period of inactivity during which pharyngeal pumping, feeding and locomotion become arrested (Cassada and Russell 1975); (b) apolysis is characterized by the separation of connections between the old cuticle basal zone and the underlying epidermis, the new cuticle is produced in the hypodermis and is secreted into the extracellular space between it and the old cuticle; (c) ecdysis results in the shedding of the old cuticle. Immediately prior to ecdysis, the pharynx begins twitching, and gland secretions are thought to be released to aid dissolving the old cuticle. Finally, the larval or young adult pushes with its head to break out of the old cuticle. The duration of molting stage displays a large diversity across distinct species, e.g., 3 to 4 hours for *Panagrellus silusiae* and *C. elegans* (Samoiloff and Pasternak 1969; Cassada and Russell 1975; Singh and Sulston 1978), 72 hours for *Hemicycliophora arenaria* (Johnson et al. 1970), and even several days for *Rotylenchus reniformis* (Bird and Akhurst 1983; Bird 1984).

One example of extracellular structure in *C. elegans* is the pharyngeal grinder, which ruptures bacteria during feeding, and is replaced with larger substitutes during each molting. Comprehensive ultrastructural analysis of pharyngeal muscle and the cuticular grinder revealed the deconstruction of the larval grinder and the construction of the adult grinder during the fourth larval stage (L4)-to-adult transition (Sparacio et al. 2020). In the Diplogastridae family, to which *P. pacificus* belongs, the grinder has been lost with the concomitant reduction of muscle cells in the postcorpus, but they have gained teeth at the anterior of pharynx (Riebesell and Sommer 2017). This suggests that remodeling of teeth in *P. pacificus* might be executed at each molt stage.

1.3 High-resolution temporal of transcriptome sequencing

A major objective of evolutionary genetics is to link gene regulatory changes to adaptive evolution of complex phenotypes. Variation in gene expression patterns often plays a key part in the evolution of morphological phenotypes (Stern and Orgogozo 2008) as well as other complex traits (Kleinjan and van Heyningen 2005; Wray 2007). The tremendous developments in genomic technologies in the last decades allows biologists for the first time to employ sophisticated high-throughput approaches to study of gene expression patterns and associated phenotypic diversity. Current throughputs approaches are orders of magnitude higher than what was possible at the dawn of DNA sequencing technologies.

1.3.1 Development of transcriptome sequencing

The first-generation Sanger-sequencing was established after the structure of DNA was discovered (Watson and Crick 1953; Sanger et al. 1977). Although the Sanger sequencing technology contributed to the first human genome project, the technology had limitations due to its labor-intensive processes and high cost. Second-generation sequencing approaches, also known as short-read NGS, such as on the Illumina or Ion Torrent platforms, were introduced between 2004 and 2006 (Mardis 2013). Second-generation approaches leveraged the power of nanotechnology, were characterized by shorter reads and stronger ability in sequencing millions of DNA molecules in parallel with shorter time and lower price in contrast to the first-generation sequencing (Tucker et al. 2009). Second-generation sequencing technologies can be broadly divided into DNA sequencing (DNA-Seq) and RNA sequencing (RNA-Seq).

Transcriptome sequencing or RNA sequencing (RNA-seq) quantifies transcriptome in given samples and thus contributes to a deeper understanding of RNA biology, form questions such as when and where transcription occurs to the folding and intermolecular interactions that govern RNA function. The first complete transcriptome was sequenced in bacteriophage MS2 by Walter Fiers and his colleagues in 1970s (Fiers et al. 1976). The discovery of reverse transcriptase allowed to convert the mRNA and ncRNA to stabilizer complementary DNA (cDNA) and made it possible to sequence RNA by sequencing its cDNA (Mizutani and Temin 1970). Microarray technology was established as a powerful and systematic approach to quantitatively detect a large number of known genes through hybridization (Wilhelm and Landry 2009). As one of the most important applications of next-generation

sequencing technique, RNA-seq overcomes the drawbacks of microarray and has become an indispensable method for transcriptome-wide analysis of different gene expression and differential splicing of mRNAs.

1.3.2 Adaptation of low input RNAs-seq method for nematodes

The RNA-seq technique has been prevalently applied in nematodes over the past decade. Most recently, high-resolution spatial maps of gene expression was made possible through single-cell transcriptional profiling even in microscopic organisms such as *C. elegans* (Cao et al. 2017). High-resolution temporal gene expression catalogs have been established in *C. elegans*, reflecting the expression patterns during development (Levin et al. 2012; Kim et al. 2013; Hendriks et al. 2014; Meeuse et al. 2020). However, such methods are restricted to organisms that can be massively cultured and are able to develop in synchrony. Most nematodes are small worms that lack sufficient RNA for standard RNA-seq methods without pooling hundred to thousands of individuals. Low input RNA-seq approaches have been increasingly developed as a response to the growing prevalence of low input RNAs-seq studies based on small tissues, single microorganisms, and single cells. The single cell RNA-seq protocol, Smart-seq2 has originally been adapted for single nematode RNA-sequencing (Macchietto et al. 2017) and has recently been revised (Chang et al. 2021). Single nematode RNA-seq could in theory be adopted in many organisms and allow us to obtain high resolution temporal transcriptome at single individual level.

Hatching and ecdysis (the shedding of the cuticle) are major landmarks of molting animals that define the chronological age of individual animals. The developmental age of nematodes can be precisely defined as hours post-hatching, highlighted by these specific landmarks. Additionally, the advances in nematode model organisms, including short lifespan and small size of genome, allows for sequencing numerous single-individual transcriptomes with enough deeper sequencing but at low cost. Thus, it is possible to apply both the single-worm RNA-sequencing approach and the chronological age of individual to establish a high-resolution catalog of transcriptome at single-worm resolution; a catalog that encompasses the entire postembryonic development process. Such transcriptomic maps will provide the platform for both tracking the gene expression pattern of individual genes and aiding in the exploration of the development function of genes of interest.

1.3.3 Gene expression oscillations

Spatiotemporal control of gene expression is crucial for orchestrated differentiation of multiple cells within the developing animal, and also subject to evolutionary changes. The mechanism that controls the patterning of embryo through spatial gene expression gradients was well described in *Drosophila* (Johnston and Nüsslein-Volhard 1992), and since, the regulation of nematode larval development via the temporal expression gradients of heterochronic genes has been established (Ruvkun and Giusto 1989; Johnstone and Barry 1996; Moss et al. 2007). Gene expression oscillations are thought to act as transcriptional clocks that can drive numerous rhythmic biological events, and were found in many biological systems, such as the circadian rhythms in metabolism and behavior (Panda et al. 2002), vertebrate somitogenesis (Oates et al. 2012), plant lateral root branching (Moreno-Risueno et al. 2010), and *Caenorhabditis elegans* larval development (Kim et al. 2013; Hendriks et al. 2014; Meeuse et al. 2020).

Transcriptional clocks temporally control biological processes, thereby allowing the organisms to synchronize with environmental cues. In yeast metabolic cycles, transcripts involved in growth require an oxidative environment for their high expression under high oxygen condition, followed by a reductive phase during which the cell cycle or mitochondrial biogenesis proceeds (Tu et al. 2005; Tu and McKnight 2006). In photosynthetic cyanobacteria, the periodic expression of the circadian clock-controlled genes allows the organism to produce energy and grow, while temporally separating incompatible chemical steps, i.e., photosynthesis and oxidative metabolism in the oxygen-generating condition of daylight and oxygen-inhibited nitrogen fixation under the more reduced night environment (Lenz and Søgaard-Andersen 2011). Similarly, the circadian clock-regulated transcripts is also described as acting to separate metabolic processes in plants (Harmer et al. 2000). In *C. elegans*, three systematic and quantitative studies on temporal transcriptomics during larval development have been performed by using RNA-seq technology (Kim et al. 2013; Hendriks et al. 2014; Meeuse et al. 2020), which revealed that extensive ultradian oscillations periodically expressed during *C. elegans* postembryonic development. These studies on *C. elegans* oscillations indicate similar oscillations in other molting animals, e.g., the rhythmic molting and metamorphosis in *Drosophila* and moths.

2 Aims of this thesis

The main objective of my research is to elucidate the genetic factors governing structural bases of mouth form in the nematode *P. pacificus*. To address this question, I simultaneously pursued several lines of research:

- i) I developed single-worm transcriptome sequencing in *P. pacificus*, aiming at profiling transcriptome from single nematodes at a given developmental stage and a particular culture condition.
- ii) I performed a 1- to 2-h high-resolution catalog of gene expression throughout postembryonic development to enable us to track gene expression patterns of single genes, and to characterize genes with oscillating expression patterns
- iii) I conducted a comparative analysis of oscillatory gene expression between the two model organisms *P. pacificus* and *C. elegans* and associated evolutionary dynamics to decipher the evolutionary mechanism of temporal gene expression.
- iv) I conducted a large-scale forward genetic screen to discover uncharacterized genetic players regulating structural bases of mouth form, and to provide a comprehensive understanding of regulation of mouth-form development.
- v) I investigated the roles of chitin synthase genes in *P. pacificus* teeth formation through reverse genetics, and explored whether the inhibitor nikkomycin Z reduces activity of chitin synthase in nematodes.

3 Results

3.1 Single worm transcriptomics identifies a developmental core network of oscillating genes with deep conservation across nematodes

Sun, S., Rödelberger, C. and Sommer, R.J.

Genome Research, 31(9), pp.1590-1601.

3.1.1. Synopsis

High-resolution spatial and temporal maps of gene expression have tremendously helped towards a comprehensive understanding of animal development and evolution. In nematodes, the small body size represented a major challenge for such studies, but recent advancements have helped overcome this limitation. Here, we implemented single worm transcriptomics (SWT) in the nematode model organism *Pristionchus pacificus* to generate a high-resolution temporal map of the developmental transcriptome. Furthermore, we identified 2,982 (10%) of genes exhibited oscillatory expression in *P. pacificus*, and revealed that revealed an overrepresentation of ancient gene classes among oscillating genes. Importantly, several of these collagen genes are regulated by the developmental switch gene *eud-1* indicating a potential function in the regulation of mouth-form plasticity, a key developmental process in this facultative predatory nematode. This analysis provides i) an updated protocol for SWT in nematodes that is applicable to many microscopic species, ii) a 1 – 2 hours high-resolution catalogue of *P. pacificus* gene expression throughout postembryonic development, and iii) a comparative analysis of oscillatory gene expression between the two model organisms *P. pacificus* and *C. elegans* and associated evolutionary dynamics.

3.1.2. Own contribution

I conceived and designed all experiments with Dr. Ralf Sommer. I conducted all the experimental studies. I performed all of the bioinformatic analyses with the help of Dr. Christian Rödelberger. I participated in writing of the manuscript. Thus, I estimate my contribution at 90%.

3.2 The oscillating Mucin-type protein DPY-6 has a conserved role in nematode mouth and cuticle formation

Sun, S.[†], Theska, T.[†], Witte, H., Ragsdale, E.J. and Sommer, R.J.
Genetics, iyab233

3.2.1 Synopsis

Nematodes show an extraordinary diversity of mouth structures and strikingly different feeding strategies, which has enabled an invasion of all ecosystems. However, nearly nothing is known about the molecular mechanisms associated with the nematode mouth form. Here, we used a large-scale genetic screen to identify genes involved in mouth formation. We identified *Ppa-dpy-6* to encode a Mucin-type hydrogel-forming protein that is macroscopically involved in the specification of the cheilostom, the anterior part of the mouth. We leveraged the power of geometric morphometrics to characterize these defects further and found additional defects that affect mouth form. Additionally, *Ppa-dpy-6* is shorter than wild-type with a typical Dumpy phenotype, indicating a role in the formation of the external cuticle. In this publication, we identified the Mucin-type protein DPY-6 as the first protein involved in the specification of nematode stoma with a conserved role in cuticle disposition. This concomitant phenotype of the cheilostom and cuticle provides the first molecular support for the continuity of these structures and for the separation of the cheilostom from the rest of the stoma.

3.2.3 Own contribution

I performed the large-scale genetic screen, generated all genomic libraries, and conducted the mapping experiments. In addition, I performed all RNA sequencing data collection, oscillating pattern analysis, and differential gene expression analysis, as well as measuring worm body sizes. I was involved in hypothesizing and designing the experiments, as well as writing relevant parts of the manuscript. Thus, my contribution to this manuscript was around 60%.

3.3 Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell

Sieriebriennikov, B., Sun, S., Lightfoot, J.W., Witte, H., Moreno, E., Rödelsperger, C. and Sommer, R.J.

PLoS genetics, 16(4), p.e1008687.

3.3.1 Synopsis

Developmental plasticity is ubiquitous and attracting considerable attention in the context of adaptation to climate change and as a facilitator of evolutionary novelty. However, little is known about the associated molecular mechanisms in nematodes. Here, we leveraged the power of suppressor screen genetics to identify the conserved nuclear hormone receptor (NHR) NHR-1 as a second transcription factor regulating the development of environmentally-induced predatory morphs in the nematode *Pristionchus pacificus*. Next, transcriptomic analysis revealed that two conserved nuclear hormone receptors (NHR-40 and NHR-1) have only 24 common target genes. Their systematic CRISPR/Cas9 knockouts and expression analysis indicated extreme functional redundancy and showed that they are strikingly expressed in a single pharyngeal gland cell, g1D., whose morphological remodeling accompanied the evolution of teeth and predation. This study enhances the understanding of the GRN regulating mouth-form plasticity, elucidates the evolutionary dynamics of underlying genes and links rapid gene evolution with morphological innovations associated with plasticity. The identification of gene regulatory networks (GRN) controlling plasticity will provide an understanding of development in novel environments and enable testing theories about the long-term evolutionary significance of plasticity.

3.3.2 Own contribution

I designed and created reporter lines for Astacin genes, performed a suppressor screen in the *nhr-1* mutant background, and participated in mouth-form phenotyping. Overall, my contribution to this study was 20%.

3.4 Improving transgenesis efficiency and CRISPR-associated tools through codon optimization and native intron addition in *Pristionchus nematodes*

Han, Z. †, Lo, W.S. †, Lightfoot, J.W., Witte, H., Sun, S. and Sommer, R.J.
Genetics, 216(4), pp.947-956.

3.4.1 Synopsis

P. pacificus suffers from a much less efficient transgenesis system compared with *C. elegans*. To improve the efficiencies of *P. pacificus* transgenics, we firstly investigated the genomic features of 10 nematode species from three of the major clades representing all different lifestyles, and uncovered that there were drastically different codon usage bias and intron composition among these species. Then, we performed the processes of codon optimization and native intron addition in *P. pacificus*, which result in dramatical improvement of the transgenesis efficiency. In addition, we leverage the power of the improved transgenesis approach to develop a fluorescent co-injection marker for CRISPR/Cas9 screening. In this study, we dramatically improved the efficiencies of *P. pacificus* transgenics, and developed a fluorescent co-injection marker for screening of the potential (CRISPR)/Cas9-edited candidates.

3.4.2 Own contribution

I participated in the investigation of fluorescent transgenic lines. I estimate my contribution at 10%.

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

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

Scientific reports 9.1 (2019): 1-9.

3.5.1 Synopsis

Improvement of gene annotation in the nematode model organism *P. pacificus* is crucial for gene function studies using reverse genetics as well as identification of orthologs of candidate genes in other species. In this paper, we employed an integrative approach using comparative genomic and transcriptomic data combined with community-based curation to improve the *P. pacificus* gene annotations. First, a comparative assessment of 22 nematode genomes revealed that *P. pacificus* has one of the best available nematode genomes. Then, we identified *C. elegans* genes that homologs and orthologs are not or incorrectly annotated in the *P. pacificus* genome. Finally, a community-based manual curation of suspicious gene models reveals thousands of missing homologs and hidden orthologs. Our small-scale crowdsourcing has drastically improved gene annotations, and the manual gene annotation with comparable quality will facilitate numerous functional studies across multiple model organisms. This pilot study can be extended to even larger gene sets, which would raise the quality of gene annotations to the next level.

3.5.2 Own contribution

I participated in the group discussions in which the manuscript was conceived.

I estimate my contribution at 5%.

3.6 Chitin contributes to the formation of a feeding structure in a predatory nematode

Shuai Sun, Hanh Witte & Ralf J. Sommer

3.6.1 Synopsis

The formation of teeth-like denticles in independent lineages of nematodes represent major innovations with a key role for the evolution of predation and parasitism and the diversity of nematode ecosystem functioning. These teeth are histologically and developmentally different from vertebrate teeth and most importantly, their biochemical composition remains elusive. Here we show a crucial function of chitin in the formation of the teeth-like denticles in *P. pacificus*. Phylogenesis analysis revealed two conserved chitin synthase exist in the nematode evolution. CRISPR/Cas9-based gene knockouts of the highly conserved chitin synthase *Ppa-chs-2* result in larval lethal or viable, but toothless mutants, depending on the exact molecular lesion in the gene. Similarly, the toothless phenotype was obtained under the treatment with the chitin-synthase inhibitor Nikkomycin Z, which consistent with the conclusion that chitin is essential for the formation of the teeth-like denticles of *P. pacificus*. Toothless animals survive to adulthood and feed on various bacteria, but they are incapable of predation. We provide support for a quantitative role of chitin in the formation of the complex feeding apparatus in nematodes.

3.6.2 Own contribution

I conducted all the experiments. I performed microinjection for CRISPR and Nikkomycin Z with the help of Hanh Witte. I was involved in hypothesizing and designing the experiments, as well as writing relevant parts of the manuscript. Thus, my contribution to this manuscript was around 90%.

3.7 Protocol for single-worm transcriptomics in *Pristionchus pacificus*, step by step

Shuai Sun, Ralf J. Sommer

3.7.1 Synopsis

Adaptations and improvements of low input RNA-seq technology for microscopic organisms, nematodes, allows for the analysis of gene expression heterogeneity at the individual nematode resolution. The single-worm transcriptomics (SWT) provides a robust platform to perform a large-scale high-resolution transcriptomic analysis in the nematode model organism of *P. pacificus*. Here, we established a step-by-step protocol for i) defining the chronological age of individual animals for establishing the high-resolution time course, ii) improving processes of SWT library preparation, iii) supporting a strategy for library sequencing regarding the sequencing cost and the number of detected genes, iv) providing a simple pipeline for analyzing SWT dataset. In conclusion, this protocol contains a detailed step-by-step procedure with many useful recommendations, and will allow investigators without the background in library preparation for exploring transcriptomics across different species or conditions at single-worm resolution.

3.7.2 Own contribution

I developed this protocol and conducted all the experiments. I was involved in hypothesizing and designing the experiments, as well as writing relevant parts of the manuscript. Thus, my contribution to this manuscript was around 95%.

4 Discussion

In this thesis, I updated the single-worm transcriptome sequencing (SWT-seq) protocols and introduced the SWT-seq approach for the nematode model organism *P. pacificus*. I leveraged power of the SWT-seq to establish a high-resolution catalog of gene expression during post-embryonic development in *P. pacificus*, which provides one of the most useful resources for pursuing countless mechanistic studies and comparative analysis in the *Pristionchus* community. The study of gene expression oscillation revealed that thousands of oscillatory genes are expressed periodically during post-embryonic development and uncovered the evolutionary dynamics of oscillating genes in the nematode phylum. Moreover, to investigate the developmental functions of oscillations in the mouth-form plasticity of *P. pacificus*, I perform both forward and reverse genetics. In the forward genetic experiment, I conducted a large-scale genetic screen and isolated six mutants that displayed concomitant phenotype of defective mouth form and altered body shape. Using the advantages of genome-editing technology and geometric morphometric protocol, I and my colleague identified the first structural component of stoma in nematodes, which simultaneously regulates the development of cheilostom of the stoma and cuticle formation. In the reverse genetic project, I investigated the role of chitin in the *P. pacificus* teeth formation. My results suggested that two chitin synthase genes are highly conserved across nematodes. The C-terminus of *Ppa-chs-2* is essential for the teeth formation, mutants in this region result in a viable but teethless phenotype. Strikingly, animals treated with chitin-synthase inhibitor Nikkomycin Z produced the same teethless phenotype. Collectively, my studies focused on (a) the SWT methods development, (b) the evolution of oscillating genes, and (c) the functions of developmental oscillations in the mouth-form plasticity of *P. pacificus*.

4.1 Future use of the single-worm transcriptome sequencing

I developed the SWT in *P. pacificus* (Sun et al. 2021). Compared with the original protocols (Macchietto et al. 2017; Chang et al. 2021), I modified the method in multiple ways, including (a) changing the process of worm lysis, (b) applying lower cDNA input, (c) reducing amplification cycles, and (d) pooling larger samples. A series of

comparative quality controls between SWT and STD (standard RNA-seq) samples revealed that the representation of gene expression is comparable between SWT and STD approaches, demonstrating that SWT provides a robust technology for profiling global gene expression. Obviously, our modified SWT protocol can be easily adapted to the majority of nematode species, and SWT will be a powerful tool for many transcriptomic comparative studies, including differential expression analysis between different mouth-form mutants and different abiotic or biotic environmental conditions.

Importantly, the selection of comparable development time points is crucial to perform transcriptomic comparison using SWT method, due to the different developmental speed of nematodes with different genotypes and/or in different treatments. Further, numerous genes with stage-specific expression pattern are broadly pervasive, i.e., nearly 10% of the *P. pacificus* genome, 18% of the *C. elegans* genome (Hendriks et al. 2014; Meeuse et al. 2020) exhibit oscillating gene expression patterns during postembryonic development, which synchronize with the molting cycle and demonstrate dramatical changes of gene expression within short time periods. We combined two singular development markers, hatching and ecdysis, and were able to provide an atlas that includes 38 developmental time points and covers the first 58 hours of development after hatching. These reference time points can be easily obtained for mutant strains or similarly, experimental treatments, which allows precise selection of comparable time points for SWT sequencing. Note that in our experiments, three biological replicates were performed but it might be reasonable to include more biological replicates for SWT sequencing, since more biological replicates of given development time points can partially overcome the slight time differences among individual replicates for a given time point.

4.2 Developmental oscillations in *P. pacificus*

We focused on developmental oscillations in gene expression and identified 2,964 genes, 10.3 % of the *P. pacificus* genome, with oscillating expression patterns. What functions might these gene-expression oscillations serve? Multiple cuticle-related genes were found that peak preferentially with each molt in *C. elegans* (Hendriks et al. 2014; Meeuse et al. 2020), supporting the early conclusion that periodic transcription of genes is required for molting and cuticle formation (Johnstone and Barry 1996;

Frand et al. 2005). The periodic expression of transcripts is not only involved in molting or cuticle formation, but is also related to other biological processes, e.g., steroid biogenesis, gonad morphogenesis, and embryonic development (Hendriks et al. 2014; Meeuse et al. 2020). Our analysis showed that *P. pacificus* oscillations are strongly enriched in the collagen gene family. Surprisingly, we found these oscillating collagens are distributed in 10 phase classes, indicating that temporal-specific collagens are related to cuticle formation, both during molting and in the intermolt stage. Furthermore, the differential gene expression analysis between the *eud-1* mutant and wild-type animals revealed that 38 *eud-1*-dependent oscillatory collagen genes have potential roles in the mouth-form gene regulatory network. To verify the function of these mouth-form-specific collagens requires leveraging genome-editing and molecular biology tools, in order to identify the collagen related phenotype of the structural stoma in *P. pacificus*.

The comparison of oscillating patterns between *P. pacificus* and *C. elegans* revealed a hitherto hidden evolutionary dimension of developmental oscillation in nematodes. Relying on the previous efforts on the genomic and transcriptomic comparisons between *P. pacificus* and *C. elegans* (Dieterich et al. 2008; Rödelberger et al. 2014, 2017, 2018; Prabh et al. 2018), we employed phylostratigraphic analysis to study the age structure of oscillating genes. It was electrifying to learn that oscillating genes are highly enriched in ancient gene clusters and to identify a highly conserved core network of 528 one-to-one orthologous genes. We hypothesize that *P. pacificus*-*C. elegans* shared oscillations (PCO) have highly conserved developmental functions.

4.3 The oscillating *dpy-6* is involved in nematode cuticle and mouth formation

We identified the Mucin-type hydrogel-forming protein DPY-6 as first structural component involved in the formation of the cheilostom in both *P. pacificus* and *C. elegans*. Interestingly, *dpy-6* was found to have a conserved spatiotemporal expression pattern in *P. pacificus* and *C. elegans*. Both, *Ppa-dpy-6* and *Cel-dpy-6* are expressed in the hypodermal syncytia hyp1, hyp5 and hyp6/7, the anterior and posterior arcade cells, and periodically highly expressed during intermolt stages, indicating their involvement in the formation of the cuticle and the cheilostom during intermolt stages. This result supports the idea that the larval growth that does occur

between molts is facilitated by physical stretching of the flexible cuticle (Knight et al. 2002). In addition, one could conclude that the collagenous cuticle continuously grows during the intermolt stages. Finally, it is an example of a pair of orthologous genes, *Ppa-dpy-6* and *Cel-dpy-6*, with a conserved developmental function in the formation of the cheilostom and cuticle through the conserved spatiotemporal gene expression.

The discovery of *Ppa-DPY-6* as a structural component of the nematode stoma represents a major milestone in our path to fully understand the gene regulatory network (GRN) involved in *P. pacificus* mouth-form development. In the last decade, the comprehensive analyses of mouth-form plasticity and its genetic, epigenetic, and environmental regulation have provided a systematic understanding of the molecular mechanism underlying the mouth-form plasticity in *P. pacificus* (for review see Sommer, 2020). This study revealed the crucial role of *Ppa-dpy-6* in the formation of the cheilostom, the most anterior part of the stoma, and contributes considerably to our understanding of the GRN of *P. pacificus* mouth-form plasticity. Finally, it should be noted that this study places structural components at the terminal branch of the gene regulatory network and opens the road for future studies on the nanotechnology of teeth formation. Also, it is important to note that in contrast to arthropods, i.e., spider webs, little is known about biomineralization in nematodes.

4.4 The role of chitin synthase 2 in the mouth-form plasticity

We showed a major role of chitin in the formation of the teeth-like denticles in *P. pacificus*. Phylogenetic analysis revealed that two chitin synthase genes were highly conserved through representative nematodes, indicating a potential role of chitin in the evolutionary diversification of feeding structures in nematodes. CRISPR/Cas9-based gene knockouts demonstrated that mutations in the C-terminus of *Ppa-chs-2* result in a teethless phenotype with severe malformation of the posterior mouth parts. Similarly, treatment with Nikkomycin Z of J4 juveniles resulted in specific defects in teeth formation, suggesting that chitin synthase activity and thus, chitin is essential for the formation of the teeth-like denticle *P. pacificus*. The viable teethless phenotype of the *Ppa-chs-2* mutants or Nikkomycin Z treated animals might provide a first indication for the different contributions of chitin for the formation of various parts of the nematode stoma. The spatiotemporal expression analysis of *Ppa-chs-2* profiled that *Ppa-chs-2* was highly expressed in the pm1 muscle cells 1-2 hours before each molt, strongly

suggested this essential function to be in the formation of the stoma and the teeth in particular. In conclusion, our data provide the first example of the complete perturbation of teeth formation in nematodes, and will lead to novel activities in various research areas.

5 Bibliography

- Aguinaldo AMA, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**: 489–493.
- Andersen SO. 1979. Biochemistry of insect cuticle. *Annu Rev Entomol* **24**: 29–59.
- Andersen SO, Hojrup P, Roepstorff P. 1995. Insect cuticular proteins. *Insect Biochem Mol Biol* **25**: 153–176.
- Athanasouli M, Witte H, Weiler C, Loschko T, Eberhardt G, Sommer RJ, Rödelsperger C. 2020. Comparative genomics and community curation further improve gene annotations in the nematode *Pristionchus pacificus*. *BMC Genomics* **21**: 1–9.
- Bento G, Ogawa A, Sommer RJ. 2010. Co-option of the hormone-signalling module daftachronic acid–DAF-12 in nematode evolution. *Nature* **466**: 494–497.
- Bird AF. 1984. Growth and moulting in nematodes: Moulting and development of the hatched larva of *Rotylenchulus reniformis*. *Parasitology* **89**: 107–120.
- Bird AF, Akhurst RJ. 1983. The nature of the intestinal vesicle in nematodes of the family steinernematidae. *International Journal for Parasitology* **13**: 599–606.
- Borchert N, Dieterich C, Krug K, Schütz W, Jung S, Nordheim A, Sommer RJ, Macek B. 2010. Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. *Genome Res* **20**: 837–846.
- Brenner S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics* **77**: 71–94.
- Brydon LJ, Gooday GW, Chappell LH, King TP. 1987. Chitin in egg shells of *Onchocerca gibsoni* and *Onchocerca volvulus*. *Molecular and Biochemical Parasitology* **25**: 267–272.
- Bui LT, Ragsdale EJ. 2019. Multiple Plasticity Regulators Reveal Targets Specifying an Induced Predatory Form in Nematodes. *Molecular Biology and Evolution* **36**: 2387–2399.
- Cabib E. 1991. Differential inhibition of chitin synthetases 1 and 2 from *Saccharomyces cerevisiae* by polyoxin D and nikkomycins. *Antimicrobial Agents and Chemotherapy* **35**: 170–173.

- Cabib E, Arroyo J. 2013. How carbohydrates sculpt cells: chemical control of morphogenesis in the yeast cell wall. *Nature Reviews Microbiology* **11**: 648–655.
- Cabib E, Roh D-H, Schmidt M, Crotti LB, Varma A. 2001. The yeast cell wall and septum as paradigms of cell growth and morphogenesis. *Journal of Biological Chemistry* **276**: 19679–19682.
- 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 (1979)* **357**: 661–667.
- Casasa S, Biddle JF, Koutsovoulos GD, Ragsdale EJ. 2021. Polyphenism of a Novel Trait Integrated Rapidly Evolving Genes into Ancestrally Plastic Networks. *Mol Biol Evol* **38**: 331–343.
- Cassada RC, Russell RL. 1975. The dauerlarva, a post-embryonic developmental variant of the nematode *Caenorhabditis elegans*. *Developmental Biology* **46**: 326–342.
- Chang D, Serra L, Lu D, Mortazavi A, Dillman A. 2021. A revised adaptation of the smart-seq2 protocol for single-nematode RNA-seq. In *Methods in Molecular Biology*, Vol. 2170 of, pp. 79–99, Humana Press Inc.
- Choi WJ, Santos B, Durán A, Cabib E. 1994. Are yeast chitin synthases regulated at the transcriptional or the posttranslational level? *Molecular and Cellular Biology* **14**: 7685–7694.
- Cohen E. 2001. Chitin synthesis and inhibition: a revisit. *Pest Management Science* **57**: 946–950.
- Cohen JD, Sundaram M v. 2020. *C. elegans* Apical Extracellular Matrices Shape Epithelia. *Journal of Developmental Biology* **8**.
- Cos T, Ford RA, Trilla JA, Duran A, Cabib E, Roncero C. 1998. Molecular analysis of Chs3p participation in chitin synthase III activity. *European Journal of Biochemistry* **256**: 419–426.
- Coutinho PM, Deleury E, Davies GJ, Henrissat B. 2003. An Evolving Hierarchical Family Classification for Glycosyltransferases. *Journal of Molecular Biology* **328**: 307–317.

- Cox GN, Hirsh D. 1985. Stage-specific patterns of collagen gene expression during development of *Caenorhabditis elegans*. *Mol Cell Biol* **5**: 363–372.
- Dähn U, Hagenmaier H, Höhne H, König WA, Wolf G, Zähler H. 1976. Stoffwechselprodukte von mikroorganismen. *Archives of Microbiology* **107**: 143–160.
- Davis MW, Birnie AJ, Chan AC, Page AP, Jorgensen EM. 2004. A conserved metalloprotease mediates ecdysis in *Caenorhabditis elegans*. *Development* **131**: 6001–6008.
- de Ley P, van de Velde MC, Mounport D, Baujard P, Coomans A. 1995. Ultrastructure of the Stoma in Cephalobidae, Panagrolaimidae and Rhabditidae, With a Proposal for a Revised Stoma Terminology in Rhabditida (Nematoda). *Nematologica* **41**: 153–182.
- Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, Dinkelacker I, Fulton L, Fulton R, Godfrey J, Minx P, et al. 2008. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nature Genetics* **40**: 1193–1198.
- Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**: 745–749.
- Félix MA, Hill RJ, Schwarz H, Sternberg PW, Sudhaus W, Sommer RJ. 1999. *Pristionchus pacificus*, a nematode with only three juvenile stages, displays major heterochronic changes relative to *Caenorhabditis elegans*. *Proceedings of the Royal Society B: Biological Sciences* **266**: 1617–1621.
- Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, den Berghe A, et al. 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* **260**: 500–507.
- Fournet F, Sannier Ch, Moriniere M, Porcheron P, Monteny N. 1995. Effects of Two Insect Growth Regulators on Ecdysteroid Production in *Aedes aegypti* (Diptera: Culicidae). *Journal of Medical Entomology* **32**: 588–593.

- Frand AR, Russel S, Ruvkun G. 2005. Functional genomic analysis of *C. elegans* molting. *PLoS Biology* **3**.
- Fürst von Lieven A. 2005. The embryonic moult in diplogastrids (Nematoda) - Homology of developmental stages and heterochrony as a prerequisite for morphological diversity. *Zool Anz* **244**: 79–91.
- Gaughran JP, Lai MH, Kirsch DR, Silverman SJ. 1994. Nikkomycin Z is a specific inhibitor of *Saccharomyces cerevisiae* chitin synthase isozyme Chs3 in vitro and in vivo. *Journal of Bacteriology* **176**: 5857–5860.
- George-Raizen JB, Shockley KR, Trojanowski NF, Lamb AL, Raizen DM. 2014. Dynamically-expressed prion-like proteins form a cuticle in the pharynx of *Caenorhabditis elegans*. *Biology Open* **3**: 1139–1149.
- Georgopapadakou NH, Walsh TJ. 1996. Antifungal agents: chemotherapeutic targets and immunologic strategies. *Antimicrob Agents Chemother* **40**: 279–291.
- Grant W, Viney M, others. 2011. The dauer phenomenon. In *Molecular and physiological basis of nematode survival*, pp. 99–125, CABI Publishing, Wallingford.
- Han Y, Shi W, Guo C, Zhao X, Liu S, Wang Y, Su W, Zha S, Wu H, Chai X, et al. 2016. Characteristics of chitin synthase (CHS) gene and its function in polyspermy blocking in the blood clam *Tegillarca granosa*. *Journal of Molluscan Studies* **82**: 550–557.
- Han Z, Lo WS, Lightfoot JW, Witte H, Sun S, Sommer RJ. 2020. Improving transgenesis efficiency and CRISPR-associated tools through codon optimization and native intron addition in *pristionchus* nematodes. *Genetics* **216**: 947–956.
- Harmer SL, Hogenesch JB, Straume M, Chang H-S, Han B, Zhu T, Wang X, Kreps JA, Kay SA. 2000. Orchestrated Transcription of Key Pathways in *Arabidopsis* by the Circadian Clock. *Science (1979)* **290**: 2110–2113.
- Harris MT, Fuhrman JA. 2002. Structure and expression of chitin synthase in the parasitic nematode *Dirofilaria immitis*. *Molecular and Biochemical Parasitology* **122**: 231–234.

- Harris MT, Lai K, Arnold K, Martinez HF, Specht CA, Fuhrman JA. 2000. Chitin synthase in the filarial parasite, *Brugia malayi*. *Molecular and Biochemical Parasitology* **111**: 351–362.
- Harry CJ, Messar SM, Ragsdale EJ. 2021. Comparative reconstruction of the predatory feeding structures of the polyphenic nematode *Pristionchus pacificus*. *bioRxiv*.
- Hendriks GJ, Gaidatzis D, Aeschimann F, Großhans H. 2014. Extensive Oscillatory Gene Expression during *C. elegans* Larval Development. *Molecular Cell* **53**: 380–392.
- Herrmann M, Kienle S, Rochat J, Mayer WE, Sommer RJ. 2010. Haplotype diversity of the nematode *Pristionchus pacificus* on Réunion in the Indian Ocean suggests multiple independent invasions. *Biological Journal of the Linnean Society* **100**: 170–179.
- Herrmann M, Mayer WE, Hong RL, Kienle S, Minasaki R, Sommer RJ. 2007. The nematode *Pristionchus pacificus* (Nematoda: Diplogastridae) is associated with the oriental beetle *Exomala orientalis* (Coleoptera: Scarabaeidae) in Japan. *Zoolog Sci* **24**: 883–889.
- Hishida R, Ishihara T, Kondo K, Katsura I. 1996. hch-1, a gene required for normal hatching and normal migration of a neuroblast in *C. elegans*, encodes a protein related to TOLLOID and BMP-1. *EMBO J* **15**: 4111–4122.
- Hopkins TL, Kramer KJ. 1992. Insect cuticle sclerotization. *Annu Rev Entomol* **37**: 273–302.
- Hori M, Kakiki K, Suzuki S, Misato T. 1971. Studies on the Mode of Action of Polyoxins: Part III. Relation of Polyoxin Structure to Chitin Synthetase Inhibition. *Agricultural and Biological Chemistry* **35**: 1280–1291.
- Howard RJ, Giacomelli M, Lozano-Fernandez J, Edgecombe GD, Fleming JF, Kristensen RM, Ma X, Olesen J, Sørensen M v., Thomsen PF, et al. 2022. The Ediacaran Origin of Ecdysozoa: Integrating Fossil and Phylogenomic Data. *J Geol Soc London* jgs2021-107.

- Johnson PW, van Gundy SD, Thomson WW. 1970. Cuticle formation in *Hemicyclophora arenaria*, *Aphelenchus avenae* and *Hirschmanniella gracilis*. *Journal of Nematology* **2**: 59.
- Johnston DS, Nüsslein-Volhard C. 1992. The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**: 201–219.
- Johnstone IL. 1994. The cuticle of the nematode *Caenorhabditis elegans*: a complex collagen structure. *Bioessays* **16**: 171–178.
- Johnstone IL, Barry JD. 1996. Temporal reiteration of a precise gene expression pattern during nematode development. *The EMBO Journal* **15**: 3633–3639.
- Jürgens G, Wieschaus E, Nüsslein-Volhard C, Kluding H. 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. *Wilhelm Roux Arch Dev Biol* **193**: 283–295.
- Kanzaki N, Herrmann M, Weiler C, Röseler W, Theska T, Berger J, Rödelsperger C, Sommer RJ. 2021. Nine new *Pristionchus* (Nematoda: Diplogastridae) species from China.
- Kim DH, Grün D, van Oudenaarden A. 2013. Dampening of expression oscillations by synchronous regulation of a microRNA and its target. *Nature Genetics* **45**: 1337–1345.
- Kim MK, Park HS, Kim CH, Park HM, Choi W. 2002. Inhibitory effect of nikkomycin Z on chitin synthases in *Candida albicans*. *Yeast* **19**: 341–349.
- Kleinjan DA, van Heyningen V. 2005. Long-Range Control of Gene Expression: Emerging Mechanisms and Disruption in Disease. *The American Journal of Human Genetics* **76**: 8–32.
- Knight CG, Patel MN, Azevedo RBR, Leroi AM. 2002. A novel mode of ecdysozoan growth in *Caenorhabditis elegans*. *Evol Dev* **4**: 16–27.
- Kramer JM, Johnson JJ, Edgar RS, Basch C, Roberts S. 1988. The *sqt-1* gene of *C. elegans* encodes a collagen critical for organismal morphogenesis. *Cell* **55**: 555–565.

- Kusch M, Edgar RS. 1986. Genetic studies of unusual loci that affect body shape of the nematode *Caenorhabditis elegans* and may code for cuticle structural proteins. *Genetics* **113**: 621–639.
- Lee DL. 2002. Cuticle, moulting and exsheathment. *The biology of nematodes* 171–209.
- Lenz P, Søgaard-Andersen L. 2011. Temporal and spatial oscillations in bacteria. *Nature Reviews Microbiology* **9**: 565–577.
- Levin M, Hashimshony T, Wagner F, Yanai I. 2012. Developmental Milestones Punctuate Gene Expression in the *Caenorhabditis* Embryo. *Developmental Cell* **22**: 1101–1108.
- Macchietto M, Angdembe D, Heidarpour N, Serra L, Rodriguez B, El-Ali N, Mortazavi A. 2017. Comparative transcriptomics of *Steinernema* and *Caenorhabditis* single embryos reveals orthologous gene expression convergence during late embryogenesis. *Genome Biol Evol* **9**: 2681–2696.
- Mardis ER. 2013. Next-generation sequencing platforms. *Annual review of analytical chemistry* **6**: 287–303.
- Meeuse MW, Hauser YP, Morales Moya LJ, Hendriks G, Eglinger J, Bogaarts G, Tsiaris C, Großhans H. 2020. Developmental function and state transitions of a gene expression oscillator in *Caenorhabditis elegans*. *Molecular Systems Biology* **16**: 1–21.
- Merzendorfer H. 2006. Insect chitin synthases: a review. *Journal of Comparative Physiology B* **176**: 1–15.
- Merzendorfer H, Zimoch L. 2003. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *Journal of Experimental Biology* **206**: 4393–4412.
- Mizutani S, Temin HM. 1970. An RNA-dependent DNA polymerase in virions of Rous sarcoma virus. In *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 35 of, pp. 847–849.
- Möhrle V, Roos U, Bormann C. 1995. Identification of cellular proteins involved in nikkomycin production in *Streptomyces tendae* Tü901. *Molecular Microbiology* **15**: 561–571.

- Moreno-Risueno MA, Norman JM van, Moreno A, Zhang J, Ahnert SE, Benfey PN. 2010. Oscillating Gene Expression Determines Competence for Periodic *Arabidopsis* Root Branching. *Science (1979)* **329**: 1306–1311.
- Moss T, Langlois F, Gagnon-Kugler T, Stefanovsky V. 2007. A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cellular and molecular life sciences* **64**: 29–49.
- Moussian B, Schwarz H, Bartoszewski S, Nüsslein-Volhard C. 2005. Involvement of chitin in exoskeleton morphogenesis in *Drosophila melanogaster*. *Journal of Morphology* **264**: 117–130.
- Mylyharju J, Kivirikko KI. 2004. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends in Genetics* **20**: 33–43.
- Nagahashi S, Sudoh M, Ono N, Sawada R, Yamaguchi E, Uchida Y, Mio T, Takagi M, Arisawa M, Yamada-Okabe H. 1995. Characterization of chitin synthase 2 of *Saccharomyces cerevisiae*: Implication of two highly conserved domains as possible catalytic sites. *Journal of Biological Chemistry* **270**: 13961–13967.
- Nakagawa Y, Matsumura F. 1994. Diflubenzuron affects gamma-thioGTP stimulated Ca²⁺ transport in vitro in intracellular vesicles from the integument of the newly molted american cockroach, *Periplaneta americana* L. *Insect Biochemistry and Molecular Biology* **24**: 1009–1015.
- Nakayama K-I, Ishita Y, Chihara T, Okumura M. 2020. Screening for CRISPR/Cas9-induced mutations using a co-injection marker in the nematode *Pristionchus pacificus*. *Dev Genes Evol* **230**: 257–264.
- Novelli J, Ahmed S, 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**: 1259–1273.
- Novelli J, Page AP, Hodgkin J. 2006. The C terminus of collagen SQT-3 has complex and essential functions in nematode collagen assembly. *Genetics* **172**: 2253–2267.
- Oates AC, Morelli LG, Ares S. 2012. Patterning embryos with oscillations: structure, function and dynamics of the vertebrate segmentation clock. *Development* **139**: 625–639.

- Ogawa A, Streit A, Antebi A, Sommer RJ. 2009. A conserved endocrine mechanism controls the formation of dauer and infective larvae in nematodes. *Current Biology* **19**: 67–71.
- Ostrowski S, Dierick HA, Bejsovec A. 2002. Genetic control of cuticle formation during embryonic development of *Drosophila melanogaster*. *Genetics* **161**: 171–182.
- Page A. 2007. The cuticle. *WormBook* 1–15.
- Page AP, Winter AD. 2003. Enzymes involved in the biogenesis of the nematode cuticle. Vol. 53 of *Advances in Parasitology*, pp. 85–148, Academic Press
- Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M, Schultz PG, Kay SA, Takahashi JS, Hogenesch JB. 2002. Coordinated Transcription of Key Pathways in the Mouse by the Circadian Clock. *Cell* **109**: 307–320.
- 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 Research* **28**: 1664–1674.
- Prockop DJ, Kivirikko KI. 1995. Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem* **64**: 403–434.
- Ragsdale EJ, Kanzaki N, Röseler W, Herrmann M, Sommer RJ. 2013a. Three new species of *Pristionchus* (Nematoda: Diplogastridae) show morphological divergence through evolutionary intermediates of a novel feeding-structure polymorphism. *Zoological Journal of the Linnean Society* **168**: 671–698.
- Ragsdale EJ, Müller MR, Rödelsperger C, Sommer RJ. 2013b. A Developmental Switch Coupled to the Evolution of Plasticity Acts through a Sulfatase. *Cell* **155**: 922–933.
- Riebesell M, Sommer RJ. 2017. Three-dimensional reconstruction of the pharyngeal gland cells in the predatory nematode *Pristionchus pacificus*. *Journal of Morphology* **278**: 1656–1666.
- 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*. *Scientific Reports* **9**: 1–9.

- Rödelsperger C, Meyer JM, Prabh N, Lanz C, Bemm F, Sommer RJ. 2017. Single-Molecule Sequencing Reveals the Chromosome-Scale Genomic Architecture of the Nematode Model Organism *Pristionchus pacificus*. *Cell Rep* **21**: 834–844.
- Rödelsperger C, Neher RA, Weller AM, Eberhardt G, Witte H, Mayer WE, Dieterich C, Sommer RJ. 2014. Characterization of genetic diversity in the nematode *Pristionchus pacificus* from population-scale resequencing data. *Genetics* **196**: 1153–1165.
- 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. *Current Biology* **28**: 3123-3127.e5.
- Roncero C. 2002. The genetic complexity of chitin synthesis in fungi. *Curr Genet* **41**: 367–378.
- Ruvkun G, Giusto J. 1989. The *Caenorhabditis elegans* heterochronic gene *lin-14* encodes a nuclear protein that forms a temporal developmental switch. *Nature* **338**: 313–319.
- Samoiloff MR, Pasternak J. 1969. Nematode morphogenesis: fine structure of the molting cycles in *Panagrellus silusiae* (de Man 1913) Goodey 1945. *Canadian Journal of Zoology* **47**: 639–643.
- Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences* **74**: 5463–5467.
- Schlager B, Wang X, Braach G, Sommer RJ. 2009. Molecular cloning of a dominant roller mutant and establishment of DNA-mediated transformation in the nematode *Pristionchus pacificus*. *genesis* **47**: 300–304.
- Schönitzer V, Eichner N, Clausen-Schaumann H, Weiss IM. 2011. Transmembrane myosin chitin synthase involved in mollusc shell formation produced in *Dictyostelium* is active. *Biochemical and Biophysical Research Communications* **415**: 586–590.
- Seroby V, Ragsdale EJ, Sommer RJ. 2014. Adaptive value of a predatory mouth-form in a dimorphic nematode. *Proceedings of the Royal Society B: Biological Sciences* **281**: 20141334–20141334.

- Sieriebriennikov B, Prabh N, Dardiry M, Witte H, Röseler W, Kieninger MR, Rödelsperger C, Sommer RJ. 2018. A Developmental Switch Generating Phenotypic Plasticity Is Part of a Conserved Multi-gene Locus. *Cell Rep* **23**: 2835-2843.e4.
- Sieriebriennikov B, Sun S, Lightfoot JW, Witte H, Moreno E, Rödelsperger C, Sommer RJ. 2020. Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell. *PLoS Genetics* **16**: 1–27.
- Singh RN, Sulston JE. 1978. Some Observations On Moulting in *Caenorhabditis Elegans*. *Nematologica* **24**: 63–71.
- Sinha A, Sommer RJ, Dieterich C. 2012. Divergent gene expression in the conserved dauer stage of the nematodes *Pristionchus pacificus* and *Caenorhabditis elegans*. *BMC Genomics* **13**: 1–17.
- Sommer R, Carta LK, Kim S, Sternberg PW. 1996. Morphological, genetic and molecular description of *Pristionchus pacificus*. *Fundamental and applied Nematology* **19**: 511–521.
- Sommer RJ. 2020. Phenotypic Plasticity: From Theory and Genetics to Current and Future Challenges. *Genetics* **215**: 1 LP – 13.
- Sommer RJ, McGaughan A. 2013. The nematode *Pristionchus pacificus* as a model system for integrative studies in evolutionary biology. *Mol Ecol* **22**: 2380–2393.
- Sommer RJ, Streit A. 2011. Comparative Genetics and Genomics of Nematodes: Genome Structure, Development, and Lifestyle. *Annual Review of Genetics* **45**: 1–20.
- Sparacio AP, Trojanowski NF, Snetselaar K, Nelson MD, Raizen DM. 2020. Teething during sleep: Ultrastructural analysis of pharyngeal muscle and cuticular grinder during the molt in *Caenorhabditis elegans*. *PLoS ONE* **15**.
- Steppek G, McCormack G, Page AP. 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**: 533–542.
- Stern DL, Orgogozo V. 2008. THE LOCI OF EVOLUTION: HOW PREDICTABLE IS GENETIC EVOLUTION? *Evolution (N Y)* **62**: 2155–2177.

- Sudhaus W. 2010. Preadaptive plateau in Rhabditida (Nematoda) allowed the repeated evolution of zooparasites, with an outlook on evolution of life cycles within Spiroascarida. *Palaeodiversity (Stuttg)* **3**: 117–130.
- Sugumaran M. 1998. Unified mechanism for sclerotization of insect cuticle. *Advances in Insect Physiology* **27**: 229–334.
- Sun S, Rödelsperger C, Sommer RJ. 2021. Single worm transcriptomics identifies a developmental core network of oscillating genes with deep conservation across nematodes. *Genome Research* **31**: 1590–1601.
- Susoy V, Herrmann M, Kanzaki N, Kruger M, Nguyen CN, Rödelsperger C, Röseler W, Weiler C, Giblin-Davis RM, Ragsdale EJ, et al. 2016. Large-scale diversification without genetic isolation in nematode symbionts of figs. *Science Advances* **2**: e1501031.
- Susoy V, Ragsdale EJ, Kanzaki N, Sommer RJ. 2015. Rapid diversification associated with a macroevolutionary pulse of developmental plasticity. *Elife* **2015**: 1–39.
- Tian H, Schlager B, Xiao H, Sommer RJ. 2008. Wnt signaling induces vulva development in the nematode *Pristionchus pacificus*. *Current Biology* **18**: 142–146.
- Tu BP, Kudlicki A, Rowicka M, McKnight SL. 2005. Logic of the Yeast Metabolic Cycle: Temporal Compartmentalization of Cellular Processes. *Science (1979)* **310**: 1152–1158.
- Tu BP, McKnight SL. 2006. Metabolic cycles as an underlying basis of biological oscillations. *Nature reviews Molecular cell biology* **7**: 696–701.
- Tucker T, Marra M, Friedman JM. 2009. Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine. *The American Journal of Human Genetics* **85**: 142–154.
- Veronico P, Gray L, Jones J, Bazzicalupo P, Arbucci S, Cortese M, di Vito M, de Giorgi C. 2001. Nematode chitin synthases: gene structure, expression and function in *Caenorhabditis elegans* and the plant parasitic nematode *Meloidogyne artiellia*. *Molecular Genetics and Genomics* **266**: 28–34.
- von Lieven AF, Sudhaus W. 2000. Comparative and functional morphology of the buccal cavity of Diplogastrina (Nematoda) and a first outline of the phylogeny of

- this taxon*. *Journal of Zoological Systematics and Evolutionary Research* **38**: 37–63.
- von Mende N, Bird DM, Albert PS, Riddle DL. 1988. dpy-13: a nematode collagen gene that affects body shape. *Cell* **55**: 567–576.
- Walker LA, Lee KK, Munro CA, Gow NAR. 2015. Caspofungin Treatment of *Aspergillus fumigatus* Results in ChsG-Dependent Upregulation of Chitin Synthesis and the Formation of Chitin-Rich Microcolonies. *Antimicrobial Agents and Chemotherapy* **59**: 5932–5941.
- Watson JD, Crick FHC. 1953. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**: 737–738.
- 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. *Scientific Reports* **7**: 1–12.
- 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 Research* **28**: 1675–1687.
- Wharton D. 1980. Nematode egg-shells. *Parasitology* **81**: 447–463.
- Wigglesworth VB. 1957. The physiology of insect cuticle. *Annual Review of Entomology* **2**: 37–54.
- Wilecki M, Lightfoot JW, Susoy V, Sommer RJ. 2015. Predatory feeding behaviour in *Pristionchus* nematodes is dependent on phenotypic plasticity and induced by serotonin. *Journal of Experimental Biology* **218**: 1306–1313.
- Wilhelm BT, Landry J-R. 2009. RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* **48**: 249–257.
- Wilson PAG. 1976. Nematode growth patterns and the moulting cycle: the population growth profile. *Journal of Zoology* **179**: 135–151.
- 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*. *Development Genes and Evolution* **225**: 55–62.

- Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics* **8**: 206–216.
- Wright KA, Thomson JN. 1981. The buccal capsule of *Caenorhabditis elegans* (Nematoda: Rhabditoidea): an ultrastructural study. *Canadian Journal of Zoology* **59**: 1952–1961.
- Wright TRF. 1987. The genetics of biogenic amine metabolism, sclerotization, and melanization in *Drosophila melanogaster*. *Adv Genet* **24**: 127–222.
- Zhang Y, Foster JM, Nelson LS, Ma D, Carlow CKS. 2005. The chitin synthase genes *chs-1* and *chs-2* are essential for *C. elegans* development and responsible for chitin deposition in the eggshell and pharynx, respectively. *Developmental Biology* **285**: 330–339.

6 Appendix

Research

Single worm transcriptomics identifies a developmental core network of oscillating genes with deep conservation across nematodes

Shuai Sun, Christian Rödelberger, and Ralf J. Sommer

Max-Planck Institute for Developmental Biology, Department for Integrative Evolutionary Biology, 72076 Tübingen, Germany

High-resolution spatial and temporal maps of gene expression have facilitated a comprehensive understanding of animal development and evolution. In nematodes, the small body size represented a major challenge for such studies, but recent advancements have helped overcome this limitation. Here, we have implemented single worm transcriptomics (SWT) in the nematode model organism *Pristionchus pacificus* to provide a high-resolution map of the developmental transcriptome. We selected 38 time points from hatching of the J2 larvae to young adults to perform transcriptome analysis over 60 h of post-embryonic development. A mean sequencing depth of 4.5 million read pairs allowed the detection of more than 23,135 (80%) of all genes. Nearly 3000 (10%) genes showed oscillatory expression with discrete expression levels, phases, and amplitudes. Gene age analysis revealed an overrepresentation of ancient gene classes among oscillating genes, and around one-third of them have 1:1 orthologs in *C. elegans*. One important gene family overrepresented among oscillating genes is collagens. Several of these collagen genes are regulated by the developmental switch gene *eud-1*, indicating a potential function in the regulation of mouth-form plasticity, a key developmental process in this facultative predatory nematode. Together, our analysis provides (1) an updated protocol for SWT in nematodes that is applicable to many microscopic species, (2) a 1- to 2-h high-resolution catalog of *P. pacificus* gene expression throughout postembryonic development, and (3) a comparative analysis of oscillatory gene expression between the two model organisms *P. pacificus* and *C. elegans* and associated evolutionary dynamics.

[Supplemental material is available for this article.]

The evolution of novelty and phenotypic divergence depends on two major molecular processes. First, the emergence of new genes through gene duplication, divergence, and de novo formation constantly change the genomic composition of organisms (for review, see Rödelberger et al. 2019b). Second, changes in the expression of individual genes during development can lead to phenotypic divergence. In general, transcriptomes show highly dynamic spatiotemporal patterns that are orchestrated by multiple regulatory events. However, a detailed understanding of transcriptional changes and their significance for evolution requires sophisticated molecular tools and comparative analyses. In recent years, the ENCODE and modENCODE consortia have generated unprecedented amounts of RNA sequencing data with increasing resolution, which also allowed for the comparison across distant species (Gerstein et al. 2014). Most recently, high-resolution spatial maps of gene expression are made possible through single-cell transcriptional profiling even in animals with small body size such as *Caenorhabditis elegans* (Cao et al. 2017). High temporal resolution throughout embryonic and/or postembryonic development requires the analysis of many parallel samples. In *C. elegans*, the use of highly synchronized cultures has provided the first comprehensive maps of gene expression throughout development (Levin et al. 2012; Kim et al. 2013; Hendriks et al. 2014; Meeuse et al. 2020). However, such methods are restricted to organisms that can be massively cultured and that stay synchronous throughout development. One alternative methodology is single worm transcriptomics (SWT), a method that has originally been intro-

duced in two entomopathogenic nematodes of the genus *Steinernema* (Macchietto et al. 2017) and recently has been revised (Chang et al. 2021). Although such a method could in theory be adopted in many organisms, to the best of our knowledge, it has not yet been widely applied. Here, we implemented SWT in the model nematode *Pristionchus pacificus* with a new updated protocol that uses (1) a different method of worm lysis, (2) even further reduced cDNA input, and (3) reduced PCR amplification cycles to provide a high-resolution map of the transcriptome throughout postembryonic development.

P. pacificus is a well-established model system in evolutionary biology first described in 1996 (Sommer et al. 1996). Based on its rapid growth with a generation time of only 4 d in the laboratory and its hermaphroditic mode of reproduction, forward and reverse genetic techniques including CRISPR-associated engineering have been successfully implemented in this nematode (Witte et al. 2015; Han et al. 2020; Nakayama et al. 2020). Although original work focused on the evolution of developmental processes and comparison to *C. elegans* (Wang and Sommer 2011), more recent studies developed *P. pacificus* as a model system for developmental plasticity and the evolution of novelty and complexity (Sommer 2020). In contrast to *C. elegans*, *P. pacificus* forms two alternative mouth-forms with different teeth-like denticles, enabling different feeding strategies (Bento et al. 2010). So-called “stenostomatous” (St) animals have a single tooth and are strict bacterial feeders, whereas “eurystomatous” (Eu) animals have two teeth and can

Corresponding author: ralf.sommer@tuebingen.mpg.de

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.275303.121>.

© 2021 Sun et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

additionally feed on fungi and other nematodes by predation (Ragsdale et al. 2013). By now, the gene regulatory network controlling mouth-form plasticity has been identified (Sieriebriennikov et al. 2018, 2020), and the long-term evolutionary consequences of mouth-form plasticity for the evolution of novelty were carefully investigated (Susoy et al. 2015, 2016).

The development of teeth-like denticles, mouth-form plasticity, and predation are characters that are shared by several taxa of the Diplogastridae, the family to which *P. pacificus* belongs, but they are absent in *C. elegans* and its relatives (Susoy et al. 2015). Importantly, all diplogastrid nematodes share another developmental feature unknown from *C. elegans*. Specifically, *P. pacificus* hatches in the J2 juvenile stage after the first molt occurs in the egg. This heterochronic feature has severe consequences for the development of these organisms (Fürst von Lieven 2005). Specifically, there are only three free-living preadult larval stages, the J2, J3, and J4 stages, respectively. Thus, the *P. pacificus* J2, J3, and J4 stages correspond to the *C. elegans* L2, L3, and L4 stages, respectively. Consequently, there are only three molts to adulthood after hatching versus four molts in *C. elegans*. Finally, the time of development spent in the egg after fertilization is extended in *P. pacificus* to nearly 24 h (20°C), whereas hatching occurs after 18 h in *C. elegans* (Wood 1988). In consequence, postembryonic development in *P. pacificus* is concluded after 45 h of development and three molting cycles.

P. pacificus and *C. elegans* belong to different nematode families, the Diplogastridae and Rhabditidae, respectively (Kanzaki and Giblin-Davis 2015). Both species reproduce as hermaphrodites, which made them important model organisms with the associated establishment of extensive technology platforms as indicated above. *P. pacificus* and *C. elegans* are thought to have shared a last common ancestor ~100 mya (Prabh et al. 2018; Werner et al. 2018). Their genomes differ in size and the total number of genes, with the 100-Mb genome of *C. elegans* containing 20,040 genes and the 168-Mb genome of *P. pacificus* containing 28,896 genes (Athanasouli et al. 2020). Importantly, 30% of the *P. pacificus* genes have 1:1 orthologs in *C. elegans* (Rödelsperger et al. 2019a). *P. pacificus* is one of around 50 species in the genus *Pristionchus*, all of which are available as living cultures for molecular investigations (Kanzaki et al. 2021). Building on this high phylogenetic resolution, recent transcriptomic studies and deep taxon sampling revealed high evolutionary dynamics of novel gene families in *Pristionchus* evolution (Prabh et al. 2018; Rödelsperger et al. 2018). Together, the availability of various integrative genomics platforms and the large number of culturable species with its high phylogenetic resolution make *Pristionchus* a primary study system for animal and nematode evolution, the evolutionary dynamics of gene families, and new gene origin (for review, see Rödelsperger et al. 2019b). This study aims at establishing SWT for *P. pacificus* to generate a high-resolution temporal map of gene expression throughout its postembryonic development and at focusing on the investigation of oscillatory gene expression patterns.

Results

Detailed characterization of *P. pacificus* postembryonic development highlights major heterochrony in comparison to *C. elegans*

To establish a robust developmental framework for studying the temporal dynamics of gene expression in *P. pacificus* at high reso-

lution, we first followed postembryonic development in highly synchronized cultures, starting with the J2 stage that hatches from the egg ~24 h after fertilization. When fed with *E. coli* OP50 on NGM agar plates at 20°C, *P. pacificus* needs ~46 h to reach adulthood. We first measured the duration of the intermolt period of the three juvenile stages J2, J3, and J4 and the time required for molting (Fig. 1A; Supplemental Fig. S1). There was little difference in the duration of the three molts, which consisted of an ~2-h lethargus–apolysis phase followed by a relatively short ecdysis of 10–15 min. These time points are similar to those measured in *C. elegans* under similar culture conditions (Kim et al. 2013; Hendriks et al. 2014). In contrast, the length of the intermolt differed substantially between stages and species. The *P. pacificus* J2 stage is by far the longest with 16 h, whereas the J3 and J4 stage are 9.6 and 12 h, respectively. In comparison to *C. elegans*, all juvenile stages are longer in *P. pacificus*, reflecting the somewhat longer overall generation time. However, the first free-living juvenile stages, *C. elegans* L1 and *P. pacificus* J2, are substantially longer than all others. In contrast, the *P. pacificus* J1 stage in the egg shell is drastically reduced with the J1 cuticle being formed 16–17 h after fertilization and hatching of the J2 larva after 24 h (Fürst von Lieven 2005). Thus, the *P. pacificus* J1 takes only ~8 h, whereas the L1 stage of *C. elegans* lasts for ~20 h (Kim et al. 2013). This recharacterization of postembryonic development is largely consistent with previous studies (Félix et al. 1999; Fürst von Lieven 2005) and highlights major differences in developmental timing between *C. elegans* and *P. pacificus*.

SWT provides a robust platform for studying the temporal dynamics of gene expression

With this timetable in hand, we decided to use 38 time points for transcriptomic analysis (Fig. 1B; Supplemental Table S1). These 38 time points cover the first 58 h after hatching of the J2 larvae with a 1- to 2-h resolution. Seven of them are after the final molt to provide some information about transcriptional changes within young adults. We conducted SWT using a modified protocol originally generated by Macchietto and coworkers (for details, see Methods) (Macchietto et al. 2017). We performed all experiments in triplicate and ran 114 independent samples on three lanes on an Illumina HiSeq 3000 (Fig. 1C; Supplemental Fig. S2). We obtained a mean sequencing depth of 4.5 million read pairs and only six out of 114 samples had fewer than 2 million read pairs (Supplemental Fig. S2C). We subsampled the raw data of one SWT and one standard RNA-seq (STD) library and profiled the distribution of the number of detected genes among nine levels of sequencing depth. We found that 2 million reads are sufficient to detect 95% of genes that are found with 12 million reads (Supplemental Fig. S2B). We removed the six libraries with fewer than 2 million PE reads in the following analysis. To compare gene expression between SWT and STD samples, we first investigated exon coverage at selected gene loci (Fig. 1D) and found that the gene structure of most genes showed complete agreement between STD and SWT (Supplemental Fig. S2D; Supplemental Table S13). Furthermore, we observed similar density distributions of gene expression levels between SWT and newly generated STD libraries (Fig. 1E). Finally, we found a strong correlation of gene expression when comparing the data set of the last lethargus of the fourth molt with a Pearson's correlation of $R^2 = 0.86$ (Fig. 1F). Note that we found similar correlations between the STD replicates ($R^2 = 0.97$) and between the SWT replicates ($R^2 = 0.92$) (Supplemental Fig. S2A). Thus, our modified protocol with (1) a different type of worm lysis, (2) lower cDNA

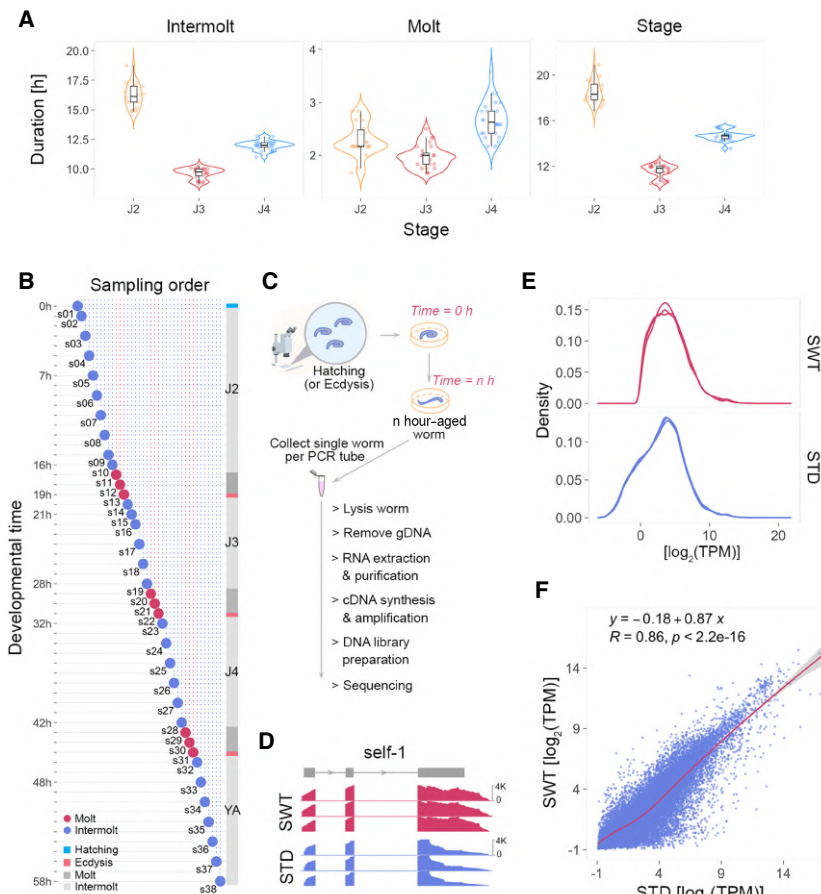


Figure 1. Single worm transcriptomic (SWT) sequencing in *P. pacificus*. (A) The violin plots display the duration of intermolt, molt, and the whole stage for all three juvenile stages (J2, J3, and J4). The head structure of the wild type at hatching, lethargus, and ecdysis is shown in Supplemental Figure S1. (B) The schematic shows the distribution of time points for SWT samples, which cover 58 h of postembryonic development from hatching to young adults. The three time points of each molt indicated in red represent the substages of molting: lethargus, apolysis, and ecdysis, respectively (for visualization, see Supplemental Fig. S1). Animals were collected hourly or every 2 h in the molt or intermolt stage. (C) Schematic summary of the workflow for SWT. (D) RNA-seq read coverage is visualized at the *self-1* locus using the Integrative Genomics Viewer (IGV) (Robinson et al. 2011) as one example for comparing SWT and standard RNA-seq (STD). (E) Density of gene expression of SWT and STD samples, six replicates each. (F) The scatterplot shows the correlation of the mean TPM of individual genes between SWT and STD data at the lethargus of the fourth molt. Three biological replicates for each method.

input, (3) reduced amplification cycles, and (4) larger sample pooling allows more high-throughput investigations and reduces associated costs without a strong decrease in data accuracy. Together, these data reveal that SWT in *P. pacificus* provides a robust platform for studying the temporal dynamics of gene expression throughout postembryonic development.

SWT establishes the most highly resolved developmental expression atlas of *P. pacificus*

Across all 38 time points, we detected 23,135 reliably expressed genes (REGs) with an average of aligned paired-end (PE) reads higher than or equal to three in at least one time point (Supplemental Table S2). More than 92% of these genes show expression evidence in at least two replicates of a single timepoint (Supple-

mental Fig. S2E), and ~95% of these genes are expressed in at least two time points (Supplemental Fig. S2F). This represents 80.1% of the currently annotated 28,896 genes of the *P. pacificus* genome (Athanasouli et al. 2020). Cross-correlation plots of gene expression profiles revealed two major developmental clusters representing early larval stages (6–38 h) and late larvae, including young adults (42–58 h) (Fig. 2A). This is consistent with a previous analysis of the developmental transcriptome of *P. pacificus*, in which principal component analysis (PCA) and hierarchical clustering also separated between early larvae and late larvae/adults. The differences between early and late transcriptomes can be largely explained by the onset of sexual maturation, as gene clusters with high expression at late stages were significantly enriched for reproductive genes such as major sperm proteins and vitellogenins (Baskaran et al. 2015). However, because only five developmental stages were studied previously, our current data set represents the most highly resolved developmental transcriptome data of *P. pacificus*. This high-resolution catalog of gene expression in *P. pacificus* can assist in many future investigations and the comparative analysis of *P. pacificus* with other organisms. As the general features of the developmental transcriptome of *P. pacificus* were characterized previously, here, we focus on oscillatory gene expression, which can only be studied using the current data set with high temporal resolution.

Thousands of genes oscillate during postembryonic development

The cross-correlation analysis between different time points also revealed signals for periodic expression, suggesting

that a fraction of genes show oscillating gene expression (Fig. 2A).

Indeed, when we performed PCA, we found that three of the four top principal components that explained 22% of the variation were associated with oscillatory patterns (Fig. 2B). PC1 reflects developmental genes with pronounced expression differences between early and late larval stages. In contrast, PC2–PC4 display an oscillatory pattern, which is in large parts synchronized with the molting cycle (Fig. 2B). In total, 2964 genes show oscillating expression patterns, whereas 20,171 are nonoscillating (Fig. 2C; Supplemental Fig. S3A). Most of the oscillating genes show up to three expression peaks (Fig. 2C), which is consistent with the fact that *P. pacificus* undergoes three larval molts after hatching from the egg. Further analysis also showed hundreds of peaks in the molting period, suggesting that the oscillating genes involved

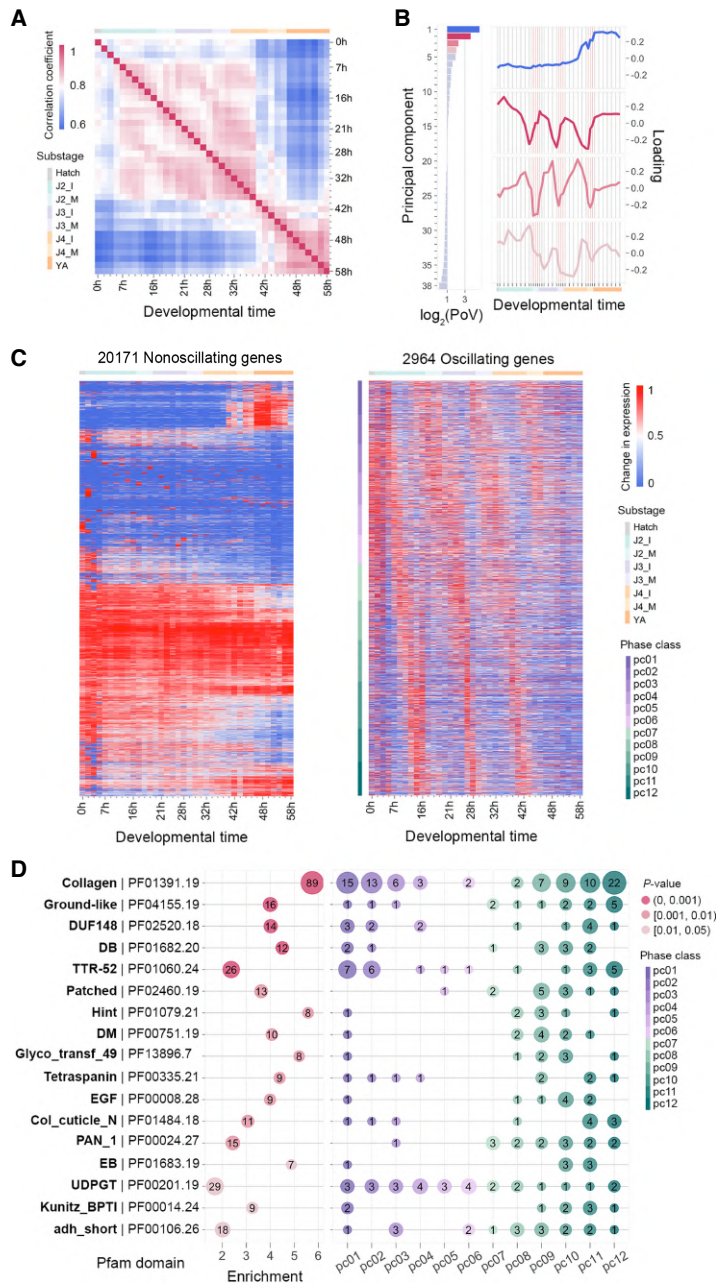


Figure 2. Thousands of oscillating genes are synchronized with the molting cycle in *P. pacificus*. (A) The heatmap displays the cross-correlation of \log_2 -transformed expression patterns of all reliably expressed genes ($n = 23,135$) obtained from the 38 time points. The time course was broken down into seven substages: J2 intermolt (J2_I), J2 molt (J2_M), J3 intermolt (J3_I), J3 molt (J3_M), J4 intermolt (J4_I), J4 molt (J4_M), and young adults (YA). (B) Proportion of variance (PoV) of gene expression by 38 principal components (PCs; left graph) and changes (loadings) of expression for each of the four top PCs (right graph). (C) The heatmaps visualize changes in gene expression throughout development for nonoscillating genes ($n = 20,171$; left) and oscillating genes ($n = 2964$; right). The mean TPM was calculated from three biological replicates and was normalized by its maximum of each expression trace. The order of rows in the heatmap with the oscillating genes was sorted by phase class, which indicated a phase difference of 30, corresponding to a peak shift by ~ 1 h. (D) Oscillating genes were tested for overrepresentation of protein domains. The left part shows the 17 protein domains that are most significantly enriched among oscillating genes, the x-axis presents the enrichment score, and the y-axis shows the name of protein domains. The size of circles corresponds to the number of oscillatory genes with a given protein domain, and the three different colors indicate significance levels as measured by Fisher's exact test. The right graph profiles the overrepresented protein domains according to the 12 phase classes. The number of oscillatory genes in each phase class is marked.

in these expression peaks have specific functions during molting (see Methods) (Supplemental Fig. S3E,F). However, the developmental oscillations are not totally synchronized but instead occur in multiple phases (Supplemental Fig. S3B).

Overrepresentation analysis of protein domains showed the strongest enrichment for collagens, which confirms the observed correlation to molting (Fig. 2D). We found an overrepresentation of collagens in 10 phase classes, suggesting that specific collagens are expressed at different time points of development and that not all collagens are exclusively expressed before molting. Indeed, visualization of collagen expression levels across development confirms the existence of “collagen classes” that are expressed at specific times in the intermolt cycle (Supplemental Fig. S3D). Specifically, 89 of 151 collagen genes of *P. pacificus* displayed oscillation patterns (Fig. 2D; Supplemental Fig. S3D), and oscillating collagens showed a significantly higher expression level than the nonoscillating collagens (Supplemental Fig. S3C). Also, 38 of these oscillating collagens showed expression peaks associated with molting events ($P < 10^{-4}$, Fisher's exact test) (Supplemental Fig. S3F). Taken together, our analysis indicates that thousands of genes in *P. pacificus* show oscillatory gene expression during post-embryonic development with a strong enrichment of several protein domains including collagens.

Developmental oscillations are mediated by ancient gene classes

To test if developmental oscillations represent a highly conserved process or are driven by novel genes, we performed an analysis of gene ages of all oscillating genes. Specifically, we defined gene age classes based on the presence of orthologs (see Methods) in distantly related species using comparative genomic data of six *Pristionchus* species, two close relatives of the genera *Parapristionchus* and *Micoletzkyia* and four more distantly related species, including *C. elegans* (Fig. 3A). The associated ladder-like phylogeny results in 13 branches (Fig. 3A, p01–p13) that separate different age classes and, thus, allows one to date the emergence of new genes. For example, p01–p03 contain genes only known from *P. pacificus* and its closest relatives *Pristionchus expectatus* and *Pristionchus arcanus*. In contrast, genes in p12 and p13 are highly

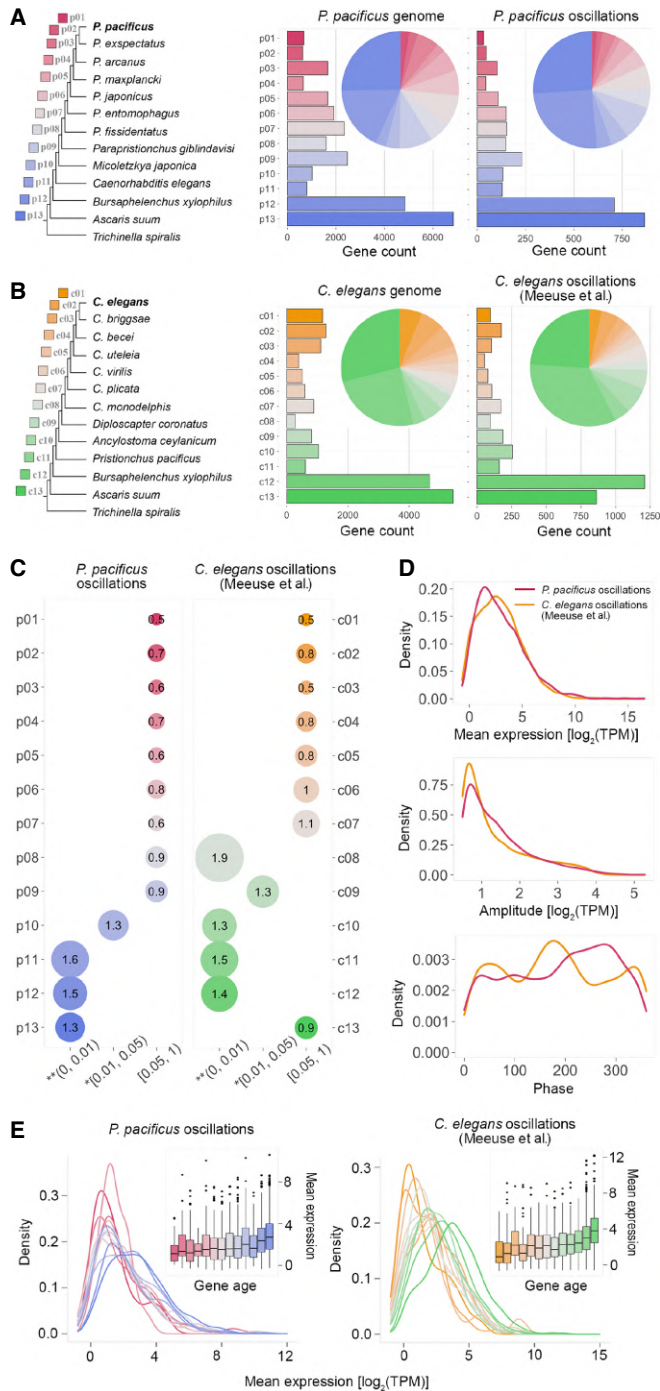


Figure 3. Ancient gene classes are enriched for oscillatory genes. (A) *P. pacificus* genes were assigned to age classes based on the most distant ortholog in a given orthology cluster. The age classes are labeled in a red (young)–blue (old) gradient. The distribution of gene number among the 13 age classes for the *P. pacificus* genome ($n = 28,896$) is shown in the central graph, and the distribution for the *P. pacificus* oscillating genes ($n = 2964$) is shown in the right part. (B) A similar analysis for the *C. elegans* oscillating genes set of Meeuse and coworkers (2020). (C) The graphs indicate the overrepresentation of gene age classes among the *P. pacificus* oscillatory genes (left) and the *C. elegans* oscillating genes sets of Meeuse et al. (2020). The size of the circles indicates the enrichment score; the x-axis shows the significance level measured by Fisher's exact test. (D) The plots show the distribution of average expression (top), amplitude (middle), and phase (bottom) of the oscillating genes in *P. pacificus* (red) and *C. elegans* (yellow). (E) The plots show the distribution of the mean expression of oscillating genes across 13 gene age classes in both species.

conserved and are even shared with distantly related nematodes of clade IV (*Bursaphelenchus xylophilus*) and clade iii (*Ascaris suum*). Gene age analysis indicates that most oscillating genes in *P. pacificus* are highly conserved in evolution (Fig. 3A). Further analysis reveals that oscillating genes in age classes p10, p11, and p12 are indeed strongly overrepresented, whereas younger gene classes contain fewer oscillating genes than expected by chance (Fig. 3C). Next, we reanalyzed the developmental transcriptomes of *C. elegans* and repeated the gene age analysis using seven *Caenorhabditis* species and a related outgroup, including *P. pacificus*, using the data of Kim et al. (2013) and Meeuse et al. (2020; Fig. 3B; Supplemental Fig. S3A). Again, we found that the majority of oscillating genes in *C. elegans* belong to older age classes, in this case in c08–c12 (Fig. 3C; Supplemental Fig. S3B). Unlike the divergence in the phase distribution between *P. pacificus* and *C. elegans*, the distribution of the average expression and amplitude are largely conserved in both species (Fig. 3D; Supplemental Fig. S3C). Genes in different age classes show different mean expression levels in both organisms. Specifically, genes in older age classes show a higher expression level in all three analyzed data sets (Fig. 3E; Supplemental Fig. S3D). Together, these results indicate that the majority of oscillating genes belong to old age classes and are conserved over larger evolutionary distance.

A core network of oscillating genes in both species

The observation that oscillatory genes are highly enriched in ancient gene classes does not necessarily indicate that the gene expression patterns are also conserved. To explicitly test this, we identified orthologous clusters between *P. pacificus* and *C. elegans*. Of the total 28,896 *P. pacificus* genes, only 6594 (22%) were found as 1:1 orthologs of *C. elegans* genes (Fig. 4A; Supplemental Fig. S6A), which reflects the high level of sequence divergence since the separation of both lineages ~100 mya (Prabh et al. 2018; Werner et al. 2018). Of the 2964 oscillating genes, 1002 (33.8%) have one-to-one orthologs (Fig. 4A; Supplemental Fig. S6B). In *C. elegans*, 1285 (34.4%) of the 3739 oscillating genes have one-to-one orthologs (Fig. 4A; Supplemental Fig. S6B). Comparing the 1:1 orthologous genes with oscillatory gene expression patterns in both species

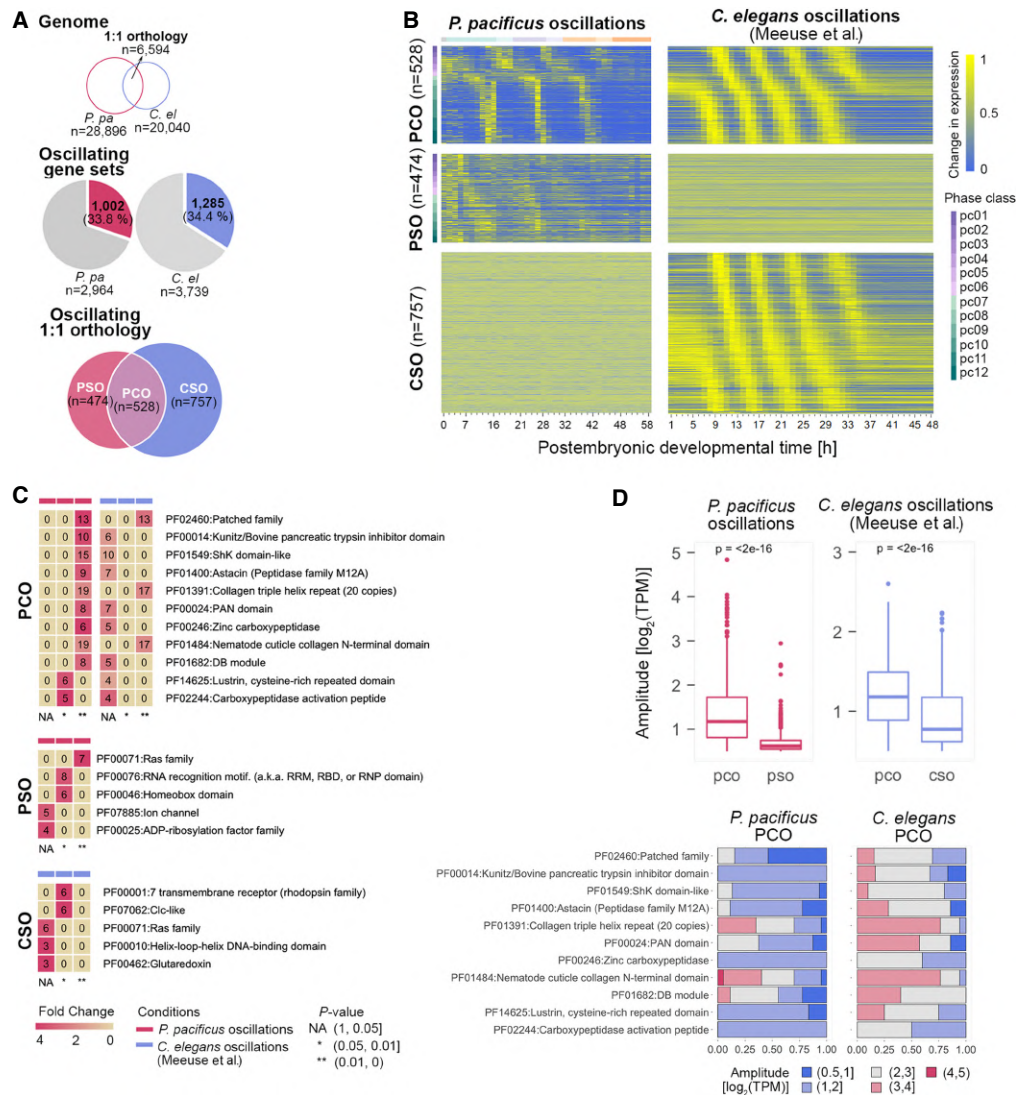


Figure 4. Orthologous oscillatory genes between *P. pacificus* and *C. elegans*. (A) A total of 6594 one-to-one orthologous gene pairs were identified between *P. pacificus* and *C. elegans*. In *P. pacificus*, the oscillating gene set contains 1002 one-to-one orthologous genes; in *C. elegans*, it includes 1285 one-to-one orthologous genes. Of these genes, 528 one-to-one orthologous genes are oscillating in both species (defined as *P. pacificus* conserved oscillations; PCO), 474 orthologous genes show a *P. pacificus*-specific oscillating pattern (PSO), and 757 orthologous genes were classified as *C. elegans*-specific oscillations (CSO). (B) The heatmaps show the developmental expression of the three gene classes (PCO, PSO, and CSO) in *P. pacificus* and *C. elegans*. (C) The matrices show the levels of significance for the tests of protein domain overrepresentation among the three oscillating gene classes. The color scale indicates the fold change; the labeling of matrix cells shows the number of oscillating genes containing a specific protein domain. (D) The box plots display the variation in the amplitude of oscillation between PCO and PSO in *P. pacificus* (red) and between PCO and CSO in *C. elegans* (blue). The bar plots show the distribution of genes with five different amplitude classes among top 11 protein domains highly enriched in PCO.

revealed that 528 (52.7%) of them showed conserved oscillation (Fig. 4A,B; Supplemental Fig. S5A–C). Even for oscillating *P. pacificus* genes with multiple orthologs in *C. elegans*, comparative analysis shows that in around half of the cases, at least one of the orthologs shows shared oscillation in *C. elegans* (Supplemental Fig. S6D,E). This level of conserved gene expression is in strong contrast to a recent analysis of spatial transcriptomes, where only a few dozens of one-to-one orthologs showed shared regional expression between both species (Rödelsperger et al. 2020).

The gene set of one-to-one orthologs with shared oscillatory expression likely represents a highly conserved core network medi-

ating developmental oscillations in nematodes. We defined this gene set as the class of *P. pacificus*–*C. elegans* conserved oscillation (PCO) (Fig. 4A,B; Supplemental Table S7) and tested for overrepresentation of protein domains. This identified Patched receptors as one of the most highly enriched gene families (Fig. 4C). Patched receptors have been shown to have undergone gene family expansions early in the nematode phylum, and RNAi-mediated knock-down experiments in *C. elegans* revealed that they affect multiple aspects of development, including molting (Zugasti et al. 2005). Comparing the amplitude among PCOs, *P. pacificus*-specific oscillation (PSO) and *C. elegans*-specific oscillation (CSO) gene sets

revealed that highly conserved oscillatory patterns (PCO) displayed greater amplitude than both PSO and CSO (Fig. 4D; Supplemental Fig. S5C). The two protein domains collagen and nematode cuticle collagen showed a high percentage of genes with greater amplitudes, reflecting the essential role of collagens in the molting cycle (Fig. 4D). Taken together, our comparative analysis of oscillatory expression patterns identifies gene families with functional relevance in nematode development.

Regional expression analysis reveals a set of collagens with potential function in the cuticle remodeling of the head and tail region during molting

To determine where oscillatory gene expression occurred, we used the spatial transcriptomes of *P. pacificus* to study the spatial expression of oscillating genes (Rödelsperger et al. 2020). In total, we identified 380 genes with oscillating and regional expression (Supplemental Table S11), and these genes cover all 11 regions in *P. pacificus* (Fig. 5A). We found the oscillations presenting in the head region (P1, P2, and P3) and tail region (P11) displayed gene expression with higher mean expression and greater amplitude (Fig. 5A). Overrepresentation analysis of protein domains of these regional oscillatory genes showed the strongest enrichment for collagens. The 11 collagens specifically distributed in the head region (P2 and P3) and preferentially displayed in the molting-related phase classes (PC01, PC04, and PC12) (Fig. 5B,C), suggesting that these regional oscillating collagens presumably have an important function in the cuticle remodeling of head (including mouth-form and pharyngeal) and tail during the molting stages.

The novel gene *eud-1* was integrated as an upstream regulator of multiple oscillatory core genes

The major morphological differences between the two phenotypically plastic mouth-forms in *P. pacificus*, the St and the Eu form, suggest a role of multiple developmental genes in controlling these alternative ontogenic paths. The gene *eud-1*, which is the master regulator of mouth-form plasticity in *P. pacificus* (Ragsdale et al. 2013), is a novel gene that is derived from recent duplications in the diplogastrid lineage (Casasa et al. 2021). *eud-1* showed a graded expression pattern during larval development and is highly expressed at the second molt (Supplemental Fig. S7A). Hypothesizing that *eud-1* may regulate other developmental and possibly also oscillatory genes, we performed SWT on a *eud-1*-mutant line to identify differentially expressed genes at two selected time-points (20 h and 28 h) (Supplemental Fig. S7B). In total, 205 genes were found to be significantly differentially expressed between the wild-type and *eud-1*-mutant conditions (Fig. 6A; Supplemental Table S12). When we compared this gene set with oscillatory genes, we found that 126 (61.5%) differentially expressed genes are indeed oscillatory genes ($P < 2.2 \times 10^{-16}$, Fisher's exact test) (Fig. 6B; Supplemental Fig. S7C). Moreover, 40 members of the oscillatory core network genes are among the set of differentially expressed genes (Fig. 5B). Specifically, 38 collagen genes that are regulated by *eud-1* showed oscillations in 11 different phases, and 11 of these 38 collagens are highly enriched in phase 2, which corresponds to the molting stage (Fig. 6C). These findings may indicate that the *eud-1*-dependent oscillatory collagen genes might represent mouth-form-specific collagens of *P. pacificus*. More generally, these findings imply that the taxonomically restricted *eud-1* gene had been integrated into an ancient regulatory network and controls the expression of highly conserved oscillatory genes.

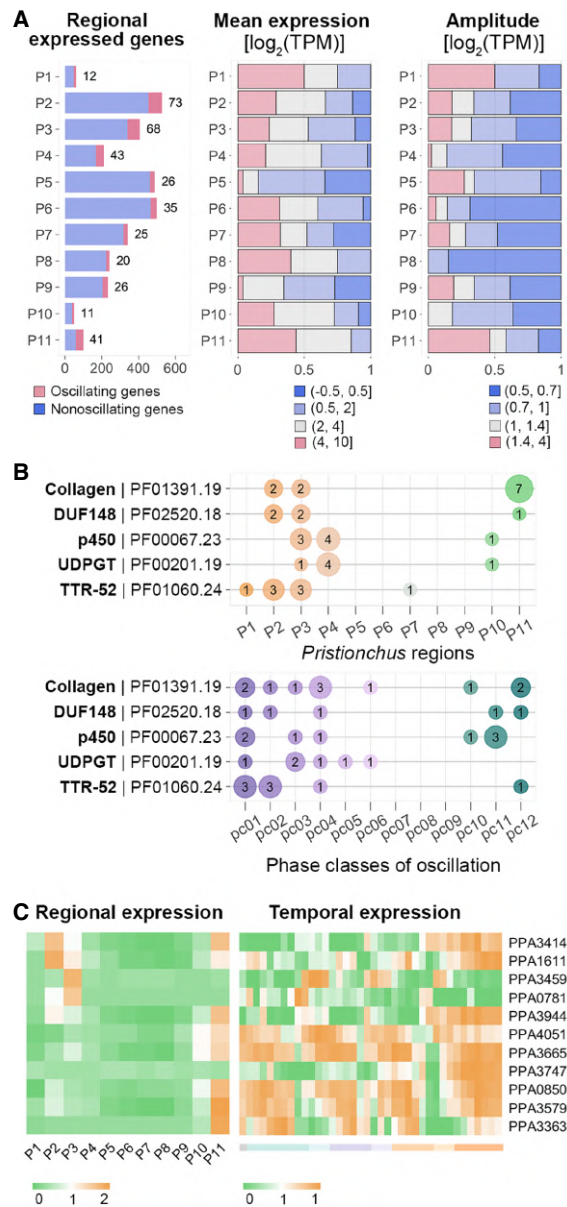


Figure 5. Oscillations are found in multiple *P. pacificus* body regions. (A) The bar plots show the numbers of oscillating genes across the 11 *P. pacificus* regions (left) as defined by spatial transcriptomics, the mean expression level of oscillatory genes with regional expression (middle), and four different amplitude classes (right). (B) The plots show the regional expression and phases of the top five overrepresented protein domains. Circles indicate the number of oscillatory genes with expression in a specific region (upper part) and their assignment into 12 phase classes (lower part). (C) The left heatmap presents the median z-score-normalized expression values of the 12 oscillatory regional collagens across the 11 regions. The heatmap on the right shows their changes in gene expression throughout the entire postembryonic development.

Discussion

In this study, we have implemented SWT in the nematode model organism *P. pacificus*, providing a high-resolution developmental transcriptome with 38 time points from hatching of the J2 larvae

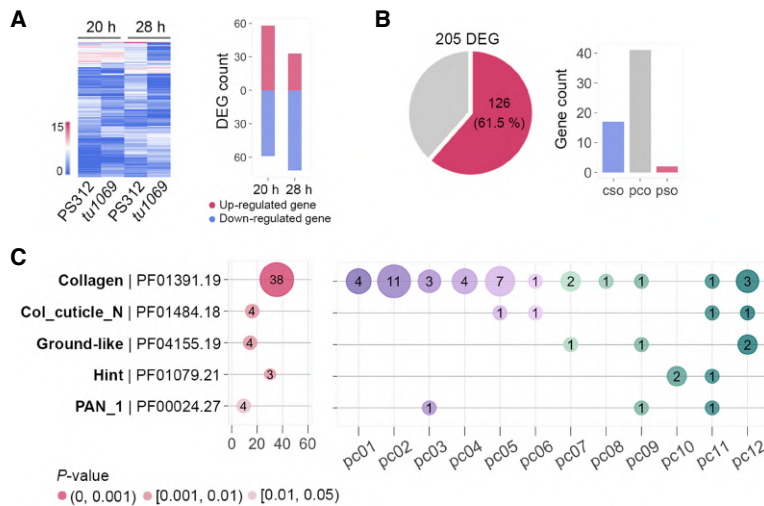


Figure 6. Integration of *eud-1* as upstream regulator of oscillatory core genes. (A) The heatmap shows the expression levels of 205 differentially expressed genes (DEGs) between the *eud-1(tu1069)*-mutant and wild-type (PS312) animals at 20 h and 28 h. The bar plot indicates the number of up-/down-regulated genes. (B) The pie chart displays the overlap between DEGs and oscillating genes. The bar plot shows the number of the oscillating DEGs among the three oscillating gene classes. (C) Protein domain enrichment analysis reveals that the collagen gene family is highly enriched in the 205 DEGs, the x-axis presents the enrichment score, and the y-axis shows the name of protein domains. The size of the circles corresponds to the number of oscillatory genes in a given protein domain, and the three different colors indicate different significance levels as measured by the Fisher's exact test. The right graph profiles the overrepresented protein domains according to the 12 phase classes. The number of oscillatory genes in each phase class is marked.

to young adults. This study combines methods development, comparative analysis, and the evolution of oscillating genes in nematodes. Nematodes are the largest phylum of multicellular animals with an expected number of 1–10 million species (Lambshhead and Boucher 2003). Besides the well-studied model organism *C. elegans*, the nematodes contain many important parasites of plants, livestock, other animals, and humans. However, although many powerful tools including integrative genomic technologies are available in *C. elegans*, similar studies are often hard to perform in other nematodes. The small body size, the often-limited availability of specimen, and the inability to grow pure cultures under laboratory conditions are some of the major factors that make the application of integrative genomic techniques difficult in many species. After the original work of Macchietto and coworkers in *Steinernema* (2017), our study establishes an updated protocol of SWT for an unrelated nematode. This protocol was inspired by our need for high-throughput investigations at lower costs. In particular, our work on phenotypic plasticity requires the parallel investigation of large numbers of samples for environmental perturbation and in experimental evolution approaches. Although our read depth is somewhat lower than in the original protocol, the representation of gene expression is comparable between SWT and STD approaches. Therefore, this protocol will allow scaling up of transcriptomic approaches to large sample sizes at low cost. Also, we are confident that this protocol can be applied to the majority of nematode species, if live material of different post-embryonic stages can be obtained. Such transcriptomic maps may be of importance for developing drug targets against parasites. Additionally, they may potentially help in diagnostics. Besides these technical aspects, this study has two major conclusions.

First, the gene expression atlas of 38 developmental time points provides an important reference point for future studies in

P. pacificus. Given the easiness of the methodology, selected developmental stages of mutant animals can be compared to the wild-type catalog presented in this study. This will represent a powerful tool for many mechanistic studies in which the selection of appropriate developmental stages will be superior over mixed stages. For example, the comprehensive analysis of mouth-form plasticity and its genetic, epigenetic, and environmental regulation has resulted in a large number of mutants that can now be compared by SWT (for review, see Sommer 2020). The differential expression analysis of the mouth-form developmental switch gene *eud-1* (Fig. 5) can serve as a proof of principle for future studies. Similarly, changing abiotic and biotic environmental conditions, including bacterial diet (Werner et al. 2017; Akduman et al. 2020), and their influence on gene expression can now be studied with high accuracy. Thus, our high-resolution catalog of gene expression can assist in many future investigations as well as in the comparative analysis of *P. pacificus* with other organisms.

Second, and most importantly, this study focused on developmental oscillations in gene expression. Although the finding of oscillating gene expression per se is not at all surprising, the comparison of these patterns between *P. pacificus* and *C. elegans* allows for the first time to add an evolutionary dimension to developmental oscillations in nematodes. In general, genomic and transcriptomic comparisons between *P. pacificus* and *C. elegans* build on a large history (Dieterich et al. 2008; Rödelsperger et al. 2014, 2017, 2018; Prabh et al. 2018). But most importantly, the work by Prabh and coworkers (2018) established comparative genomic analysis of six *Pristionchus* species with more distant relatives, including *C. elegans* (Fig. 3A). The associated ladder-like phylogeny allows the distinction of different age classes and the emergence of new genes. Oscillatory genes are highly enriched in ancient gene classes and allowed us to identify a highly conserved core network of 528 one-to-one orthologous genes. This *P. pacificus*–*C. elegans* shared oscillation (PCO) network will be useful for future studies when selecting individual genes for the identification of the regulatory aspects of expression oscillations and their conservation.

Perhaps the most unexpected finding is the diversity of collagen expression in *P. pacificus*. For nematodes with their exoskeleton, collagens are one of the most important gene families. The current 151 collagen genes annotated in the genome can be grouped into two different classes. First, 62 of these genes do not show oscillatory gene expression. Second, among the 89 oscillating collagen genes, subsets expressed in 10 distinct phase classes can be identified (Supplemental Fig. S2). Of the 89 oscillating collagen genes, 38 showed expression peaks during the molting periods (Supplemental Fig. S3F), suggesting that the synthesis of phase-specific collagens could make different contributions to cuticle growth during the molting and intermolt periods. One unexpected finding was the observation that *eud-1*-dependent oscillating collagens are expressed in different phases and that

11 of these collagens show a strong enrichment in molting (phase class 2). Future studies might indicate that such collagens have an important role during mouth-form development in *P. pacificus* and might have a specific function in the stoma. In any case, the comparison of wild-type and *eud-1*-mutant animals reveal the power of the SWT for future studies in developmental biology.

Methods

Maintenance of worm cultures

All studies reported in this manuscript have been conducted by using the wild-type strain *P. pacificus* PS312, originally isolated in Pasadena, California, in 1988 (Sommer et al. 1996). In *P. pacificus*, we use the term juvenile and designate stages as J2, J3, and J4, as typical for nematodes. Note that in *C. elegans*, juveniles are named L1–L4 for historical reasons. Stock cultures of two strains used in this study, PS312 and *eud-1*(*tu1069*), were reared at 20°C on nematode growth medium (NGM) (Sieriebriennikov et al. 2020).

Nematode synchronization and collection

Bleached egg populations were seeded onto plates to obtain synchronized worm cultures (Werner et al. 2017). Synchronized eggs-J1 were observed under the Discovery V20 microscopy and at hatching (Supplemental Fig. S1A) were isolated to a new 6-cm agar plate with 50 μ L OP50 as starting time point. For later juvenile stages, the ecdysis (Supplemental Fig. S1D,G,J) within each molt was chosen as starting time point for animal collection. At the desired time point, single animals were picked into one PCR tube containing 3 μ L nuclear-free water and transferred into liquid nitrogen immediately (Fig. 1C). Such samples could be stored at –80°C for up to a month. In total, worms were collected at 38 time points with three biological replicates, covering the entire postembryonic development and young adulthood (Fig. 1B).

SWT library preparation

We started sample preparation by 3 \times freeze-thawing with liquid nitrogen. Afterward, 3 μ L of lysis buffer (low input cDNA synthesis & amplification module; NEB E6421S) was added, and samples were incubated for 40 min at 65°C, followed by 1 min at 85°C. To remove genomic DNA, we used the TURBO DNA-free kit (Invitrogen AM1907), which allows efficient digestion of DNA within 20 min. Also, the genome digest is performed in the same PCR tube without cleaning up the sample through a column. Importantly, we replaced the 10 \times TURBO DNase buffer with the 5 \times SuperScript II first-strand buffer (Thermo Fisher Scientific, Invitrogen 18064014), owing to the former containing higher concentration of Mg²⁺, which might inhibit the efficiency of reverse transcription. In the next step, we applied RNA-SPRI beads (Agencourt RNAClean XP, protocol 001298v001) to extract mRNA and clean up the samples. A Single Cell/Low Input cDNA Synthesis & Amplification Module (NEB E6421S) was used to generate full-length cDNA and cDNA amplification with 15 cycles of PCR amplification. The Nextera DNA Flex Library Prep Kit (Illumina) was used to prepare the DNA library (Fig. 1C). To avoid potential amplification biases by different PCR cycles, we used 12 PCR cycles for cDNA amplification and eight PCR cycles for DNA (library) amplification in all samples. Together, this protocol contains four major modifications relative to the work of Macchietto et al. (2017) and Chang et al. (2021): (1) We lysed worms through freeze-thawing; (2) we removed genomic DNA and performed this reaction in the same PCR tube as the cDNA synthesis reaction; (3) for the library preparation, we reduced the input of cDNA (to

~1 ng); and (4) we only amplified the final library through eight cycles of PCR reaction. Raw reads of the 114 SWT samples were submitted to the European Nucleotide Archive (see Data access).

STD library preparation

To prepare three STD libraries for comparison to SWT (Fig. 1F), we used the lethargus stage of the fourth molt. We observed individual worms at lethargus (Supplemental Fig. S1H) under the discovery microscopy and picked appropriate animals to one PCR tube containing 3 μ L nuclease-free water. After 3 \times freeze-thawing steps with liquid nitrogen, more than 70 PCR tubes containing single animals at the lethargus stage were pooled into one 1.5-mL tube as one replicate. Three independent biological replicates were prepared in different experiments. To extract RNA, worms suspended in TRIzol were frozen and thawed three times in liquid nitrogen, debris was pelleted for 10–15 min at 14,000 rpm at 4°C, and 200 μ L of chloroform was added to the supernatant. After vigorous vortexing and incubation at room temperature for 5 min, tubes were microfuged for 15 min at 14,000 rpm at 4°C. The aqueous phase was combined with an equal volume of 100% ethanol, RNA was purified using RNA Clean & Concentrator kit (Zymo Research), and its integrity was verified using RNA Nano chips on the Bioanalyzer 2100 instrument (Agilent). We built libraries using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). Raw sequencing of the three newly generated STD samples were submitted to the European Nucleotide Archive (see Data access).

Transcriptome data for comparative analysis across protocols and species

This study contains 38 substages covering the first 58 h after hatching of the J2 larvae with a 1- to 2-h resolution. Three substages in each molt, and seven substages after the final molt provide information about transcriptional changes during the molting stage and within young adult animals (Supplemental Table S1). To estimate the similarities of our SWT and STD libraries, we used three SWT samples of the lethargus during the fourth molt and the three STD samples as described above. To examine how much sequencing depth is required to detect most of the expressed genes for SWT and STD libraries, we selected one library containing more than 12 million read pairs from SWT and STD data set; subsampled the raw reads to nine subsamples with different sequencing depth covering 0.5 million, 1 million, 2 million, 3 million, 4 million, 6 million, 8 million, 10 million, and 12 million; and profiled the distribution of number of detected genes among the nine classes of sequencing depth (Supplemental Fig. 2B). For comparison with *C. elegans*, we used the data sets of Kim et al. (2013) with a time course containing 26 samples at 20°C (sample DH2 for L1–L3, sample DH5 for L4) covering 48 h of postembryonic development (Supplemental Table S1). Raw reads were downloaded from the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena/browser/home>) accession PRJNA212741. Additionally, we compared our data to the one of Meeuse et al. (2020) with a time course including 48 samples at 25°C (TP1–TP13(TC1, L1–L2), TP14–TP48(TC2, L1–YA)) (Supplemental Table S1). Raw reads were obtained from the NCBI Sequence Read Archive (SRA; <https://www.ncbi.nlm.nih.gov/sra>) accession SRP195783.

RNA-seq read alignment and estimation of relative abundance

All libraries were sequenced at 150-bp PE on an Illumina HiSeq 3000. Raw reads were trimmed by the program cutadapt (Martin 2011) version 2.10 with parameters “-q 30,25 --gc-content = 50, --minimum-length 25:25.” Read pairs were aligned by the HISAT2

(version 2.1.0) (Kim et al. 2015) software to the *P. pacificus* genome assembly (version El Paco) using the gene annotations (version El Paco gene annotation 3) as additional information. HISAT2 was run with the additional parameter "--rna-strandness RF." Mapped reads were summarized to the genomic features by the featureCounts (version 1.6.4) with parameters "-a -F GTF -g Parent -d 50 -D 500." Ninety-five percent expressed genes with 12 million PE reads were detected with at least 2 million PE reads; we removed the six libraries with fewer than 2 million PE reads in the further analysis, including ENA samples ERS5595844, ERS5595845, ERS5595849, ERS5595862, ERS5595942, ERS5595945. The six time points (1 h, 3 h, 5 h, 13 h, 52 h, and 54 h) contained two replicates; another 32 time points included three replicates with enough sequencing depth. Gene expression value, transcripts per million (TPM), was computed through normalization by library size and feature effective length. The mean TPM was calculated from three biological replicates; 23,135 REGs were identified as their mean PE reads are greater than three in at least one of the 38 time points (Supplemental Table S2).

Raw reads of the two *C. elegans* time courses were mapped to the *C. elegans* genome (WS275) using HISAT2 (version 2.1.0) with parameters "--rna-strandness F." Mapped reads were summarized to the longest isoforms by the featureCounts (version 1.6.4) (Liao et al. 2014) with parameters "-a -F GTF -g Parent -M." Gene expression value (TPM) was calculated by the package *t-arae/ngscmdr* (version 0.1.0.181203) in R (R Core Team 2020). Genes with more than 10 read counts in at least one time point were named as REGs, 17,446 REGs and 17,409 REGs were identified in the study of Kim et al. (2013) and Meeuse et al. (2020), respectively.

Pairwise correlation analyses and principal component analysis

log₂-transformed gene expression levels of 23,135 REG were used to compute the pairwise correlations between transcriptomes at different developmental time points using the R command `cor` (data, use = "pairwise.complete.obs", method = "Pearson"). To further show the strong temporal relationship between our samples, we performed a PCA using the function `princomp` (Sigg and Buhmann 2008) in R with default parameters. The loadings corresponding to the PC2, PC3, and PC4 appeared to be periodical patterns.

Classification of oscillatory genes by Meta2D

To classify the oscillatory genes in *P. pacificus*, we applied the MetaCycle (Wu et al. 2016) package (version 1.2.0) in R, 20 time points from 5 h to 44 h covering the entire J3 stage with 2-h intervals were picked to predict oscillation genes. The parameters (cycMethod = c("ARS", "JTK"), minper = 7, maxperiod = 13, ARSdefaultPer = 10, weightedPerPha = TRUE) were used to perform the meta2d algorithm. We predicted the oscillations for the Kim et al. (2013) data sets with the parameters (cycMethod = c("ARS", "JTK"), minper = 6, maxperiod = 12, ARSdefaultPer = 9, weightedPerPha = TRUE) considering this time course under 20°C. We reanalyzed the oscillating genes in the Meeuse et al. (2020) data sets by using the same parameters. In total, we identified 2964 oscillations in *P. pacificus* (Supplemental Table S3) and 2221 oscillations in the Kim et al. (2013) data set (Supplemental Table S4) by using the two filters, amplitude ≥ 0.5 and FDR < 0.05. 3739 oscillations that classified by Meeuse et al. (2020) were used to perform further comparison.

Identification of gene expression peaks

To identify the peaks of temporal gene expression during the time-course, we applied the "findpeaks" function (with nups = 2, ndowns = 2) of the package "pracma" (version 2.2.9; #37) in R. To quantify the number of the expression peaks coinciding

with the molting cycle, we defined the molting periods as the intervals spanning 1 h before until 1 h after molting. These time points are for the second molt "TP-2M" (16 h, 17 h, 18 h, 19 h, and 20 h), the third molt "TP-3M" (28 h, 29 h, 30 h, 31 h, 32 h), and the fourth molt "TP-4M" (42 h, 43 h, 44 h, 45 h, 46 h). Similarly, they are for the intermolt periods "TP-2I" (3 h, 5 h, 7 h, 9 h, 11 h, 13 h, 15 h), "TP-3I" (21 h, 22 h, 24 h, 26h) and "TP-4I" (34 h, 36 h, 38 h, 40 h). Time points for young adult (TP-YA) include 48 h, 50 h, 52 h, and 54 h. To study the distribution of oscillation peaks during three molting events, we defined seven different classes based on the peak distribution. C1, C2, or C3 indicates oscillating genes that show only one peak during TP-2M, TP-3M, or TP-4M, respectively. C4, C5, or C6, represents oscillating genes that display two peaks during TP-2M and TP-3M, TP-2M and TP-4M, or TP-3M and TP-4M, respectively. C7 indicates oscillating genes that show three peaks during the three molting periods.

Pfam domain prediction and overrepresentation analysis in *P. pacificus* oscillation genes

To predict Pfam domains of all genes in the *P. pacificus* genome, we applied the HMMER (Mistry et al. 2013) version 3.3 with parameters "hmmsearch --tblout Pfam-A.hmm"; 17,664 of all 28,896 genes contain predicted Pfam domains (*e*-value of full seq < 0.01) (Supplemental Table S8). Overrepresented Pfam domains in the set of oscillating genes were determined by calculating the fold enrichment of the number of overlapping genes compared to what to expect by chance given the number of genes in a particular Pfam domains and the number of oscillating genes with only expressed genes considered. To minimize the large enrichments that would otherwise be caused by Pfam domains with small numbers of genes, we added a pseudocount of 12 to the total number of genes for a given Pfam domain before calculating the actual ratio. The *P*-value was generated by the Fisher's exact test.

Gene age analysis

For *C. elegans*, we performed gene age analysis by obtaining annotated protein sets for *C. elegans* (WS275), *Caenorhabditis briggsae* (WS275), *Caenorhabditis becei* (QG2083_v1), *Caenorhabditis uteleia* (JU2585_v1), *Caenorhabditis virilis* (JU1968_v1), *Caenorhabditis plicata* (SB355_v1), *Caenorhabditis monodelphis* (JU1667_v1), *Diploscaptor coronatus* (WBPS14), *Ancylostoma ceylanicum* (WS248), *P. pacificus* (El_Paco_v3), *B. xylophilus* (WS248), *A. suum* (WS248), and *Trichinella spiralis* (WS248) from WormBase (<https://wormbase.org>) and from <http://www.pristionchus.org>. The set of *C. elegans* proteins (the longest isoform per gene) was taken as query sequences, and gene age classes were defined from the OrthoFinder (version 2.3.12) (Emms and Kelly 2020), DIAMOND (version 0.9.31, with parameters BLASTP -e 0.001) (Buchfink et al. 2015) searches against all other species and clustered orthogroups by the MCL algorithm (Enright et al. 2002) with default parameters; 27,132 (93.9%) *P. pacificus* annotated genes and 18,750 (93.6%) *C. elegans* annotated genes were assigned into orthogroups, respectively (Supplemental Fig. S3E). Further, we used a phylostratigraphy approach (Domazet-Lošo et al. 2007; Prabh et al. 2018; Prabh and Rödelisperger 2019; Rödelisperger et al. 2020), to trace the origin of the assigned genes on the phylogeny in *P. pacificus* and *C. elegans*. Based on the presence of the most distant outgroup species in a given orthologous cluster, we identified 13 gene age classes in both species (Fig. 3A,B; Supplemental Tables S5, S6). Note that singletons or the unsigned genes in orthogroup (6.1% of *P. pacificus* genes and 6.4% of *C. elegans* genes) were removed from this analysis considering they are

different from any gene age class as in previous analysis (Prabh et al. 2018). Gene age analysis for *P. pacificus* was performed using the nine data sets of predicted proteins from the recently sequenced diplogastrid genomes (Prabh et al. 2018) and proteins from *C. elegans*, *B. xylophilus*, *A. suum*, and *T. spiralis*.

One-to-one ortholog prediction between *P. pacificus* and *C. elegans*

For *C. elegans*, 20,040 protein sequences were extracted from the gene annotation (WS275) by taking the longest isoform of each annotated gene. Similarly, 28,896 protein sequences of the last version of *P. pacificus* (El_Paco_v3) were used to identify the one-to-one orthologs. We applied the OrthoFinder (version 2.3.12) program to prediction, DIAMOND (version 0.9.31, with parameters BLASTP -e 0.001) searches against protein data from both species were performed, and the orthogroups were clustered by the MCL algorithm with default parameters. In total, 6594 one-to-one orthologs, 529 one-to-many (*Ppa* vs. *Cel*) orthologs, 756 one-to-many (*Cel* vs. *Ppa*) orthologs, and 187 many-to-many (*Ppa* vs. *Cel*) orthologs were classified from the resulting orthologous clusters.

Gene Ontology (GO) analysis for orthologous oscillatory genes

We applied the DAVID (Dennis et al. 2003) v6.8 web-accessible programs to do the Gene Ontology (GO) analysis by using the UniProt accessions of the *C. elegans* genes, as obtained from the gene annotation file (WS275). For UniProt accession annotation of *P. pacificus* genes, we first applied DIAMOND (version 0.9.31, with parameters BLASTP -e 0.001) searches against protein sequences from both species and obtained the best-reciprocal pairs with lowest cross-species e-value, and then we annotated the UniProt accession of *P. pacificus* gene based on the UniProt accession of *C. elegans* gene that is the best-reciprocal hit. The GO analysis includes biological process (BP), molecular function (MF), and cellular component (CC), and we used gene enrichment and EASE scores from the output of DAVID for visualization (Supplemental Fig. S4D; Supplemental Table S9).

Regional expression analysis of oscillatory genes

The spatial transcriptomic analysis defined 11 regions across the anterior posterior axis, which correspond to anatomical structures such as head, intestine, germline, and tail (Rödelsperger et al. 2020). The corresponding gene sets were defined based on a relative enrichment over the mean expression (z -score > 1). As the spatial transcriptome was analyzed for a reference annotation that was based on a strand-specific transcriptome assembly (Rödelsperger et al. 2016, 2018), we applied the following set of rules to assign the genes of the last version of *P. pacificus* (El_Paco_v3): (1) applied DIAMOND (version 0.9.31, with parameters blastx -e 0.00001) searches against protein sequences of the last version of *P. pacificus* (El_Paco_v3) and obtained the best-hits with the lowest e-value, (2) used BEDTools (version 2.26.0, with parameters intersect -wa -wb -s) to confirm the same genomic location, and (3) selected the longest predicted transcript as the final corresponding locus to the El_Paco version 3. This could automatically assign 3155 of the 3502 regional genes to the corresponding gene identifiers in the latest annotation (El_Paco gene annotation, version 3) (Supplemental Table S10). We identified 380 *P. pacificus* genes with oscillatory and regional expression by overlapping the 2964 oscillations in this study and 3155 regional expression genes (Supplemental Table S11).

Differential expression analysis

To identify differentially expressed genes regulated by *eud-1*, we picked two time points, 20 h (the first hour after the second

molt) and 28 h (the last hour before the third molt), to perform the gene expression comparison between the *eud-1(tu1609)* mutant and wild type (PS312). These experiments were performed independent of the wild-type SWT for logistic reasons. Differential expression analysis was performed in R (version 4.0.3) using DESeq2 (version 1.18.1) (Love et al. 2014). We applied an adjusted *P*-value cut-off of 0.05 and a fold change cutoff of two to identify the DEG genes. Raw reads of the six SWT samples of the *eud-1(tu1069)* mutant were submitted to the European Nucleotide Archive (see Data access).

Data access

Raw and processed data sets from this study have been submitted to the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena>) under accession numbers PRJEB42613, PRJEB42633, and PRJEB42635.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank the members of the Sommer laboratory for thoughtful comments of experiments, results, and interpretation. We thank Dr. Wen-Sui Lo for her assistance with the bioinformatic analysis and Dr. Kohta Yoshida for his help with the SWT protocol. S.S. was funded by a fellowship from the Chinese Sponsorship Council (CSC). This work was funded by the Max-Planck Society.

Author contributions: S.S. and R.J.S. conceived and designed all experiments. S.S. conducted all experimental studies. S.S. and C.R. performed all bioinformatic analysis. All three authors contributed to writing.

References

- Akduman N, Lightfoot JW, Röseler W, Witte H, Lo WS, Rödelsperger C, Sommer RJ. 2020. Bacterial vitamin B₁₂ production enhances nematode predatory behavior. *ISME Journal* **14**: 1494–1507. doi:10.1038/s41396-020-0626-2
- Athanasouli M, Witte H, Weiler C, Loschko T, Eberhardt G, Sommer RJ, Rödelsperger C. 2020. Comparative genomics and community curation further improve gene annotations in the nematode *Pristionchus pacificus*. *BMC Genomics* **21**: 708. doi:10.1186/s12864-020-07100-0
- Baskaran P, Rödelsperger C, Prabh N, Serobyany V, Markov G, Hirsekorn A, Dieterich C. 2015. Ancient gene duplications have shaped developmental stage-specific expression in *Pristionchus pacificus*. *BMC Evol Biol* **15**: 185. doi:10.1186/s12862-015-0466-2
- Bento G, Ogawa A, Sommer RJ. 2010. Co-option of the hormone-signalling module dafachronic acid-DAF-12 in nematode evolution. *Nature* **466**: 494–497. doi:10.1038/nature09164
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59–60. doi:10.1038/nmeth.3176
- Cao J, Packer JS, Ramani V, Cusanovich DA, Huyenh 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**: 661–667. doi:10.1126/science.aam8940
- Casasa S, Biddle JF, Koutsovoulos GD, Ragsdale EJ. 2021. Polyphenism of a novel trait integrated rapidly evolving genes into ancestrally plastic networks. *Mol Biol Evol* **38**: 331–343. doi:10.1093/molbev/msaa235
- Chang D, Serra L, Lu D, Mortazavi A, Dillman A. 2021. A revised adaptation of the smart-Seq2 protocol for single-nematode RNA-seq. *Methods Mol Biol* **2170**: 79–99. doi:10.1007/978-1-0716-0743-5_6
- Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* **4**: R60. doi:10.1186/gb-2003-4-9-r60
- Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, Dinkelacker I, Fulton L, Fulton R, Godfrey J, Minx P, et al. 2008. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat Genet* **40**: 1193–1198. doi:10.1038/ng.227

- Domazet-Lošo T, Brajković J, Tautz D. 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet* **23**: 533–539. doi:10.1016/j.tig.2007.08.014
- Emms D, Kelly S. 2020. OrthoFinder manual: accurate inference of orthologues and orthogroups made easy! *Genome Biol* **20**: 238.
- Enright AJ, Dongen Sv, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucl. Acid Res.* **30**: 1575–1584. doi:10.1093/nar/30.7.1575
- Félix MA, Hill RJ, Schwarz H, Sternberg PW, Sudhaus W, Sommer RJ. 1999. *Pristionchus pacificus*, a nematode with only three juvenile stages, displays major heterochronic changes relative to *Caenorhabditis elegans*. *Proc Roy Soc B* **266**: 1617–1621. doi:10.1098/rspb.1999.0823
- Fürst von Lieven A. 2005. The embryonic moult in diplogastrids (Nematoda): homology of developmental stages and heterochrony as a prerequisite for morphological diversity. *Zool Anz* **244**: 79–91. doi:10.1016/j.jcz.2005.05.001
- Gerstein MB, Rozowsky J, Yan K-K, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Jessica Li J, et al. 2014. Comparative analysis of the transcriptome across distant species. *Nature* **512**: 445–448. doi:10.1038/nature13424
- Han Z, Lo WS, Lightfoot JW, Witte H, Sun S, Sommer RJ. 2020. Improving transgenesis efficiency and CRISPR-associated tools through codon optimization and native intron addition in *Pristionchus* nematodes. *Genetics* **216**: 947–956. doi:10.1534/genetics.120.303785
- Hendriks GJ, Gaidatzis D, Aeschimann F, Großhans H. 2014. Extensive oscillatory gene expression during *C. elegans* larval development. *Mol Cell* **53**: 380–392. doi:10.1016/j.molcel.2013.12.013
- Kanzaki N, Giblin-Davis RM. 2015. Diplogastrid systematics and phylogeny. In *Pristionchus pacificus*, a nematode model for comparative and evolutionary biology (ed. Sommer RJ), pp. 43–76. Brill, Leiden.
- Kanzaki N, Herrmann M, Weiler C, Röseler W, Theska T, Berger J, Rödelsperger C, Sommer RJ. 2021. Nine new *Pristionchus* (nematoda: Diplogastridae) species from China. *Zootaxa* **4943**: zootaxa.4943.1.1. doi:10.11646/zootaxa.4943.1.1
- Kim DH, Grün D, van Oudenaarden A. 2013. Dampening of expression oscillations by synchronous regulation of a microRNA and its target. *Nat Genet* **45**: 1337–1344. doi:10.1038/ng.2763
- Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**: 357–360. doi:10.1038/nmeth.3317
- Lambshhead PJD, Boucher G. 2003. Marine nematode deep-sea biodiversity: hyperdiverse or hype? *J Biogeogr* **30**: 475–485. doi:10.1046/j.1365-2699.2003.00843.x
- Levin M, Hashimshony T, Wagner F, Yanai I. 2012. Developmental milestones punctuate gene expression in the *Caenorhabditis* embryo. *Dev Cell* **22**: 1101–1108. doi:10.1016/j.devcel.2012.04.004
- Liao Y, Smyth GK, Shi W. 2014. Sequence analysis feature counts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930. doi:10.1093/bioinformatics/btt656
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. doi:10.1186/s13059-014-0550-8
- Macchietto M, Angdemby D, Heidarpour N, Serra L, Rodriguez B, El-Ali N, Mortazavi A. 2017. Comparative transcriptomics of *Steinernema* and *Caenorhabditis* single embryos reveals orthologous gene expression convergence during late embryogenesis. *Genome Biol Evol* **9**: 2681–2696. doi:10.1093/gbe/evx195
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**: 10–12. doi:10.14806/ej.17.1.200
- Meeuse MW, Hauser YP, Morales Moya LJ, Hendriks G, Eglinger J, Bogaarts G, Tsiarlis C, Großhans H. 2020. Developmental function and state transitions of a gene expression oscillator in *Caenorhabditis elegans*. *Mol Syst Biol* **16**: e9498. doi:10.15252/msb.209975
- 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**: e121. doi:10.1093/nar/gkt263
- Nakayama K-I, Ishita Y, Chihara T, Okumura M. 2020. Screening for CRISPR/Cas9-induced mutations using a co-injection marker in the nematode *Pristionchus pacificus*. *Dev Genes Evol* **230**: 257–264. doi:10.1007/s00427-020-00651-y
- 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. doi:10.1534/g3.119.400326
- 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**: 1664–1674. doi:10.1101/gr.234971.118
- Ragsdale EJ, Müller MR, Rödelsperger C, Sommer RJ. 2013. A developmental switch coupled to the evolution of plasticity acts through a sulfatase. *Cell* **155**: 922–933. doi:10.1016/j.cell.2013.09.054
- R Core Team. 2020. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26. doi:10.1038/nbt.1754
- Rödelsperger C, Neher RA, Weller AM, Eberhardt G, Witte H, Mayer WE, Dieterich C, Sommer RJ. 2014. Characterization of genetic diversity in the nematode *Pristionchus pacificus* from population-scale resequencing data. *Genetics* **196**: 1153–1165. doi:10.1534/genetics.113.159855
- Rödelsperger C, Meyer JM, Prabh N, Lanz C, Bemm F, Sommer RJ. 2017. Single-molecule sequencing reveals the chromosome-scale genomic architecture of the nematode model organism *Pristionchus pacificus*. *Cell Rep* **21**: 834–844. doi:10.1016/j.celrep.2017.09.077
- 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**: 3123–3127.e5. doi:10.1016/j.cub.2018.07.041
- Rödelsperger C, Athanasouli M, Lenuzzi M, Theska T, Sun S, Dardiry M, Wighard S, Hu W, Sharma DR, Han Z. 2019a. Crowdsourcing and the feasibility of manual gene annotation: a pilot study in the nematode *Pristionchus pacificus*. *Sci Rep* **9**: 18789. doi:10.1038/s41598-019-55359-5
- Rödelsperger C, Prabh N, Sommer RJ. 2019b. New gene origin and deep taxon phylogenomics: opportunities and challenges. *Trends Genet* **35**: 914–922. doi:10.1016/j.tig.2019.08.007
- Rödelsperger C, Ebbing A, Sharma DR, Okumura M, Sommer RJ, Korswagen HC. 2020. Spatial transcriptomics of nematodes identifies sperm cells as a source of genomic novelty and rapid evolution. *Mol Biol Evol* **38**: 229–243. doi:10.1093/molbev/msaa207
- Sieriebriennikov B, Prabh N, Dardiry M, Witte H, Röseler W, Kiener MR, Rödelsperger C, Sommer RJ. 2018. A developmental switch generating phenotypic plasticity is part of a conserved multi-gene locus. *Cell Rep* **23**: 2835–2843.e4. doi:10.1016/j.celrep.2018.05.008
- Sieriebriennikov B, Sun S, Lightfoot JW, Witte H, Moreno E, Rödelsperger C, Sommer RJ. 2020. Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell. *PLoS Genet* **16**: e1008687. doi:10.1371/journal.pgen.1008687
- Sigg CD, Buhmann JM. 2008. Expectation-maximization for sparse and non-negative PCA. In *Proceedings of the 25th International Conference on Machine Learning*, pp. 960–967. Association for Computing Memory, New York.
- Sommer RJ. 2020. Phenotypic plasticity: from theory and genetics to current and future challenges. *Genetics* **215**: 1–13. doi:10.1534/genetics.120.303163
- Sommer RJ, Carta LK, Kim SY, Sternberg PW. 1996. Morphological, genetic and molecular description of *Pristionchus pacificus* sp. n. (Nematoda: Neodiplogastridae). *Fund Appl Nemat* **19**: 511–521.
- Susoy V, Ragsdale EJ, Kanzaki N, Sommer RJ. 2015. Rapid diversification associated with a macroevolutionary pulse of developmental plasticity. *eLife* **4**: e05463. doi:10.7554/eLife.05463
- Susoy V, Herrmann M, Kanzaki N, Kruger M, Nguyen CN, Rödelsperger C, Röseler W, Weiler C, Giblin-Davis RM, Ragsdale EJ, et al. 2016. Large-scale diversification without genetic isolation in nematode symbionts of figs. *Sci Adv* **2**: e1501031. doi:10.1126/sciadv.1501031
- Wang X, Sommer RJ. 2011. Antagonism of LIN-17/frizzled and LIN-18/RyK in nematode vulva induction reveals evolutionary alterations in core developmental pathways. *PLoS Biol* **9**: e1001110. doi:10.1371/journal.pbio.1001110
- 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**: 7207. doi:10.1038/s41598-017-07455-7
- 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**: 1675–1687. doi:10.1101/gr.234872.118
- 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**: 55–62. doi:10.1007/s00427-014-0486-8
- Wood W. 1988. Introduction to *C. elegans* biology. In *The nematode Caenorhabditis elegans* (ed. Wood WB). Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Wu G, Anafi RC, Hughes ME, Kornacker K, Hogenesch JB. 2016. MetaCycle: an integrated R package to evaluate periodicity in large scale data. *Bioinformatics* **32**: 3351–3353. doi:10.1093/bioinformatics/btw405
- Zugasti O, Rajan J, Kuwabara PE. 2005. The function and expansion of the patched- and hedgehog-related homologs in *C. elegans*. *Genome Res* **15**: 1402–1410. doi:10.1101/gr.3935405

Received January 25, 2021; accepted in revised form July 14, 2021.



Single worm transcriptomics identifies a developmental core network of oscillating genes with deep conservation across nematodes

Shuai Sun, Christian Rödelberger and Ralf J. Sommer

Genome Res. 2021 31: 1590-1601 originally published online July 22, 2021
Access the most recent version at doi:[10.1101/gr.275303.121](https://doi.org/10.1101/gr.275303.121)

Supplemental Material <http://genome.cshlp.org/content/suppl/2021/08/23/gr.275303.121.DC1>

References This article cites 52 articles, 8 of which can be accessed free at:
<http://genome.cshlp.org/content/31/9/1590.full.html#ref-list-1>

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Affordable, Accurate
Sequencing.

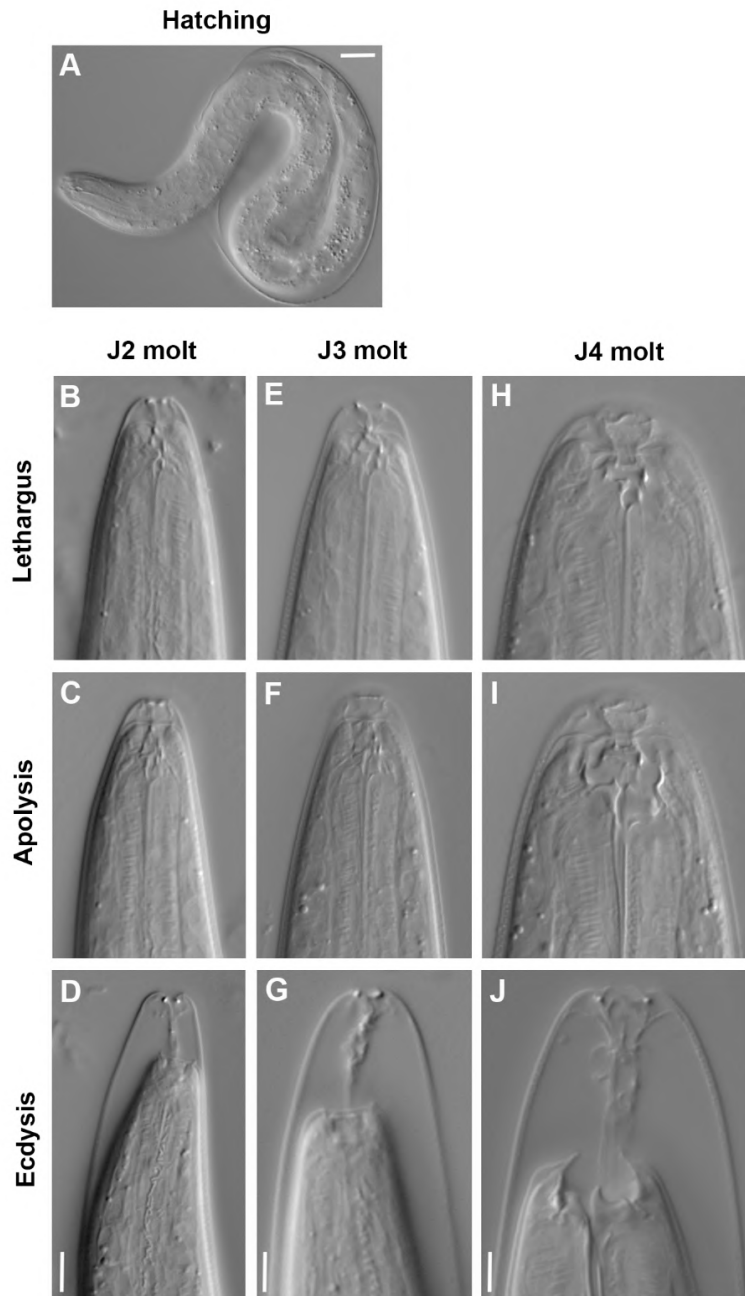


To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

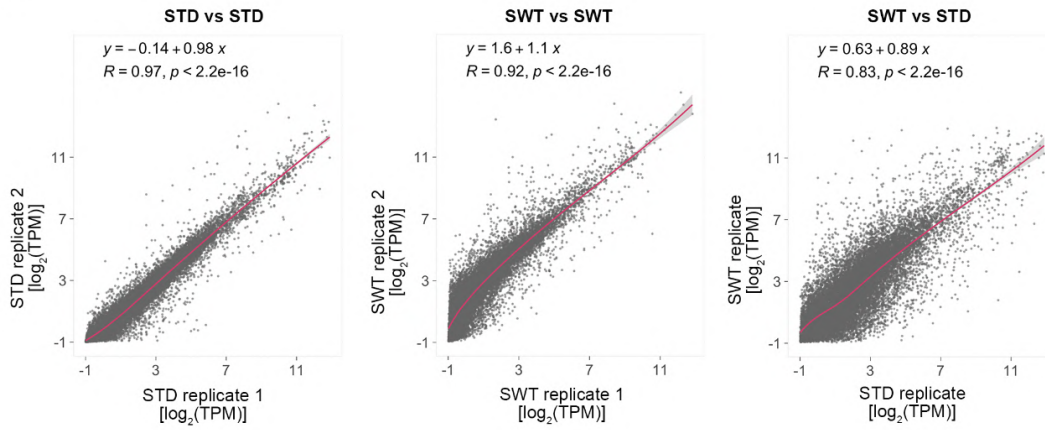
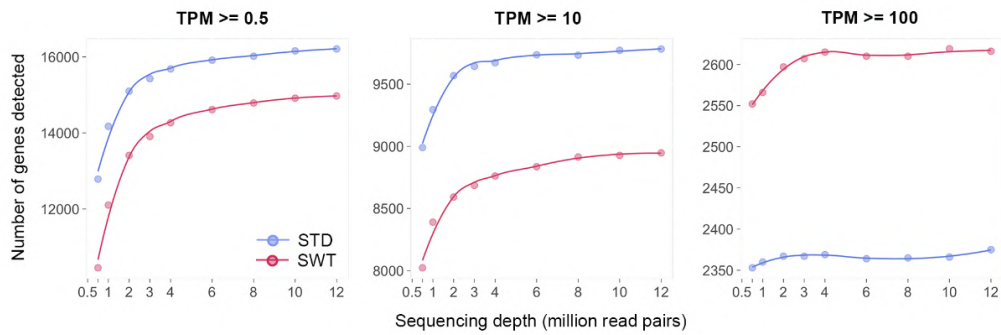
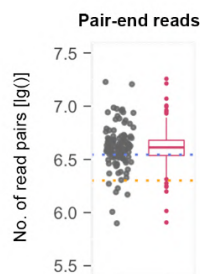
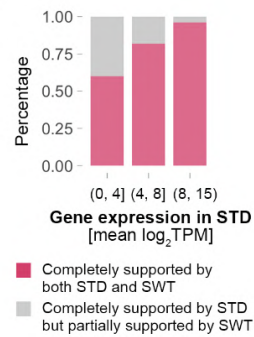
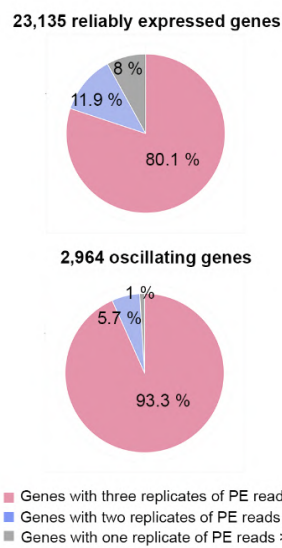
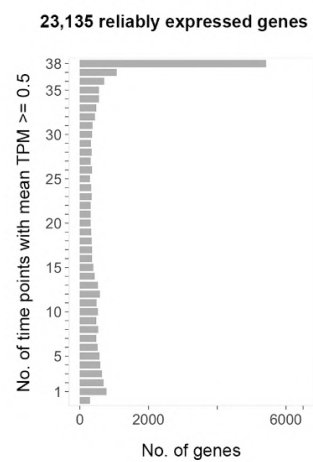
Supplementary Material

**Single worm transcriptomics identifies a developmental core network
of oscillating genes with deep conservation across nematodes**

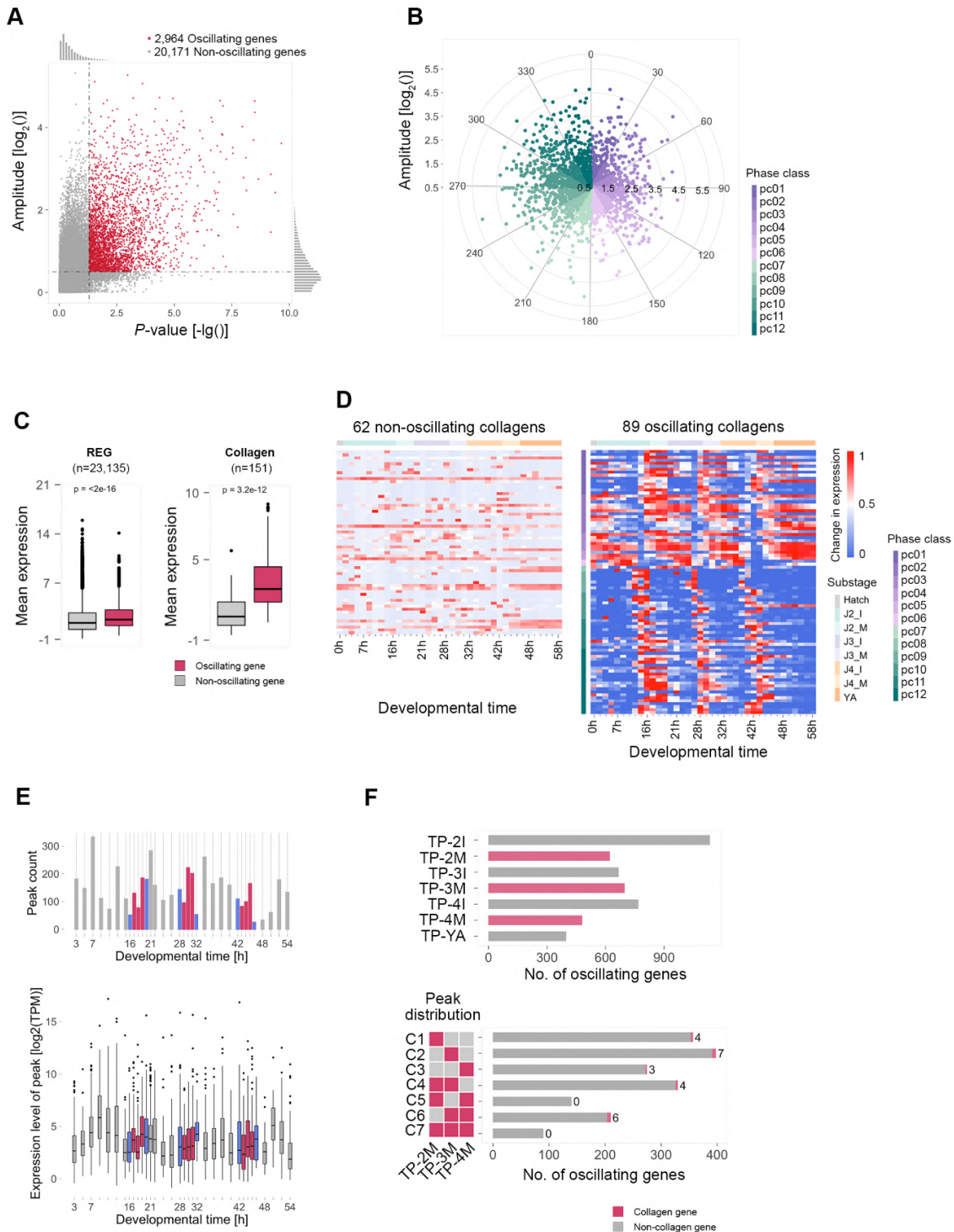
Shuai Sun, Christian Rödelsperger, and Ralf J. Sommer



Supplemental Figure S1. Developmental markers were used to define the duration of juvenile stage in *P. pacificus*, related to Fig 1. Images showing the head part of the animals at the hatching (A), the lethargus (a period of inactivity during which pharyngeal pumping, feeding, and locomotion become arrested) of the 2nd, 3rd, 4th molt (B, E, H), the apolysis of the 2nd, 3rd, 4th molt (C, F, I), and the ecdysis (rupture and shedding of the old cuticle) (Lee 2002) of the 2nd, 3rd, 4th molt (D, G, J), scalar bar is 10 μ m.

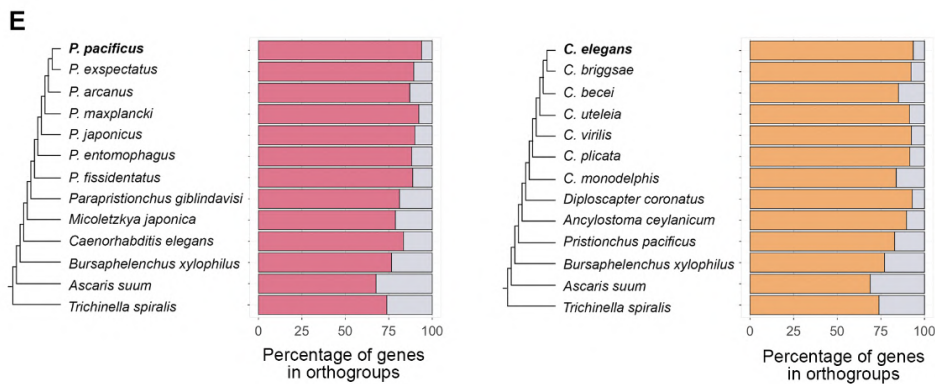
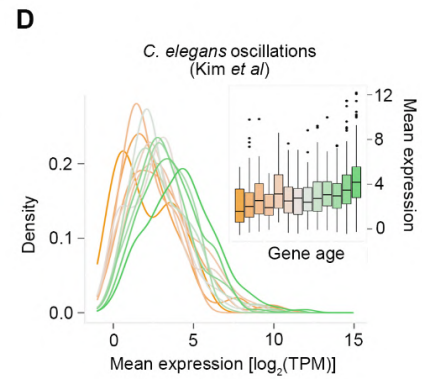
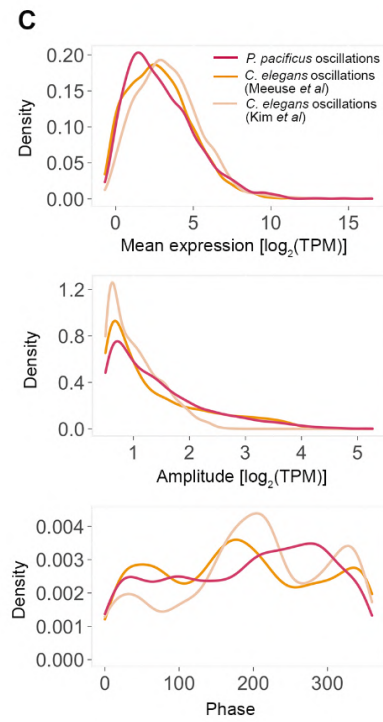
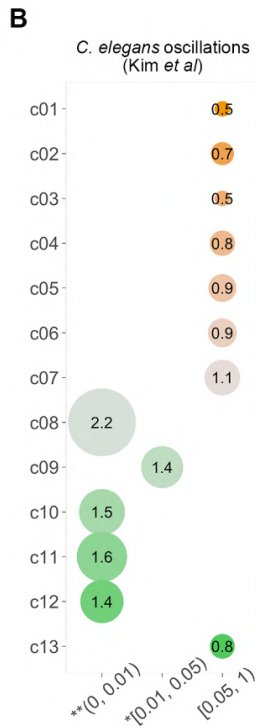
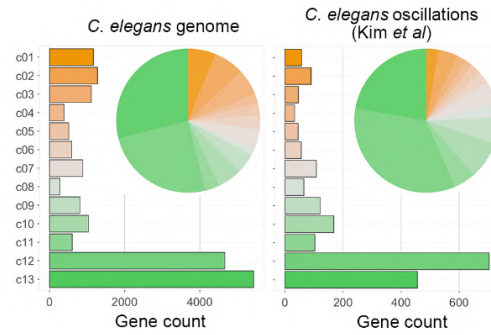
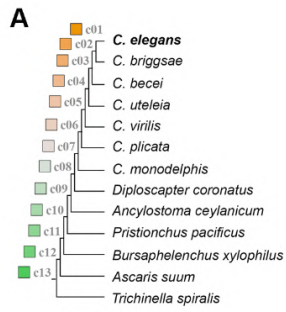
A**B****C****D****E****F**

Supplemental Figure S2. Evaluation of single worm transcriptomic (SWT) libraries, related to Fig 1. (A) The scatter plots present the $\log_2(\text{TPM})$ of all genes between individual STD replicates (left), SWT replicates (middle), and STD replicate and SWT replicate (right). The replicates of STD and SWT are at the lethargus of the fourth molt. (B) The line plots show the number of detected genes as a function of sequencing depth for individual STD and SWT replicates with more than 12 million reads. Subsampling of reads was done for nine levels of sequencing depth and the number of detected genes was assessed for three different expression levels (TPM ≥ 0.5 (left), TPM ≥ 10 (middle), TPM ≥ 100 (right))- (C) The boxplot displays the number of read pairs (\log_{10}) for all 114 samples, the blue line corresponds to the mean sequencing depth of 4.5 M read pairs, the orange line shows the sequencing depth of 2 M read pairs. (D) The bar plot shows the percentage of expressed genes in STD samples with completely (red) or partially (grey) supported by SWT-seq, 100 genes were randomly chosen from three different gene groups with distinct gene expression level in the STD samples (Supplemental Table S13) and the degree of gene structure support was manually assessed after inspection of alignment files in a genome browser. (E) The pie charts present the degree of the genes with three replicates of TPM ≥ 1 (red), the genes with two replicates of TPM ≥ 1 (blue), the genes with one replicate of TPM ≥ 1 (grey), in the 23,135 reliably expression gene set (upper) and the 2,964 oscillating gene set (lower). (F) The barplot shows the number of time points where the reliably expressed genes displayed the TPM is equal or greater than 0.5.

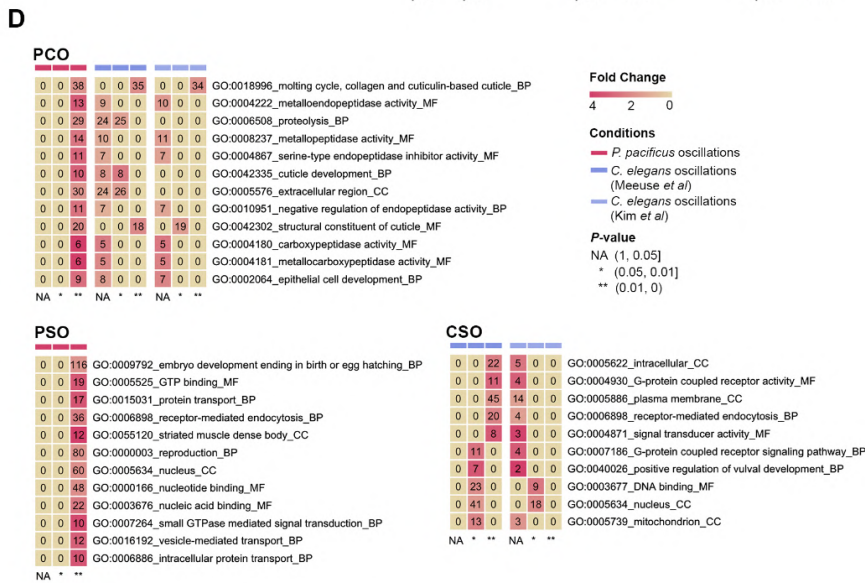
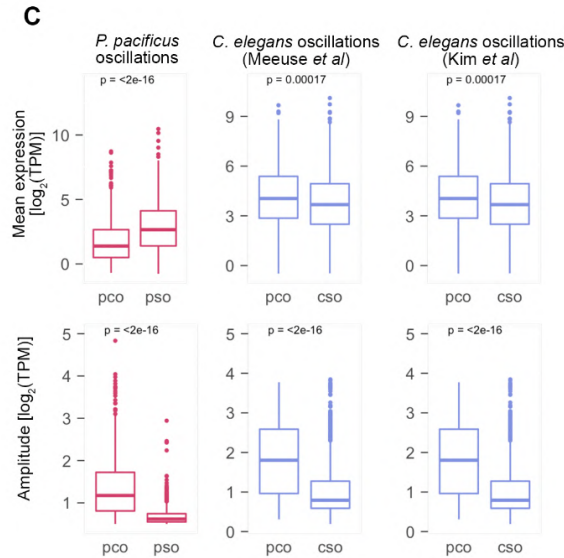
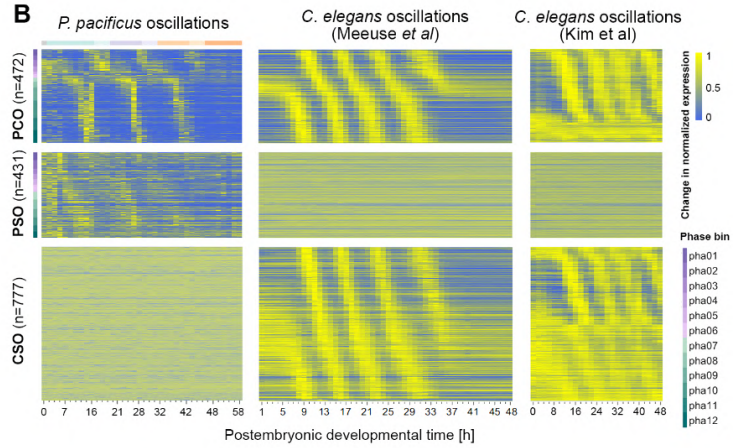
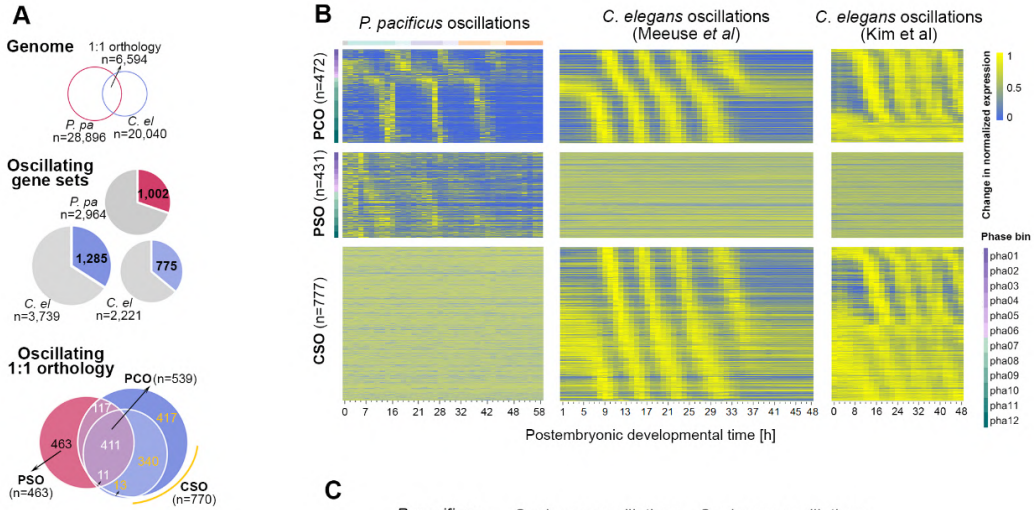


Supplemental Figure S3. Thousands of oscillating genes are synchronized with the molting cycle in *P. pacificus*, related to Fig 2. (A) Scatter plot classifying genes with oscillating expression in red (n=2,964). Two indicators, $\log_2(\text{meta2d_amplitude}) \geq 0.5$ and $P\text{-value} < 0.05$,

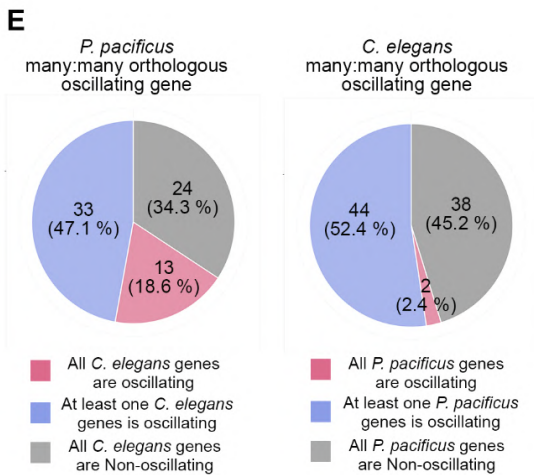
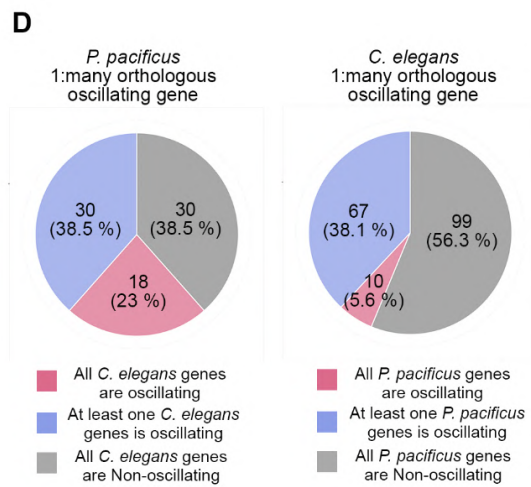
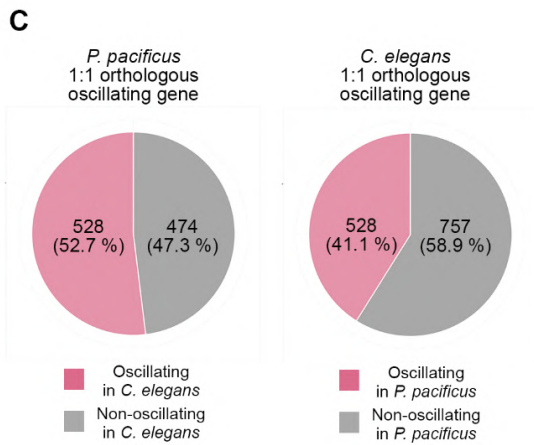
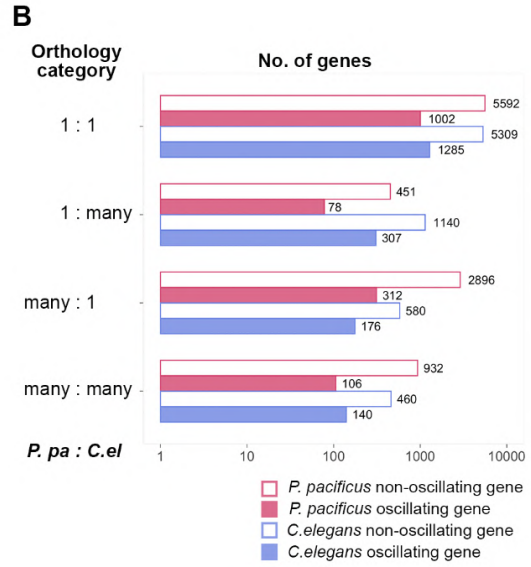
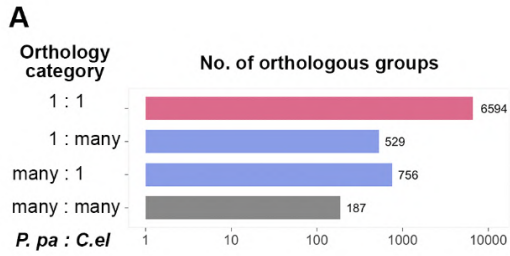
were used for gene classification. Genes below either indicator were grouped into the “non-oscillating genes” (grey, n=20,171). (B) Radar chart profiling oscillation amplitude over the phase of peak expression. Each point indicates the \log_2 -transformed amplitude, and phase classes. (C) The boxplots show the variation in mean expression of genes between oscillating and non-oscillating groups in the 23,135 REG gene set (left) and in the 151 collagen family (right). (D) The heatmaps visualize the changes in gene expression traces throughout development for non-oscillating collagens (n=62, right) and oscillating collagens (n=89, right). The mean TPM was calculated from three biological replicates and was normalized by its maximum of each expression trace. The order of rows in the heatmap with the oscillating genes was sorted by phase class, which indicated a phase difference of 30, corresponding to a peak shift by roughly one hour. (E) The upper barplot shows the number of peaks across 34 time points. The three substages in each molt are colored in red, and the hour before and after molt are defined as molting periods (see Materials and Methods) in blue. The lower boxplot profiles the distribution of the gene expression levels at the time point of peaks among 34 time points. (F) The upper barplot displays the number of the oscillation genes which exhibit expression peaks during the molting periods, inter-molting periods. The lower barplot shows the number of collagens (red) and non-collagens (grey) among seven peak classes (C1-C7, see Materials and Methods for details), The color code in the heatmap, red means peak and grey means not peak.



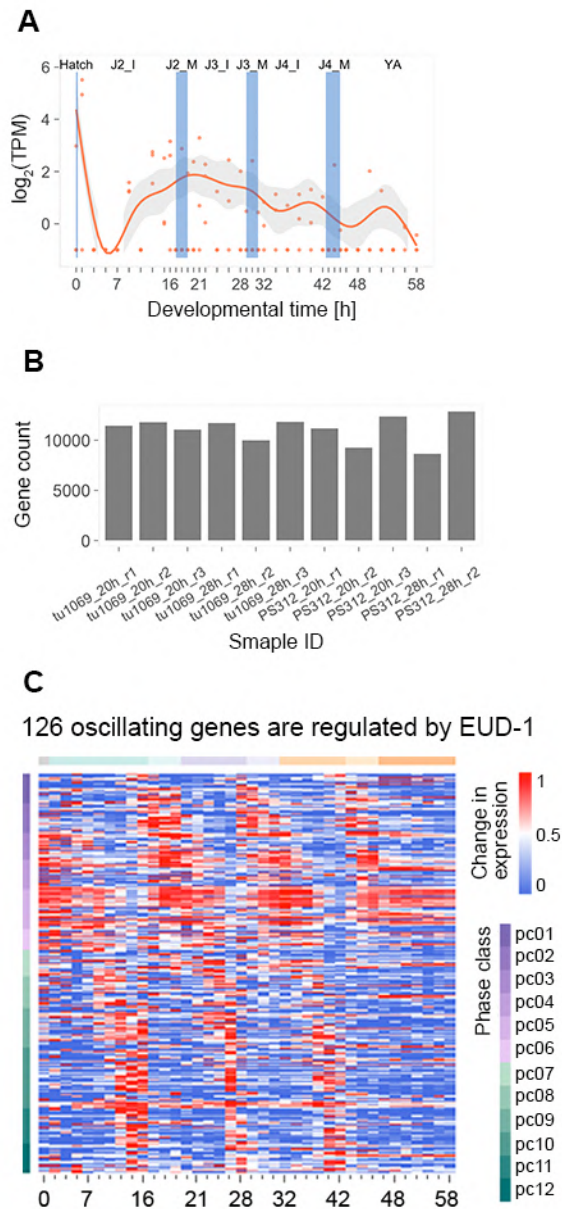
Supplemental Figure S4. Ancient gene classes are enriched for oscillatory genes, related to Fig 3. (A) *C. elegans* genes were assigned to age classes based on the most distant ortholog in a given orthology cluster. The age classes are labeled in a yellow (young) - green (old) gradient. The distribution of gene number among the 13 age classes for the *C. elegans* genome (n = 20,040) is shown in the central graph, and the distribution for the *C. elegans* oscillating genes of Kim et al (2013) (n = 2,221) is shown in the right part. (B) The graphs indicate the overrepresentation of gene age classes among the *C. elegans* oscillating genes sets of Kim et al (2020). The size of the circles indicates the enrichment score, the x-axis shows the significance level measured by Fisher's exact test. (C) The plots show the distribution of average expression (top), amplitude (middle), and phase (bottom) of the oscillating genes in *P. pacificus* (red), the *C. elegans* oscillating genes of Meeuse et al (2020) (dark yellow), and the *C. elegans* oscillating genes of Kim et al (2013) (light yellow). (D) The plots show the distribution of the mean expression of the *C. elegans* oscillating genes of Kim et al (2013) across 13 gene age classes. (E) The barplots visualize the percentage of genes in orthogroups (assigned genes) across 13 species on the phylogeny in *P. pacificus* (left) and *C. elegans* (right). 27,132 (93.9 %) *P. pacificus* annotated genes and 18,750 (93.6 %) *C. elegans* annotated genes were assigned into orthogroups, respectively.



Supplemental Figure S5. Orthologous oscillatory genes among three oscillating gene sets, related to Fig 4 (A) A total of 6,594 one-to-one orthologous gene pairs were identified between *P. pacificus* and *C. elegans*. In *P. pacificus*, the oscillating gene set contains 1,002 one-to-one orthologous genes, in *C. elegans* oscillating genes of Meeuse et al (2020) it includes 1,285 one-to-one orthologous genes, and in *C. elegans* oscillating genes of Kim et al (2013) it includes 775 one-to-one orthologous genes. Of these genes, 539 one-to-one orthologous genes are oscillating in both species (defined as *P. pacificus* conserved oscillations, 'PCO'), 463 orthologous genes show *P. pacificus*-specific oscillating pattern (PSO), and 770 orthologous genes were classified as *C. elegans*-specific oscillations (CSO). (B) The heatmaps show the developmental expression of the three gene classes (PCO, PSO and CSO) in *P. pacificus* and *C. elegans*. (C) The box plots display the variation in the mean expression (top) and the amplitude (bottom) of oscillation between PCO and PSO in *P. pacificus* (red) and between PCO and CSO in *C. elegans* (blue). (D) The matrices show the levels of significance for the tests of GO (gene ontology) overrepresentation among the three oscillating gene classes. The color scale indicates the fold change, the labeling of matrix cells shows the number of oscillating genes containing a specific GO term, the three sub-ontologies containing biological process (BP), molecular function (MF), and cellular component (CC).




Supplemental Figure S6. Evolution of oscillating expression of orthologous groups, related to Fig 4 (A) Four orthologous group classes were identified between *P. pacificus* and *C. elegans*. (B) The barplots show the distribution of *P. pacificus* oscillation genes (red) and *C. elegans* oscillation genes (blue) among the four orthologous group classes. (C) 469 of one-to-one orthologous groups exhibit a shared oscillatory pattern. (D) The pie charts show the degree of shared oscillation for one-to-many orthologous groups. (E) The pie charts show the degree of shared oscillation for many-to-many orthologous groups.



Supplemental Figure S7. 205 oscillating genes are regulated by the *eud-1*, related to Fig 5.

(A) The dynamic expression of the *eud-1* gene during postembryonic stages. *eud-1* gene exhibited a graded expression pattern with highly expressed during late J2 and entire J3 stage rather than an oscillating pattern. (B) The bar plot shows that enough expressed genes were detected (> 8,500 expressed genes) in every sample indicating that the sequencing depth was sufficient for DEG analysis. (C) The heatmap shows the dynamic expression pattern of the 126 oscillations that are regulated by the *eud-1* locus.

The oscillating Mucin-type protein DPY-6 has a conserved role in nematode mouth and cuticle formation

Shuai Sun ^{1,†}, Tobias Theska ^{1,†}, Hanh Witte¹, Erik J. Ragsdale ², and Ralf J. Sommer ^{1,*}

¹Department for Integrative Evolutionary Biology, Max Planck Institute for Biology Tübingen, 72076 Tübingen, Germany,

²Department of Biology, Indiana University, Bloomington, IN 47405, USA

[†]These authors contributed equally to the work.

*Corresponding author: Department for Integrative Evolutionary Biology, Max Planck Institute for Biology Tübingen, Max-Planck Ring 9, Tübingen 72076, Germany. Email: ralf.sommer@tuebingen.mpg.de

Abstract

Nematodes show an extraordinary diversity of mouth structures and strikingly different feeding strategies, which has enabled an invasion of all ecosystems. However, nearly nothing is known about the structural and molecular architecture of the nematode mouth (stoma). *Pristionchus pacificus* is an intensively studied nematode that exhibits unique life history traits, including predation, teeth-like denticle formation, and mouth-form plasticity. Here, we used a large-scale genetic screen to identify genes involved in mouth formation. We identified *Ppa-dpy-6* to encode a Mucin-type hydrogel-forming protein that is macroscopically involved in the specification of the cheilostom, the anterior part of the mouth. We used a recently developed protocol for geometric morphometrics of miniature animals to characterize these defects further and found additional defects that affect mouth form, shape, and size resulting in an overall malformation of the mouth. Additionally, *Ppa-dpy-6* is shorter than wild-type with a typical *Dumpy* phenotype, indicating a role in the formation of the external cuticle. This concomitant phenotype of the cheilostom and cuticle provides the first molecular support for the continuity of these structures and for the separation of the cheilostom from the rest of the stoma. In *Caenorhabditis elegans*, *dpy-6* was an early mapping mutant but its molecular identity was only determined during genome-wide RNAi screens and not further investigated. Strikingly, geometric morphometric analysis revealed previously unrecognized cheilostom and gymnostom defects in *Cel-dpy-6* mutants. Thus, the Mucin-type protein DPY-6 represents to the best of our knowledge, the first protein involved in nematode mouth formation with a conserved role in cuticle deposition. This study opens new research avenues to characterize the molecular composition of the nematode mouth, which is associated with extreme ecological diversification.

Keywords: nematode stoma; *Pristionchus pacificus*; *Caenorhabditis elegans*; *dpy-6*; developmental plasticity; evo devo

Introduction

Nematodes are extremely abundant, estimated to account for 80% of all land animals and are found in all ecosystems and on all continents (van den Hoogen et al. 2019). Nonetheless, their cylindrical body is surprisingly uniform with the exception of the head and mouth region (stoma), which has undergone striking adaptations during evolution (Wright 1976; Malakhov 1994). This resulted in a diversity of feeding structures and strategies in free-living and predatory species as well as in parasites of plants, animals, and humans. For example, the human hook worm *Ancylostoma duodenale* has a mouth structure with 2 plates that form 2 strong teeth each, which enable the parasite to attach to the villi in the small intestine causing severe tissue damage (Anderson 2000). These and other examples indicate that the mouth and head structures are the site of major evolutionary innovation in nematodes. However, little is known about the structural architecture of the nematode mouth and associated genetic and molecular processes in any species. This is largely because in *Caenorhabditis elegans*, the prime nematode model organism, the

mouth structures have found little attention except for associated neuronal and behavioral traits. Also, the stoma is formed in parts by miniature cells that escape visual observation by differential interference contrast (DIC) microscopy, the usual microscopical technique used in *C. elegans* research (Hall and Altun 2008; Burr and Baldwin 2016).

In general, the cuticle of nematodes is formed by a syncytial layer of cells, inconsistently named epidermal or hypodermal (Wright 1976; Lee 2002). The nematode cuticle is an exoskeleton that in most species consists of 3 layers of fibrillar proteins with many evolutionary adaptations. In *C. elegans*, the major proteins in the cuticle are collagen-like proteins encoded by a large gene family with more than 150 members (Cohen and Sundaram 2020). When mutated, some of the collagen-encoding genes or collagen-modifying enzymes result in a *Dumpy* (shorter than wild-type) (*Dpy*) phenotype. There are more than 30 *Dpy* mutants in *C. elegans*, which have played an important role for genetic mapping (Brenner 1974). Other components of the cuticle are cuticulins, glycoproteins, and lectins, many of which have been

Received: June 15, 2021. Accepted: December 13, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America. All rights reserved.

For permissions, please email: journals.permissions@oup.com

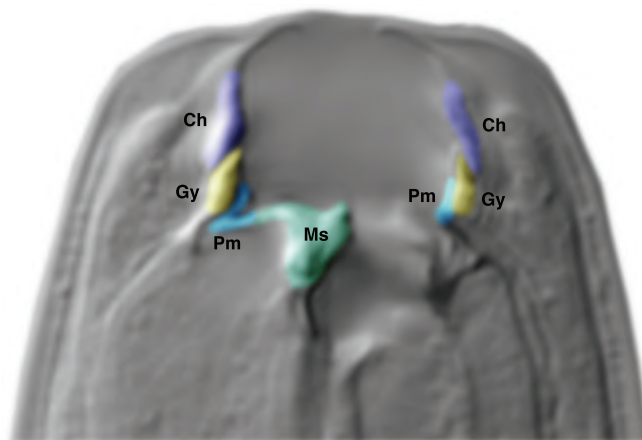
investigated in parasitic species as potential drug targets (Page et al. 2014). Thus, a substantial knowledge exists about the composition of the nematode cuticle, which is in strong contrast to the composition of the mouth.

The mouth of nematodes is the major opening of the cuticle but traditionally, the mouth is considered to be fully separated from the external cuticle (Wright 1976). This view is supported by the fact that no defects in mouth structures have been reported in *C. elegans* Dpy mutants. In general, the stoma consists of the “buccal cavity” as the lumen that extends from the mouth opening to the anterior end of the pharynx and the “buccal capsule” as its cuticular lining (Fig. 1) (Fürst von Lieven and Sudhaus 2000). The cellular architecture underlying these structures has been elucidated from transmission electron microscopy after original work in *C. elegans* (Albertson and Thomson 1976; Wright and Thomson 1981). Many structural modifications have been identified from comparative studies, which are in parts controversial as the identity and homology of epidermal and pharyngeal muscle cells that constitute the buccal capsule remains uncertain (De Ley et al. 1995; Baldwin et al. 1997; Fürst von Lieven and Sudhaus

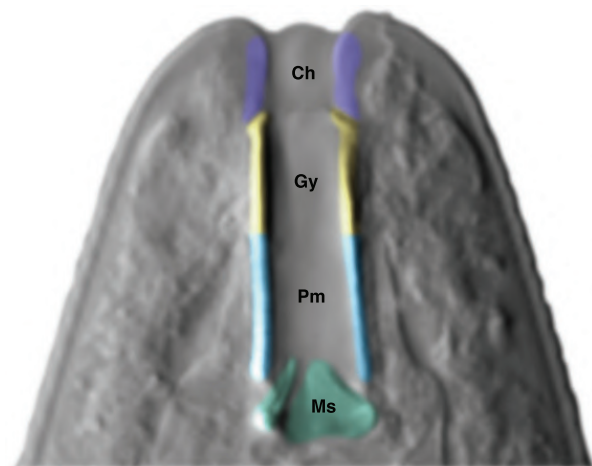
2000; Ragsdale and Baldwin 2010). In the buccal capsule 6 regions can be distinguished (Fig. 1) (De Ley et al. 1995). First, the anterior most “cheilostom” is surrounded by labial cells and is thought to be continuous with the cuticle of the worm. According to this interpretation, the cheilostom is separated from the more posterior parts of the buccal cavity. However, no molecular support for this hypothesis is available and no *C. elegans* cuticle mutants are known to have defects in the cheilostom. Second, the “gymnostom” is the middle part of the buccal cavity and is surrounded by arcade epidermis. Finally, the “stegostom” is the posterior-most part of the buccal cavity, which is further subdivided into “pro-”, “meso-”, “meta-”, and “telo-”stegostom (De Ley et al. 1995). This region is formed by interradiial muscle cells and often overlapping because these regions are telescoped. Importantly, the stegostom is a region of major evolutionary diversification as it forms teeth-like denticles in *Pristionchus pacificus* and other Diplogastriidae that enable predation (Fig. 1, a and c) (Fürst von Lieven and Sudhaus 2000; Susoy et al. 2015).

Pristionchus pacificus was introduced as a model organism for comparative and evolutionary studies with genetic and

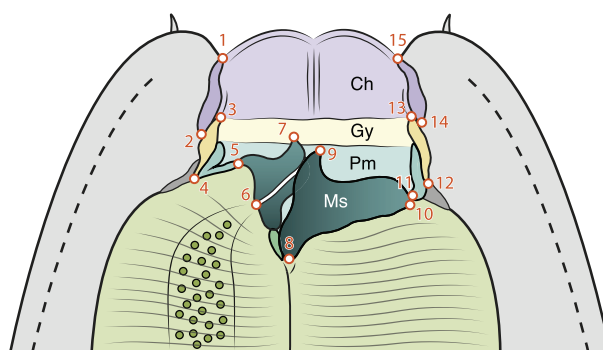
(a) Microscopic image (DIC) of *Pristionchus pacificus*



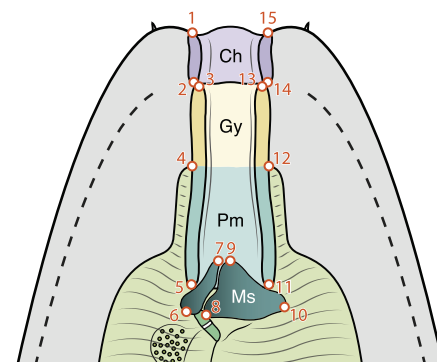
(b) Microscopic image (DIC) of *Caenorhabditis elegans*



(c) Landmark configuration in *Pristionchus pacificus*



(d) Landmark configuration in *Caenorhabditis elegans*



■ Cheilostom (Ch) ■ Gymnostom (Gy) ■ Promesostegostom (Pm) ■ Metastegostom (Ms) ○ Landmark

Fig. 1. Structural overview of the nematode stoma. a) DIC image of the *Pristionchus pacificus* mouth in right-lateral perspective (sagittal plane). b) DIC image of the *Caenorhabditis elegans* mouth in right-lateral perspective (sagittal plane). c) Schematic illustration of the *P. pacificus* mouth in right-lateral perspective, including the landmark configuration used for geometric morphometric analysis. d) Schematic illustration of the *C. elegans* mouth in right-lateral perspective, including the landmark configuration used for geometric morphometric analysis.

molecular tools in the mid-1990s (Sommer et al. 1996). This self-fertilizing species has a 4-day generation time when fed *Escherichia coli* under laboratory conditions and has a technological platform related to *C. elegans*. Its genome was originally sequenced in 2008 with single molecule re-sequencing facilitating reverse genetic and genome editing (Dieterich et al. 2008; Rödelsperger et al. 2017; Han et al. 2020; Nakayama et al. 2020). In the last decade, much research on *P. pacificus* focused on the formation of teeth-like denticles in the mouth that result in intra-guild predation (Bento et al. 2010; Quach and Chalasani 2020). Importantly, these teeth-like denticles represent an example of phenotypic plasticity as they are expressed in 2 alternative forms. The “stenostomatous” (St) mouth form has a single dorsal tooth with a narrow buccal cavity, whereas the “eurystomatous” (Eu) mouth form has 2 teeth and a broader buccal cavity (Fig. 1, a and c). Specifically, Eu animals have a large right ventrosublateral tooth, curved dorsal tooth, and the anterior tip of their promesostegostom is located posterior to the anterior tip of their gymnostom, whereas the anterior tips of both elements are approximately on the same level in St animals. While Eu animals are omnivorous feeders that can use bacteria, fungi, and nematodes as food source, St animals are strict bacterial feeders. Many studies have identified (1) the environmental influences, (2) the developmental switches, and (3) the gene regulatory network associated with mouth-form plasticity and the associated evolutionary processes (Ragsdale et al. 2013; Werner et al. 2017; Sieriebriennikov et al. 2018, 2020; Bui and Ragsdale 2019; Casasa et al. 2021). Here, we present a large-scale genetic screen to identify genes involved in mouth formation and identified *Ppa-dpy-6* to encode a Mucin-type gel-forming protein that is involved in the specification of the stoma.

Materials and methods

Maintenance of worm cultures and genetic crosses

Stock cultures of all strains were reared at room temperature (20–25°C) on nematode growth medium (NGM) in 6 cm Petri dishes, as outlined before (Sommer et al. 1996). *Escherichia coli* OP50 was used as food source. Bacteria were grown overnight at 37°C in LB medium, and 400 µl of the overnight culture was pipetted on NGM agar plates and left for several days at room temperature to grow bacterial lawns. Nematodes were passed on these lawns and propagated by passing various numbers of mixed developmental stages.

Mouth-form phenotyping

Adult hermaphrodites were immobilized on 5% Noble Agar pads with 0.3% NaN₃ added as an anaesthetic, and examined using DIC microscopy. Animals that had a large right ventrosublateral tooth, curved dorsal tooth, and the anterior tip of the promesostegostom posterior to the anterior tip of the gymnostom were classified as Eu morphs. Animals that did not exhibit these 3 characters simultaneously were classified as St morphs.

Worm size measurement

To study body length of wild-type and mutant animals, we used 7 development stages after bleaching of J1/eggs: 24 h (J2), 36 h (J3), 48 h (J4), 72 h (1-day-old adults), 96 h (2-day-old adults), 120 h (3-day-old adults), 144 h (4-day-old adults), and performed between 10 and 25 replicates. We found a strong increase in body size during the 3 juvenile stages (J2, J3, and J4) but little growth during adulthood (Supplementary Fig. 3a). The maximum body length

(1.06 mm) was reached at 4 days adulthood. For the comparison of body length between *Dpy* mutants and wild-type, we singled out the 5-day-old adults to NGM plates without bacteria. Bright field images of the worms were taken using 0.63× objective of ZEISS SteREO Discovery V16 and the AxioCam camera with parameters “magnification 40×, field of view 5.8 mm, resolution 1.1 µm, depth of field 26 µm.” Images were analyzed using the Wormsizer plug-in for Image J/Fiji (Schindelin et al. 2012). The raw data of worm size measurement in this study can be found in Supplementary Data 1.

Quantification of morphological change in mutant animals

We quantified morphological differences in the mouths of wild-type and mutant worms, using our recently published protocol (Theska et al. 2020), which combines landmark-based geometric morphometrics with model-based clustering (a detailed account of our analysis is provided in the Supplementary Material). In short, we obtained image stacks of the nematode mouths in lateral position and recorded the X and Y coordinates of 15 fixed landmarks (Fig. 1, c and d and Supplementary Table 4) which capture the overall form of the mouth using FIJI (ver. 2.1.0) (Schindelin et al. 2012). We generated landmark data for 193 animals and created 2 separate data sets: (1) a *P. pacificus* (exclusively Eu animals) data set ($n=118$) comprising the wild-type strain (PS312), a frameshift mutant of *Ppa-dpy-6* (*tu1645*), and a “rescue” strain RS4003 (*tu1696*); and (2) a *C. elegans* data set ($n=75$) comprising the wild-type N2 and the *Cel-dpy-6* mutant CB14. All steps of the geometric morphometric analysis of mouth form, shape, and size were performed in R (ver. 4.1.1) (R Core Team 2021), using a combination of the GEOMORPH (ver. 4.0.1) (Adams and Otárola-Castillo 2013; Adams et al. 2021) and MORPHO (ver. 2.9) (Schlager et al. 2017, 2021) packages. Shape and form differences were visualized using PCA; differences in mouth size were approximated by the centroid size and visualized in box-plots. To test whether mouth form, shape, and/or size changed due to a mutation in *dpy-6*, we performed distance-based permutational MANOVA (PERMANOVA) (Anderson 2001, 2014) with a randomized residual permutation procedure (RRPP) (Collyer et al. 2015; Collyer and Adams 2018, 2021), using linear model fits produced with the *procD.lm* function of GEOMORPH. Pairwise comparisons among strains were performed using the *pairwise* function of the RRPP package (ver. 1.1.0; Collyer et al. 2015, 2018, 2021). To identify groups of specimens with similar morphology, we performed model-based clustering with the *Mclust* function of the MCLUST R package (ver. 5.4.7) (Li and Durbin 2009; Fraley et al. 2021). We only used “meaningful” principal components of shape or form variation (identified with MORPHO’s *getMeaningfulPCs* function in Schlager 2017; Schlager et al. 2021) as input variables for clustering, in order to avoid overparameterization of the multivariate normal models. To assess static allometry in the adult nematode mouths, we generated common and group-specific allometry model fits with *procD.lm* and compared them in a homogeneity of slopes (HOS) test using the *anova* function of the R STATS package (R Core Team 2021). The compatibility of our data with a respective allometry hypothesis was tested using the aforementioned PERMANOVA-RRPP approach (Collyer et al. 2015). Allometric trajectories were visualized by plotting the first principal component (PC1) of a matrix containing predicted shapes from a multivariate regression of shape on size against the log-transformed centroid size of each specimen, using the *plotAllometry* function of GEOMORPH (Adams and Otárola-Castillo 2013; Adams et al. 2021).

EMS mutagenesis

To obtain Dumpy mutants with a short body size, we performed a large-scale genetic screen using ethyl methanesulfonate (EMS) mutagenesis in *P. pacificus* PS312. Specifically, we incubated a mixture of J4 larvae and young adults in M9 buffer (3 g/l KH_2PO_4 , 6 g/l Na_2HPO_4 , 5 g/l NaCl, 1 mM MgSO_4) with 47 mM EMS for 4 h (Sommer et al. 1996). A total of 1,500 healthy young adults were singled out on new plate in several rounds of mutagenesis. After animals had laid approximately 20 eggs, parents were removed and F1 progeny were allowed to reach maturity. A total of 5–10 F1 animals (which contain heterozygous mutants) from each P0 plate were then singled out. In total, 9,000 F1 animals were isolated and F2 progeny (which contained a mixture of genotypes, including homozygous mutants) were allowed to develop to adulthood. We isolated *Dpy* mutants that are shorter than wild-type using a discovery V20 stereomicroscope (Zeiss). Around 200 *Dpy* mutants were isolated and re-screened for mouth-form defects in the F3 or F4 generation using DIC microscopy. In total, we isolated 6 EMS mutants displaying a simultaneous *Dpy* and mouth-form phenotype (Fig. 2a and Supplementary Fig. 1a).

Mapping of *tu743*

For mapping, the EMS mutant *tu743* was backcrossed 3 times to the PS312 wild-type strain. The backcrossed line was sequenced by generating a NGS library using Low Input Library Prep kit (Clontech) and sequenced on the Illumina HiSeq3000 platform. Raw Illumina reads of the *tu743* mutant and of a mapping panel were aligned to the El Paco assembly of the *P. pacificus* genome

(Dieterich et al. 2008; Rödelsperger et al. 2017) by using the BWA software package (ver. 0.6) with parameter “bwa mem -M” (Li and Durbin 2009). Mutations were called by using the samtools (ver. 1.10) software with the parameter “bcftools mpileup—max-depth 1000 | bcftools call -cv” (Li et al. 2009). In total, 98 nonsynonymous/nonsense mutations (Supplementary Table 2) and 7 mutations in introns near splice sites of annotated genes (El Paco gene annotations v3) were identified in the candidate interval by a previously described custom variant classification software (Rae et al. 2012). The single-worm transcriptome of entire postembryonic stage in *P. pacificus* was downloaded from the European Nucleotide Archive with the study accession number PRJEB42613 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB42613>).

CRISPR/Cas9 mutagenesis

We followed the previously published protocol for *P. pacificus* with subsequently introduced modifications (Han et al. 2020; Nakayama et al. 2020). All target-specific CRISPR RNAs (crRNAs) were designed to target 20 bp upstream of the protospacer adjacent motifs (PAMs). A total of 10 μl of the 100 μM stock of crRNA (CRISPR/Cas9 RNA; IDT) was combined with 10 μl of the 100 μM stock of tracrRNA (catalog # 1072534; IDT), denatured at 95°C for 5 min, and allowed to cool down to room temperature and anneal. The hybridization product was combined with Cas9 protein (catalog# 1081058; IDT) and incubated at room temperature for 5 min. The mix was diluted with Tris-EDTA buffer to a final concentration of 18.1 μM for the RNA hybrid and 2.5 μM for Cas9. For the induction of specific site-directed mutations via CRISPR/Cas9, a ssDNA oligo template was included in

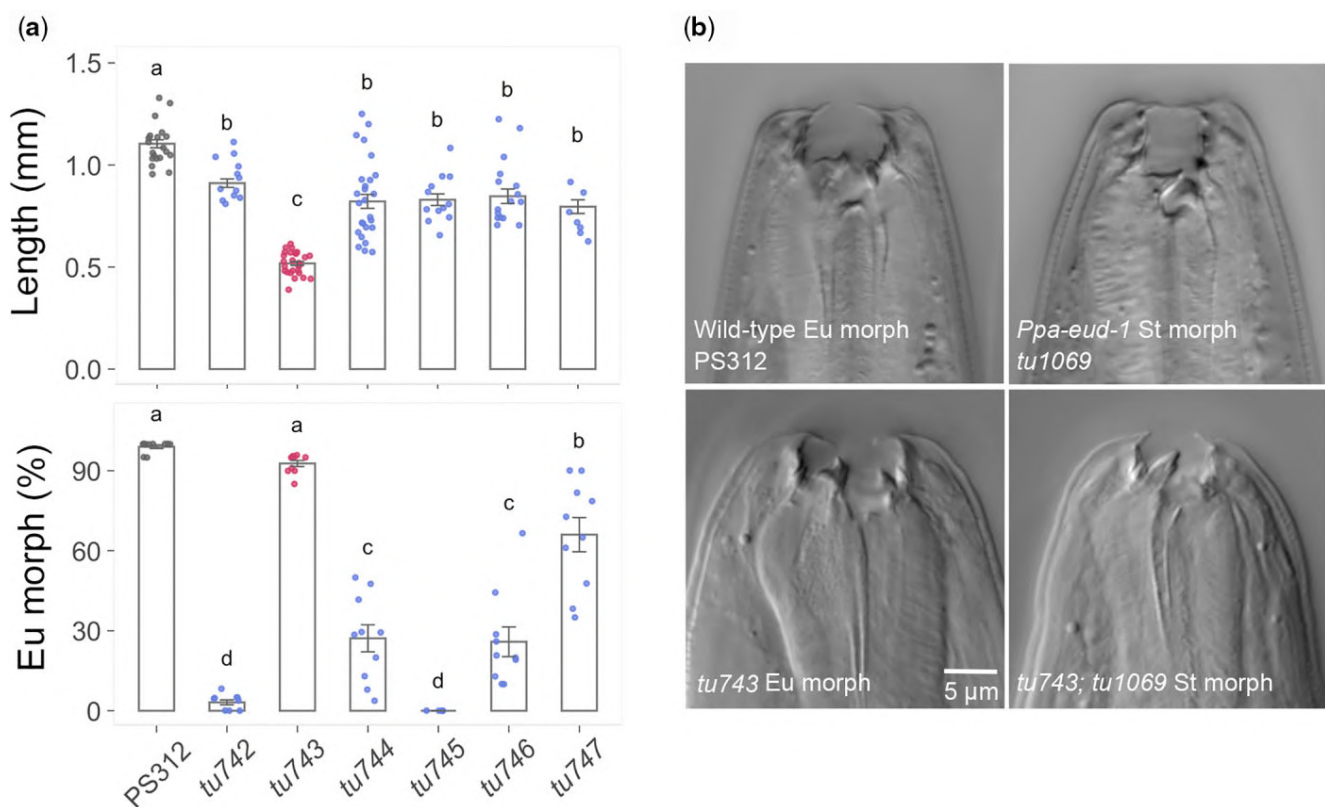


Fig. 2. Dumpy and mouth-form phenotypes of 6 novel *P. pacificus* mutants. a) The upper bar plot shows differences in the length of the worm body among wild-type (PS312) and the 6 mutant strains. The lower bar plot displays changes in mouth-form ratios between strains. Turkey's multiple comparison were tested by the Turkey's HSD test, with letters upon each volume indicating significant differences between the means, P -adjusted value < 0.05 , $n > 20$. b) Mouth form of wild-type Eu, St morph of the *Ppa-eud-1* mutant (*tu1069*), Eu morph with reduced cheilostom of the *tu743* mutant strain, and the St morph with a reduced cheilostom of *tu743* in a *Ppa-eud-1* double mutant. Additional images in 2 focal planes are shown in Supplementary Fig. 1b.

the mix at a concentration of 4 μ M. The repair template contained the desired modifications flanked by 50bp homology arms either side of the edited sequences. The plasmid carrying the *Ppa-ef1-3* promoter and modified TurboRFP sequences (Han et al. 2020) was used as co-injection marker. sgRNAs, associated primers and repair templates for generating mutants and specific targeted knock ins utilized in this study can be found in [Supplementary Table 5](#).

Injections were performed on a Zeiss Axiovert microscope (Zeiss, Germany) coupled to an Eppendorf TransferMan micro-manipulator and Eppendorf FemtoJet injector (Eppendorf AG., Hamburg, Germany). The microinjection mixture was injected in the gonad rachis of approximately 1-day-old adult hermaphrodites. Eggs laid by injected animals within a 12–16-h period post injection were recovered. After 2 days, the P0 plates containing the F1 animals with fluorescent signal of co-injection marker were isolated, and 8–10 F1 progenies from the isolated P0 plates were singled out on individual plates. After F1 animals have laid eggs, they were placed in 10 μ l of single worm lysis buffer (10 mM Tris-HCl at pH 8.3, 50 mM KCl, 2.5 mM MgCl₂, 0.45% NP-40, 0.45% Tween 20, 120 μ g/ml Proteinase K), and incubated in a thermocycler at 65°C for 1 h, followed by heat deactivation of the proteinase at 95°C for 10 min. The resulting lysate was used as a template in subsequent PCR steps. The genotype of the F1 animals were subsequently analyzed via Sanger sequencing and mutations identified before re-isolation in homozygosis.

Genetic transformation

Transcriptional reporter lines of *Ppa-dpy-6* were generated by PCR amplification of a 2.2-kb upstream region of the predicted start codon ([Supplementary Table 6](#)) using Gibson assembly (New England Biolabs). The resulting product was cloned into a pUC19 vector containing GFP (Han et al. 2020) and the *Ppa-rpl-23* 3'UTR. For injection, we used 10 ng/ml of the NotI-digested *Ppa-dpy-6::GFP* plasmid with 60 ng/ml of genomic carrier DNA, also digested with NotI. We found the same expression pattern in 2 independent lines. For the reporter construct of *Cel-dpy-6*, we cloned a 2.2-kb upstream region of the predicted start codon in N2 ([Supplementary Table 6](#)) into a pUC19 vector containing *gfp* and *Cel-rpl-23* 3'UTR via Gibson assembly (New England Biolabs). The injection mix for transformation contained 10 ng/ μ l of the reporter construct and 60 ng/ μ l genomic DNA. The same expression pattern was confirmed in 3 independent lines.

Analysis of oscillatory gene expression of *Ppa-dpy-6* and *Cel-dpy-6*

To profile the temporal gene expression of *Ppa-dpy-6* and *Cel-dpy-6* during postembryogenesis, we extracted the FPKM of *Ppa-dpy-6* at 38 time points with 3 biological replicates (Sun et al. 2021) and the RPKM of *Cel-dpy-6* at 48 time points (Meeuse et al. 2020). Mean FPKM at each time point was calculated for profiling the gene expression of *Ppa-dpy-6*, and RPKM at 48 time points was used to display the gene expression of *Cel-dpy-6*. The meta2d_Base and meta2d_AMP of oscillating gene set, which was generated from the results of the METACYCLE package, were used to both, profile the distribution of the mean gene expression and amplitude of all oscillating genes and the comparison of the mean expression and amplitude between *dpy-6* and the average of oscillating gene sets in both species.

Comparison of gene expression of *Ppa-dpy-6* between wild-type and *eud-1* mutant animals

We collected the transcriptome data at 20h posthatching for PS312 wild-type and *eud-1(tu1069)* mutant animals from

European Nucleotide Archive under the study accession number PRJEB42633 to compare the gene expression of *Ppa-dpy-6* between *Ppa-eud-1* mutant and wild-type (Sun et al. 2021). The RNA-seq data for the comparison of *Ppa-dpy-6* expression between the *Ppa-nhr-40* mutant and wild-type, and the *Ppa-nhr-1* mutant and wild-type, were downloaded from the European Nucleotide Archive with the study accession number PRJEB34615. We reperformed differential expression analysis using DESEQ2 (ver. 1.18.1) (Love et al. 2014) and applied an adjusted P-value cut-off of 0.05 and a fold change cut-off of 2 to identify differentially expressed genes. To compare the expression of *Ppa-dpy-6* between mouth-form mutants and wild-type, we extracted the FPKM values of *Ppa-dpy-6* ([Supplementary Table 7](#)) from each data set and applied a comparison by using the Tukey HSD test function in the AGRICOLAE package (version 1.3.3) with default parameters.

Results

A genetic screen for Dumpy mutants with accompanying mouth-form defects

To isolate genetic mutants defective in *P. pacificus* stomatal structures, we performed a combinatorial screen by first obtaining *Dpy* mutants that are shorter than wild-type, followed by re-screening these mutants for mouth-form defects. This rationale was based on preliminary findings that some available *Dpy* and *Uncoordinated Movement (Unc)* mutants showed concomitant defects of stomatal structures (Müller 2010). From a screen of 1,500 mutagenized animals, a total of 9,000 F1 progeny were singled out to individual plates. This resulted in the isolation of around 200 *Dpy* mutants with a phenotypic range similar to previous studies (Kenning et al. 2004). During re-screening, we found 6 mutants to have a concomitant mouth-form phenotype ([Fig. 2a](#), [Supplementary Fig. 1a](#), and [Supplementary Tables 1 and 2](#)). We focus on the mutant RS3887(*tu743*), which is severely *Dpy* and has a strong morphological phenotype in the stoma. First inspection by light microscopy of RS3887(*tu743*) mutant animals indicated a strong reduction in the size of the cheilostom in these otherwise Eu animals ([Fig. 2b](#)). To investigate the role of RS3887(*tu743*) in St animals, we crossed this line into the *Ppa-eud-1* mutant background that results in an all-St phenotype (Ragsdale et al. 2013). In *eud-1(tu1069); tu743* double mutants, we found similar defects in the cheilostom ([Fig. 2b](#) and [Supplementary Figs. 1b and 3b](#)). Thus, RS3887(*tu743*) affects the cheilostom in Eu and St animals and represents the first mutant with a specific defect in the nematode stoma.

RS3887 carries a mutation in the Mucin-type gene *Ppa-dpy-6*

To identify the molecular lesion responsible for the cheilostom defect of RS3887, we backcrossed this mutant line to wild-type animals and performed whole-genome sequencing (see Materials and Methods for detailed methodology). We found a total of 109 single-nucleotide polymorphisms (SNPs) resulting in nonsynonymous or nonsense mutations in annotated genes or intronic changes near splice sites in these backcrossed animals ([Supplementary Fig. 2a](#) and [Supplementary Table 3](#)). This included 1 SNP in the *P. pacificus* ortholog of *Cel-dpy-6*. As RS3887 shows a severe *Dpy* phenotype, *dpy-6* was a strong candidate for the causal gene behind the combined *Dpy* and cheilostom-defective phenotype. Note that the C>T SNP in *Ppa-dpy-6* is located at the 3'-splice site of intron 10 ([Supplementary Fig. 2c](#)). Therefore, we used CRISPR/Cas9-mediated gene knockouts and engineering to confirm the phenotypic consequences of this

Table 1. *Ppa-dpy-6* associated phenotypes.

	Genotype	Mutagen	Mutation		Body shape	Cheilostom	% Eu
a	PS312	n.a.	n.a.	n.a.	WT	WT	>95
b	RS3887 (tu743)	EMS	C>T intronic	11th exon	Dumpy	Short	>95
c	<i>Ppa-dpy-6</i> (tu1643)	CRISPR	Frameshift	3rd exon	Dumpy	Short	>95
	<i>Ppa-dpy-6</i> (tu1644)	CRISPR	Frameshift	3rd exon	Dumpy	Short	>95
	<i>Ppa-dpy-6</i> (tu1645)	CRISPR	Frameshift	3rd exon	Dumpy	Short	>95
d	<i>Ppa-dpy-6</i> (tu1646)	CRISPR	C>T intronic	11th exon	Dumpy	Short	>95
	<i>Ppa-dpy-6</i> (tu1658)	CRISPR	Frameshift	11th exon	Dumpy	Short	>95
e	<i>Ppa-dpy-6</i> (tu1659)	CRISPR	Frameshift	11th exon	Dumpy	Short	>95
	<i>Ppa-dpy-6</i> (tu1660)	CRISPR	Frameshift	11th exon	Dumpy	Short	>95
	RS3887 reversion	CRISPR	T>C intronic	11th exon	WT	WT	>95
g	<i>Ppa-dpy-6</i> & <i>Ppa-eud-1</i>	EMS, CRISPR and cross	C>T intronic	11th exon	Dumpy	Short	<5

N > 50 for all strains. % Eu, percent eury stomatous animals; n.a., not applicable; WT, wild-type.

mutation. First, we created 3 mutant lines in *Ppa-dpy-6* using a sgRNA in exon 3, all of which resulted in frameshift mutations. All 3 alleles *Ppa-dpy-6* (tu1643, tu1644, and tu1645) are strongly *Dpy*, have a reduced cheilostom, but do not affect the mouth-form ratio (Table 1a–c, Supplementary Fig. 3c, and Supplementary Table 1). Thus, all 3 *Ppa-dpy-6* alleles have a phenotype that is indistinguishable of the original RS3887(tu743) mutant. Second, we reconstituted the original C>T mutation in intron 10 by CRISPR/Cas9 engineering and this mutant strain has a similar *Dpy* and cheilostom phenotype (Table 1d and Supplementary Fig. 3c). Three mutant lines with frameshift mutations in exon 11 were obtained from the same injection. All 3 alleles *Ppa-dpy-6* (tu1658, tu1659, and tu1660) have the same phenotype as RS3887 (Table 1e and Supplementary Fig. 3c). Finally, we reverted the C>T mutation back to wild-type in the RS3887 mutant background. The resulting strain RS4003 has a normal body shape and a normal cheilostom (Table 1f and Supplementary Fig. 3c). We conclude that mutations in *Ppa-dpy-6* result in strongly reduced body size and concomitant defects in the cheilostom. Thus, *Ppa-dpy-6* represents the first gene involved in the specification of the morphology of the stoma.

In *C. elegans*, *dpy-6* was cloned as part of genome-wide RNA interference screens, but was not further characterized (Simmer et al. 2003). DPY-6 shows sequence similarities to Mucins, a family of glycosylated hydrogel-forming proteins conserved throughout the animal kingdom. In humans, Mucins play important roles as mucus barriers in wet epithelial surfaces of the body, such as eyes, lungs, and the stomach with a proposed function as biological hydrogel that coats surface structures (Witten and Ribbeck 2017; Wagner et al. 2018), which would be compatible with the phenotype of *Ppa-dpy-6* in the cuticle and cheilostom.

Geometric morphometrics reveal mutant phenotype in mouth form, shape, and size

To analyze the mouth-form defects of *Ppa-dpy-6* with greater accuracy, we used our recently established protocol for geometric morphometric analyses in microscopic animals (Theska et al. 2020) and found a consistent phenotype in mouth shape (see Supplementary Material for detailed procedure). First, a PCA of mouth shape revealed that *Ppa-dpy-6*(tu1645) mutants occupy a separate aspect of the shape space when compared to wild-type (PS312) and rescued animals (RS4003, tu1696) (Fig. 3a). Note that wild-type and rescued animals occupy strongly overlapping areas of the shape space, indicating that their mouth shapes are very similar. Looking at the variation described by the individual principal components revealed that PC1 described 77.1% of the total variation, rendering this shape data almost 1-dimensional. However, PC1 is not the only PC that contains biologically

relevant information. In fact, we found 3 PCs (PC1 = 77.1%, PC2 = 4.0%, and PC3 = 3.0%) to be “meaningful” according to the *getMeaningfulPCs* function of MORPHO (threshold value = 1.29, based on $n = 118$). Our PCA indicates (1) a clearly different mouth shape in *Ppa-dpy-6*(1645) mutants compared to wild-type animals and (2) a practically complete reversal of the mutant shape when the function of *Ppa-dpy-6*(1645) is restored with CRISPR. This is also supported by multivariate statistical analysis, which shows that the independent variable “strain” strongly influences *P. pacificus* mouth shape as a response variable (PERMANOVA-RRPP: shape ~ strain, $Z = 5.9$, and $P = 0.0001$). Pairwise comparisons of distances between the group means of individual strains revealed that our shape data are incompatible with the null hypotheses that no mouth shape differences exist between *Ppa-dpy-6*(tu1645) mutants and wild-type or rescued animals (wild-type strain vs. *Ppa-dpy-6*(1645): $Z = 3.5$ and $P = 0.0001$, rescue strain vs. *Ppa-dpy-6*(1645): $Z = 3.6$ and $P = 0.0001$), but compatible with the null hypothesis that no clear differences exist in mouth shape between wild-type and rescued animals (wild-type strain vs. rescue strain: $Z = 1.2$ and $P = 0.1322$). Finally, we used model-based clustering to identify how many separate groups can be identified in the shape data and found only 2 separate clusters, one that contains all *Ppa-dpy-6* mutants and another containing all wild-type and rescued animals (Fig. 3b). Thus, PCA, PERMANOVA, and model-based clustering analyses reveal clear mouth shape alterations in *Ppa-dpy-6* mutants.

Second, we estimated the positive and negative extreme shapes for PC1 using wireframe representations of the *P. pacificus* mouth. We found that the most prominent mouth shape difference is an extremely shortened and slightly posteriorly shifted cheilostom in *Ppa-dpy-6*(1645) mutants. This is also reflected in very large relative contributions of landmarks 1 and 15 (anterior tips of the cheilostom) to PC1 (LM1 = 20.0% and LM15 = 20.7% of the described variation). Another obvious change is the anteriorly shifted (relative to the rest of the buccal cavity), but similarly sized, metastegostomal tooth (LM6 = 7.6% and LM7 = 12.3%, LM 9 = 7.9% of the variation described by PC1) (Fig. 3c).

Third, we investigated whether mouth size influences mouth shape, that is, whether there is allometry. A HOS test revealed that a common allometry model (shape ~ size + strain) is most appropriate to fit the shape data. Thus, we found that the slopes of the allometric trajectories among all 3 strains are similar, with mouth size moderately influencing mouth shape and “strain” being the major factor of influence (PERMANOVA-RRPP: size, $Z = 3.3$, and $P = 0.0001$; strain, $Z = 5.3$, and $P = 0.0001$). We conclude that our data are incompatible with a null hypothesis of isometry (see Supplementary Material for more details), but compatible with the alternative hypothesis of allometry. Additionally,

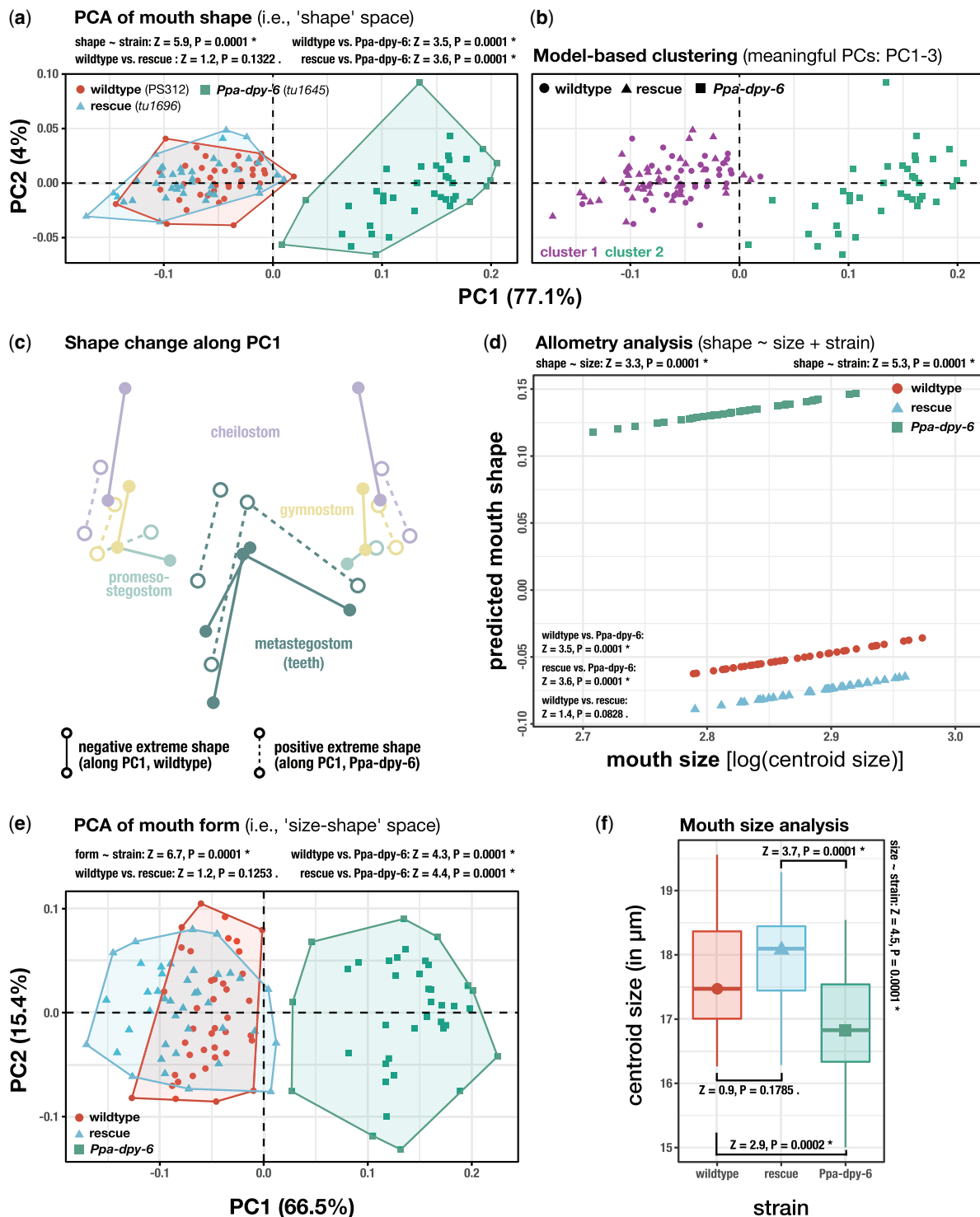


Fig. 3. Geometric morphometric analysis of mouth shape, size, and form in *P. pacificus* (exclusively Eu animals of each strain). a) PCA of the mouth shape for the wild-type (PS312), the *Ppa-dpy-6*(*tu1645*) mutant, and the rescue strain (RS4003, *tu1696*). The plot shows the distribution of all specimens in a shape space. b) Clusters identified in the *P. pacificus* shape data set. Clusters were found by model-based clustering, using all “meaningful” principal components as input variables. c) Superimposed wireframe representations of extreme shapes along PC1. d) Allometric trajectories for each of the *P. pacificus* strains. The allometry model that was used to fit the data is indicated above the plot. The y-axis corresponds the PC1 of a matrix of predicted values, which were obtained by a multivariate regression of shape on size. e) PCA of the mouth forms of each *P. pacificus* strain. The plot shows the distribution of all specimens in a “size-shape” (i.e. form) space. f) Boxplot showing the centroid sizes for all specimens of each *P. pacificus* strain. The box indicates the interquartile range (IQR, i.e. the range from the first to the third quartile) and the solid line indicates the median centroid size. Lower whisker extends from the lower quartile to the smallest value which is at most 1.5 IQR away from the lower quartile. Upper whisker extends from the upper quartile to the largest value which is at most 1.5 IQR away from the upper quartile. Specimens which fall beyond the whiskers (in either direction) are plotted as individual dots and are regarded to be outliers. Relative effect sizes (Z-scores) and P-values obtained via PERMANOVA-RRPP are indicated where appropriate. Asterisks “*” indicate incompatibility with a null hypothesis, while periods “.” indicate compatibility (see Supplementary Material for details).

pairwise comparisons of LS means among strains revealed that our allometry data are compatible with the null hypothesis of no blatant differences between wild-type and rescued worms (wild-type strain vs. rescue strain: $Z = 1.4$ and $P = 0.0828$), while we could find clear differences in the LS means of *Ppa-dpy-6* mutants on the one hand, and wild-type and rescued animals on the other (wild-type strain vs. *Ppa-dpy-6*: $Z = 3.5$ and $P = 0.0001$, rescue strain vs. *Ppa-dpy-6*: $Z = 3.6$ and $P = 0.0001$). Most importantly, the allometric trajectories of *Ppa-dpy-6* and wild-type animals have very different intercepts along the y-axis (Fig. 3d). Thus, while all 3 strains have similar allometric slopes, they do not share a general underlying trajectory and the mutant mouth shape cannot be obtained by scaling up or down the size of the wild-type mouth shape along such a general allometric trajectory.

Fourth, we analyzed morphometric differences in mouth form. Some clarification is appropriate here: previously, we colloquially used the term “mouth form” to refer to the 2 different mouth morphologies (Eu and St) that can be found in *P. pacificus*. In this section however, we strictly use the morphometric denotation of *form*. Briefly, in traditional morphometrics size and shape are treated as logically separate aspects of a biological structure and it is common practice to analyze variation in size separate from variation in shape (Mitteroecker et al. 2013). When taken together, the shape and the size of a structure define its *form*. Thus, analyses which focus on differences in *form* address the entirety of morphological variation in a biological structure, rather than variation in a specific structural aspect like shape or size. In modern morphometrics, multiple tools are available to investigate biological form and here we performed a “size-shape” PCA *sensu* Mitteroecker et al. (Mitteroecker et al. 2004, see [Supplementary Material](#) for additional information). We plotted the first 2 principal components of mouth-form variation against each other to visualize a form space (Fig. 3e). PC1 of the “size-shape” PCA describes less variation (66.5%) than PC1 of the shape PCA (77.1%), while PC2 of the “size-shape” PCA describes much more variation (15.4%) than PC2 of the shape PCA (4%). We were able to recover a clear separation of *Ppa-dpy-6* mutant mouth forms from wild-type and rescued animals, similar to what we found in the shape PCA (Fig. 3, a and e). This is also reflected in the statistical analysis of the mouth-form data set, which indicated that *Ppa-dpy-6* mutants differ from wild-type animals more prominently in mouth form ($Z = 6.7$, Fig. 3e) than in mouth shape ($Z = 5.9$, Fig. 3a). Pairwise comparisons of distances between group means revealed that our form data are incompatible with the null hypothesis that there are no mouth form differences between *Ppa-dpy-6* mutants, and wild-type or rescued animals (wild-type strain vs. *Ppa-dpy-6*(1645): $Z = 4.3$, $P = 0.0001$, rescue strain vs. *Ppa-dpy-6*(1645): $Z = 4.4$ and $P = 0.0001$). Similar to the results from our shape analysis, our form data are compatible with the null hypothesis that there are no discernable mouth form differences between wild-type and rescued animals (wild-type vs. rescue strain: $Z = 1.2$, $P = 0.1253$), indicating a reversal of the mutant mouth form to the wild-type situation after CRISPR-repair of the mutant *Ppa-dpy-6*(1645) allele. This is further corroborated by model-based clustering: we were able to identify 2 separate clusters in our form space which almost perfectly correspond to *Ppa-dpy-6*(1645) mutant animals on the one hand, and wild-type and rescued animals on the other ([Supplementary Fig. 3](#)).

Lastly, we investigated mouth size using the centroid size as proxy. We found that, on average, wild-type and rescued worms tend to have larger mouths than both *Ppa-dpy-6* mutants (Fig. 3f). The median centroid sizes of wild-type and rescued animals are

17.47 and 18.10 μm , while it is 16.83 μm in *Ppa-dpy-6*(1645) mutants. Thus, the median centroid size of wild-type animals and the first (lower) quartile of rescued animals correspond to the third (upper) quartile of *Ppa-dpy-6*(1645) mutants (17.54 μm) and the median centroid size of *Ppa-dpy-6*(1645) mutants is smaller than the first (lower) quartile of both, wild-type and rescued animals (Fig. 3e). Statistical analysis indicates that our size data are incompatible with the null hypothesis of no mouth size differences among the 3 strains (PERMANOVA-RRPP: mouth size \sim strain: $Z = 4.5$ and $P = 0.0001$). Interestingly, pairwise comparisons of distances between group means revealed that these data are compatible with the null hypothesis that no blatant mouth size differences exist between wild-type and rescued animals ($Z = 0.9$ and $P = 0.1785$), while it is incompatible with the null hypotheses of no mouth size differences between the *Ppa-dpy-6* mutant strain and the wild-type or rescue strain (*Ppa-dpy-6*(1645) vs. wild-type strain: $Z = 2.9$ and $P = 0.0002$, *Ppa-dpy-6*(1645) vs. rescue strain: $Z = 3.7$, $P = 0.0001$). Thus, we were able to fully reverse the observed mutant phenotypes in mouth form, mouth shape, and mouth size by repairing the mutant allele of *Ppa-dpy-6*(1645) with CRISPR. The high resolution of the phenotypic analyses we applied here highlight the power and necessity for modern geometric morphometrics in evo devo studies on miniature animals.

***Ppa-dpy-6* is an oscillatory gene, requires *eud-1*-dependent environmental signaling, and is expressed in the hypodermis**

Given the prominent stoma phenotype, we next wanted to study gene regulation of *Ppa-dpy-6*. Single-worm transcriptomics (SWT) in *P. pacificus* has recently provided a high-resolution map of the postembryonic developmental transcriptome (Sun et al. 2021). Of the more than 28,000 predicted genes of *P. pacificus*, nearly 3,000 show oscillatory expression with discrete phases, amplitudes, and expression levels (Sun et al. 2021). This includes many of the body-wall collagens that are often transcribed before molting and are involved in cuticle formation. Therefore, we analyzed the expression profile of *Ppa-dpy-6* and found strong oscillation in expression with the highest peak 4 h before the second, third, and fourth molt (Fig. 4a). This expression profile is consistent with a role of *Ppa-dpy-6* in the synthesis of the new cuticle. When we profiled the expression of *Cel-dpy-6* during postembryogenesis using the data of Meeuse et al. (2020), we found that the expression of *Cel-dpy-6* also displayed an oscillating pattern coupled to the molting cycle (Fig. 4b). Interestingly, profiling the distribution of the mean expression and relative amplitude of oscillating genes in both species, we found that *Ppa-dpy-6* and *Cel-dpy-6* exhibited higher gene expression levels and greater oscillating amplitudes than the mean value of the oscillating gene sets (Fig. 4c). Together, these temporally conserved gene expression patterns of *Ppa-dpy-6* and *Cel-dpy-6* indicate a highly conserved role in cuticle formation.

Next, we wanted to know if *Ppa-dpy-6* is controlled by the mouth-form plasticity gene regulatory network that has recently been identified (Sieriebriennikov et al. 2020). The sulfatase-encoding gene *eud-1* is a major developmental switch involved in environmental sensing and developmental reprogramming (Ragsdale et al. 2013; Sommer 2020). SWT of *eud-1* mutant animals have recently been performed at 20 and 28 h after hatching (Sun et al. 2021). When we compared *Ppa-dpy-6* expression in wild-type and *Ppa-eud-1* mutant animals, we found a strong upregulation of *Ppa-dpy-6* in the mutant condition (Fig. 4d). Thus, *Ppa-dpy-6* expression is sensitive to *Ppa-eud-1*-dependent environmental signaling. In contrast, we found no difference in *Ppa-dpy-6* expression when

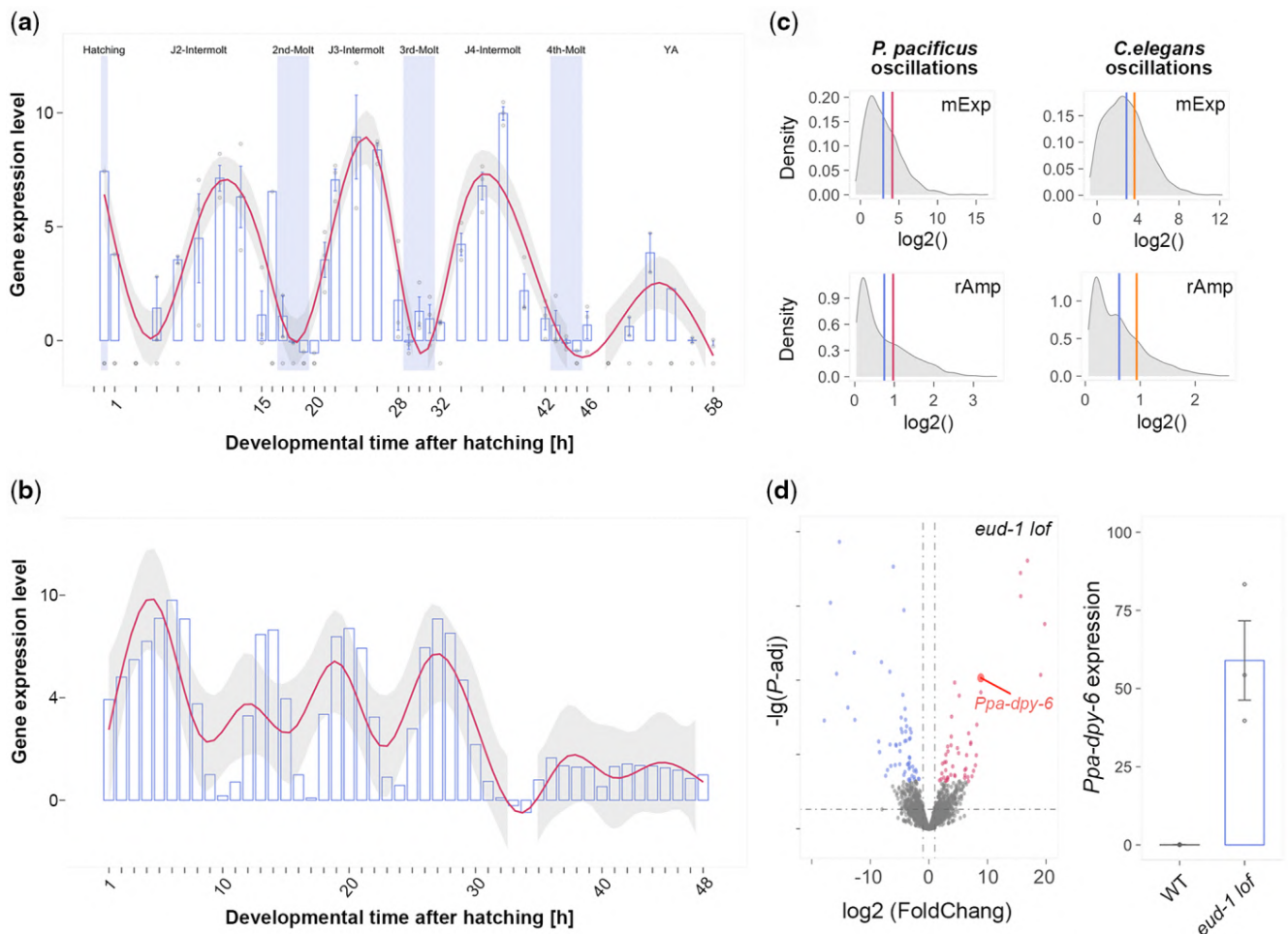


Fig. 4. Temporal expression patterns of *Ppa-dpy-6* and expression of *Ppa-dpy-6* in the *Ppa-eud-1* mutant condition. a) Dynamic expression of *Ppa-dpy-6* during postembryonic development. X-axis represents the 38 time points after hatching, the y-axis is the $\log_2()$ of FPKM, 3 replicates per time point. The data were extracted from the single-worm transcriptome data set of Sun et al. (2021). b) Bar plot profiles of the dynamic expression of *Cel-dpy-6* during postembryogenesis using the data set of Meeuse et al. (2020). c) Line plots profiling the comparison of the mean expression and the relative amplitude between *Ppa-dpy-6* (red) and the mean of the *P. pacificus* oscillatory gene set (blue), and between *Cel-dpy-6* (yellow) and the mean of *C. elegans* oscillating gene set (blue). d) Volcano plot revealing that *Ppa-dpy-6* is one among the strongest differentially expressed genes (DEG) in the comparison between *Ppa-eud-1* mutant and wild-type conditions. Downregulated genes in blue, upregulated genes in red, and nonsignificantly expressed genes in gray. Bar plot displaying differences in gene expression of *Ppa-dpy-6* in *Ppa-eud-1* and wild-type condition. All samples used are at 20 h after hatching. Turkeys multiple comparison were tested by the Turkeys HSD test, with the letters upon each volume indicating significant differences between the means, P-adjusted value <0.05 .

comparing the transcriptomes of wild-type animals and those of the 2 nuclear hormone receptor (NHR) mutants *Ppa-nhr-1* and *Ppa-nhr-40* (Sieriebriennikov et al. 2020) (Supplementary Fig. 6, a and b). This finding indicates that during evolution toward the *Pristionchus* lineage *dpy-6* has come under the control of the *eud-1* switch locus as this gene does not exist in *C. elegans*. Furthermore, these data suggest that not all *eud-1*-dependent environmental sensing goes through *Ppa-nhr-1* and *Ppa-nhr-40* and that other downstream transcription factors likely exist.

Finally, to study the expression pattern of *Ppa-dpy-6*, we generated a 2-kb promoter fusion construct with GFP. *Ppa-dpy-6* is continuously expressed in postembryogenesis and adult stages (Fig. 5). We observed expression in the hypodermal syncytia hyp1, hyp5, hyp6/7, and the anterior and posterior arcade cells (Fig. 5 and Supplementary Video 1), the characterization of which is based on the recent reconstruction of the head hypodermis in *P. pacificus* (Harry et al. 2021). Interestingly, when we generated a *Cel-dpy-6* reporter construct, we found a similar expression pattern except for *C. elegans* also showing expression in hyp4

(Table 2). We conclude that the oscillatory gene expression and the large overlap in gene expression of *Ppa-dpy-6* and *Cel-dpy-6* are consistent with a conserved role in cuticle formation. Therefore, we finally used geometric morphometrics of *Cel-dpy-6* mutants to search for a stoma phenotype, as no defects in the mouth were previously reported.

***Cel-dpy-6* mutants have defects in the cheilostom and gymnostom**

We compared the mouth shapes of wild-type (N2) and *Cel-dpy-6(e14)* mutants (CB14) and found a strong mutant phenotype that is reminiscent to the one observed for *Ppa-dpy-6*. Specifically, PCA of mouth shapes revealed that *Cel-dpy-6(e14)* mutants occupy a different part of the shape space than wild-type animals with both strains being almost completely separated along PC1 (Fig. 6a). Similar to our findings in the *Ppa-dpy-6* analysis, the first 3 principal components of shape variation in *C. elegans* (PC1 = 38.5%, PC2 = 23.7%, and PC3 = 7.7%) can be regarded as “meaningful” according to the *getMeaningfulPCs* function of

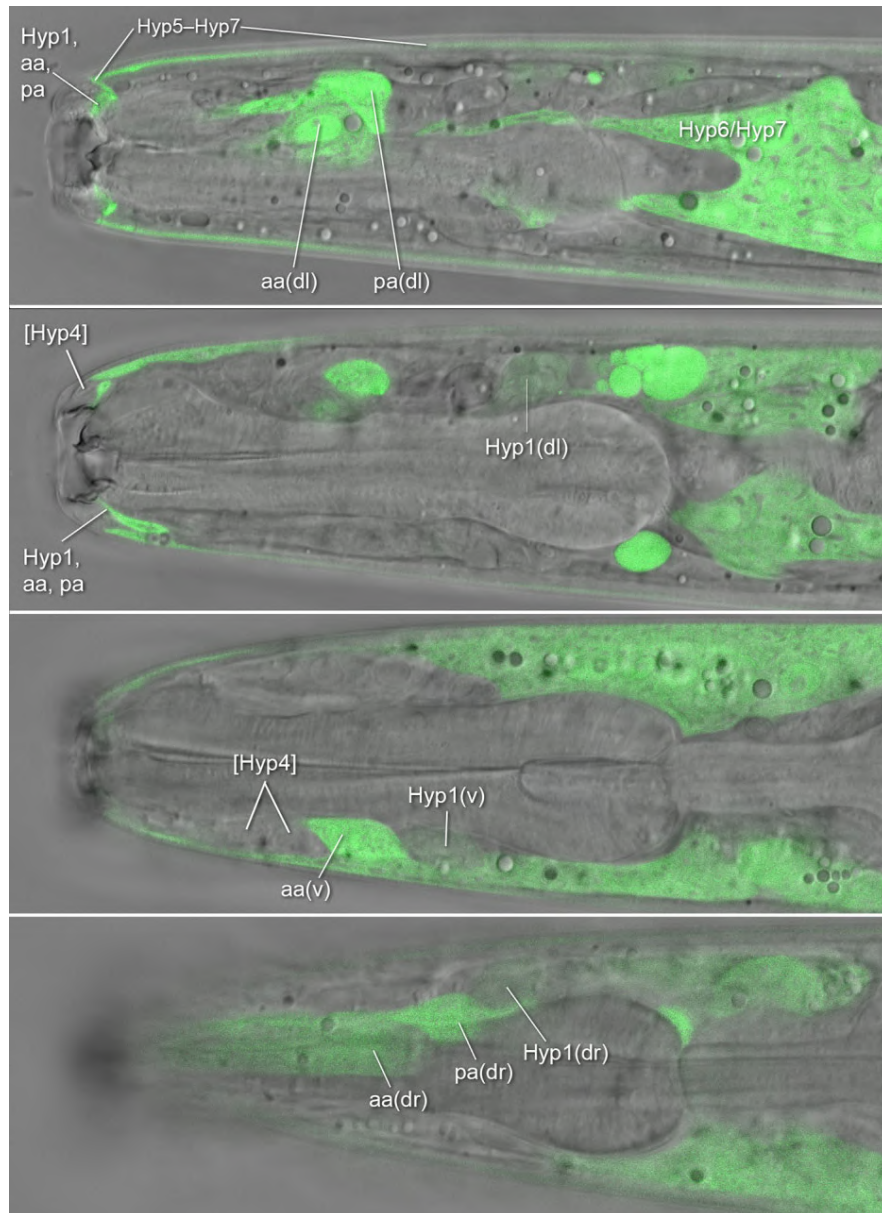


Fig. 5. Spatial expression patterns of *Ppa-dpy-6*. Spatial expression of *Ppa-dpy-6* from a 2.2-kb *Ppa-dpy-6* promoter element driving GFP expression in different focal planes. aa, anterior arcade cell; pa, posterior arcade cell; Hyp, hypodermal syncytia; dr, dorsal right; dl, dorsal left; v, ventral. Images were analyzed using ImageJ software and Arivis Vision 4D software (Arivis AG, Berlin, Germany).

Table 2. *dpy-6* expression in anterior epithelial syncytia in *P. pacificus* and *C. elegans*

	<i>P. pacificus</i>	<i>C. elegans</i>
Posterior arcade	+	+
Anterior arcade	+	+
Hyp1	+	+
Hyp2	-	-
Hyp3	-	-
Hyp4	-	+
Hyp5	+	+
Hyp6/7	+	+

MORPHO (threshold value = 1.38, based on $n = 75$). The presence of a *Cel-dpy-6(e14)* mutant phenotype in mouth shape is further supported by multivariate statistical analysis (PERMANOVA-RRPP:

shape ~ strain, $Z = 5.4$ and $P = 0.0001$). Pairwise comparisons of distances between group means revealed that shape data are incompatible with the null hypothesis of no mouth shape differences between the *Cel-dpy-6(e14)* and wild-type ($Z = 5.1$, $P = 0.0001$). This conclusion is further strengthened by model-based clustering, which identified separate wild-type and mutant clusters in the shape data (Fig. 5b). We conclude that mutating the one-to-one ortholog of *Ppa-dpy-6* in *C. elegans* also results in a clearly identifiable mutant mouth shape defect, just as observed for *Ppa-dpy-6* mutants.

Given this conclusion, we wondered whether the shape differences between wild-type and *Cel-dpy-6* mutants actually stem from morphological changes in the same structures for which we observed shape differences in *P. pacificus*. Comparing the estimated extreme shapes along PC1 revealed that 2 components of the mouth are strongly affected by the genetic perturbation.

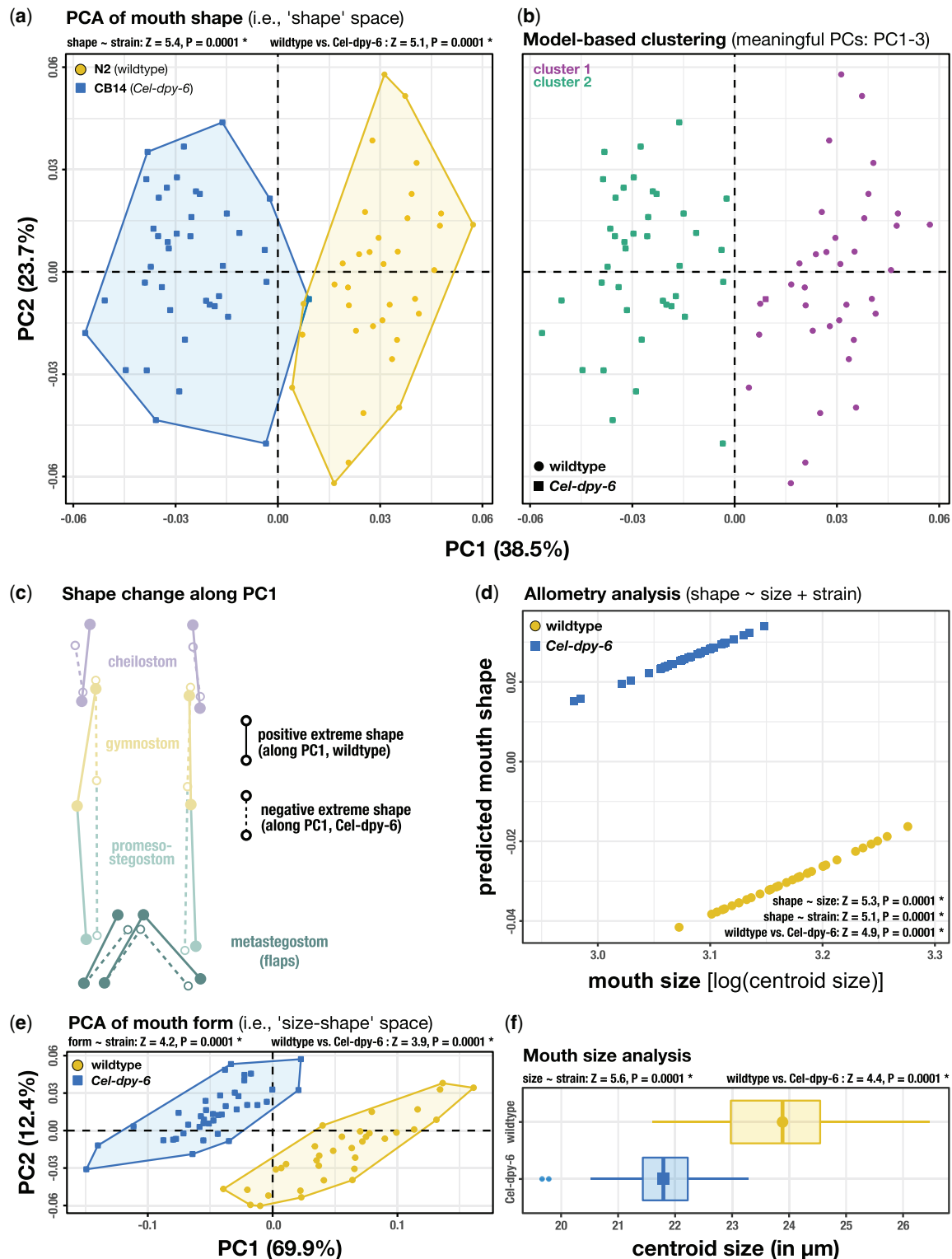


Fig. 6. Geometric morphometric analysis of mouth shape, size and form in *C. elegans*. a) PCA of the mouth shape for the wild-type (N2) and the *Cel-dpy-6* mutant strain (CB14). The plot shows the distribution of all specimens in a shape space. b) Clusters identified in the *C. elegans* shape data set. Clusters were found by model-based clustering, using all “meaningful” principal components as input variables. c) Superimposed wireframe representations of extreme shapes along PC1. d) Allometric trajectories for the *C. elegans* strains. The allometry model that was used to fit the data is indicated above the plot. The y-axis corresponds to the PC1 of a matrix of predicted values which were obtained by a multivariate regression of shape on size. e) PCA of the mouth forms of *C. elegans*. The plot shows the distribution of all specimens in a “size-shape” (i.e. form) space. f) Boxplot showing the centroid sizes for all specimens of both *C. elegans* strains. The box indicates the interquartile range (IQR, i.e. the range from the first to the third quartile) and the solid line indicates the median centroid size. Lower whisker extends from the lower quartile to the smallest value which is at most 1.5 IQR away from the lower quartile. Upper whisker extends from the upper quartile to the largest value which is at most 1.5 IQR away from the upper quartile. Specimens which fall beyond the whiskers (in either direction) are plotted as individual dots and are regarded to be outliers. Relative effect sizes (Z-scores) and P-values obtained via PERMANOVA-RRPP are indicated where appropriate. Asterisks “*” indicate incompatibility with a null hypothesis, while periods “.” indicate compatibility (see Supplementary Material for details).

First, the cheilostom of *Cel-dpy-6(e14)* mutants is shortened along the anterior–posterior axis, and the mouth opening is wider than in wild-type worms. Second, the gymnostom of *Cel-dpy-6(e14)* mutants is shortened along the anterior-posterior axis and the main cuticular tube of the mouth (gymnostom+promesostegostom) is narrower than in wild-type worms (Fig. 6, a–c). These observations are also reflected in very large relative contributions of landmarks 1 and 15 (anterior tips of the cheilostom) to PC1 (LM1 = 17.1% and LM15 = 12.4% of the described variation), and of landmarks 4 and 12 (posterior tips of the gymnostom) also to PC1 (LM4 = 26.4% and LM15 = 8.5% of the described variation). Thus, we were able to identify a similar phenotype in cheilostom shape in *dpy-6* mutants of both nematode species, but only *C. elegans* shows a phenotype in gymnostom shape.

Next, we performed an allometry analysis similar to the one we used for *P. pacificus*. The HOS test revealed a common allometry model as most appropriate to fit *C. elegans* shape data, with both mouth size and strain being highly influential factors on mouth shape as a phenotypic response variable (PERMANOVA-RRPP: shape ~ size + strain; size, $Z = 5.3$ and $P = 0.0001$; strain, $Z = 5.1$ and $P = 0.0001$). We conclude that the *C. elegans* shape data set are incompatible with a null hypothesis of isometry, and instead, is compatible with the alternative hypothesis of allometry (Fig. 5d), a finding in line with our observations for *P. pacificus*. Interestingly, the *dpy-6(e14)* mutation apparently does not change the allometric trend in adult stomatal growth of both nematode species (i.e. the slope of the trajectory; see [Supplementary Material](#) for more details), but it changes adult mouth shape and size. The mouth shape phenotype in *Cel-dpy-6(e14)* mutants cannot be explained by an allometric scaling effect only, as this would require similar y-intercepts of mutant and wild-type trajectories, a finding that is congruent with the observation in *P. pacificus* (Figs. 3d and 6d).

Finally, we investigated whether there are phenotypes in mouth form and size in *Cel-dpy-6* mutants. The “size-shape” PCA revealed that wild-type animals and mutants occupy nonoverlapping aspects in the form space created by PC1 and PC2 (Fig. 6e). We found that the first 3 principal components of form variation can be considered “meaningful” (threshold value: 1.38, $n = 75$). Given that PC1 explains such a large amount of form variation (69.9%) and that there is a considerable overlap in shape variation along PC1, we consider the overall mouth form difference between wild-type and mutant to be less obvious than the difference in shape (Fig. 6, a and e). Therefore, the *C. elegans* strains show the opposite pattern to *P. pacificus* in that they differ more prominently in mouth shape ($Z = 5.4$) than in mouth form ($Z = 4.2$, compare Figs. 3, a and e and 6, a and e). Furthermore, pairwise comparison of the *C. elegans* strains revealed that our form data are incompatible with the null hypothesis that there are no mouth form differences between them (wild-type strain vs. *Cel-dpy-6*: $Z = 3.9$, $P = 0.0001$). When analyzing mouth size, we found that (on average) wild-type worms have larger mouths than *Cel-dpy-6(e14)* as the median centroid size of wild-type animals is 23.88 μm , while it is 21.79 μm for the *Cel-dpy-6(e14)* mutants (Fig. 6f). Additionally, the first (lower) quartile of wild-type mouth sizes (22.98 μm) is larger than the third (upper) quartile of mutant mouth sizes (22.22 μm), which we interpret as indicative of a clear mouth size phenotype in mutant animals. This is further supported by the statistical analysis, which found that our data are incompatible with the null hypothesis that there are no differences in mouth size (PERMANOVA-RRPP: mouth size ~ strain: $Z = 5.6$ and $P = 0.0001$). These findings perfectly correspond to the observations for *P. pacificus* (compare Figs. 3f and 6f).

In conclusion, *Ppa-dpy-6* and *Cel-dpy-6* show overlapping defects in all structural aspects of stoma morphology with some minor differences between species as revealed by our geometric morphometric analysis. This overall conclusion supports *dpy-6* as an important and evolutionary conserved gene controlling nematode mouth formation—the first one to be characterized in greater detail.

Discussion

We have identified the Mucin-type hydrogel-forming protein DPY-6 to play a major role in the formation of the cheilostom in both *P. pacificus* and *C. elegans*. The associated morphological defects indicate that the cheilostom and body wall cuticle share common structural components, a finding that supports the traditional notion that the cheilostom is continuous with the external cuticle. In both species, we observed *dpy-6* expression in the hypodermal syncytia *hyp1*, *hyp5*, and *hyp6/7* and the anterior and posterior arcade cells indicating their involvement in the formation of the cuticle and the cheilostom. Strikingly, however, except for *dpy-6* no other *Dpy* mutants in *C. elegans* show defects of the cheilostom. Thus, the cheilostom is likely different from the cuticle in its exact chemical composition. While the conserved role of DPY-6 in cuticle deposition can easily be deduced from the strong *Dpy* phenotype, the full understanding of its role in stoma formation required sophisticated geometric morphometric analyses. Note that the cheilostom defect of *Ppa-dpy-6* is visible in standard light microscopy, whereas the *Cel-dpy-6* phenotype was more subtle and could be accessed properly only through geometric morphometrics. This highlights the power of modern quantification methods for morphological differences in miniature animals and indicates that these techniques would aid similar research programs in other organisms, such as rotifers or tardigrades. Together, the identification of the first structural component of the nematode stoma will initiate 3 new research directions.

First, our genetic analysis of *dpy-6* is consistent with its gene product serving as an initiation point for the identification of additional proteins of the *Pristionchus* stoma. This will hopefully help to address some of the ultimate questions associated with mouth-form plasticity in *P. pacificus*. Are there qualitative differences—that is novel structural components—in the teeth of *P. pacificus* or other members of the Diplogastridae that are unknown from other free-living nematodes? Also, are the differences between the Eu and the St mouth form of quantitative or qualitative nature? We speculate that qualitative differences between the stoma of *P. pacificus* and other nematodes such as *C. elegans* do exist and future studies will hopefully identify their molecular nature.

Second, the molecular nature of DPY-6 will result in nanotechnological approaches. DPY-6 shows high sequence similarity to human mucins, a group of glycoproteins that shape the barrier function of the mucus (Phani et al. 2018; Werlang et al. 2019). As typical for mucins, DPY-6 is a high-molecular weight protein with a large number of tandemly repeated serine and threonine residues in its backbone that serve as O-glycosylation sites. When hydrated, mucins form hydrogels with crosslinking domains including disulfide bonds supplied by cysteine residues (Wagner et al. 2018). Additionally, the glycosylation will further increase the complexity of protein aggregates. Importantly, multiple structural variations exist in human gel-forming mucins that are not fully understood. Therefore, future work on *Ppa-dpy-6* using the simplicity of the nematode body plan and phenotypic readout

might help to characterize the different domains of mucin proteins. In addition, *in vitro* studies of *Ppa*-DPY-6 can reveal its nanotechnological properties similar to investigations of the spider's web as another example of a biological hydrogel (Witten and Ribbeck 2017).

Finally, the identification of *Ppa-dpy-6* as a structural component of the stoma extends the knowledge of the gene regulatory network (GRN) involved in *P. pacificus* teeth formation and the evolutionary age structure of these genes. Many genes involved in environmental sensing and subsequent developmental switching have resulted from lineage-specific gene duplications, that is, the *eud-1*/sulfatase of *P. pacificus* is one of 2 duplicates of the single *sul-2* gene in *C. elegans* (Bui and Ragsdale 2019; Casasa et al. 2021). Interestingly and in contrast to these findings, the 2 known transcription factors *Ppa-NHR-1* and *Ppa-NHR-40* acting downstream in the GRN are highly conserved in nematode evolution. *NHR-1* and *NHR-40* of *P. pacificus* have 1:1 orthologs in *C. elegans*, while their known regulatory targets are again rapidly evolving (Sieriebriennikov et al. 2020). Ultimately, the identification of the structural components of teeth formation will represent a major milestone for a full understanding of the GRN of mouth-form plasticity. In conclusion, the identification of the first structural component of the nematode stoma will lead to novel activities in various research areas and provides an important framework for future investigations.

Data availability

Raw and processed data sets from this study have been submitted to the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena>) under the accession number PRJEB45271. The landmark data used for geometric morphometric analyses are provided in [Supplementary Data 1 and 2](#).

Supplemental material is available at *GENETICS* online.

Acknowledgments

We are grateful to Gabi Eberhardt and Tobias Loschko for their assistance with the mutagenesis screens and Jürgen Berger for taking the SEM image of the *P. pacificus* mouth. We thank Dr Michael Werner, Dr Christian Rödelsperger, Dr Neel Prabh, Dr Adrian Streit, Tess Renahan, and Metta Riebesell for discussion. SS performed the EMS screen, did the mapping experiments, obtained and analyzed the RNA sequencing data, measured worm body sizes, and created [Figs. 2 and 4](#). TT obtained all landmark data, performed the complete geometric morphometric and clustering analyses, and created [Figs. 1, 3, and 5](#). HW generated and injected the plasmid constructs for transgenesis and CRISPR-RNP-complexes for mutagenesis. EJR identified the cells in the transcriptional reporter lines. RJS conceptualized and supervised the project. SS, TT, and RJS wrote the manuscript.

Funding

This study was funded by the Max Planck Society. SS was supported by the China Scholarship Council. TT was supported by the IMPRS "From Molecules to Organisms."

Conflicts of interest statement

The authors declare no conflict of interest.

Literature cited

- Adams D, Collyer M, Kaliontzopoulou A, Baken E. Geomorph: Software for Geometric Morphometric Analyses. R Package Version 4.0.1; 2021. <https://cran.r-project.org/package=geomorph>.
- Adams DC, Otárola-Castillo E., geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods Ecol Evol.* 2013;4(4):393–399.
- Albertson DG, Thomson JN. The pharynx of *Caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci.* 1976;275(938):299–325.
- Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 2001;26(1):32–46.
- Anderson MJ. Permutational multivariate analysis of variance (PERMANOVA). In: Balakrishnan N, editor. New Jersey: Wiley & Sons. Wiley Statsref: Statistics Reference Online; 2014. p. 1–15. <https://onlinelibrary.wiley.com/doi/full/10.1002/9781118445112.stat07841>.
- Anderson RC. Nematode Parasites of Vertebrates. Oxford, UK: CABI Publishing; 2000.
- Baldwin JG, Giblin-Davis RM, Eddleman CD, Williams DS, Vida JT, Thomas WK. The buccal capsule of *Aduncospiculum halicti* (Nematoda: Diplogastridae): an ultrastructural and molecular phylogenetic study. *Can J Zool.* 1997;75(3):407–423.
- Bento G, Ogawa A, Sommer RJ. Co-option of the endocrine signaling module dafachronic acid-DAF-12 in nematode evolution. *Nature* 2010;466(7305):494–497.
- Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics* 1974;77(1):71–94.
- Bui LT, Ragsdale EJ. Multiple plasticity regulators reveal targets specifying an induced predatory form in nematodes. *Mol. Biol. Evol.* 2019;36(11):2387–2399.
- Burr AHJ, Baldwin JG. The nematode stoma: homology of cell architecture with improved understanding by confocal microscopy of labeled boundaries. *J Morphol.* 2016;277(9):1168–1186.
- Casasa S, Biddle JF, Koutsovoulos GD, Ragsdale EJ. Polyphenism of a novel trait integrated rapidly evolving genes into ancestrally plastic networks. *Mol Biol Evol.* 2021;38(2):331–343.
- Cohen JD, Sundaram MV. *C. elegans* apical extracellular matrix shape epithelia. *J Dev Biol.* 2020;8(4):23.
- Collyer ML, Sekora DJ, Adams DC. A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity (Edinb)* 2015;115(4):357–365.
- Collyer ML, Adams DC. RRPP: an R package for fitting linear models to high-dimensional data using residual randomization. *Methods Ecol Evol.* 2018;9(7):1772–1779.
- Collyer ML, Adams DC. RRPP: Linear Model Evaluation with Randomized Residuals in a Permutation Procedure. R Package Version 1.1.0; 2021. <https://cran.r-project.org/package=RRPP>.
- De Ley P, Van de Velde MC, Mounport D, Baujard P, Coomans A. Ultrastructure of the stoma in Cephalobidae, Panagrolaimidae and Rhabditidae, with a proposal for a revised stoma terminology in Rhabditida (Nematoda). *Nematologica* 1995;41(1–4):153–182.
- Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, Dinkelacker I, Fulton L, Fulton R, Godfrey J, Minx P, et al. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat Genet.* 2008;40(10):1193–1198.
- Fraley C, Raftery A, Scrucca L, Murphy TB, Fop M. mclust: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. R Package Version 5.4.7.; 2021. <https://cran.r-project.org/web/packages/mclust>.
- Fürst von Lieven A, Sudhaus W. Comparative and functional morphology of the buccal cavity of Diplogastridae (Nematoda) and a

- first outline of the phylogeny of this taxon. *J Zool. Syst.* 2000; 38(1):37–63.
- Hall DH, Altun ZF. *C. elegans Atlas*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2008.
- Han Z, Lo W-S, Lightfoot JW, Witte H, Sun S, Sommer RJ. Improving transgenesis efficiency and CRISPR-associated tools through codon optimization and native intron addition in nematodes. *Genetics* 2020;216(4):947–956.
- Harry CJ, Messar SM, Ragsdale EJ. Comparative reconstruction of the predatory feeding structure of the polyphonic nematode *Pristionchus pacificus*. *bioRxiv* 2021. <https://doi.org/10.1101/2021.10.15.464383>.
- Kenning C, Kipping I, Sommer RJ. Mutations with altered gross-morphology in the nematode *Pristionchus pacificus*. *Genesis* 2004; 40(3):176–183.
- Lee DL. *The Biology of Nematodes*. London: Taylor & Francis; 2002.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25(14): 1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25(16):2078–2079.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
- Malakhov VV. *Nematodes*. Washington: Smithsonian Institution Press; 1994.
- Meeuse MW, Hauser YP, Morales Moya LJ, Hendriks G-J, Eglinger J, Bogaarts G, Tsiairis C, Großhans H. Developmental function and state transitions of a gene expression oscillator in *Caenorhabditis elegans*. *Mol. Syst. Biol.* 2020;16:1–21.
- Mitteroecker P, Gunz P, Bernhard M, Schaefer K, Bookstein FL. Comparison of cranial ontogenetic trajectories among great apes and humans. *J Hum Evol.* 2004;46(6):679–698.
- Mitteroecker P, Gunz P, Windhager S, Schaefer K. A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. *HYSTRIX*. 2013;24: 59–66.
- Müller MR. Genetic aspects of the development of the mouth form dimorphism in *Pristionchus pacificus*. Diploma thesis, University of Tübingen, Tübingen; 2010.
- Nakayama K-I, Ishita Y, Chihara T, Okumura M. Screening for CRISPR/Cas9-induced mutations using a co-injection marker in the nematode *Pristionchus pacificus*. *Dev Genes Evol.* 2020;230(3): 257–264.
- Page AP, Stepek G, Winter AD, Pertab D. Enzymology of the nematode cuticle: a potential drug target? *Int J Parasitol Drugs Drug Resist.* 2014;4(2):133–141.
- Phani V, Shivakumara TN, Davies KG, Rao U. Knockdown of a mucin-like gene in *Meloidogyne incognita* (Nematoda) decreases attachment of endospores of *Pasteuria penetrans* to the infective juveniles and reduces nematode fecundity. *Mol Plant Pathol.* 2018;19(11):2370–2383.
- Quach KT, Chalasan SH. Intraguild predation between *Pristionchus pacificus* and *Caenorhabditis elegans*: a complex interaction with the potential for aggressive behavior. *J Neurogenetics.* 2020. <https://doi.org/10.1080/01677063.2020.1833004>.
- Rae R, Witte H, Rödelberger C, Sommer RJ. The importance of being regular: *Caenorhabditis elegans* and *Pristionchus pacificus* defecation mutants are hypersusceptible to bacterial pathogens. *Int J Parasitol.* 2012;42(8):747–753.
- Ragsdale EJ, Baldwin JG. Resolving phylogenetic incongruence to articulate homology and phenotypic evolution: a case study from Nematoda. *Proc Biol Sci.* 2010;277(1686):1299–1307.
- Ragsdale EJ, Müller MR, Rödelberger C, Sommer RJ. A developmental switch coupled to the evolution of plasticity acts through a sulfatase. *Cell* 2013;155(4):922–933.
- R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; European Environment Agency, 2021.
- Rödelberger C, Meyer JM, Prabh N, Lanz C, Bemm F, Sommer RJ. Single-molecule sequencing reveals the chromosome-scale genomic architecture of the nematode model organism *Pristionchus pacificus*. *Cell Rep.* 2017;21(3):834–844.
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 2012;9(7):676–682.
- Schlager S. Morpho and Rvcg-shape analysis in R: R-packages for geometric morphometrics, shape analysis and surface manipulations. In: G Zheng, S Li G Szekely, editors. *Statistical Shape and Deformation Analysis*. Academic Press; 2017. p. 217–256.
- Schlager S, Jefferis G, Ian D. Morpho: Calculations and Visualisations Related to Geometric Morphometrics. R Package Version 2.9; 2021. <https://cran.r-project.org/web/packages/Morpho>.
- Sieriebriennikov B, Prabh N, Dardiry M, Witte H, Röseler W, Kieninger MR, Rödelberger C, Sommer RJ. A developmental switch regulating phenotypic plasticity is a part of a conserved multi-gene locus. *Cell Rep.* 2018;23(10):2835–2843.
- Sieriebriennikov B, Sun S, Lightfoot JW, Witte H, Moreno E, Rödelberger C, Sommer RJ. Conserved hormone-receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell. *PLoS Genet.* 2020;16(4):e1008687.
- Simmer F, Moorman C, van der Linden AM, Kuijk E, van den Berghe PVE, Kamath RS, Fraser AG, Ahringer J, Plasterk RHA. Genome-wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveals novel gene functions. *PLoS Biol.* 2003;1(1):e12.
- Sommer RJ. Phenotypic plasticity: from theory and genetics to current and future challenges. *Genetics* 2020;215(1):1–13.
- Sommer RJ, Carta LK, Kim S-Y, Sternberg PW. Morphological, genetic and molecular description of *Pristionchus pacificus* sp. n. (Nematoda, Diplogastridae). *Fundam. Appl. Nematol.* 1996;19: 511–521.
- Sun S, Rödelberger C, Sommer RJ. Single worm transcriptomics identifies a developmental core network of oscillating genes with deep conservation across nematodes. *Genome Res.* 2021;31(9): 1590–1601.
- Susoy V, Ragsdale EJ, Kanzaki N, Sommer RJ. Rapid diversification associated with a macroevolutionary pulse of developmental plasticity. *eLIFE* 2015;4:e05463.
- Theska T, Sieriebriennikov B, Wighard SS, Werner M, Sommer RJ. Geometric morphometrics of microscopic animals as exemplified by model nematodes. *Nat Protoc.* 2020;15(8):2611–2644.
- van den Hoogen J, Geisen S, Routh D, Ferris H, Traunspurger W, Wardle DA, de Goede RGM, Adams BJ, Ahmad W, Andriuzzi WS, et al. A global database of soil nematodes abundance and functional group composition. *Nature* 2019;572(7768):194–198.
- Wagner CE, Wheeler KM, Ribbeck K. Mucins and their role in shaping the function of mucus barriers. *Annu Rev Cell Dev Biol.* 2018;34: 189–215.
- Werlang C, Carcarmo-Oyarce G, Ribbeck K. Engineering mucus to study and influence the microbiome. *Nat Rev Mater.* 2019;4(2): 134–145.

- Werner MS, Sieriebriennikov B, Loschko T, Namdeo S, Lenuzzi M, Dardiry M, Renahan T, Sharma DR, Sommer RJ. Environmental influence on *Pristionchus pacificus* mouth-form through different culture methods. *Sci Rep.* 2017;7(1):7207.
- Witten J, Ribbeck K. The particle in the spider's web: transport through biological hydrogels. *Nanoscale* 2017;9(24):8080–8095.
- Wright KA. Functional organization of the nematode's head. In: NA Croll, editor. *The Organization of Nematodes*. London (UK): Academic Press; 1976. p. 71–106.
- Wright KA, Thomson JN. The buccal capsule of *Caenorhabditis elegans* (Nematoda: Rhabditidae): an ultrastructural study. *Can J Zool.* 1981;59(10):1952–1961.

Communicating editor: B. Grant

Supplementary Material

The oscillating Mucin-type protein DPY-6 has a conserved role in nematode mouth and cuticle formation

Shuai Sun, Tobias Theska, Hanh Witte, Erik J. Ragsdale & Ralf J. Sommer

Content

Geometric morphometrics – detailed methodology	2
Supplementary Figures 1-6	9
Supplementary Tables 1-31	15

Detailed methods for morphological analyses:

In order to quantify phenotypic differences in the mouths of wild-type and mutant worms, we used our recently published protocol for shape analysis in microscopic animals (Theska et al. 2020). We generally follow the exact descriptions provided in this protocol, with only a few minor modifications. In the following paragraphs we provide a detailed account on how we handled our data, which settings we used to analyze morphological differences in this data, and which aspects we added to the previously published pipeline.

Replicability of our landmark data

First, we assessed the replicability of the landmark configurations we defined for both species (see Table S4) by randomly sampling 10 adult hermaphrodites of both *P. pacificus* [Eu] and *C. elegans*, encrypting their names in the image stacks, and obtaining landmark data from each of these specimens in three technical replicates. Our landmark annotator took 15 min breaks in between labeling individual worms and each worm was labeled only once per day. The three technical replicates of each landmark configuration were acquired on three subsequent days (i.e., after 24h breaks). We then calculated the measurement error (%ME)² for each of the two species of interest, following the procedure described by Claude (Claude 2008). Additionally, we perform a GPA on these 'replicate' data sets and visualize the amount of shape variation among replicates in a PCA plot (Figure S5). We found that our landmark configurations are highly similar (i.e., the replicates of each specimen are very close to each other in the shape space, while the main shape variation results from natural shape differences among worms) and we were able to confirm a small %ME for both *P. pacificus* (16.3%), and *C. elegans* (14.4%). This is further supported by multivariate statistical analysis: for both species the PERMANOVA-RRPP (see below) approach revealed that the effect that the factor 'specimen' has on the phenotypic response variable 'mouth shape' is much stronger than the effect that the factor 'replicate' has (*P. pacificus*: $shape \sim specimen$: $Z = 9.2$ and $P = 0.000999$, $shape \sim replicate$: $Z = -2.4$ and $P = 0.993$; *C. elegans*: $shape \sim specimen$: $Z = 8.8$ and $P = 0.000999$, $shape \sim replicate$: $Z = -3.0$ and $P = 1.0$; see Tables S8-11). Thus, we conclude that our landmark data is highly replicable.

Additional information on the geometric morphometric analysis of shapes

Initially, we generated landmark data for 200 animals (40 specimens per strain) and created two separate data sets: (I) a *P. pacificus* data set comprising the wild-type strain (PS312), a frameshift mutant of *Ppa-dpy-6* (*tu1645*), and the 'rescue' strain RS4003 (*tu1696*); and (II) a *C. elegans* data set comprising the wild-type N2 and the *Cel-dpy-6* mutant CB14. Analyses were performed in R (ver. 4.1.1) (R Core Team 2021) on a MacBook Pro (13-inch, 2020, macOS Catalina ver. 10.15.6) equipped with a 2 GHz Quad-Core Intel Core i5 processor, 16GB RAM (3733 MHz LPDDR4X), and an Intel Iris Plus graphics card (1536 MB). General Procrustes analysis (GPA) was performed with the *gpa* function of the GEOMORPH package (ver. 4.0.1.) (Adams and Otarola-Castill 2013; Adams *et al.* 2021). Procrustes coordinates were projected into a linear tangent space to obtain Euclidean distances for statistical analysis. We performed GPA on each of the two landmark data sets and checked for outlier specimens using the *plotOutliers* function of GEOMORPH. In the *C. elegans* data set, we identified and subsequently removed five outliers among the wild-type animals, and in the *P. pacificus* data set we identified and removed two outliers among the *Ppa-dpy-6* mutants. We re-ran the GPA on these 'outlier-corrected' landmark data sets (*C. elegans*: n=75; *P. pacificus*: n=118, Supplementary Data 2,3) and used the resulting shape data for all downstream analyses. Differences in mouth shape were visualized by plotting the first two principal components (PCs) of variation in a shape space. Additionally, wireframe representations of the nematode mouth were generated for the positive and negative extreme shapes along PC1 to visualize the differences which separate wild-type and mutant mouths in the shape space (Theska *et al.* 2020).

Multivariate statistical analysis of phenotypic data using PERMANOVA with RRPP

Phenotypic differences (i.e., mouth shape, form, and size) were assessed utilizing a distance-based permutational MANOVA (PERMANOVA) (Anderson *et al.* 2014; 2021) with a randomized residual permutation procedure (RRPP) (Collyer *et al.* 2015; 2018; 2021). This test assesses the terms of a statistical model using the Euclidean distances among specimens as a source of information, rather than explained covariance matrices among variables (Anderson *et al.* 2014;2021). RRPP is an approach that uses a resampling experiment to randomize the row vectors of a matrix containing residuals from a reduced model to calculate pseudorandom values for estimation of effects from a full model (Collyer *et al.* 2015; 2018;

2021). For simplicity, we only reported relative effect sizes (Z -scores) and P -values in the main text, but additional statistics are provided in the Supplementary Tables below. The Z -scores for all PERMANOVA effects were calculated as the number of standard deviations of observed F -values from empirically generated sample distributions of random values found via RRPP (i.e., a relative effect size of $Z = 2.0$ has a signal that is two times larger than the standard deviation). The obtained P -values represent the probability of finding a random value larger than the observed value.

Before statistical analysis, we considered any PERMANOVA effect as 'incompatible with the null hypothesis' if (I) the relative effect size was at least twice as large as the standard deviation (i.e., Z -score ≥ 2.0) and (II) the P -value was below a type I error rate of $\alpha = 0.05$ (i.e., 5%). We assessed the compatibility of the data with a null hypothesis through RRPP in 10,000 iterations using the *procD.lm* function of GEOMORPH under an arbitrarily set seed of '12345'. Generally, an association (i.e., $Z \geq 2.0$ and $P < 0.05$) between a phenotypic response variable 'Y' (here: mouth shape, mouth (centroid) size, or mouth form) and the independent variable 'X' (here: strain or, for allometry, centroid size – see below) suggests that our data is incompatible with the null hypothesis of 'no difference in Y between the wild-type and the mutant strains' and indicates compatibility with the alternative hypothesis that 'there is a mutant phenotype in Y'. The absence of an association between the response variable of interest and the independent variable 'strain' (i.e., $Z < 2.0$ and $P \geq 0.05$) indicates compatibility of our data with the null hypothesis or a lack of statistical power to detect phenotypic differences in 'Y' between wild-type and mutant animals. In our figures, incompatibility of our phenotypic data with a corresponding null hypothesis is indicated with an asterisk '*' and compatibility with a period '.'. Pairwise comparisons of LS means among groups were performed with the *pairwise* function of the RRPP package (ver. 1.1.0). The full statistics (including degrees of freedom, Sum of Squares, Mean Squares, coefficients of determination, F statistics, relative effect sizes, and P -values) for all PERMANOVA-RRPP assessments of mouth form, shape, and size are reported in the table supplements 12-17. Full summary statistics of the corresponding pairwise comparisons are reported in the table supplements 18-23.

Additional information on geometric morphometric analysis of mouth form and size

In addition to the analyses of mouth shape and mouth allometry, we also assessed differences in mouth form among specimens through a 'size-shape' PCA (*sensu* Mitteroecker *et al.*). We simply appended a size variable (here, the log-transformed centroid size) to our shape data matrices, in order to obtain form data sets (form is defined as shape plus size, see Klingenberg 2016). We subsequently performed a PCA on the form data sets to visualize morphological differences in a 'size-shape space'. Additionally, we assessed differences in mouth size using the centroid size of the original shape data set as source of information. The centroid size is the square root of the sum of squared distances of each landmark to the corresponding centroid in the landmark configuration, and thus provides information on the overall size of the nematode mouth (Theska *et al.* 2020).

Additional information on group structure analysis by clustering

We used model-based clustering to reveal the number and structure of separate groups with similar morphology in our shape and form data sets. This approach is opposite to the PERMANOVA-RRPP in the sense that it does not define groups (e.g., strains) *a priori* to than subsequently test for differences among these predefined groups. The clustering analysis was performed in R (ver. 4.1.1.) using the MCLUST (ver. 5.4.7) package (Scrucca *et al.* 2016; Fraley *et al.* 2021). All procedures were carried out exactly following the previous protocol¹. To avoid overparameterization of our models, we only used 'meaningful' PCs of each shape data set as input for the clustering approaches. 'Meaningful' PCs were determined with the *getMeaningfulPCs* function of the MORPHO package (ver. 2.9) (Schlager *et al.* 2017; 2021). This function calculates a threshold value that is based on a log-likelihood ratio and dependent on the sample size (e.g., an estimated threshold of 2.0 means that a PC can be considered 'meaningful' if it describes at least two times as much variation as the next PC) (Schlager *et al.* 2021; Bookstein 2014). We calculated the meaningful PCs separately for the shape PCA and size-shape PCAs we performed on both data sets.

Analyses of static allometric trajectories

Allometry, the relationship of biological shape with size, is a major factor driving morphological diversification (Klingenberg 2016; Esquerre *et al.* 2017). Thus, it is not surprising that allometry is frequently found in shape data sets, often underlying the variation

described by PC1 of the PCA performed on the shape data (Mitteroecker *et al.* 2004; Adams and Nistri 2010). Since *dumpy*-mutants are known to exhibit clear differences in adult size relative to wildtype worms, we wondered whether potential alterations of mouth morphology in these mutants can be explained by size as well. Here, we assessed the relationship of the adult mouth shape and its log-transformed centroid size (i.e., static allometry in adult worms) for each of the strains under investigation in form of a homogeneity of slopes (HOS) test based on the same distance-based PERMANOVA approach we used for shape analysis. We followed the exact descriptions to perform this test which are given in the "GEOMORPH version 3.1.0 assistance" vignette provided by M. Collyer (version from 01. April 2021, see 'How to perform allometry analyses' paragraph in <https://cran.r-project.org/web/packages/geomorph/vignettes/geomorph.assistance.html>). In short, we created linear model fits (using *procD.lm* of GEOMORPH) for a simple null model of allometry (shape ~ size) and two complex allometry model fits: (I) a common allometry model (shape ~ size + strain) and (II) a group-specific allometry model (shape ~ size * strain). Then, we compared the simple null model of allometry to the common allometry model using the *anova* function of the R STATS package. A large Z-score and a small P-value indicate a compatibility of our shape data with the common allometry model, suggesting that its underlying linear model are more appropriate to fit the data. Thus, in such a case, we favored the common allometry model over the simple null model and we subsequently compared the common allometry model (as the new null model) to the group-specific allometry model, using the same ANOVA approach. Again, we interpreted a large Z-score together with a small P-value to indicate a higher compatibility of our shape data with the group-specific allometry model (which assumes allometric slopes to be different among the strains), than with the common allometry model (which suggests that different strains have similar allometric growth patterns). Thus, low compatibility of our shape data with the group-specific allometry model (i.e., small Z-scores along with large P-values) indicates that allometric trajectories are similar across strains or that there is a lack of statistical power to detect group-specific allometric trends.

Our allometry model comparisons (i.e., the HOS test) revealed that for both data sets a common allometry model is the most appropriate (table supplements 24,25 and 28,29) to fit the shape data. Thus, within both species, the slopes of the static allometric trajectories of each strain are similar. Therefore, we chose the common allometry model for each of our two

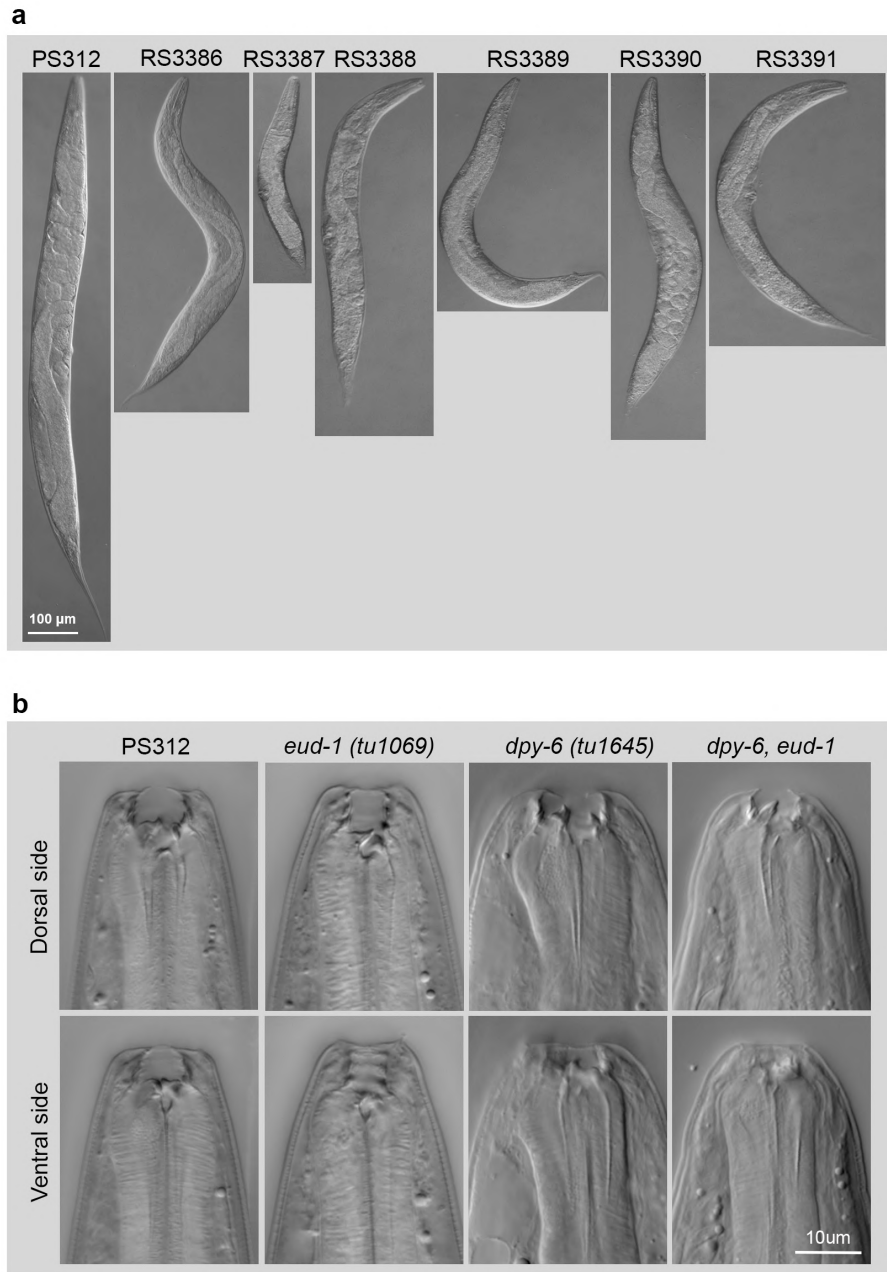
shape data sets and assessed the compatibility of our data with the respective allometry hypothesis through the aforementioned PERMANOVA-RRPP approach with 10,000 iterations using the *procD.lm* function of GEOMORPH under an arbitrarily set seed of '12345'. An association of shape and size ($Z \geq 2.0$ and $P < 0.05$) was interpreted as incompatible with the null hypothesis of 'isometric growth' and indicative of allometry (i.e., that the mouth shape changes with size), while the absence of an association ($Z < 2.0$ and $P \geq 0.05$) was regarded to be indicative of isometry or a lack of statistical power to detect allometry. The full statistical results of the allometry analysis for both species are provided in table supplements 26 and 30. Given that the HOS test revealed similar allometric slopes, we performed pairwise comparisons of the allometric data for our nematode strains by focusing on the LS means of these groups, rather than on slope features (e.g., slopes length, angle, or gradient). The summary statistics of these pairwise comparisons are reported in the table supplements 27 and 31.

References

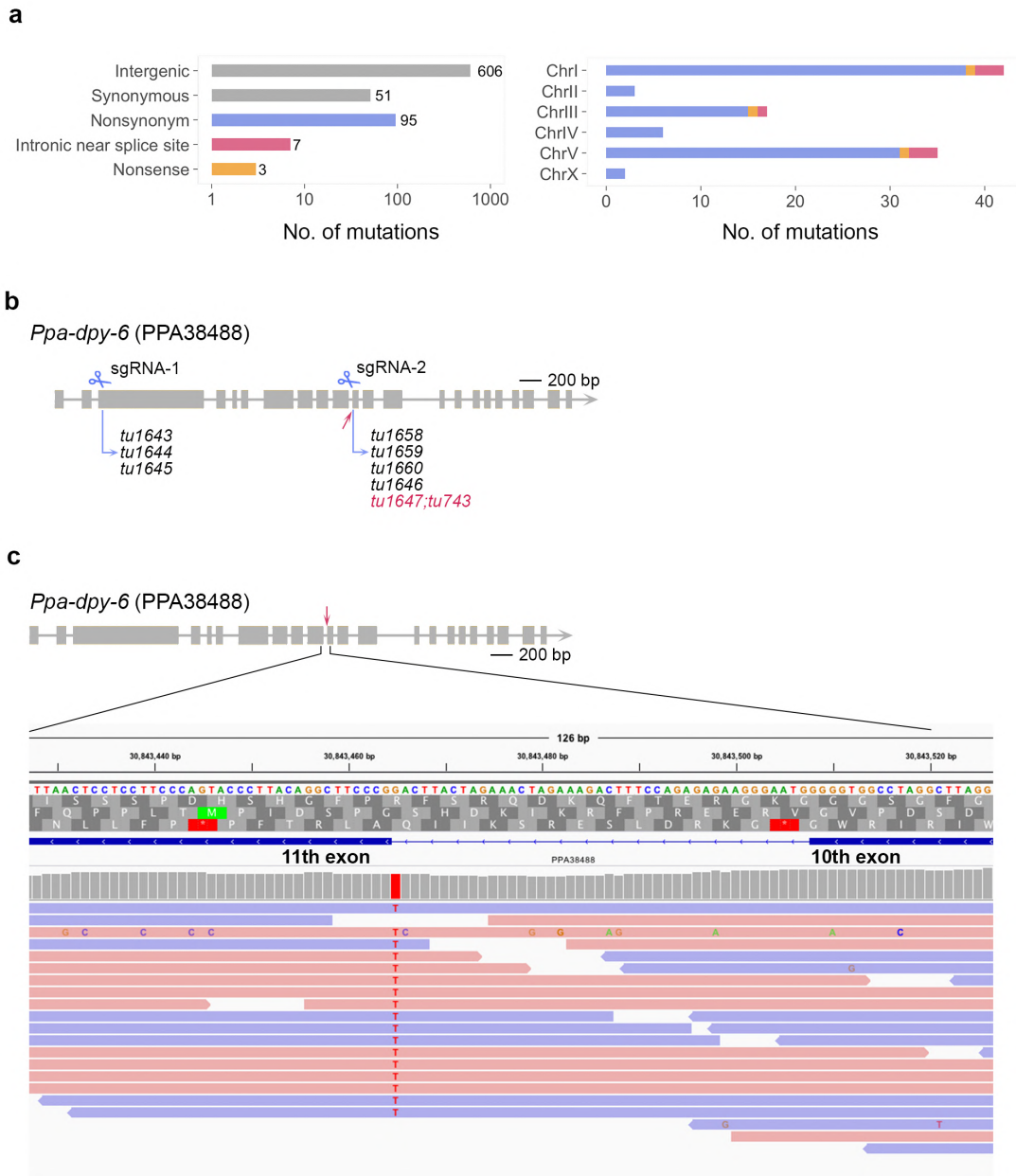
- Theska, T., Sieriebriennikov, B., Wighard, S. S., Werner, M.S., and R. J. Sommer, 2020 Geometric morphometrics of microscopic animals as exemplified by model nematodes. *Nat. Protoc.* **15**:2611–2644.
- Yezerinac, S., Loughheed, S. C, and P. Handford P.1992.Measurement error and morphometric studies: statistical power and observer experience. *Syst. Biol.* **41**:471–482.
- Claude, J., 2008 *Morphometrics with R*. (Springer).
- R Core Team, 2021 *R: A language and environment for statistical computing* (R Foundation for Statistical Computing). Ver 4.1.1.
- Adams, D. C., and E. Otárola-Castillo, 2013 geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods Ecol. Evol.* **4**:393–399.
- Adams, D., Collyer, M., Kaliontzopoulou, A, and E. Baken, 2021 Geomorph: Software for geometric morphometric analyses. *R package version* 4.0.1. <https://cran.r-project.org/package=geomorph>
- Anderson, M. J., 2021 A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**:32–46.
- Anderson, M. J., 2014 Permutational multivariate analysis of variance (PERMANOVA). in *Wiley Statsref: Statistics Reference Online* (eds. Balakrishnan, N. et al.) 1–15.
<https://onlinelibrary.wiley.com/doi/full/10.1002/9781118445112.stat07841>.
- Collyer, M. L., Sekora, D. J., and D. C. Adams, 2015 A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity* **115**:357–365.
- Collyer, M. L., and D. C. Adams, 2018 RRPP: An R package for fitting linear models to high-dimensional data using residual randomization. *Methods Ecol. Evol.* **9**:1772–1779.

- Collyer, M., and D. C. Adams, 2021. RRPP: linear model evaluation with randomized residuals in a permutation procedure. *R package version 1.1.0*. <https://cran.r-project.org/package=RRPP>.
- Mitteroecker, P., Gunz, P., Bernhard, M., Schaefer, K., and F. L. Bookstein, 2004 Comparison of cranial ontogenetic trajectories among great apes and humans. *J. Hum. Evol.* **46**:679–698.
- Klingenberg, C., 2016 P. Size, shape, and form: concepts of allometry in geometric morphometrics. *Dev. Genes Evol.* **226**:113–137.
- Scrucca, L., Fop, M., Murphy, T. B., and A. E. Raftery, 2016 mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J.* **8**:289.
- Fraley, C., Raftery, A., Scrucca, L., Murphy, T. B., and M. Fop, 2021 mclust: Normal mixture modeling for model-based clustering, classification, and density estimation. *R package version 5.4.7*. <https://cran.r-project.org/web/packages/mclust>.
- Schlager, S., 2017 Morpho and Rvcg–Shape Analysis in R: R-Packages for geometric morphometrics, shape analysis and surface manipulations. in *Statistical Shape and Deformation Analysis* (eds. Zheng, G., Li, S. & Szekely, G.) 217–256 (Academic Press).
- Schlager, S., Jefferis, G., and D. Ian, 2021 Morpho: Calculations and Visualisations Related to Geometric Morphometrics. *R package version 2.9*. <https://cran.r-project.org/web/packages/Morpho>.
- Bookstein, F. L., 2014 *Measuring and reasoning: Numerical inference in the sciences*. (Cambridge University Press).
- Esquerré, D., Sherratt, E., and J. S. Keogh, 2017 Evolution of extreme ontogenetic allometric diversity and heterochrony in pythons, a clade of giant and dwarf snakes. *Evolution* **71**:2829–2844.
- Adams, D. C., and A. Nistri, 2010 Ontogenetic convergence and evolution of foot morphology in European cave salamanders (Family: Plethodontidae). *BMC Evol. Biol.* **10**:1–10.

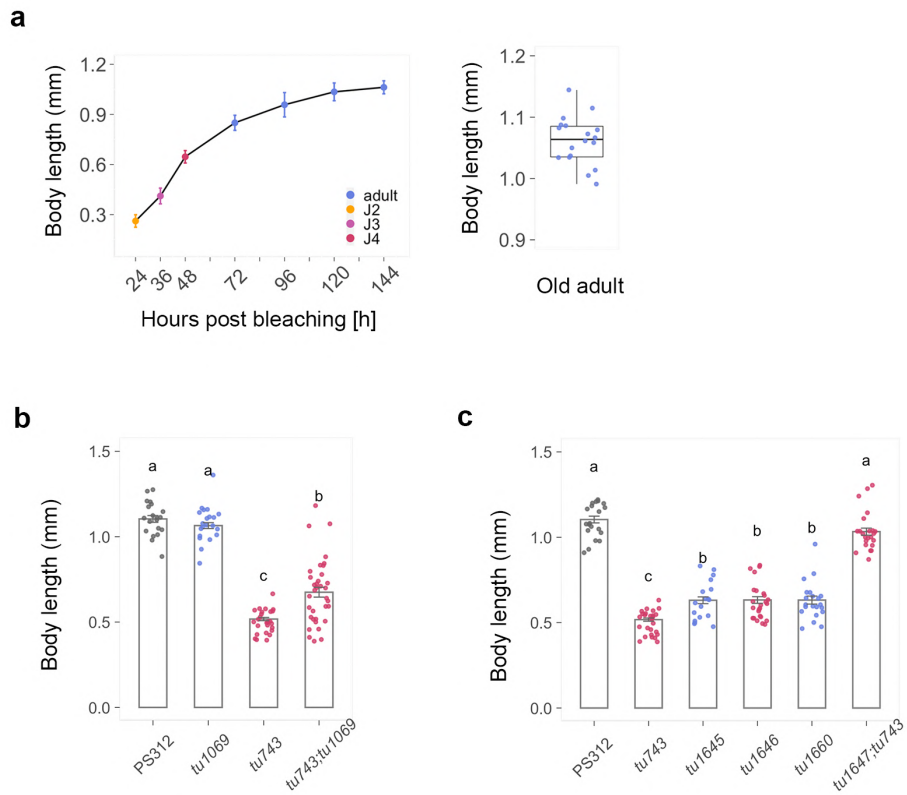
Supplementary Figures



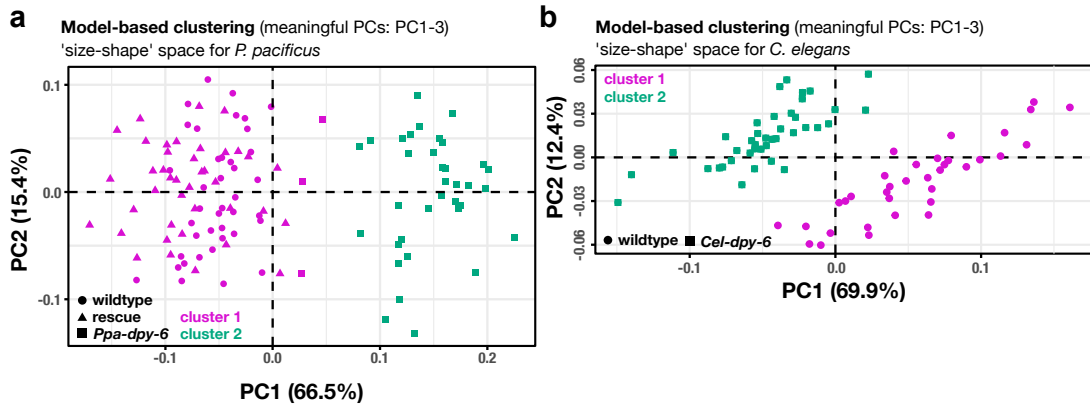
Supplementary Figure 1. Additional images. **a**, Body size of wild-type (PS312) and the six mutant strains (designated as RS3386-RS3391). **b**, Mouth structure of wild-type eurystomatous (Eu) morph, stenostomatous (St) morph of the *Ppa-eud-1 (tu1069)* mutant, Eu morph with reduced cheilostom of the *Ppa-dpy-6(tu1645)* mutant (strain RS3887), and the St morph of the *Ppa-dpy-6; Ppa-eud-1* double mutant with reduced cheilostom in two focal planes.



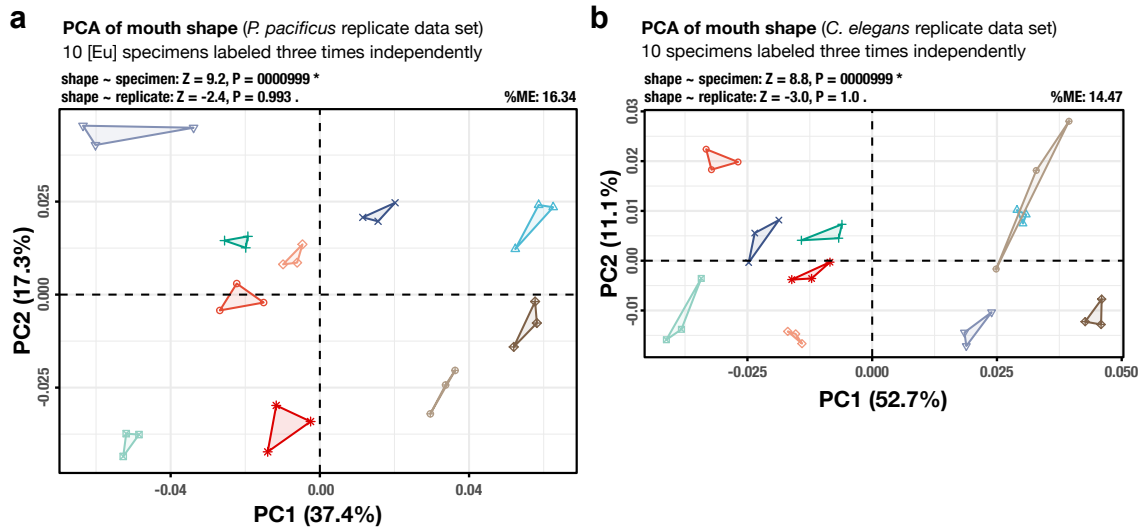
Supplementary Figure 2. Mapping of *tu743* and CRISPR/Cas9 mutagenesis on *Ppa-dpy-6*. **a**, The left bar plot shows the distribution of single nucleotide polymorphisms in the *tu743* mutant relative to PS312 wild-type with five classes, including intergenic, synonymous, nonsynonym, intronic near splice site, and nonsense mutations. The right plot displays the distribution of the three classes, which result in changes in the amino acid sequences among six chromosomes. **b**, Schematic gene model of *Ppa-dpy-6* and the design of two different sgRNAs and CRISPR mutants obtained from these sgRNA. **c**, Screenshot of the IGV profiles of the reads of the RS3887 *tu743* mutant strain at the *Ppa-dpy-6* locus indicating the T > C substitution in intron 10 at the splice site of *Ppa-dpy-6*.



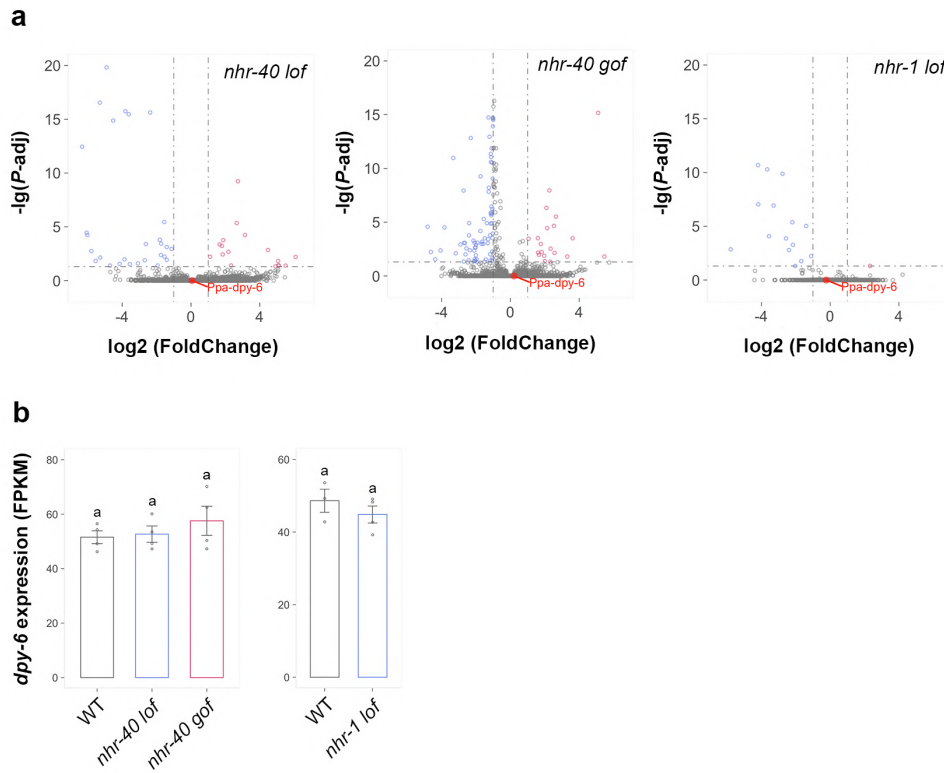
Supplementary Figure 3. Body length of *Ppa-dpy-6* associated mutants. **a**, The left line plot presents the differences in the length of the whole body of wild-type strain (PS312) at seven distinct developmental time points, containing post-bleaching 24 h (J2), 36 h (J3), 48 h (J4), 72 h (one-day old adult), 96 h (two-day old adult), 120 h (three-day old adult), and 144 h (four-day old adult). Worms were cultured on OP50 at 20 °C, more than 20 replicates for each time point. The right box plot displays the distribution of the body length of old adults (> five-day old adult), $n > 20$. **b**, Bar plot shows the length of worm body between wild-type (PS312), *Ppa-eud-1* mutant (*tu1069*), *Ppa-dpy-6* (*tu743*), and the double mutant (*tu743; tu1069*). **c**, Bar plot displays length of worm body of wild-type (PS312), and the five *Ppa-dpy-6* mutant strains. Turkey's multiple comparison was tested by the Turkey's HSD test, letters upon each volume indicate significant difference between the means, p -adjusted value < 0.05 , $n > 20$.



Supplementary Figure 4. Model-based clustering performed on mouth **form** data sets. **a**, clusters identified in the *P. pacificus* form data set. **b**, clusters identified in the *C. elegans* form data set. Only 'meaningful' principal components of form variation were used as input variables.



Supplementary Figure 5. Replicability of landmark data for the nematode mouths. **a**, PCA of shape variation in the *P. pacificus* replicate data set. **b**, PCA of shape variation in the *C. elegans* replicate data set. Different colors and symbols represent different specimens. Replicates of the same specimens are connected by convex hulls. %ME, percent measurement error according to Yezerinac *et al.*² (i.e., the amount of shape variation in the replicate data set which is due to annotation imprecision). Z-scores (relative effect sizes) and P-values were obtained with PERMANOVA-RRPP (see details above, Supplementary Tables S8-S11).



Supplementary Figure 6. Gene expression of *Ppa-dpy-6* in nuclear-hormone-receptor (NHR) mutant backgrounds. **a**, Volcano plots reveals that *Ppa-dpy-6* is not among the differentially expressed genes (DEG) in *nhr-40* and *nhr-1* loss-of-function or *nhr-40* gain-of-function mutants. Downregulated genes in blue, upregulated genes in red, and non-differentially expressed genes in grey. Fold change cutoff is 2, adjusted P-value cut-off is 0.05. **b**, Comparison of gene expression of *Ppa-dpy-6* in *nhr-40* loss-of-function, *nhr-40* gain-of-function mutant, and wild-type (left), and in *nhr-1* loss of function mutant and wild-type (right). All samples used in **a** and **b** are at mouth-form determination period (Sieriebriennikov et al. 2020), *lof* = loss-of-function, *gof* = gain-of-function. Tukey's multiple comparison were tested by the Tukey's HSD test, and the letters upon each volume means significant difference between the means, p-adjusted value < 0.05.

Supplementary Tables

Supplementary Table 1. Detailed information on *Ppa-dpy-6* associated mutants

Strain ID (RS)	Strain ID (tu)	Mutagen	Background strain	Gene locus (El_Paco_V3)	Modifications	Source
RS3886	<i>tu742</i>	EMS				This paper
RS3887	<i>tu743</i>	EMS				This paper
RS3888	<i>tu744</i>	EMS				This paper
RS3889	<i>tu745</i>	EMS				This paper
RS3890	<i>tu746</i>	EMS				This paper
RS3891	<i>tu747</i>	EMS				This paper
RS3899	<i>tu1643</i>	CRISPR	PS312	PPA38488	8 bp deletion in the third exon of the <i>Ppa-dpy-6</i>	This paper
RS3900	<i>tu1644</i>	CRISPR	PS312	PPA38488	28 bp insertion in the third exon of the <i>Ppa-dpy-6</i>	This paper
RS3905	<i>tu1645</i>	CRISPR	PS312	PPA38488	2 bp deletion in the third exon of the <i>Ppa-dpy-6</i>	This paper
RS3910	<i>tu1646</i>	CRISPR	PS312	PPA38488	create the C>T substitution of the <i>Ppa-dpy-6</i> in PS312	This paper
RS3901	<i>tu1647;tu743</i>	CRISPR	<i>tu743</i>	PPA38488	rescue <i>Ppa-dpy-6</i> in <i>tu743</i> strain	This paper
RS3914	<i>tu1658</i>	CRISPR	PS312	PPA38488	1 bp insertion in the 11th exon of the <i>Ppa-dpy-6</i>	This paper
RS3915	<i>tu1659</i>	CRISPR	PS312	PPA38488	37 bp insertion in the 11th exon of the <i>Ppa-dpy-6</i>	This paper
RS3916	<i>tu1660</i>	CRISPR	PS312	PPA38488	4 bp deletion in the 11th exon of the <i>Ppa-dpy-6</i>	This paper
RS3111	<i>tu1069</i>	CRISPR	PS312	PPA43535	10 bp deletion in the third exon of the <i>Ppa-eud-1</i>	Sieriebriennikov et al. 2018
RS3924	<i>RS3887;tu1069</i>	EMS, CRISPR		PPA38488; PPA43535	Introduce the CRISPR mutation of <i>Ppa-eud-1</i> into the RS3887	This paper

Supplementary Table 2. Mouth-form ratio of the EMS strains *tu742-tu747*

Strain ID	Individual ID	No. of Eu	No. of St	% Eu	Strain ID	Individual ID	No. of Eu	No. of St	% Eu
PS312	PS312_ind0 1	20	0	100.0	<i>tu745</i>	<i>tu745_ind0 1</i>	0	20	0.0
PS312	PS312_ind0 2	19	1	95.0	<i>tu745</i>	<i>tu745_ind0 2</i>	0	21	0.0
PS312	PS312_ind0 3	20	0	100.0	<i>tu745</i>	<i>tu745_ind0 3</i>	0	16	0.0
PS312	PS312_ind0 4	20	0	100.0	<i>tu745</i>	<i>tu745_ind0 4</i>	0	20	0.0
PS312	PS312_ind0 5	20	0	100.0	<i>tu745</i>	<i>tu745_ind0 5</i>	0	20	0.0
PS312	PS312_ind0 6	20	0	100.0	<i>tu745</i>	<i>tu745_ind0 6</i>	0	20	0.0
PS312	PS312_ind0 7	19	1	95.0	<i>tu745</i>	<i>tu745_ind0 7</i>	0	20	0.0
PS312	PS312_ind0 8	20	0	100.0	<i>tu745</i>	<i>tu745_ind0 8</i>	0	20	0.0
PS312	PS312_ind0 9	20	0	100.0	<i>tu745</i>	<i>tu745_ind0 9</i>	0	21	0.0
PS312	PS312_ind1 0	20	0	100.0	<i>tu745</i>	<i>tu745_ind1 0</i>	0	20	0.0
<i>tu742</i>	<i>tu742_ind0 1</i>	0	20	0.0	<i>tu746</i>	<i>tu746_ind0 1</i>	4	27	12.9
<i>tu742</i>	<i>tu742_ind0 2</i>	1	25	3.8	<i>tu746</i>	<i>tu746_ind0 2</i>	6	17	26.1
<i>tu742</i>	<i>tu742_ind0 3</i>	2	22	8.3	<i>tu746</i>	<i>tu746_ind0 3</i>	5	21	19.2
<i>tu742</i>	<i>tu742_ind0 4</i>	1	19	5.0	<i>tu746</i>	<i>tu746_ind0 4</i>	5	19	20.8
<i>tu742</i>	<i>tu742_ind0 5</i>	0	20	0.0	<i>tu746</i>	<i>tu746_ind0 5</i>	6	15	28.6
<i>tu742</i>	<i>tu742_ind0 6</i>	1	19	5.0	<i>tu746</i>	<i>tu746_ind0 6</i>	2	18	10.0
<i>tu742</i>	<i>tu742_ind0 7</i>	0	20	0.0	<i>tu746</i>	<i>tu746_ind0 7</i>	14	7	66.7
<i>tu742</i>	<i>tu742_ind0 8</i>	1	20	4.8	<i>tu746</i>	<i>tu746_ind0 8</i>	4	16	20.0
<i>tu742</i>	<i>tu742_ind0 9</i>	1	21	4.5	<i>tu746</i>	<i>tu746_ind0 9</i>	8	10	44.4
<i>tu742</i>	<i>tu742_ind1 0</i>	0	20	0.0	<i>tu746</i>	<i>tu746_ind1 0</i>	2	18	10.0
<i>tu743</i>	<i>tu743_ind0 1</i>	18	2	90.0	<i>tu747</i>	<i>tu747_ind0 1</i>	16	6	72.7
<i>tu743</i>	<i>tu743_ind0 2</i>	19	1	95.0	<i>tu747</i>	<i>tu747_ind0 2</i>	22	6	78.6
<i>tu743</i>	<i>tu743_ind0 3</i>	19	1	95.0	<i>tu747</i>	<i>tu747_ind0 3</i>	13	21	38.2
<i>tu743</i>	<i>tu743_ind0 4</i>	17	3	85.0	<i>tu747</i>	<i>tu747_ind0 4</i>	7	13	35.0
<i>tu743</i>	<i>tu743_ind0 5</i>	23	1	95.8	<i>tu747</i>	<i>tu747_ind0 5</i>	13	7	65.0
<i>tu743</i>	<i>tu743_ind0 6</i>	19	1	95.0	<i>tu747</i>	<i>tu747_ind0 6</i>	18	4	81.8
<i>tu743</i>	<i>tu743_ind0 7</i>	20	2	90.9	<i>tu747</i>	<i>tu747_ind0 7</i>	18	2	90.0
<i>tu743</i>	<i>tu743_ind0 8</i>	21	1	95.5	<i>tu747</i>	<i>tu747_ind0 8</i>	11	7	61.1

tu743	tu743_ind09	18	2	90.0
tu743	tu743_ind10	19	1	95.0
tu744	tu744_ind01	9	9	50.0
tu744	tu744_ind02	5	12	29.4
tu744	tu744_ind03	8	19	29.6
tu744	tu744_ind04	1	25	3.8
tu744	tu744_ind05	10	11	47.6
tu744	tu744_ind06	6	15	28.6
tu744	tu744_ind07	2	23	8.0
tu744	tu744_ind08	3	20	13.0
tu744	tu744_ind09	10	14	41.7
tu744	tu744_ind10	4	16	20.0

tu747	tu747_ind09	11	12	47.8
tu747	tu747_ind10	18	2	90.0

Supplementary Table 3. List of non-synonymous substitutions, nonsense substitutions and the substitutions in intronic near splice site in *tu743* mutant

Chromosome	Position	Ref	Mut	Gene ID (El Paco V3)	Classification	Putative <i>C. elegans</i> ortholog
ChrI	1626620	G	A	ppa_stranded_DN18867_c0_g1_i1	intronic_near_splice_sit	spd-1
ChrI	3084346	C	T	PPA38488	intronic_near_splice_sit	dpy-6
ChrI	3824021	C	T	PPA24475	intronic_near_splice_sit	
ChrIII	129246	G	A	PPA17835	intronic_near_splice_sit	let-805
ChrV	5869190	C	T	PPA03558	intronic_near_splice_sit	F56F10.1
ChrV	7549180	G	A	ppa_stranded_DN18023_c0_g1_i1	intronic_near_splice_sit	ser-5
ChrV	2243076	C	T	PPA23050	intronic_near_splice_sit	rad-8
ChrI	3852296	T	A	PPA29201	nonsense	dhc-3
ChrIII	1287249	A	T	PPA02442	nonsense	glr-2
ChrV	1843872	C	T	PPA10610	nonsense	
ChrI	4742880	C	A	PPA42121	nonsynonym	asd-2
ChrI	6745893	C	T	PPA02066	nonsynonym	
ChrI	7070181	G	A	PPA28207	nonsynonym	
ChrI	8637570	A	G	PPA45065	nonsynonym	
ChrI	9040509	T	C	ppa_stranded_DN30480_c0_g1_i3	nonsynonym	sipa-1
ChrI	9226073	G	A	PPA01881	nonsynonym	C29G2.6
ChrI	1024332	G	A	PPA25009	nonsynonym	CELE_F41G3.10
ChrI	1070290	G	A	PPA10389	nonsynonym	CELE_F40F9.10
ChrI	1138090	G	A	PPA37714	nonsynonym	CELE_T10G3.3
ChrI	1227163	A	T	ppa_stranded_DN12886_c0_g1_i1	nonsynonym	CELE_Y73F8A.35
ChrI	1835444	C	T	PPA18108	nonsynonym	
ChrI	1999454	G	A	Contig0-snapTAU.622	nonsynonym	CELE_T07D3.4
ChrI	2154374	G	A	PPA31671	nonsynonym	lbp-1
ChrI	2296623	G	A	PPA07314	nonsynonym	
ChrI	2464863	A	T	ppa_stranded_DN29041_c0_g1_i1	nonsynonym	CELE_Y4C6B.4
ChrI	2577954	G	T	PPA07095	nonsynonym	pgrn-1
ChrI	2617752	G	A	ppa_stranded_DN27869_c0_g1_i1	nonsynonym	CELE_Y53F4B.9
ChrI	2866459	G	C	ppa_stranded_DN25223_c0_g2_i1	nonsynonym	ivns-1
ChrI	2869864	G	A	ppa_stranded_DN27915_c1_g2_i2	nonsynonym	F46H5.3
ChrI	3092251	T	A	PPA21692	nonsynonym	air-2
ChrI	3115168	C	T	PPA38482	nonsynonym	npp-11
ChrI	3121515	C	T	PPA38476	nonsynonym	CELE_F53B6.4
ChrI	3183154	C	T	ppa_stranded_DN24132_c1_g1_i1	nonsynonym	nas-10
ChrI	3188118	C	T	ppa_stranded_DN11105_c0_g1_i1	nonsynonym	C15C6.2
ChrI	3188119	T	A	ppa_stranded_DN11105_c0_g1_i1	nonsynonym	C15C6.2
ChrI	3192976	C	T	ppa_stranded_DN27249_c0_g1_i1	nonsynonym	tsp-11
ChrI	3228929	C	T	PPA23730	nonsynonym	pgp-10
ChrI	3331629	T	A	ppa_stranded_DN21200_c0_g1_i1	nonsynonym	CELE_W02D9.2
ChrI	3360265	G	A	PPA00025	nonsynonym	CELE_Y57G11C.9
ChrI	3636447	C	T	PPA00476	nonsynonym	nhr-28
ChrI	3685136	C	A	PPA38259	nonsynonym	glh-3
ChrI	3696854	T	A	PPA33913	nonsynonym	
ChrI	3763901	C	A	PPA38349	nonsynonym	zipt-15
ChrI	3831847	G	A	ppa_stranded_DN12571_c0_g1_i1	nonsynonym	
ChrI	3836733	C	T	PPA24496	nonsynonym	
ChrI	3851548	C	T	PPA29201	nonsynonym	dhc-3
ChrI	3852305	C	T	PPA29201	nonsynonym	dhc-3
ChrI	3928562	C	T	ppa_stranded_DN29329_c0_g1_i3	nonsynonym	CELE_F22G12.5
ChrII	1239789	T	C	ppa_stranded_DN30137_c0_g2_i3	nonsynonym	CELE_K07C5.3
ChrII	1798426	C	T	PPA14720	nonsynonym	
ChrII	2116389	T	A	PPA36530	nonsynonym	elo-1
ChrIII	316553	G	C	ppa_stranded_DN28612_c0_g1_i1	nonsynonym	CELE_Y75B8A.7
ChrIII	1510368	C	T	PPA04029	nonsynonym	srsx-35
ChrIII	4294991	C	A	PPA02947	nonsynonym	
ChrIII	5073718	C	T	PPA05259	nonsynonym	mrps-10
ChrIII	7864948	C	T	PPA15345	nonsynonym	
ChrIII	1016633	C	A	ppa_stranded_DN27642_c0_g2_i1	nonsynonym	let-716
ChrIII	1016633	C	A	ppa_stranded_DN27642_c0_g2_i2	nonsynonym	let-716
ChrIII	1032331	A	C	PPA03169	nonsynonym	cdh-4

ChrIII	1055834	C	T	PPA40084	nonsynonym	CELE_T28D6.5
ChrIII	1166376	C	T	PPA25490	nonsynonym	C13B9.2
ChrIII	1236173	C	T	ppa_stranded_DN28219_c0_g1_i1	nonsynonym	doxa-1
ChrIII	1292407	C	T	ppa_stranded_DN16501_c0_g1_i4	nonsynonym	sftb-1
ChrIII	1315230	G	A	PPA36678	nonsynonym	
ChrIII	2026517	G	A	ppa_stranded_DN29773_c2_g1_i3	nonsynonym	R05D3.12
ChrIII	2030171	T	C	ppa_stranded_DN25495_c1_g2_i1	nonsynonym	ztf-18
ChrIV	1825046	C	G	PPA14698	nonsynonym	plc-4
ChrIV	1851637	G	T	ppa_stranded_DN30915_c1_g1_i6	nonsynonym	wdfy-3
ChrIV	2366706	A	G	PPA42887	nonsynonym	
ChrIV	2632444	C	A	PPA42712	nonsynonym	
ChrIV	2720828	T	A	PPA13673	nonsynonym	zfn-207
ChrIV	3056791	C	T	Iso_D.12473.1	nonsynonym	
ChrV	1859556	C	G	PPA33781	nonsynonym	pept-1
ChrV	3697367	T	C	PPA30982	nonsynonym	
ChrV	5149268	A	C	PPA33807	nonsynonym	
ChrV	6089722	G	A	PPA01242	nonsynonym	ced-7
ChrV	6356139	T	C	PPA27650	nonsynonym	CELE_Y92H12BR.7
ChrV	6437626	G	A	PPA43437	nonsynonym	
ChrV	7865942	C	G	PPA03142	nonsynonym	CELE_W02A2.5
ChrV	8238020	C	G	ppa_stranded_DN18776_c0_g1_i1	nonsynonym	
ChrV	9333361	C	T	PPA26749	nonsynonym	ubc-25
ChrV	9693645	T	G	PPA31261	nonsynonym	
ChrV	1078286	G	A	ppa_stranded_DN31053_c0_g1_i2	nonsynonym	ccep-290
ChrV	1227677	A	T	ppa_stranded_DN18733_c0_g1_i1	nonsynonym	lin-41
ChrV	1228818	G	A	PPA09266	nonsynonym	ppfr-1
ChrV	1252591	G	A	PPA39971	nonsynonym	slc-25a42
ChrV	1276287	G	A	PPA39682	nonsynonym	abtm-1
ChrV	1324471	C	T	PPA08122	nonsynonym	ant-1.1
ChrV	1410204	C	T	PPA35687	nonsynonym	
ChrV	1559072	G	C	PPA12037	nonsynonym	kin-32
ChrV	1560884	G	T	PPA12041	nonsynonym	cutl-11
ChrV	1582504	C	T	PPA11165	nonsynonym	bath-36
ChrV	1631799	C	T	ppa_stranded_DN31518_c0_g1_i2	nonsynonym	atm-1
ChrV	1649669	C	T	PPA11055	nonsynonym	noah-1
ChrV	1727803	C	T	PPA35811	nonsynonym	
ChrV	1831972	C	T	PPA35854	nonsynonym	
ChrV	1886102	T	A	ppa_stranded_DN24465_c0_g3_i2	nonsynonym	lim-7
ChrV	1982665	C	T	ppa_stranded_DN31403_c0_g1_i5	nonsynonym	
ChrV	1994990	G	T	PPA35952	nonsynonym	
ChrV	2054492	T	A	ppa_stranded_DN29783_c1_g1_i3	nonsynonym	let-607
ChrV	2054492	T	A	ppa_stranded_DN29783_c1_g1_i4	nonsynonym	let-607
ChrV	2209488	T	A	ppa_stranded_DN24948_c0_g1_i1	nonsynonym	rad-54.l
ChrV	2327107	C	A	PPA34021	nonsynonym	
ChrX	18682	A	G	PPA09538	nonsynonym	CELE_K06A9.1
ChrX	20645	A	G	PPA09538	nonsynonym	CELE_K06A9.1

Supplementary Table 4. Description on the 15 fixed (i.e., non-sliding) landmarks used to quantify morphological differences in *Pristionchus pacificus* and *Caenorhabditis elegans* strains.

landmark	<i>Pristionchus pacificus</i>	<i>Caenorhabditis elegans</i>
1	anterior tip of the dorsal cheilostom	anterior tip of the dorsal cheilostom
2	posterior tip of the dorsal cheilostom	posterior tip of the dorsal cheilostom
3	anterior tip of the dorsal gymnostom	anterior tip of the dorsal gymnostom
4	posterior tip of the dorsal gymnostom	posterior tip of the dorsal gymnostom
5	posterior tip of the dorsal promesostegostom	posterior tip of the dorsal promesostegostom
6	posterior base point of the dorsal tooth (metastegostom)	posterior base point of the dorsal flap (metastegostom)
7	anterior tip of the dorsal tooth (metastegostom)	anterior tip of the dorsal flap (metastegostom)
8	lateral base point of the right ventrosublateral tooth (metastegostom)	lateral base point of the right ventrosublateral flap (metastegostom)
9	anterior tip of the right ventrosublateral tooth (metastegostom)	anterior tip of the right ventrosublateral flap (metastegostom)
10	ventral base point of the right ventrosublateral tooth (metastegostom)	ventral base point of the right ventrosublateral flap (metastegostom)
11	posterior tip of the ventral promesostegostom	posterior tip of the ventral promesostegostom
12	posterior tip of the ventral gymnostom	posterior tip of the ventral gymnostom
13	anterior tip of the ventral gymnostom	anterior tip of the ventral gymnostom
14	posterior tip of the ventral cheilostom	posterior tip of the ventral cheilostom
15	anterior tip of the ventral cheilostom	anterior tip of the ventral cheilostom

Supplementary Table 5. CRISPR/Cas9 sgRNA target sequences and primer information used to generate all *Ppa-dpy-6* (PPA38488) mutants including specific targeted modifications

Strain	Modification	Type of primer	Sequences
PS312	<i>Ppa-dpy-6</i> knock out from the third exon	CRISPR sgRNA	ATCGTTGTCGTCCTTCAGA
		Forward Primer	CATTCCAGATTTCACCATCGTCAAGAGGATAAAG
		Reverse Primer	GGCATCTCTGTGCCTTCAACTCATCCAAT
		Sequencing Primer	CATTCCAGATTTCACCATCGTCAAGAGGATAAAG
PS312	<i>Ppa-dpy-6</i> knock out from the 10th exon	CRISPR sgRNA	GGGTCATGGGAATGTCCGAA
		Forward Primer	TCAGACTGAGTGGCCATCGGCA
		Reverse Primer	GATGAATAAGTCACTGACGGGTCC
		Sequencing Primer	TCAGACTGAGTGGCCATCGGCA
PS312	Create the substitution of C > T in the intronic splice site of <i>Ppa-dpy-6</i> gene	CRISPR sgRNA	GGGTCATGGGAATGTCCGAA
		Forward Primer	TCAGACTGAGTGGCCATCGGCA
		Reverse Primer	GATGAATAAGTCACTGACGGGTCC
		Sequencing Primer	TCAGACTGAGTGGCCATCGGCA
		Repair Templates	CTTTCTTCATCCTCTTCAGCTTGCTTGTCCATGTCGAATTGAGGAGGAAGGGT CATGGGAATGTCCGAAGGGCATGAATGATCTTTGATCTTTCTGAAAGGTCTCT CTTCCCTTACCCAC
tu743	Repair the substitution of T > C in the intronic splice site of <i>Ppa-dpy-6</i> gene	CRISPR sgRNA	GGGTCATGGGAATGTCCGAA
		Forward Primer	TCAGACTGAGTGGCCATCGGCA
		Reverse Primer	GATGAATAAGTCACTGACGGGTCC
		Sequencing Primer	TCAGACTGAGTGGCCATCGGCA
		Repair Templates	CTTTCTTCATCCTCTTCAGCTTGCTTGTCCATGTCGAATTGAGGAGGAAGGGT CATGGGAATGTCCGAAGGGCCTGAATGATCTTTGATCTTTCTGAAAGGTCTCT CTTCCCTTACCCAC

Supplementary Table 6. Description of transgenic constructs

Strain ID (RS)	Strain ID (tu)	Gene ID	Background strain	Promoter	CDS	3'UTR
RS3892	tuEx340	<i>Ppa-dpy-6</i> , PPA38488	PS312	2,189 bp upstream sequences of the first ATG of PPA38488	GFP sequence (Han et al. 2020)	<i>rpl-23</i>
RS3893	tuEx341	<i>Ppa-dpy-6</i> , PPA38488	PS312	2,189 bp upstream sequences of the first ATG of PPA38488	GFP sequence (Han et al. 2020)	<i>rpl-23</i>
RS3894	tuEx342	<i>Cel-dpy-6</i> , F16F9.2	N2	2,275 bp upstream sequences of the first ATG of F16F9.2	GFP sequence	<i>rpl-23</i>
RS3895	tuEx343	<i>Cel-dpy-6</i> , F16F9.2	N2	2,275 bp upstream sequences of the first ATG of F16F9.2	GFP sequence	<i>rpl-23</i>
RS3896	tuEx344	<i>Cel-dpy-6</i> , F16F9.2	N2	2,275 bp upstream sequences of the first ATG of F16F9.2	GFP sequence	<i>rpl-23</i>

Supplementary Table 7. FPKM values used in the analysis of differential expression gene

Comparison the gene expression of <i>Ppa-dpy-6</i> in between <i>Ppa-eud-1</i> mutant and WT				
Strain name	Strain description	Replicate_ID	Developmental stage	FPKM
<i>tu1069</i>	<i>Ppa-eud-1</i> loss of function strain	Replicate_1	20 hours after hatching	39.7
<i>tu1069</i>	<i>Ppa-eud-1</i> loss of function strain	Replicate_2	21 hours after hatching	54.3
<i>tu1069</i>	<i>Ppa-eud-1</i> loss of function strain	Replicate_3	22 hours after hatching	82.9
PS312	Wild-type strain	Replicate_1	23 hours after hatching	0.0
PS312	Wild-type strain	Replicate_2	24 hours after hatching	0.0
PS312	Wild-type strain	Replicate_3	25 hours after hatching	0.2
Comparison the gene expression of <i>Ppa-dpy-6</i> in between <i>Ppa-nhr-40</i> mutants and WT				
Strain name	Strain description	Replicate_ID	Developmental stage	FPKM
<i>tu505</i>	<i>Ppa-nhr-40</i> gain of function strain	Replicate_1	Mouth form determination	70.2
<i>tu505</i>	<i>Ppa-nhr-40</i> gain of function strain	Replicate_2	Mouth form determination	50.3
<i>iub6</i>	<i>Ppa-nhr-40</i> gain of function strain	Replicate_3	Mouth form determination	47.3
<i>iub6</i>	<i>Ppa-nhr-40</i> gain of function strain	Replicate_4	Mouth form determination	62.5
<i>tu1418</i>	<i>Ppa-nhr-40</i> loss of function strain	Replicate_1	Mouth form determination	49.2
<i>tu1418</i>	<i>Ppa-nhr-40</i> loss of function strain	Replicate_2	Mouth form determination	47.2
<i>tu1423</i>	<i>Ppa-nhr-40</i> loss of function strain	Replicate_3	Mouth form determination	53.5
<i>tu1423</i>	<i>Ppa-nhr-40</i> loss of function strain	Replicate_4	Mouth form determination	60.7
PS312	Wild-type strain	Replicate_1	Mouth form determination	54.3
PS312	Wild-type strain	Replicate_2	Mouth form determination	49.1
PS312	Wild-type strain	Replicate_3	Mouth form determination	46.2
PS312	Wild-type strain	Replicate_4	Mouth form determination	56.5
Comparison the gene expression of <i>Ppa-dpy-6</i> in between <i>Ppa-nhr-1</i> mutant and WT				
Strain name	Strain description	Replicate_ID	Developmental stage	FPKM
<i>tu1163</i>	<i>Ppa-nhr-1</i> loss of function strain	Replicate_1	Mouth form determination	42.7
<i>tu1163</i>	<i>Ppa-nhr-1</i> loss of function strain	Replicate_2	Mouth form determination	39.2
<i>tu1164</i>	<i>Ppa-nhr-1</i> loss of function strain	Replicate_3	Mouth form determination	49.1
<i>tu1164</i>	<i>Ppa-nhr-1</i> loss of function strain	Replicate_4	Mouth form determination	48.3
PS312	Wild-type strain	Replicate_1	Mouth form determination	53.8
PS312	Wild-type strain	Replicate_2	Mouth form determination	49.2
PS312	Wild-type strain	Replicate_3	Mouth form determination	42.8

Mouth form determination: Combined stage of 24 h and 48 h post-bleaching

Supplementary Table 8. PERMANOVA–RRPP summary statistics for **replicate factor of *P. pacificus* mouth shape replicability** (n=10, replicates=3). Linear model used: *shape* ~ *replicate*. Iterations used: 1,001. Seed used: 12345. Effect type: F, Z, relative effect sizes; P, probability of finding a random value larger than the observed F value (P-value).

source	df	SS	MS	R ²	F	Z	P (>F)
replicate	2	0.002887	0.0014435	0.02475	0.3427	-2.4493	0.993
residuals	27	0.113739	0.0042126	0.97525			
total	29	0.116626					

Supplementary Table 9. PERMANOVA–RRPP summary statistics for **specimen factor of *P. pacificus* mouth shape replicability** (n=10, replicates=3). Linear model used: *shape* ~ *specimen*. Iterations used: 1,001. Seed used: 12345. Effect type: F, Z, relative effect sizes; P, probability of finding a random value larger than the observed F value (P-value).

source	df	SS	MS	R ²	F	Z	P (>F)
specimen	9	0.102674	0.0114083	0.88037	16.354	9.2699	0.000999
residuals	20	0.013952	0.0006976	0.11963			
total	29	0.116626					

Supplementary Table 10. PERMANOVA–RRPP summary statistics for **replicate factor of *C. elegans* mouth shape replicability** (n=10, replicates=3). Linear model used: *shape* ~ *replicate*. Iterations used: 1,001. Seed used: 12345. Effect type: F, Z, relative effect sizes; P, probability of finding a random value larger than the observed F value (P-value).

source	df	SS	MS	R ²	F	Z	P (>F)
replicate	2	0.000646	0.00032314	0.01452	0.1989	-3.0027	1.0
residuals	27	0.043859	0.00162443	0.98548			
total	29	0.044506					

Supplementary Table 11. PERMANOVA–RRPP summary statistics for **specimen factor of *C. elegans* mouth shape replicability** (n=10, replicates=3). Linear model used: *shape* ~ *specimen*. Iterations used: 1,001. Seed used: 12345. Effect type: F, Z, relative effect sizes; P, probability of finding a random value larger than the observed F value (P-value).

source	df	SS	MS	R ²	F	Z	P (>F)
specimen	9	0.039784	0.0044205	0.89392	18.726	8.8816	0.000999
residuals	20	0.004721	0.0002361	0.10608			
total	29	0.044506					

Supplementary Table 12. PERMANOVA–RRPP summary statistics for ***P. pacificus* mouth shapes** (n=118). Linear model used: *shape* ~ *strain*. Iterations used: 10,000. Seed used: 12345. Effect type: F, Z, relative effect sizes; P, probability of finding a random value larger than the observed F value (P-value).

source	df	SS	MS	R ²	F	Z	P (>F)
strain	2	1.01955	0.50977	0.65646	109.87	5.9632	0.0001
residuals	115	0.53355	0.00464	0.34354			
total	117	1.55310					

Supplementary Table 13. PERMANOVA–RRPP summary statistics for *C. elegans* mouth shapes (n=75). Linear model used: *shape* ~ *strain*. Iterations used: 10,000. Seed used: 12345. Effect type: F.

source	df	SS	MS	R ²	F	Z	P (>F)
strain	1	0.059076	0.059076	0.32253	34.754	5.4348	0.0001
residuals	73	0.124089	0.001700	0.67747			
total	74	0.183165					

Supplementary Table 14. PERMANOVA–RRPP summary statistics for *P. pacificus* mouth forms (n=118). Linear model used: *form* ~ *strain*. Iterations used: 10,000. Seed used: 12345. Effect type: F.

source	df	SS	MS	R ²	F	Z	P (>F)
strain	2	1.10374	0.55187	0.58068	79.626	6.7103	0.0001
residuals	115	0.79704	0.00693	0.41932			
total	117	1.90078					

Supplementary Table 15. PERMANOVA–RRPP summary statistics for *C. elegans* mouth forms (n=75). Linear model used: *form* ~ *strain*. Iterations used: 10,000. Seed used: 12345. Effect type: F.

source	df	SS	MS	R ²	F	Z	P (>F)
strain	1	0.21504	0.215038	0.46015	62.222	4.2637	0.0001
residuals	73	0.25229	0.003456	0.53985			
total	74	0.46732					

Supplementary Table 16. PERMANOVA–RRPP summary statistics for *P. pacificus* mouth sizes (n=118). Linear model used: *centroid size* ~ *strain*. Iterations used: 10,000. Seed used: 12345. Effect type: F.

source	df	SS	MS	R ²	F	Z	P (>F)
strain	2	25.122	12.5609	0.23849	18.008	4.5533	0.0001
residuals	115	80.214	0.6975	0.76151			
total	117	105.335					

Supplementary Table 17. PERMANOVA–RRPP summary statistics for *C. elegans* mouth sizes (n=75). Linear model used: *centroid size* ~ *strain*. Iterations used: 10,000. Seed used: 12345. Effect type: F.

source	df	SS	MS	R ²	F	Z	P (>F)
strain	1	82.069	82.069	0.54601	87.798	5.6049	0.0001
residuals	73	68.237	0.935	0.45399			
total	74	150.305					

Supplementary Table 18. Summary statistics of pairwise comparisons among *P. pacificus* mouth shapes (n=118). Pairwise comparisons of LS means were performed with the *pairwise* function of RRPP. Abbreviations: d, distance between the means; UCL, upper confidence limit; Z, relative effect size (Z-score); P, probability of finding a random value larger than the observed distance value (P-value).

pairwise comparisons	d	UCL (95%)	Z	P > d
wildtype vs. rescue	0.03609841	0.04582869	1.217962	0.1322
wildtype vs. <i>Ppa-dpy-6</i>	0.18483771	0.04642146	3.593817	0.0001
<i>Ppa-dpy-6</i> vs. rescue	0.20883560	0.04610816	3.642244	0.0001

Supplementary Table 19. Summary statistics of pairwise comparisons between *C. elegans* mouth shapes (n=75). Abbreviations: d, distance between the means; UCL, upper confidence limit; Z, relative effect size (Z-score); P, probability of finding a random value larger than the observed distance value (P-value).

pairwise comparisons	d	UCL (95%)	Z	P > d
wildtype vs. <i>Cel-dpy-6</i>	0.05625639	0.01726215	5.129054	0.0001

Supplementary Table 20. Summary statistics of pairwise comparisons among *P. pacificus* mouth forms (n=118). Pairwise comparisons of LS means were performed with the *pairwise* function of RRPP. Abbreviations: d, distance between the means; UCL, upper confidence limit; Z, relative effect size (Z-score); P, probability of finding a random value larger than the observed distance value (P-value).

pairwise comparisons	d	UCL (95%)	Z	P > d
wildtype vs. rescue	0.0395137	0.04839635	1.217225	0.1322
wildtype vs. <i>Ppa-dpy-6</i>	0.1908064	0.04897315	4.329126	0.0001
<i>Ppa-dpy-6</i> vs. rescue	0.2182532	0.04899521	4.466567	0.0001

Supplementary Table 21. Summary statistics of pairwise comparisons between *C. elegans* mouth forms (n=75). Pairwise comparisons of LS means were performed with the *pairwise* function of RRPP. Abbreviations: d, distance between the means; UCL, upper confidence limit; Z, relative effect size (Z-score); P, probability of finding a random value larger than the observed distance value (P-value).

pairwise comparisons	d	UCL (95%)	Z	P > d
wildtype vs. <i>Cel-dpy-6</i>	0.1073307	0.03117189	3.926253	0.0001

Supplementary Table 22. Summary statistics of pairwise comparisons among *P. pacificus* mouth sizes (n=118). Pairwise comparisons of LS means were performed with the *pairwise* function of RRPP. Abbreviations: d, distance between the means; UCL, upper confidence limit; Z, relative effect size (Z-score); P, probability of finding a random value larger than the observed distance value (P-value).

pairwise comparisons	d	UCL (95%)	Z	P > d
wildtype vs. rescue	0.2785290	0.4114308	0.9472539	0.1785
wildtype vs. <i>Ppa-dpy-6</i>	0.8172373	0.4218938	2.9842151	0.0002
<i>Ppa-dpy-6</i> vs. rescue	1.0957663	0.4193543	3.7446847	0.0001

Supplementary Table 23. Summary statistics of pairwise comparisons between *C. elegans* mouth sizes (n=75). Pairwise comparisons of LS means were performed with the *pairwise* function of RRPP. Abbreviations: d, distance between the means; UCL, upper confidence limit; Z, relative effect size (Z-score); P, probability of finding a random value larger than the observed distance value (P-value).

pairwise comparisons	d	UCL (95%)	Z	P > d
wildtype vs. <i>Cel-dpy-6</i>	2.096794	0.6345145	4.489541	0.0001

Supplementary Table 24. ANOVA summary statistics for a comparison of a simple allometry model ($shape \sim \log(size)$) and a common allometry model ($shape \sim \log(size) + strain$) for the *P. pacificus* data set (n=118). Iterations used: 1,000. Effect type: F. RRPP = T. Size refers to the centroid sizes contained in the shape data. *procd.lm* was used to fit the data, *anova* was used for the model comparison.

model	res.df	df	rSS	SS	MS	R ²	F	Z	P
shape ~ log(size) (Null)	116	1	1.36576			0.00000			
shape ~ log(size) + strain	114	2	0.51853	0.84723	0.42362	0.54551	93.134	5.1362	0.001
total	117		1.55310						

Supplementary Table 25. ANOVA summary statistics for a comparison of a common allometry model ($shape \sim \log(size) + strain$) and a group-specific allometry model ($shape \sim \log(size) * strain$) for the *P. pacificus* data set (n=118). Iterations used: 1,000. Effect type: F. RRPP = T. Size refers to the centroid sizes contained in the shape data. *procd.lm* was used to fit the data, *anova* was used for the model comparison.

model	res. df	df	rSS	SS	MS	R ²	F	Z	P
shape ~ log(size) + strain (Null)	114	1	0.51853			0.000000			
shape ~ log(size) * strain	112	2	0.50132	0.017212	0.0086059	0.011082	1.9227	1.7334	0.044
total	117		1.55310						

Supplementary Table 26. PERMANOVA–RRPP summary statistics for *P. pacificus* mouth allometry (n=118). Linear model used: $shape \sim \log(centroid\ size) + strain$. Iterations used: 10,000. Seed used: 12345. Effect type: F. c.size = centroid size. *procd.lm* was used to fit the data.

source	df	SS	MS	R ²	F	Z	P (>F)
log(c.size)	1	0.18476	0.18476	0.11896	40.592	3.310	0.0001
strain	2	0.84944	0.42472	0.54693	93.311	5.317	0.0001
residuals	114	0.51889	0.00455	0.33410			
total	117	1.55310					

Supplementary Table 27. Summary statistics of pairwise comparisons among *P. pacificus* mouth allometries (n=118). Pairwise comparisons of LS means were performed with the *pairwise* function of RRPP. Abbreviations: d, distance between the LS means; UCL, upper confidence limit; Z, relative effect size (Z-score); P, probability of finding a random value larger than the observed distance value (P-value).

pairwise comparisons	d	UCL (95%)	Z	P > d
wildtype vs. rescue	0.03860415	0.04338278	1.450365	0.0828
wildtype vs. <i>Ppa-dpy-6</i>	0.19129182	0.04631384	3.556634	0.0001
<i>Ppa-dpy-6</i> vs. rescue	0.21818263	0.04884414	3.625766	0.0001

Supplementary Table 28. ANOVA summary statistics for a comparison of a simple allometry model ($shape \sim \log(size)$) and a common allometry model ($shape \sim \log(size) + strain$) for the *C. elegans* data set (n=75). Iterations used: 1,000. Effect type: F. RRPP = T. Size refers to the centroid sizes contained in the shape data. *procD.lm* was used to fit the data, *anova* was used for the model comparison.

model	res.df	df	rSS	SS	MS	R ²	F	Z	P
shape ~ log(size) (Null)	73	1	0.143393			0.00000			
shape ~ log(size) + strain	72	1	0.095601	0.04779	0.04779	0.26092	35.993	5.9397	0.001
total	74		0.183165						

Supplementary Table 29. ANOVA summary statistics for a comparison of a common allometry model ($shape \sim \log(size) + strain$) and a group-specific allometry model ($shape \sim \log(size) * strain$) for the *C. elegans* data set (n=75). Iterations used: 1,000. Effect type: F. RRPP = T. Size refers to the centroid sizes contained in the shape data. *procD.lm* was used to fit the data, *anova* was used for the model comparison.

model	res.df	df	rSS	SS	MS	R ²	F	Z	P
shape ~ log(size) + strain (Null)	72	1	0.0956			0.00000			
shape ~ log(size) * strain	71	1	0.0948	0.000765	0.000765	0.00417	0.573	-1.164	0.87
total	74		0.1831						

Supplementary Table 30. PERMANOVA–RRPP summary statistics for *C. elegans* mouth allometry (n=75). Linear model used: $shape \sim \log(centroid\ size) + strain$. Iterations used: 10,000. Seed used: 12345. Effect type: F. c.size = centroid size. *procD.lm* was used to fit the data.

source	df	SS	MS	R ²	F	Z	P (>F)
log(c.size)	1	0.039762	0.039762	0.21709	30.024	5.3888	0.0001
strain	1	0.048048	0.048048	0.26232	36.280	5.1584	0.0001
residuals	72	0.095355	0.001324	0.52059			
total	74	0.183165					

Supplementary Table 31. Summary statistics of pairwise comparisons among *C. elegans* mouth allometries (n=75). Pairwise comparisons of LS means were performed with the *pairwise* function of RRPP. Abbreviations: d, distance between the LS means; UCL, upper confidence limit; Z, relative effect size (Z-score); P, probability of finding a random value larger than the observed distance value (P-value).

pairwise comparisons	d	UCL (95%)	Z	P > d
wildtype vs. <i>Cel-dpy-6</i>	0.07529769	0.02232566	4.92008	0.0001

RESEARCH ARTICLE

Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell

Bogdan Sieriebriennikov [‡], Shuai Sun, James W. Lightfoot , Hanh Witte, Eduardo Moreno, Christian Rödelsperger , Ralf J. Sommer ^{*}

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

[‡] Current address: Department of Biology, New York University, New York, NY, United States of America

^{*} ralf.sommer@tuebingen.mpg.de



Abstract

Environment shapes development through a phenomenon called developmental plasticity. Deciphering its genetic basis has potential to shed light on the origin of novel traits and adaptation to environmental change. However, molecular studies are scarce, and little is known about molecular mechanisms associated with plasticity. We investigated the gene regulatory network controlling predatory vs. non-predatory dimorphism in the nematode *Pristionchus pacificus* and found that it consists of genes of extremely different age classes. We isolated mutants in the conserved nuclear hormone receptor *nhr-1* with previously unseen phenotypic effects. They disrupt mouth-form determination and result in animals combining features of both wild-type morphs. In contrast, mutants in another conserved nuclear hormone receptor *nhr-40* display altered morph ratios, but no intermediate morphology. Despite divergent modes of control, NHR-1 and NHR-40 share transcriptional targets, which encode extracellular proteins that have no orthologs in *Caenorhabditis elegans* and result from lineage-specific expansions. An array of transcriptional reporters revealed co-expression of all tested targets in the same pharyngeal gland cell. Major morphological changes in this gland cell accompanied the evolution of teeth and predation, linking rapid gene turnover with morphological innovations. Thus, the origin of feeding plasticity involved novelty at the level of genes, cells and behavior.

OPEN ACCESS

Citation: Sieriebriennikov B, Sun S, Lightfoot JW, Witte H, Moreno E, Rödelsperger C, et al. (2020) Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell. *PLoS Genet* 16(4): e1008687. <https://doi.org/10.1371/journal.pgen.1008687>

Editor: Artyom Kopp, University of California Davis, UNITED STATES

Received: December 11, 2019

Accepted: February 20, 2020

Published: April 13, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1008687>

Copyright: © 2020 Sieriebriennikov et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Raw reads from the RNA-seq experiment have been deposited at the European Nucleotide Archive, study accession

Author summary

Rather than following a pre-determined genetic “blueprint”, organisms can adjust their development when they perceive relevant environmental signals—a phenomenon called plasticity. This improves performance in changing environment and may also affect how species evolve. To learn how plasticity works on the mechanistic genetic level, we investigated the roundworm *Pristionchus pacificus*. It may develop either as a toothed predator or as a narrow-mouthed microbe-eater depending on food source and population density, an ability that evolved less than 100 million years ago. Previous studies identified switch

number PRJEB34615 (<http://www.ebi.ac.uk/ena/data/view/PRJEB34615>). Nucleotide sequences of the fragments used to create transgenic constructs are provided in [S1 Data](#). Phylogenetic trees and alignments used to generate them are provided in [S2 Data](#). Data used to generate the boxplot in [Fig 2B](#) and the stacked barplot in [Fig 3B](#) are provided in [S3](#) and [S4 Data](#), respectively. Landmark coordinates used to conduct geometric morphometric analysis are provided in [S5 Data](#).

Funding: This work was supported by institutional funds of the Max-Planck Society (to R.J.S.) and the China Scholarship Council (to S.S.). Both funders, the Max-Planck Society and the China Scholarship Council did not play any role in the study design, data collection and analysis, decision to publish, or the preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

genes, whose inactivation or overactivation forces either predatory or non-predatory development. Here, we identified the first core gene, which is required for the specification of both morphologies. It encodes a transcription factor, whose inactivation creates animals that appear intermediate between predators and non-predators. We queried which genes are simultaneously controlled by this previously unknown regulator and by a closely related protein that acts as a classical switch. All of the co-regulated genes were recently born and are acting in a single cell that was strongly modified when predator vs. non-predator plasticity evolved. We suggest that conserved regulators of different classes enlisted novel genes in a refurbished cell to regulate a novel plastic trait.

Introduction

Developmental plasticity is the ability to generate different phenotypes in response to environmental input [1]. As a result, even genetically identical individuals may develop distinct phenotypes, the most extreme example being castes in social insects [2]. Developmental plasticity is attracting considerable attention in the context of adaptation to climate change [3–6] and as a facilitator of evolutionary novelty [7–11]. However, the role of plasticity in evolution has been contentious [6,12] because the genetic and epigenetic underpinnings of plastic traits have long remained elusive. Nonetheless, recent studies have begun to elucidate associated molecular mechanisms in insects and nematodes [13–16]. Ultimately, the identification of gene regulatory networks (GRN) controlling plasticity will provide an understanding of development in novel environments and enable the testing of theories about the long-term evolutionary significance of plasticity.

The free-living nematode *Pristionchus pacificus* has recently been established as a model to study plasticity [13]. These worms can develop two alternative mouth forms, called eurystomatous (Eu) and stenostomatous (St) mouth forms, respectively. Eu morphs have a wide buccal cavity and two large opposed teeth enabling predation on other nematodes, while St morphs have a narrow buccal cavity and one tooth limiting their diet to microbial sources [17,18] ([Fig 1A–1C](#), [S1A Fig](#)). The wild-type *P. pacificus* strain PS312 preferentially forms Eu morphs in standard culture conditions on agar plates, but becomes predominantly St in liquid culture [19]. Additionally, nematode-derived modular metabolites excreted by adult animals induce the predatory Eu morph [20,21]. A forward genetic screen identified the sulfatase gene *eud-1* as a developmental switch confirming long-standing predictions that plastic traits are regulated by binary switches [18]. Subsequent studies implicated several other enzyme-encoding genes, such as *nag-1*, *nag-2*, and *sult-1/seud-1* in regulating mouth-form plasticity [22–25]. Additionally, the chromatin modifier genes *lxy-12* and *mbd-2* influence *eud-1* expression [26]. In contrast, only one transcription factor, the nuclear hormone receptor (NHR) NHR-40, was so far found to regulate mouth-form fate [27], and no downstream targets have been identified ([Fig 1D](#)).

Here, we leveraged the power of suppressor screen genetics to identify the conserved nuclear hormone receptor NHR-1 as a second transcription factor controlling mouth-form development. It differs from *nhr-40* and all the other genes identified to date in that *nhr-1* mutants develop a morphology that combines features of the two morphs, consistent with disrupted mouth-form determination. Furthermore, transcriptomic profiling revealed that NHR-40 and NHR-1 share transcriptional targets, which exhibit functional redundancy and are expressed in a single pharyngeal gland cell, g1D. This cell has undergone extreme morphological remodeling in nematode evolution, which is associated with the emergence of teeth and

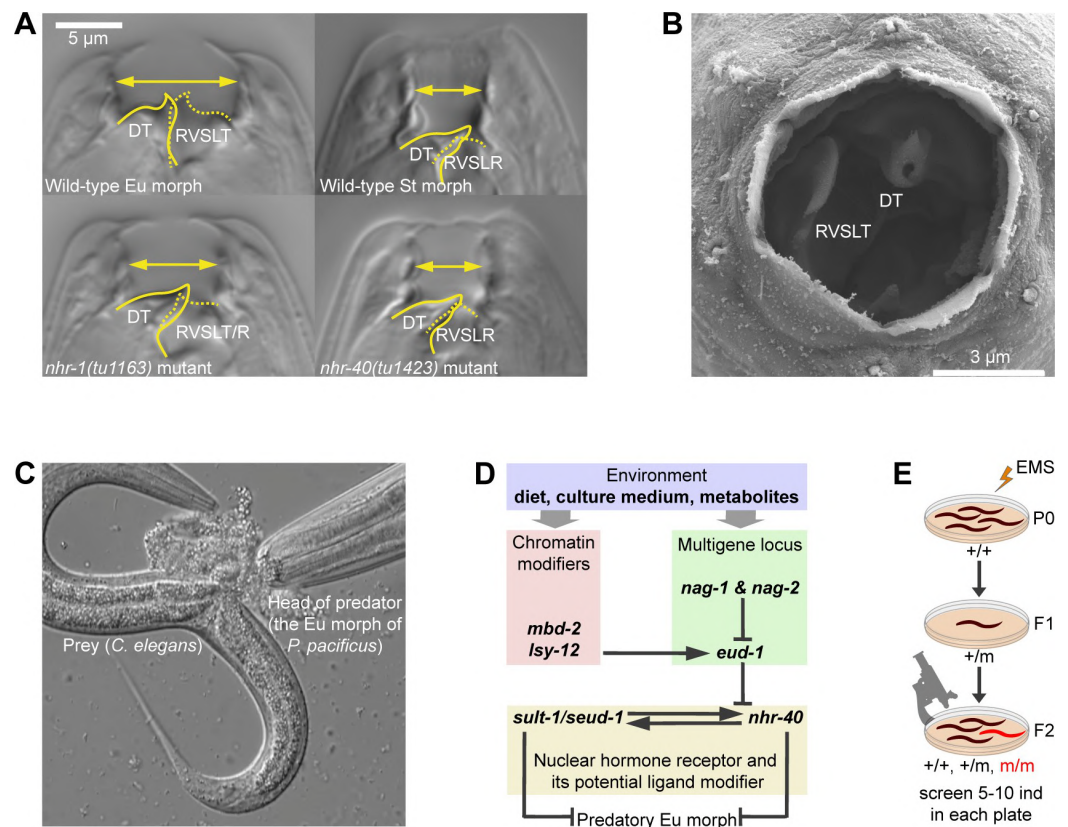


Fig 1. Mouth-form plasticity in *P. pacificus*. (A) Mouth structure of wild-type eurytomatous (Eu) morph, wild-type stenotomatous (St) morph, *nhr-1* mutant, and *nhr-40* mutant. Unlabeled images in two focal planes are shown in S1A Fig. (B) Scanning electron microscopy image of the mouth opening of the Eu morph. (C) The Eu morph devouring its prey. (D) Putative gene regulatory network controlling mouth-form plasticity in *P. pacificus*. (E) Design of the suppressor screen. DT = dorsal tooth, RVSLT = right ventrosublateral tooth, RVSLR = right ventrosublateral ridge, EMS = ethyl methanesulfonate.

<https://doi.org/10.1371/journal.pgen.1008687.g001>

predatory feeding. Interestingly, *nhr-1* and *nhr-40* are well conserved, whereas all target genes are rapidly evolving and have no orthologs in *C. elegans*. This study enhances the understanding of the GRN regulating mouth-form plasticity, elucidates the evolutionary dynamics of underlying genes and links morphological innovations with rapid gene evolution.

Results

Suppressor screen in *nhr-40* identifies another NHR gene regulating mouth-form development

While our previous studies have identified various components involved in the regulation of mouth form plasticity, most of these genes are expressed in neurons responsible for environmental sensing and we had yet to find factors acting in the tissues forming the mouth structure. Therefore, we looked for more downstream factors by conducting a suppressor screen in the mutant background of *nhr-40*. This is the most downstream gene in the current GRN controlling *P. pacificus* mouth-form plasticity and it encodes a transcription factor [27]. We mutagenized *nhr-40(tu505)* worms, which are all-Eu, and isolated one allele, *tu515*, that had a no-Eu phenotype (Fig 1E, Table 1).

Table 1. Mouth-form frequencies in wild type and mutant lines.

Medium	Genotype	Eu, %	N
NGM agar	wild type PS312	98	650
NGM agar	<i>nhr-40(tu505)</i>	100	100
NGM agar	<i>nhr-40(tu505) tu515</i>	0	136
NGM agar	<i>tu515</i>	0	136
NGM agar	<i>nhr-1(tu1163)</i>	0	133
NGM agar	<i>nhr-1(tu1164)</i>	0	140
NGM agar	<i>nhr-1(tu1163)/tu515</i>	0	70
NGM agar	<i>nhr-1(tu1163);tuEx305[nhr-1(+);egl-20p::TurboRFP]</i>	85	110
NGM agar	<i>nhr-1(tu1163);tuEx310[nhr-1(+);egl-20p::TurboRFP]</i>	86	112
NGM agar	<i>nhr-1(tu1163);tuEx328[nhr-1(+);HA;egl-20p::TurboRFP]</i>	86	150
NGM agar	<i>nhr-40(tu505) nhr-1(tu1163)</i>	2	134
NGM agar	<i>nhr-40(tu1418)</i>	0	150
NGM agar	<i>nhr-40(tu1419)</i>	0	150
NGM agar	<i>nhr-40(tu1420)</i>	0	150
NGM agar	<i>nhr-40(tu1423)</i>	0	150
NGM agar	<i>nhr-40(iub6)</i>	100	100
NGM agar	<i>nhr-40(tu1421)</i>	100	150
NGM agar	<i>nhr-40(tu1422)</i>	100	100
NGM agar	duodecuple Astacin mutant ^a	98	55
NGM agar	quintuple CAP mutant ^b	94	50
NGM agar	<i>PPA04200(tu1213) PPA39293(tu1214)</i>	100	50
NGM agar	<i>PPA04200(tu1216) PPA39293(tu1217)</i>	100	50
NGM agar	<i>PPA27560(tu1475)</i>	100	51
NGM agar	<i>PPA27560(tu1476)</i>	100	53
NGM agar	<i>PPA30108(tu1230)</i>	100	50
NGM agar	<i>PPA30108(tu1231)</i>	100	50
NGM agar	<i>PPA30435(tu1477)</i>	100	48
NGM agar	<i>PPA30435(tu1478)</i>	98	54
NGM agar	<i>PPA38892(tu1473)</i>	100	50
NGM agar	<i>PPA38892(tu1474)</i>	100	50
S-medium	wild type PS312	5	850
S-medium	<i>nhr-40(tu505)</i>	100	150
S-medium	<i>nhr-40(tu1418)</i>	0	150
S-medium	<i>nhr-40(tu1419)</i>	0	150
S-medium	<i>nhr-40(tu1420)</i>	0	150
S-medium	<i>nhr-40(tu1423)</i>	0	150
S-medium	<i>nhr-40(iub6)</i>	100	150
S-medium	<i>nhr-40(tu1421)</i>	100	150
S-medium	<i>nhr-40(tu1422)</i>	100	150

N = total number of animals examined

^aThe genotype of the duodecuple Astacin mutant is *PPA03932(tu1259) PPA32730(tu1503);PPA05669(tu1316) PPA05618(tu1317) PPA21987(tu1329) PPA16331(tu1339) PPA27985(tu1340) PPA34430(tu1341) PPA20266(tu1385) PPA42924(tu1386);PPA05955(tu1481) PPA42525(tu1482)*.

^bThe genotype of the quintuple CAP mutant is *tuDf6[PPA21912 PPA29522 PPA21910] tuDf7[PPA05611 PPA39470] tuDf8[PPA13058 PPA39735]*.

<https://doi.org/10.1371/journal.pgen.1008687.t001>

The phenotype was fully penetrant, both in the presence of *nhr-40(tu505)* and after outcrossing, *i.e.* Eu animals were never observed under any culture condition. Thus, *tu515* represents a novel factor influencing the mouth-form ratio. Interestingly, however, *tu515* mutants also exhibited a non-canonical mouth morphology (Fig 1A, S1A and S3 Figs). In contrast to all previously isolated mutants, which either display altered mouth-form frequencies or an aberrant morphology, *tu515* individuals develop a morphology that combines normal features of the two morphs with no apparent dimorphism. Specifically, *tu515* mutants closely resemble the St morph in that they have a flattened dorsal tooth, lack a fully developed right ventrosublateral tooth, and the anterior tip of the promesostegostom aligns with the anterior tip of the gymnostom plate. However, the width of the mouth and the curvature of the dorsal tooth appear intermediate between Eu and St, and the right ventrosublateral ridge is frequently enlarged and resembles an underdeveloped tooth of the Eu morph (Fig 1A, S1A Fig). Therefore, while other known mutants affect mouth-form determination by changing the preferred developmental trajectory, *tu515* is the first mutant that disrupts determination, resulting in non-canonical morphology that resembles the St morph but combines features of both morphs.

To map *tu515*, we performed bulked segregant analysis. We examined the list of non-synonymous and nonsense mutations within the candidate region on the X chromosome (S2B Fig, S1 Table) and discovered a non-synonymous mutation in another NHR-encoding gene, *nhr-1*. The substitution changed the sequence of a highly conserved FFRR motif within the DNA recognition helix [28] to FFRW, which may cause the loss of DNA-binding activity. We performed the following experiments to verify that *nhr-1* is the suppressor of *nhr-40(tu505)*. First, we created *nhr-1* mutants using CRISPR/Cas9 by generating frameshift mutations at the beginning of the ligand-binding domain (LBD). The resulting alleles *tu1163* and *tu1164* exhibited a no-Eu phenotype and the same morphological abnormalities as *tu515* (Fig 1A, S1A Fig, Table 1). Second, we crossed the *tu1163* and *tu515* mutants and established that *tu1163/tu515* trans-heterozygotes were no-Eu showing that the two mutants do not complement each other (Table 1). Third, we overexpressed the complementary DNA (cDNA) of *nhr-1* driven by the *nhr-1* promoter region in the *nhr-1(tu1163)* mutant background and obtained an almost complete rescue (Table 1). Fourth, we crossed *nhr-1(tu1163)* with *nhr-40(tu505)* and observed a highly penetrant no-Eu phenotype in double mutant animals, similar to the phenotype of *tu515 nhr-40(tu505)* mutants (Table 1). Taken together, frameshift alleles of *nhr-1* and the original suppressor allele *tu515* exhibit the same phenotype, do not complement each other, and have identical epistatic interactions with *nhr-40(tu505)*. Therefore, we conclude that *nhr-1* is the suppressor of *nhr-40(tu505)*.

Reverse genetic analysis of *nhr-40* results in all-stenostomatous mutants

The available alleles of *nhr-1* and *nhr-40* have different phenotypes with regard to mouth-form frequency and morphology. This is surprising because NHRs often form heterodimers [29], in which case *loss-of-function* phenotypes of interacting partners are identical. Two different hypotheses could explain our observations. First, *nhr-1* and *nhr-40* may indeed have different functions. Second, the three available alleles of *nhr-40* (*tu505*, *iub6*, *iub5*), all of which are non-synonymous substitutions outside of the DNA-binding domain (DBD) [27], may represent *gain-of-function* alleles. Our previous analysis had suggested that these alleles are *loss-of-function* based on the phenotype of *nhr-40* overexpression, which resulted in all-St animals [27]. However, we recently realized that in *C. elegans*, overexpression of *Cel-nhr-40* and *loss-of-function* of *Cel-nhr-40* induced by RNAi and a deletion mutation all cause similar developmental defects [30]. This may occur if NHR-40 inhibits its own transcription [31] or if the

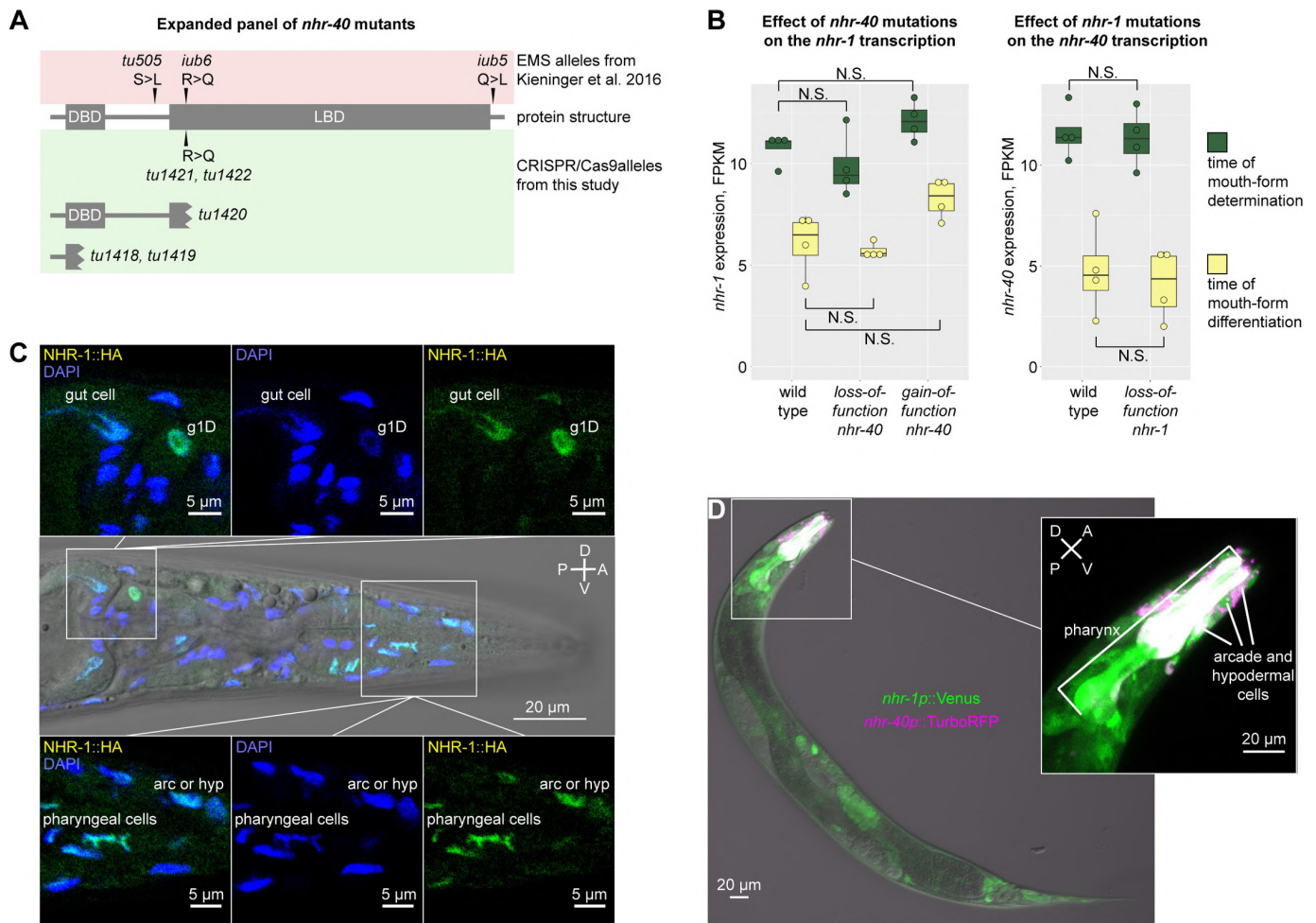


Fig 2. Reverse genetics, transcriptomics and expression patterns of *nhr-40* and *nhr-1*. (A) Protein structure of NHR-40 in wild-type and mutant animals. (B) Expression levels of *nhr-40* and *nhr-1* in wild type and mutants as revealed by transcriptomic profiling. (C) Antibody staining against the HA epitope in an *nhr-1* rescue line. (D) Expression patterns of *nhr-40* and *nhr-1* transcriptional reporters in a double reporter line. TurboRFP (magenta) and Venus (green) channels are presented as maximum intensity projections. Co-expression results in white color. D = dorsal, V = ventral, A = anterior, P = posterior, N.S. = not significant, FPKM = Fragments Per Kilobase of transcript per Million mapped reads.

<https://doi.org/10.1371/journal.pgen.1008687.g002>

concatenated coding sequence of the rescue construct acts as a substrate to induce RNAi [30]. Therefore, we investigated *nhr-40* in *P. pacificus* further, and generated nonsense alleles using CRISPR/Cas9.

We introduced mutations in two different locations in *nhr-40* (Fig 2A). The alleles *tu1418* and *tu1419* truncate the DBD. The *tu1420* allele contains a frameshift at the beginning of the LBD while leaving the DBD intact. We phenotyped the newly obtained mutants in liquid S-medium, which represses the Eu morph, and on agar plates, which induces it [19]. All frameshift alleles had a completely penetrant all-St phenotype in both culture conditions, which is opposite to the original ethyl methanesulfonate (EMS) alleles (Table 1). The newly obtained *nhr-40* mutants displayed no morphological abnormalities such as those observed in *nhr-1* mutants. Moreover, *nhr-1* mutants were epistatic to such mutants with respect to abnormal mouth morphology (S4 Fig). Additionally, we created a null allele, *tu1423*, which contains a 13 kb deletion or rearrangement of the locus (S2A Fig). This null allele again had a completely

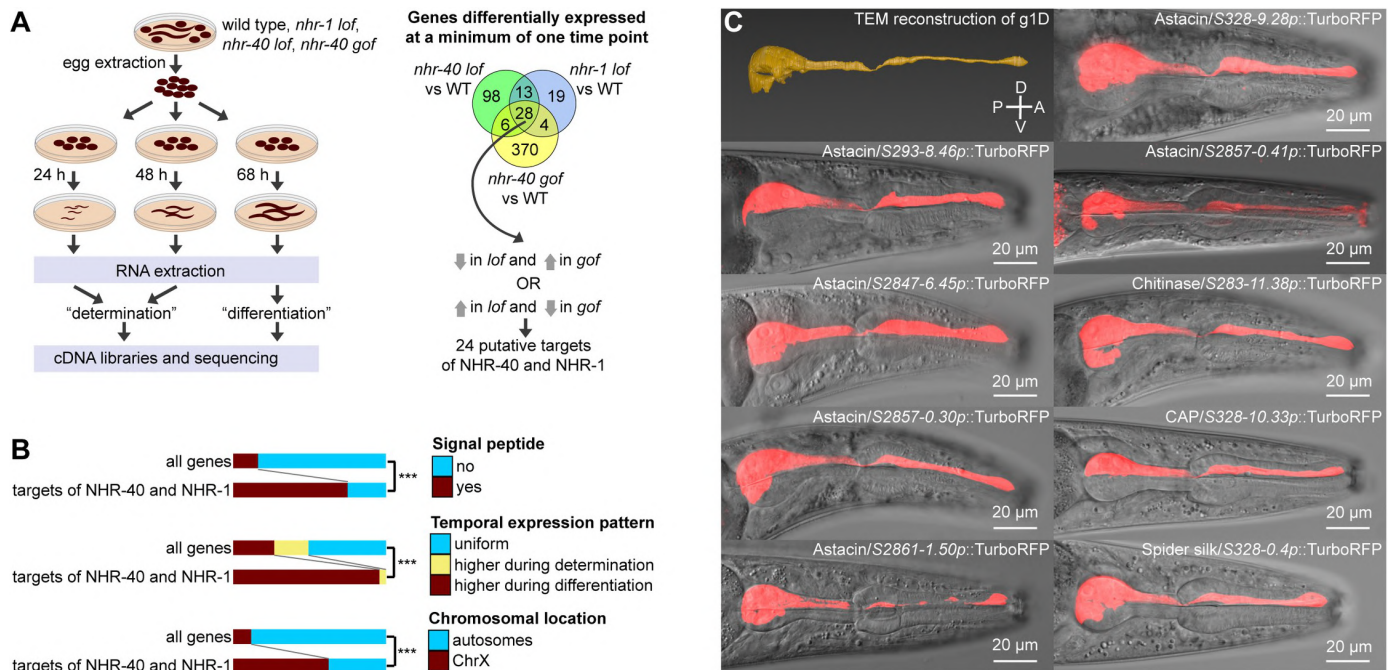


Fig 3. Target genes of NHR-40 and NHR-1. (A) Experimental setup of transcriptomics experiment and selection criteria to identify target genes. (B) Trends among target genes compared to genome-wide pattern. (C) Transmission electron microscopy reconstruction of the dorsal pharyngeal gland cell (g1D) [53] and expression patterns of transcriptional reporters for nine selected targets of NHR-40 and NHR-1. TurboRFP channel is presented as standard deviation projections. *lof* = loss-of-function, *gof* = gain-of-function, *** = $p < 0.001$, D = dorsal, V = ventral, A = anterior, P = posterior.

<https://doi.org/10.1371/journal.pgen.1008687.g003>

penetrant all-St phenotype and showed no morphological abnormalities (Table 1). To eliminate the possibility that the phenotype of the EMS mutants was caused by random mutations outside *nhr-40*, we introduced a nucleotide substitution identical to *iub6* via homology-directed repair (Fig 2A). Indeed, the two resulting alleles, *tu1421* and *tu1422*, had an all-Eu phenotype, identical to that of *iub6* and other EMS alleles, and opposite to that of the frameshift alleles (Table 1). Thus, frameshift mutations in DBD, LBD, and the deletion/rearrangement of the entire gene have an opposite phenotype to that of the three previously isolated non-synonymous substitutions. We conclude that *tu505*, *iub6*, *iub5*, *tu1421* and *tu1422* are gain-of-function alleles.

NHR-40 and NHR-1 interact post-transcriptionally

In GRNs, transcription factors may activate or repress each other transcriptionally [32–36], or alternatively, they may interact at the post-transcriptional level. The latter includes indirect interactions, such as independent binding to the same promoters [37], or ligand-mediated interactions [38]. To distinguish if *nhr-1* and *nhr-40* interact at the transcriptional or post-transcriptional level, we analyzed the transcriptomes of wild type, *nhr-1 loss-of-function*, *nhr-40 loss-of-function* and *nhr-40 gain-of-function* mutants at two developmental stages (Fig 3A). RNA collected from J2-J4 larvae is enriched with transcripts expressed at the time of mouth-form determination, as environmental manipulation during this time window affects morph frequency [39]. RNA collected from J4 larvae and adults is enriched with transcripts expressed at the time of mouth-form differentiation, because cuticularized mouthparts that distinguish the two morphs are believed to be secreted during the J4-adult molt [40]. We found that at both time points, *nhr-40* transcript levels were not affected by loss-of-function of *nhr-1*.

Similarly, *nhr-1* transcript levels were not affected by *loss-of-function* of *nhr-40*, although they were slightly, but not significantly increased by *nhr-40 gain-of-function* (Fig 2B). Thus, at the transcriptional level, both *nhr* genes remain unaffected by the *loss-of-function* of the other gene. Therefore, NHR-40 and NHR-1 may interact at the post-transcriptional level, although the possibility remains that their transcriptional interaction in specific cells is masked in whole-animal transcriptome data.

***nhr-40* and *nhr-1* are expressed at the site of polyphenism**

Next, we wanted to determine the expression pattern of *nhr-1* and *nhr-40* and test if they were co-expressed. We took three complementary approaches to establish the expression pattern of *nhr-1*. First, we created transcriptional reporters comprising the presumptive promoter region upstream of the potential start site in the second exon fused with TurboRFP or Venus. The resulting expression pattern was broad with the strongest expression in the head, including both muscle and gland cells of the pharynx, and what may be the hypodermal and arcade cells (Fig 2D, S1B Fig). Second, we performed antibody staining against an HA epitope tag in the *nhr-1* rescue line described above. We observed a similar expression pattern that was predictably localized to the nuclei (Fig 2C). Finally, we used CRISPR/Cas9 to “knock in” an HA tag in the endogenous *nhr-1* locus at the C-terminus of the coding sequence. Antibody staining against HA revealed a similar expression pattern, but with a weaker signal due to the lower number of copies of endogenous DNA (S1C Fig). Together, these results show that NHR-1 localizes to nuclei of multiple cells in the head region, with strong expression in pharyngeal muscle cells, which presumably secrete structural components of the teeth.

To explore whether NHR-40 and NHR-1 are expressed in overlapping tissues, we created a double reporter line, in which the *nhr-40* promoter is fused to TurboRFP and the *nhr-1* promoter to Venus. We observed a strong and consistent expression of *nhr-40* in the head. Specifically, it localized to the pharyngeal muscle cells and cells whose cell body position is consistent with them being arcade or hypodermal cells (Fig 2D, S1D Fig). *nhr-40* and *nhr-1* signals co-localized in a subset of presumptive hypodermal and arcade cells, and in the pharyngeal muscles. In contrast, only *nhr-1* was expressed in the dorsal pharyngeal gland cell g1D (Fig 2D, S1D and S1E Fig). In summary, while the expression of *nhr-40* is more restricted than the expression of *nhr-1*, the two genes display robust co-localization in several cell types.

Common transcriptional targets of NHR-40 and NHR-1 encode extracellular proteins expressed during mouth-form differentiation

Since NHR-40 and NHR-1 are co-expressed and regulate the same phenotype, we speculate that they regulate a set of common target genes, even though such regulation may be indirect. We analyzed the full list of genes differentially expressed between the wild type and mutant samples from the experiments described above. Given the pleiotropic action of NHR-40 and NHR-1, we applied the following selection criteria. We only retained genes whose transcript levels at either of the two examined time points were simultaneously altered in *nhr-1*, *nhr-40 loss-of-function*, and *nhr-40 gain-of-function* mutants (Fig 3A). Only 28 genes satisfied this criterion, and their expression changed in the same direction in the *loss-of-function* mutants of *nhr-1* and *nhr-40*. We further retained those genes whose expression changed in one direction in the *loss-of-function* mutants of *nhr-1* and *nhr-40*, and in the opposite direction in the *gain-of-function* mutants of *nhr-40* (Fig 3A), resulting in a list of 24 genes, provided in Table 2. Interestingly, the expression of 23 of them decreased in the *loss-of-function* mutants (Table 2).

Table 2. List of targets of NHR-40 and NHR-1.

Chr	Wormbase WS268 identifier	El Paco annotation v1 identifier	Predicted PFAM domains	LFC, <i>nhr-1</i> vs. WT, DT	LFC, <i>nhr-1</i> vs. WT, DF	LFC, <i>nhr-40 lof</i> vs. WT, DT	LFC, <i>nhr-40 lof</i> vs. WT, DF	LFC, <i>nhr-40 gof</i> vs. WT, DT	LFC, <i>nhr-40 gof</i> vs. WT, DF
X	PPA05669	UMM-S328-9.28-mRNA-1	Astacin	-6.2	-8.0	-7.2	-7.8	2.1	1.1
IV	PPA42525	UMM-S2847-7.46-mRNA-1	Astacin	-4.8	-4.0	-7.8	-4.7	1.4	NS
IV	PPA05955	UMM-S2847-6.45-mRNA-1	Astacin	-4.0	-3.7	-6.6	-6.0	1.6	1.1
X	PPA05618	UMM-S328-7.47-mRNA-1	Astacin	-3.6	-2.9	-7.7	-7.0	1.5	1.0
X	PPA16331	UMA-S293-8.46-mRNA-1	Astacin	-2.7	-4.1	-3.5	-4.2	1.7	NS
X	PPA39735	UMM-S328-10.33-mRNA-1	CAP	-2.2	-2.5	-5.1	-5.3	1.4	NS
I	PPA32730	UMM-S57-4.91-mRNA-1	Astacin	-2.2	-1.6	-3.7	-3.0	1.7	0.8
X	PPA13058	UMM-S328-10.78-mRNA-1	CAP	-2.1	-2.4	-4.8	-4.8	1.4	NS
IV	PPA39293	UMM-S283-11.38-mRNA-1	Glyco_hydro_18	-1.8	-1.0	-3.0	-1.9	0.9	NS
X	PPA29522	UMM-S322-3.5-mRNA-1	CAP	-1.6	NS	-3.1	-1.5	1.0	NS
X	PPA39470	UMM-S293-11.30-mRNA-1	CAP	-1.5	-2.2	-4.4	-4.9	1.4	NS
X	PPA21910	UMA-S322-3.38-mRNA-1	CAP	-1.4	NS	-2.3	-1.2	0.9	NS
IV	PPA04200	UMM-S283-11.45-mRNA-1	Glyco_hydro_18; MFS_1	-1.3	-0.8	-2.1	-1.3	0.9	NS
X	PPA21987	UMA-S322-7.39-mRNA-1	Astacin	-1.1	-1.1	-1.5	NS	0.9	0.9
X	PPA27985	UMS-S2861-1.50-mRNA-1	Astacin	-1.0	-1.6	-1.4	-1.8	0.9	0.9
X	PPA30108	UMS-S328-0.4-mRNA-1	none	-0.9	-2.0	-1.7	-3.1	1.4	0.7
II	PPA27560	UMS-S10-46.25-mRNA-1	none	NS	-2.5	NS	-1.5	1.6	NS
I	PPA30435	UMM-S57-36.5-mRNA-1	Lectin_C	NS	-2.5	-6.4	-7.5	1.7	0.8
X	PPA34430	UMA-S2861-1.27-mRNA-1	Astacin	NS	-2.0	-1.4	-2.2	1.0	0.9
X	PPA38892	UMM-S250-3.76-mRNA-1	ShK	NS	-2.0	-1.8	-2.2	1.0	NS
X	PPA20266	UMM-S2857-0.30-mRNA-1	Astacin	NS	-1.9	-1.4	-2.5	1.1	NS
X	PPA42924	UMM-S2857-0.41-mRNA-1	Astacin	NS	-1.1	NS	-1.7	0.8	NS
I	PPA03932	UMM-S7-5.16-mRNA-1	Astacin	NS	-1.1	NS	-1.6	1.1	0.8
IV	PPA06264	UMA-S2838-46.74-mRNA-1	adh_short; KR; THF_DHG_CYH_C	NS	2.2	NS	3.0	-2.0	NS

Chr = chromosome, LFC = log fold change, WT = wild type, *lof* = loss of function, *gof* = gain of function, DT = determination, DF = differentiation, NS = not significant.

<https://doi.org/10.1371/journal.pgen.1008687.t002>

We hypothesized that if the making of cuticularized mouthparts involves these genes, they must encode extracellular proteins, and their expression is likely to be biased towards the time of mouth-form differentiation. To verify the extracellular function of the target proteins, we predicted signal peptides and compared the list of targets with the genome-wide pattern. Indeed, we found that the targets of NHR-40 and NHR-1 are significantly enriched with genes containing signal peptides (Fig 3B). To examine a potential temporal expression bias, we compared the wild-type transcriptomes at the time of mouth-form determination and mouth-form differentiation. While most genes in the genome (51%) showed uniform expression at the two time points, 23 of the 24 targets of NHR-40 and NHR-1 were more highly expressed at the time of mouth-form differentiation (Fig 3B). Surprisingly, we also observed a third trend in our data set. While only 12% of all genes in the genome are located on the X chromosome, 15 of the 24 targets of NHR-40 and NHR-1 were X-linked (Fig 3B), which resembled the bias in the chromosomal location of hermaphrodite-specific somatically expressed genes in *C. elegans* [41].

To explore the potential functions of the NHR-40 and NHR-1 targets, we used information about their annotated protein domains. Surprisingly, 12 of the 24 genes contain an Astacin domain (Table 2) typical of secreted or membrane-anchored Zinc-dependent endopeptidases [42]. Of the 40 Astacin-containing genes in *C. elegans*, only *dpy-31*, *nas-6* and *nas-7* have known functions, whereby mutations in these genes result in abnormal cuticle synthesis [43,44]. Another five of the 24 NHR targets encode a CAP (cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1) domain (Table 2), which is contained in extracellular proteins with diverse functions [45–47], including the proteolytic modification of extracellular matrix [48]. Two genes belong to the glycoside hydrolases family 18 (Table 2), which includes chitinases and chitinase-like proteins [49] that may modify the cuticle, as chitin is the main component of the cuticle in nematodes [50]. Finally, the NHR target list includes an unannotated protein, PPA30108 (Table 2), which contains multiple GGF and GGR repeats, similar to some structural proteins of spider silk [51,52]. Thus, the examination of the domain composition of the targets of NHR-40 and NHR-1 suggests that many encode enzymes that may directly modify the cuticle.

A duodecuple Astacin mutant shows no mouth-form abnormalities

Next, we tested if mutations in the identified genes affected mouth-form frequency or morphology. We therefore performed systematic CRISPR/Cas9 knockout experiments of the 23 genes downregulated in the *loss-of-function* mutants. To compensate for potential redundancy between paralogous genes encoding identical domains, we produced lines in which all such genes are inactivated simultaneously. For example, rather than generating 12 strains with mutations affecting single Astacin-encoding genes, we produced a duodecuple mutant line, in which we sequentially knocked out all 12 genes (Table 1). We phenotyped the mutants both on agar plates and in liquid S-medium. However, we detected no significant change in mouth-form frequencies and no recapitulation of the morphological defects of *nhr-1*. Similarly, we produced a quintuple CAP mutant and double chitinase mutants and observed no change in mouth-form frequency or morphology (Table 1). We speculate that this may be caused by the extreme redundancy in the factors involved. For instance, despite mutagenizing 12 Astacin-encoding genes, there are more than 60 such genes in the genome. Consistent with this, in a phenotypic screen of Astacin genes in *C. elegans*, the majority showed no detectable phenotypes and the function of one, *nas-7*, was only elucidated due to its enhancement of a weakly penetrant allele of *nas-6* [44]. Alternatively, it is also possible that some examined genes function in other tissues unrelated to mouth morphology. Therefore, we next studied the spatial expression of selected downstream target genes.

Downstream targets genes are expressed in the same pharyngeal gland cell

We selected six of the 12 Astacin genes, one chitinase gene, one CAP gene, and the gene bearing similarity to spider silk proteins, and created transcriptional reporters by fusing their promoters with TurboRFP. Remarkably, all reporter lines showed expression in the same single cell, the dorsal pharyngeal gland cell g1D (Fig 3C). In contrast, we found no expression in the pharyngeal muscles or other expression foci of *nhr-40* and *nhr-1*. Thus, all analyzed targets are co-expressed with *nhr-1* in g1D (Fig 3C, S1E Fig). The recent reconstruction of the pharyngeal gland cell system of *P. pacificus* [53] revealed that the cell body of g1D is located at the posterior end of the pharynx. It sends a long process through the entire pharynx to the anterior tip where it connects, via a short duct in the cuticle, to a channel in the dorsal tooth which opens into the buccal cavity (Figs 1B and 3C). Importantly, the process of g1D is surrounded by pharyngeal muscle cells which directly underlie the teeth. Therefore, we hypothesize that the enzymes excreted from g1D act on the structural components that are themselves secreted by the pharyngeal muscles.

Expansion of the pharyngeal gland cells is concomitant with the emergence of teeth

The expression of the targets of NHR-40 and NHR-1 in g1D is remarkable, because g1D is the site of a major evolutionary innovation in the family Diplogastridae, to which *P. pacificus* belongs. The pharynx in free-living nematodes of the order Rhabditida and the outgroup [54] family Teratocephalidae is divided into two parts. The anterior part, called the corpus, is muscular, and in some lineages ends with a dilation, called the median bulb. The posterior part, called the postcorpus, is divided into a narrow isthmus and a dilation, called the terminal bulb, which contains muscle cells and three to five gland cells. The terminal bulb contains muscular valves that form a specialized cuticular structure, the grinder, which helps fragment food particles [55] (Fig 4). Phylogenetic reconstruction indicates that the outgroup Teratocephalidae, and the rhabditid families Cephalobidae and Rhabditidae retained the ancestral character states, whereby they have a grinder, but no teeth [56–58]. In contrast, Diplogastridae have no grinder, but they have concomitantly gained teeth at the base of the family [7,59]. The acquisition of teeth and the loss of the grinder were accompanied by the reduction of the muscle cells

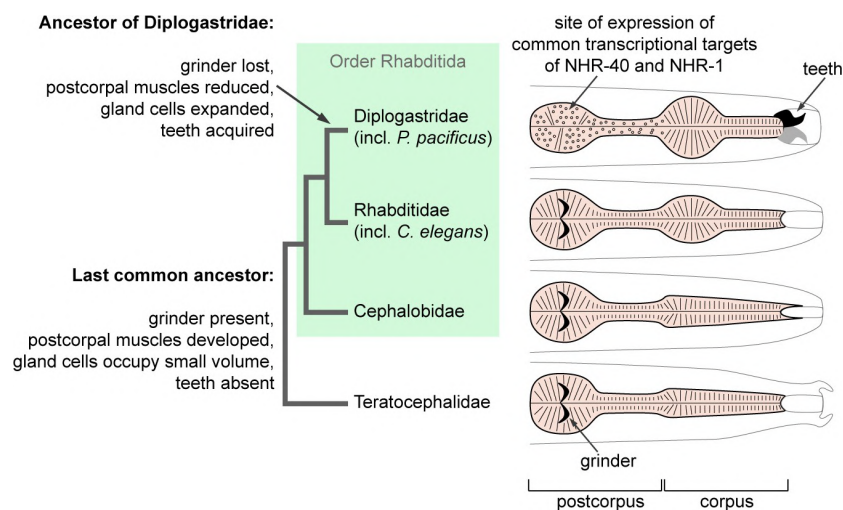


Fig 4. Evolution of pharynx morphology in the order Rhabditida.

<https://doi.org/10.1371/journal.pgen.1008687.g004>

in the postcorpus, and an expansion of three gland cells g1D, g1VL, and g1VR, one in each sector of the trilaterally symmetrical pharynx [53,59] (Fig 4). While the exact role of pharyngeal gland cells in *C. elegans* and other nematodes has remained elusive [55], we speculate that the functional remodeling of g1D, in which the target genes of NHR-40 and NHR-1 are expressed, may be a prerequisite for the formation of teeth and the evolution of predation. Therefore, we investigated the evolutionary dynamics of the identified genes expressed in this cell.

Conserved transcription factors regulate fast-evolving target genes

To investigate if the morphological lineage-specific evolutionary innovation in *P. pacificus* and Diplogastridae is associated with taxonomically restricted genes, we reconstructed the phylogeny of NHR genes and their identified targets. This is an important evolutionary question as recent genomic studies involving deep taxon sampling revealed high evolutionary dynamics of novel gene families in *Pristionchus*, with only one third of all genes having 1:1 orthologs between *P. pacificus* and *C. elegans* [60,61]. First, we reconstructed the phylogeny of NHR genes. We identified similar numbers of NHR genes in the genomes of *P. pacificus* and *C. elegans*—254 and 266 genes, respectively. In the phylogenetic tree (Fig 5A), most clades contained genes from predominantly or exclusively one of the two species. These genes likely result from lineage-specific duplications and losses, a phenomenon commonly seen in nematode gene families [62]. *nhr-40* and *nhr-1*, however, belonged to one of the few clades that contained a mixture of genes from both species, with many genes displaying a 1:1 orthology relationship. Indeed, the *P. pacificus* and *C. elegans* copies of *nhr-40* and *nhr-1* showed 1:1 orthology with 100% bootstrap support (Fig 5A). Importantly, *nhr-40* and *nhr-1* are also extremely closely related to each other (Fig 5A). Thus, in the overall context of NHR evolution, *nhr-40* and *nhr-1* are closely related duplicates that have been conserved since the divergence of *P. pacificus* and *C. elegans*.

The conservation of *nhr-40* and *nhr-1* is in stark contrast to the evolutionary history of their downstream targets. To reconstruct the phylogenies of the Astacin, CAP and chitinase genes (Fig 5B–5D), we used functional domains rather than complete genes to facilitate the alignment of genes with different domain architectures. Similar to the case of NHRs, all three gene families exhibit strong signatures of lineage-specific expansions. Furthermore, all target genes containing Astacin, CAP and chitinase domains belonged to such lineage-specific clades (Fig 5B–5D). These findings suggest that the targets of NHR-40 and NHR-1 undergo rapid turnover. This is further supported by the phylogeny of CAP genes within the genus *Pristionchus*. Specifically, the five targets identified in *P. pacificus* clustered separately from the homologs in the early branching species *P. fissidentatus* with 94% bootstrap support (Fig 5E). Thus, two conserved NHRs target rapidly evolving downstream genes of multiple gene families. We speculate that the striking co-expression of the target genes might result from an ancient regulatory linkage between the NHRs and the promoters of the ancestral target genes. Alternatively, however, these transcription factors might have captured new promoters by mutations creating binding sites for them, a possibility that will become directly testable once the target sequences of NHR-1 and NHR-40 have been identified.

Discussion

In this study, we expanded the GRN controlling predatory vs. non-predatory plasticity in *P. pacificus*, thereby enhancing the molecular understanding of plasticity. We uncovered novel genetic factors and genomic features at two regulatory levels, which allowed linking rapid gene evolution with morphological innovations associated with plasticity. First, we identified a mutation in the nuclear receptor gene *nhr-1*, which disrupts mouth-form determination. Most

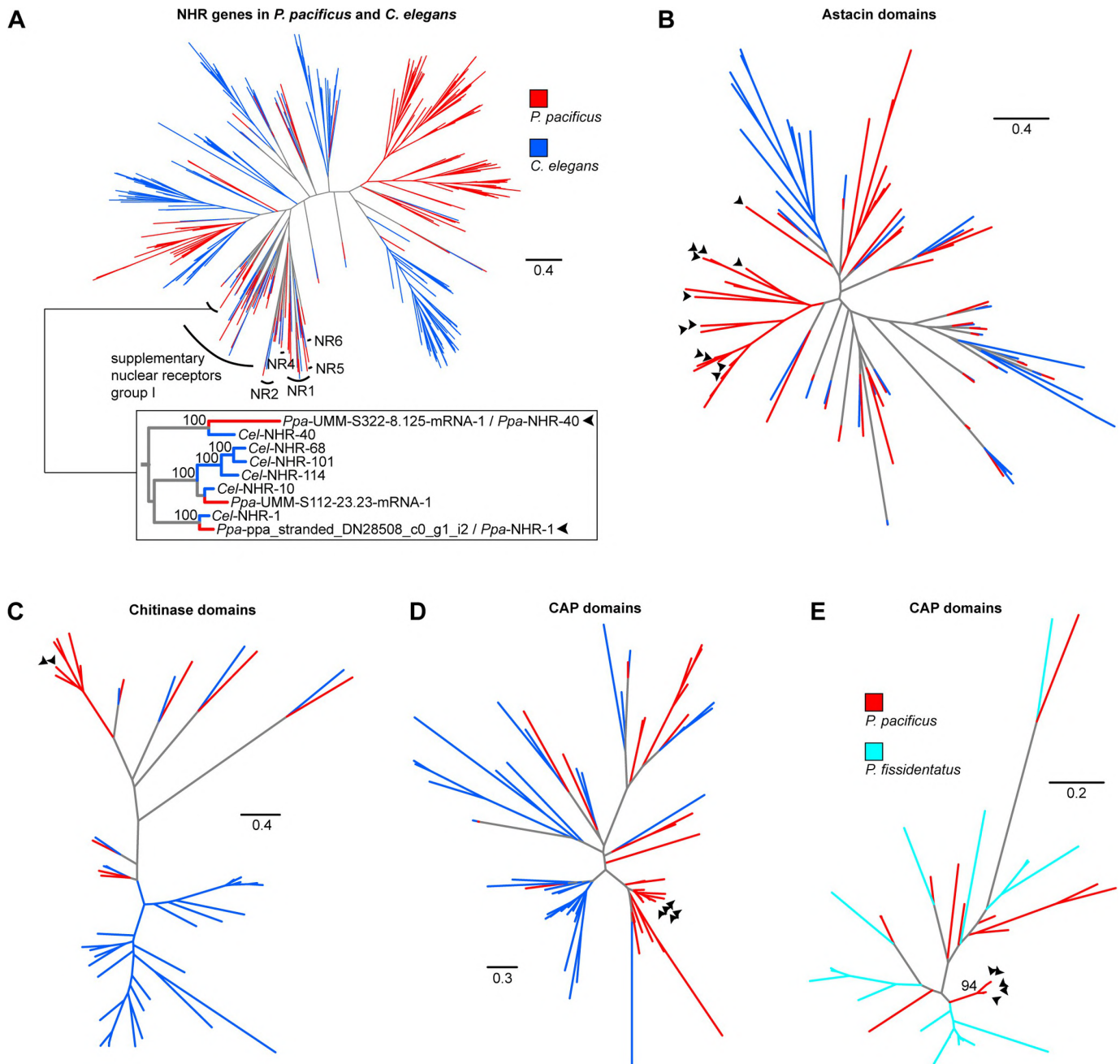


Fig 5. Evolution of *nhr-40*, *nhr-1*, and their target genes. Arrowheads point at the genes of interest. Protein-based trees of NHR genes (A), Astacin domains (B), chitinase domains (C), and CAP domains (D) in *P. pacificus* and *C. elegans*. (E) Nucleotide-based tree of the CAP domains from a poorly-resolved protein-based subtree of all predicted CAP domains in *P. pacificus* and *P. fissidentatus*.

<https://doi.org/10.1371/journal.pgen.1008687.g005>

previously identified genes, such as *eud-1* or *sult-1/seud-1*, influence the determination process by affecting the preferred developmental trajectory, but the resulting morphology exhibits no observable differences to the corresponding wild-type morphology [18,23,24]. On the other hand, interfering with heat shock protein activity, including a mutation in *daf-21/Hsp90*, produces aberrant morphologies while maintaining the dimorphism [63]. In contrast to both

classes of genetic interventions, mutations in *nhr-1* lead to a morphology that combines features of normal Eu and St morphs, with no apparent dimorphism (S3 Fig). Therefore, we speculate that NHR-1 is required for mouth-form determination and the specification of both morphs. By contrast, we showed that *gain-* and *loss-of-function* mutations in *nhr-40* result in all-Eu and all-St phenotypes respectively, reminiscent of the role of *daf-12*, another *nhr* gene, in controlling dauer plasticity in *C. elegans* [64]. Different phenotypic effects of *nhr-1* and *nhr-40* are also consistent with the lack of evidence of transcriptional regulation of one factor by the other. Except for DAF-12 in *C. elegans*, no single nematode NHR has been de-orphanised. Therefore, the identification of the potential ligands of NHR-1 and NHR-40 may reveal additional layers of regulation and elucidate their cross-talk. Indeed, recent studies suggested that cytosolic sulfotransferases, including *sult-1/seud-1* in *P. pacificus* and its homolog *ssu-1* in *C. elegans*, may regulate NHRs by modifying their ligands [23,24,65].

Second, the transcriptomic analysis of *nhr-1* and *nhr-40* mutants revealed an unexpectedly small number of downstream targets. While cell-specific signals may be masked in whole-animal transcriptome data, and our selection criteria excluded genes affected by the *gain-of-function* of *nhr-40* in other ways than by exhibiting increased transcript levels, having a small list of target genes enabled a systematic analysis of their genomic location, function and expression. Surprisingly, the majority of targets were located on the X-chromosome, which parallels X-linkage of many previously identified genes associated with mouth-form, including both *nhr-40* and *nhr-1*, and additionally the multigene switch locus comprising *eud-1*, *nag-1* and *nag-2*. While the exact meaning of this phenomenon remains unclear, the X chromosome in *C. elegans* is enriched with hermaphrodite-biased somatically expressed genes [41]. Accordingly, the incidence of Eu morphs is higher in *P. pacificus* hermaphrodites than in males [39], which may be reflected in the chromosomal distributions of the genes associated with the Eu morph.

Both the absence of phenotypes in duodecuple and quintuple mutants, and the restricted expression of all tested genes in the same cell g1D are compatible with extreme functional redundancy. Such redundancy might result from features of genome evolution that are common to nematodes and other animals. Studies over the last decade revealed that nematode genomes are gene-rich and exhibit high rates of gene birth and death [60,66,67]. In particular, enzyme-encoding genes are subject to high evolutionary dynamics [62]. Therefore, the position of genes in GRNs may determine the speed and direction of their evolution. Consistent with this idea, many genes encoding proteins of signal transduction and their terminal transcription factors are highly conserved across animals [68–70]. In this study, we complement this knowledge by showing that the downstream targets of conserved transcription factors are indeed fast evolving genes. Importantly, their expression focus, the g1D cell, also underwent a major evolutionary change, whereby its structural and functional remodeling accompanied the emergence of teeth in the family Diplogastridae. Thus, our study demonstrates that fast-evolving genes are expressed in a fast-evolving cell, linking morphological innovations with rapid gene evolution.

Materials and methods

Maintenance of worm cultures and genetic crosses

Stock cultures of all strains used in this study were reared at room temperature (20–25°C) on nematode growth medium (NGM) (1.7% agar, 2.5 g/L tryptone, 3 g/L NaCl, 1 mM CaCl₂, 1 mM MgSO₄, 5 mg/L cholesterol, 25 mM KPO₄ buffer at pH 6.0) in 6 cm Petri dishes, as outlined in the *C. elegans* maintenance protocol [71]. *Escherichia coli* OP50 was used as food source. Bacteria were grown overnight at 37°C in L Broth (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, pH adjusted to 7.0), and 400 μL of the overnight culture was pipetted on NGM

agar plates and left for several days at room temperature to grow bacterial lawns. *P. pacificus* were passed on these lawns and propagated by passing various numbers of mixed developmental stages. To cross worms, agar plates were spotted with 10 μ L of the *E. coli* culture, and five to six males and one or two hermaphrodites were transferred to the plate and allowed to mate. Males were removed after two days of mating.

Mouth form phenotyping

We phenotyped worms in two culture conditions. Rearing *P. pacificus* on solid NGM induces the Eu morph and facilitates identification of Eu-deficient (all-St) phenotypes. Conversely, growing worms in liquid S-medium (5.85 g/L NaCl, 1 g/L K_2HPO_4 , 6 g/L KH_2PO_4 , 5 mg/L cholesterol, 3 mM $CaCl_2$, 3 mM $MgSO_4$, 18.6 mg/L disodium EDTA, 6.9 mg/L $FeSO_4 \cdot 7H_2O$, 2 mg/L $MnCl_2 \cdot 4H_2O$, 2.9 mg/L $ZnSO_4 \cdot 7H_2O$, 0.25 mg/L $CuSO_4 \cdot 5H_2O$ and 10 mM Potassium citrate buffer at pH 6.0) represses the Eu morph and facilitates identification of Eu-constitutive (all-Eu) phenotypes [19,71]. As food source, S-medium contained *E. coli* OP50 in the amount corresponding to 100 mL of an overnight culture with OD_{600} 0.5 per 10 mL of medium. We started phenotyping by isolating eggs from stock culture plates, which contained large numbers of gravid hermaphrodites and eggs deposited on the agar surface [71]. To isolate eggs, we washed worms and eggs from plates with water, and incubated them in a mixture of 0.5 M NaOH and household bleach at 1:5 final dilution for 10 min with regular vortexing to disintegrate vermiform stages. Remaining eggs were pelleted at 1,300 g for 30 sec, washed with 5 mL of water, pelleted again, resuspended in water and pipetted on agar plates or into S-medium. Agar plates were left at room temperature (20–25°C) for 3–5 days and 25 mL Erlenmeyer flasks with liquid medium were shaken at 22°C, 180 rpm for 4–6 days. Adult hermaphrodites were immobilized on 5% Noble Agar pads with 0.3% NaN_3 added as an anaesthetic, and examined using differential interference contrast (DIC) microscopy. Animals that had a large right ventrosublateral tooth, curved dorsal tooth, and the anterior tip of the promesostegostom posterior to the anterior tip of the gymnostom plate were classified as Eu morphs. Animals that did not exhibit these three characters simultaneously were classified as St morphs, although there was a distinction between the morphology of *nhr-1* mutants and of other all-St mutants (S1A Fig).

Geometric morphometric analysis

We reused the published [63] landmark data for the wild-type strain RS2333 and the *daf-21* (*tu519*) mutant. We complemented this data set with newly collected data for the *nhr-1* (*tu1163*) mutant, whereby we imaged young adults mounted on microscope slides on 5% Noble agar pads containing 0.3% NaN_3 as an anaesthetic. Only individuals with their right body side facing upwards were imaged. We took stack images of the anterior tip of the head, and recorded X and Y coordinates of 20 landmarks identical to the ones used in the previous study [63] using FIJI [72]. Procrustes alignment and PCA were done in R (ver. 3.4.4) [73] using geomorph package [74].

CRISPR/Cas9 mutagenesis

We followed the previously published protocol for *P. pacificus* [75] with subsequently introduced modifications [76]. All target-specific CRISPR RNAs (crRNAs) were designed to target 20 bp upstream of the protospacer adjacent motifs (PAMs). We purchased crRNAs and universal trans-activating CRISPR RNA (tracrRNA) from Integrated DNA Technologies (Alt-R product line). 10 μ L of the 100 μ M stock of crRNA was combined with 10 μ L of the 100 μ M stock of tracrRNA, denatured at 95°C for 5 min, and allowed to cool down to room

temperature and anneal. The hybridization product was combined with Cas9 protein (purchased from New England Biolabs or Integrated DNA Technologies) and incubated at room temperature for 5 min. The mix was diluted with Tris-EDTA buffer to a final concentration of 18.1 μM for the RNA hybrid and 2.5 μM for Cas9. When site-directed mutations were introduced via homology-directed repair, a ssDNA oligo template designed on the same strand as the gRNA was included in the mix at a final concentration of 4 μM . The diluted mixture was injected in the gonad rachis of approximately one day old adult hermaphrodites.

Eggs laid by injected animals within a 12–16 h period post injection were recovered, and the F1 progeny were singled out upon reaching maturity. After F1 animals have laid eggs, they were placed in 10 μL of single worm lysis buffer (10 mM Tris-HCl at pH 8.3, 50 mM KCl, 2.5 mM MgCl_2 , 0.45% NP-40, 0.45% Tween 20, 120 $\mu\text{g}/\text{ml}$ Proteinase K), frozen and thawed once, and incubated in a thermocycler at 65°C for 1 h, followed by heat deactivation of the proteinase at 95°C for 10 min. The resulting lysate was used as a template in subsequent PCR steps. Where possible, molecular lesions at the crRNA target sites were detected by melting curve analysis on a LightCycler 480 Instrument II (Roche) of PCR amplicons obtained using LightCycler 480 High Resolution Melting Master (Roche). Presence of mutations in candidate amplicons was further verified by Sanger sequencing. Alternatively, PCR was done using Taq PCR Master Mix (Qiagen) and all the F1 were Sanger sequenced.

To detect large rearrangements, we conducted whole genome re-sequencing of lines for which no PCR amplicon containing the crRNA target site could be obtained. For most such lines, we extracted genomic DNA using GenElute Mammalian Genomic DNA Miniprep Kit (Merck), whereby we modified the tissue digestion step by raising the Proteinase K concentration to 2 mg/mL, and prepared next-generation sequencing (NGS) libraries using Nextera DNA Flex Library Prep Kit (Illumina). For the *nhr-40* null mutant line, we followed a recently introduced cost-effective alternative procedure [77] with several modifications. Single worms were placed in 10 μL water, and frozen and thawed 3 times in liquid nitrogen. Then, we added 10 μL 2x single worm lysis buffer (20 mM Tris-HCl at pH 8.3, 100 mM KCl, 5 mM MgCl_2 , 0.9% NP-40, 0.9% Tween 20, 240 $\mu\text{g}/\text{ml}$ Proteinase K) and incubated the tubes in a thermocycler at 65°C for 1 h. After a clean-up using HighPrep beads (MagBio Genomics), DNA was eluted in 7 μL Tris buffer at pH 8.0. Then, 100 pg of DNA was diluted with water to the total volume of 9 μL , mixed with 2 μL 5X TAPS-DMF buffer (50 mM TAPS at pH 8.5, 25 mM MgCl_2 , 50% DMF) and 1 μL Tn5 transposase from Nextera DNA Library Prep Kit (Illumina) diluted beforehand 1:25 in dialysis buffer (100 mM HEPES at pH 7.2, 0.2 M NaCl, 0.2 mM EDTA, 0.2% Triton X-100, 20% glycerol). The mixture was incubated for 14 min at 55°C. Tagmented DNA was amplified using Q5 HotStart High-Fidelity DNA Polymerase (New England Biolabs) for 14 cycles, whereby adapters and indices were added as primer overhangs, and size-selected for 250–550 bp fragments using HighPrep beads (MagBio Genomics). NGS libraries prepared using both methods were sequenced in a paired-end run of a HiSeq 3000 machine (Illumina). Reads were mapped to the El Paco assembly of the *P. pacificus* genome [78] using Bowtie 2 (ver. 2.3.4.1) [79]. We visually inspected read coverage in the loci of interest using IGV [80] to identify the precise regions in which coverage was close to zero.

EMS mutagenesis

To induce heritable mutations in *P. pacificus*, we incubated a mixture of J4 larvae and young adults in M9 buffer (3 g/L KH_2PO_4 , 6 g/L Na_2HPO_4 , 5 g/L NaCl, 1 mM MgSO_4) with 47 mM ethyl methanesulfonate (EMS) for 4 h [81]. Subsequently, the worms were allowed to recover on agar plates with bacteria (see above), and 40–120 actively moving J4 larvae were singled out. After the animals have laid approximately 20 eggs, they were killed, and F1 progeny were

allowed to develop and reach maturity. F1 animals (which contained heterozygous mutants) were then singled out, and F2 progeny (which contained a mixture of genotypes, including homozygous mutants) were allowed to develop until adulthood. In each F1 plate, we determined the mouth form in 5–10 F2 individuals using Discovery V20 stereomicroscope (Zeiss). If at least one individual appeared to have a mouth form different from that of the background strain, such an animal was transferred to a fresh plate and its progeny was screened again using DIC until we gained confidence that a homozygous line was isolated. In the screen for suppressors of *nhr-40*, we mutagenized *nhr-40(tu505)* worms, which are all-Eu, screened approximately 1,000 F1 plates, and isolated one no-Eu allele, *tu515*. In an attempt to identify further downstream target genes, we conducted two suppressor screens in the *nhr-1(tu1163)* mutant background and screened approximately 3,800 F1 plates in total, but found no Eu individuals.

Mapping of *tu515*

We crossed the *tu515* mutant, produced in the background of the RS2333 strain (a derivative of the PS312 strain), to a highly-Eu wild type strain PS1843. The resulting males were crossed to a strain RS2089, which is a derivative of PS1843 containing a morphological marker mutation causing the Dumpy phenotype. The progeny were allowed to segregate and 100 no-Eu lines were established. Four individuals from each line were pooled and genomic DNA was extracted from the pool using the MasterPure Complete DNA and RNA Purification Kit (Epicentre). Additionally, genomic DNA was extracted from the *tu515* line. NGS libraries were prepared using Low Input Library Prep kit (Clontech) and sequenced on Illumina HiSeq3000. Raw Illumina reads of the *tu515* mutant and of a mapping panel were aligned to the El Paco assembly of the *P. pacificus* genome (strain PS312) [78] by the *aln* and *sampe* programs of the BWA software package (ver. 0.7.17-r1188) [82]. Initial mutations were called with the *samtools* (ver. 1.7) *mpileup* command [83]. The same program was used to measure PS312 allele frequencies in the mapping panel at variant positions with regard to whole genome sequencing data of the PS1843 strain [78]. S2B Fig shows that large regions between the positions 5 Mb and 16 Mb of the *P. pacificus* chromosome X exhibit high frequency of the PS312 alleles (the mutant background) in the mapping panel. In total, 28 non-synonymous/nonsense mutations (S1 Table) in annotated genes (El Paco gene annotations v1, Wormbase release WS268) were identified in the candidate interval by a previously described custom variant classification software [84].

Transgenesis

To identify putative promoter regions, which included 5' untranslated regions (UTR) and may have included the beginning of coding sequences, we manually re-annotated the 5' ends of predicted genes of interest using RNA-seq data and the information about predicted signal peptides. The ATG codon preceding the signal peptide or the last ATG codon in the second exon was designated as the putative start codon. As a general rule, the promoter region included a sequence spanning from the 3' end of the closest upstream gene on the same strand to the start codon, but if the upstream gene was located further than 2 kb away, a 1.5–2 kb region upstream of the identified start codon was designated as the putative promoter. In the case of inverted tandem duplicates in the head-to-head orientation, the 5' end of the promoter region was approximately in the middle between the start codons of the two genes. For the reporter constructs, we used the previously published coding sequences of TurboRFP [85] and Venus [27] fused with the 3' UTR of the ribosomal gene *rpl-23* [85]. For the *nhr-1* rescue construct, we used the native coding sequence, in which we replaced native introns with synthetic

introns, fused with the native 3' UTR. As the latter fragment could not be amplified from genomic or complementary DNA in one piece, we purchased a corresponding gBlocks fragments (Integrated DNA Technologies). FASTA sequences of all promoter regions, coding sequences and 3' UTRs are provided in [S1 Data](#).

Plasmids carrying reporter and rescue constructs, listed in [S2 Table](#), were created by Gibson assembly using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) or a homemade master mix [86]. Small modifications, such as deletions and insertions under 70 bp, were introduced using Q5 Site-Directed Mutagenesis kit (New England Biolabs). Injection mix for transformation was created by digesting the plasmid of interest, the marker plasmid carrying a tail-bound reporter *egl-20p::TurboRFP* (if applicable), and genomic DNA with FastDigest restriction enzymes (Thermo Fisher Scientific), whereby genomic DNA was cut with an enzyme(s) that had the same cutting site(s) as the enzyme(s) used to digest the plasmids. Digested DNA was purified using Wizard SV Gel and PCR Clean-Up system (Promega), and the components were mixed in the following ratios. Injection mixes with rescue constructs contained 1 ng/ μ L rescue construct, 10 ng/ μ L marker, and 50 ng/ μ L genomic DNA. Injection mixes with reporter constructs contained 10 ng/ μ L reporter construct, 10 ng/ μ L marker, and 60 ng/ μ L genomic DNA. The mix was injected in the gonad rachs of approximately 1 day old hermaphrodites, and their progeny was screened for fluorescent animals [85].

Antibody staining

We followed a previously published protocol [87] with minor modifications. Animals were washed from mature plates with phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄ at pH 7.4), passed over a 5–20 μ m nylon filter, concentrated at the bottom of a 2 mL tube and chilled on ice. We then added chilled fixative (15 mM Na-PIPES at pH 7.4, 80 mM KCl, 20 mM NaCl, 10 mM Na₂EGTA, 5 mM Spermidine-HCl, 2% paraformaldehyde, 40% MeOH), froze the worms in liquid nitrogen and thawed them on ice for 1–2 h with occasional inversion. Subsequently, the animals were washed twice with Tris-Triton buffer (100 mM Tris-HCl at pH 7.4, 1 mM EDTA, 1% Triton X-100), incubated in Tris-Triton buffer with 1% β -mercaptoethanol in a thermomixer at 600 rpm for 2 h at 37°C, washed once in borate buffer (25 mM H₃BO₃, 12.5 mM NaOH), incubated in borate buffer with 10 mM dithiothreitol in a thermomixer at 600 rpm for 15 min at room temperature, washed once in borate buffer, incubated in borate buffer with ~0.3% H₂O₂ in a thermomixer at 600 rpm for 15 min at room temperature, and washed once more in borate buffer. Next, the worms were washed three times with antibody buffer B (0.1% bovine serum albumin, 0.5% Triton X-100, 0.05% NaN₃, 1 mM EDTA in PBS) on a rocking wheel, incubated with a dye-conjugated antibody (Thermo Fisher Scientific, cat. # 26183-D550 and cat. # 26183-D488) diluted 1:25 in antibody buffer A (1% bovine serum albumin, 0.5% Triton X-100, 0.05% NaN₃, 1 mM EDTA in PBS) on a rocking wheel in the dark for 3 h at room temperature or overnight at 4°C, washed three times with antibody buffer B and mounted on slides in a 1:1 mixture of PBS and Vectashield (Vector Laboratories) with 1 μ g/mL DAPI added. Slides were imaged using a Leica SP8 confocal microscope.

RNA-seq analysis

To obtain a sufficient number of eggs, we passed young adult hermaphrodites to new agar plates with 5–10 animals per plate. After their F1 progeny have laid eggs (5–6 days), they were bleached (see above), then resuspended in 400 μ L water per starting plate, pipetted onto multiple fresh plates with 100 μ L suspension per fresh plate and placed at 20°C. Animals were collected at 24 h (corresponding to J2 and J3 larvae), 48 h (J3 and J4 larvae) and 68 h (J4 instar

larvae and young adults) post-bleaching by adding some water to the plates, scraping off the bacterial lawns with worms in them using disposable cell spreaders and passing the resulting suspension through a 5 μm nylon filter, which efficiently separated worms from bacteria. Worms were washed from the filter into 1.5 mL tubes, pelleted in a table-top centrifuge at the maximum speed setting, after which the supernatant was removed and 1 mL TRIzol (Invitrogen) was added to the worm pellets. Tubes were flash-frozen in liquid nitrogen and stored at -80°C for up to a month. To extract RNA, worms suspended in TRIzol were frozen and thawed three times in liquid nitrogen, debris were pelleted for 10–15 min at 14,000 rpm at 4°C , and 200 μL of chloroform was added to the supernatant. After vigorous vortexing and incubation at room temperature ($20\text{--}25^{\circ}\text{C}$) for 5 min, tubes were rotated for 15 min at 14,000 rpm at 4°C . The aqueous phase was combined with an equal volume of 100% ethanol, RNA was purified using RNA Clean & Concentrator Kit (Zymo Research) and its integrity was verified using RNA Nano chips on the Bioanalyzer 2100 instrument (Agilent).

To analyse the transcriptome at the time of mouth form determination, we combined 500 ng RNA isolated at 24 h with 500 ng RNA isolated at 48 h post-bleaching, and proceeded to make libraries using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). To analyse the transcriptome at the time of mouth form differentiation, we prepared libraries from 1 μg of RNA isolated at 68 h post-bleaching. For wild type strain PS312, four biological replicates were collected at different time points. For the mutants, two replicates of two independent alleles were collected at two different time points, and these were treated as four biological replicates. Specifically, we sequenced the following alleles: *nhr-1* (*tu1163*) *loss-of-function*, *nhr-1* (*tu1164*) *loss-of-function*, *nhr-40* (*tu505*) *gain-of-function*, *nhr-40* (*iub6*) *gain-of-function*, *nhr-40* (*tu1418*) *loss-of-function*, *nhr-40* (*tu1423*) *null*.

Libraries were sequenced in two paired-end runs of a HiSeq 3000 machine, whereby we aimed at 10–20 mln reads per library. Raw sequences have been deposited in the European Nucleotide Archive with the study accession number PRJEB34615 (<http://www.ebi.ac.uk/ena/data/view/PRJEB34615>). The fourth biological replicate of wild-type PS312 and all replicates of the *nhr-40* *loss-of-function/null* mutants were sequenced in a different run than the other samples. To ensure that batch effects were negligible, we additionally re-sequenced the first three replicates of wild-type PS312 in the same run and verified that coordinates in PCA conducted using complete transcriptomes were minimally altered when comparing the same samples sequenced in the two runs. Reads were mapped to the El Paco assembly of the *P. pacificus* genome [78] using STAR (ver. 020201) [88]. Differential expression analysis was carried out in R (ver. 3.4.4) [73] using Bioconductor (ver. 3.6) [89] and DESeq2 (ver. 1.18.1) [90], whereby we counted reads mapping to El Paco v1 gene predictions [78]. We applied an adjusted p-value cutoff of 0.05 and no fold change cutoff. Alignments and coverage were visualized in IGV [80].

To examine the transcript levels of *nhr-1* and *nhr-40*, we repeated differential expression analysis, whereby we counted reads mapping to Trinity-assembled transcripts generated from previously published RNA-seq data [26] because the El Paco v1 gene prediction for *nhr-1* was incorrect in that it was a fusion of multiple neighboring genes. To test the differences in FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values for *nhr-1* and *nhr-40* in different mutants at each of the two time points, we performed t-test as implemented in the *t.test* function in R (ver. 3.4.4) [73] and applied false discovery rate (FDR) correction to the p-values obtained. Prior to conducting the t-test, we verified the assumptions for parametric statistics by performing Shapiro-Wilk test for normality (*shapiro.test* function) and Levene test for homoscedasticity (*levene.test* function of the *car* package [91]). Signal peptides were predicted using SignalP (ver. 4.1) [92]. To compare relative numbers of genes in different categories listed in Fig 3B, we used chi-squared test as implemented in the *chisq.test* function in R (ver. 3.4.4) [73].

Phylogenetic reconstructions

To identify NHR, CAP, and chitinase genes in the *C. elegans* genome, we retrieved the current version (PRJNA13758) of predicted proteins and domains from the <http://wormbase.org> website and selected genes that contained “IPR001628”, “CAP domain”, and “IPR001223” as predicted InterPro domains, respectively. The list of Astacin genes was taken from an earlier study [93] and the corresponding gene predictions were manually retrieved from the <http://wormbase.org> website. To identify NHR, Astacin, CAP, and chitinase genes in the *P. pacificus* genome, we predicted domains in the El Paco v1 version of gene predictions [78] using HMMER (ver. 3.1b2) software in conjunction with the PFAM profile database [94] and selected genes that contained “PF00105”, “Astacin”, “CAP”, and “PF00704” as predicted PFAM domains, respectively. Manual inspection of the retrieved NHR genes in *P. pacificus* revealed that many of the gene predictions represent fusions of multiple neighboring genes. Therefore, we used the information about the predicted domains, RNA-seq data generated in this study, and Illumina and PacBio RNA-seq datasets generated earlier [26,95,96] to manually reannotate the NHR gene predictions in *P. pacificus*. We submitted the improved annotations to <http://wormbase.org> as part of a larger set of manually curated gene annotations [97]. For the tree of CAP domains in *P. pacificus* and *P. fissidentatus*, we predicted domains in the Pinocchio versions of gene predictions for both genomes [60] and selected genes that contained “PF00188” as a predicted PFAM domain. In the case of NHR genes, complete sequences were aligned, while in the case of other gene families, functional domains extracted using HMMER (ver. 3.1b2) were aligned to facilitate the alignment of genes with divergent domain architecture. Alignments were done in MAFFT (ver. 7.310) [98] and maximum likelihood trees were built using RAxML (ver. 8.2.11) [99]. Protein-based trees were generated with the following parameters: `-f a -m PROTGAMMAAUTO -N 100`. In the case of CAP domains in *P. pacificus* and *P. fissidentatus*, we first generated a protein-based tree and identified a poorly resolved subtree containing the genes of interest. To increase the number of informative sites, we extracted corresponding nucleotide sequences, aligned them in MAFFT and built a tree in RAxML with the following parameters: `-f a -m GTRCAT -N 100`. Obtained phylogenetic trees were visualized using FigTree (ver. 1.4.2). All phylogenetic trees and corresponding alignments are provided in [S2 Data](#).

Supporting information

S1 Fig. Additional images. (A) The mouth of the wild-type eury stomatous (Eu) morph, wild-type stenostomatous (St) morph, *nhr-1* mutant, and *nhr-40* mutant in two focal planes. (B) Expression pattern of an *nhr-1* transcriptional reporter in a young larva. TurboRFP channel is presented as a maximum intensity projection. (C) Antibody staining against the HA epitope in a line, in which the tag was “knocked in” into the endogenous locus. Fluorescent channel is presented as a maximum intensity projection. (D) Expression patterns of *nhr-40* and *nhr-1* transcriptional reporters in a double reporter line. TurboRFP (magenta) and Venus (green) channels are presented as standard deviation and maximum intensity projections, respectively. Co-expression results in white color. (E) Expression patterns of *nhr-40* and *nhr-1* transcriptional reporters in a double reporter line. TurboRFP is encoded as magenta, Venus as green. Co-expression results in white color. D = dorsal, V = ventral, A = anterior, P = posterior. (TIF)

S2 Fig. Additional bioinformatic analyses. (A) Whole-genome re-sequencing and RNA-seq of the null allele of *nhr-40*. (B) bulked segregant analysis of the suppressor of *nhr-40* (*tu505*) with the location of *nhr-1* marked with a dotted line. See [S1 Table](#) for the list of non-

synonymous and nonsense substitutions within the candidate region.
(TIF)

S3 Fig. Geometric morphometric analysis of selected strains. Geometric morphometric analysis of 20 landmarks in the mouth of the wild-type strain RS2333, the *nhr-1(tu1163)* mutant isolated in this study, and the *daf-21/Hsp90(tu519)* mutant previously shown to exhibit an aberrant mouth morphology while maintaining the dimorphism. Each point in the PCA plot corresponds to a single animal. Deformation grids at the extremes of the two axes display differences to the mean shape of all individuals.

(TIF)

S4 Fig. The effect of *nhr-1* mutations on mouth morphology is epistatic to that of *nhr-40* mutations. The mouth of the *nhr-1 loss-of-function* mutant, *nhr-40 null* mutant and the double *nhr-40 loss-of-function nhr-1 loss-of-function* mutant in two focal planes. D = dorsal, V = ventral, A = anterior, P = posterior.

(TIF)

S1 Table. Candidate substitutions. List of non-synonymous and nonsense substitutions within the candidate region on chromosome X identified through the bulked segregant analysis of the suppressor of *nhr-40(tu505)*.

(XLSX)

S2 Table. Description of transgenic constructs. Promoters include 5' UTRs and may include coding exons, and introns. See [S1 Data](#) for the sequences of the listed elements. CDS = coding sequence, UTR = untranslated sequence.

(XLSX)

S1 Data. FASTA file of nucleotide sequences used to create transgenic constructs.

(FAS)

S2 Data. Phylogenetic trees from [Fig 5](#) and alignments used to generate them.

(ZIP)

S3 Data. Data used to build the graphs in [Fig 2B](#).

(CSV)

S4 Data. Data used to build the graphs in [Fig 3B](#).

(TXT)

S5 Data. Landmark coordinates used for the analysis shown in [S3 Fig](#).

(TXT)

Acknowledgments

We are grateful to Gabi Eberhardt and Tobias Loschko for their assistance with the mutagenesis screens and Jürgen Berger for taking the SEM image of the *P. pacificus* mouth. We thank Dr. Michael Werner, Dr. Neel Prabh, Dr. Adrian Streit, and Metta Riebesell for the discussion.

Author Contributions

Conceptualization: Ralf J. Sommer.

Data curation: Bogdan Sieriebriennikov, Christian Rödelsperger, Ralf J. Sommer.

Formal analysis: Bogdan Sieriebriennikov, Christian Rödelsperger.

Funding acquisition: Ralf J. Sommer.

Investigation: Bogdan Sieriebriennikov, Shuai Sun, James W. Lightfoot, Hanh Witte, Eduardo Moreno.

Methodology: Bogdan Sieriebriennikov, Shuai Sun, James W. Lightfoot, Hanh Witte, Christian Rödelsperger.

Project administration: Ralf J. Sommer.

Resources: Hanh Witte.

Software: Bogdan Sieriebriennikov, Christian Rödelsperger.

Supervision: Ralf J. Sommer.

Validation: Shuai Sun, James W. Lightfoot, Eduardo Moreno.

Visualization: Bogdan Sieriebriennikov.

Writing – original draft: Bogdan Sieriebriennikov, Ralf J. Sommer.

Writing – review & editing: Bogdan Sieriebriennikov, James W. Lightfoot, Ralf J. Sommer.

References

1. West-Eberhard MJ. Developmental plasticity and evolution. Oxford; New York: Oxford University Press; 2003.
2. Corona M, Libbrecht R, Wheeler DE. Molecular mechanisms of phenotypic plasticity in social insects. *Curr Opin Insect Sci.* 2016; 13: 55–60. <https://doi.org/10.1016/j.cois.2015.12.003> PMID: 27436553
3. Nicotra AB, Atkin OK, Bonser SP, Davidson AM, Finnegan EJ, Mathesius U, et al. Plant phenotypic plasticity in a changing climate. *Trends Plant Sci.* 2010; 15: 684–692. <https://doi.org/10.1016/j.tplants.2010.09.008> PMID: 20970368
4. Seebacher F, White CR, Franklin CE. Physiological plasticity increases resilience of ectothermic animals to climate change. *Nat Clim Chang.* 2014; 5: 61.
5. Charmantier A, McCleery RH, Cole LR, Perrins C, Kruuk LEB, Sheldon BC. Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science.* 2008; 320: 800–803. <https://doi.org/10.1126/science.1157174> PMID: 18467590
6. Oostra V, Saastamoinen M, Zwaan BJ, Wheat CW. Strong phenotypic plasticity limits potential for evolutionary responses to climate change. *Nat Commun.* 2018; 9: 1005. <https://doi.org/10.1038/s41467-018-03384-9> PMID: 29520061
7. Susoy V, Ragsdale EJ, Kanzaki N, Sommer RJ. Rapid diversification associated with a macroevolutionary pulse of developmental plasticity. *Elife.* 2015; 4: e05463.
8. West-Eberhard MJ. Developmental plasticity and the origin of species differences. *Proc Natl Acad Sci U S A.* 2005; 102 Suppl 1: 6543–6549.
9. Corl A, Bi K, Luke C, Challa AS, Stern AJ, Sinervo B, et al. The Genetic Basis of Adaptation following Plastic Changes in Coloration in a Novel Environment. *Curr Biol.* 2018; 28: 2970–2977.e7. <https://doi.org/10.1016/j.cub.2018.06.075> PMID: 30197088
10. Wund MA, Baker JA, Clancy B, Golub JL, Foster SA. A test of the “flexible stem” model of evolution: ancestral plasticity, genetic accommodation, and morphological divergence in the threespine stickleback radiation. *Am Nat.* 2008; 172: 449–462. <https://doi.org/10.1086/590966> PMID: 18729721
11. Levis NA, Pfennig DW. Phenotypic plasticity, canalization, and the origins of novelty: Evidence and mechanisms from amphibians. *Semin Cell Dev Biol.* 2019; 88: 80–90. <https://doi.org/10.1016/j.semdb.2018.01.012> PMID: 29408711
12. Ghalambor CK, Hoke KL, Ruell EW, Fischer EK, Reznick DN, Hughes KA. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature.* 2015; 525: 372–375. <https://doi.org/10.1038/nature15256> PMID: 26331546
13. Sommer RJ, Dardiry M, Lenuzzi M, Namdeo S, Renahan T, Sieriebriennikov B, et al. The genetics of phenotypic plasticity in nematode feeding structures. *Open Biol.* 2017; 7: 160332. <https://doi.org/10.1098/rsob.160332> PMID: 28298309

14. Projecto-Garcia J, Biddle JF, Ragsdale EJ. Decoding the architecture and origins of mechanisms for developmental polyphenism. *Curr Opin Genet Dev*. 2017; 47: 1–8. <https://doi.org/10.1016/j.gde.2017.07.015> PMID: 28810163
15. Opachaloemphan C, Yan H, Leibholz A, Desplan C, Reinberg D. Recent Advances in Behavioral (Epi) Genetics in Eusocial Insects. *Annu Rev Genet*. 2018; 52: 489–510. <https://doi.org/10.1146/annurev-genet-120116-024456> PMID: 30208294
16. Moczek AP, Sultan S, Foster S, Ledón-Rettig C, Dworkin I, Nijhout HF, et al. The role of developmental plasticity in evolutionary innovation. *Proc Biol Sci*. 2011; 278: 2705–2713. <https://doi.org/10.1098/rspb.2011.0971> PMID: 21676977
17. Bento G, Ogawa A, Sommer RJ. Co-option of the hormone-signalling module dafachronic acid-DAF-12 in nematode evolution. *Nature*. 2010; 466: 494–499. <https://doi.org/10.1038/nature09164> PMID: 20592728
18. Ragsdale EJ, Müller MR, Rödelsperger C, Sommer RJ. A Developmental Switch Coupled to the Evolution of Plasticity Acts through a Sulfatase. *Cell*. 2013; 155: 922–933. <https://doi.org/10.1016/j.cell.2013.09.054> PMID: 24209628
19. Werner MS, Sieriebriennikov B, Loschko T, Namdeo S, Lenuzzi M, Dardiry M, et al. Environmental influence on *Pristionchus pacificus* mouth form through different culture methods. *Sci Rep*. 2017; 7: 7207. <https://doi.org/10.1038/s41598-017-07455-7> PMID: 28775277
20. Bose N, Ogawa A, von Reuss SH, Yim JJ, Ragsdale EJ, Sommer RJ, et al. Complex small-molecule architectures regulate phenotypic plasticity in a nematode. *Angew Chem Int Ed Engl*. 2012; 51: 12438–12443. <https://doi.org/10.1002/anie.201206797> PMID: 23161728
21. Werner MS, Claaßen MH, Renahan T, Dardiry M, Sommer RJ. Adult Influence on Juvenile Phenotypes by Stage-Specific Pheromone Production. *iScience*. 2018; 10: 123–134. <https://doi.org/10.1016/j.isci.2018.11.027> PMID: 30513394
22. Sieriebriennikov B, Prabh N, Dardiry M, Witte H, Röseler W, Kieninger MR, et al. A Developmental Switch Generating Phenotypic Plasticity Is Part of a Conserved Multi-gene Locus. *Cell Rep*. 2018; 23: 2835–2843.e4. <https://doi.org/10.1016/j.celrep.2018.05.008> PMID: 29874571
23. Namdeo S, Moreno E, Rödelsperger C, Baskaran P, Witte H, Sommer RJ. Two independent sulfation processes regulate mouth-form plasticity in the nematode *Pristionchus pacificus*. *Development*. 2018; 145: dev166272. <https://doi.org/10.1242/dev.166272> PMID: 29967123
24. Bui LT, Ivers NA, Ragsdale EJ. A sulfotransferase dosage-dependently regulates mouthpart polyphenism in the nematode *Pristionchus pacificus*. *Nat Commun*. 2018; 9: 4119. <https://doi.org/10.1038/s41467-018-05612-8> PMID: 30297689
25. Bui LT, Ragsdale EJ. Multiple plasticity regulators reveal targets specifying an induced predatory form in nematodes. *Molecular Biology and Evolution*. 2019. <https://doi.org/10.1093/molbev/msz171> PMID: 31364718
26. Serobyán V, Xiao H, Namdeo S, Rödelsperger C, Sieriebriennikov B, Witte H, et al. Chromatin remodeling and antisense-mediated up-regulation of the developmental switch gene *eud-1* control predatory feeding plasticity. *Nat Commun*. 2016; 7: 12337. <https://doi.org/10.1038/ncomms12337> PMID: 27487725
27. Kieninger MR, Ivers NA, Rödelsperger C, Markov GV, Sommer RJ, Ragsdale EJ. 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: 2174–2179. <https://doi.org/10.1016/j.cub.2016.06.018> PMID: 27451902
28. Sluder AE, Mathews SW, Hough D, Yin VP, Maina CV. The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. *Genome Res*. 1999; 9: 103–120. PMID: 10022975
29. Evans RM, Mangelsdorf DJ. Nuclear Receptors, RXR, and the Big Bang. *Cell*. 2014; 157: 255–266. <https://doi.org/10.1016/j.cell.2014.03.012> PMID: 24679540
30. Brožová E, Simecková K, Kostrouch Z, Rall JE, Kostrouchová M. NHR-40, a *Caenorhabditis elegans* supplementary nuclear receptor, regulates embryonic and early larval development. *Mech Dev*. 2006; 123: 689–701. <https://doi.org/10.1016/j.mod.2006.06.006> PMID: 16920335
31. Crews ST, Pearson JC. Transcriptional autoregulation in development. *Curr Biol*. 2009; 19: R241–6. <https://doi.org/10.1016/j.cub.2009.01.015> PMID: 19321138
32. Mangan S, Alon U. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A*. 2003; 100: 11980–11985. <https://doi.org/10.1073/pnas.2133841100> PMID: 14530388
33. Macneil LT, Walhout AJM. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res*. 2011; 21: 645–657. <https://doi.org/10.1101/gr.097378.109> PMID: 21324878

34. Bulcha JT, Giese GE, Ali MZ, Lee Y-U, Walker MD, Holdorf AD, et al. A Persistence Detector for Metabolic Network Rewiring in an Animal. *Cell Rep*. 2019; 26: 460–468.e4. <https://doi.org/10.1016/j.celrep.2018.12.064> PMID: 30625328
35. Abdusselamoglu MD, Eroglu E, Burkard TR, Knoblich JA. The transcription factor odd-paired regulates temporal identity in transit-amplifying neural progenitors via an incoherent feed-forward loop. *Elife*. 2019; 8: e46566. <https://doi.org/10.7554/eLife.46566> PMID: 31329099
36. Taylor-Teeple M, Lin L, de Lucas M, Turco G, Toal TW, Gaudinier A, et al. An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. *Nature*. 2015; 517: 571–575. <https://doi.org/10.1038/nature14099> PMID: 25533953
37. Jaumouillé E, Machado Almeida P, Stähli P, Koch R, Nagoshi E. Transcriptional regulation via nuclear receptor crosstalk required for the *Drosophila* circadian clock. *Curr Biol*. 2015; 25: 1502–1508. <https://doi.org/10.1016/j.cub.2015.04.017> PMID: 26004759
38. Anbalagan M, Huderson B, Murphy L, Rowan BG. Post-translational modifications of nuclear receptors and human disease. *Nucl Recept Signal*. 2012; 10: e001. <https://doi.org/10.1621/nrs.10001> PMID: 22438791
39. Seroby V, Ragsdale EJ, Müller MR, Sommer RJ. Feeding plasticity in the nematode *Pristionchus pacificus* is influenced by sex and social context and is linked to developmental speed. *Evol Dev*. 2013; 15: 161–170. <https://doi.org/10.1111/ede.12030> PMID: 23607300
40. Hirschmann H. Über das Vorkommen zweier Mundhöhlentypen bei *Diplogaster lheritieri* Maupas und *Diplogaster biformis* n. sp. und die Entstehung dieser hermaphroditischen Art aus *Diplogaster lheritieri*. *Zool Jb, Abt System, Ökol u Geogr*. 1951; 80: 132–170.
41. Reinke V, Gil IS, Ward S, Kazmer K. Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development*. 2004; 131: 311–323. <https://doi.org/10.1242/dev.00914> PMID: 14668411
42. Gomis-Rüth FX, Trillo-Muyo S, Stöcker W. Functional and structural insights into astacin metalloproteinases. *Biol Chem*. 2012; 393: 1027–1041. <https://doi.org/10.1515/hsz-2012-0149> PMID: 23092796
43. Novelli J, Ahmed S, Hodgkin J. 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*. 2004; 168: 1259–1273. <https://doi.org/10.1534/genetics.104.027953> PMID: 15579684
44. Park J-O, Pan J, Möhrlen F, Schupp M-O, Johnsen R, Baillie DL, et al. Characterization of the astacin family of metalloproteinases in *C. elegans*. *BMC Dev Biol*. 2010; 10: 14. <https://doi.org/10.1186/1471-213X-10-14> PMID: 20109220
45. Li Y, Zhao Y, Su M, Glover K, Chakravarthy S, Colbert CL, et al. Structural insights into the interaction of the conserved mammalian proteins GAPR-1 and Beclin 1, a key autophagy protein. *Acta Crystallogr D Struct Biol*. 2017; 73: 775–792. <https://doi.org/10.1107/S2059798317011822> PMID: 28876241
46. Darwiche R, Kelleher A, Hudspeth EM, Schneider R, Asojo OA. Structural and functional characterization of the CAP domain of pathogen-related yeast 1 (Pry1) protein. *Sci Rep*. 2016; 6: 28838. <https://doi.org/10.1038/srep28838> PMID: 27344972
47. Choudhary V, Schneider R. Pathogen-Related Yeast (PRY) proteins and members of the CAP superfamily are secreted sterol-binding proteins. *Proc Natl Acad Sci U S A*. 2012; 109: 16882–16887. <https://doi.org/10.1073/pnas.1209086109> PMID: 23027975
48. Gibbs GM, Roelants K, O'Bryan MK. The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins—roles in reproduction, cancer, and immune defense. *Endocr Rev*. 2008; 29: 865–897. <https://doi.org/10.1210/er.2008-0032> PMID: 18824526
49. Henrissat B, Callebaut I, Fabrega S, Lehn P, Moron JP, Davies G. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci U S A*. 1995; 92: 7090–7094. <https://doi.org/10.1073/pnas.92.15.7090> PMID: 7624375
50. Lints R, Hall DH. The cuticle. *WormAtlas*. 2009.
51. Tokareva O, Jacobsen M, Buehler M, Wong J, Kaplan DL. Structure-function-property-design interplay in biopolymers: spider silk. *Acta Biomater*. 2014; 10: 1612–1626. <https://doi.org/10.1016/j.actbio.2013.08.020> PMID: 23962644
52. Guan J, Vollrath F, Porter D. Two mechanisms for supercontraction in *Nephila* spider dragline silk. *Biomacromolecules*. 2011; 12: 4030–4035. <https://doi.org/10.1021/bm201032v> PMID: 21951163
53. Riebesell M, Sommer RJ. Three-dimensional reconstruction of the pharyngeal gland cells in the predatory nematode *Pristionchus pacificus*. *Journal of Morphology*. 2017; 278: 1656–1666. <https://doi.org/10.1002/jmor.20739> PMID: 28898441
54. van Megen H, van den Elsen S, Holterman M, Karssen G, Mooyman P, Bongers T, et al. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology*. 2009; 11: 927–950.

55. Altun ZF, Hall DH. Alimentary System, Pharynx. WormAtlas. 2009.
56. Zhang YC, Baldwin JG. Ultrastructure of the postcorpus of the esophagus of *Teratocephalus lirellus* (Teratocephalida) and its use for interpreting character evolution in Secernentea (Nematoda). *Can J Zool.* 2001; 79: 16–25.
57. Zhang YC, Baldwin JG. Ultrastructure of the post-corpus of *Zeldia punctata* (Cephalobina) for analysis of the evolutionary framework of nematodes related to *Caenorhabditis elegans* (Rhabditina). *Proceedings of the Royal Society of London Series B: Biological Sciences.* 2000; 267: 1229–1238. <https://doi.org/10.1098/rspb.2000.1132> PMID: 10902689
58. Chiang J- TA, Steciuk M, Shtonda B, Avery L. Evolution of pharyngeal behaviors and neuronal functions in free-living soil nematodes. *J Exp Biol.* 2006; 209: 1859–1873. <https://doi.org/10.1242/jeb.02165> PMID: 16651552
59. Zhang YC, Baldwin JG. Ultrastructure of the Esophagus of *Diplenteron* sp. (Diplogasterida) to Test Hypotheses of Homology with Rhabditida and Tylenchida. *J Nematol.* 1999; 31: 1–19. PMID: 19270870
60. Prabh N, Roeseler W, Witte H, Eberhardt G, Sommer RJ, Rödelsperger C. Deep taxon sampling reveals the evolutionary dynamics of novel gene families in *Pristionchus* nematodes. *Genome Res.* 2018; 28: 1664–1674. <https://doi.org/10.1101/gr.234971.118> PMID: 30232197
61. Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, Dinkelacker I, et al. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat Genet.* 2008; 40: 1193–1198. <https://doi.org/10.1038/ng.227> PMID: 18806794
62. Markov GV, Baskaran P, Sommer RJ. The same or not the same: lineage-specific gene expansions and homology relationships in multigene families in nematodes. *J Mol Evol.* 2015; 80: 18–36. <https://doi.org/10.1007/s00239-014-9651-y> PMID: 25323991
63. Sieriebriennikov B, Markov GV, Witte H, Sommer RJ. The Role of DAF-21/Hsp90 in Mouth-Form Plasticity in *Pristionchus pacificus*. *Mol Biol Evol.* 2017; 34: 1644–1653. <https://doi.org/10.1093/molbev/msx106> PMID: 28333289
64. Antebi A, Yeh WH, Tait D, Hedgecock EM, Riddle DL. *daf-12* encodes a nuclear receptor that regulates the dauer diapause and developmental age in *C. elegans*. *Genes Dev.* 2000; 14: 1512–1527. PMID: 10859169
65. Burton NO, Dwivedi VK, Burkhardt KB, Kaplan REW, Baugh LR, Horvitz HR. Neurohormonal signaling via a sulfotransferase antagonizes insulin-like signaling to regulate a *Caenorhabditis elegans* stress response. *Nat Commun.* 2018; 9: 5152. <https://doi.org/10.1038/s41467-018-07640-w> PMID: 30514845
66. Rödelsperger C, Streit A, Sommer RJ. *Structure, Function and Evolution of The Nematode Genome.* eLS. Chichester, UK: John Wiley & Sons, Ltd; 2013.
67. Mitreva M, Blaxter ML, Bird DM, McCarter JP. Comparative genomics of nematodes. *Trends Genet.* 2005; 21: 573–581. <https://doi.org/10.1016/j.tig.2005.08.003> PMID: 16099532
68. Manning G, Plowman GD, Hunter T, Sudarsanam S. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci.* 2002; 27: 514–520. [https://doi.org/10.1016/s0968-0004\(02\)02179-5](https://doi.org/10.1016/s0968-0004(02)02179-5) PMID: 12368087
69. Goodrich LV, Johnson RL, Milenkovic L, McMahon JA, Scott MP. Conservation of the *hedgehog/patched* signaling pathway from flies to mice: induction of a mouse *patched* gene by Hedgehog. *Genes Dev.* 1996; 10: 301–312. <https://doi.org/10.1101/gad.10.3.301> PMID: 8595881
70. Logan CY, Nusse R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol.* 2004; 20: 781–810. <https://doi.org/10.1146/annurev.cellbio.20.010403.113126> PMID: 15473860
71. Stiernagle T. Maintenance of *C. elegans* (February 11, 2006). In: *The C. elegans research community*, editor. WormBook. 2016.
72. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 2012; 9: 676–682. <https://doi.org/10.1038/nmeth.2019> PMID: 22743772
73. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2016.
74. Adams DC, Otarola-Castillo E. geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods Ecol Evol.* 2013; 4: 393–399.
75. Witte H, Moreno E, Rödelsperger C, Kim J, Kim JS, Streit A, et al. Gene inactivation using the CRISPR/Cas9 system in the nematode *Pristionchus pacificus*. *Dev Genes Evol.* 2015; 225: 55–62. <https://doi.org/10.1007/s00427-014-0486-8> PMID: 25548084
76. Lightfoot JW, Wilecki M, Rödelsperger C, Moreno E, Susoy V, Witte H, et al. Small peptide-mediated self-recognition prevents cannibalism in predatory nematodes. *Science.* 2019; 364: 86–89. <https://doi.org/10.1126/science.aav9856> PMID: 30948551

77. Zhou S, Fu X, Pei P, Kucka M, Liu J, Tang L, et al. Characterization of a non-sexual population of *Strongyloides stercoralis* with hybrid 18S rDNA haplotypes in Guangxi, Southern China. *PLoS Negl Trop Dis*. 2019; 13: e0007396. <https://doi.org/10.1371/journal.pntd.0007396> PMID: 31059500
78. Rödelsperger C, Meyer JM, Prabh N, Lanz C, Bemm F, Sommer RJ. Single-Molecule Sequencing Reveals the Chromosome-Scale Genomic Architecture of the Nematode Model Organism *Pristionchus pacificus*. *Cell Rep*. 2017; 21: 834–844. <https://doi.org/10.1016/j.celrep.2017.09.077> PMID: 29045848
79. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9: 357–359. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
80. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011; 29: 24–26. <https://doi.org/10.1038/nbt.1754> PMID: 21221095
81. Pires-daSilva A. *Pristionchus pacificus* protocols (March 14, 2013). In: The *C. elegans* research community, editor. *WormBook*. 2013.
82. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
83. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
84. Rae R, Witte H, Rödelsperger C, Sommer RJ. The importance of being regular: *Caenorhabditis elegans* and *Pristionchus pacificus* defecation mutants are hypersusceptible to bacterial pathogens. *Int J Parasitol*. 2012; 42: 747–753. <https://doi.org/10.1016/j.ijpara.2012.05.005> PMID: 22705203
85. Schlager B, Wang X, Braach G, Sommer RJ. Molecular cloning of a dominant roller mutant and establishment of DNA-mediated transformation in the nematode *Pristionchus pacificus*. *Genesis*. 2009; 47: 300–304. <https://doi.org/10.1002/dvg.20499> PMID: 19298013
86. OpenWetWare contributors. Gibson Assembly. In: OpenWetWare [Internet]. 7 May 2018 [cited 25 Jul 2018]. Available: https://openwetware.org/mediawiki/index.php?title=Gibson_Assembly&oldid=1043969
87. Bettinger JC, Lee K, Rougvie AE. Stage-specific accumulation of the terminal differentiation factor LIN-29 during *Caenorhabditis elegans* development. *Development*. 1996; 122: 2517–2527. PMID: 8756296
88. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; 29: 15–21. <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886
89. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004; 5: R80. <https://doi.org/10.1186/gb-2004-5-10-r80> PMID: 15461798
90. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15: 550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
91. Fox J, Weisberg S. *An R Companion to Applied Regression*. Thousand Oaks CA: Sage; 2019. Available: <https://socialsciences.mcmaster.ca/fox/Books/Companion/>
92. Nielsen H. Predicting Secretory Proteins with SignalP. *Methods Mol Biol*. 2017; 1611: 59–73. https://doi.org/10.1007/978-1-4939-7015-5_6 PMID: 28451972
93. Möhrle F, Hutter H, Zwilling R. The astacin protein family in *Caenorhabditis elegans*. *Eur J Biochem*. 2003; 270: 4909–4920. <https://doi.org/10.1046/j.1432-1033.2003.03891.x> PMID: 14653817
94. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016; 44: D279–85. <https://doi.org/10.1093/nar/gkv1344> PMID: 26673716
95. Sinha A, Langnick C, Sommer RJ, Dieterich C. 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: 1386–1397. <https://doi.org/10.1261/rna.041954.113> PMID: 25015138
96. Werner MS, Sieriebriennikov B, Prabh N, Loschko T, Lanz C, Sommer RJ. Young genes have distinct gene structure, epigenetic profiles, and transcriptional regulation. *Genome Res*. 2018; 28: 1675–1687. <https://doi.org/10.1101/gr.234872.118> PMID: 30232198
97. Rödelsperger C, Athanasouli M, Lenuzzi M, Theska T, Sun S, Dardiry M, Wighard S, Hu W, Sharma DR, Han Z. Crowdsourcing and the feasibility of manual gene annotation: A pilot study in the nematode *Pristionchus pacificus*. *Sci. Rep*. 2019; 9: 18789. <https://doi.org/10.1038/s41598-019-55359-5> PMID: 31827189
98. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30: 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690

99. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> PMID: [24451623](https://pubmed.ncbi.nlm.nih.gov/24451623/)

Improving Transgenesis Efficiency and CRISPR-Associated Tools Through Codon Optimization and Native Intron Addition in *Pristionchus* Nematodes

Ziduan Han,¹ Wen-Sui Lo,¹ James W. Lightfoot, Hanh Witte, Shuai Sun, and Ralf J. Sommer²

Max Planck Institute for Developmental Biology, Tuebingen 72076, Germany

ORCID IDs: 0000-0001-8341-0678 (Z.H.); 0000-0002-2438-0015 (W.-S.L.); 0000-0002-5835-2135 (J.W.L.); 0000-0003-1503-7749 (R.J.S.)

ABSTRACT A lack of appropriate molecular tools is one obstacle that prevents in-depth mechanistic studies in many organisms. Transgenesis, clustered regularly interspaced short palindromic repeats (CRISPR)-associated engineering, and related tools are fundamental in the modern life sciences, but their applications are still limited to a few model organisms. In the phylum Nematoda, transgenesis can only be performed in a handful of species other than *Caenorhabditis elegans*, and additionally, other species suffer from significantly lower transgenesis efficiencies. We hypothesized that this may in part be due to incompatibilities of transgenes in the recipient organisms. Therefore, we investigated the genomic features of 10 nematode species from three of the major clades representing all different lifestyles. We found that these species show drastically different codon usage bias and intron composition. With these findings, we used the species *Pristionchus pacificus* as a proof of concept for codon optimization and native intron addition. Indeed, we were able to significantly improve transgenesis efficiency, a principle that may be usable in other nematode species. In addition, with the improved transgenes, we developed a fluorescent co-injection marker in *P. pacificus* for the detection of CRISPR-edited individuals, which helps considerably to reduce associated time and costs.

KEYWORDS nematodes; *C. elegans*; *P. pacificus*; transgenesis; CRISPR editing; codon usage bias; intron-mediated enhancement; parasitic nematodes

The utilization of transgenes has proven fundamental to many aspects of molecular biology and for functional genomic studies (Rubin and Spradling 1982; Mello *et al.* 1991; Chalfie *et al.* 1994; Clough and Bent 1998; Hutter 2012). For instance, easily applied and efficient transgenic methods have been instrumental in furthering our understanding of biological pathways and dissecting associated phenotypes. Additionally, it has facilitated the visualization of gene expression patterns and protein localization through the usage of fluorescent proteins such as GFP in a swathe of organisms (Chalfie *et al.* 1994). However, limiting factors for the successful establishment of transgenesis in an organism

are the differing regulatory strategies and mechanisms found between species. In accordance with this and despite their ubiquitous usage, efficient transgenesis tools are frequently restricted to canonical model organisms.

Gene expression, including transgene expression, is regulated by a multitude of factors, including at the transcriptional and translational levels. One such regulatory mechanism is through codon usage bias (CUB). Here, the degenerate nature of the nucleotide triplet code ensures that each amino acid can be encoded by several synonymous codons, with the exception of the amino acids methionine and tryptophan (Sharp and Li 1987). Correspondingly, organism genomes show their own distinct usage of the code. This codon bias is more pronounced in genes with elevated expression levels. Specifically, highly expressed genes strongly favor a specific set of codons with the favored codons contributing to a more efficient translation process through faster ribosome elongation (Duret and Mouchiroud 1999; Plotkin and Kudla 2011). Further, artificial manipulation of the CUB can also alter gene expression dramatically (Redemann *et al.* 2011). In addition to CUB, regulatory regions of a gene are also thought to be

Copyright © 2020 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.120.303785>

Manuscript received September 9, 2020; accepted for publication October 14, 2020; published Early Online October 15, 2020.

Available freely online through the author-supported open access option.

Supplemental material available at figshare: <https://doi.org/10.25386/genetics.13090322>.

¹These authors contributed equally to this work.

²Corresponding author: Max Planck Institute for Developmental Biology, Department for Integrative Evolutionary Biology, Max-Planck-Ring 9, 72076 Tübingen, Germany. E-mail: ralf.sommer@tuebingen.mpg.de

crucial for transcriptional control as evidence suggests a relationship between the exon–intron structure of a gene and its expression through a process termed “intron-mediated enhancement” (IME). Here, intron density positively correlates with both the level and extent of a gene’s expression (Castillo-Davis *et al.* 2002). As such, these phenomena have been exploited for the enhancement of molecular tools including improving transgenesis in a number of well-studied model organisms (Brinster *et al.* 1988; Bischof *et al.* 2007).

One such organism is the nematode *Caenorhabditis elegans*, where an abundance of molecular tools, including transgenesis, are available and its CUB and exon–intron structures are well characterized (Ragle *et al.* 2015). In particular, IME in *C. elegans* is strongly influenced by the position, number, and sequence of introns, and introns positioned near the 5′ end of a gene shows the greatest contribution to this effect (Okkema *et al.* 1993; Crane *et al.* 2019). Further, replacing native codons with favored codons increases the translation level of a protein (Redemann *et al.* 2011). However, *C. elegans* is far from the only nematode of significance in the phylum, with an array of parasitic nematodes of both animals and plants, as well as other free-living nematodes, also now frequently used for research. Despite this, transgenesis has only been successfully applied to a few nematode species outside of the genus *Caenorhabditis* (Higazi *et al.* 2002; Li *et al.* 2006; Schlager *et al.* 2009; Lok 2012), with problems arising due to efficient delivery of DNA materials to the gonad (Evans 2006) and compatibility of the DNA to the endogenous genetic machinery of the recipient. Therefore, despite recent advancements (Adams *et al.* 2019), low efficiency is still the bottleneck for most transgenic experiments in other nematode species.

In addition to *C. elegans*, another distantly related free-living nematode frequently used for research is *Pristionchus pacificus* (Sommer *et al.* 1996). This nematode has been established as a model system to study evolutionary developmental biology and, more specifically, the evolution of novelty. This is due to the nature of its mouth structure, which is phenotypically plastic and demonstrates two distinct variants. One morph exhibits two teeth while the other contains only a single tooth. The genetic network behind this developmental decision has been extensively studied and is heavily influenced by the nematode’s environment (Ragsdale *et al.* 2013; Kieninger *et al.* 2016; Bui *et al.* 2018; Sieriebriennikov *et al.* 2020). The presence of teeth in *P. pacificus* facilitates an additional behavior as they are capable of predated the larvae of other nematodes. Here, it has been observed that the mouth-form dimorphism strongly correlates with the predation behavior, as only the morphs possessing two teeth are active predators, whereas the single-toothed morphs are strict bacterial feeders (Wilecki *et al.* 2015; Moreno *et al.* 2019; Akduman *et al.* 2020). Furthermore, the predatory behavior coincides with the existence of a self-recognition system (Lightfoot *et al.* 2019) and environmental responses distinct from *C. elegans* (Hong and Sommer 2006; Moreno *et al.* 2016, 2017).

Outside of *C. elegans*, *P. pacificus* is arguably the most advanced nematode system in terms of the availability of

molecular tools (Schlager *et al.* 2009; Witte *et al.* 2015; Okumura *et al.* 2017; Loer *et al.* 2019). However, previous methodologies resulted in low efficiencies of *P. pacificus* transgenics, with on average one to three F1 Roller(s) per 40 injected P0s (Schlager *et al.* 2009). Thus, *P. pacificus* suffers from a much less efficient transgenesis system compared with *C. elegans* for several potential reasons. First, it relies on the formation of complex arrays, which incorporate transgene DNA, genomic DNA fragments that must come from *P. pacificus* itself, and a co-injection marker, to be carried as heritable chromosome fragments. Second, the current versions of fluorescent proteins utilized in *P. pacificus* (Schlager *et al.* 2009) have not been adapted to its specific CUB and no attempts have yet been made to improve these fluorescent proteins further by investigating any potential IME. Together, these factors likely contribute to the varying degrees of generational transmission observed in *P. pacificus* transgenesis experiments and will likely hinder the successful development of other transgenic techniques including additional fluorescent proteins, calcium imaging, and optogenetics.

In this study, using publicly available data sets, we first computed the CUB and global intron structure in 10 nematode species to investigate the conservation of these factors across the phylum, making use of the most recent genomic and transcriptomic data sets. For this, we selected species living in different ecosystems including parasites of animals and plants. As each nematode species shows a distinct CUB and potential IME, we focused on *P. pacificus* and utilized these factors together with its spliced leaders (SLs), a specific but conserved transcriptional regulatory element in nematodes (Denker *et al.* 2002), to improve the efficiency of transgenesis in this species. Finally, with the improved transgenesis in *P. pacificus*, we established a new method using a fluorescent co-injection marker to identify potential clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-edited candidates, reducing the workload and cost for CRISPR/Cas9 screening.

Materials and Methods

Obtaining genome annotations and transcription profiles

We collected published annotations and transcriptomes of 10 nematode species representing three of the five major nematode clades (Blaxter *et al.* 1998): *C. elegans* (Lee *et al.* 2018; Liu *et al.* 2019) [WormBase web site (<https://wormbase.org>), release WS271 2019], *C. briggsae* (Grün *et al.* 2014), *Haemonchus contortus* (Laing *et al.* 2013), *P. pacificus* (Prabh *et al.* 2018; Rödelberger *et al.* 2019), *P. fissionatus* (Prabh *et al.* 2018; Rödelberger *et al.* 2018), *Strongyloides ratti* (Hunt *et al.* 2016), *Globodera pallida* (Cotton *et al.* 2014), *Bursaphelenchus xylophilus* (Kikuchi *et al.* 2011; Tanaka *et al.* 2019), *Brugia malayi* (Choi *et al.* 2011; Foster *et al.* 2020), and *Ascaris suum* (Wang *et al.* 2011, 2017). To acquire the expression profiles of *P. pacificus* and *P. fissionatus*, we retrieved RNA-sequencing (RNA-seq) data sets of *P. pacificus* and *P. fissionatus* from the Sequence

Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra/>; Supplemental Material, Table S1). We mapped raw reads to the reference genome of each species (Table S1) using Hisat2 (Kim *et al.* 2015) with default parameters, and quantified the numbers of reads mapping to each annotated gene using the package featureCounts (Liao *et al.* 2014). For other species, the gene expression results were directly downloaded from WormBase (Howe *et al.* 2016, 2017). Detailed information of the metadata is summarized in Table S1. For *C. elegans* and *A. suum*, whose annotations included isoform data, only the longest transcripts were used in downstream analyses.

Codon usage computation

To identify the CUBs of genes with different expression levels, the percentage codon usage for each gene was calculated using cusp from EMBOSS suite (Rice *et al.* 2000). We optimized the codon of proteins based on the most preferred codons of genes with high expression levels in *P. pacificus*.

P. pacificus trans-spliced messenger RNA identification

To identify the *P. pacificus* transcripts that contain SLs, we first performed RNA-seq using a ribosomal RNA (rRNA) depletion library. Briefly, total RNA of *P. pacificus* was extracted via Direct-zol RNA Miniprep (Zymo Research) and a RiboZero rRNA Removal Kit (Human/Mouse/Rat; Illumina), and RNA libraries were constructed using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina). Sequencing was carried out on an Illumina HiSeq 3000 sequencer with one-sixth of a lane. We used Trinity (Grabherr *et al.* 2011) for *de novo* transcriptome assembly, and identified the transcripts with SLs by the consensus SL sequences at 5' ends (SL1: TACC CAAGTTTGAG; and SL2: CAGTATCTCAAG) (Guiliano and Blaxter 2006). We used MEME SUITE (Bailey *et al.* 2009) to identify the motifs of 3' sequences of the *trans*-splice sites.

Statistics

We performed the chi-square test to test whether the frequencies of synonymous codons in the most highly expressed genes (11th bin) were deviated from the frequencies of genome-wide synonymous codons. We performed the one-tailed Kolmogorov–Smirnov test to compare the intron length distributions between *C. elegans* and the other species. We calculated the Pearson's correlation coefficient to measure the linear relationship between intron length and gene expression level. We performed the Wilcoxon signed-rank test to test whether genes that contained SLs were different in expression level from genes without SLs.

Plasmid construction and microinjection

The optimized *egl-20p::GFP* and *egl-20p::TurboRFP* (red fluorescent protein) were modified based on a pUC19 backbone from a previous study (Schlager *et al.* 2009). Full sequences of these plasmids in text files can be found in the supplemental materials. Modified GFP and TurboRFP sequences were synthesized from Integrated DNA Technologies (IDT; Coralville, IA) and cloned into the pUC19 backbone using Gibson

Assembly Master Mix (New England Biolabs, Beverly, MA) following the manufacturers' protocols. Plasmids were extracted using the QIAprep Spin Miniprep kit (QIAGEN, Valencia, CA).

Three introns from the rRNA gene *Ppa-rps-1* (gene ID: PPA18896; El_paco annotation_v2) (Rödelsperger *et al.* 2019) were added into the sequence of GFP or turboRFP from 5' to 3' and were roughly evenly spaced (“Fire Lab Vector Kit 1995”): intron 1, gtgagcatttctgtgtgtaatgggggtgtgaaaactcatgggattcctaaccttaatttttcag; intron 2, gtaagtcgtatacattagcgggtgctttacgtgatccggggtttgttttgagagaggatatttataataaataatatttcag; and intron 3, gtgagtgtgtcaaatattaagtacatgaaacttttctcag. For the two-intron codon-optimized *egl-20p::GFP* and *egl-20p::TurboRFP*, intron 1 (the most 5' intron) was removed using a Q5 Site-Directed Mutagenesis kit (New England Biolabs), and both intron 1 and 2 were removed for the one-intron *egl-20p::GFP* and *egl-20p::TurboRFP*.

P. pacificus microinjections were performed following the standard protocol (Schlager *et al.* 2009; Witte *et al.* 2015). Plasmids were diluted to 50 ng/μl for microinjection using TE buffer. Well-fed *P. pacificus* (strain PS312) young hermaphrodites (preferably not carrying any eggs) were used for injections. The injection mix for the co-injection marker-assisted CRISPR/Cas9 editing was modified from those of Witte *et al.* (2015) and Dokshin *et al.* (2018), and the mix contained 0.5 μg/μl Cas9 nuclease (catalog# 1081058; IDT), 0.1 μg/μl *trans*-activating CRISPR RNA (catalog# 1072534; IDT), 0.056 μg/μl guide RNA (CRISPR/Cas9 RNA; IDT), and 0.05 μg/μl co-injecting plasmid. Potential CRISPR-edited alleles were amplified by PCR and sequenced using Sanger sequencing. Alternatively, the PCR amplicons were run on a 4% TBE agarose gel to detect heteroduplex formation (Bhattacharya and Van Meir 2019).

Data availability

The raw sequence data of *P. pacificus* rRNA-depleted RNA-seq have been deposited at the SRA under BioProject identified PRJNA658248. Supplemental material available at figshare: <https://doi.org/10.25386/genetics.13090322>.

Results

Codon usage is divergent among nematode species

To enhance transgene expression in diverse nematode species, we first obtained a comprehensive view of CUB in nematodes. For that, we calculated CUB in 10 species of eight nematode genera, representing three of the five major clades of the phylum Nematoda (Figure 1A and Figure S1; Blaxter *et al.* 1998). For a given amino acid, we found a favored codon in highly expressed genes in every species. The frequencies of different codons in the most highly expressed genes (11th bin) deviate significantly from their genome-wide frequencies (for all comparisons, $0 < P < 4.19 \times 10^{-11}$, chi-square test). There was no clear pattern between CUBs and phylogenetic relationships or lifestyles (free-living or parasitic). However, genome-wide GC content may be

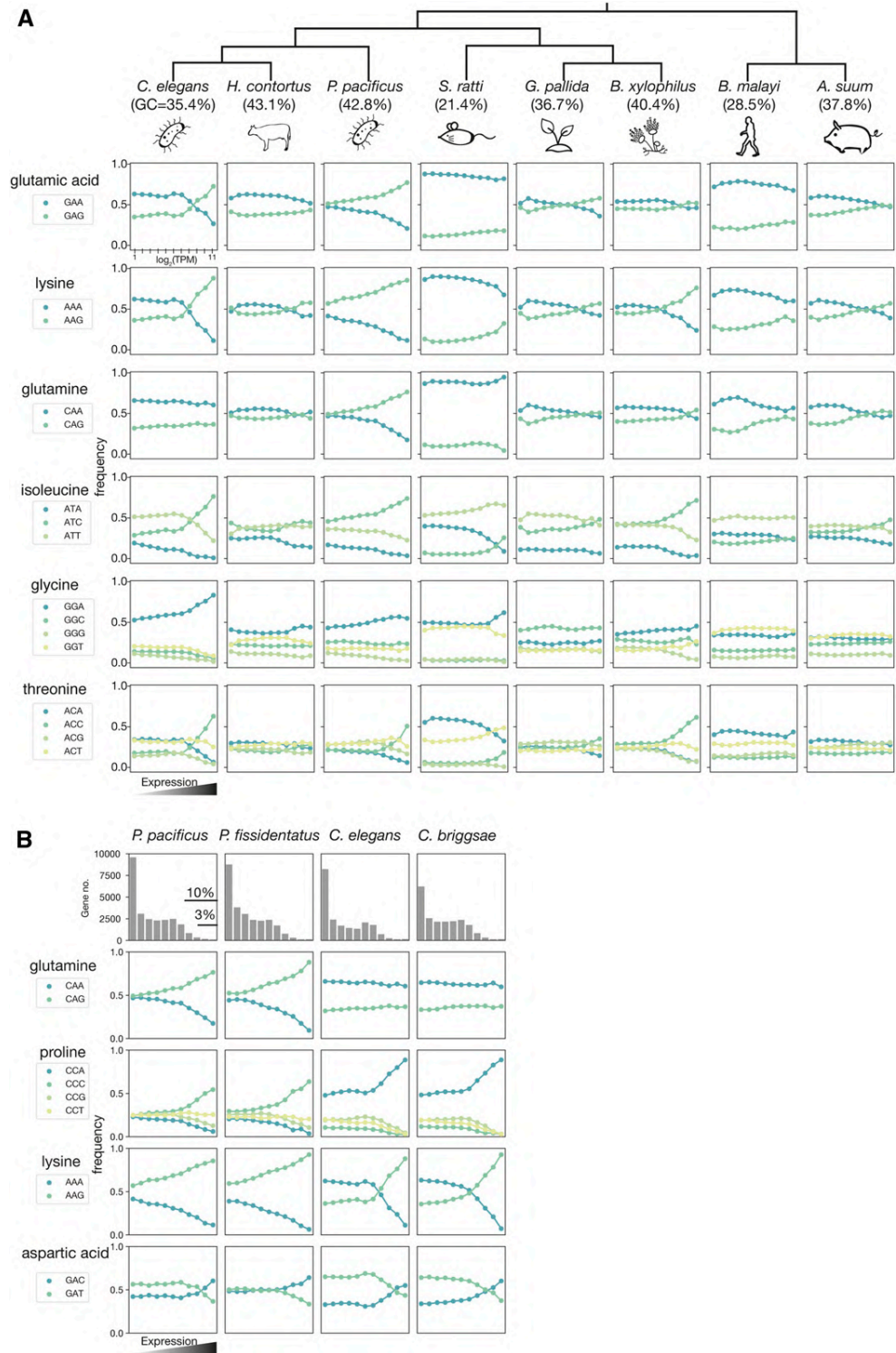


Figure 1 Codon preferences in nematodes species as a function of expression levels. (A) The codon usage bias of *C. elegans*, *H. contortus*, *P. pacificus*, *S. ratti*, *G. pallida*, *B. xylophilus*, *B. malayi*, and *A. suum*. The protein-coding genes are binned based on the transcripts per kilobase million value from expression level low to high with a \log_2 scale into 11 bins (x-axis). The dots represent the average codon usage frequency of a given bin. (B) Gene grouping and codon usage bias for *P. pacificus*, *P. fissionatus*, *C. elegans*, and *C. briggsae*. Figures in the first row show the number of genes (y-axis) grouped from low to high expression with a \log_2 scale into 11 bins (same as A).

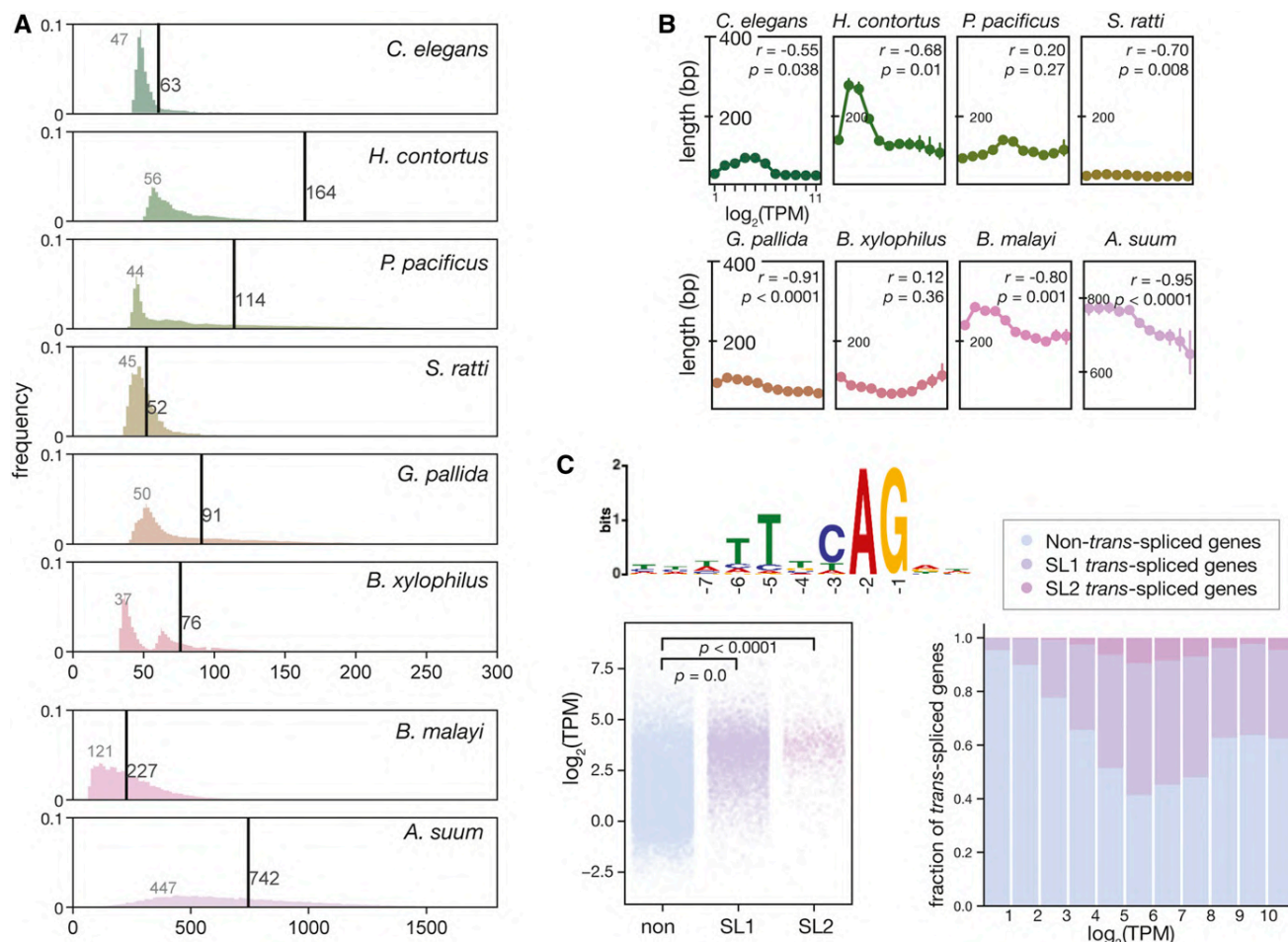


Figure 2 Global intron length distribution in diverse nematodes. (A) The intron length distribution of eight nematode species. Vertical lines indicate the median lengths of introns of each species, while the numbers in gray indicate the modes of intron lengths (bp). (B) Median intron length as a function of the gene expression. The protein-coding genes are binned by expression level from low to high with a \log_2 scale into 11 bins. (C) Elevated expression level of SL1-operated genes. Consensus sequence of the SL1 *trans*-splice sites in *P. pacificus* (top). SL1- and SL2-spliced genes have a higher expression level than those that are not *trans*-spliced (left). The proportion of *trans*-spliced genes is positively associated with expression level (right). SL1, spliced leader 1; TPM, transcripts per kilobase million.

one major factor correlating with the CUB (Mitreva *et al.* 2006). For example, *S. ratti* and *B. malayi* have low-GC-content genomes, and subsequently the codon usage is also biased toward AT-rich codons. In these species, GCA and GCT are more preferred than GCC and GCG for alanine. Intriguingly, species with a similar GC content can still exhibit drastically different patterns in CUB. For example, *P. pacificus* and *H. contortus*, which both have $\sim 43\%$ GC content, show differing codon preferences for coding proline and alanine (Figure S1). Thus, our new analysis of CUB confirms previous studies that codon usage is divergent among nematode species. Note that the species considered here belong to very different nematode taxa and are phylogenetically only distantly related.

Codon usage adaptation is conserved within genera

To study the evolution of codon usage between more closely related nematodes, we focused on two well-studied nematode genera *Caenorhabditis* and *Pristionchus*. *C. elegans* and *P.*

pacificus share a common ancestor around 100 million years ago, and they have a distinct CUB. The most dramatic examples of this can be seen in the amino acids glutamine, glutamic acid, and lysine, where *P. pacificus* and *C. elegans* favor the opposing codons. However, within the genus *Pristionchus* the CUB appears conserved, as in *P. fissidentatus*, a basal species in the *Pristionchus* genus (Rödelsperger *et al.* 2018), and we found it shares a highly similar CUB with that observed in *P. pacificus* (Figure 1B and Figure S2). Similarly, in *Caenorhabditis*, the CUB is conserved between *C. elegans* and *C. briggsae* (Figure S3). This finding strongly suggests that codon usage adaptation evolved more ancestrally than the speciation events within the genera *Pristionchus* and *Caenorhabditis*, and that CUB is conserved between closely related species.

Global intron structure and SL1 frequency shows distinct patterns

As the presence and distributions of introns also contribute to gene regulation (Castillo-Davis *et al.* 2002), we next

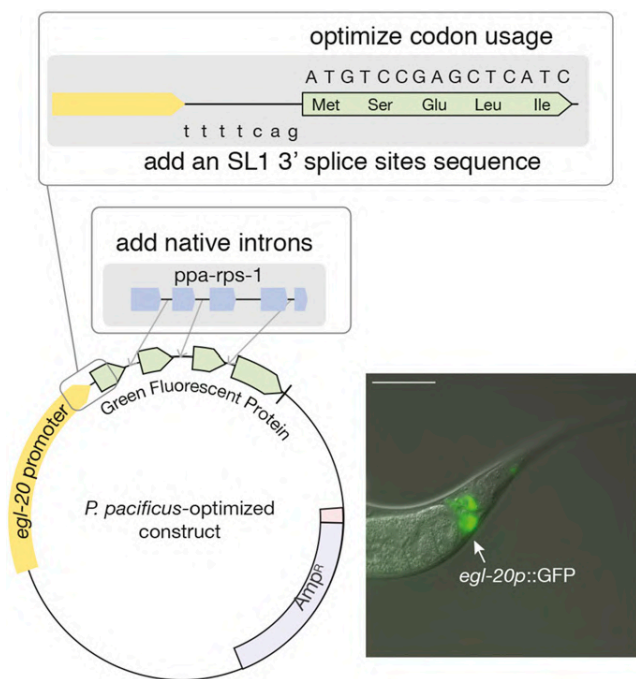


Figure 3 Optimized transgenic plasmids based on *P. pacificus* genomic features. An illustration of the construct structure for codon-optimized and native intron addition in *egl-20p::GFP/turboRFP* (left). An overlay of DIC and GFP image of *egl-20p::GFP* (left). Bar, 50 μm. AMP^R, ampicillin resistance; GFP, green fluorescent protein.

investigated global intron composition to potentially understand the IME of genes across nematodes. Unexpectedly, introns across eight annotated genomes showed distinct features in terms of the general pattern observed (for all comparisons $P = 0.0$, Kolmogorov–Smirnov test) and the median intron size. When intron length was plotted by frequency, a unimodal pattern was detected in *C. elegans*, *H. contortus*, *P. pacificus*, and *S. ratti*, whereas a bimodal pattern was observed in *B. xylophilus* (Figure 2A). Further, in the clade three nematodes (*B. malayi* and *A. suum*) intron length appeared to be much longer and with a wider distribution.

Further analysis of the average intron length revealed that the introns of *C. elegans* have a mode (the most abundant number) of 47 nt and a median of 63 nt in length (Figure 2A), while the distantly related *S. ratti* has even shorter introns with a median of 52 nt and a more homogeneous distribution. In *H. contortus*, *P. pacificus*, *B. xylophilus*, and *G. pallida*, the distribution of intron length shows a greater range compared with *C. elegans*, although an accumulation of introns with a size between 40 and 60 nt is also detectable. When comparing intron size with gene expression level, we found that in *C. elegans* ($r = -0.55$, $P = 0.038$, Pearson's correlation), *H. contortus* ($r = -0.68$, $P = 0.010$, Pearson's correlation), *S. ratti* ($r = -0.70$, $P = 0.008$, Pearson's correlation), *G. pallida* ($r = -0.91$, $P < 0.001$, Pearson's correlation), *B. malayi* ($r = -0.91$, $P = 0.001$, Pearson's correlation), and *A. suum* ($r = -0.95$, $P < 0.001$, Pearson's correlation), the

intron size was negatively correlated with gene expression level. However, this correlation was not observed in the other species (Figure 2B).

Finally, we investigated another gene regulatory element, SL1. Nematodes have a specific *trans*-splicing mechanism at the 5' end of many premature messenger RNAs (mRNAs), which is trimmed and replaced by an SL sequence (Denker *et al.* 2002). This mechanism is thought to increase translation (Yang *et al.* 2017). Using a *P. pacificus* rRNA-depleted RNA-seq library instead of deeply sequenced mRNA-enriched RNA-seq data sets, which are traditionally used in *C. elegans* (Allen *et al.* 2011), we identified a total of 5982 genes in *P. pacificus* that were SL1-operated, and 922 genes that were SL2-operated. These genes have an SL1 3' splice site with a consensus sequence "TTTCAG" (Figure 2C), which is also conserved in *C. elegans* (Yang *et al.* 2017). Globally, higher expression levels were observed in *P. pacificus* genes associated with SL1 compared with genes without splicing leaders ($P = 0.0$, Wilcoxon signed-rank test). Therefore, this suggests that SL1 increases translation in *P. pacificus*, a similar phenomenon to that observed in *C. elegans* (Yang *et al.* 2017). While the published nematode data sets are not sufficient for us to survey the SL1 *trans*-spliced genes of other nematode species, given the fact that the sequences of SLs are conserved among nematodes (Guiliano and Blaxter 2006), the SL *trans*-splicing could be a highly conserved mechanism in the Nematoda phylum.

Optimization of GFP and TurboRFP sequences and increased transgenesis efficiency

With our observations of the large variations in CUB and potential IME regulating gene expression across nematodes, we decided to focus on a single species and attempt to improve its transgenesis efficiency. Therefore, we focused on establishing two fluorescent proteins for use in the free-living nematode *P. pacificus*. These were based on the previously utilized TurboRFP (Schlager *et al.* 2009) and on GFP (Fire Lab Vector Kit 1995), which are commonly utilized across the *C. elegans* community. In *P. pacificus*, TurboRFP has been used to successfully produce transgenic lines; however, this was only at a low transmission efficiency. The GFP previously used in *P. pacificus* was optimized according to *C. elegans*' CUB and hardly generated detectable fluorescence. Therefore, we replaced the codons in these two fluorescent proteins with two sets of codon usages: the CUB found associated with the top 10% most highly expressed genes and with the top 3% most highly expressed genes of *P. pacificus* (Table S2). Alongside this, we also attempted to optimize both fluorescent proteins further through the addition of native introns to increase its transcription. We selected the native introns of the gene *Ppa-rps-1* as it is highly expressed through all life stages and has four relatively short introns. The three shorter introns of *Ppa-rps-1* were added into the reading frame of the codon-optimized GFP and TurboRFP. Finally, we added an SL1 3' splice site sequence immediately upstream of the start codon of both fluorescent proteins (illustrated in Figure 3).

Table 1 Improved transgenesis efficiency using transcriptional reporter constructs with codon optimization and intron addition in *P. pacificus*

Construct	Number of introns	Injected P0s	Number of P0s with fluorescent F1s	Efficiency (%)
<i>egl-20p::GFP</i>	3	49	11	22 ($P = 0.02$)
<i>egl-20p::TurboRFP</i>	3	55	16	29 ($P = 0.003$)
<i>egl-20p::GFP</i>	2	40	4	10 ($P = 0.39$)
<i>egl-20p::GFP</i>	1	12	0	0 ($P = 0.48$)
<i>egl-20p::TurboRFP</i>	1	18	0	0 ($P = 0.34$)
<i>Ppa-prl-1</i> ^a	NA	NA	NA	5

The GFP and TurboRFP sequences were optimized using *P. pacificus* favored codons (from top 10% highly expressed) with addition of native introns.

^a Data from Schlager *et al.* 2009, summarized from over 3000 P0 injections. Chi-square tests were performed between *Ppa-prl-1* and optimized constructs.

In a first set of experiments, we performed all three optimization steps (CUB, native intron addition, and SL1 3' splice site sequence) simultaneously and used the previously established *egl-20* promoter to drive fluorescent protein expression. We were able to obtain GFP transcriptional reporter lines with robust and intense signals (Figure 3). More importantly, we considerably improved the efficiency of transgenesis of both GFP (PZH008) and TurboRFP (PZH009) constructs ($P = 0.02$ and $P = 0.003$, respectively; Table 1). Note that we still experienced variability in the efficiency; possibly due to factors such as injector and age of the specimen, the efficiency increased to >20% of injected animals. While we did not systematically test all variables individually due to the enormous costs that would have been associated with such studies, we confirmed the increase in efficiency by the subsequent removal of introns. Indeed, intron removal coincided with a decrease in transgenic efficiency (Table 1). Together, we found that the codon-optimized three-intron GFP and TurboRFP had greater efficiency compared with the previous nonoptimized TurboRFP (Schlager *et al.* 2009). However, for unknown reasons, utilizing the CUB of

the top 3% highly expressed genes did not further increase the efficiency (Table S3).

Fluorescent co-injection marker-assisted CRISPR genome editing

With the establishment of reliable and robust transgenic markers in *P. pacificus*, we next attempted to implement these tools to reduce the workload and the cost of screening potential CRISPR/Cas9 alleles. Therefore, we tried to establish a method that employed the optimized fluorescent markers to identify potential mutants induced with CRISPR/Cas9 (Figure 4A). A fluorescent marker can indicate well-injected specimens, which carry an increased likelihood of successfully induced CRISPR/Cas9 mutations. Therefore, using the *egl-20p::TurboRFP* (PZH009) as a CRISPR/Cas9 co-injection marker, our experienced injectors obtained between 1 and 5 P0s (on average 2.5) producing RFP-positive F1 progeny from 30 well-injected nematodes (Figure 4B). Furthermore, progeny cooccurring on RFP injection marker-positive plates also frequently carried CRISPR/Cas9-induced mutations at high efficiency (77% of the identified plates),

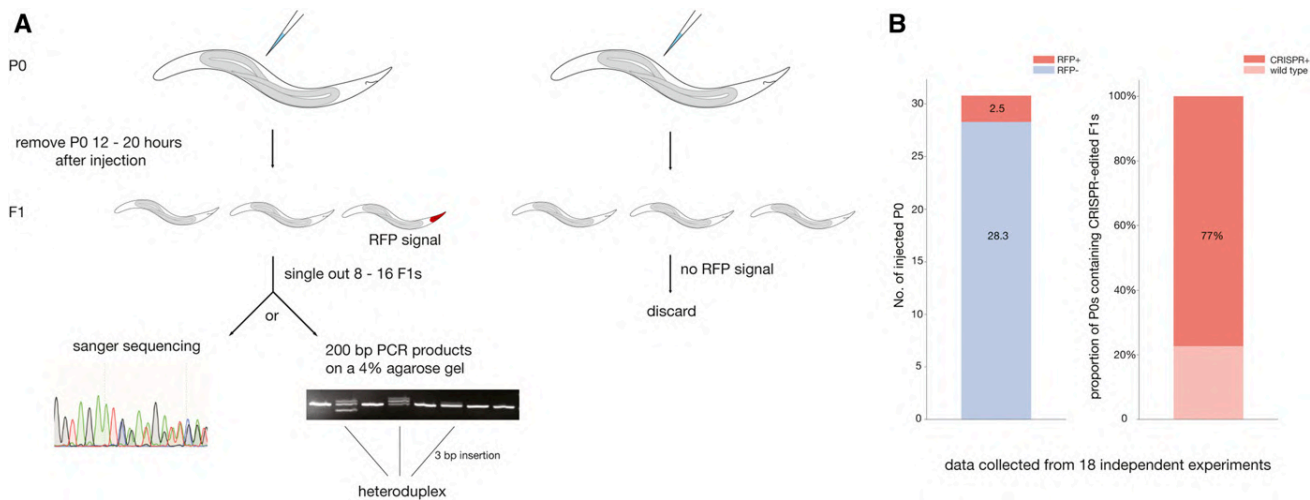


Figure 4 Newly established fluorescent co-injection marker-assisted CRISPR genome editing in *P. pacificus*. (A) An illustration of the workflow for CRISPR genome editing in *P. pacificus*. (B) Using an *egl-20p::TurboRFP* construct (PZH009) as a co-injection marker, an average of 2.5 P0s had RFP + F1s from 30 well-injected *P. pacificus* P0s (left). (B) Next, 8–16 F1s were selected from each P0 with RFP + F1s to detect CRISPR alleles. There was a 77% chance that the P0-contained RFP + F1s also contained CRISPR-edited F1s (right). These data were accumulated from 18 independent experiments. Note, there were two additional experiments with no RFP + F1 detected, but CRISPR editing still occurred. CRISPR, clustered regularly interspaced short palindromic repeats; RFP, red fluorescent protein.

allowing the number of progeny necessary to be screened to isolate a CRISPR/Cas9 mutant to be greatly reduced (Figure 4B). Thus, the improved fluorescent-based co-injection marker strongly assisted the detection of CRISPR-generated edits in *P. pacificus*. We would like to note here that this fluorescent marker-assisted CRISPR method is compatible with knockouts and shorter repair templates (<120 nt), but does not seem to work with longer repair templates.

Discussion

The usage of transgenic tools is fundamental to successful studies in molecular biology; however, their efficiency is not uniform between organisms. This is, in part, likely due to differences in gene regulatory mechanisms between different species. In canonical model organisms, the development of efficient transgenic tools is aided by the existence of large scientific communities capable of refining and optimizing their application; however, this is not usually possible in other systems. Although the delivery of DNA to the germline can be an obstacle in nematode species (Kranse *et al.* 2020), delivery via microinjection is not a hindrance in *Pristionchus*, since we have generally achieved a higher efficiency for CRISPR knockouts compared with transgenesis. Here, we have revealed large differences in CUB and IME across nematodes, which likely contribute to gene regulatory differences between species. As a proof of principle, we investigated a single nematode species, *P. pacificus*, whereby we have successfully exploited its favored CUB and IME to develop *P. pacificus*-adapted fluorescent transgenic proteins. Additionally, we have shown that these adapted proteins containing *P. pacificus* gene regulatory requirements demonstrate a dramatically increased expression efficiency. It has recently been shown in *C. elegans* that the 5' intron contributes the most to the elevated level of gene expression (Crane *et al.* 2019). Our results in *P. pacificus* agree with this finding because transgenesis efficiency decreased when the 5' intron was removed. Transgenes with constructs that were modified using the top 3% CUB did not further improve efficiency. We can only speculate that this might be due to the most highly favored codons causing ribosomal traffic jams (Plotkin and Kudla 2011). Nevertheless, these improvements allow transgenes to be utilized as co-injection markers to reduce the screening time and costs of CRISPR/Cas9 genome editing. Thus, our method provides an alternative to the existing *Pristionchus* co-CRISPR method, in which the identification of CRISPR candidates relies on a Dpy phenotype (Nakayama *et al.* 2020).

By means of an initial bioinformatic analysis of the species-specific CUB and IME, our experiments demonstrate the potential to develop optimized transgenic tools and explore distinctive attributes that were not previously possible. While we did not systematically test the specific contributions of CUB, IME, and SL1, they likely all play important roles in transcription and translation for the increased transgenesis efficiency in *P. pacificus* (Redemann *et al.* 2011; Yang *et al.*

2017; Crane *et al.* 2019), but it is important to note that this principle could be further utilized to optimize genetically encoded calcium indicators and optogenetic tools to explore *Pristionchus*-specific behaviors, and genetic ablation methods to investigate aspects of anatomy and physiology. We hypothesize that, using knowledge of species-specific genomic features, it is possible to establish transgenic tool kits in other free-living nematodes, and additionally in parasitic nematode systems that have a significant impact on world health (Brindley *et al.* 2009) and crop production (Nicol *et al.* 2011).

Acknowledgments

The authors would like to thank the members of the Sommer laboratory for collecting CRISPR editing data. Adrian Streit provided insightful thoughts about this manuscript. This work was supported by the Max Planck Society (funding awarded to R.J.S.); an Alexander von Humboldt Foundation Postdoctoral fellowship (awarded to Z.H.), and a Chinese Scholarship Council Ph.D. fellowship (awarded to S.S.). Construct maps in Snapgene files are available upon request.

Literature Cited

- Adams, S., P. Pathak, H. Shao, J. B. Lok, and A. Pires-daSilva, 2019 Liposome-based transfection enhances RNAi and CRISPR-mediated mutagenesis in non-model nematode systems. *Sci. Rep.* 9: 483. <https://doi.org/10.1038/s41598-018-37036-1>
- Akduman, N., J. W. Lightfoot, W. Röseler, H. Witte, W.-S. Lo *et al.*, 2020 Bacterial vitamin B₁₂ production enhances nematode predatory behavior. *ISME J.* 14: 1494–1507 (erratum: *ISME J.* 14: 1911). <https://doi.org/10.1038/s41396-020-0626-2>
- Allen, M. A., L. W. Hillier, R. H. Waterston, and T. Blumenthal, 2011 A global analysis of *C. elegans* trans-splicing. *Genome Res.* 21: 255–264. <https://doi.org/10.1101/gr.113811.110>
- Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant *et al.*, 2009 MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37: W202–W208. <https://doi.org/10.1093/nar/gkp335>
- Bhattacharya, D., and E. G. Van Meir, 2019 A simple genotyping method to detect small CRISPR-Cas9 induced indels by agarose gel electrophoresis. *Sci. Rep.* 9: 4437. <https://doi.org/10.1038/s41598-019-39950-4>
- Bischof, J., R. K. Maeda, M. Hediger, F. Karch, and K. Basler, 2007 An optimized transgenesis system for *Drosophila* using germ-line-specific C31 integrases. *Proc. Natl. Acad. Sci. USA* 104: 3312–3317. <https://doi.org/10.1073/pnas.0611511104>
- Blaxter, M. L., P. De Ley, J. R. Garey, L. X. Liu, P. Scheldeman *et al.*, 1998 A molecular evolutionary framework for the phylum Nematoda. *Nature* 392: 71–75. <https://doi.org/10.1038/32160>
- Brindley, P. J., M. Mitreva, E. Ghedin, and S. Lustigman, 2009 Helminth genomics: the implications for human health. *PLoS Negl. Trop. Dis.* 3: e538. <https://doi.org/10.1371/journal.pntd.0000538>
- Brinster, R. L., J. M. Allen, R. R. Behringer, R. E. Gelinas, and R. D. Palmiter, 1988 Introns increase transcriptional efficiency in transgenic mice. *Proc. Natl. Acad. Sci. USA* 85: 836–840. <https://doi.org/10.1073/pnas.85.3.836>

- Bui, L. T., N. A. Ivers, and E. J. Ragsdale, 2018 A sulfotransferase dosage-dependently regulates mouthpart polyphenism in the nematode *Pristionchus pacificus*. *Nat. Commun.* 9: 4119. <https://doi.org/10.1038/s41467-018-05612-8>
- Castillo-Davis, C. I., S. L. Mekhedov, D. L. Hartl, E. V. Koonin, and F. A. Kondrashov, 2002 Selection for short introns in highly expressed genes. *Nat. Genet.* 31: 415–418. <https://doi.org/10.1038/ng940>
- Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher, 1994 Green fluorescent protein as a marker for gene expression. *Science* 263: 802–805. <https://doi.org/10.1126/science.8303295>
- Choi, Y.-J., E. Ghedin, M. Berriman, J. McQuillan, N. Holroyd *et al.*, 2011 A deep sequencing approach to comparatively analyze the transcriptome of lifecycle stages of the filarial worm, *Brugia malayi*. *PLoS Negl. Trop. Dis.* 5: e1409. <https://doi.org/10.1371/journal.pntd.0001409>
- Clough, S. J., and A. F. Bent, 1998 Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16: 735–743. <https://doi.org/10.1046/j.1365-313x.1998.00343.x>
- Cotton, J. A., C. J. Lilley, L. M. Jones, T. Kikuchi, A. J. Reid *et al.*, 2014 The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biol.* 15: R43. <https://doi.org/10.1186/gb-2014-15-3-r43>
- Crane, M. M., B. Sands, C. Battaglia, B. Johnson, S. Yun *et al.*, 2019 In vivo measurements reveal a single 5'-intron is sufficient to increase protein expression level in *Caenorhabditis elegans*. *Sci. Rep.* 9: 9192. <https://doi.org/10.1038/s41598-019-45517-0>
- Denker, J. A., D. M. Zuckerman, P. A. Maroney, and T. W. Nilsen, 2002 New components of the spliced leader RNP required for nematode trans-splicing. *Nature* 417: 667–670. <https://doi.org/10.1038/nature00783>
- Dokshin, G. A., K. S. Ghanta, K. M. Piscopo, and C. C. Mello, 2018 Robust genome editing with short single-stranded and long, partially single-stranded DNA donors in *Caenorhabditis elegans*. *Genetics* 210: 781–787. <https://doi.org/10.1534/genetics.118.301532>
- Duret, L., and D. Mouchiroud, 1999 Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 96: 4482–4487. <https://doi.org/10.1073/pnas.96.8.4482>
- Evans, T. C., 2006 Transformation and microinjection (April 6, 2006), *WormBook*, ed. The *C. elegans* Research Community *WormBook*, doi/10.1895/wormbook.1.108.1, <http://www.wormbook.org>. <https://doi.org/10.1895/wormbook.1.108.1>
- Fire, A. Fire Lab *C. elegans* Vector Kit 1995. <https://media.addgene.org/cms/files/Vec95.pdf>
- Foster, J. M., A. Grote, J. Mattick, A. Tracey, Y.-C. Tsai *et al.*, 2020 Sex chromosome evolution in parasitic nematodes of humans. *Nat. Commun.* 11: 1964. <https://doi.org/10.1038/s41467-020-15654-6>
- Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29: 644–652. <https://doi.org/10.1038/nbt.1883>
- Grün, D., M. Kirchner, N. Thierfelder, M. Stoeckius, M. Selbach *et al.*, 2014 Conservation of mRNA and protein expression during development of *C. elegans*. *Cell Rep.* 6: 565–577. <https://doi.org/10.1016/j.celrep.2014.01.001>
- Guiliano, D. B., and M. L. Blaxter, 2006 Operon conservation and the evolution of trans-splicing in the phylum Nematoda. *PLoS Genet.* 2: e198. <https://doi.org/10.1371/journal.pgen.0020198>
- Higazi, T. B., A. Merriweather, L. Shu, R. Davis, and T. R. Unnasch, 2002 *Brugia malayi*: transient transfection by microinjection and particle bombardment. *Exp. Parasitol.* 100: 95–102. [https://doi.org/10.1016/S0014-4894\(02\)00004-8](https://doi.org/10.1016/S0014-4894(02)00004-8)
- Hong, R. L., and R. J. Sommer, 2006 *Pristionchus pacificus*: a well-rounded nematode. *Bioessays* 28: 651–659. <https://doi.org/10.1002/bies.20404>
- Howe, K. L., B. J. Bolt, S. Cain, J. Chan, W. J. Chen *et al.*, 2016 WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res.* 44: D774–D780. <https://doi.org/10.1093/nar/gkv1217>
- Howe, K. L., B. J. Bolt, M. Shafie, P. Kersey, and M. Berriman, 2017 WormBase ParaSite - a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* 215: 2–10. <https://doi.org/10.1016/j.molbiopara.2016.11.005>
- Hunt, V. L., I. J. Tsai, A. Coghlan, A. J. Reid, N. Holroyd *et al.*, 2016 The genomic basis of parasitism in the Strongyloides clade of nematodes. *Nat. Genet.* 48: 299–307. <https://doi.org/10.1038/ng.3495>
- Hutter, H., 2012 Fluorescent protein methods: strategies and applications. *Methods Cell Biol.* 107: 67–92. <https://doi.org/10.1016/B978-0-12-394620-1.00003-5>
- Kieninger, M. R., N. A. Ivers, C. Rödelberger, G. V. Markov, R. J. Sommer *et al.*, 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: 2174–2179. <https://doi.org/10.1016/j.cub.2016.06.018>
- Kikuchi, T., J. A. Cotton, J. J. Dalzell, K. Hasegawa, N. Kanzaki *et al.*, 2011 Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathog.* 7: e1002219. <https://doi.org/10.1371/journal.ppat.1002219>
- Kim, D., B. Langmead, and S. L. Salzberg, 2015 HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12: 357–360. <https://doi.org/10.1038/nmeth.3317>
- Kranse O., H. Beasley, S. Adams, A. P. da Silva, C. Bell, *et al.*, 2020 Towards genetic modification of plant-parasitic nematodes: delivery of macromolecules to adults and expression of exogenous mRNA in second stage juveniles. *bioRxiv*. doi.org/10.1101/2020.07.15.193052 (Preprint posted July 15, 2020).
- Laing, R., T. Kikuchi, A. Martinelli, I. J. Tsai, R. N. Beech *et al.*, 2013 The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery. *Genome Biol.* 14: R88. <https://doi.org/10.1186/gb-2013-14-8-r88>
- Lee, R. Y. N., K. L. Howe, T. W. Harris, V. Arnaboldi, S. Cain *et al.*, 2018 WormBase 2017: molting into a new stage. *Nucleic Acids Res.* 46: D869–D874. <https://doi.org/10.1093/nar/gkx998>
- Liao, Y., G. K. Smyth, and W. Shi, 2014 featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930. <https://doi.org/10.1093/bioinformatics/btt656>
- Lightfoot, J. W., M. Wilecki, C. Rödelberger, E. Moreno, V. Susoy *et al.*, 2019 Small peptide-mediated self-recognition prevents cannibalism in predatory nematodes. *Science* 364: 86–89. <https://doi.org/10.1126/science.aav9856>
- Li, X., H. C. Massey, Jr., T. J. Nolan, G. A. Schad, K. Kraus *et al.*, 2006 Successful transgenesis of the parasitic nematode *Strongyloides stercoralis* requires endogenous non-coding control elements. *Int. J. Parasitol.* 36: 671–679. <https://doi.org/10.1016/j.ijpara.2005.12.007>
- Liu, Y., K. G. Kaval, A. van Hoof, and D. A. Garsin, 2019 Heme peroxidase HPX-2 protects *Caenorhabditis elegans* from pathogens. *PLoS Genet.* 15: e1007944. <https://doi.org/10.1371/journal.pgen.1007944>
- Loer, C., H. Witte, R. Sommer, and O. Hobert, 2019 An antibody staining protocol variation for nematodes that adds heat-induced antigen retrieval (HIAR). *MicroPubl Biol* 2019. Available at:

- <https://www.micropublication.org/journals/biology/micropub.biology.000135/>. <https://doi.org/10.17912/micropub.biology.000135>
- Lok, J. B., 2012 Nucleic acid transfection and transgenesis in parasitic nematodes. *Parasitology* 139: 574–588. <https://doi.org/10.1017/S0031182011001387>
- Mello, C. C., J. M. Kramer, D. Stinchcomb, and V. Ambros, 1991 Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* 10: 3959–3970. <https://doi.org/10.1002/j.1460-2075.1991.tb04966.x>
- Mitreva, M., M. C. Wendl, J. Martin, T. Wylie, Y. Yin *et al.*, 2006 Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Biol.* 7: R75. <https://doi.org/10.1186/gb-2006-7-8-r75>
- Moreno, E., A. McGaughran, C. Rödelberger, M. Zimmer, and R. J. Sommer, 2016 Oxygen-induced social behaviours in *Pristionchus pacificus* have a distinct evolutionary history and genetic regulation from *Caenorhabditis elegans*. *Proc. Biol. Sci.* 283: 20152263. <https://doi.org/10.1098/rspb.2015.2263>
- Moreno, E., B. Sieriebriennikov, H. Witte, C. Rödelberger, J. W. Lightfoot *et al.*, 2017 Regulation of hyperoxia-induced social behaviour in *Pristionchus pacificus* nematodes requires a novel cilia-mediated environmental input. *Sci. Rep.* 7: 17550. <https://doi.org/10.1038/s41598-017-18019-0>
- Moreno, E., J. W. Lightfoot, M. Lenuzzi, and R. J. Sommer, 2019 Cilia drive developmental plasticity and are essential for efficient prey detection in predatory nematodes. *Proc. Biol. Sci.* 286: 20191089. <https://doi.org/10.1098/rspb.2019.1089>
- Nakayama, K.-I., Y. Ishita, T. Chihara, and M. Okumura, 2020 Screening for CRISPR/Cas9-induced mutations using a co-injection marker in the nematode *Pristionchus pacificus*. *Dev. Genes Evol.* 230: 257–264. <https://doi.org/10.1007/s00427-020-00651-y>
- Nicol, J. M., S. J. Turner, D. L. Coyne, L. den Nijs, S. Hockland *et al.*, 2011 Current nematode threats to world agriculture, pp. 21–43 in *Genomics and Molecular Genetics of Plant-Nematode Interactions*, edited by J. Jones, G. Gheysen, and C. Fenoll. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-007-0434-3_2
- Okkema, P. G., S. W. Harrison, V. Plunger, A. Aryana, and A. Fire, 1993 Sequence requirements for myosin gene expression and regulation in *Caenorhabditis elegans*. *Genetics* 135: 385–404.
- Okumura, M., M. Wilecki, and R. J. Sommer, 2017 Serotonin drives predatory feeding behavior via synchronous feeding rhythms in the nematode *Pristionchus pacificus*. *G3 (Bethesda)* 7: 3745–3755. <https://doi.org/10.1534/g3.117.300263>
- Plotkin, J. B., and G. Kudla, 2011 Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12: 32–42. <https://doi.org/10.1038/nrg2899>
- Prabh, N., W. Roeseler, H. Witte, G. Eberhardt, R. J. Sommer *et al.*, 2018 Deep taxon sampling reveals the evolutionary dynamics of novel gene families in *Pristionchus* nematodes. *Genome Res.* 28: 1664–1674. <https://doi.org/10.1101/gr.234971.118>
- Ragle, J. M., S. Katzman, T. F. Akers, S. Barberan-Soler, and A. M. Zahler, 2015 Coordinated tissue-specific regulation of adjacent alternative 3' splice sites in *C. elegans*. *Genome Res.* 25: 982–994. <https://doi.org/10.1101/gr.186783.114>
- Ragsdale, E. J., M. R. Müller, C. Rödelberger, and R. J. Sommer, 2013 A developmental switch coupled to the evolution of plasticity acts through a sulfatase. *Cell* 155: 922–933. <https://doi.org/10.1016/j.cell.2013.09.054>
- Redemann, S., S. Schloissnig, S. Ernst, A. Pozniakowsky, S. Ayloo *et al.*, 2011 Codon adaptation-based control of protein expression in *C. elegans*. *Nat. Methods* 8: 250–252. <https://doi.org/10.1038/nmeth.1565>
- Rice, P., I. Longden, and A. Bleasby, 2000 EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16: 276–277. [https://doi.org/10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2)
- Rödelberger, C., W. Röseler, N. Prabh, K. Yoshida, C. Weiler *et al.*, 2018 Phylotranscriptomics of *Pristionchus* nematodes reveals parallel gene loss in six hermaphroditic lineages. *Curr. Biol.* 28: 3123–3127.e5. <https://doi.org/10.1016/j.cub.2018.07.041>
- Rödelberger, C., M. Athanasouli, M. Lenuzzi, T. Theska, S. Sun *et al.*, 2019 Crowdsourcing and the feasibility of manual gene annotation: a pilot study in the nematode *Pristionchus pacificus*. *Sci. Rep.* 9: 18789. <https://doi.org/10.1038/s41598-019-55359-5>
- Rubin, G. M., and A. C. Spradling, 1982 Genetic transformation of *Drosophila* with transposable element vectors. *Science* 218: 348–353. <https://doi.org/10.1126/science.6289436>
- Schlager, B., X. Wang, G. Braach, and R. J. Sommer, 2009 Molecular cloning of a dominant roller mutant and establishment of DNA-mediated transformation in the nematode *Pristionchus pacificus*. *Genesis* 47: 300–304. <https://doi.org/10.1002/dvg.20499>
- Sharp, P. M., and W. H. Li, 1987 The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15: 1281–1295. <https://doi.org/10.1093/nar/15.3.1281>
- Sieriebriennikov, B., S. Sun, J. W. Lightfoot, H. Witte, E. Moreno *et al.*, 2020 Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell. *PLoS Genet.* 16: e1008687. <https://doi.org/10.1371/journal.pgen.1008687>
- Sommer, R. J., L. K. Carta, S. Kim, and P. W. Sternberg, 1996 Morphological, genetic and molecular description of *Pristionchus pacificus*. *Fundam. Appl. Nematol.* 6: 511–521.
- Tanaka, S. E., M. Dayi, Y. Maeda, I. J. Tsai, R. Tanaka *et al.*, 2019 Stage-specific transcriptome of *Bursaphelenchus xylophilus* reveals temporal regulation of effector genes and roles of the dauer-like stages in the lifecycle. *Sci. Rep.* 9: 6080. <https://doi.org/10.1038/s41598-019-42570-7>
- Wang, J., B. Czech, A. Crunk, A. Wallace, M. Mitreva *et al.*, 2011 Deep small RNA sequencing from the nematode *Ascaris* reveals conservation, functional diversification, and novel developmental profiles. *Genome Res.* 21: 1462–1477. <https://doi.org/10.1101/gr.121426.111>
- Wang, J., S. Gao, Y. Mostovoy, Y. Kang, M. Zagoskin *et al.*, 2017 Comparative genome analysis of programmed DNA elimination in nematodes. *Genome Res.* 27: 2001–2014. <https://doi.org/10.1101/gr.225730.117>
- Wilecki, M., J. W. Lightfoot, V. Susoy, and R. J. Sommer, 2015 Predatory feeding behaviour in *Pristionchus* nematodes is dependent on phenotypic plasticity and induced by serotonin. *J. Exp. Biol.* 218: 1306–1313. <https://doi.org/10.1242/jeb.118620>
- Witte, H., E. Moreno, C. Rödelberger, J. Kim, J.-S. Kim *et al.*, 2015 Gene inactivation using the CRISPR/Cas9 system in the nematode *Pristionchus pacificus*. *Dev. Genes Evol.* 225: 55–62. <https://doi.org/10.1007/s00427-014-0486-8>
- Yang, Y.-F., X. Zhang, X. Ma, T. Zhao, Q. Sun *et al.*, 2017 Trans-splicing enhances translational efficiency in *C. elegans*. *Genome Res.* 27: 1525–1535. <https://doi.org/10.1101/gr.202150.115>

Communicating editor: O. Hobert

OPEN

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

Christian Rödel*perger, 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 evolution³ 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 community³⁸. 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

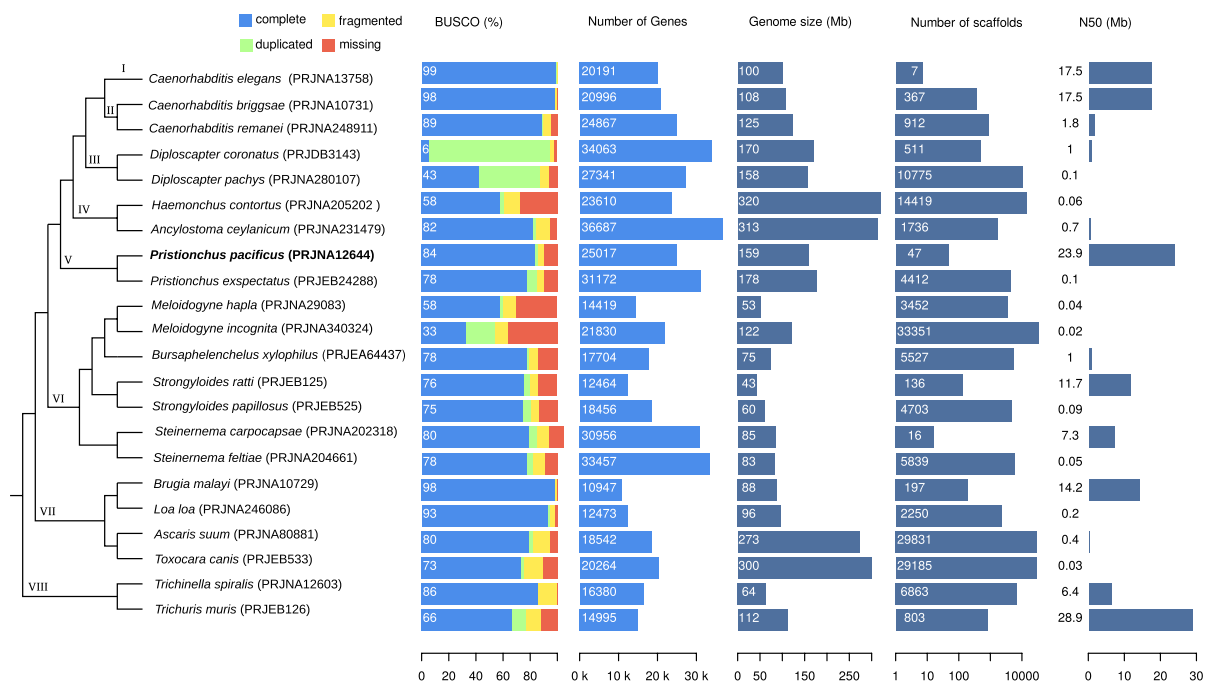


Figure 1. Comparative assessment of nematode genome quality. Genomic data for 22 nematode species was obtained from WormBase ParaSite (release WBPS13) and evaluated based on completeness level of gene annotations and genome assembly contiguity. The barplots show the results of a benchmarking of single copy orthologs (BUSCO⁴⁰) 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 nematodes³⁹, 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 level³⁸.

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 phylum³⁹. 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 universal single copy orthologs (BUSCO⁴⁰) 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 nature⁴¹. 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 duplication⁴². 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 map¹⁵, 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^{11,13,43}. 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 BUSCO⁴⁰ 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
<i>de novo</i> 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 improvement¹⁶. 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 reads⁴⁴. 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 cultures⁴⁵. 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 overlap⁴⁵. 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 to 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

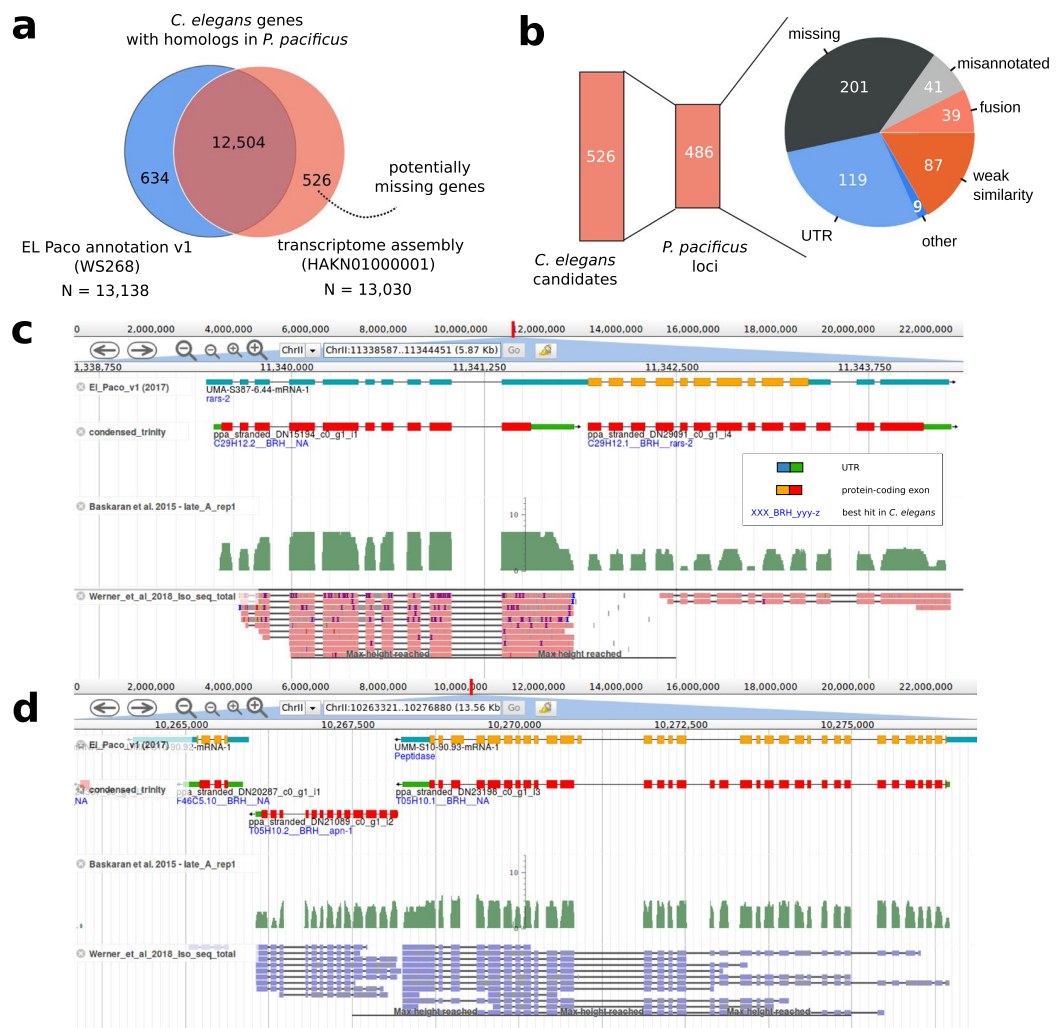


Figure 2. Identification of missing genes. **(a)** 526 potentially missing genes were identified based on *C. elegans* genes with homologs in the transcriptome assembly but not in current gene annotations. **(b)** The 526 missing gene candidates were located in 486 *P. pacificus* loci that were classified 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. pacificus* 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:

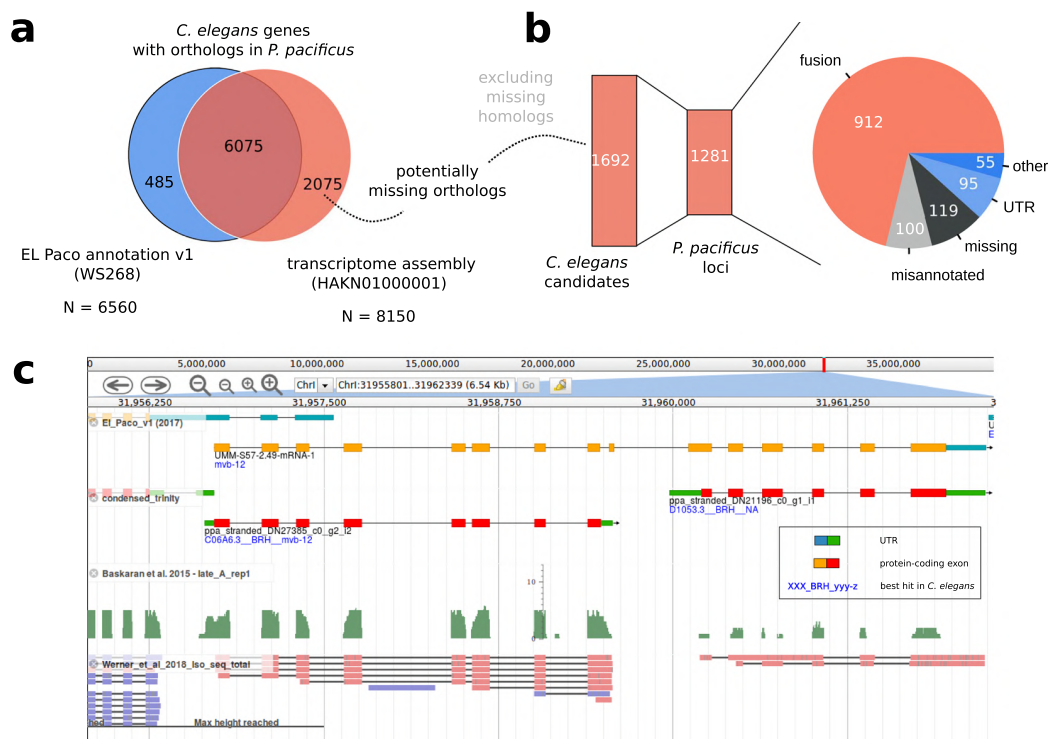


Figure 3. Community-based curation of hidden orthologs. (a) We identified 2075 putative *C. elegans* one-to-one orthologs that were specific to the *P. pacificus* transcriptome assembly. (b) Community-based curation classified most of the corresponding gene loci as artificial gene fusions. (c) Non-overlapping 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 annotations¹⁵ (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 analysis⁴⁷ 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 tree⁴⁷. 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 orthAgogue⁴⁸. 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 account⁴⁹. 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 analysis⁴⁶. While 53 of the previously confirmed one-to-one orthologs were captured as best reciprocal hits, only 33 were also identified

by orthoAgogue. 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* genome¹⁵ 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 genomes³⁸, 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 compact⁹ and genes frequently overlap^{37,45}. Consequently, manual annotation of restricted gene sets has been proposed and applied previously to circumvent this problem³⁸. 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 genes⁵⁵ 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 conditions⁵⁹, sexes¹⁶, 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 data⁴⁴ and a *de novo* assembly of strand-specific RNA-seq data¹⁶ into our genome browser (implemented in jbrowse⁶²) 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 exonerate⁶³ (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 (Hybrid1⁵⁶ and TAU2011⁵⁷) 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 assembly¹⁶. 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 orthoAgogue was run with default setting on the same input files⁴⁸.

Data availability

The strand-specific *de novo* transcriptome was submitted to the European Nucleotide Archive under the accession HAKN01000001¹⁶ and the Iso-seq data was submitted to the European Nucleotide Archive under the accessions ERX2315712 and ERX2315713⁴⁴. 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

- 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).
- 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).
- 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).
- Bose, N. *et al.* Complex small-molecule architectures regulate phenotypic plasticity in a nematode. *Angew. Chem. Int. Ed Engl.* **51**, 12438–12443 (2012).
- 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).
- Falcke, J. M. *et al.* Linking genomic and metabolomic natural variation uncovers nematode pheromone biosynthesis. *Cell Chem Biol* **25**, 787–796.e12 (2018).
- 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).
- Prabh, N. *et al.* Deep taxon sampling reveals the evolutionary dynamics of novel gene families in *Pristionchus* nematodes. *Genome Res.* **28**, 1664–1674 (2018).
- 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).
- 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).
- Susoy, V. *et al.* Large-scale diversification without genetic isolation in nematode symbionts of figs. *Sci Adv* **2**, e1501031 (2016).
- 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).
- Witte, H. *et al.* Gene inactivation using the CRISPR/Cas9 system in the nematode *Pristionchus pacificus*. *Dev. Genes Evol.* **225**, 55–62 (2015).
- Srinivasan, J. *et al.* A bacterial artificial chromosome-based genetic linkage map of the nematode *Pristionchus pacificus*. *Genetics* **162**, 129–134 (2002).
- 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).
- Rödelsperger, C. *et al.* Phylotranscriptomics of *Pristionchus* nematodes reveals parallel gene loss in six hermaphroditic lineages. *Curr. Biol.* **28**, 3123–3127.e5 (2018).
- 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).
- 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).
- 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).
- 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).
- Liu, Z. *et al.* Predator-secreted sulfolipids induce defensive responses in *C. elegans*. *Nature Communications* **9** (2018).
- Moreno, E. *et al.* DAF-19/RFX controls ciliogenesis and influences oxygen-induced social behaviors in *Pristionchus pacificus*. *Evol. Dev.* **20**, 233–243 (2018).
- 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).
- 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).
- Namdeo, S. *et al.* Two independent sulfation processes regulate mouth-form plasticity in the nematode. *Development* **145** (2018).
- Rödelsperger, C. Comparative genomics of gene loss and gain in *Caenorhabditis* and Other Nematodes. In *Methods in Molecular Biology* 419–432 (2018).
- Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
- Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–9 (2006).
- Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **1962**, 65–95 (2019).
- Cantarel, B. L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
- Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
- Lightfoot, J. W. *et al.* Small peptide-mediated self-recognition prevents cannibalism in predatory nematodes. *Science* **364**, 86–89 (2019).
- 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).
- 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).
- 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).
- 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).
- Prabh, N. & Rödelsperger, C. Divergence, and mixed origin contribute to the emergence of orphan genes in nematodes. *G3* **9**, 2277–2286 (2019).
- Salzberg, S. L. Next-generation genome annotation: we still struggle to get it right. *Genome Biol.* **20**, 92 (2019).
- 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).
- 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).
- McLean, F., Berger, D., Laetsch, D. R., Schwartz, H. T. & Blaxter, M. Improving the annotation of the *Heterorhabditis bacteriophora* genome. *Gigascience* **7** (2018).
- Hiraki, H. *et al.* Genome analysis of *Diploscapter coronatus*: insights into molecular peculiarities of a nematode with parthenogenetic reproduction. *BMC Genomics* **18**, 478 (2017).
- 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., Serobyán, 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. orthoAgogue: 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. Serobyán, 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

Improving Transgenesis Efficiency and CRISPR-Associated Tools Through Codon Optimization and Native Intron Addition in *Pristionchus* Nematodes

Ziduan Han,¹ Wen-Sui Lo,¹ James W. Lightfoot, Hanh Witte, Shuai Sun, and Ralf J. Sommer²

Max Planck Institute for Developmental Biology, Tuebingen 72076, Germany

ORCID IDs: 0000-0001-8341-0678 (Z.H.); 0000-0002-2438-0015 (W.-S.L.); 0000-0002-5835-2135 (J.W.L.); 0000-0003-1503-7749 (R.J.S.)

ABSTRACT A lack of appropriate molecular tools is one obstacle that prevents in-depth mechanistic studies in many organisms. Transgenesis, clustered regularly interspaced short palindromic repeats (CRISPR)-associated engineering, and related tools are fundamental in the modern life sciences, but their applications are still limited to a few model organisms. In the phylum Nematoda, transgenesis can only be performed in a handful of species other than *Caenorhabditis elegans*, and additionally, other species suffer from significantly lower transgenesis efficiencies. We hypothesized that this may in part be due to incompatibilities of transgenes in the recipient organisms. Therefore, we investigated the genomic features of 10 nematode species from three of the major clades representing all different lifestyles. We found that these species show drastically different codon usage bias and intron composition. With these findings, we used the species *Pristionchus pacificus* as a proof of concept for codon optimization and native intron addition. Indeed, we were able to significantly improve transgenesis efficiency, a principle that may be usable in other nematode species. In addition, with the improved transgenes, we developed a fluorescent co-injection marker in *P. pacificus* for the detection of CRISPR-edited individuals, which helps considerably to reduce associated time and costs.

KEYWORDS nematodes; *C. elegans*; *P. pacificus*; transgenesis; CRISPR editing; codon usage bias; intron-mediated enhancement; parasitic nematodes

The utilization of transgenes has proven fundamental to many aspects of molecular biology and for functional genomic studies (Rubin and Spradling 1982; Mello *et al.* 1991; Chalfie *et al.* 1994; Clough and Bent 1998; Hutter 2012). For instance, easily applied and efficient transgenic methods have been instrumental in furthering our understanding of biological pathways and dissecting associated phenotypes. Additionally, it has facilitated the visualization of gene expression patterns and protein localization through the usage of fluorescent proteins such as GFP in a swathe of organisms (Chalfie *et al.* 1994). However, limiting factors for the successful establishment of transgenesis in an organism

are the differing regulatory strategies and mechanisms found between species. In accordance with this and despite their ubiquitous usage, efficient transgenesis tools are frequently restricted to canonical model organisms.

Gene expression, including transgene expression, is regulated by a multitude of factors, including at the transcriptional and translational levels. One such regulatory mechanism is through codon usage bias (CUB). Here, the degenerate nature of the nucleotide triplet code ensures that each amino acid can be encoded by several synonymous codons, with the exception of the amino acids methionine and tryptophan (Sharp and Li 1987). Correspondingly, organism genomes show their own distinct usage of the code. This codon bias is more pronounced in genes with elevated expression levels. Specifically, highly expressed genes strongly favor a specific set of codons with the favored codons contributing to a more efficient translation process through faster ribosome elongation (Duret and Mouchiroud 1999; Plotkin and Kudla 2011). Further, artificial manipulation of the CUB can also alter gene expression dramatically (Redemann *et al.* 2011). In addition to CUB, regulatory regions of a gene are also thought to be

Copyright © 2020 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.120.303785>

Manuscript received September 9, 2020; accepted for publication October 14, 2020; published Early Online October 15, 2020.

Available freely online through the author-supported open access option.

Supplemental material available at figshare: <https://doi.org/10.25386/genetics.13090322>.

¹These authors contributed equally to this work.

²Corresponding author: Max Planck Institute for Developmental Biology, Department for Integrative Evolutionary Biology, Max-Planck-Ring 9, 72076 Tübingen, Germany. E-mail: ralf.sommer@tuebingen.mpg.de

crucial for transcriptional control as evidence suggests a relationship between the exon–intron structure of a gene and its expression through a process termed “intron-mediated enhancement” (IME). Here, intron density positively correlates with both the level and extent of a gene’s expression (Castillo-Davis *et al.* 2002). As such, these phenomena have been exploited for the enhancement of molecular tools including improving transgenesis in a number of well-studied model organisms (Brinster *et al.* 1988; Bischof *et al.* 2007).

One such organism is the nematode *Caenorhabditis elegans*, where an abundance of molecular tools, including transgenesis, are available and its CUB and exon–intron structures are well characterized (Ragle *et al.* 2015). In particular, IME in *C. elegans* is strongly influenced by the position, number, and sequence of introns, and introns positioned near the 5’ end of a gene shows the greatest contribution to this effect (Okkema *et al.* 1993; Crane *et al.* 2019). Further, replacing native codons with favored codons increases the translation level of a protein (Redemann *et al.* 2011). However, *C. elegans* is far from the only nematode of significance in the phylum, with an array of parasitic nematodes of both animals and plants, as well as other free-living nematodes, also now frequently used for research. Despite this, transgenesis has only been successfully applied to a few nematode species outside of the genus *Caenorhabditis* (Higazi *et al.* 2002; Li *et al.* 2006; Schlager *et al.* 2009; Lok 2012), with problems arising due to efficient delivery of DNA materials to the gonad (Evans 2006) and compatibility of the DNA to the endogenous genetic machinery of the recipient. Therefore, despite recent advancements (Adams *et al.* 2019), low efficiency is still the bottleneck for most transgenic experiments in other nematode species.

In addition to *C. elegans*, another distantly related free-living nematode frequently used for research is *Pristionchus pacificus* (Sommer *et al.* 1996). This nematode has been established as a model system to study evolutionary developmental biology and, more specifically, the evolution of novelty. This is due to the nature of its mouth structure, which is phenotypically plastic and demonstrates two distinct variants. One morph exhibits two teeth while the other contains only a single tooth. The genetic network behind this developmental decision has been extensively studied and is heavily influenced by the nematode’s environment (Ragsdale *et al.* 2013; Kieninger *et al.* 2016; Bui *et al.* 2018; Sieriebriennikov *et al.* 2020). The presence of teeth in *P. pacificus* facilitates an additional behavior as they are capable of preying on the larvae of other nematodes. Here, it has been observed that the mouth-form dimorphism strongly correlates with the predation behavior, as only the morphs possessing two teeth are active predators, whereas the single-toothed morphs are strict bacterial feeders (Wilecki *et al.* 2015; Moreno *et al.* 2019; Akduman *et al.* 2020). Furthermore, the predatory behavior coincides with the existence of a self-recognition system (Lightfoot *et al.* 2019) and environmental responses distinct from *C. elegans* (Hong and Sommer 2006; Moreno *et al.* 2016, 2017).

Outside of *C. elegans*, *P. pacificus* is arguably the most advanced nematode system in terms of the availability of

molecular tools (Schlager *et al.* 2009; Witte *et al.* 2015; Okumura *et al.* 2017; Loer *et al.* 2019). However, previous methodologies resulted in low efficiencies of *P. pacificus* transgenics, with on average one to three F1 Roller(s) per 40 injected P0s (Schlager *et al.* 2009). Thus, *P. pacificus* suffers from a much less efficient transgenesis system compared with *C. elegans* for several potential reasons. First, it relies on the formation of complex arrays, which incorporate transgene DNA, genomic DNA fragments that must come from *P. pacificus* itself, and a co-injection marker, to be carried as heritable chromosome fragments. Second, the current versions of fluorescent proteins utilized in *P. pacificus* (Schlager *et al.* 2009) have not been adapted to its specific CUB and no attempts have yet been made to improve these fluorescent proteins further by investigating any potential IME. Together, these factors likely contribute to the varying degrees of generational transmission observed in *P. pacificus* transgenesis experiments and will likely hinder the successful development of other transgenic techniques including additional fluorescent proteins, calcium imaging, and optogenetics.

In this study, using publicly available data sets, we first computed the CUB and global intron structure in 10 nematode species to investigate the conservation of these factors across the phylum, making use of the most recent genomic and transcriptomic data sets. For this, we selected species living in different ecosystems including parasites of animals and plants. As each nematode species shows a distinct CUB and potential IME, we focused on *P. pacificus* and utilized these factors together with its spliced leaders (SLs), a specific but conserved transcriptional regulatory element in nematodes (Denker *et al.* 2002), to improve the efficiency of transgenesis in this species. Finally, with the improved transgenesis in *P. pacificus*, we established a new method using a fluorescent co-injection marker to identify potential clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-edited candidates, reducing the workload and cost for CRISPR/Cas9 screening.

Materials and Methods

Obtaining genome annotations and transcription profiles

We collected published annotations and transcriptomes of 10 nematode species representing three of the five major nematode clades (Blaxter *et al.* 1998): *C. elegans* (Lee *et al.* 2018; Liu *et al.* 2019) [WormBase web site (<https://wormbase.org>), release WS271 2019], *C. briggsae* (Grün *et al.* 2014), *Haemonchus contortus* (Laing *et al.* 2013), *P. pacificus* (Prabh *et al.* 2018; Rödelberger *et al.* 2019), *P. fissidentatus* (Prabh *et al.* 2018; Rödelberger *et al.* 2018), *Strongyloides ratti* (Hunt *et al.* 2016), *Globodera pallida* (Cotton *et al.* 2014), *Bursaphelenchus xylophilus* (Kikuchi *et al.* 2011; Tanaka *et al.* 2019), *Brugia malayi* (Choi *et al.* 2011; Foster *et al.* 2020), and *Ascaris suum* (Wang *et al.* 2011, 2017). To acquire the expression profiles of *P. pacificus* and *P. fissidentatus*, we retrieved RNA-sequencing (RNA-seq) data sets of *P. pacificus* and *P. fissidentatus* from the Sequence

Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra/>; Supplemental Material, Table S1). We mapped raw reads to the reference genome of each species (Table S1) using Hisat2 (Kim *et al.* 2015) with default parameters, and quantified the numbers of reads mapping to each annotated gene using the package featureCounts (Liao *et al.* 2014). For other species, the gene expression results were directly downloaded from WormBase (Howe *et al.* 2016, 2017). Detailed information of the metadata is summarized in Table S1. For *C. elegans* and *A. suum*, whose annotations included isoform data, only the longest transcripts were used in downstream analyses.

Codon usage computation

To identify the CUBs of genes with different expression levels, the percentage codon usage for each gene was calculated using cusp from EMBOSS suite (Rice *et al.* 2000). We optimized the codon of proteins based on the most preferred codons of genes with high expression levels in *P. pacificus*.

P. pacificus trans-spliced messenger RNA identification

To identify the *P. pacificus* transcripts that contain SLs, we first performed RNA-seq using a ribosomal RNA (rRNA) depletion library. Briefly, total RNA of *P. pacificus* was extracted via Direct-zol RNA Miniprep (Zymo Research) and a Ribo-Zero rRNA Removal Kit (Human/Mouse/Rat; Illumina), and RNA libraries were constructed using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina). Sequencing was carried out on an Illumina HiSeq 3000 sequencer with one-sixth of a lane. We used Trinity (Grabherr *et al.* 2011) for *de novo* transcriptome assembly, and identified the transcripts with SLs by the consensus SL sequences at 5' ends (SL1: TACC CAAGTTTGAG; and SL2: CAGTATCTCAAG) (Guiliano and Blaxter 2006). We used MEME SUITE (Bailey *et al.* 2009) to identify the motifs of 3' sequences of the *trans*-splice sites.

Statistics

We performed the chi-square test to test whether the frequencies of synonymous codons in the most highly expressed genes (11th bin) were deviated from the frequencies of genome-wide synonymous codons. We performed the one-tailed Kolmogorov–Smirnov test to compare the intron length distributions between *C. elegans* and the other species. We calculated the Pearson's correlation coefficient to measure the linear relationship between intron length and gene expression level. We performed the Wilcoxon signed-rank test to test whether genes that contained SLs were different in expression level from genes without SLs.

Plasmid construction and microinjection

The optimized *egl-20p::GFP* and *egl-20p::TurboRFP* (red fluorescent protein) were modified based on a pUC19 backbone from a previous study (Schlager *et al.* 2009). Full sequences of these plasmids in text files can be found in the supplemental materials. Modified GFP and TurboRFP sequences were synthesized from Integrated DNA Technologies (IDT; Coralville, IA) and cloned into the pUC19 backbone using Gibson

Assembly Master Mix (New England Biolabs, Beverly, MA) following the manufacturers' protocols. Plasmids were extracted using the QIAprep Spin Miniprep kit (QIAGEN, Valencia, CA).

Three introns from the rRNA gene *Ppa-rps-1* (gene ID: PPA18896; El paco annotation_v2) (Rödelsperger *et al.* 2019) were added into the sequence of GFP or turboRFP from 5' to 3' and were roughly evenly spaced ("Fire Lab Vector Kit 1995"): intron 1, gtgagcattcttgggttgatgggggttgaaaacttcatgggattcctaacctattaatttttcag; intron 2, gtaagtcgtatacattgagggtgtctttacgtgatatccggggttggtttgagagaggagatatttataataaatataatttcag; and intron 3, gtgagtgctgcaaatattaagtacatgaaacttttctcag. For the two-intron codon-optimized *egl-20p::GFP* and *egl-20p::TurboRFP*, intron 1 (the most 5' intron) was removed using a Q5 Site-Directed Mutagenesis kit (New England Biolabs), and both intron 1 and 2 were removed for the one-intron *egl-20p::GFP* and *egl-20p::TurboRFP*.

P. pacificus microinjections were performed following the standard protocol (Schlager *et al.* 2009; Witte *et al.* 2015). Plasmids were diluted to 50 ng/μl for microinjection using TE buffer. Well-fed *P. pacificus* (strain PS312) young hermaphrodites (preferably not carrying any eggs) were used for injections. The injection mix for the co-injection marker-assisted CRISPR/Cas9 editing was modified from those of Witte *et al.* (2015) and Dokshin *et al.* (2018), and the mix contained 0.5 μg/μl Cas9 nuclease (catalog# 1081058; IDT), 0.1 μg/μl *trans*-activating CRISPR RNA (catalog# 1072534; IDT), 0.056 μg/μl guide RNA (CRISPR/Cas9 RNA; IDT), and 0.05 μg/μl co-injecting plasmid. Potential CRISPR-edited alleles were amplified by PCR and sequenced using Sanger sequencing. Alternatively, the PCR amplicons were run on a 4% TBE agarose gel to detect heteroduplex formation (Bhattacharya and Van Meir 2019).

Data availability

The raw sequence data of *P. pacificus* rRNA-depleted RNA-seq have been deposited at the SRA under BioProject identified PRJNA658248. Supplemental material available at figshare: <https://doi.org/10.25386/genetics.13090322>.

Results

Codon usage is divergent among nematode species

To enhance transgene expression in diverse nematode species, we first obtained a comprehensive view of CUB in nematodes. For that, we calculated CUB in 10 species of eight nematode genera, representing three of the five major clades of the phylum Nematoda (Figure 1A and Figure S1; Blaxter *et al.* 1998). For a given amino acid, we found a favored codon in highly expressed genes in every species. The frequencies of different codons in the most highly expressed genes (11th bin) deviate significantly from their genome-wide frequencies (for all comparisons, $0 < P < 4.19 \times 10^{-11}$, chi-square test). There was no clear pattern between CUBs and phylogenetic relationships or lifestyles (free-living or parasitic). However, genome-wide GC content may be

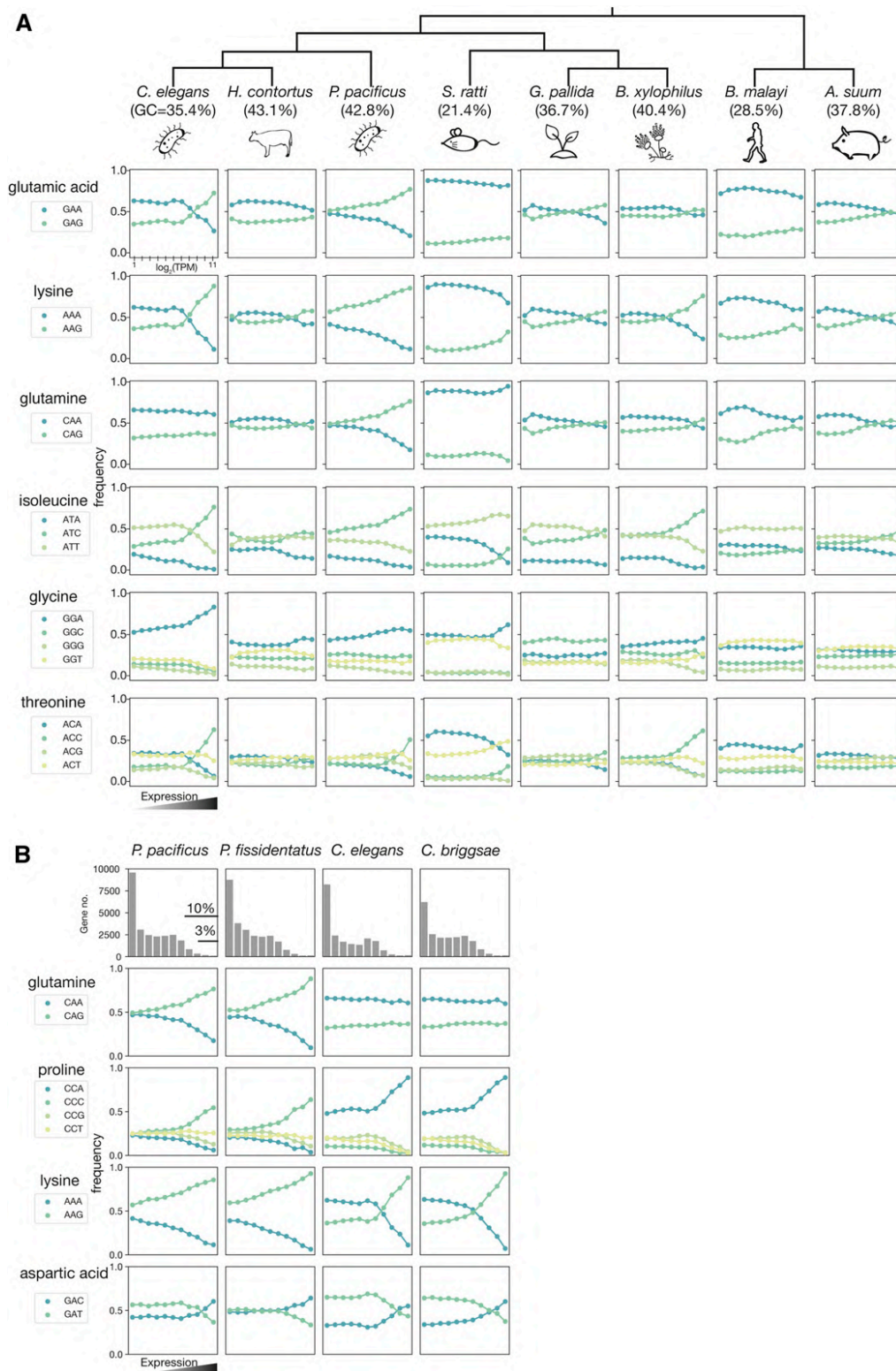


Figure 1 Codon preferences in nematodes species as a function of expression levels. (A) The codon usage bias of *C. elegans*, *H. contortus*, *P. pacificus*, *S. ratti*, *G. pallida*, *B. xylophilus*, *B. malayi*, and *A. suum*. The protein-coding genes are binned based on the transcripts per kilobase million value from expression level low to high with a \log_2 scale into 11 bins (x -axis). The dots represent the average codon usage frequency of a given bin. (B) Gene grouping and codon usage bias for *P. pacificus*, *P. fissidentatus*, *C. elegans*, and *C. briggsae*. Figures in the first row show the number of genes (y -axis) grouped from low to high expression with a \log_2 scale into 11 bins (same as A).

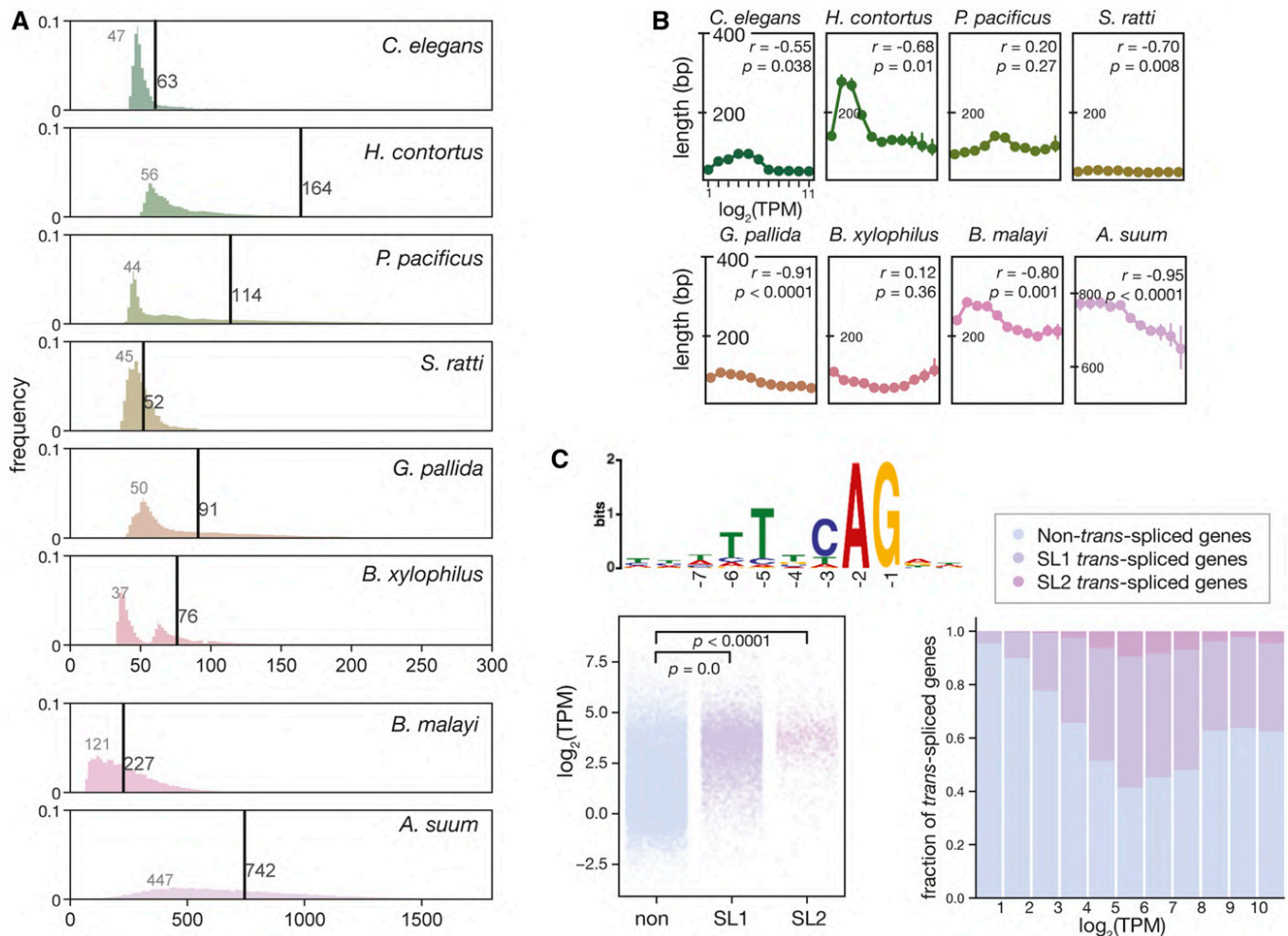


Figure 2 Global intron length distribution in diverse nematodes. (A) The intron length distribution of eight nematode species. Vertical lines indicate the median lengths of introns of each species, while the numbers in gray indicate the modes of intron lengths (bp). (B) Median intron length as a function of the gene expression. The protein-coding genes are binned by expression level from low to high with a log₂ scale into 11 bins. (C) Elevated expression level of SL1-operated genes. Consensus sequence of the SL1 *trans*-splice sites in *P. pacificus* (top). SL1- and SL2-spliced genes have a higher expression level than those that are not *trans*-spliced (left). The proportion of *trans*-spliced genes is positively associated with expression level (right). SL1, spliced leader 1; TPM, transcripts per kilobase million.

one major factor correlating with the CUB (Mitreva *et al.* 2006). For example, *S. ratti* and *B. malayi* have low-GC-content genomes, and subsequently the codon usage is also biased toward AT-rich codons. In these species, GCA and GCT are more preferred than GCC and GCG for alanine. Intriguingly, species with a similar GC content can still exhibit drastically different patterns in CUB. For example, *P. pacificus* and *H. contortus*, which both have ~43% GC content, show differing codon preferences for coding proline and alanine (Figure S1). Thus, our new analysis of CUB confirms previous studies that codon usage is divergent among nematode species. Note that the species considered here belong to very different nematode taxa and are phylogenetically only distantly related.

Codon usage adaptation is conserved within genera

To study the evolution of codon usage between more closely related nematodes, we focused on two well-studied nematode genera *Caenorhabditis* and *Pristionchus*. *C. elegans* and *P.*

pacificus share a common ancestor around 100 million years ago, and they have a distinct CUB. The most dramatic examples of this can be seen in the amino acids glutamine, glutamic acid, and lysine, where *P. pacificus* and *C. elegans* favor the opposing codons. However, within the genus *Pristionchus* the CUB appears conserved, as in *P. fissidentatus*, a basal species in the *Pristionchus* genus (Rödelsperger *et al.* 2018), and we found it shares a highly similar CUB with that observed in *P. pacificus* (Figure 1B and Figure S2). Similarly, in *Caenorhabditis*, the CUB is conserved between *C. elegans* and *C. briggsae* (Figure S3). This finding strongly suggests that codon usage adaptation evolved more ancestrally than the speciation events within the genera *Pristionchus* and *Caenorhabditis*, and that CUB is conserved between closely related species.

Global intron structure and SL1 frequency shows distinct patterns

As the presence and distributions of introns also contribute to gene regulation (Castillo-Davis *et al.* 2002), we next

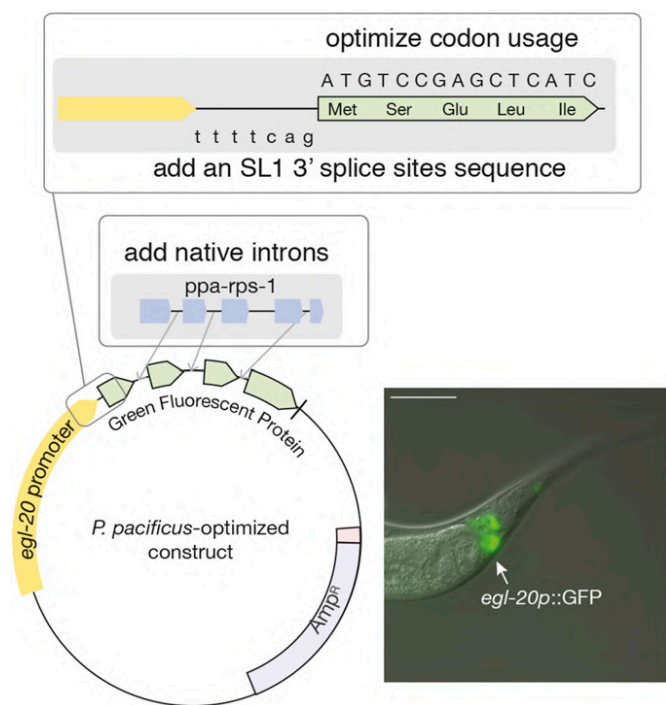


Figure 3 Optimized transgenic plasmids based on *P. pacificus* genomic features. An illustration of the construct structure for codon-optimized and native intron addition in *egl-20p::GFP/turboRFP* (left). An overlay of DIC and GFP image of *egl-20p::GFP* (left). Bar, 50 μm . AMP^R, ampicillin resistance; GFP, green fluorescent protein.

investigated global intron composition to potentially understand the IME of genes across nematodes. Unexpectedly, introns across eight annotated genomes showed distinct features in terms of the general pattern observed (for all comparisons $P = 0.0$, Kolmogorov–Smirnov test) and the median intron size. When intron length was plotted by frequency, a unimodal pattern was detected in *C. elegans*, *H. contortus*, *P. pacificus*, and *S. ratti*, whereas a bimodal pattern was observed in *B. xylophilus* (Figure 2A). Further, in the clade three nematodes (*B. malayi* and *A. suum*) intron length appeared to be much longer and with a wider distribution.

Further analysis of the average intron length revealed that the introns of *C. elegans* have a mode (the most abundant number) of 47 nt and a median of 63 nt in length (Figure 2A), while the distantly related *S. ratti* has even shorter introns with a median of 52 nt and a more homogeneous distribution. In *H. contortus*, *P. pacificus*, *B. xylophilus*, and *G. pallida*, the distribution of intron length shows a greater range compared with *C. elegans*, although an accumulation of introns with a size between 40 and 60 nt is also detectable. When comparing intron size with gene expression level, we found that in *C. elegans* ($r = -0.55$, $P = 0.038$, Pearson's correlation), *H. contortus* ($r = -0.68$, $P = 0.010$, Pearson's correlation), *S. ratti* ($r = -0.70$, $P = 0.008$, Pearson's correlation), *G. pallida* ($r = -0.91$, $P < 0.001$, Pearson's correlation), *B. malayi* ($r = -0.91$, $P = 0.001$, Pearson's correlation), and *A. suum* ($r = -0.95$, $P < 0.001$, Pearson's correlation), the

intron size was negatively correlated with gene expression level. However, this correlation was not observed in the other species (Figure 2B).

Finally, we investigated another gene regulatory element, SL1. Nematodes have a specific *trans*-splicing mechanism at the 5' end of many premature messenger RNAs (mRNAs), which is trimmed and replaced by an SL sequence (Denker *et al.* 2002). This mechanism is thought to increase translation (Yang *et al.* 2017). Using a *P. pacificus* rRNA-depleted RNA-seq library instead of deeply sequenced mRNA-enriched RNA-seq data sets, which are traditionally used in *C. elegans* (Allen *et al.* 2011), we identified a total of 5982 genes in *P. pacificus* that were SL1-operated, and 922 genes that were SL2-operated. These genes have an SL1 3' splice site with a consensus sequence "TTTCAG" (Figure 2C), which is also conserved in *C. elegans* (Yang *et al.* 2017). Globally, higher expression levels were observed in *P. pacificus* genes associated with SL1 compared with genes without splicing leaders ($P = 0.0$, Wilcoxon signed-rank test). Therefore, this suggests that SL1 increases translation in *P. pacificus*, a similar phenomenon to that observed in *C. elegans* (Yang *et al.* 2017). While the published nematode data sets are not sufficient for us to survey the SL1 *trans*-spliced genes of other nematode species, given the fact that the sequences of SLs are conserved among nematodes (Guiliano and Blaxter 2006), the SL *trans*-splicing could be a highly conserved mechanism in the Nematoda phylum.

Optimization of GFP and TurboRFP sequences and increased transgenesis efficiency

With our observations of the large variations in CUB and potential IME regulating gene expression across nematodes, we decided to focus on a single species and attempt to improve its transgenesis efficiency. Therefore, we focused on establishing two fluorescent proteins for use in the free-living nematode *P. pacificus*. These were based on the previously utilized TurboRFP (Schlager *et al.* 2009) and on GFP (Fire Lab Vector Kit 1995), which are commonly utilized across the *C. elegans* community. In *P. pacificus*, TurboRFP has been used to successfully produce transgenic lines; however, this was only at a low transmission efficiency. The GFP previously used in *P. pacificus* was optimized according to *C. elegans*' CUB and hardly generated detectable fluorescence. Therefore, we replaced the codons in these two fluorescent proteins with two sets of codon usages: the CUB found associated with the top 10% most highly expressed genes and with the top 3% most highly expressed genes of *P. pacificus* (Table S2). Alongside this, we also attempted to optimize both fluorescent proteins further through the addition of native introns to increase its transcription. We selected the native introns of the gene *Ppa-rps-1* as it is highly expressed through all life stages and has four relatively short introns. The three shorter introns of *Ppa-rps-1* were added into the reading frame of the codon-optimized GFP and TurboRFP. Finally, we added an SL1 3' splice site sequence immediately upstream of the start codon of both fluorescent proteins (illustrated in Figure 3).

Table 1 Improved transgenesis efficiency using transcriptional reporter constructs with codon optimization and intron addition in *P. pacificus*

Construct	Number of introns	Injected P0s	Number of P0s with fluorescent F1s	Efficiency (%)
<i>egl-20p::GFP</i>	3	49	11	22 ($P = 0.02$)
<i>egl-20p::TurboRFP</i>	3	55	16	29 ($P = 0.003$)
<i>egl-20p::GFP</i>	2	40	4	10 ($P = 0.39$)
<i>egl-20p::GFP</i>	1	12	0	0 ($P = 0.48$)
<i>egl-20p::TurboRFP</i>	1	18	0	0 ($P = 0.34$)
<i>Ppa-prl-1</i> ^a	NA	NA	NA	5

The GFP and TurboRFP sequences were optimized using *P. pacificus* favored codons (from top 10% highly expressed) with addition of native introns.

^a Data from Schlager *et al.* 2009, summarized from over 3000 P0 injections. Chi-square tests were performed between *Ppa-prl-1* and optimized constructs.

In a first set of experiments, we performed all three optimization steps (CUB, native intron addition, and SL1 3' splice site sequence) simultaneously and used the previously established *egl-20* promoter to drive fluorescent protein expression. We were able to obtain GFP transcriptional reporter lines with robust and intense signals (Figure 3). More importantly, we considerably improved the efficiency of transgenesis of both GFP (PZH008) and TurboRFP (PZH009) constructs ($P = 0.02$ and $P = 0.003$, respectively; Table 1). Note that we still experienced variability in the efficiency; possibly due to factors such as injector and age of the specimen, the efficiency increased to >20% of injected animals. While we did not systematically test all variables individually due to the enormous costs that would have been associated with such studies, we confirmed the increase in efficiency by the subsequent removal of introns. Indeed, intron removal coincided with a decrease in transgenic efficiency (Table 1). Together, we found that the codon-optimized three-intron GFP and TurboRFP had greater efficiency compared with the previous nonoptimized TurboRFP (Schlager *et al.* 2009). However, for unknown reasons, utilizing the CUB of

the top 3% highly expressed genes did not further increase the efficiency (Table S3).

Fluorescent co-injection marker-assisted CRISPR genome editing

With the establishment of reliable and robust transgenic markers in *P. pacificus*, we next attempted to implement these tools to reduce the workload and the cost of screening potential CRISPR/Cas9 alleles. Therefore, we tried to establish a method that employed the optimized fluorescent markers to identify potential mutants induced with CRISPR/Cas9 (Figure 4A). A fluorescent marker can indicate well-injected specimens, which carry an increased likelihood of successfully induced CRISPR/Cas9 mutations. Therefore, using the *egl-20p::TurboRFP* (PZH009) as a CRISPR/Cas9 co-injection marker, our experienced injectors obtained between 1 and 5 P0s (on average 2.5) producing RFP-positive F1 progeny from 30 well-injected nematodes (Figure 4B). Furthermore, progeny cooccurring on RFP injection marker-positive plates also frequently carried CRISPR/Cas9-induced mutations at high efficiency (77% of the identified plates),

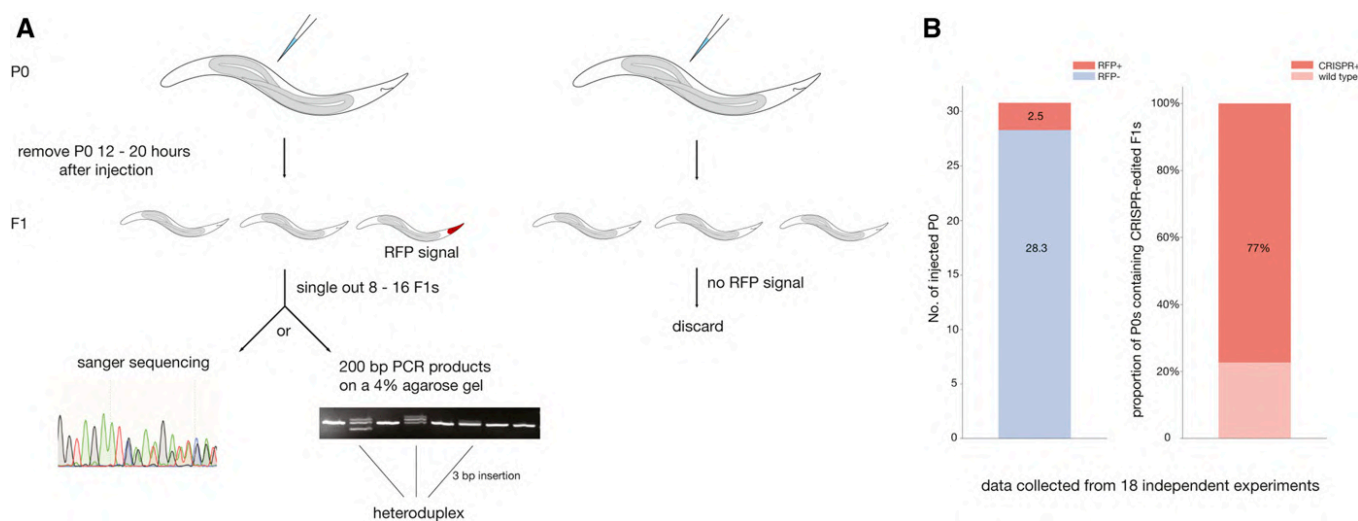


Figure 4 Newly established fluorescent co-injection marker-assisted CRISPR genome editing in *P. pacificus*. (A) An illustration of the workflow for CRISPR genome editing in *P. pacificus*. (B) Using an *egl-20p::TurboRFP* construct (PZH009) as a co-injection marker, an average of 2.5 P0s had RFP + F1s from 30 well-injected *P. pacificus* P0s (left). (B) Next, 8–16 F1s were selected from each P0 with RFP + F1s to detect CRISPR alleles. There was a 77% chance that the P0-contained RFP + F1s also contained CRISPR-edited F1s (right). These data were accumulated from 18 independent experiments. Note, there were two additional experiments with no RFP + F1 detected, but CRISPR editing still occurred. CRISPR, clustered regularly interspaced short palindromic repeats; RFP, red fluorescent protein.

allowing the number of progeny necessary to be screened to isolate a CRISPR/Cas9 mutant to be greatly reduced (Figure 4B). Thus, the improved fluorescent-based co-injection marker strongly assisted the detection of CRISPR-generated edits in *P. pacificus*. We would like to note here that this fluorescent marker-assisted CRISPR method is compatible with knockouts and shorter repair templates (<120 nt), but does not seem to work with longer repair templates.

Discussion

The usage of transgenic tools is fundamental to successful studies in molecular biology; however, their efficiency is not uniform between organisms. This is, in part, likely due to differences in gene regulatory mechanisms between different species. In canonical model organisms, the development of efficient transgenic tools is aided by the existence of large scientific communities capable of refining and optimizing their application; however, this is not usually possible in other systems. Although the delivery of DNA to the germline can be an obstacle in nematode species (Kranse *et al.* 2020), delivery via microinjection is not a hindrance in *Pristionchus*, since we have generally achieved a higher efficiency for CRISPR knockouts compared with transgenesis. Here, we have revealed large differences in CUB and IME across nematodes, which likely contribute to gene regulatory differences between species. As a proof of principle, we investigated a single nematode species, *P. pacificus*, whereby we have successfully exploited its favored CUB and IME to develop *P. pacificus*-adapted fluorescent transgenic proteins. Additionally, we have shown that these adapted proteins containing *P. pacificus* gene regulatory requirements demonstrate a dramatically increased expression efficiency. It has recently been shown in *C. elegans* that the 5' intron contributes the most to the elevated level of gene expression (Crane *et al.* 2019). Our results in *P. pacificus* agree with this finding because transgenesis efficiency decreased when the 5' intron was removed. Transgenes with constructs that were modified using the top 3% CUB did not further improve efficiency. We can only speculate that this might be due to the most highly favored codons causing ribosomal traffic jams (Plotkin and Kudla 2011). Nevertheless, these improvements allow transgenes to be utilized as co-injection markers to reduce the screening time and costs of CRISPR/Cas9 genome editing. Thus, our method provides an alternative to the existing *Pristionchus* co-CRISPR method, in which the identification of CRISPR candidates relies on a Dpy phenotype (Nakayama *et al.* 2020).

By means of an initial bioinformatic analysis of the species-specific CUB and IME, our experiments demonstrate the potential to develop optimized transgenic tools and explore distinctive attributes that were not previously possible. While we did not systematically test the specific contributions of CUB, IME, and SL1, they likely all play important roles in transcription and translation for the increased transgenesis efficiency in *P. pacificus* (Redemann *et al.* 2011; Yang *et al.*

2017; Crane *et al.* 2019), but it is important to note that this principle could be further utilized to optimize genetically encoded calcium indicators and optogenetic tools to explore *Pristionchus*-specific behaviors, and genetic ablation methods to investigate aspects of anatomy and physiology. We hypothesize that, using knowledge of species-specific genomic features, it is possible to establish transgenic tool kits in other free-living nematodes, and additionally in parasitic nematode systems that have a significant impact on world health (Brindley *et al.* 2009) and crop production (Nicol *et al.* 2011).

Acknowledgments

The authors would like to thank the members of the Sommer laboratory for collecting CRISPR editing data. Adrian Streit provided insightful thoughts about this manuscript. This work was supported by the Max Planck Society (funding awarded to R.J.S.); an Alexander von Humboldt Foundation Postdoctoral fellowship (awarded to Z.H.), and a Chinese Scholarship Council Ph.D. fellowship (awarded to S.S.). Construct maps in Snapgene files are available upon request.

Literature Cited

- Adams, S., P. Pathak, H. Shao, J. B. Lok, and A. Pires-daSilva, 2019 Liposome-based transfection enhances RNAi and CRISPR-mediated mutagenesis in non-model nematode systems. *Sci. Rep.* 9: 483. <https://doi.org/10.1038/s41598-018-37036-1>
- Akduman, N., J. W. Lightfoot, W. Röseler, H. Witte, W.-S. Lo *et al.*, 2020 Bacterial vitamin B₁₂ production enhances nematode predatory behavior. *ISME J.* 14: 1494–1507 (erratum: *ISME J.* 14: 1911). <https://doi.org/10.1038/s41396-020-0626-2>
- Allen, M. A., L. W. Hillier, R. H. Waterston, and T. Blumenthal, 2011 A global analysis of *C. elegans* trans-splicing. *Genome Res.* 21: 255–264. <https://doi.org/10.1101/gr.113811.110>
- Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant *et al.*, 2009 MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37: W202–W208. <https://doi.org/10.1093/nar/gkp335>
- Bhattacharya, D., and E. G. Van Meir, 2019 A simple genotyping method to detect small CRISPR-Cas9 induced indels by agarose gel electrophoresis. *Sci. Rep.* 9: 4437. <https://doi.org/10.1038/s41598-019-39950-4>
- Bischof, J., R. K. Maeda, M. Hediger, F. Karch, and K. Basler, 2007 An optimized transgenesis system for *Drosophila* using germ-line-specific C31 integrases. *Proc. Natl. Acad. Sci. USA* 104: 3312–3317. <https://doi.org/10.1073/pnas.0611511104>
- Blaxter, M. L., P. De Ley, J. R. Garey, L. X. Liu, P. Scheldeman *et al.*, 1998 A molecular evolutionary framework for the phylum Nematoda. *Nature* 392: 71–75. <https://doi.org/10.1038/32160>
- Brindley, P. J., M. Mitreva, E. Ghedin, and S. Lustigman, 2009 Helminth genomics: the implications for human health. *PLoS Negl. Trop. Dis.* 3: e538. <https://doi.org/10.1371/journal.pntd.0000538>
- Brinster, R. L., J. M. Allen, R. R. Behringer, R. E. Gelinas, and R. D. Palmiter, 1988 Introns increase transcriptional efficiency in transgenic mice. *Proc. Natl. Acad. Sci. USA* 85: 836–840. <https://doi.org/10.1073/pnas.85.3.836>

- Bui, L. T., N. A. Ivers, and E. J. Ragsdale, 2018 A sulfotransferase dosage-dependently regulates mouthpart polyphenism in the nematode *Pristionchus pacificus*. *Nat. Commun.* 9: 4119. <https://doi.org/10.1038/s41467-018-05612-8>
- Castillo-Davis, C. I., S. L. Mekhedov, D. L. Hartl, E. V. Koonin, and F. A. Kondrashov, 2002 Selection for short introns in highly expressed genes. *Nat. Genet.* 31: 415–418. <https://doi.org/10.1038/ng940>
- Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher, 1994 Green fluorescent protein as a marker for gene expression. *Science* 263: 802–805. <https://doi.org/10.1126/science.8303295>
- Choi, Y.-J., E. Ghedin, M. Berriman, J. McQuillan, N. Holroyd *et al.*, 2011 A deep sequencing approach to comparatively analyze the transcriptome of lifecycle stages of the filarial worm, *Brugia malayi*. *PLoS Negl. Trop. Dis.* 5: e1409. <https://doi.org/10.1371/journal.pntd.0001409>
- Clough, S. J., and A. F. Bent, 1998 Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16: 735–743. <https://doi.org/10.1046/j.1365-313x.1998.00343.x>
- Cotton, J. A., C. J. Lilley, L. M. Jones, T. Kikuchi, A. J. Reid *et al.*, 2014 The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biol.* 15: R43. <https://doi.org/10.1186/gb-2014-15-3-r43>
- Crane, M. M., B. Sands, C. Battaglia, B. Johnson, S. Yun *et al.*, 2019 In vivo measurements reveal a single 5'-intron is sufficient to increase protein expression level in *Caenorhabditis elegans*. *Sci. Rep.* 9: 9192. <https://doi.org/10.1038/s41598-019-45517-0>
- Denker, J. A., D. M. Zuckerman, P. A. Maroney, and T. W. Nilsen, 2002 New components of the spliced leader RNP required for nematode trans-splicing. *Nature* 417: 667–670. <https://doi.org/10.1038/nature00783>
- Dokshin, G. A., K. S. Ghanta, K. M. Piscopo, and C. C. Mello, 2018 Robust genome editing with short single-stranded and long, partially single-stranded DNA donors in *Caenorhabditis elegans*. *Genetics* 210: 781–787. <https://doi.org/10.1534/genetics.118.301532>
- Duret, L., and D. Mouchiroud, 1999 Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 96: 4482–4487. <https://doi.org/10.1073/pnas.96.8.4482>
- Evans, T. C., 2006 Transformation and microinjection (April 6, 2006), *WormBook*, ed. The *C. elegans* Research Community *WormBook*, doi:10.1895/wormbook.1.108.1, <http://www.wormbook.org>. <https://doi.org/10.1895/wormbook.1.108.1>
- Fire, A. Fire Lab *C. elegans* Vector Kit 1995. <https://media.addgene.org/cms/files/Vec95.pdf>
- Foster, J. M., A. Grote, J. Mattick, A. Tracey, Y.-C. Tsai *et al.*, 2020 Sex chromosome evolution in parasitic nematodes of humans. *Nat. Commun.* 11: 1964. <https://doi.org/10.1038/s41467-020-15654-6>
- Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29: 644–652. <https://doi.org/10.1038/nbt.1883>
- Grün, D., M. Kirchner, N. Thierfelder, M. Stoeckius, M. Selbach *et al.*, 2014 Conservation of mRNA and protein expression during development of *C. elegans*. *Cell Rep.* 6: 565–577. <https://doi.org/10.1016/j.celrep.2014.01.001>
- Guiliano, D. B., and M. L. Blaxter, 2006 Operon conservation and the evolution of trans-splicing in the phylum Nematoda. *PLoS Genet.* 2: e198. <https://doi.org/10.1371/journal.pgen.0020198>
- Higazi, T. B., A. Merriweather, L. Shu, R. Davis, and T. R. Unnasch, 2002 *Brugia malayi*: transient transfection by microinjection and particle bombardment. *Exp. Parasitol.* 100: 95–102. [https://doi.org/10.1016/S0014-4894\(02\)00004-8](https://doi.org/10.1016/S0014-4894(02)00004-8)
- Hong, R. L., and R. J. Sommer, 2006 *Pristionchus pacificus*: a well-rounded nematode. *Bioessays* 28: 651–659. <https://doi.org/10.1002/bies.20404>
- Howe, K. L., B. J. Bolt, S. Cain, J. Chan, W. J. Chen *et al.*, 2016 WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res.* 44: D774–D780. <https://doi.org/10.1093/nar/gkv1217>
- Howe, K. L., B. J. Bolt, M. Shafie, P. Kersey, and M. Berriman, 2017 WormBase ParaSite - a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* 215: 2–10. <https://doi.org/10.1016/j.molbiopara.2016.11.005>
- Hunt, V. L., I. J. Tsai, A. Coghlan, A. J. Reid, N. Holroyd *et al.*, 2016 The genomic basis of parasitism in the Strongyloides clade of nematodes. *Nat. Genet.* 48: 299–307. <https://doi.org/10.1038/ng.3495>
- Hutter, H., 2012 Fluorescent protein methods: strategies and applications. *Methods Cell Biol.* 107: 67–92. <https://doi.org/10.1016/B978-0-12-394620-1.00003-5>
- Kieninger, M. R., N. A. Ivers, C. Rödelberger, G. V. Markov, R. J. Sommer *et al.*, 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: 2174–2179. <https://doi.org/10.1016/j.cub.2016.06.018>
- Kikuchi, T., J. A. Cotton, J. J. Dalzell, K. Hasegawa, N. Kanzaki *et al.*, 2011 Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathog.* 7: e1002219. <https://doi.org/10.1371/journal.ppat.1002219>
- Kim, D., B. Langmead, and S. L. Salzberg, 2015 HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12: 357–360. <https://doi.org/10.1038/nmeth.3317>
- Kranse O., H. Beasley, S. Adams, A. P. da Silva, C. Bell, *et al.*, 2020 Towards genetic modification of plant-parasitic nematodes: delivery of macromolecules to adults and expression of exogenous mRNA in second stage juveniles. *bioRxiv*. doi.org/10.1101/2020.07.15.193052 (Preprint posted July 15, 2020).
- Laing, R., T. Kikuchi, A. Martinelli, I. J. Tsai, R. N. Beech *et al.*, 2013 The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery. *Genome Biol.* 14: R88. <https://doi.org/10.1186/gb-2013-14-8-r88>
- Lee, R. Y. N., K. L. Howe, T. W. Harris, V. Arnaboldi, S. Cain *et al.*, 2018 WormBase 2017: molting into a new stage. *Nucleic Acids Res.* 46: D869–D874. <https://doi.org/10.1093/nar/gkx998>
- Liao, Y., G. K. Smyth, and W. Shi, 2014 featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930. <https://doi.org/10.1093/bioinformatics/btt656>
- Lightfoot, J. W., M. Wilecki, C. Rödelberger, E. Moreno, V. Susoy *et al.*, 2019 Small peptide-mediated self-recognition prevents cannibalism in predatory nematodes. *Science* 364: 86–89. <https://doi.org/10.1126/science.aav9856>
- Li, X., H. C. Massey, Jr., T. J. Nolan, G. A. Schad, K. Kraus *et al.*, 2006 Successful transgenesis of the parasitic nematode *Strongyloides stercoralis* requires endogenous non-coding control elements. *Int. J. Parasitol.* 36: 671–679. <https://doi.org/10.1016/j.ijpara.2005.12.007>
- Liu, Y., K. G. Kaval, A. van Hoof, and D. A. Garsin, 2019 Heme peroxidase HPX-2 protects *Caenorhabditis elegans* from pathogens. *PLoS Genet.* 15: e1007944. <https://doi.org/10.1371/journal.pgen.1007944>
- Loer, C., H. Witte, R. Sommer, and O. Hobert, 2019 An antibody staining protocol variation for nematodes that adds heat-induced antigen retrieval (HIAR). *MicroPubl Biol* 2019. Available at:

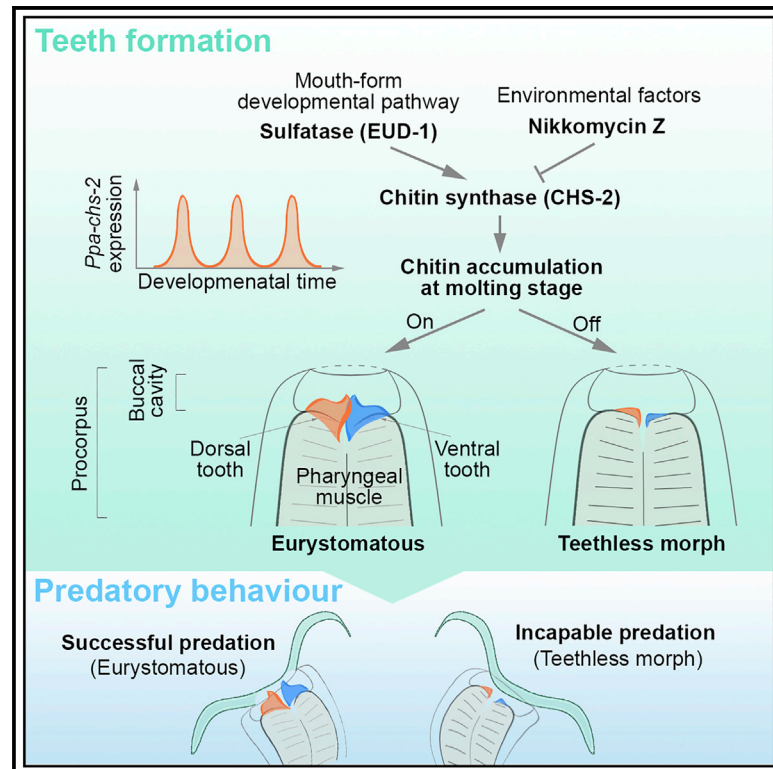
- <https://www.micropublication.org/journals/biology/micropub.biology.000135/>. <https://doi.org/10.17912/micropub.biology.000135>
- Lok, J. B., 2012 Nucleic acid transfection and transgenesis in parasitic nematodes. *Parasitology* 139: 574–588. <https://doi.org/10.1017/S0031182011001387>
- Mello, C. C., J. M. Kramer, D. Stinchcomb, and V. Ambros, 1991 Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* 10: 3959–3970. <https://doi.org/10.1002/j.1460-2075.1991.tb04966.x>
- Mitreva, M., M. C. Wendl, J. Martin, T. Wylie, Y. Yin *et al.*, 2006 Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Biol.* 7: R75. <https://doi.org/10.1186/gb-2006-7-8-r75>
- Moreno, E., A. McGaughran, C. Rödelberger, M. Zimmer, and R. J. Sommer, 2016 Oxygen-induced social behaviours in *Pristionchus pacificus* have a distinct evolutionary history and genetic regulation from *Caenorhabditis elegans*. *Proc. Biol. Sci.* 283: 20152263. <https://doi.org/10.1098/rspb.2015.2263>
- Moreno, E., B. Sieriebriennikov, H. Witte, C. Rödelberger, J. W. Lightfoot *et al.*, 2017 Regulation of hyperoxia-induced social behaviour in *Pristionchus pacificus* nematodes requires a novel cilia-mediated environmental input. *Sci. Rep.* 7: 17550. <https://doi.org/10.1038/s41598-017-18019-0>
- Moreno, E., J. W. Lightfoot, M. Lenuzzi, and R. J. Sommer, 2019 Cilia drive developmental plasticity and are essential for efficient prey detection in predatory nematodes. *Proc. Biol. Sci.* 286: 20191089. <https://doi.org/10.1098/rspb.2019.1089>
- Nakayama, K.-I., Y. Ishita, T. Chihara, and M. Okumura, 2020 Screening for CRISPR/Cas9-induced mutations using a co-injection marker in the nematode *Pristionchus pacificus*. *Dev. Genes Evol.* 230: 257–264. <https://doi.org/10.1007/s00427-020-00651-y>
- Nicol, J. M., S. J. Turner, D. L. Coyne, L. den Nijs, S. Hockland *et al.*, 2011 Current nematode threats to world agriculture, pp. 21–43 in *Genomics and Molecular Genetics of Plant-Nematode Interactions*, edited by J., Jones, G. Gheysen, and C. Fenoll. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-007-0434-3_2
- Okkema, P. G., S. W. Harrison, V. Plunger, A. Aryana, and A. Fire, 1993 Sequence requirements for myosin gene expression and regulation in *Caenorhabditis elegans*. *Genetics* 135: 385–404.
- Okumura, M., M. Wilecki, and R. J. Sommer, 2017 Serotonin drives predatory feeding behavior via synchronous feeding rhythms in the nematode *Pristionchus pacificus*. *G3 (Bethesda)* 7: 3745–3755. <https://doi.org/10.1534/g3.117.300263>
- Plotkin, J. B., and G. Kudla, 2011 Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12: 32–42. <https://doi.org/10.1038/nrg2899>
- Prabh, N., W. Roeseler, H. Witte, G. Eberhardt, R. J. Sommer *et al.*, 2018 Deep taxon sampling reveals the evolutionary dynamics of novel gene families in *Pristionchus* nematodes. *Genome Res.* 28: 1664–1674. <https://doi.org/10.1101/gr.234971.118>
- Ragle, J. M., S. Katzman, T. F. Akers, S. Barberan-Soler, and A. M. Zahler, 2015 Coordinated tissue-specific regulation of adjacent alternative 3' splice sites in *C. elegans*. *Genome Res.* 25: 982–994. <https://doi.org/10.1101/gr.186783.114>
- Ragsdale, E. J., M. R. Müller, C. Rödelberger, and R. J. Sommer, 2013 A developmental switch coupled to the evolution of plasticity acts through a sulfatase. *Cell* 155: 922–933. <https://doi.org/10.1016/j.cell.2013.09.054>
- Redemann, S., S. Schloissnig, S. Ernst, A. Pozniakowsky, S. Ayloo *et al.*, 2011 Codon adaptation-based control of protein expression in *C. elegans*. *Nat. Methods* 8: 250–252. <https://doi.org/10.1038/nmeth.1565>
- Rice, P., I. Longden, and A. Bleasby, 2000 EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16: 276–277. [https://doi.org/10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2)
- Rödelberger, C., W. Röseler, N. Prabh, K. Yoshida, C. Weiler *et al.*, 2018 Phylotranscriptomics of *Pristionchus* nematodes reveals parallel gene loss in six hermaphroditic lineages. *Curr. Biol.* 28: 3123–3127.e5. <https://doi.org/10.1016/j.cub.2018.07.041>
- Rödelberger, C., M. Athanasouli, M. Lenuzzi, T. Theska, S. Sun *et al.*, 2019 Crowdsourcing and the feasibility of manual gene annotation: a pilot study in the nematode *Pristionchus pacificus*. *Sci. Rep.* 9: 18789. <https://doi.org/10.1038/s41598-019-55359-5>
- Rubin, G. M., and A. C. Spradling, 1982 Genetic transformation of *Drosophila* with transposable element vectors. *Science* 218: 348–353. <https://doi.org/10.1126/science.6289436>
- Schlager, B., X. Wang, G. Braach, and R. J. Sommer, 2009 Molecular cloning of a dominant roller mutant and establishment of DNA-mediated transformation in the nematode *Pristionchus pacificus*. *Genesis* 47: 300–304. <https://doi.org/10.1002/dvg.20499>
- Sharp, P. M., and W. H. Li, 1987 The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15: 1281–1295. <https://doi.org/10.1093/nar/15.3.1281>
- Sieriebriennikov, B., S. Sun, J. W. Lightfoot, H. Witte, E. Moreno *et al.*, 2020 Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell. *PLoS Genet.* 16: e1008687. <https://doi.org/10.1371/journal.pgen.1008687>
- Sommer, R. J., L. K. Carta, S. Kim, and P. W. Sternberg, 1996 Morphological, genetic and molecular description of *Pristionchus pacificus*. *Fundam. Appl. Nematol.* 6: 511–521.
- Tanaka, S. E., M. Dayi, Y. Maeda, I. J. Tsai, R. Tanaka *et al.*, 2019 Stage-specific transcriptome of *Bursaphelenchus xylophilus* reveals temporal regulation of effector genes and roles of the dauer-like stages in the lifecycle. *Sci. Rep.* 9: 6080. <https://doi.org/10.1038/s41598-019-42570-7>
- Wang, J., B. Czech, A. Crunk, A. Wallace, M. Mitreva *et al.*, 2011 Deep small RNA sequencing from the nematode *Ascaris* reveals conservation, functional diversification, and novel developmental profiles. *Genome Res.* 21: 1462–1477. <https://doi.org/10.1101/gr.121426.111>
- Wang, J., S. Gao, Y. Mostovoy, Y. Kang, M. Zagoskin *et al.*, 2017 Comparative genome analysis of programmed DNA elimination in nematodes. *Genome Res.* 27: 2001–2014. <https://doi.org/10.1101/gr.225730.117>
- Wilecki, M., J. W. Lightfoot, V. Susoy, and R. J. Sommer, 2015 Predatory feeding behaviour in *Pristionchus* nematodes is dependent on phenotypic plasticity and induced by serotonin. *J. Exp. Biol.* 218: 1306–1313. <https://doi.org/10.1242/jeb.118620>
- Witte, H., E. Moreno, C. Rödelberger, J. Kim, J.-S. Kim *et al.*, 2015 Gene inactivation using the CRISPR/Cas9 system in the nematode *Pristionchus pacificus*. *Dev. Genes Evol.* 225: 55–62. <https://doi.org/10.1007/s00427-014-0486-8>
- Yang, Y.-F., X. Zhang, X. Ma, T. Zhao, Q. Sun *et al.*, 2017 Trans-splicing enhances translational efficiency in *C. elegans*. *Genome Res.* 27: 1525–1535. <https://doi.org/10.1101/gr.202150.115>

Communicating editor: O. Hobert

Current Biology

Chitin contributes to the formation of a feeding structure in a predatory nematode

Graphical abstract



Authors

Shuai Sun, Hanh Witte, Ralf J. Sommer

Correspondence

ralf.sommer@tuebingen.mpg.de

In brief

Sun et al. find that mutations of chitin-synthase 2 in the nematode *P. pacificus* are larval lethal or teethless. They show that the chitin-synthase inhibitor Nikkomycin Z also results in teethless animals supporting a role of chitin in nematode teeth formation. Chitin-synthase 2 acts downstream of the plasticity switch gene *eud-1*.

Highlights

- Chitin contributes to teeth formation in *P. pacificus*
- Chitin-synthase inhibitor Nikkomycin Z treatment results in teethless animals
- Teethless animals can sense prey nematodes but are unsuccessful at killing

Article

Chitin contributes to the formation of a feeding structure in a predatory nematode

Q1 Shuai Sun,¹ Hanh Witte,¹ and Ralf J. Sommer^{1,2,*}

¹Max Planck Institute for Biology, Department for Integrative Evolutionary Biology, Max-Planck-Ring 9, Tuebingen 72076, Germany

²Lead contact

*Correspondence: ralf.sommer@tuebingen.mpg.de

<https://doi.org/10.1016/j.cub.2022.11.011>

SUMMARY

Some nematode predators and parasites form teeth-like denticles that are histologically different from vertebrate teeth, but their biochemical composition remains elusive. Here, we show a role of chitin in the formation of teeth-like denticles in *Pristionchus pacificus*, a model system for studying predation and feeding structure plasticity. *Pristionchus* forms two alternative mouth morphs with one tooth or two teeth, respectively. The *P. pacificus* genome encodes two chitin synthases, with the highly conserved *chs-2* gene being composed of 60 exons forming at least four isoforms. Generating CRISPR-Cas9-based gene knockouts, we found that *Ppa-chs-2* mutations that eliminate the chitin-synthase domain are lethal. However, mutations in the C terminus result in viable but toothless worms, with severe malformation of the mouth. Similarly, treatment with the chitin-synthase inhibitor Nikkomycin Z also results in toothless animals. Toothless worms can feed on various bacterial food sources but are incapable of predation. High-resolution transcriptomics revealed that *Ppa-chs-2* expression is controlled by the sulfatase-encoding developmental switch *Ppa-eud-1*. This study indicates a key role of chitin in the formation of teeth-like denticles and the complex feeding apparatus in nematodes.

Q2

Q5 Q4 Q3 INTRODUCTION
Q8

Nematodes with their cylindrical and uniform body are extremely abundant and found in all ecosystems.¹ This evolutionary success is largely due to striking adaptations in the head and mouth region (stoma), resulting in a diversity of feeding structures and strategies. For example, the formation of teeth-like denticles in independent nematode lineages is associated with predation and parasitism. The stoma of the human hookworm *Ancylostoma duodenale* has two plates that form two strong teeth, which enable the attachment to the small intestine in the host, causing severe tissue damage.² The free-living soil nematode *Pristionchus pacificus* (*Ppa*) forms two alternative mouth forms, an example of developmental (phenotypic) plasticity. Specifically, the *stenostomatous* (*St*) morph with a single tooth and a narrow stoma is a strict bacterial feeder, whereas the *eurystomatous* (*Eu*) form with two teeth and a wider stoma is a potential predator of other nematodes and fungi (Figures 1A–1C). *P. pacificus* mouth-form dimorphism has become a major model system for the investigation of the molecular mechanisms regulating developmental plasticity.^{6,7} Indeed, various forward and reverse genetic and experimental approaches have identified a complex gene-regulatory network (GRN) controlling mouth-form plasticity.^{8–13} However, despite the molecular understanding of the GRN that controls mouth-form plasticity and the available insight into the ecological and evolutionary significance of these feeding structure variations,^{14–16} little is known about the structural and biochemical architecture of nematode teeth and mouth structures in any species.

From ultrastructural analysis it has been suggested that, except for the anteriormost part (called the cheilostom), the nematode mouth is fully separated from the cuticular exoskeleton.¹⁷ Genetic and molecular studies have confirmed this hypothesis because mouth structures do not contain collagen, which at the same time represents the major component of the cuticle.¹⁸ However, a recent study identified the mucin-type hydrogel-forming protein DPY-6 as a conserved protein component of the mouth with a concomitant role in cuticle formation in *P. pacificus* and *Caenorhabditis elegans* (*Cel*).³ In humans, mucin-type proteins are glycoproteins that form a barrier of the mucus.¹⁹ In both *P. pacificus* and *C. elegans*, *dpy-6* mutant animals have a strongly reduced cheilostom, confirming original hypotheses based on ultrastructural analysis. At the same time, *P. pacificus* teeth are formed normally in *Ppa-dpy-6* mutants indicating that other proteins or macromolecules must form the basis of the teeth in the mouth.

Chitin is a major component of the nematode eggshell and has also been shown to contribute to the grinder of *C. elegans*.^{20,21} The grinder is the major structure for the lysis of bacteria in *C. elegans* and many other free-living nematodes and is part of the pharynx. In addition, chitin was detected in the parasite *Oesophagostomum dentatum* in the pharyngeal cuticle by lectin-labeling studies,²⁰ whereas similar studies in other nematode species did not detect chitin. Thus, except for the nematode eggshell the strongest evidence for a role of chitin comes from the *C. elegans* grinder. It is important to note that the presence of the grinder and teeth-like denticles are mutually exclusive; i.e., *P. pacificus* and its relatives that have gained teeth-like denticles exhibit a concomitant loss of

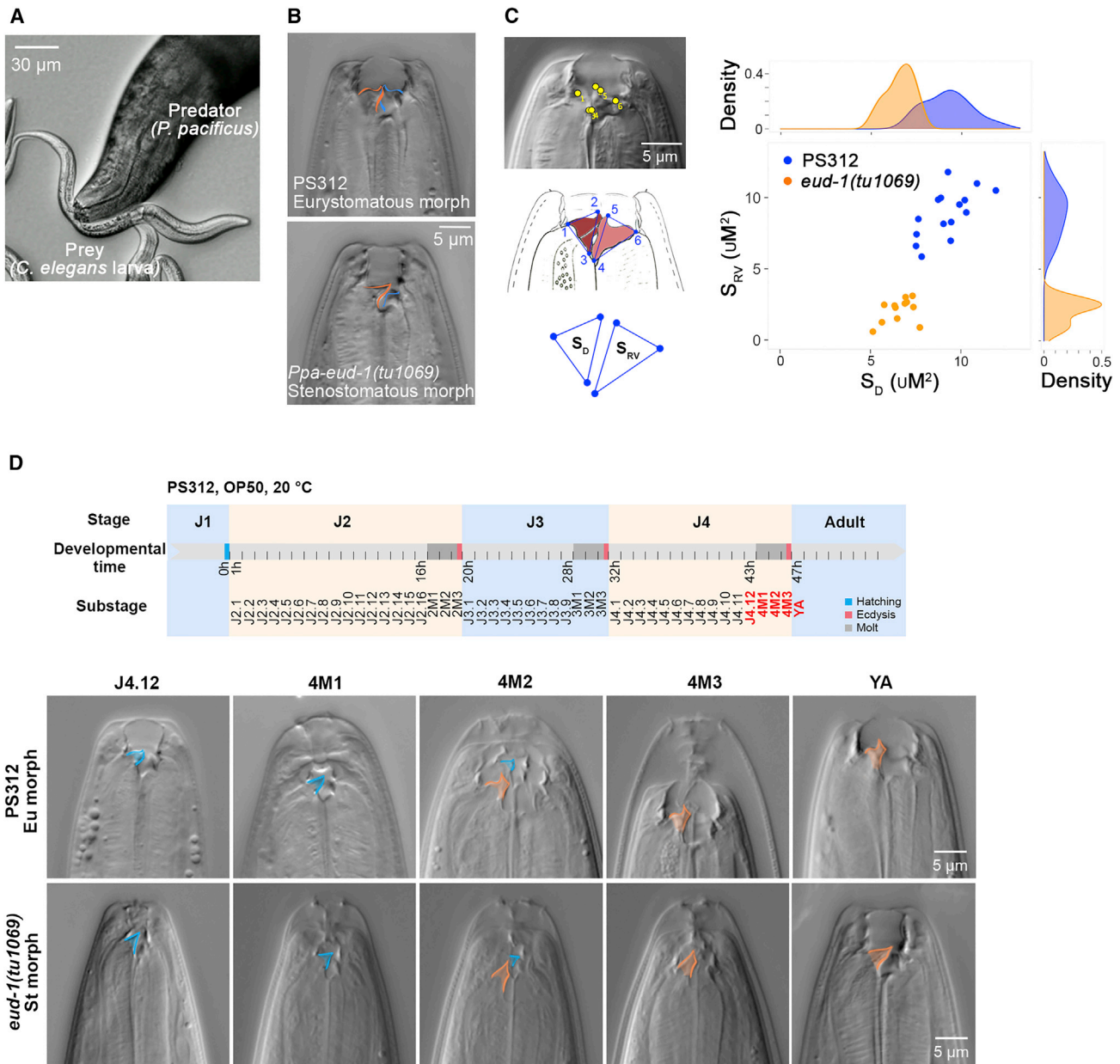


Figure 1. *P. pacificus* mouth-form plasticity and mouth-form reconstruction

(A) Predatory behavior of a eurystomatous *P. pacificus* adult preying on a larva of *Caenorhabditis elegans*.

(B) The two alternative mouth forms of *P. pacificus*, eurystomatous (Eu) and stenostomatous (St) morphs in median plane. Top: the omnivorous Eu morph, here shown for a wild-type animal (PS312), has a wide mouth, a claw-shaped dorsal tooth (in orange), and a hooked ventrosublateral tooth (in blue). Unlabeled images in two focal planes are shown in [Figure S1A](#). Bottom: the bacterial-feeding St morph, here shown for a *Ppa-eud-1* mutant animal, has a narrow mouth with a single dorsal tooth (in orange) and a minute denticle on the right ventrosublateral part of the stegostom (in blue).

(C) Quantitative measurement of teeth size illustrating the differences between Eu and St morphs. Left: six landmarks used for teeth size measurement ([STAR Methods](#); [Figure 2](#)). Right: point plot shows the comparison of teeth size between wild-type (PS312) and *Ppa-eud-1(tu1069)* mutants. S_D and S_{RV} indicate the size of the dorsal tooth and right subventral tooth, respectively.

(D) Mouth-form reconstruction during the last molting stage. The top panel presents a timetable of the life history from hatching to young adults in *P. pacificus* (20 °C, OP50, modified from Sun et al.⁴). The five substages in red indicate the five time points corresponding to the last molt stage. In the bottom panel, the upper five images show the morphological changes at five substages in a PS312 Eu animal. J4.12 is 1 h before the J4 intermolt stage. 4M1 represents the fourth molt lethargus, a period of inactivity during which pharyngeal pumping, feeding, and locomotion become arrested. 4M2 represents the fourth molt apolysis, in which the new teeth appear and massive cuticle is synthesized to replace the old cuticle lining in the body wall. 4M3 represents the fourth molt ecdysis, a short period in which the old cuticle is shed and the molting process is completed. YA (young adult). The dorsal teeth in blue and orange indicate the old J4 dorsal tooth (blue), and the new adult dorsal tooth (orange), respectively. Scale bars, 5 μm . The lower five images present similar stages for the St morph in a *Ppa-eud-1(tu1069)* mutant background.

See also [Figure S1](#) and [Videos S1](#) and [S2](#).

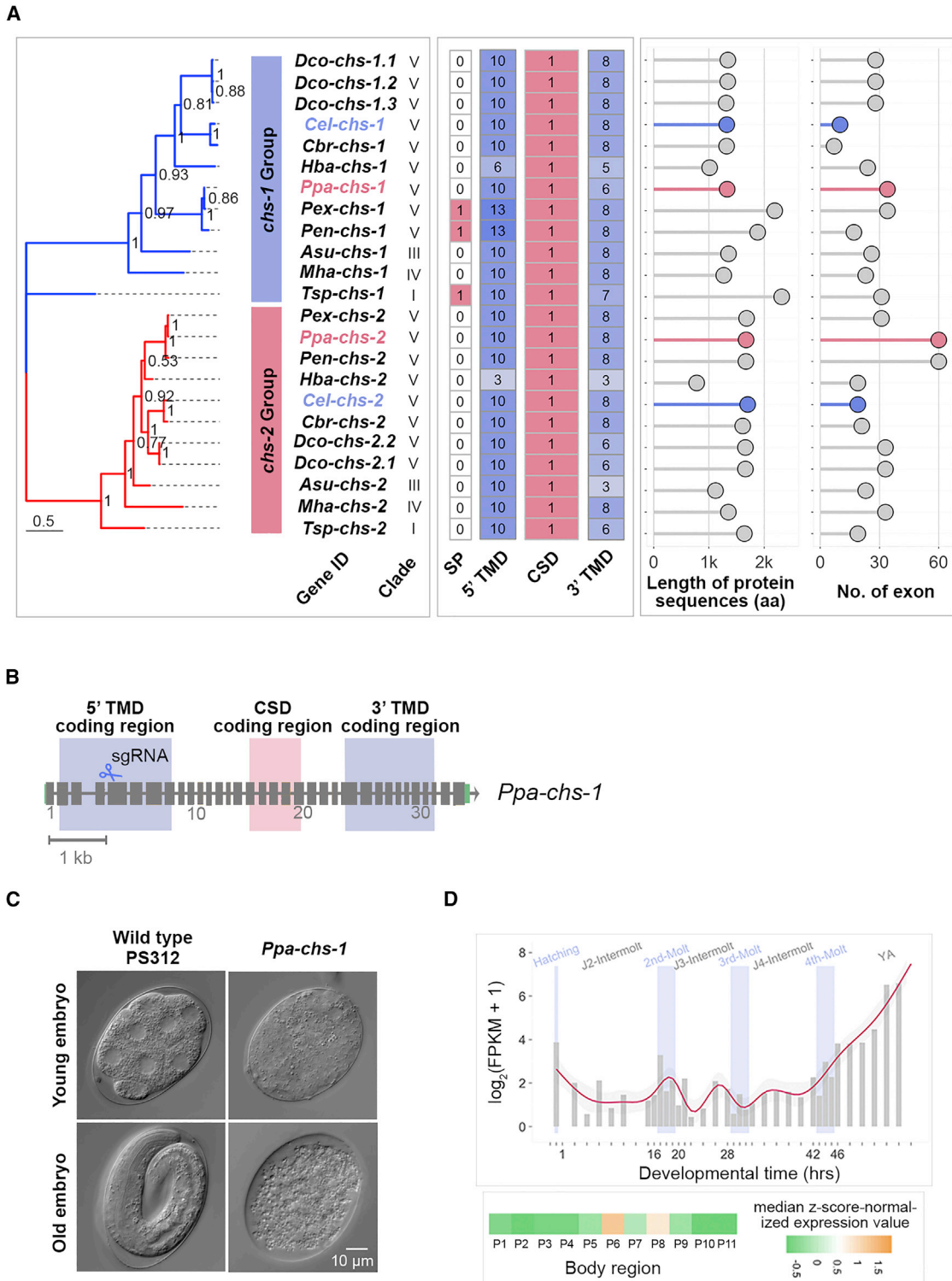


Figure 2. Chitin-synthase phylogeny and the role of *Ppa-chs-1* in eggshell formation

(A) Evolution of the chitin-synthase genes in nematodes. Left panel shows a phylogeny of 23 chitin-synthase genes (*chs*) that are clustered into two groups, *chs-1* and *chs-2*. Values at nodes represent approximate likelihood ratio test support values. The four columns in the middle panel display the distribution of four predicted protein domains of the CHS proteins, including the signal peptide (SP), transmembrane domains close to the N terminus of the gene (5' TMD), chitin-synthase domain (CSD), and transmembrane domains at the C terminus of the gene (3' TMD). Numbers shown in the box represent copy numbers of the protein domains (SP and CSD) and the number of membrane spans for the TMD. The right panel shows the length of the protein sequences and number of annotated

(legend continued on next page)

the grinder.¹⁴ However, the grinder is not part of the mouth and is neither related to nor homologous to nematode teeth. Therefore, there is currently no evidence for a role of chitin in the formation of nematode teeth and thus, the structural composition of nematode teeth remains elusive.

Here, we show a major role of chitin in the formation of the teeth-like denticles in *P. pacificus* through two complementary approaches. First, CRISPR-Cas9-based gene knockouts of the highly conserved chitin-synthase *Ppa-chs-2* result in lethal or viable, but toothless mutants, depending on the exact molecular lesion in the gene. Second, treatment with the chitin-synthase inhibitor Nikkomycin Z also results in toothless animals. Toothless animals survive to adulthood and feed on various bacteria, but they are incapable of predation. High-resolution transcriptomics revealed that the oscillating gene *Ppa-chs-2* is regulated by the key developmental switch *Ppa-eud-1* at the last molt. This study provides the first evidence for a role of chitin in the formation of nematode teeth.

RESULTS

Pristionchus teeth-like denticles are rebuilt during molting

Nematodes are ecdysozoans and have four larval stages.²¹ The collagenous cuticle is replaced with a newly synthesized cuticle 5 times during development, and the old cuticle is removed by molting. Besides the external cuticle the extracellular mouth parts are also shed during the molting process. To reconstruct and properly visualize the formation of the new teeth during the last molt, we recorded the dynamic morphological changes in the head region using high-resolution time-lapse video differential interference contrast (DIC) microscopy (Figures 1D, S1, and S2; Video S1). We used high-resolution developmental staging that was previously established for transcriptomic studies.⁴ Pumping cessation (PC) of the pharynx starts at the beginning of the lethargus and was chosen as starting point for the observation. In wild-type *P. pacificus* animals (PS312), we found that teeth reconstruction of Eu animals was completed within 150 min post-PC (20°C). In the lethargus stage (4M1; 0–60 min post-PC), the old teeth were detached from the underlying pharyngeal muscle (pm) cells and the frame of the new teeth appeared (Figure 1D). In apolysis (4M2; 60–120 min post-PC), the size of the new teeth increased by accumulating new extracellular material. Finally, in ecdysis (4M3; 120–150 min post-PC), the old cuticle was removed through a series of body motions. Note that in DIC microscopy, the cheilostomatal elements remain more robust through the molt than the posterior elements. Unfortunately, transmission-electron-microscopical data are not available to confirm this observation. We used mutants in the

developmental switch gene *Ppa-eud-1*, which results in all-St cultures⁷ following the reconstruction of the St morph. While *Ppa-eud-1* mutant animals revealed a significant difference in the duration of teeth reconstruction, the overall order of events was similar (Figures 1D and S1B; Video S2). Thus, both Eu and St animals shed their teeth and other cuticular mouth structures together with the external cuticle during molting.

Two chitin-synthase genes are highly conserved in nematode evolution

Next, we wanted to test the hypothesis that chitin is part of the mouth and/or the teeth in *P. pacificus*. One important inroad into the biology and function of chitin is through the analysis of chitin-synthase genes (*chs*).²² In fungi, the evolution of the *chs* family strongly correlates with morphogenesis and the adaptation to ecological niches with up to seven classes of fungal chitin synthases.²³ In contrast, the genome of the nematode *C. elegans* contains only two *chs* genes that both contain a chitin-synthase-2 domain (CS2, PF03142). RNA interference (RNAi) experiments indicated a role of *chs-1* in eggshell formation.^{24,25} The second gene, *Cel-chs-2*, is exclusively expressed in the pharynx and RNAi against *Cel-chs-2* resulted in aberrant phenotypes of the pharyngeal grinder, whereas the mouth was unaffected in these RNAi-treated animals.^{24–26}

We performed bioinformatic analysis of potential *chs* in representative nematodes. Based on the conservation of the CS2 domain (PF03142), we identified a total of 23 genes in the annotated genomes of 10 species from four nematode clades (clade I, III, IV, and V), including free-living (i.e., *C. elegans*, *P. pacificus*), animal (i.e., *Ascaris suum* [Asu] and *Trichinella spiralis* [Tsp]), and plant (*Meloidogyne hapla* [Mha]) parasitic species (Figure 2A; STAR Methods). Maximum likelihood analysis revealed two distinct orthologous clades with most species containing a single copy for both *chs-1*- and *chs-2*-type genes (Figure 2A). The exception is *Diploscapter coronatus* (Dco), which includes three copies related to *chs-1* and two copies related to *chs-2* (Figure 2A). This phylogenetic analysis allows two conclusions. First, the ecological diversification of nematodes is not associated with an evolutionary multiplication of *chs* as is observed in fungi. Second, *chs-2* in particular is highly conserved throughout nematodes, while its function remains currently unknown.

P. pacificus chs-1 is essential for eggshell formation

To study a potential role of chitin in the formation of nematode teeth-like denticles and the mouth of *P. pacificus* in general, we used the well-developed CRISPR-Cas9 engineering.^{27–29} *P. pacificus* contains clear orthologs of *chs-1* and *chs-2*, like most other clade V nematodes (Figure 2A). Therefore, we first targeted *Ppa-chs-1* and obtained several homozygous *Ppa-chs-1* mutants in the

exons. Species used for phylogenetic reconstruction: *Trichinella spiralis* (Tsp), *Ascaris suum* (Asu), *Meloidogyne hapla* (Mha), *Diploscapter coronatus* (Dco), *Heterorhabdites bacteriophora* (Hba), *Caenorhabdites briggsae* (Cbr), *Caenorhabdites elegans* (Cel), *Pristionchus entomophagus* (Pen), *Pristionchus exspectatus* (Pex), and *Pristionchus pacificus* (Ppa).

(B) Gene model for *Ppa-chs-1* (PPA22049). The transmembrane domains and the chitin-synthase domain are marked in blue and red, respectively. The sgRNA for CRISPR-induced gene knockout was designed in exon 5.

(C) Differential interference contrast images showing the defective development of the progeny of homozygous *Ppa-chs-1* mutants.

(D) Temporal and spatial gene expression of *Ppa-chs-1* deduced from the data of Sun et al.^{3,4} and Rödelsperger et al.⁵ Barplot profile of the temporal gene expression ($\log_2(\text{FPKM} + 1)$) of *Ppa-chs-1* across postembryonic development. The heatmap at the bottom shows that *Ppa-chs-1* is highly expressed in the central body region P6 and P8.

See also Data S1.

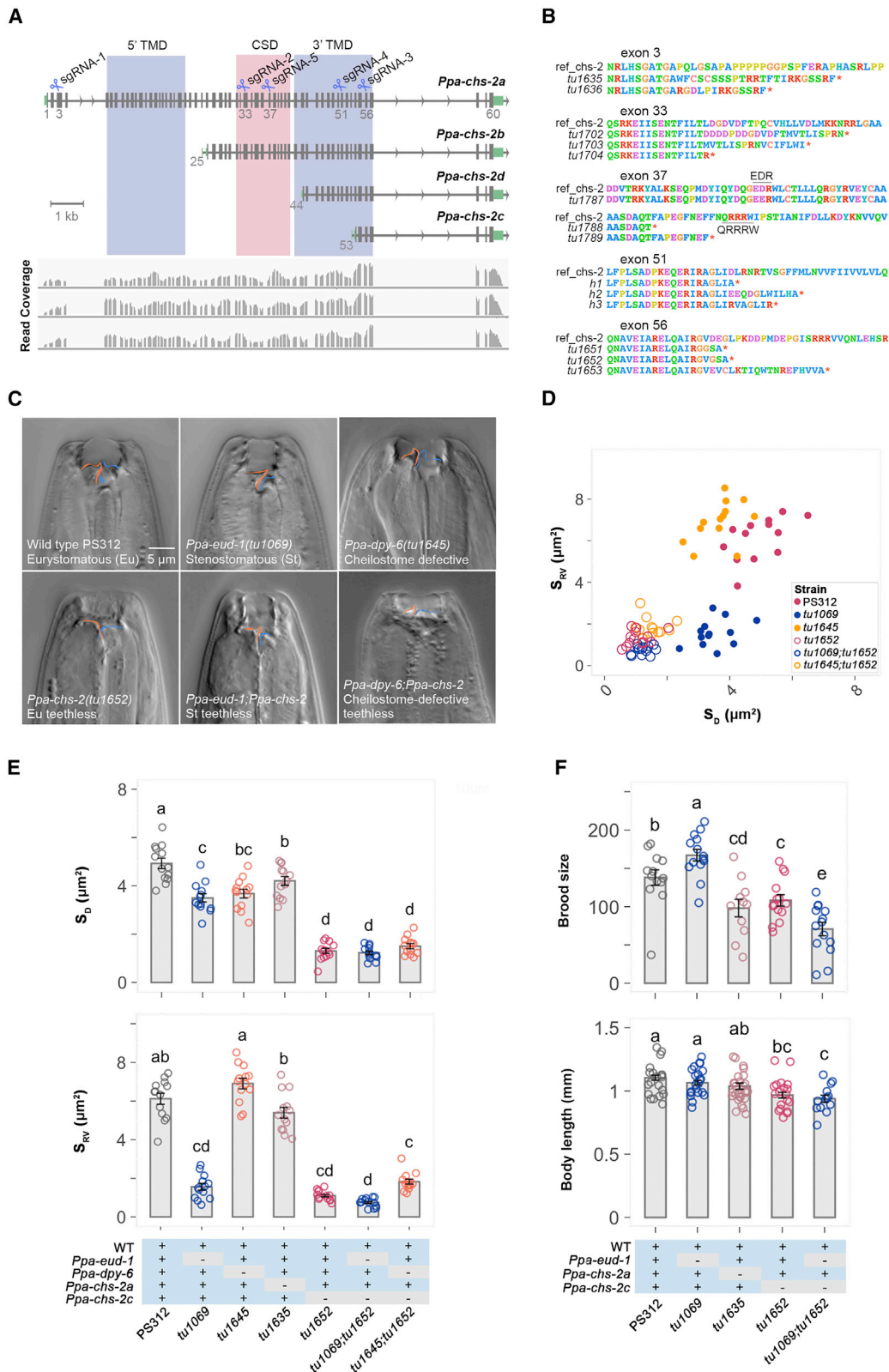


Figure 3. CRISPR-based analysis of *Ppa-CHS-2* reveals an essential role in teeth formation

(A) Gene structure of *Ppa-chs-2* (PPA39501) with 60 exons and at least four different isoforms (*Ppa-chs-2a-d*) confirmed by RACE-PCR. The coding region of the transmembrane domains (5'/3' TMD) and chitin-synthase domain (CSD) are marked in blue and red, respectively. The longest isoform *Ppa-chs-2a* contains all 60 (legend continued on next page)

the F2 generation (Figure 2B). However, we were unable to establish propagating viable mutant lines. Instead, all eggs of homozygous *Ppa-chs-1(tu1634)* mutants have an abnormal egg shape and fail to develop (Figure 2C). We used the previously generated single-worm, transcriptomic, high-resolution developmental gene-expression atlas with 38 postembryonic time points at 1–2 h resolution for proper gene-expression analysis.⁴ This analysis revealed that *Ppa-chs-1* is highly expressed in the adult stage, whereas expression is low at juvenile stages (Figure 2D).⁴ Furthermore, the analysis of the available TOMO-seq data of adult hermaphrodites⁵ indicates that *Ppa-chs-1* is specifically expressed in the gonadal region P6–P8 (Figure 2D). Thus, *Ppa-chs-1* has an essential role in eggshell formation, similar to the previously described role of *chs-1* in eggshell formation of *C. elegans*, *B. malayi*, and *D. immitis*.^{25,30,31}

Mutations in *Ppa-chs-2* form an allelic series including toothless phenotypes

Next, we wanted to investigate *Ppa-chs-2*, the ortholog of *C. elegans chs-2*, which when mutated results in defects in the grinder. *Ppa-chs-2* exhibits an unusually complex gene structure with a total of 60 exons and the chitin-synthase domain (CSD) being encoded in exons 31–42 (Figure 3A). Previous transcriptomic studies suggested three potential isoforms of *Ppa-chs-2*, including one short isoform that misses the chitin-synthase domain (Figures 3A and S3).^{9,29,32} Using RACE cDNA amplification protocols, we were able to confirm the existence of these three *Ppa-chs-2* isoforms consisting of 60, 36, and 9 exons, respectively (Figures 3A and S3). Additionally, we also amplified a fourth isoform that was not annotated in previous transcriptomic studies and also misses the CSD (Figure 3A). Thus, we conclude the existence of at least four isoforms of *Ppa-chs-2*, which we named *Ppa-chs-2a* to *Ppa-chs-2d* (Figure 3A).

To discriminate between potentially different functions of these isoforms, we used CRISPR sgRNAs in exons, 3, 33, 37, 51, and 56, respectively, and isolated a total of 14 independent mutant alleles (Figures 3A and 3B). These mutations form an allelic series with phenotypes all the way to lethality. First, two mutations in exon 3 (*tu1635* and *tu1636* with 8- and 38-bp deletions, respectively) affect only *Ppa-chs-2a* and have no visible mouth-form phenotype. Although both alleles result in premature stop codons in front of the CSD, they are both viable (Figure 3B).

However, these mutants displayed a slight decrease in the size of the dorsal tooth (DT) (Figure 3E), and a significant reduction in brood size (Figure 3F). These results suggest that the *Ppa-chs-2a* isoform is nonessential but might have a redundant function with other *Ppa-chs-2* isoforms. Second, we generated six alleles with mutations in the CSD with sgRNAs targeting exons 33 and exon 37 (Figure 3B). Specifically, we obtained one allele with a substitution in the putative acceptor saccharide-binding site [E(D985E)R] (*tu1787*), and two alleles with mutations in the product-binding motif QRRRW, resulting in frameshift mutations (*tu1788* and *tu1789*) (Figure 3B). All of these mutations exhibit larval lethality, suggesting that the *Ppa-chs-2b* isoform is essential (Figure S3D). Specifically, lethality is seen in young J2 stages and is fully penetrant. Third, when we generated mutations in the C terminal exon 56, we obtained three alleles that are all viable but toothless (*tu1651*, *tu1652*, and *tu1653*) (Figures 3C–3E). All three of these alleles show severe defects of the posterior part of the stoma and the absence of the teeth in J3, J4, and adult stages (Figures 3C and S1D). These results suggest that *Ppa-chs-2* is essential for teeth formation in *P. pacificus* and likely essential for the proper formation of the stoma. However, it remains unclear whether the different phenotypes—lethality versus toothless—of mutations induced in exons 33, 37, and 56, respectively, are due to different functions of the isoforms with or without the CSD. Therefore, we finally generated CRISPR mutations in exon 51, which is specific to the isoforms *Ppa-chs-2a*, *Ppa-chs-2b*, and *Ppa-chs-2d*. We were able to generate three alleles (Figure 3B), all of which show a lethal phenotype similar to the six mutations in the CSD. While these results suggest that *Ppa-chs-2* is essential for larval development and might have a key role in the formation of the mouth, they do not allow to distinguish if the function of *Ppa-CHS-2* in teeth formation requires the CSD encoded in exons 31–42. Theoretically, *Ppa-chs-2c* might have a function other than synthesizing chitin, an issue that will be further addressed in complementary studies described below.

As these mutants were generated in the *P. pacificus* PS312 background, the majority of animals have the Eu mouth form. To determine if the St morph is also affected, we first generated *Ppa-eud-1(tu1069)*, *Ppa-chs-2(tu1652)* double mutants as *Ppa-eud-1* mutants are completely St.⁷ Indeed, *Ppa-eud-1(tu1069)*, *Ppa-chs-2(tu1652)* mutant animals are also toothless

predicted exons and the chitin-synthase domain. The second isoform *Ppa-chs-2b* excludes the N-terminal region, which contains the first 24 exons and encodes only parts of the transmembrane domain at the N terminus. In contrast, the third isoform *Ppa-chs-2c* contains only the last 8 exons, which includes the last part of the transmembrane domain, but not the chitin-synthase domain. The fourth isoform *Ppa-chs-2d* starts from the 44th exon and encodes most of the C-terminal transmembrane domain. For the fifth isoform, sgRNA1-5 was used to target different isoforms. Bottom: three tracking channels of RNA-seq read coverage suggest different expression between N- and C-terminal exons of the *Ppa-chs-2* locus at the fourth molt. See also Figure S2.

(B) Fourteen specific mutations induced in *Ppa-chs-2* by targeting different exons.

(C) Mouth-form images show the effects of *Ppa-chs-2* on teeth formation. The upper three images present the Eu morph of wild-type PS312, the St morph of *Ppa-eud-1(tu1069)* mutants, and the cheilostom-defective morph of *Ppa-dpy-6(tu1645)*. The three lower images display the toothless phenotype of Eu, St, and cheilostom-defective animals, respectively. The dorsal and right subventral tooth are labeled in orange and blue, respectively. Scale bars, 5 μ m. Unlabeled images are shown in Figure S1B.

(D) *Ppa-chs-2* significantly decreased teeth size of wild-type and mouth-form mutants. Point plot shows the comparison of teeth size in wild type (PS312); *Ppa-eud-1(tu1069)*; *Ppa-dpy-6(tu1645)*, *Ppa-chs-2(tu1653)*; *Ppa-eud-1(tu1069)*, *Ppa-chs-2(tu1653)*; and *Ppa-dpy-6(tu1645)*, *Ppa-chs-2(tu1653)* animals. S_D and S_{RV} indicate the size of the dorsal tooth and right subventral tooth, respectively.

(E and F) *Ppa-chs-2*-associated mutants display dramatically reduced size of the dorsal and right subventral tooth, brood size, and small changes in body length. Tukey's multiple comparisons were tested by the Tukey's HSD test, with the letters upon each volume indicating significant differences between the means, p adjusted value < 0.05.

See also Figure S3 and Data S1 and S2.

and completely viable (Figures 3C–3F). Finally, we tested genetic interactions between *Ppa-chs-2* and *Ppa-dpy-6* mutations, the latter of which display cheilostom defects. We found that a *Ppa-chs-2(tu1652)*, *Ppa-dpy-6(tu1645)* double mutant exhibits a toothless and cheilostom-defective phenotype and, despite the strong aberrations in their mouths, are still viable (Figures 3C–3E).

Nikkomycin Z supports a role of chitin in teeth formation

To overcome the limitations of our genetic analysis due to the lethality of *Ppa-chs-2* alleles with mutations in the chitin-synthase domain, we used an inhibitor that specifically targets the CSD. In general, several inhibitors of chitin synthases have been used primarily for the control of insect pests and disease-causing fungi. Specifically, peptidyl nucleosides from a variety of *Streptomyces* species function as substrate analogs and inhibit chitin synthases.²² One such inhibitor of CHS-2-type proteins is Nikkomycin Z, which we have used as a potential inhibitor of *Ppa-chs-2*. We first supplemented nematode agar plates with Nikkomycin Z but were unable to observe any visible phenotypes (STAR Methods). Therefore, we directly injected Nikkomycin Z in late J4 larval stage animals, 2–3 h before the final molt (Figure 4A). For this, we used two different sites of injection in the animal. First, we injected into the pm cells in the metacarpus. Second, we injected into the intercellular space in the isthmus, the body region adjacent to the metacarpus (Figure 4A). We observed a survival rate of 50% if animals were injected into the metacarpus, but nearly 100% of survival when injecting into the isthmus (Figure 4B). Most importantly, there was no difference in survival when injecting Nikkomycin Z or water controls (Figure 4B). Thus, animals do survive the injection with the chitin-synthase inhibitor Nikkomycin Z.

Injection of J4 juveniles with Nikkomycin Z resulted in specific defects in teeth formation (Figure 4). We found between 75% and 95% toothless animals when injecting Nikkomycin Z at a concentration of 10 mM in the intracellular space and the metacarpus, respectively (Figure 4B). Other animals showed the development of small teeth, which might be due to the fact that the amount of injected inhibitor cannot be completely controlled. Note that further morphological analysis suggests that the outline of the stoma in these toothless animals is otherwise normal (Figure 4C). When we injected different concentrations of Nikkomycin Z, we found a strong correlation between the concentration of Nikkomycin Z and the size of the teeth in both morphs. Specifically, teeth size displayed a strong reduction along with the increase of the Nikkomycin Z concentration (Figure 4D). The Nikkomycin Z concentration-dependence of teeth size suggests that the amount of chitin synthase is essential for the construction of the proper teeth in *P. pacificus*. Interestingly, injection with 0.1 mM Nikkomycin Z did not affect teeth size of the Eu morph but significantly reduced teeth size of St morphs (Figures 4D, 4E, and S4). These results might provide first evidence that different amounts of chitin are required for teeth formation in the two alternative mouth forms. These findings indicate that chitin-synthase activity, and thus chitin is essential for the formation of the teeth-like denticles of *P. pacificus*. However, similar injection experiments with a chitinase (Sigma, catalog no. C6137) at two concentrations (1 and 2 mg/mL) did not affect teeth formation. Also, previous mutant analysis of some *P. pacificus* chitinases did not affect teeth

formation,⁹ which might result from the complex chemical nature of chitin in the teeth. Nonetheless, together with the mutant analysis described above, our results from Nikkomycin Z injections strongly support a role of chitin in teeth and stoma formation.

Ppa-chs-2 is highly expressed in the pharyngeal muscle cells and epithelial cells, and peaks before the larval molts

To study the expression of *Ppa-chs-2*, we used two complementary approaches. First, we generated a transcriptional reporter of *Ppa-chs-2* using a 1,926-bp fragment upstream of *Ppa-chs-2* to study its spatial distribution. This fragment covers the complete region upstream of the first potential start ATG of *Ppa-chs-2*. We obtained two independent transgenic lines (*tuEx351* and *tuEx352*), both of which are expressed in the “pm” cells pm1 (including pm1D, pm1VR, and pm1VL), which are connected to the DT and the right subventral tooth (RVT) (Figures 5A and S5A; Video S3).³³ In addition, *Ppa-chs-2* is expressed in the epithelial (“e”) cells e2, including e2DL, e2DR, and e2V, which line the two subdorsal and ventral apices of the mouth, and the e3 cells (including e3D, e3VL, and e3VR) that line thin strips of cuticle that attach to the teeth and left subventral ridge to the gymnostom (Figures 5A and S5A; Video S3).³³ This spatial restriction supports a role of chitin in the formation of the teeth and the stoma and is consistent with the Nikkomycin Z inhibitor studies described above. Second, we analyzed the *Ppa-chs-2* expression profile in the recently generated single-worm transcriptome (SWT) atlas with its 38 distinct time points.⁴ This analysis covers complete postembryonic development with a 1–2 h temporal resolution. We found that *Ppa-chs-2* expression is highly oscillating (ARS_p value = 0.026 and ARS_amplitude = 1.439)⁴ and peaks 1–2 h before the molt (Figure 5B). Note that all three molts, in the J2, J3, and J4 stages show similar restricted expression peaks of *Ppa-chs-2*. This temporal expression pattern is consistent with a role of *Ppa-chs-2* in the remodeling of the teeth during the larval molts. Conclusively, the spatiotemporal expression pattern supports the essential role of *Ppa-chs-2* in teeth formation during molts.

Ppa-eud-1 mediates gene expression of *Ppa-chs-2* at the last molt

The expression of the transcriptional reporter *Ppa-chs-2::GFP* was found to be lower in many *Ppa-eud-1* mutant animals (*tuEx352;tu1069*) when compared with wild type (Figure S5B). This observation suggests that *Ppa-chs-2* expression might be regulated by *Ppa-eud-1*. Indeed, the developmental switch gene *Ppa-eud-1* was recently found to act upstream of several genes with oscillatory gene expression at early developmental time points, i.e., 20 and 28 h after hatching in the second and third larval molts.⁴ To elucidate if *Ppa-chs-2* is also regulated by *Ppa-eud-1* during the last molting stage, we performed high-resolution transcriptomics at the fourth molt between wild-type and *Ppa-eud-1(tu1069)* mutant animals (see STAR Methods for staging procedures). We selected the four substages J4.12, 4M1, 4M2, and 4M3, which represent the different phases of the fourth molt (Figure 1D) and sequenced all samples in triplicate. Principal component analysis (PCA) of gene expression revealed that the transcriptomes were grouped into four clusters corresponding to the four developmental substages (Figure 5C). In total, we

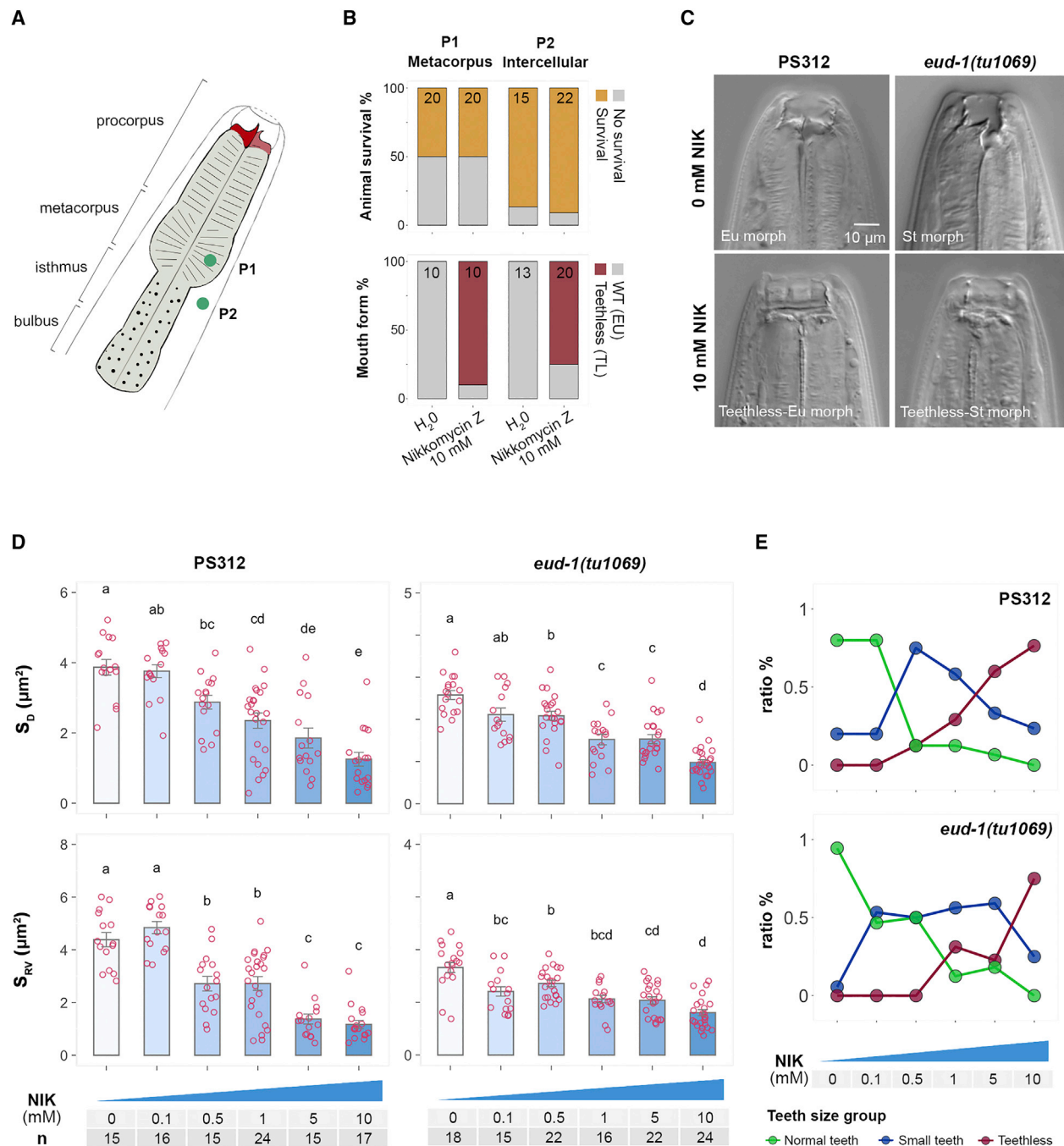


Figure 4. *P. pacificus* teeth size manipulation through the inhibitor Nikkomycin Z

(A) Schematic of head region with the two injection sites for Nikkomycin Z. P1 presents the pharyngeal muscle cells in the metacarpus; P2 indicates the intercellular space in the isthmus.

(B) Barplots display the comparison of the different injection sites for animal survival (upper) and teethless phenotype (lower). Water injection was used as control. Numbers in the bars indicate the numbers of injected animal (upper bars) and the numbers of surviving animals (lower bars).

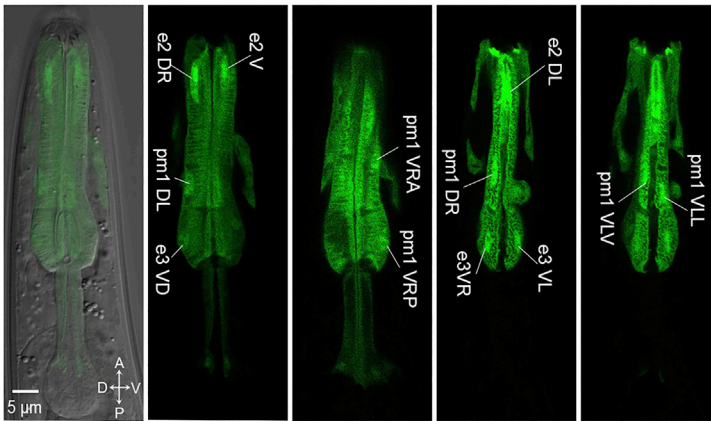
(C) Nomarski images present the mouth form of PS312 wild-type (Eu morph) and *Ppa-eud-1(tu1069)* mutants (St morph) under two injection conditions of Nikkomycin Z: 0 and 10 mM. Nikkomycin Z was injected at the late J4 stage and images were taken 24 h after injection.

(D) Barplots present the changes in teeth size of wild-type PS312 and *Ppa-eud-1(tu1069)* mutants among six different injection conditions of Nikkomycin Z. The condition 0 mM (water) serves as control. n indicates the number of animals that survived injection and could be scored. S_D and S_{RV} indicate the size of dorsal tooth and right subventral tooth, respectively.

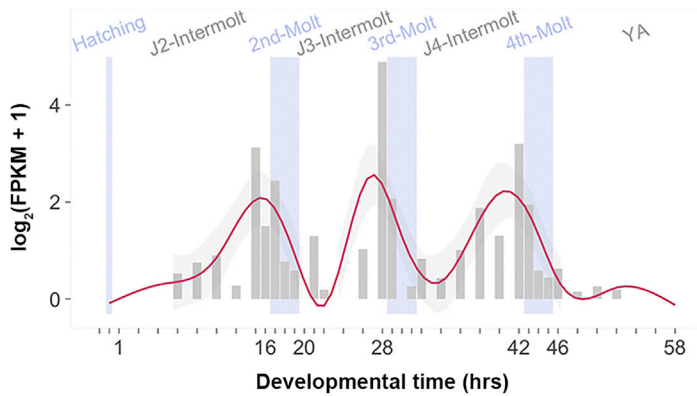
(E) Line plots profile density distribution of the three teeth size groups among the six Nikkomycin Z concentrations in wild-type strain PS312 and *Ppa-eud-1* mutants. See Figure S4 for details on characterization of the three teeth size groups.

See also Figure S4 and Data S2.

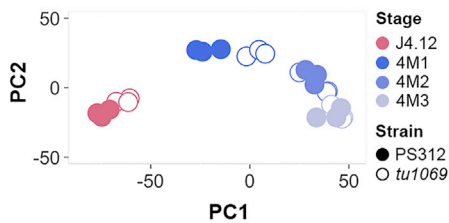
A



B



C



E

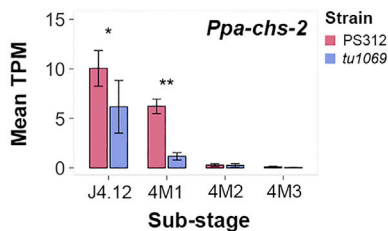


Figure 5. Spatiotemporal gene expression of *Ppa-chs-2*

(A) Expression pattern of *Ppa-chs-2* in the pm1, e2, and e3 cells at the fourth lethargus stage (4M1). The image on the left shows the overlay of a DIC image and a sum-of-slices projection of the GFP channel. The four images on the right display the position of the nuclei of the pm1, e2, and e3 cells in the GFP channel. Pharyngeal muscle (pm), epithelial (e), dorsal tooth (DT), right subventral tooth (RVT). The three pm1 cells include pm1D, pm1VR, and pm1VL, the three e2 cells are e2DL, e2DR and e2V, and three e3 cells are e3D, e3VL, and e3VR. D, dorsal; V, ventral; A, anterior; P, posterior.

(B) Barplot profiles of the temporal expression ($\log_2(\text{FPKM} + 1)$) of *Ppa-chs-2* across postembryonic development.

(C) Point plot displays 24 transcriptomes clustered into four groups corresponding to the four substages. (D) Heatmap showing the differences in gene expression of 2,298 differentially expressed genes (DEGs) between wild-type and *Ppa-eud-1* mutants. Gene-expression values of individual genes were normalized by the maximum among all conditions.

(E) Barplot shows that *Ppa-chs-2* was downregulated in *Ppa-eud-1* mutants during the fourth molting stage. * $p < 0.05$, ** $p < 0.01$.

See also [Figure S5](#), [Table S1](#), and [Video S3](#).

D

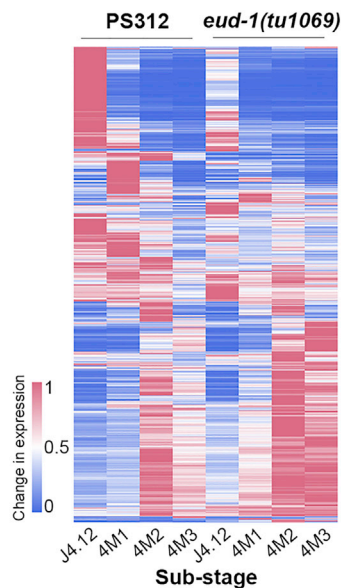


Table 1. Summary of predation behavior of the *chs-2* mutants

Strain ID	Genotype	Mouth-form type	Number of corpses (N2)/2 h
PS312	wild type	Eu	78.2 (\pm 11.4)
<i>tu1635</i>	<i>Ppa-chs-2a</i>	Eu	59.7 (\pm 19.2)
<i>tu1636</i>	<i>Ppa-chs-2a</i>	Eu	64.5 (\pm 15.9)
<i>tu1651</i>	<i>Ppa-chs-2c</i>	Eu toothless	0
<i>tu1652</i>	<i>Ppa-chs-2c</i>	Eu toothless	0
<i>tu1653</i>	<i>Ppa-chs-2c</i>	Eu toothless	0
<i>tu1069</i>	<i>Ppa-eud-1</i>	St	0
<i>tu1069</i> ; <i>tu1652</i>	<i>Ppa-eud-1</i> ; <i>Ppa-chs-2c</i>	St toothless	0

Note: The rightmost column presents the average number of *C. elegans* corpses in the assay for 2 h. Numbers in parentheses indicate standard deviation. Three independent biological replicates were investigated in each predator strain. Eu, eury stomatous; St, stenostomatous.

See also [Video S4](#).

found 15,329 reliably expressed genes (REGs) with 2,298 differentially expressed genes (DEGs) (Figures 5D and S5D–S5F). Indeed, *Ppa-chs-2* was downregulated at substages J4.12 and 4M1 in *Ppa-eud-1* mutants (Figure 5E). These data suggest that the oscillating gene *Ppa-chs-2* is regulated by the developmental switch gene *Ppa-eud-1*.

Teethless animals are incapable of predation of other nematodes

The viable toothless phenotypes of the *Ppa-chs-2* mutant alleles *tu1651*, *tu1652*, and *tu1653* or Nikkomycin-Z-treated animals are the first example of the complete perturbation of teeth formation in nematodes. Therefore, we finally wanted to investigate the behavioral consequences for such toothless animals. As indicated in the supplemental video, toothless *Ppa-chs-2*(*tu1653*) mutant animals are incapable of preying on *C. elegans* larval prey (Table 1; Video S4). Importantly, however, toothless animals still attack the prey, suggesting that the sensory recognition of prey worms is still active in these mutants. Thus, *Ppa-chs-2* mutant toothless animals still recognize prey nematodes as non-self but are incapable of successful predation.

DISCUSSION

The complex feeding apparatus of *P. pacificus* has become a model system for the genetic and epigenetic analysis of polyphenisms and developmental plasticity in general. While the identification of molecular and environmental factors regulating mouth-form plasticity has identified a complex GRN, there were two important shortcomings to our understanding of teeth-like denticles and the nematode mouth. First, the structural basis of teeth and the stoma was completely unclear. Second, the cellular architecture and the exact cells forming the various parts of the feeding apparatus of *P. pacificus* had not been identified, although such studies were performed previously in outgroup nematodes.^{34–36} Fortunately, the reconstruction of the *P. pacificus* feeding structure has recently clarified the cellular contributions of epithelial and myoepithelial cells and provided important homologies to the pattern in *C. elegans*.³³ Similarly,

the characterization of the mucin-type hydrogel-forming protein DPY-6 represents the first protein to form part of the stoma.³ However, teeth formation is unaffected in *Ppa-dpy-6* mutants, leaving the structural basis of the extracellular teeth open. Here, we provide unprecedented evidence for a role of chitin in teeth formation. This study allows three conclusions but is also limited by several uncertainties to be discussed below.

First, some *Ppa-chs-2* mutants and Nikkomycin-Z-treated animals result in toothless worms, strongly supporting a role of chitin in teeth formation. Chitin is one of the most abundant biological polysaccharides and forms the exoskeleton of insects and other arthropods, the other major ecdysozoan phylum with an exoskeleton.^{37,38} While chitin has been shown to be a component of the eggshell in several nematodes,^{25,30,31} its contribution to other nematode tissues remained unclear and controversial for decades. In particular, no experimental evidence had been available to support a role of chitin in head and mouth formation. The larval lethality of the majority of *Ppa-chs-2* mutants generated by CRISPR indicates an essential role of chitin in *P. pacificus*. The expression of *Ppa-chs-2* in the pm1, e2, and e3 cells strongly suggests this essential function to lie in the formation of the stoma and the teeth in particular. The recent cellular reconstruction revealed that the three pm1 muscle cells insert on the DT, the subventral tooth and the subventral ridge.³³ The additional finding that some *Ppa-chs-2* mutants and Nikkomycin-Z-treated animals are toothless provide a first indication of the different contributions of chitin for the formation of various parts of the nematode stoma. However, several uncertainties limit our understanding of the quantitative contribution of chitin to mouth formation and require future studies. Most importantly, the unusually complex gene structure of *Ppa-chs-2* with 60 exons and various isoforms—two of which miss the CSD—prevents a full assignment of biochemical functions to the *Ppa-CHS-2* protein generated from the different isoforms. Future cloning and expression studies are necessary to identify the exact function of the *Ppa-CHS-2* proteins generated by the different isoforms.

Second, our findings provide evidence for a potential role of chitin in the evolutionary diversification of feeding structures in nematodes. While currently of speculative nature, a role of chitin in stomatal diversification might become testable given the strong conservation of the *chs-2* gene, as indicated from our phylogenetic analysis. The availability of CRISPR technology in a growing number of nematode species and taxa, might open the opportunity to study *chs-2*-type genes in additional species in the near future. If these studies would confirm the role of chitin in the evolution of nematode head structures, such findings would strengthen the importance of chitin as the most important macromolecule for the diversification of invertebrate animals. While the role of chitin for insect and arthropod evolution is long appreciated, missing experimental evidence in other systems has previously prevented such conclusions.

Finally, with the characterization of chitin and the hydrogel-forming protein DPY-6 two structural components of the nematode stoma that may serve as a scaffold for additional factors have been identified. It is likely that additional proteins contribute to stoma formation and to genetic and biochemical studies using *Ppa-dpy-6* as a starting point may help identify such factors. In addition, the understanding of mucin-type hydrogels in humans

suggests post-translational glycosylation to play a key role in mucins obtaining their final biochemical and biophysical properties. Therefore, future studies have to involve nanotechnological approaches to better understand the biophysics of nematode teeth formation and function. It should also be noted that it remains still unknown if and to what extent the two alternative mouth forms of *P. pacificus* depend on qualitative and/or quantitative differences in their structural composition. The different expression of *Ppa-chs-2* in Eu wild-type and *Ppa-eud-1* mutant St animals already suggests that chitin might contribute in a quantitative manner to the two mouth forms. One way to obtain insight in this direction will be to perform labeling studies. However, all previous attempts have been inconclusive as they have been for lectin-binding studies in other nematodes.

In summary, with the identification of the role of chitin in teeth formation in *P. pacificus*, this study, together with the recent findings on the cellular basis of the *P. pacificus* stoma and *Ppa-DPY-6*, provides promising inroads into a mechanistic understanding of the structural composition of the nematode mouth. These studies can also serve as a foundation for a general and comprehensive perspective of the nematode head.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Mouth form phenotyping
 - Phylogenetic analysis of chitin synthase genes in nematode
 - Alternative splicing of *Ppa-chs-2*
 - CRISPR/Cas9 mutagenesis
 - Genetic transformation
 - Spatiotemporal gene expression profiling using published transcriptomics
 - Tooth size measurement in *P. pacificus*
 - Body length measurement
 - Overall self-fecundity measurement
 - Microinjection nikkomyacin Z in *P. pacificus*
 - Bulk RNA sequencing and transcriptome data analysis
 - Predatory behavior assay
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2022.11.011>.

ACKNOWLEDGMENTS

We are grateful to Dr. Adrian Streit and Dr. Christian Rödelsperger for their suggestions on the CRISPR mutants. We also thank Dr. Catia Igreja for her discussion with the microinjection of Nikkomycin Z and Tobias Theska for his

assistance with the cell determination of *Ppa-chs-2* expression. The study was funded by the Max Planck Society, and S.S. was supported by the China Q10 Scholarship Council.

AUTHOR CONTRIBUTIONS

Conceptualization, S.S. and R.J.S.; methodology, S.S. and H.W.; investigation, S.S.; writing – original draft, R.J.S. and S.S.; writing – review & editing, R.J.S. and S.S.; formal analysis, S.S.; funding acquisition, R.J.S.; supervision, R.J.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 28, 2022

Revised: September 20, 2022

Accepted: November 4, 2022

Published: December 1, 2022

REFERENCES

1. van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D.A., de Goede, R.G.M., Adams, B.J., Ahmad, W., Andriuzzi, W.S., et al. (2019). Soil nematode abundance and functional group composition at a global scale. *Nature* 572, 194–198.
2. Anderson, R.C. (2000). *Nematode Parasites of Vertebrates: Their Development and Transmission* (CABI Publishing).
3. Sun, S., Theska, T., Witte, H., Ragsdale, E.J., and Sommer, R.J. (2022). The oscillating Mucin-type protein DPY-6 has a conserved role in nematode mouth and cuticle formation. *Genetics* 220, iyab233.
4. 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 Res.* 31, 1590–1601.
5. Rödelsperger, C., Ebbing, A., Sharma, D.R., Okumura, M., Sommer, R.J., and Korswagen, H.C. (2020). Spatial transcriptomics of nematodes identifies sperm cells as a source of genomic novelty and rapid evolution. *Mol. Biol. Evol.* 1, 1–35.
6. 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, 494–497.
7. Ragsdale, E.J., Müller, M.R., Rödelsperger, C., and Sommer, R.J. (2013). A developmental switch coupled to the evolution of plasticity acts through a sulfatase. *Cell* 155, 922–933.
8. Kieninger, M.R., Ivers, N.A., Rödelsperger, C., Markov, G.V.v., Sommer, R.J., and Ragsdale, E.J. (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, 2174–2179.
9. Serobyan, V., Xiao, H., Namdeo, S., Rödelsperger, C., Sieriebriennikov, B., Witte, H., Röseler, W., and Sommer, R.J. (2016). Chromatin remodeling and antisense-mediated up-regulation of the developmental switch gene *eud-1* control predatory feeding plasticity. *Nat. Commun.* 7, 12337.
10. Bui, L.T., Ivers, N.A., and Ragsdale, E.J. (2018). A sulfotransferase dosage-dependently regulates mouthpart polyphenism in the nematode *Pristionchus pacificus*. *Nat. Commun.* 9, 4119.
11. Bui, L.T., and Ragsdale, E.J. (2019). Multiple plasticity regulators reveal targets specifying an induced predatory form in nematodes. *Mol. Biol. Evol.* 36, 2387–2399.
12. Sieriebriennikov, B., Sun, S., Lightfoot, J.W., Witte, H., Moreno, E., Rödelsperger, C., and Sommer, R.J. (2020). Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell. *PLoS Genet.* 16, 1–27.

13. Casasa, S., Biddle, J.F., Koutsovoulos, G.D., and Ragsdale, E.J. (2021). Polyphenism of a novel trait integrated rapidly evolving genes into ancestrally plastic networks. *Mol. Biol. Evol.* **38**, 331–343.
14. Susoy, V., Ragsdale, E.J., Kanzaki, N., and Sommer, R.J. (2015). Rapid diversification associated with a macroevolutionary pulse of developmental plasticity. *eLife* **4**, e05463.
15. Renahan, T., Lo, W.S., Werner, M.S., Rochat, J., Herrmann, M., and Sommer, R.J. (2021). Nematode biphasic ‘boom and bust’ dynamics are dependent on host bacterial load while linking dauer and mouth-form polyphenisms. *Environ. Microbiol.* **23**, 5102–5113.
16. Renahan, T., and Sommer, R.J. (2021). Nematode interactions on beetle hosts indicate a role of mouth-form plasticity in resource competition. *Front. Ecol. Evol.* **9**, 1–9.
17. Wright, K.A., and Thomson, J.N. (1981). The buccal capsule of *Caenorhabditis elegans* (Nematoda: Rhabditoidea): an ultrastructural study. *Can. J. Zool.* **59**, 1952–1961.
18. Cohen, J.D., and Sundaram, M.V. (2020). *C. elegans* apical extracellular matrices shape epithelia. *J. Dev. Biol.* **8**, 8.4.
19. Phani, V., Shivakumara, T.N., Davies, K.G., and Rao, U. (2018). Knockdown of a mucin-like gene in *Meloidogyne incognita* (Nematoda) decreases attachment of endospores of *Pasteuria penetrans* to the infective juveniles and reduces nematode fecundity. *Mol. Plant Pathol.* **19**, 2370–2383.
20. Neuhaus, B., Bresciani, J., and Peters, W. (1997). Ultrastructure of the pharyngeal cuticle and lectin labelling with wheat germ agglutinin-gold conjugate indicating Chitin in the pharyngeal cuticle of *Oesophagostomum dentatum* (Strongylida, Nematoda). *Acta Zool.* **78**, 205–213.
21. Page, A.P., and Johnstone, I.J. (2007). The cuticle. *WormBook*, 1–15. <https://doi.org/10.1895/wormbook.1.138.1>.
22. Merzendorfer, H. (2006). Insect chitin synthases: a review. *J. Comp. Physiol. B* **176**, 1–15.
23. Liu, R., Xu, C., Zhang, Q., Wang, S., and Fang, W. (2017). Evolution of the chitin synthase gene family correlates with fungal morphogenesis and adaptation to ecological niches. *Sci. Rep.* **7**, 44527.
24. Veronico, P., Gray, L.J., Jones, J.T., Bazzicalupo, P., Arbucci, S., Cortese, M.R., di Vito, M., and de Giorgi, C. (2001). Nematode chitin synthases: gene structure, expression and function in *Caenorhabditis elegans* and the plant parasitic nematode *Meloidogyne artiellia*. *Mol. Genet. Genomics* **266**, 28–34.
25. Zhang, Y., Foster, J.M., Nelson, L.S., Ma, D., and Carlow, C.K.S. (2005). The chitin synthase genes *chs-1* and *chs-2* are essential for *C. elegans* development and responsible for chitin deposition in the eggshell and pharynx, respectively. *Dev. Biol.* **285**, 330–339.
26. Sparacio, A.P., Trojanowski, N.F., Snetselaer, K., Nelson, M.D., and Raizen, D.M. (2020). Teething during sleep: ultrastructural analysis of pharyngeal muscle and cuticular grinder during the molt in *Caenorhabditis elegans*. *PLoS One* **15**, e0233059.
27. Witte, H., Moreno, E., Rödelserperger, C., Kim, J., Kim, J.S., Streit, A., and Sommer, R.J. (2015). Gene inactivation using the CRISPR/Cas9 system in the nematode *Pristionchus pacificus*. *Dev. Genes Evol.* **225**, 55–62.
28. Nakayama, K.I., Ishita, Y., Chihara, T., and Okumura, M. (2020). Screening for CRISPR/Cas9-induced mutations using a co-injection marker in the nematode *Pristionchus pacificus*. *Dev. Genes Evol.* **230**, 257–264.
29. Han, Z., Lo, W.S., Lightfoot, J.W., Witte, H., Sun, S., and Sommer, R.J. (2020). Improving transgenesis efficiency and CRISPR-associated tools through codon optimization and native intron addition in *Pristionchus* nematodes. *Genetics* **216**, 947–956.
30. Harris, M.T., and Fuhrman, J.A. (2002). Structure and expression of chitin synthase in the parasitic nematode *Dirofilaria immitis*. *Mol. Biochem. Parasitol.* **122**, 231–234.
31. Harris, M.T., Lai, K., Arnold, K., Martinez, H.F., Specht, C.A., and Fuhrman, J.A. (2000). Chitin synthase in the filarial parasite, *Brugia malayi*. *Mol. Biochem. Parasitol.* **111**, 351–362.
32. Werner, M.S., Sieriebriennikov, B., Prabh, N., Loschko, T., Lanz, C., and Sommer, R.J. (2018). Young genes have distinct gene structure, epigenetic profiles, and transcriptional regulation. *Genome Res.* **28**, 1675–1687.
33. Harry, C.J., Messar, S.M., and Ragsdale, E.J. (2022). Comparative reconstruction of the predatory feeding structures of the polyphenic nematode *Pristionchus pacificus*. *Evol. Dev.* **24**, 16–36.
34. Baldwin, J.G., Eddleman, C.D., Giblin-Davis, R.M., Williams, D.S., Vida, J.T., and Thomas, W.K. (1997). The buccal capsule of *Aduncospiculum halicti* (Nemata: Diplogasterina): an ultrastructural and molecular phylogenetic study. *Can. J. Zool.* **75**, 407–423.
35. Bumbarger, D.J., Crum, J., Ellisman, M.H., and Baldwin, J.G. (2006). Three-dimensional reconstruction of the nose epidermal cells in the microbial feeding nematode, *Acrobeles complexus* (Nematoda: Rhabditida). *J. Morphol.* **267**, 1257–1272.
36. Ragsdale, E.J., Crum, J., Ellisman, M.H., and Baldwin, J.G. (2008). Three-dimensional reconstruction of the stomatostylet and anterior epidermis in the nematode *Aphelenchus avenae* (Nematoda: Aphelenchidae) with implications for the evolution of plant parasitism. *J. Morphol.* **269**, 1181–1196.
37. Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C., and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. *Wilehm Roux Arch. Dev. Biol.* **193**, 283–295.
38. Merzendorfer, H., and Zimoch, L. (2003). Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *J. Exp. Biol.* **206**, 4393–4412.
39. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12.
40. Kim, D., Langmead, B., and Salzberg, S.L. (2015). hisAt: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360.
41. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.
42. Sigg, C.D., and Buhmann, J.M. (2008). Expectation-maximization for sparse and non-negative PCA. In *Proceedings of the 25th International Conference on Machine Learning*, pp. 960–967.
43. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
44. Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **4**, P3.
45. R Core Team (2020). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
46. Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121.
47. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.
48. Subramanian, B., Gao, S., Lercher, M.J., Hu, S., and Chen, W.H. (2019). Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* **47**, W270–W275.
49. Moore, B.T., Jordan, J.M., and Baugh, L.R. (2013). WormSizer: high-throughput analysis of nematode size and shape. *PLoS One* **8**, e57142.
50. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682.
51. Sommer, R., Carta, L.K., Kim, S., and Sternberg, P.W. (1996). Morphological, genetic and molecular description of *Pristionchus pacificus*. *Fundam. Appl. Nematol.* **19**, 511–521.
52. Theska, T., Sieriebriennikov, B., Wighard, S.S., Werner, M.S., and Sommer, R.J. (2020). Geometric morphometrics of microscopic animals as exemplified by model nematodes. *Nat. Protoc.* **15**, 2611–2644.

53. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
54. Werner, M.S., Sieriebriennikov, B., Loschko, T., Namdeo, S., Lenuzzi, M., Dardiry, M., Renahan, T., Sharma, D.R., and Sommer, R.J. (2017). Environmental influence on *Pristionchus pacificus* mouth form through different culture methods. *Sci. Rep.* 7, 7207.
55. Schlager, B., Wang, X., Braach, G., and Sommer, R.J. (2009). Molecular cloning of a dominant roller mutant and establishment of DNA-mediated transformation in the nematode *Pristionchus pacificus*. *Genesis* 47, 300–304.
56. Sieriebriennikov, B., Markov, G.V.v., Witte, H., and Sommer, R.J. (2017). The role of DAF-21/Hsp90 in mouth-form plasticity in *Pristionchus pacificus*. *Mol. Biol. Evol.* 34, 1644–1653.
57. Dardiry, M., Piskobulu, V., Kalirad, A., and Sommer, R.J. (2022). Experimental and theoretical support for costs of plasticity and phenotype in a nematode cannibalistic trait. Preprint at bioRxiv. <https://doi.org/10.1101/2022.02.28.482339>.
58. Lightfoot, J.W., Wilecki, M., Rödelsperger, C., Moreno, E., Susoy, V., Witte, H., and Sommer, R.J. (2019). Small peptide-mediated self-recognition prevents cannibalism in predatory nematodes. *Science* 364, 86–89.

Q6 Q7 STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
<i>Escherichia coli</i> : OP50	Caenorhabditis Genetics Center	WBStrain00041969
Chemicals, peptides, and recombinant proteins		
Nikkomycin Z	Sigma	Cat#59456-70-1
Chitinase	Sigma	Cat#C6137
tracrRNA	Integrated DNA Technologies	Cat#1072534
Deposited data		
Bulk RNA-seq of the lethargus in the last molt in <i>P. pacificus</i> wild type PS312	Sun et al. ⁴	ENA: PRJEB42635
Bulk RNA-seq of the last molt in <i>P. pacificus</i> wild type PS312 and <i>eud-1(tu1069)</i>	This paper	ENA: PRJEB53545
Experimental models: Organisms/strains		
<i>Pristionchus pacificus</i> : strain PS312	Sommer Lab	RRID:WB-STRAIN:PS312
<i>Pristionchus pacificus</i> : strain RS3897: <i>Ppa-chs-1(tu1634)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS3898: <i>Ppa-chs-2(tu1635)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS3907: <i>Ppa-chs-2(tu1636)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS4007: <i>Ppa-chs-2(tu1702)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS4008: <i>Ppa-chs-2(tu1703)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS4009: <i>Ppa-chs-2(tu1704)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS3911: <i>Ppa-chs-2(tu1651)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS3912: <i>Ppa-chs-2(tu1652)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS3913: <i>Ppa-chs-2(tu1653)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain h1: <i>Ppa-chs-2(h1)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain h2: <i>Ppa-chs-2(h2)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain h3: <i>Ppa-chs-2(h3)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS4117: <i>Ppa-chs-2(tu1787)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS4118: <i>Ppa-chs-2(tu1788)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS4119: <i>Ppa-chs-2(tu1789)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS3905: <i>Ppa-dpy-6(tu1645)</i>	Sun et al. ³	N/A
<i>Pristionchus pacificus</i> : strain RS3917: <i>Ppa-dpy-6;Ppa-chs-2(tu1645;tu1652)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS3111: <i>Ppa-eud-1(tu1069)</i>	Sieriebriennikov et al. ¹²	N/A
<i>Pristionchus pacificus</i> : strain RS3919: <i>Ppa-eud-1;Ppa-chs-2(tu1069;tu1652)</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS3930: <i>tuEx351[Ppa-chs-2p::GFP::rpl-23utr; Ppa-egl-20p::TurboRFP::rpl-23utr]</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS3931: <i>tuEx352[Ppa-chs-2p::GFP::rpl-23utr; Ppa-egl-20p::TurboRFP::rpl-23utr]</i>	This paper	N/A
<i>Pristionchus pacificus</i> : strain RS4093: <i>tuEx352[Ppa-chs-2p::GFP::rpl-23utr; Ppa-egl-20p::TurboRFP::rpl-23utr; Ppa-eud-1(tu1069)</i>	This paper	N/A
<i>Caenorhabditis elegans</i> : N2: wild type	Caenorhabditis Genetics Center	NA
<i>Caenorhabditis elegans</i> : strain RS4176: <i>tuEx369[Ppa-chs-2p::Ppa-chs-2cgs::Ppa-chs-2utr; Cel-sur-5p::GFP]</i>	This paper	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Caenorhabditis elegans</i> : strain RS4177: <i>tuEx370[Ppa-chs-2p::Ppa-chs-2cds::Ppa-chs-2utr;Cel-sur-5p::GFP]</i>	This paper	N/A
<i>Caenorhabditis elegans</i> : strain RS4178: <i>tuEx371[Ppa-chs-2p::Ppa-chs-2cds::Ppa-chs-2utr;Cel-sur-5p::GFP]</i>	This paper	N/A
Oligonucleotides		
sgRNA target sequence: exon 5 of <i>Ppa-chs-1</i> : TTGCATCATCTGCGTATATA	This paper	N/A
sgRNA target sequence: exon 3 of <i>Ppa-chs-2</i> : GCCACCGGTGCTCCACAACT	This paper	N/A
sgRNA target sequence: exon 33 of <i>Ppa-chs-2</i> : ACGTTTATCCTGACTCTCGA	This paper	N/A
sgRNA target sequence: exon 51 of <i>Ppa-chs-2</i> : GAGCAGGACTTATCGACTTG	This paper	N/A
sgRNA target sequence: exon 56 of <i>Ppa-chs-2</i> : GCGATTCGGGGAGTAGACGA	This paper	N/A
sgRNA target sequence: EDR motif in exon 36 of <i>Ppa-chs-2</i> : GACTACATCCAATATGATCA	This paper	N/A
sgRNA target sequence: QRRRW motif in exon 37 of <i>Ppa-chs-2</i> : GCTCAGACCTTCGCTCCGGA	This paper	N/A
Primer: Amplification of <i>Ppa-chs-2</i> promoter for Gibson assembly, Forward: CCTCTCCCTTAGACATCTGAAAATCCTCATTCTATGATTGATTATGTCCAGAGC	This paper	N/A
Primer: Amplification of <i>Ppa-chs-2</i> promoter for Gibson assembly, Reverse: CCTGCAGGTGACGTCGGGCAAGA ACTTCCGCCAGCAGG	This paper	N/A
Primer: Linearizing the plasmid of pUC19- <i>Ppa-egl-20::GFP::rpl-23utr</i> for Gibson assembly, Forward: CCGACGTCGA CCTGCAGG	This paper	N/A
Primer: Linearizing the plasmid of pUC19- <i>Ppa-egl-20::GFP::rpl-23utr</i> for Gibson assembly, Reverse: TTTTCAGATGTCTAAGGGAGAGG	This paper	N/A
Primer: Gene specific primer (ss0338) for 3' RACE PCR: CTAAAATTTGTTCTAGCAGTTTCACAATTGTC	This paper	N/A
Primer: Short UPM primer (ss0360) for 3' RACE PCR: AAGCAGTGGTATCAACGCAGAGT	This paper	N/A
Primer: Nest PCR, Forward (ss0298): ATGGGCGCAACTGATCCAACCTG	This paper	N/A
Primer: Nest PCR, Reverse (ss0301): CGCCTTAGGTTGATCGAATAATCG	This paper	N/A
Primer: Nest PCR, Reverse (ss0302): GATGTCGGTGCGCCTTAGGTTGATCG	This paper	N/A
Recombinant DNA		
Plasmid: pUC19- <i>Ppa-chs-2p::GFP::rpl-23utr</i>	This paper	N/A
Plasmid: pUC19- <i>Ppa-egl-20p::TurboRFP::rpl-23utr</i>	Han et al. ²⁹	N/A
Plasmid: pUC19- <i>Ppa-egl-20p::GFP::rpl-23utr</i>	Han et al. ²⁹	N/A
Software and algorithms		
cutadapt (version 2.10)	Martin et al. ³⁹	https://cutadapt.readthedocs.io/en/stable/
Hisat2 (version 2.1.0)	Kim et al. ⁴⁰	https://daehwankimlab.github.io/hisat2/
FeatureCounts (version 1.6.4)	Liao et al. ⁴¹	https://sourceforge.net/projects/subread/files/
t-arae/ngscmdr (version 0.1.0.181203)	N/A	https://rdr.io/github/t-arae/ngscmdr/
princomp	Sigg et al. ⁴²	https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/princomp

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DESeq2 (version 1.18.1)	Love et al. ⁴³	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
DAVID (v6.8)	Dennis et al. ⁴⁴	https://david.ncifcrf.gov/
R (version 4.2.2)	R Core Team ⁴⁵	https://www.r-project.org/
HMMER (version 3.3)	Mistry et al. ⁴⁶	http://hmmer.org/download.html
MAFFT (ver. 7.310)	Katoh et al. ⁴⁷	https://mafft.cbrc.jp/alignment/server/
Evolview (ver. 3)	Subramanian et al. ⁴⁸	https://www.evolgenius.info/evolview/
Wormsizer	Moore et al. ⁴⁹	https://github.com/bradtmooore/wormsizer
ImageJ	Schindelin et al. ⁵⁰	https://imagej.nih.gov/ij/download.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Ralf J. Sommer. ralf.sommer@tuebingen.mpg.de

Materials availability

P. pacificus strains and bacterial isolates generated in this work are freely available through the lead contact.

Data and code availability

Sequencing data have been deposited at European Nucleotide Archive, and accession numbers are listed in the [key resources table](#). All data and codes reported in this work are freely available through the lead contact.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Stocks of all strains were cultured at room temperature (20–25°C) on nematode growth medium (NGM) in 6 cm Petri dishes, as outlined before.⁵¹ In all experiments, *Escherichia coli* OP50 was used as food source. Bacteria were grown overnight at 37°C in LB medium, and 400 µL of the overnight culture was pipetted on NGM agar plates and left for several days at room temperature to grow bacterial lawns. Nematodes were passed on these lawns and propagated using various numbers of mixed developmental stages. The following strains were used in these studies: PS312 as *P. pacificus* wild type; *eud-1(tu1069)*, an all-St strain.

METHOD DETAILS

Mouth form phenotyping

To phenotype the mouth form of individual animals we used previously established protocols.⁵² In short, adult hermaphrodites were immobilized on 5% Noble Agar pads with 0.3% NaN₃ added as an anaesthetic, and examined using differential interference contrast (DIC) microscopy. Animals that had a large right ventrosublateral tooth, curved dorsal tooth, and the anterior tip of the promeso-stegostom posterior to the anterior tip of the gymnostom were classified as Eu morphs. Animals that did not exhibit these three characters simultaneously were classified as St morphs.

Phylogenetic analysis of chitin synthase genes in nematode

To identify putative chitin synthase genes (CHSs) in nematodes, we collected proteome data of 10 publicly available nematode genomes from WormBase (<http://www.Caenorhabditis.org>) and <http://www.pristionchus.org>. The HMMER⁴⁶ version 3.3 with parameters “hmmscan –tblout Pfam-A.hmm” was applied to predict Pfam domains (e-value of full seq < 0.01) of each of the nematode proteomes (the longest isoform per gene). As two annotated chitin synthases in the *C. elegans* genome contain the domain “Chitin_synth_2” (CS2, PF03142) rather than the domain of Chitin_synth_1 (CS1, PF01644) or Chitin_synth_1N (CS1N, PF08407),^{24,25,31} we used the CS2 domain to perform phylogenetic analysis of CHSs in nematodes. Specifically, the genes that contain the CS2 domain (best match) were marked as putative CHSs and were confirmed to contain the chitin synthase domain using the Pfam database and the SMART database. Protein sequences of the CHSs were aligned using MAFFT⁴⁷ (ver. 7.310) with default parameters and maximum likelihood trees were constructed using RAxML (ver. 8.2.11)⁵³ with the following parameters “fa -m PROTGAMMAAUTO -N 100”. The resulting phylogenetic trees were visualized using Evolview (ver. 3).⁴⁸ The protein sequences of 23 identified chitin synthase genes are provide in [Data S1](#).

Alternative splicing of *Ppa-chs-2*

Synchronous populations of PS312 eggs-J1 were obtained by bleaching⁵⁴ and pipetted onto 10cm NGM plates seeded with 800 μ l OP50. Worms were cultured at 20° C and harvested at 48 h post-bleaching when most worms reached the late J4 stage or the fourth molt. Animals were washed from plates with M9 buffer into 15 ml conical tubes, and filtered two times through a 140 μ m nylon net (Millipore) to remove contaminating bacteria, then centrifuged 1300 x g, 1 minute to pellet. Supernatant was removed, the pellet was resuspended in 300 μ l water and frozen for RNA extraction.

Total RNA was extracted using Ambion™ TRIzol reagent. TURBO DNA-free kit (Invitrogen AM1907) was used to remove gDNA. RNA purity and yield was calculated using a nanodrop spectrophotometer. The SMARTer RACE 5'/3' Kit (Cat. # 634858) was employed for reverse transcription, 5' and 3' RACE experiments with slight modifications to the manufacturer guidelines. Briefly, the first-strand cDNA was diluted to 50 ng/ μ l for the RACE experiment. 3' RACE PCR was performed using primers 3' Gene-Specific-Primer (GSP) ss0338 and Universal-Primer-Short (UPM short). Touchdown PCR was conducted by the two following PCR programs. Program 1: Initial denaturation at 94 °C for 3 min; five cycles with two steps of 94 °C for 30 sec, 72 °C for 3.5 min; five cycles with three steps, 94 °C for 30 sec, 71 °C for 30 sec, 72 °C for 3.5 min; 35 cycles with three steps, 94 °C for 30 sec, 69 °C for 30 sec, 72 °C for 3.5 min; Final extension at 72 °C for 10 min. Program 2: Initial denaturation at 94 °C for 3 min; five cycles with two steps of 94 °C for 30 sec, 72 °C for 3.5 min; five cycles with three steps, 94 °C for 30 sec, 70 °C for 30 sec, 72 °C for 3.5 min; 35 cycles with three steps, 94 °C for 30 sec, 68 °C for 30 sec, 72 °C for 3.5 min; Final extension at 72 °C for 10 min (Figure S3A). The nested PCR for the amplification of the four different isoforms was conducted using the GoTaq Long PCR Master Mix (M4021, Promega), 1 μ l production of the 3'RACE PCR was added as template, including the primer pair ss0298+ss0301 for isoform *Ppa-chs-2a* amplification, the primer pair UPM short+ss0302 for the amplification of three short isoforms *Ppa-chs-2b-d* (Figures S3B and S3C; Data S1). Amplified fragments were cloned using the TA Cloning Kit from Invitrogen (K202020) and were subsequently Sanger sequenced. The primers designed in the 3' RACE experiment is provided in [key resources table](#).

CRISPR/Cas9 mutagenesis

Procedure for CRISPR/Cas9 mutagenesis was based on previously published protocols for *P. pacificus*.^{27–29} All target-specific CRISPR RNAs (crRNAs) were synthesized to target 20 bp upstream of the protospacer adjacent motifs (PAMs), and were fused to tracrRNA (catalog# 1072534; IDT) at 95° C for 5 minutes, and subsequently allowed to cool down to room temperature and anneal. The hybridization product was combined with Cas9 protein (catalog# 1081058; IDT). After a further 5 minutes incubation at room temperature, TE buffer was added for final concentration of 18.1 μ M for the sgRNA and 12.5 μ M for Cas9. For the induction of specific site-directed mutations via CRISPR/Cas9, a ssDNA oligo template was contained in the mix at a concentration of 4 μ M. The repair template included the desired modifications flanked by 80 bp homology arms at both sides of the edited sequence. The plasmid carrying the *Ppa-efl-3* promoter and modified TurboRFP sequences was employed as co-injection marker. After two days, P0 plates containing the F1 animals with fluorescent signal of co-injection marker were isolated, and 8–10 F1 progenies from these plates were singled out on individual plates. The genotype of the F1 animals were subsequently analyzed via Sanger sequencing and mutations identified before isolation of homozygous mutant carriers. Detailed information of mutant generated in this study can be found in [key resources table](#). sgRNAs, and repair templates for generating mutants and specific targeted knock ins utilized in this study can be found in [key resources table](#).

Genetic transformation

To generate a transcriptional reporter of *Ppa-chs-2* in *P. pacificus*, we used the previously established transformation protocol.⁵⁵ A transcriptional reporter construct for *Ppa-chs-2* was generated by PCR amplification of a 1926 bp upstream region in front of the first predicted ATG start codon into a pUC19-based plasmid containing a green fluorescent protein (GFP) and the 3'UTR of the ribosomal gene *Ppa-rpl-23*. NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) was employed to perform cloning. Injection mix contained 10 ng/ml of the NotI-digested *Ppa-chs-2p::GFP* plasmid, 10 ng/ml of the NotI-digested *Ppa-egl-20p::TurboRFP* plasmid as co-injection marker, and 60 ng/ml of the NotI-digested genomic carrier DNA.

To overexpress the 16 kb *Ppa-chs-2* in *C. elegans*, an appr. 16 kb *Ppa-chs-2* fragment containing the 1926 putative promoter, all exons and introns of the gene, and the 3'UTR were amplified using Long PCR Enzyme Mix (#K0182, Thermo Scientific). The 16 kb *Ppa-chs-2* fragment (30 ng/ μ l) was injected together with the co-injection marker *Cel-sur-5::GFP* (10 ng/ μ l). The same marker gene expression was confirmed in two independent lines, but no changes in the mouth and pharynx were observed. Associated primers for generating reporter utilized in this study can be found in [key resources table](#).

Spatiotemporal gene expression profiling using published transcriptomics

To analyze the temporal expression of *Ppa-chs-1* and *Ppa-chs-2* during postembryonic development, we used the high-resolution single-worm transcriptomics dataset and profiled the gene expression pattern for both genes using the mean gene expression (FPKM) at 38 time points as previously described.⁴ The spatial gene expression of *Ppa-chs-1* was plotted using the spatial transcriptomics of *P. pacificus*.⁵ The normalized z-score value of *Ppa-chs-1* was extracted and the gene expression pattern was profiled across nine distinct body regions.

Tooth size measurement in *P. pacificus*

We developed a protocol for teeth size measurement based on landmarks, which were first described in geometric morphometric studies in *P. pacificus*.^{3,52,56} We estimated teeth size by calculating the area of the triangle based on three most lateral landmarks on each tooth (Figure S2). This triangle covers most parts of the tooth. We used five-day old adult animals to take a stack of images of the head region using Nomarski differential interference contrast microscopy under a 100x/1.4 oil objective. Coordinates of six landmarks in XY plane were recorded using ImageJ software.⁵⁰ The raw data of tooth size measurement in this study can be found in Data S2.

Body length measurement

To remove the age effect on body size for wild type and mutant strains, we selected five-day old adult worms to measure their body length, as it is known that the maximum body length is reached at four days adulthood.³ First, we synchronized the stage of animals by picking late J4 larval individuals that displayed “box-like” vulva shape. After five days, we singled out the five-day old adults on NGM plates without bacteria. Bright field images of the worms were taken using 0.63x objective of ZEISS Stereo Discovery V16 and the AxioCam camera with parameters “Magnification 40x, Field of view 5.8 mm, Resolution 1.1 μm , Depth of field 26 μm ”. Images were analyzed using the WormSizer plug-in⁴⁹ for Image J.⁵⁰ The raw data of worm size measurement in this study can be found in Data S2.

Overall self-fecundity measurement

Overall and daily self-fecundity measurements were conducted as previously described⁵⁷ with minor modifications. Briefly, synchronized eggs were obtained by bleaching.^{39,54} After synchronization, late J4 larvae with ‘box-like’ vulva shape were individually isolated on separate NGM plates seeded with 25 μl OP50. For five consecutive days, single worms were transferred to fresh plates spotted with 25 μl OP50 every 24 hours. Starting from day 6, worms were kept on the same plate for two more days and then killed. A daily readout for the first five days and a day 6 readout recording the last three days combined. After four days from removing the mother, all plates were counted for viable progeny to obtain fecundity counts. Overall self-fecundity counts were calculated as sum of six daily fecundity counts. The raw data of overall and daily self-fecundity measurements can be found in Data S2.

Microinjection nikkomycin Z in *P. pacificus*

To study the effects of the chitin synthase domain-specific inhibitor nikkomycin Z on mouth form in *P. pacificus*, we performed two different experiments. First, we used NGM plates with 25 μl of a mixture of nikkomycin Z and OP50, at three concentrations (0 μM (control), 10 μM and 100 μM nikkomycin Z (Cat. # 59456-70-1; Sigma)) of nikkomycin Z described above. Animals at J2 or middle J4 stage were transferred onto these NGM plates and cultured at 20° C. Mouth-form observation of adult animals DIC revealed no visible changes in the *P. pacificus* mouth form under any treatment condition. Second, we directly injected Nikkomycin Z in the head region of J4 larval animals, 2-3 hours before the last molt. Two injection sites were selected: P1, the pharyngeal muscle cells in the metacarpus, which corresponds to the cells expressing *Ppa-chs-2*, and P2, the intercellular space in the isthmus. Nikkomycin Z-injected animals were kept individually on NGM plate with 50 μl of OP50 and cultured at 20° C. After 24 hours, adult animals were investigated for their mouth structure and form. Nearly 90% injected animals survived the injection into P2 with 10 mM or 0 mM concentration of Nikkomycin Z, respectively. To examine the different responses to Nikkomycin Z between the PS312 strain (preferentially Eu morph) and the *eud-1(tu1069)* strain (St morph only), six different concentrations of Nikkomycin Z were used, including 0 mM (control), 0.1 mM, 0.5 mM, 1 mM, 5 mM and 10 mM were injected in P2 of late J4 worms of the two strains. In total, 15 – 26 surviving animals were investigated in the particular conditions.

To compare the differences in response to the Nikkomycin Z treatment between the two alternative mouth forms, morphs of Nikkomycin Z injected animals were classified into three different groups based on teeth size (Figures 4E and S4). Specifically, in wild type animals, the “normal teeth” group was classified by the criteria of one of two teeth (dorsal tooth S_D and subventral tooth S_{RV}) being at least larger than 4 μm^2 ($S_D > 4 \mu\text{m}^2$ or $S_{RV} > 4 \mu\text{m}^2$), the “small teeth” group was defined by the criteria of both teeth being equal to or smaller than 4 μm^2 and one of the two teeth being at least larger than 2 μm^2 ((i) $S_D \leq 4 \mu\text{m}^2$ and $S_{RV} \leq 4 \mu\text{m}^2$, (ii) either $S_D > 2 \mu\text{m}^2$ or $S_{RV} > 2 \mu\text{m}^2$) and the “teethless” group was defined by the criteria of both teeth being equal to or smaller than 2 μm^2 ($S_D \leq 2 \mu\text{m}^2$ and $S_{RV} \leq 2 \mu\text{m}^2$). Similarly, in the St morph strain of *eud-1(tu1069)* animals, the “teethless” group was defined by the criteria of $S_D \leq 1.5 \mu\text{m}^2$ and $S_{RV} \leq 1 \mu\text{m}^2$ simultaneously, whereas the “small teeth” group was defined by either $S_D > 1.5 \mu\text{m}^2$ or $S_{RV} > 1 \mu\text{m}^2$, (Figures 4E and S4). The raw data of tooth size measurement associated with Nikkomycin Z experiment can be found in Data S2.

Bulk RNA sequencing and transcriptome data analysis

To perform the high-resolution transcriptomics of the fourth molt between wild type and *Ppa-eud-1* mutants, we collected worms at the precise substage for Bulk RNA sequencing using the protocol described previously.⁴ Synchronized eggs were cultured as described above. Individual worms at the appropriate substage were picked to one PCR tube containing 3 μl nuclease-free water, then conducted to 3 x freeze-thawing steps with liquid nitrogen. 70 – 100 individuals with appropriate substage were pooled into one 1.5-mL tube as one replicate. Three independent replicates were collected for each of the four substages (J4.12, 4M1, 4M2, and 4M3) in both strains (PS312 and *Ppa-eud-1(tu1069)*). Total RNA extraction of all 24 samples was performed using the Direct-Zol RNA Mini prep kit (Zymo Research) according to the manufacturer’s guidelines. RNA library preparation was conducted using the NEBNext

Ultra II Directional RNA Library Prep Kit for Illumina (NEB). All 24 libraries were sequenced as 150-bp PE on the Illumina HiSeq3000 (Table S1).

Raw reads were trimmed by the program cutadapt⁴⁰ version 2.10. Genomic alignment to the *P. pacificus* reference genome (<http://pristionchus.org/cgi-bin/news.cgi>, version: El Paco) was performed by using the software Hisat2⁴¹ (version 2.1.0). FeatureCounts⁴⁵ (version 1.6.4) was employed to quantify transcripts based on the *P. pacificus* reference genome (version El Paco gene annotation 3). Gene expression value were calculated as TPM by the package t-arae/ngscmdr (version 0.1.0.181203) in R.⁴² Mean TPM was calculated from three replicates in each of the eight conditions (two animal strains and four sub-stages), and genes with mean TPM greater than 1 at least in one of the eight conditions were defined as reliably expressed genes (REGs). 15,392 REGs in total and more than 11,000 REGs in the particular condition were detected (Figure S5D; Table S1). Principal component analysis (PCA) on transcriptomes was conducted using the function princomp⁴³ in R with default parameters. Differential expression analysis was performed using DESeq2⁴⁴ (version 1.18.1) with two criteria, *P*-value < 0.05 and Fold change > 2. In total, 2298 differentially expressed genes (DEGs) regulated by *Pa-eud-1* at the fourth molt were identified (Figure S5E; Table S1). Gene Ontology (GO) analysis was conducted by using DAVID⁵⁸ v6.8 web-accessible programs (Figure S5F). UniProt accession of *P. pacificus* genes were annotated based on the previous published method.⁴ Detailed on GO analysis of the top 18 GO-terms to see Table S1.

Predatory behavior assay

Q9 Killing assays were conducted by following the previously published protocol.⁵⁹ Briefly, *C. elegans* prey was maintained on bacteria until freshly starved, resulting in an abundance of young larvae. These plates were washed with M9, passed twice through a 20 μ m filter, centrifuged at 377 \times g for 1 minute and deposited on to the assay plate by pipetting 2 μ l of worm pellet on to a 6 cm NGM unseeded plate. Assay plates were left for a minimum of 1 hour to allow larvae to distribute evenly over the plate. Young adult predatory nematodes were firstly transferred to an empty NGM plates for 30min to remove any excess bacteria from their bodies. Five predatory nematodes were added to assay plates. Predators were permitted to feed on the prey for two hours and the plate was subsequently screened for the presence of dead corpses. Each assay was conducted in triplicate to ensure repeatability. Kills were characterized by an opening of the prey cuticle.

QUANTIFICATION AND STATISTICAL ANALYSIS

For all the statistics analyses, Student's test with an adjective p-value using the false discovery rate (FDR) was used to compare comparisons of the differences between the control group and experimental groups. The One-way ANOVA with Tukey HSD is used for the compare more than three comparisons.

Current Biology, Volume 33

Supplemental Information

**Chitin contributes to the formation of a feeding
structure in a predatory nematode**

Shuai Sun, Hanh Witte, and Ralf J. Sommer

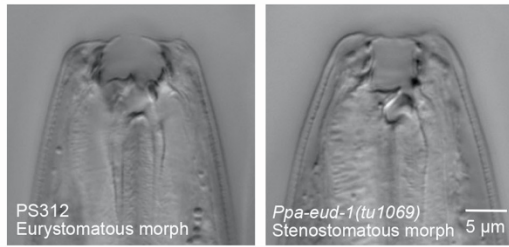
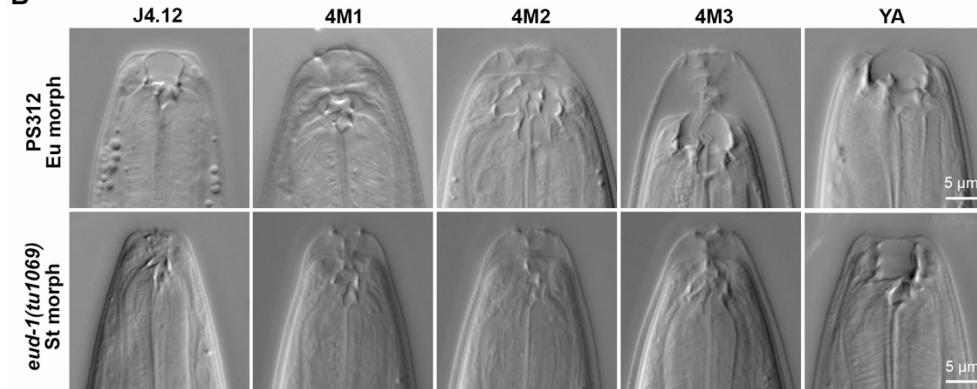
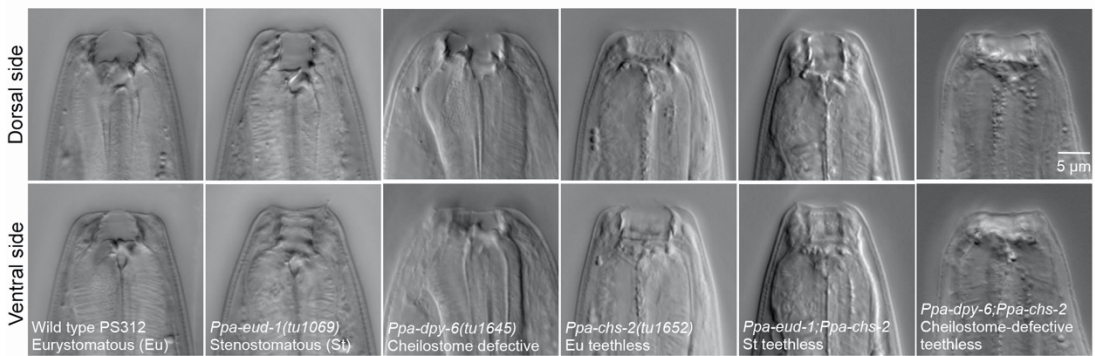
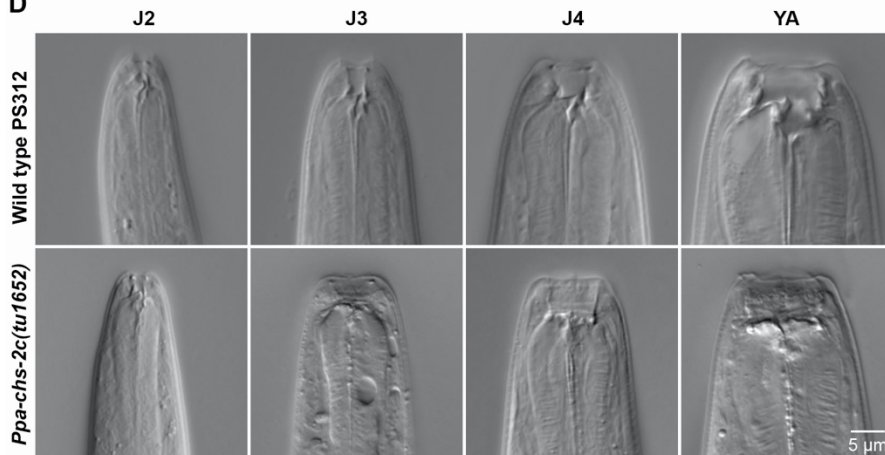
A**B****C****D**

Figure S1. Mouth-form images, related to Figure 1

Unlabeled images Figure S1A, S1B and S1C are related to Figure 1 A, B and Figure 3C, respectively. (D) Images display there are morphological differences in teeth shape between wild type PS312 and *Ppa-chs-2c(tu1652)* mutant among four distinct stages, including J2, J3, J4 and young adult. Scale bar, 5 μm .

A

Landmark	Morphological description	Type
1	lateral base point of the dorsal tooth	sliding
2	anterior tip of the dorsal tooth	fixed
3	posteriormost point of the metastegostom on dorsal side	fixed
4	lateral base point of the right ventrosublateral tooth	sliding
5	anterior tip of the right ventrosublateral tooth	fixed
6	ventral base point of the right ventrosublateral tooth	fixed

B

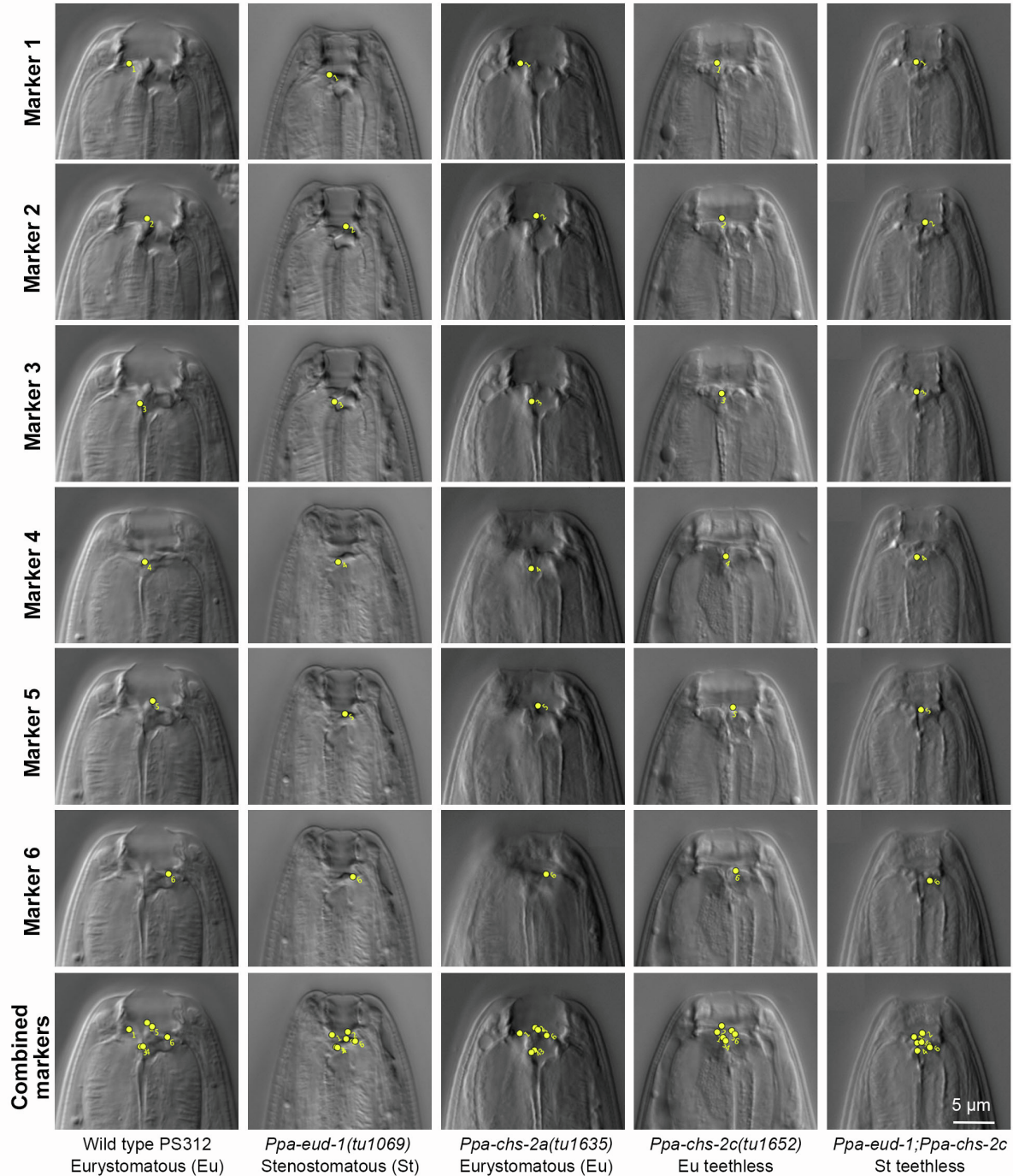
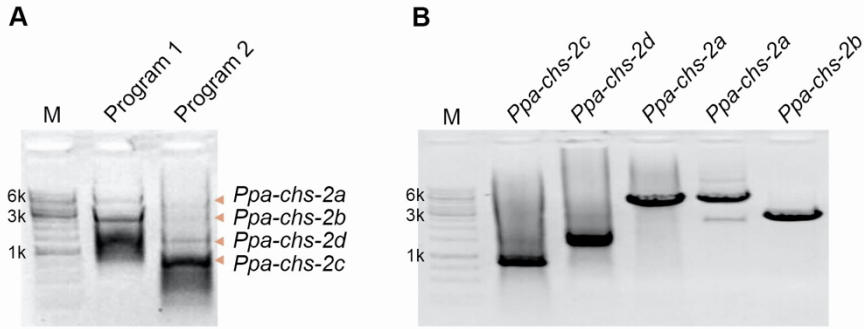


Figure S2. Landmarks used for the quantitative measurement of teeth size in *P. pacificus*, related to STAR methods (Morphometrics)

(A) Description of the six landmarks used in the measurement of tooth size. The six landmarks include Marker 1 (lateral base point of the dorsal tooth), Marker 2 (anterior tip of the dorsal tooth), Marker 3 (posteriormost point of the metastegostom on dorsal side), Marker 4 (lateral base point of the right ventrosublateral tooth), Marker 5 (anterior tip of the right ventrosublateral tooth), and Marker 6 (ventral base point of the right ventrosublateral tooth). (B) Images show the six landmarks labeled on the teeth for size measurement wild type and mutant animals. The images in the last row show the distribution of the six markers on the morph. Scale bar, 5 μm .



C
Detailed of the four isoforms in *Ppa-chs-2* locus

Isoform_ID	<i>Ppa-chs-2a</i>	<i>Ppa-chs-2b</i>	<i>Ppa-chs-2c</i>	<i>Ppa-chs-2d</i>
Isoform locus	ppa_stranded_DN30470_c0_g2_i1	TRINITY_DN178_c1_g1_j8	ppa_stranded_DN30470_c0_g2_i2	NA
Genome annotation	Trinity_2016	Trinity_2020	Trinity_2016	NA
	Trinity_2020			
	PPCAC_2021			
RACE confirmation	Yes	Yes	Yes	Yes
contain	e01 - e60	the 3' of e25, i25 & e26 - e60	the 3' of i52 & e53 - e60	the 3' of e44, e45 - e60
ATG location	e01	e26	e54	e45
CDS length bp (aa)	5016 (1672)	2955 (985)	657 (219)	1320 (440)

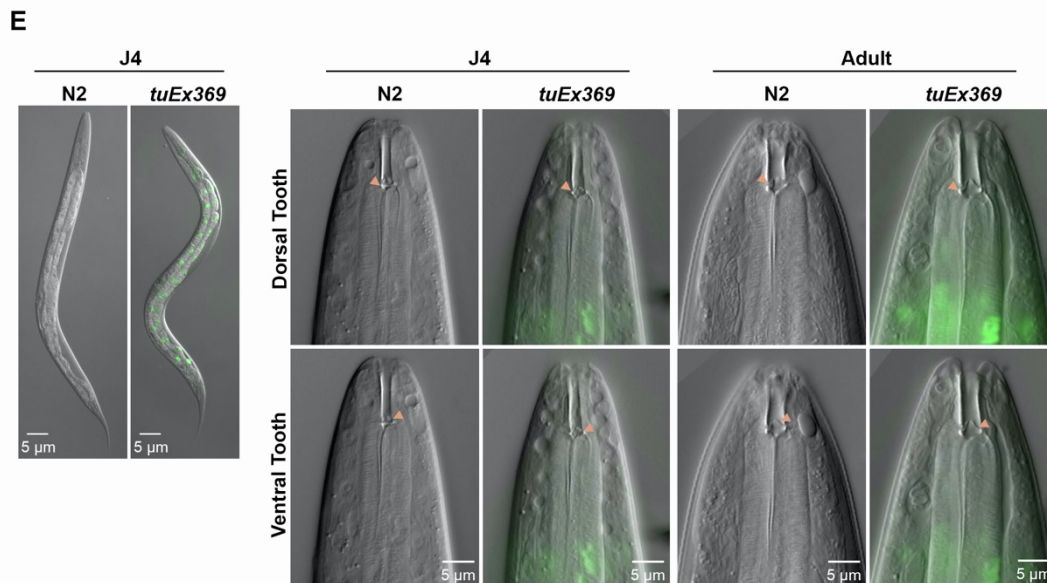
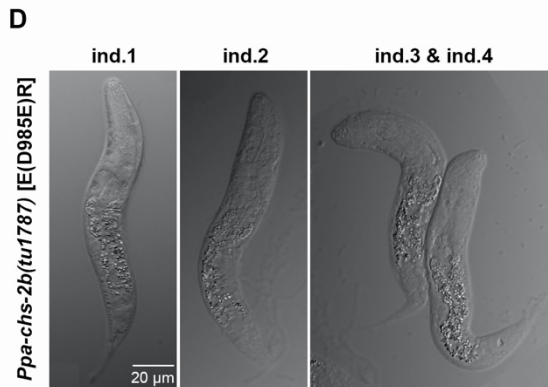


Figure S3. cDNA structure of four isoforms of *Ppa-chs-2* validated by RACE-PCR, related to Figure 3

(A) 3' RACE-PCR of *Ppa-chs-2* using primers 3' Gene-Specific-Primer (GSP) ss0338 and Universal-Primer-Short (UPM short) for touchdown PCR. M, molecular weight markers. (B) Nested-PCR amplification of isoform *Ppa-chs-2a-d* cDNA fragments, primers ss0298 and ss0301 for *Ppa-chs-2a* cDNA amplification, primers UPM short and ss0302 for *Ppa-chs-2b-d* cDNA amplification. M, molecular weight markers. (C) Detailed information of *Ppa-chs-2a-d* isoforms. *Ppa-chs-2a* was annotated by three previous transcriptome assemblies and confirmed by RACE-PCR. *Ppa-chs-2a* contains all 60 annotated exons and 1,672 amino acids are encoded in the CDS of *Ppa-chs-2a*. *Ppa-chs-2b* was confirmed by the transcriptome assembly in 2020 (Han et al. 2020) and RACE-PCR in this study. The cDNA of *Ppa-chs-2b* contains the right part of the 25th exon, 25th intron retention, and the sequence from 26th to 60th exon. *Ppa-chs-2b* encodes the chitin synthase domain and the transmembrane domains (TMD) of C-terminal but excludes the TMD in N-terminus. *Ppa-chs-2c* was the shortest isoform with transcriptome annotation by Serobyán et al (2016) and RACE-PCR validation. The cDNA of *Ppa-chs-2c* includes the right part of 52nd intron and sequence from 53rd to 60th exon, and encodes only the last copy of the C-terminal TMD. *Ppa-chs-2d* was only found in this RACE-PCR experiment. The sequence of *Ppa-chs-2d* starts from the right part of 44th exon to 60th exon, which encodes the most part of C-terminal TMD. (D) Images show four individuals of *Ppa-chs-2b(tu1787)* carrying a substitution in the putative acceptor saccharide binding site [E(D985E)R] exhibiting developmental arrest at the J2 stage. Scale bar, 20 μ m. (E) *Ppa-chs-2* overexpression in *C. elegans* did not result in *C. elegans* worms with teeth or pharyngeal defects. Left: overview of J4 animals in wild type N2 and *Ppa-chs-2* overexpression line *tuEx369*. Right: Mouth form comparison between N2 and *tuEx369* at late J4 and adult stages. The orange triangle indicates the position of dorsal or ventral tooth.

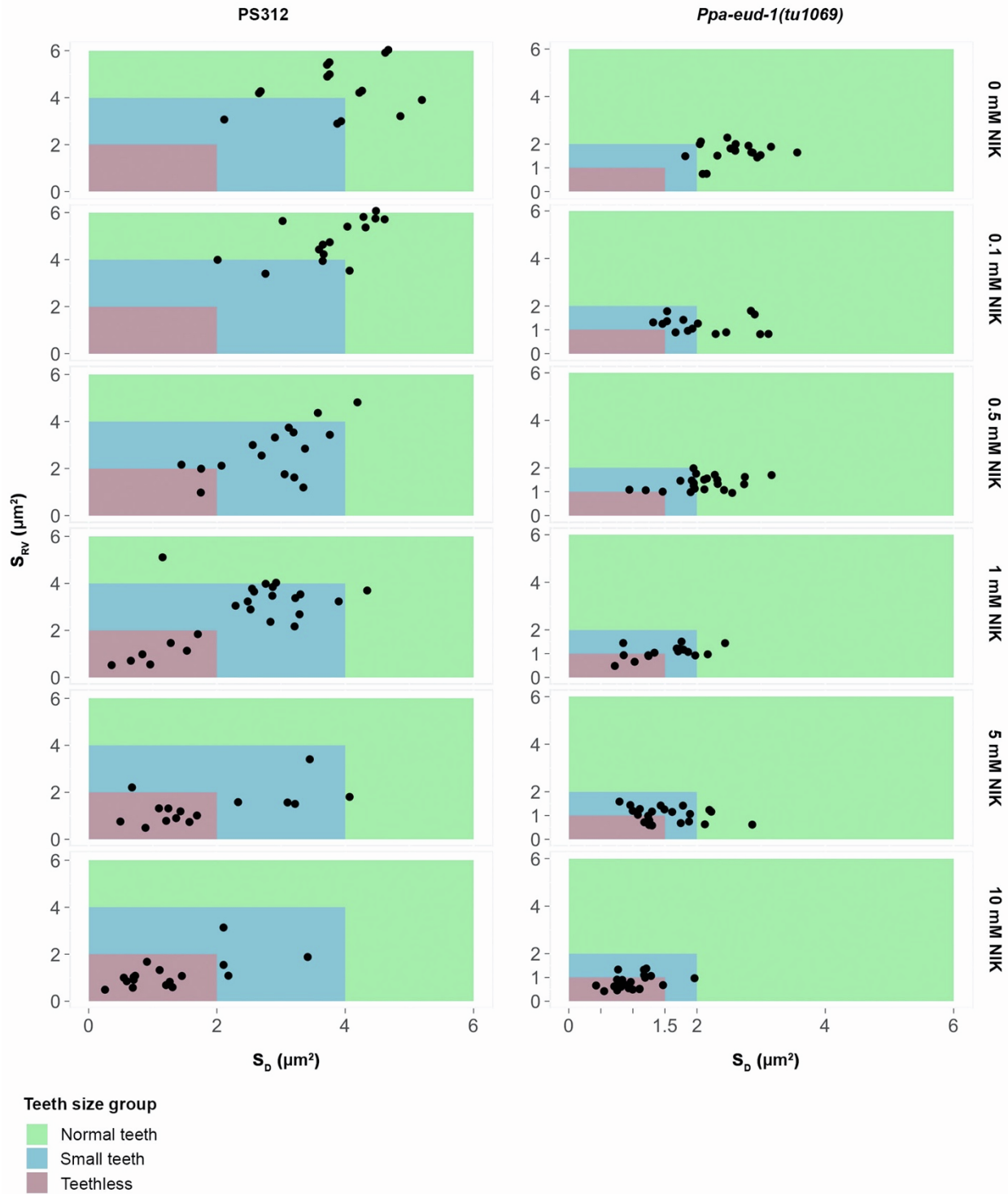


Figure S4. Characterization of three teeth size groups, related to Figure 4

The point plots show the different effects of Nikkomycin Z on Eu morphs in PS312 and the St morph in *Ppa-eud-1(tu1069)*. In PS312, the “teethless” group was classified by the criteria of both teeth (dorsal and subventral tooth) being smaller than $2 \mu\text{m}^2$, the “small teeth” group was clustered

by the criteria of both teeth being smaller than $4 \mu\text{m}^2$ and one of the two teeth being at least larger than $2 \mu\text{m}^2$, and the morph with one of two teeth being at least larger than $4 \mu\text{m}^2$ was grouped into the “normal teeth” group. Similarly, in the St morph strain *tu1069*, the “teethless” group was grouped by the criteria of $S_D \leq 1.5 \mu\text{m}^2$ and $S_{RV} \leq 1 \mu\text{m}^2$ simultaneously, the “small teeth” group was clustered by the criteria of i) both teeth are equal to or smaller than $2 \mu\text{m}^2$, and ii) either $S_D > 1.5 \mu\text{m}^2$ or $S_{RV} > 1 \mu\text{m}^2$, and the morph with one of two teeth at least being larger than $2 \mu\text{m}^2$ was characterized into the “normal teeth” group. S_D and S_{RV} indicate the size of dorsal tooth and right subventral tooth, respectively. The teeth size area of group “teethless”, “Small teeth” and “normal teeth” are labeled in pine, blue and green, respectively.

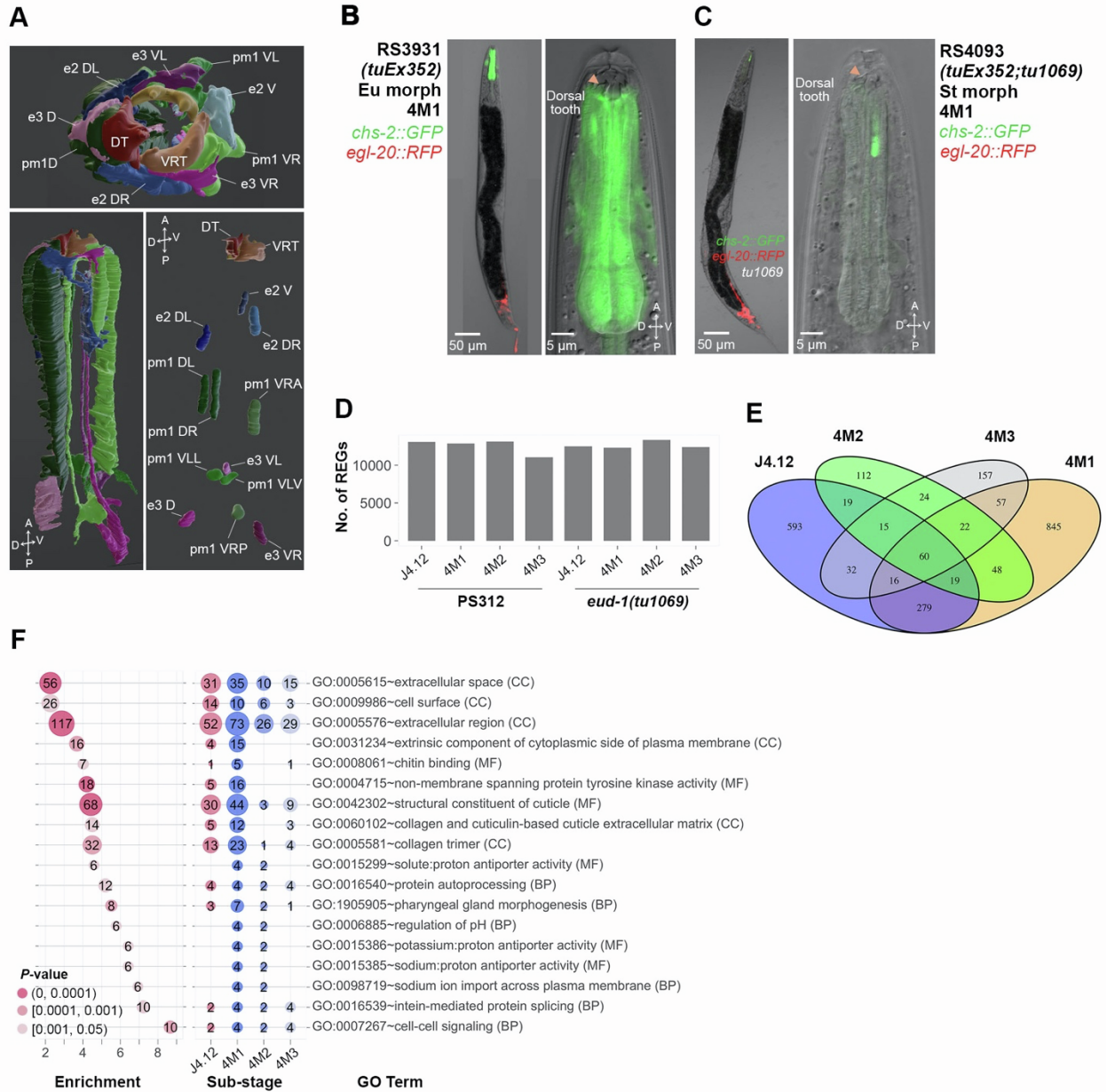


Figure S5. Spatiotemporal gene expression of *Ppa-chs-2* and its regulatory network, related to Figure 5

(A) Three-dimensional models of spatial expression pattern of *Ppa-chs-2* in the pm1, e2 and e3 cells at the fourth lethargus stage (4M1). Three-dimensional models of the *Ppa-chs-2* expression pattern was visualized using data from Harry (Harry et al. 2022) in Blender software, viewed en face (Top), in dorsal view (left bottom). The nuclei corresponding to these cells were labeled on the right bottom site. (B) *Ppa-chs-2* is highly expressed in the pm1, e2 and e3 cells at the substage 4M1 in the wild type background (*tuEx352*). Left: overlay of a differential interference contrast (DIC) image, TurboRFP and GFP channel. Right: overlay of a DIC image, sum-of-slices

projection of the GFP channel. The orange triangle indicates the position of dorsal tooth. D = dorsal, V = ventral, A = anterior, P= posterior. (C) Similar analysis for strain RS4093 (*tuEx352;Ppa-eud-1*) shows that *Ppa-chs-2* expression is significantly reduced in *Ppa-eud-1* mutants. (D) Bar plot displays the numbers of reliably expressed genes among these four substages for two strains, PS312 and *tu1069*. (E) Venn graph exhibits the distribution of the 2,298 differentially expressed genes (DEGs) among four distinct substages. (F) DEGs were tested for the tests for overrepresentation of GO (gene ontology). The left panel shows the 18 GO terms that are most significantly enriched among DEGs, the x-axis presents the enrichment score, and the y-axis shows the name of GO term. The size of circles corresponds to the number of DEGs with a given GO term, and the three different colors indicate significance levels as measured by Fisher's exact test. The three sub-ontologies containing biological process (BP), molecular function (MF), and cellular component (CC). The right panel profiles the overrepresented GO terms according to the four substages. The number of DEGs at each substage is marked.

1 **Protocol for single-worm transcriptomics in *Pristionchus pacificus*,** 2 **step by step**

3

4 Shuai Sun, Ralf Sommer

5

6 Max Planck Institute for Biology, Department for Integrative Evolutionary Biology, Max-Planck-
7 Ring 9, Tuebingen, 72076, Germany

8

9

10 **Abstract**

11

12 High-resolution spatiotemporal transcriptomic data presents the opportunity to gain key
13 insights into cellular and disease mechanisms. Adaptations and improvements of low input
14 RNA-seq technology for microscopic organisms, nematodes, allows for the analysis of gene
15 expression heterogeneity at the individual level. The single-worm transcriptomics (SWT)
16 provides a robust platform to perform a large-scale high-resolution transcriptomic analysis in the
17 nematode model organism of *P. pacificus*. Here, we provide a step-by-step protocol for i)
18 defining the chronological age of individual animals for establishing the high-resolution time
19 course, ii) modifying library preparation of SWT, iii) performing library sequencing strategy
20 regarding the sequencing cost and the number of genes detected. Together, this protocol
21 provides a detailed step-by-step procedure with many useful recommendations for both library
22 preparation and data analysis of single-worm transcriptomics in *P. pacificus*.

23

24 **Keywords:** *Pristionchus pacificus*, single-worm transcriptomics, RNA library preparation,
25 sequencing depth

26

27 **Background**

28

29 RNA-seq has seen a large variety of applications: from gene expression analysis by
30 quantitating the relative amounts of RNA sequence reads to the discovery of novel transcripts
31 or splice variants, ribosome profiling, or the detection of single nucleotide polymorphisms. The
32 low input RNA-seq technology, Smart-seq2 (Picelli et al. 2014; Trombetta et al. 2014) has
33 been introduced and improved for single nematode RNA-sequencing (Macchietto et al. 2017;
34 Serra et al. 2018; Chang et al. 2021). The methodology of single-worm transcriptomics has
35 recently been developed in the nematode model organism *Pristionchus pacificus* (Sun et al.
36 2021). Here, we developed a step-by-step protocol for single-worm transcriptomics in *P.*

37 *pacificus*. First, we provided a strategy of age synchronization for defining the chronological
38 age of individual nematodes by using two landmarks, hatching and ecdysis. Second, we
39 optimized the library preparation of SWT, including worm lysis through freeze-thawing,
40 genomic DNA removal and RNA purification, low input of cDNA (to ~1 ng) for the DNA library
41 preparation, and fewer PCR amplification cycles. Third, we provided a strategy for library
42 sequencing regarding the sequencing cost and the number of expressed genes. In the end,
43 we also supported the pipeline for analyzing of single-worm transcriptomics. In summary, this
44 step-by-step protocol includes detailed procedures with helpful recommendations, thus, allows
45 non-experienced investigators for performing high-resolution single-worm transcriptomics.

46

47 **Contents**

48

49	I Materials and reagents	4
50	II Equipment.....	4
51	III Software.....	4
52	IV Procedure.....	5
53	A. Worms Preparation.....	5
54	1. Synchronizing eggs-J1 by bleaching (eggs-J1 preparation).....	5
55	2. Distinguish substages and collection of worms	5
56	B. RNA Preparation	7
57	3. Worm lysis	7
58	4. Remove gDNA.....	7
59	5. Isolate total RNA and lysate cleanup	8
60	C. Library Preparation.....	8
61	6. cDNA synthesis & amplification	8
62	6.1. Primer Annealing for First Strand Synthesis	9
63	6.2. Reverse Transcription (RT) and Template Switching	9
64	6.3. cDNA amplification by PCR	9
65	6.4. Post-PCR bead (Ampure XP beads) clean-up.....	10
66	6.5. Estimate the library concentration and the fragment size	11
67	6.6. Normalization of amount of dsDNA before constructing DNA library.....	12
68	7. DNA library preparation	12
69	7.1. Tagment genomic DNA.....	12
70	7.2. Post tagmentation cleanup	13
71	7.3. Amplify tagmented DNA.....	14
72	7.4. Clean up libraries	15
73	D. Sequencing	16
74	8. Pooling libraries	16
75	E. Data analysis	17
76	F. Recipes.....	18
77	1. 2x SWT Lysis Buffer	18
78	2. Lysis buffer stock	18

79

80 I Materials and reagents

81

- 82 1. TURBO DNA-free™ Kit (Invitrogen, Catalog Number AM1907),
- 83 2. Superscript II reverse transcriptase kit (Thermo Fisher Scientific, Invitrogen™,
- 84 catalog number: 18064014)
- 85 3. AGENCOURT® RNACLEAN® XP (Protocol 001298v001)
- 86 4. NEBNext® Low Input cDNA Synthesis & Amplification Module (NEB #E6421S)
- 87 5. RNasin ribonuclease inhibitor (RNase inhibitor) (Promega, catalog number: N2611)
- 88 6. Nextera DNA Flex Library Prep (Illumina, 96 samples)
- 89 7. Proteinase K (QIAGEN, catalog number: 19131)
- 90 8. Betaine (BioUltra ≥ 99.0%) (Sigma-Aldrich, catalog number: 61962)
- 91 9. Magnesium chloride (MgCl₂; anhydrous) (Sigma-Aldrich, catalog number: M8266)
- 92 10. Agencourt Ampure XP beads (Beckman Coulter, catalog number: A63881)
- 93 11. Tris-HCl pH 8.0 (Thermo Fisher Scientific, Invitrogen™, catalog number: AM9850G)
- 94 12. Triton X-100 (Sigma-Aldrich, catalog number: T9284)
- 95 13. Polysorbate 20, Acros Organics™ (Tween 20) (Acros Organics, catalog number:
- 96 233362500)
- 97 14. EDTA pH 8.0 (Mediatech, catalog number: 46-034-CI)
- 98 15. 2x SWT Lysis Buffer (see Recipes)
- 99 16. 2x SWT Lysis Buffer (see Recipes)

100

101 II Equipment

102

- 103 1. 96-well plate magnet (Thermo Fisher, catalog number: 12331D)
- 104 2. Bioanalyzer (Agilent)
- 105 3. HiSeq 3000 (Illumina)
- 106 4. Qubit (Thermo Fisher, catalog number: Q32854)
- 107 5. Thermocycler
- 108 6. Thermomixer
- 109 7. Multichannel pipettes
- 110 8. Plate centrifuge

111

112 III Software

113

- 114 1. Cutadapt version 2.10 (Martin 2011)

- 115 2. HISAT version 2.1.0 (Pertea et al. 2016)
- 116 3. featureCounts (version 1.6.4) (Liao et al. 2014)
- 117 4. R (R Core Team 2020)
- 118 5. Integrated genomics viewer

119

120 **IV Procedure**

121

122 **A. Worms Preparation**

123

124 **1. Synchronizing eggs-J1 by bleaching (eggs-J1 preparation)**

125 1.1. Pick 20 young adults on new 60 mm NGM-agar plate with 200 μ l of OP50. After
126 culturing 8-9 days at 20 °C, obtain enough eggs.

127 1.2. The eggs were synchronized by washing off of plates with M9 using plastic Pasteur
128 pipettes into 15 ml conical tubes, and adding 30% final volume NaOH/bleach (0.5 ml
129 NaOH, 1 ml bleach/3.5 ml washed worms) for 9 minutes with gentle agitation every few
130 minutes.

131 1.3. Centrifuge 1,300 x g, 1 min, discard the suspension.

132 1.4. Add 3 ml M9 to the 15 ml conical tubes, place 1 min at room temperature.

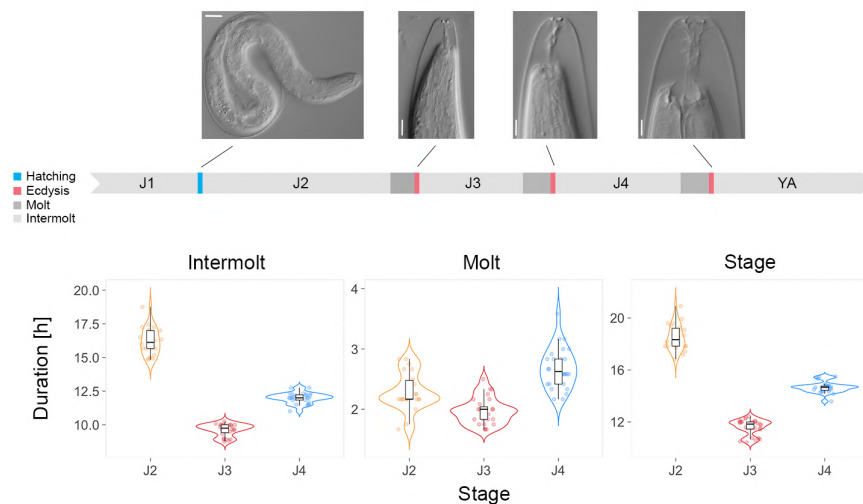
133 1.5. Centrifuge 1,300 x g, 1 min, discard the suspension.

134 1.6. Place synchronized eggs on new 60 mm NGM-agar plate with 200 μ l of OP50.

135

136 **2. Distinguish substages and collection of worms**

137 *Synchronization recommendations:* To perform high-resolution time course transcriptome
138 experiment, the hatching or ecdysis is recommended as the starting time point for the
139 sample collection, as both development events not only finished within short period time
140 (10 – 15 min), also can be easily obtained under the discovery microscopy (Fig. 1).



141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152

Figure 1. Landmarks was used to define the chronological age of individual animals in *P. pacificus*. Differential interference contrast images show the hatching, J2-ecdysis, J3-ecdysis, and J4 ecdysis. Violin plots on the bottom display the duration of each sub-stage during *P. pacificus* postembryonic development.

P. pacificus experienced three molting processes from hatching to adult, each of molt exit 2 – 3 hours (Fig. 1). Based on the precise duration of juvenile stages described above, the individual animals at the expected molting stage can be easily obtained after synchronization of bleach eggs (Table 1). Thus, the landmark, ecdysis, was single out for defining the nematode development age as hours post-hatching as stages.

Table 1. The molting animals was obtained many hours after egg bleaching.

Substages	No. of hours after bleaching
2 nd molt	16-18
3 rd molt	26-28
4 th molt	40-42

153
 154
 155
 156
 157
 158
 159
 160
 161
 162

- 2.1. After x hours (as above table mentioned), search worms during ecdysis under discovery microscopy
- 2.2. Carefully pick the single worm to new 60 mm NGM-agar plate with 50 µl of OP50, mark the time on the plate when you finish picking.
- 2.3. After few hours (depending on how old worm you want), transfer the worm to one new empty 60 mm NGM-agar plate, allow the worm crawl 1 min to remove the bacteria contamination.
- 2.4. Connect with step 3.1

163 **B. RNA Preparation**

164

165 **3. Worm lysis**

166 *Note:* Working under RNase Free Conditions

167 *Prepare:* 2X SWT Lysis Buffer (see Recipes)

168 *Expected time:* Within 40 min

169

170 3.1. Pipette 3 μ l dH₂O into 200 μ l PCR tube. Pick a single worm into each tube containing
171 water.

172 3.2. Freeze/thaw 3x with liquid nitrogen.

173 *Note:* Safe stopping point, samples can be safely stored overnight at -80°C.

174 3.3. Add 3 μ l 2x SWT Lysis Buffer to each tube, mix well by pipetting up and down 6 times
175 (avoid using vortex).

176 3.4. Incubate in PCR machine at 65°C, 30 min; 85°C, 1min; 4°C, hold on.

177

178 **4. Remove gDNA**

179 *Note:* Working under RNase Free Conditions

180 *Prepare:* TURBO DNA-free™ Kit (invitrogen, Catalog Number AM1907)

181 Superscript II reverse transcriptase kit

182 *Expected time:* Within 40 min

183

184 4.1. Prepare the DNase treatment mix (total volume 24 μ l) as follows:

185 TURBO DNase™ Enzyme 1 μ l

186 5x Superscript II first-strand buffer 2 μ l

187 Nuclease-free Water 21 μ l

188 4.2. Add the 24 μ l of DNase treatment mix (above) to 6 μ l of the lysis volume (Step 3.4),
189 mix well by pipetting up and down at least 6 times (avoid using vortex).

190 4.3. Incubate at 37°C for 20 min.

191 4.4. Resuspend the DNase Inactivation Reagent by flicking or vortexing the tube before
192 use.

193 4.5. Add 3 μ l resuspended DNase Inactivation Reagent, then mix well by pipetting up and
194 down 6 times (avoid using vortex).

195 4.6. Incubate the sample for 2 min at room temperature (22-26°C). Flick the tube 2-3 times
196 during the incubation period to redisperse the DNase Inactivation Reagent.

197 4.7. Centrifuge microcentrifuge tubes at 4°C, 12,000 x g for 1.5 min.

198 4.8. Carefully transfer 30 µl the supernatant containing the RNA to a fresh PCR tube. Do
199 not disturb the pellet of DNase Inactivation Reagent, because it can sequester divalent
200 cations and change the buffer conditions
201

202 **5. Isolate total RNA and lysate cleanup**

203 *Note:* Working under RNase Free Conditions

204 *Prepare:* AGENCOURT® RNACLEAN® XP (Protocol 001298v001)

205 *Expected time:* Within 20 min
206

207 5.1. Thoroughly vortex RNA-SPRI beads (Agencourt RNAClean XP SPRI beads) to ensure
208 a uniform suspension, incubate at room temperature for 30 min

209 5.2. Add 54 µl the RNA-SPRI beads into each 30 µl RNA sample (Step 4.8), then mix well
210 by pipetting up and down 10 times.

211 5.3. Incubate RNA and bead suspension at room temperature for 3 min.

212 5.4. Place tubes on 96-well plate magnet and incubate for 2 min. Remove supernatant from
213 each tube, being careful not to aspirate the beads.

214 5.5. Wash beads by adding 100 µl of 70% ethanol (prepared same day with nuclease-free
215 water) to each tube. Wait approximately 10 seconds and aspirate ethanol.

216 5.6. Repeat step 3.5

217 5.7. Leave the tube on the magnet and allow beads to dry at room temperature for
218 approximately 3 min.

219 5.8. Add 9 µl RNase-free water to each tube, then mix well by pipetting up and down 10
220 times.

221 5.9. Incubate samples at room temperature for 3 min.

222 5.10. Place tubes on 96-well plate magnet and incubate for 2 min.

223 5.11. Carefully transfer 7.5 µl the clean-up RNA to a fresh PCR tube.
224

225 **C. Library Preparation**

226

227 **6. cDNA synthesis & amplification**

228 *Note:* Working under RNase Free Conditions

229 *Prepare:* NEBNext® Low Input cDNA Synthesis & Amplification Module (NEB #E6421S)

230 RNasin ribonuclease inhibitor (RNase inhibitor)

231 Sample Recommendations: The RNA sample should be free of salts (e.g., Mg²⁺, or
232 guanidinium salts), divalent cation chelating agents (e.g. EDTA, EGTA, citrate).
233 Starting Material: 2 pg–200 ng poly(A) tail-containing total RNA (DNA free), RIN score ≥
234 8.0.

235

236 **6.1. Primer Annealing for First Strand Synthesis**

237 6.1.1. Prepare the primer mix (total volume 1.5 µl) in a separate tube as follows:

238 • (lilac) NEBNext Single Cell RT Primer Mix 1 µl
239 RNase Inhibitor 0.5 µl

240 6.1.2. Add the 1.5 µl of the primer mix (above) to 7.5 µl of treated RNA sample (Step
241 5.11), then mix gently by pipetting up and down at least 10 times, then centrifuge
242 briefly to collect solutions to the bottom of tubes.

243 6.1.3. Incubate for 5 minutes at 70°C in a thermocycler with the heated lid set to 105°C,
244 then hold at 4°C until next step.

245

246 **6.2. Reverse Transcription (RT) and Template Switching**

247 6.2.1. Vortex the NEBNext Single Cell RT Buffer briefly.

248 6.2.2. Prepare the RT mix (total volume 11 µl) in a separate tube as follows:

249 • (lilac) NEBNext Single Cell RT Buffer 5 µl
250 • (lilac) NEBNext Template Switching Oligo 1 µl
251 • (lilac) NEBNext Single Cell RT Enzyme Mix 2 µl
252 RNase Inhibitor 0.5 µl
253 Nuclease-free Water 2.5 µl

254 6.2.3. Mix thoroughly by pipetting up and down several times, then centrifuge briefly to
255 collect solutions to the bottom of tubes.

256 6.2.4. Combine 11 µl of the RT mix (above) with 9 µl of the annealed sample (Step 6.1.3),
257 mix well by pipetting up and down at least 10 times, and centrifuge briefly.

258 6.2.5. Incubate the reaction mix in a thermocycler with the following steps and the heated
259 lid set to 105°C,

260 90 minutes at 42°C,

261 10 minutes at 70°C,

262 Hold at 4°C.

263 Note: Safe stopping point, samples can be safely stored overnight at 4°C or –20°C.

264

265 **6.3. cDNA amplification by PCR**

266 6.3.1. Prepare cDNA amplification mix (total volume 80 µl) as follows:

267 • (orange) NEBNext Single Cell cDNA PCR Master Mix 50 µl
268 • (orange) NEBNext Single Cell cDNA PCR Primer 2 µl

269 (white) NEBNext Cell Lysis Buffer (10X) 0.5 µl
 270 Nuclease-free Water 2.5 µl

271 6.3.2. Add 80 µl cDNA amplification mix to 20 µl of the sample (Step 6.2.5), mix thoroughly
 272 by pipetting up and down at least 10 times.

273 6.3.3. Incubate the reaction in a thermocycler with the following PCR cycling conditions
 274 and the heated lid set to 105°C:

CYCLE STEP	TEMP	TIME	CYCLES
Initial Denaturation	98°C	45 sec	1
Denaturation 98°C	10 sec		
Annealing	62°C	15 sec	15
Extension	72°C	3 min	
Final Extension	72°C	5 min	1
Hold	4°C	∞	

282 6.3.4. *Recommended Number of PCR Cycles*

283 The amount of mRNA of animals at different stages is different, which means the
 284 amount of starting material (input RNA) is different. The NEB protocol recommends
 285 the number of PCR cycles based on different total RNA. However, when we do
 286 time-course experiments and we want to make sure that samples from different
 287 stages can be compared, therefore, the same number of the PCR cycles is
 288 recommended for each different substages to avoid the amplification bias
 289 generated from distinct number of PCR cycles. The amount of dsDNA can be
 290 produced under 15 PCR cycles for distinct substages as follows:

291
 292 Table 2. The amount of dsDNA was generated from individual worms
 293 at different developmental stages

Substages	PCR cycles	Amount of dsDNA, ng
J3.1-J3.3	15	7
J3.4-J3.6	15	7
J3.7-J3.9	15	14
J3.10-J3.12	15	23
J4.1-J4.3	15	24
J4.5-J4.7	15	26
J4.8-J4.11	15	25
J4.12-J4.14	15	26
YA.2	15	34

294
 295
 296 Safe stopping point: Samples can be safely stored overnight at 4°C or -20°C.

297
 298 **6.4. Post-PCR bead (Ampure XP beads) clean-up**

299 6.4.1. Take samples out of -20 °C and let them sit at room temperature for 10 min.

300 6.4.2. Allow the NEBNext Bead Reconstitution Buffer and the SPRI beads (if stored at
301 4°C) to warm to room temperature for at least 30 minutes before use. Vortex SPRI
302 beads to resuspend well and prepare fresh 80% ethanol (400 µl per sample).

303 6.4.3. Add 100 µl (1x of sample volume) resuspended beads to the PCR reaction (Step
304 6.3.3). Mix well by pipetting up and down at least 10 times. Be careful to expel all
305 of the liquid out of the tip during the last mix

306 *Size Selection Recommendations:* Double size selection is not recommended, 1x beads
307 could keep all size of dsDNA.

308 6.4.4. Incubate samples on the bench top for at least 5 minutes at room temperature.

309 6.4.5. After incubation, place sample on a magnetic bead stand for 5 min.

310 6.4.6. After 5 minutes (or when the solution is clear), carefully remove and discard the
311 supernatant. Be careful not to disturb the beads that contain cDNA.

312 *Note:* The beads will be bound to cDNA of interest.

313 6.4.6. Leave tubes on magnetic bead stand, add 200 µl of 80% ethanol to the beads 30
314 sec, and then carefully remove and discard the supernatant. Be careful not to
315 disturb the beads that contain cDNA.

316 6.4.7. Repeat Step 6.5.6 once for a total of two washes. Be sure to remove all visible
317 liquid after the second wash.

318 6.4.8. Air dry the beads for up to 5 minutes while the tube/plate is on the magnetic stand
319 with the lid open.

320 *Note:* Do not over-dry the beads. This may result in lower recovery of cDNA.

321 6.4.9. Remove the tube from the magnet and immediately add 17.5 µl of 0.1X TE (dilute
322 1X TE Buffer 1:10 in water) or EB buffer to the beads. Pipette up and down 10 times
323 to thoroughly mix. The solution should be homogenous and brown.

324 6.4.10. Incubate the sample for at least 2 min at room temperature.

325 6.4.11. Place the tube on the magnetic stand. After 5 min (or when the solution is clear),
326 transfer 16 µl to a new PCR tube.

327 *Safe stopping point: Samples can be safely stored overnight at 4°C or -20°C.*

328

329 **6.5. Estimate the library concentration and the fragment size**

330 6.5. Use 1 µl of the purified PCR product to measure the fragment size distribution using
331 the Agilent HS DNA BioAnalyzer (Fig. 1) and 1 µl to estimate the library concentration
332 using the Qubit® dsDNA HS Assay kit with Qubit® assay tubes.

333

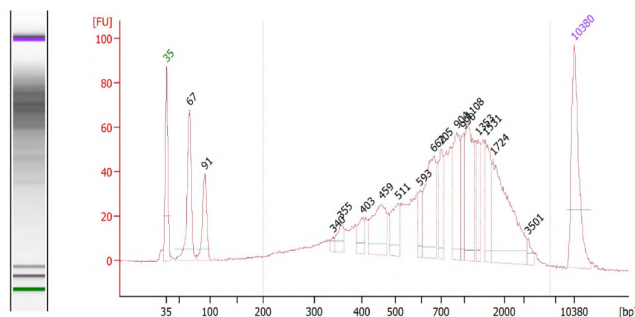


Figure 2. The output of Bioanalyzer displays cDNA size distribution.

334

335

336

337

6.6. Normalization of amount of dsDNA before constructing DNA library

338

6.6. Separate 70 ng purified PCR product (Step 6.4.11), and then add nuclease-free water to the DNA samples to bring the total volume to 30 μ l.

339

340

Normalization of the amount of dsDNA for preparing DNA library Recommendations:

341

It is crucial to normalize the amount of purified dsDNAs prior to Nextera DNA Flex library construction (next step). This ensures that dsDNA from each sample will yield tagmentation products with similar distributions of fragment length, and is also benefit to perform the next amplification under same number of PCR cycles.

342

343

344

345

346

7. DNA library preparation

347

Note: Working under RNase Free Conditions

348

Kit: Nextera DNA Flex Library Prep (Alternative selection)

349

350

7.1. Tagment genomic DNA

351

Prepare the following consumables:

352

Item (Storage), Instructions

353

BLT (2°C to 8°C), Bring to room temperature. Vortex to mix. Do not centrifuge before pipetting

354

TB1 (-25°C to -15°C), Bring to room temperature. Vortex to mix

355

356

7.1.1. Vortex BLT vigorously for 10 seconds, then visually check the beads for complete resuspend. Repeat as necessary.

357

358

7.1.2. Prepare tagmentation master mix. For each reaction use:

359

BLT 10 μ l

360

TB1 10 μ l

361

7.1.3. Vortex the tagmentation master mix thoroughly to make sure the BLT beads are evenly resuspended in the buffer.

362

363 7.1.4. Using fresh tips, transfer 20 µl of tagmentation master mix to each well containing
364 a sample.

365 7.1.5. Pipette mix the 50 µl reaction mix to resuspend.

366 7.1.6. Seal the plate.

367 7.1.7. Incubate the plate at 55°C for 15 min, followed by a 10°C hold. Use a thermal cycler
368 with a heated lid set to 100°C.

369

370 7.2. Post tagmentation cleanup

371 Prepare the following consumables:

372 Item (Storage), Instructions

373 TSB (15°C to 30°C), Check for any precipitates. If present, heat the buffer at 37°C, 10 min

374 TWB (15°C to 30°C), Use at room temperature

375

376 7.2.1. Add 10 µl TSB to the tagmentation reaction (Step 7.1.7).

377 7.2.2. Gently pipette mix the entire volume to resuspend the beads.

378 7.2.3. Seal the plate.

379 7.2.4. Incubate at 37°C for 15 min on a thermal cycler with heated lid set at 100°C, then
380 hold at 10°C. Volume of reaction is 60 µl.

381 7.2.5. Using a multichannel pipette, remove supernatant and discard.

382 7.2.6. Remove the plate from the magnet and add 100 µl TWB. Gently pipette mix until
383 beads are fully resuspended.

384 7.2.7. Place on the magnet for 3 min, or until solution is clear.

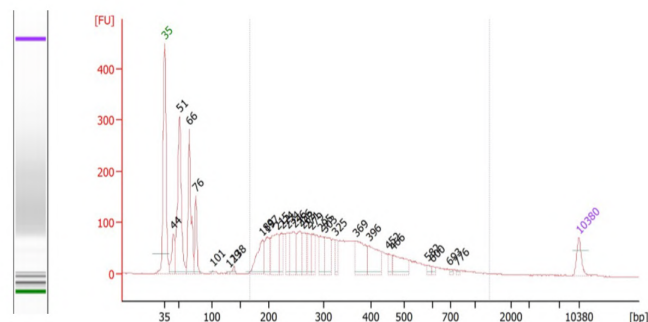
385 7.2.8. Remove supernatant with a multichannel pipette and discard.

386 7.2.9. Repeat steps 7.2.6 through 7.2.8 one more time for a total of two washes.

387 7.2.10. Remove the plate from the magnet and add 100 µl TWB. Gently pipette mix until
388 beads are fully resuspended.

389 7.2.11. Place the plate with TWB on the magnet and allow it to incubate until step 7.3.
390 The plate should incubate for at least 3 min, or until clear.

391



392

393 Figure 3. The output of Bioanalyzer displays the size distribution of tagmented DNA.

394

395 **7.3. Amplify tagmented DNA**

396 Prepare the following consumables:

397 Item (Storage), Instructions

398 EPM (-25°C to -15°C), Thaw on ice, invert to mix, then briefly centrifuge

399 Nextera DNA Flex Indexes (15°C to 30°C), Thaw at room temperature, invert to mix, then briefly
400 centrifuge

401

402 7.3.1. Prepare the PCR master mix.

403	Reagent	Volume per reaction
-----	---------	---------------------

404	EPM	20 µl
-----	-----	-------

405	Nuclease-free water	20 µl
-----	---------------------	-------

406 7.3.2. Vortex and spin down the PCR master mix.

407 7.3.3. Remove the third TWB wash from the samples while on the magnet. Use a P20
408 multichannel pipette to remove any excess liquid from the plate. Any remaining
409 foam on the well walls does not adversely affect the library.

410 7.3.4. Remove the plate from the magnet. Proceed immediately to the next step to prevent
411 excessive drying of the beads.

412 7.3.5. Add 40 µl of the PCR master mix to each sample well. Pipette mix to make sure
413 that the beads are thoroughly resuspended.

414 7.3.6. Add index adapters to each sample. Check volumes in the following table. For low-
415 plexity conditions, refer to the Index Adapters Pooling Guide (document #
416 1000000041071).

417	Index Kit Type	Kit Configuration	Volume of Index Adapter per Sample
-----	----------------	-------------------	------------------------------------

418	96 plex (dual index)	96-well plate	10 µl of primer mix
-----	----------------------	---------------	---------------------

419 7.3.7. Using a pipette set to 40 µl, pipette mix a minimum of 10 times to mix the entire
420 reaction volume.

421 7.3.8. Seal the plate, place in the thermal cycler and run the program.

422	TEMP	TIME	CYCLES
-----	------	------	--------

423	68°C	3 min	1
-----	------	-------	---

424	98°C	3 min	1
-----	------	-------	---

425	98°C	45 sec	
-----	------	--------	--

426	62°C	30 sec	5
-----	------	--------	---

427	68°C	2 min	
-----	------	-------	--

428	68°C	1 min	1
-----	------	-------	---

429	4°C	∞	
-----	-----	---	--

430 7.3.9. Remove the plate from the thermal cycler when the PCR program completes.

431 7.3.10. Centrifuge for 1 minute at 280 × g to make sure all the liquid is at the bottom of
432 the well.

433 Safe stopping point: Samples can be safely stored overnight at 4°C or –20°C.

434

435 **7.4. Clean up libraries**

436 Prepare the following consumables:

437	Item	Storage	Instructions
438	SPB	2°C to 8°C	Let stand at room temperature for 30 minutes. Vortex and invert to mix.
439	RSB	-25°C to -15°C	Thaw at room temperature, invert to mix.

440 7.4.1. Place the plate (Step 7.3.10) on the magnet for 5 minutes or until the supernatant
441 is clear.

442 7.4.2. Transfer 45 µl of the PCR supernatant into a fresh midi plate.

443 7.4.3. Vortex and invert SPB multiple times to ensure full resuspension.

444 7.4.4. Prepare a master mix of diluted SPB:

445	Reagent	Volume per reaction
446	SPB	56 µl (0.65x)
447	Nuclease-free water	40 µl

448 7.4.5. Vortex the diluted SPB master mix thoroughly and add 96 µl mix to each PCR
449 product.

450 7.4.6. Pipette mix a minimum of 10 times or until thoroughly mixed.

451 7.4.7. Incubate at room temperature for 5 minutes.

452 7.4.8. Place the midi plate on a plate magnet for 5 minutes or until supernatant is clear.

453 7.4.9. During incubation, vortex the SPB (undiluted stock tube) thoroughly, and then add
454 18.3 µl (0.9x) to each well in a new midi plate.

455 7.4.10. Transfer 125 µl of supernatant from the first midi plate into the second midi plate
456 (containing the 18.3 µl of PB).

457 Size Selection Recommendations: Double size selection is recommended, 0.65x SPB for
458 the right side selection, and 0.9x SPB for the left side selection. Using the
459 recommended selection condition, mean length of fragment of 360 – 380 bp could
460 be selected (Fig. xx). Bigger fragment size (>450 bp) is not recommended for the
461 transcriptomic sequencing, given the transcripts with short length will be missed.

462 7.4.11. Pipette mix 10 times.

463 7.4.12. Seal the second midi plate and incubate at room temperature for 5 minutes.

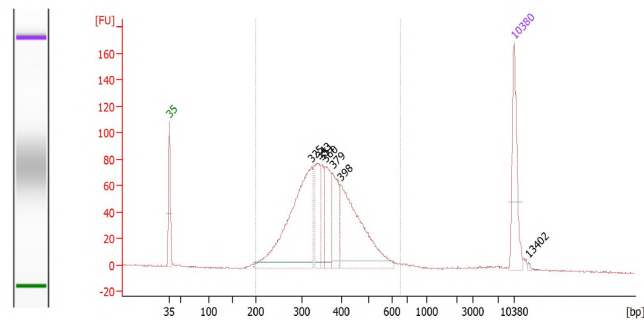
464 7.4.13. Place midi plate on a magnet for 5 minutes or until clear.

465 7.4.14. Remove and discard supernatant without disrupting the beads.

466 7.4.15. With the plate on the magnet, add 200 µl of fresh 80% ethanol without mixing and
467 incubate for 30 seconds.

- 468 7.4.16. Pipette to remove the ethanol.
- 469 7.4.17. Repeat steps 15 and 16 for a total of 2 washes.
- 470 7.4.18. Use a P20 pipette to remove any excess liquid from the midi plate.
- 471 7.4.19. Air-dry on the magnetic stand until dry (~5 minutes).
- 472 7.4.20. Remove the midi plate from the magnet and add 32 μ l of RSB to the beads.
- 473 7.4.21. Pipette mix until thoroughly resuspended.
- 474 7.4.22. Incubate at room temperature on the bench for 2 minutes.
- 475 7.4.23. Place the midi plate back on the magnet for 2 minutes or until clear.
- 476 7.4.24. Transfer 30 μ l of the supernatant into a new 96-well PCR plate.
- 477 Safe stopping point: Samples can be safely stored overnight at 4°C or -20°C.

478



479

480 Figure 4. The output of Bioanalyzer displays the fragment size distribution after selection.

481

482 D. Sequencing

483

484 8. Pooling libraries and sequencing depth

485 The number of genes can be detected from transcriptomic samples depends on i) the

486 development stages or organism tissues of RNA samples, as there are massive differentially

487 expressed genes at temporal and spatial levels; ii) sequencing depth of RNA libraries, more

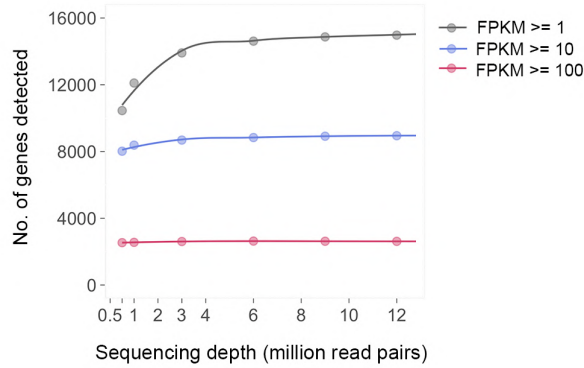
488 expressed genes were detected as the sequencing depth increase from 0 to 4 million read

489 pairs (Fig. 5), which means the gene with low expression can't be detected with low

490 sequencing depth whereby influence the following RNA-seq analysis. 4 million read pairs per

491 RNA library might be safe for RNA-seq analysis.

492



493

494

495

Figure 5. The line plots show the number of detected genes as a function of sequencing depth for SWT library with more than 12 million reads. Subsampling of reads was done for nine levels of sequencing depth and the number of detected genes was assessed for three different expression levels (FPKM >= 1 (left), FPKM >= 10 (middle), FPKM >= 100 (right)).

499

500

When sequencing was processed on the illumine Hiseq3000 platform, which generates a pair short read (150 bp) per fragment. Thus, the 2.5 raw data per sample should be required.

502

503

Table 3. Sequencing depth and no. of genes detected

No. of samples pooling in one lane	No. of read pairs per sample	No. of raw data per sample	Median No. of genes detected (count >= 1)
40	4 M	2.5 G	19,272
55	3 M	1.8 G	20,931

504

505

506 E. Data analysis

507

508

1. Raw reads were trimmed by the program cutadapt (version 2.10) with parameters “-q 30,25 --gc-content =50, --minimum-length 25:25”.

509

510

2. Read pairs were aligned by the HISAT2 (version 2.1.0) with the additional parameter “--rna-strandness RF”.

511

512

3. Mapped reads were summarized to the genomic features by the featureCounts (version 1.6.4) with parameters “-a -F GTF -g Parent -d 50 -D 500”.

513

514

4. Gene expression value, transcripts per million (TPM), was computed through normalization by library size and feature effective length. The mean TPM was calculated from three biological replicates.

515

516

517

518 **F. Recipes**

519

520

1. 2x SWT Lysis Buffer

521

Note: Lysis buffer stock can be stored indefinitely wrapped in foil at room temperature.

2x SWT Lysis Buffer	55 ul
Lysis Buffer Stock	46.8 ul
Proteinase K	3.2 ul
RNase inhibitor	5 ul

522

523

2. Lysis buffer stock

524

Note: Aliquot Triton-X (100%) to a 1.5 ml Eppendorf tube and warm to 35 °C and use

525

as described in Recipe.

Lysis Buffer Stock	1.871 ml
Tris-HCl pH 8.0 (1 M)	20 µl
Triton-X (100%)	20 µl
Tween-20 (10%)	200 µl
EDTA (0.5 M)	2 µl
Nuclease free water	1.628 ml

526

527

528

Acknowledgments

529

We thank the members of the Sommer laboratory for thoughtful comments of experiments,

530

results, and interpretation. S.S. was funded by a fellowship from the Chinese Sponsorship

531

Council (CSC). This work was funded by the Max-Planck Society. Author contributions: S.S.

532

and R.J.S. conceived and designed all experiments. S.S. conducted all experimental studies.

533

S.S. and R.J.S contributed to writing.

534

535

Competing interests

536

The authors declare no competing interests.

537

538

References

539

Chang D, Serra L, Lu D, Mortazavi A, Dillman A. 2021. A revised adaptation of the smart-

540

seq2 protocol for single-nematode RNA-seq. In *Methods in Molecular Biology*, Vol.

541

2170 of, pp. 79–99, Humana Press Inc.

542

Liao Y, Smyth GK, Shi W. 2014. Sequence analysis featureCounts: an efficient general

543

purpose program for assigning sequence reads to genomic features. **30**: 923–930.

544

Macchietto M, Angdembe D, Heidarpour N, Serra L, Rodriguez B, El-Ali N, Mortazavi A.

545

2017. Comparative transcriptomics of *Steinernema* and *Caenorhabditis* single embryos

546 reveals orthologous gene expression convergence during late embryogenesis. *Genome*
547 *Biol Evol* **9**: 2681–2696.

548 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
549 reads. *EMBnet journal* **17**: 10–12.

550 Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression
551 analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature*
552 *Protocols* **11**: 1650–1667.

553 Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. 2014. Full-length
554 RNA-seq from single cells using Smart-seq2. *Nature Protocols* **9**: 171–181.

555 R Core Team. 2020. R: A Language and Environment for Statistical Computing.

556 Serra L, Chang D, Macchietto M, Williams K, Murad R, Lu D, Dillman A, Mortazavi A. 2018.
557 Adapting the Smart-seq2 Protocol for Robust Single Worm RNA-seq. *Bio-Protocol* **8**: 1–
558 16.

559 Sun S, Rödelsperger C, Sommer RJ. 2021. Single worm transcriptomics identifies a
560 developmental core network of oscillating genes with deep conservation across
561 nematodes. *Genome Research* **31**: 1590–1601.

562 Trombetta JJ, Gennert D, Lu D, Satija R, Shalek AK, Regev A. 2014. Preparation of single-
563 cell RNA-Seq libraries for next generation sequencing. *Current Protocols in Molecular*
564 *Biology* **2014**: 4.22.1-4.22.17.

565