

**Functional characterisation of  
the new nematode model  
*Allodiplogaster sudhausi***

**Dissertation**

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

vorgelegt von  
Sara Wighard  
aus Kapstadt, Südafrika

Tübingen  
2023

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

Tag der mündlichen Qualifikation:	07.11.2023
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Ralf J. Sommer
2. Berichterstatter/-in:	Prof. Dr. Ulrike Zentgraf



# Table of Contents

<b>I. ACKNOWLEDGEMENTS</b>	<b>5</b>
<b>II. SUMMARY</b>	<b>6</b>
<b>III. DEUTSCHE ZUSAMMENFASSUNG</b>	<b>7</b>
<b>IV. LIST OF PUBLICATIONS</b>	<b>8</b>
<b>1. INTRODUCTION</b>	<b>9</b>
1.1. How do new traits arise?	9
1.2. Whole genome duplication	9
1.2.1 What is whole genome duplication?	9
1.2.2 Previous whole genome duplication events	10
1.2.3 Whole genome duplication as a driver of evolution	10
1.2.4 Whole genome duplication can increase body size	12
1.3. Phenotypic plasticity	12
1.3.1 What is phenotypic plasticity?	12
1.3.2 Plasticity in cannibalistic behaviour	13
1.3.3 Phenotypic plasticity as a driver of novelty	14
1.4. Nematodes as models for evolutionary biology	15
1.4.1 Why use nematodes?	15
1.4.2 The mouth-form regulatory network in <i>Pristionchus pacificus</i>	16
1.4.3 <i>Allodiplogaster sudhausi</i> , an early-branching diplogastrid	17
<b>2. RESEARCH OBJECTIVES</b>	<b>19</b>
<b>3. RESULTS</b>	<b>20</b>
3.1 Crowdsourcing and the feasibility of manual gene annotation: A pilot study in the nematode <i>Pristionchus pacificus</i>	20
3.1.1 Synopsis	20
3.1.2 Own contribution	20

<b>3.2 Geometric morphometrics of microscopic animals as exemplified by model nematodes</b>	<b>21</b>
3.2.1 Synopsis	21
3.2.2 Own contribution	21
<b>3.3 A New Hope: A Hermaphroditic Nematode Enables Analysis of a Recent Whole Genome Duplication</b>	
<b>Event</b>	<b>22</b>
3.3.1 Synopsis	22
3.3.2 Own contribution	22
<b>3.4. Conserved switch genes regulate a novel cannibalistic morph after whole genome duplication</b>	<b>23</b>
3.4.1 Synopsis	23
3.4.2 Own contribution	23
<b>3.5 Multiple evidence for environmental sex determination in the hermaphroditic nematode</b>	
<b><i>Allodiplogaster sudhausi</i></b>	<b>24</b>
3.5.1 Synopsis	24
3.5.2 Own contribution	24
<b>4. DISCUSSION</b>	<b>25</b>
<b>5. REFERENCES</b>	<b>29</b>
<b>6. APPENDIX</b>	<b>38</b>

# I. Acknowledgements

Over the last 5 (and a bit) years there are many people who helped me. First of all, I'd like to acknowledge all the professional guidance I received. For general help at the facility, I want to acknowledge the (former) in-house genome sequencing facility and the SEM facility for the wonderful pictures. Special acknowledgements go to our technicians in lab, who are always there to help. My biggest thanks goes to Hanh for doing the CRISPR injections. If you weren't around, this project would never have even taken off! Also, thanks for the company in lab 17. Thank you to Metta for karyotyping, Walli for sequencing and Heike for freezing. Previous lab members also provided guidance, especially Bogdan for mentoring me my first few years. That was so immensely helpful – and for your friendship during that time. Other members that gave helpful input are James, Mo, Mike and Ata. Thanks goes to everyone at the Sommer Lab, who I always learnt from in some way. There are too many people to mention, but for the good times I want to thank Marina (especially catching up after work), Peng (I'll miss having you on my right hand side), Tobi (I'll always remember the Turtle™), Suryesh, Devansh, Yinan, Radhika, Shuai, Zeeshan, Veysi and too many more to mention. I'd like to thank our group leaders, Christian Rödelsperger, Adrian Streit and Catia Igreja for always being available to talk and giving helpful advice. I extend thanks toward my TAC members, Felicity Jones and Estienne Swart for their feedback over the years.

My biggest thanks goes to my supervisor, Ralf Sommer for guiding me during this time, supporting me and helping me grow as a scientist. Also, for taking a chance on me after I sent in a cold letter. It really changed the course of my life.

During my time in Tübingen I was able to meet many amazing people and make great memories. A PhD can be a tough time, but overall I really enjoyed my time here and a big part of that is due to the people I spent my time around. You all made it better and even the not-so-great experiences helped me grow so much. So I'd like to mention the friends I made outside of lab. Special thanks to Bridgit, I feel very lucky to have you as a friend, and to all the memories we made. Thanks to these great people: Laura, Albane, Andrea, Leo, Marco, Dingwen, Minakshi, George, Christian, Meghna, Julia, Linda, and everyone who joined in soccer, bouldering, Christmas markets and food outings.

Finally, thank you to my family who supported me even after I moved so far away and continue doing so. I'll come back one day.

## II. Summary

Phenotypic or developmental plasticity is the process whereby identical genotypes can produce different phenotypes in response to environmental cues. Phenotypic plasticity can also act as a driver of evolutionary events. However, this theory only recently started to gain popularity and thus further studies are needed to better understand the process of plasticity and the genetic mechanisms regulating it. Ideal organisms to study the impact of plasticity on evolutionary processes are nematodes that belong to the diplogastrid family, which ancestrally display phenotypic plasticity in adult mouth phenotype. The majority of diplogastrid species form one-of-two mouth-form morphs as adults in response to environmental cues. The hermaphroditic diplogastrid *Pristionchus pacificus* has been well-studied, and various genes have been identified that act as master regulators of the morphs. Knock-outs of these 'switch genes' lead to constitutive expression of only one morph. Although *P. pacificus* has been well-studied, the question arose as to whether the regulatory network of mouth-form plasticity was conserved within diplogastrids as a whole. To investigate this, I examined another hermaphroditic diplogastrid that branched early in the lineage, called *Allodiplogaster sudhausi*. In the process of characterising this organism, I made a number of surprising findings, including 1) a recent whole genome duplication event, 2) the evolution of environmental sex determination and, notably, 3) evolution of a third mouth-form morph. This morph is induced in poor nutrition and on fungus and displays cannibalistic behaviour. To examine regulation of this morph, I identified two sulfatase-encoding genes in *A. sudhausi*, that arose via whole genome duplication, and are homologous to a key mouth-form switch gene in *P. pacificus*. Using CRISPR, I then showed that double mutant knock-outs in both genes prevent two of the three morphs from being formed, including the novel morph. This strongly suggests that the mouth-form genes are functionally conserved within the diplogastrid family. Further, single mutants of these genes display diverged roles in regulation of the different morphs, suggesting dosage effects in regulation of one morph and redundancy in the other. Overall, I made a number of surprising findings in *A. sudhausi* and identified unexpected features that make it a potentially useful model for evaluating the evolution of novel traits.

### III. Deutsche Zusammenfassung

Phänotypische oder entwicklungsbedingte Plastizität ist der Prozess, durch den identische Genotypen unterschiedliche Phänotypen in Reaktion auf Umweltreize hervorrufen können. Phänotypische Plastizität kann auch als Treiber von evolutionären Ereignissen wirken. Diese Theorie hat jedoch erst kürzlich an Popularität gewonnen, und daher sind weitere Studien notwendig, um den Prozess der Plastizität und die genetischen Mechanismen, die sie regulieren, besser zu verstehen. Ideale Organismen, um den Einfluss der Plastizität auf evolutionäre Prozesse zu untersuchen, sind Nematoden, die zur Familie der Diplogastriden gehören und die phänotypische Plastizität im adulten Mundphänotyp aufweisen. Die Mehrheit der Arten der Diplogastriden bildet als Erwachsene in Reaktion auf Umweltreize eine von zwei Mundformen. Der hermaphroditische Diplogastride *Pristionchus pacificus* wurde umfassend erforscht, und verschiedene Gene wurden identifiziert, die als Hauptregulatoren der Formen wirken. Knock-outs dieser "Schaltergene" führen zu einer konstitutiven Expression nur einer Form. Obwohl *P. pacificus* intensiv erforscht wurde, stellte sich die Frage, ob das regulatorische Netzwerk der Mundformplastizität bei den Diplogastriden insgesamt konserviert ist. Um diese Frage zu untersuchen, untersuchte ich einen anderen hermaphroditischen Diplogastriden, der früh im Stammbaum verzweigte und *Allodiplogaster sudhausi* genannt wird. Im Zuge der Charakterisierung dieses Organismus machte ich eine Reihe überraschender Entdeckungen, darunter 1) ein kürzlich stattgefundenes Ereignis der gesamten Genomduplikation, 2) die Evolution der umweltbedingten Geschlechtsbestimmung und insbesondere 3) die Evolution einer dritten Mundform. Diese Form wird bei schlechter Ernährung und auf Pilzen induziert und zeigt kannibalistisches Verhalten. Um die Regulation dieser Form zu untersuchen, identifizierte ich zwei Gene, die Sulfatasen kodieren, in *A. sudhausi*, die durch eine gesamte Genomduplikation entstanden sind und homolog zu einem Schlüsselgen für die Mundformumschaltung in *P. pacificus* sind. Unter Verwendung von CRISPR zeigte ich dann, dass doppelte knock-out-Mutanten in beiden Genen verhindern, dass zwei der drei Formen gebildet werden, einschließlich der neuen Form. Dies legt nahe, dass die Mundformgene innerhalb der Familie der Diplogastriden funktional konserviert sind. Außerdem weisen Einzelmutanten dieser Gene unterschiedliche Rollen in der Regulation der verschiedenen Formen auf, was auf Dosiseffekte bei der Regulation einer Form und Redundanz bei der anderen hinweist. Insgesamt machte ich eine Reihe überraschender Entdeckungen bei *A. sudhausi* und identifizierte unerwartete Merkmale, die es zu einem potenziell nützlichen Modell für die Bewertung der Evolution neuer Merkmale machen.

## IV. List of Publications

1. 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.
2. Theska T, Sieriebriennikov B, **Wighard S**, Werner MS, Sommer RJ. 2020. Geometric morphometrics of microscopic animals as exemplified by model nematodes. *Nature Protocols* 15:2611–2644.
3. **Wighard S**, Athanasouli M, Witte H, Rödelsperger C, Sommer RJ. 2022. A New Hope: A Hermaphroditic Nematode Enables Analysis of a Recent Whole Genome Duplication Event. Wolfe K, editor. *Genome Biology and Evolution* 14:evac169.
4. **Wighard S**, Witte H, Sommer RJ. 2023. Conserved switch genes regulate a novel cannibalistic morph after whole genome duplication. *bioRxiv:2023.08.22.554244*.
5. **Wighard S**, Rödelsperger C, Sommer RJ. Multiple evidence for environmental sex determination in the hermaphroditic nematode *Allodiplogaster sudhausi*. *In preparation*.

# 1. Introduction

*“ Nothing can be made from nothing “*

Lucretius - *On the Nature of Things* (circa 50 BC)

## 1.1. How do new traits arise?

A key tenet of evolutionary biology is the formation of new traits. These can then be selected for and spread throughout a population, which can eventually give rise to new species. Both Russell Wallace (1858) and Darwin (1859) surmised evolution by natural selection which fundamentally revolutionised biology. However, more than a century later, we still have gaps in our knowledge with regards to the mechanisms of evolutionary events, particularly the formation of new traits. We know mutations can occur that lead to changes in amino acid sequences, which in turn can produce different proteins and alter the phenotype. Thus, novelty arises from pre-existing material that has been modified, and indeed co-option of existing material is considered the most likely origin of novel traits (Wagner 2011). However, at the population level most mutations that occur are unlikely to become fixed (Barrick et al. 2009). The acquisition and retention of new traits is therefore rare compared to the rate of mutations that take place. Nevertheless, there are certain events that are more likely to lead to retention of traits and hence modifications at the population level. These so-called ‘drivers of evolution’ promote novelty and speciation and can lead to large-scale changes. Examples of evolutionary drivers at the molecular level include the likes of duplication events (local and genome-wide), phenotypic (developmental) plasticity and transposable elements (TEs). Two possible drivers of evolution and their subsequent effects will be the main focus of this dissertation: 1) Whole genome duplication and 2) Phenotypic plasticity.

## 1.2. Whole genome duplication

### 1.2.1 What is whole genome duplication?

Whole genome duplication (WGD), also known as polyploidisation, is the process whereby the entire genome, including the chromosomes sets and regulatory elements, is doubled. WGD can come about by one of two ways: 1) Autopolyploidisation, the doubling of the

entire genome including the chromosomes and regulatory elements, usually due to an error in mitosis, and 2) Allopolyploidisation, hybridisation between two closely related species, i.e. half the chromosomes originate from one species and the other half from another species (Parisod et al. 2010). Although WGD was already hypothesised in the 1970s as a driver of novelty and evolution (Ohno 1970), its impact is still not well understood, particularly in animals.

### **1.2.2 Previous whole genome duplication events**

WGD is believed to have occurred in most eukaryotic lineages (Wolfe 2015). The majority of documented cases occurred ancestrally, with the exact timing often poorly understood. WGD events are abundant in plants (Clark and Donoghue 2018), but comparatively rarer in animals. This can partly be put down to reproductive barriers to WGD, such as organisms with different sex chromosomes (Ohno 1970). Nevertheless, a number of ancestral WGD events have been reported in animals. For instance, at least two rounds of WGD took place in the vertebrate ancestral lineage (Dehal and Boore 2005), while additional WGD events are reported in teleost fish (Amores et al. 1998), *Xenopus laevis* (Session et al. 2016) and plant-parasitic nematodes (Blanc-Mathieu et al. 2017). Aside from multicellular organisms, WGD has been well-documented in unicellular eukaryotes, including protozoans and yeast (Aury et al. 2006; Marcet-Houben and Gabaldón 2015). These previous WGD events have brought to light interesting associations, such as TE expansion (Marburger et al. 2018), and body size increase (see section 2.4) following WGD.

### **1.2.3 Whole genome duplication as a driver of evolution**

In general, duplication events are considered one of the major sources of novelty at the molecular level. Duplication generates redundant genetic material that can potentially gain mutations, obtain a new function, and be positively selected for in a population (Ohno 1970). It is thus a prominent driver of evolution. Two principal forms of duplications of genetic material have to be distinguished. Duplication of a single gene affects the dosage relationship between genes, leading to dosage imbalance (Papp et al. 2003). Thus, the majority of small-scale gene duplications are silenced and do not end up having any effect at the population level. In contrast, in WGD all genes and regulatory elements are duplicated, meaning there is no immediate change in the dosage relationship between



genes, i.e. the balance is not affected and gene silencing is thus not common immediately after WGD (Ohno 1970; Papp et al. 2003); in fact, selection initially acts to preserve duplicated genes and maintain the dosage balance (Gout and Lynch 2015). Therefore, WGD is considered a greater evolutionary driver than local gene duplications as the genes are not as readily lost due to the constraints of dosage balance (Birchler et al. 2005; Blomme et al. 2006).

After WGD, duplicate genes are initially redundant with the same expression and role as their partner; however, over time they diverge and can have three potential fates. These are: 1) Nonfunctionalisation, where one duplicate loses its function (also known as a pseudogene) and can eventually become lost; 2) Subfunctionalisation, where each duplicate gene adopts a partial function of the parental gene, and 3) Neofunctionalisation, where one duplicate gains a novel function (Zhang 2003). The duplicates often obtain novelty by co-opting features from parental genes (Conant and Wolfe 2008). Neofunctionalization and the gain of a novel function is the most common way WGD events are inferred to drive evolution and novelty. However, it can be difficult to identify neofunctionalised genes as they can diverge so much that their origin cannot be correctly determined. Regardless, neofunctionalisation has been identified in fishes (Zakon et al. 2006; Moriyama et al. 2016), *Drosophila* (Assis and Bachtrog 2013) and *X. laevis* (Hayashi et al. 2022). Certain neofunctionalised genes have gained important evolutionary roles. For instance, neofunctionalisation plays a crucial roles in male fertility and sex ratio in mice (Kruger et al. 2019), male specification in medaka (Matsuda et al. 2002), adaptive evolution of snake venom metalloproteinases (Casewell et al. 2011), female sexual differentiation in *X. laevis* (Hayashi et al. 2022) and evolution of C4 photosynthesis in plants (Monson 2003). Therefore, the changes caused by WGD can drive evolutionary events. Despite these notable cases, it is still difficult to evaluate the overall impact of WGD as it is relatively rare (in animals) and, importantly, most recorded events are ancient (Wolfe 2001). The presence of a recent animal WGD, as will be described in this dissertation (Wighard et al. 2022), allows us to characterize and evaluate the impact of such events.

### **1.2.4 Whole genome duplication can increase body size**

Aside from generating duplicate material and allowing novelty to evolve, WGD can also lead to a pronounced increase in body size. The effect of WGD in driving size has been recorded in plants, with polyploidisation events often increasing plant size (Otto and Whitton 2000). In animals, the effect of WGD on size usually depends on the taxa affected. It is accepted that WGD increases size in invertebrates but not in vertebrates (Gregory et al. 2000; Otto and Whitton 2000). WGD directly influences animal adult size when growth is determined by cell-cell interactions or when there is a fixed number of cells (Otto 2007). The positive relationship between WGD and body size has been seen in rotifers (Walsh and Zhang 1992), water fleas (Weider 1987), flatworms and copepods (Gregory et al. 2000) and, on multiple occasions, in nematodes (Madl and Herman 1979; Triantaphyllou and Riggs 1979; Wighard et al. 2022). Additionally, it has experimentally been shown that endoreduplication drives increased body size in *Caenorhabditis* (Flemming et al. 2000). Thus, due to its effects on gene novelty and body size, WGD can drive novelty at both the organismal and genomic level.

## **1.3. Phenotypic plasticity**

### **1.3.1 What is phenotypic plasticity?**

Phenotypic or developmental plasticity is a natural phenomenon whereby an organism changes its phenotype according to environmental signals. It is defined by West-Eberhard (2003) as the “ability of an organism to react to an internal or external environmental input with a change in form, state, movement or rate of activity”. Discrete and discontinuous phenotypic plasticity that results in alternative phenotypes is referred to as a polyphenism (Nijhout 2003). In polyphenic individuals, the environmental stimulus acts as a switch (Bradshaw 1965). Phenotypic plasticity is most useful in fluctuating environments as this enables it to respond to different conditions. For this reason, it is commonly observed in plants (Griffith and Sultan 2005; Joly and Schoen 2021) as they are static and have to endure their fluctuating environment, in contrast to animals that can usually move to different environments (Bradshaw 1965). Despite this, polyphenism has been observed in a number of animal organisms, including beetles (*Onthophagus taurus*) (Moczek 1998), pea aphids (*Acyrtosiphon pisum*) (Parker and Brisson 2019), water fleas (Lüning 1992), locusts (Pener 1991) and nematodes (Von Lieven and Sudhaus 2000).

### **1.3.2 Plasticity in cannibalistic behaviour**

Cannibalism, the killing and feeding of conspecific kin, has been observed in different kingdoms, including bacteria (Claverys and Håvarstein 2007) and ciliates (Giese 1938); however, it is in animals that most studies have been performed. Cannibalism is occasionally observed in most animal species, but is usually not the predominant behaviour. This is logical as the killing of kin has obvious negative fitness costs and in its broad sense is the antithesis of Darwinian evolution - potentially reducing the chances of passing genes onto the next generation. However, there are circumstances where cannibalism can be of benefit. For instance, in environments with limited resources, many species display cannibalistic behaviour as a means to survive (Elgar and Crespi 1992). If there is no food source other than fellow kin, cannibalism is a means of increasing the survivability of a few to better enable them to reproduce and pass their genes onto the next generation. In insects, cannibalism is density-dependent. It reduces population density, minimizing competition for limited resources and improving the fitness of the cannibal (Richardson et al. 2010). Cannibalism is commonly observed in locusts and crickets (Simpson et al. 2006; Bazazi et al. 2008) and is even hypothesized to promote evolution of polyphenism in locusts (Guttal et al. 2012). Sexual cannibalism also occurs in arthropods, where females eat their mates after mating in the case of praying mantids and certain arachnids (Elgar and Schneider 2004). The feeding provides the female with additional nutrients and increases the likelihood of successful reproduction. There are thus multiple ways cannibalism takes place, which is in agreement with the theory of natural selection.

An additional form of cannibalism is the evolution of cannibalistic morphs, a form of phenotypic plasticity and a polyphenism. Cannibalism as a polyphenic trait circumvents negative fitness consequences that would be seen in an organism that continuously cannibalises. Cannibalistic morphs are formed in response to environmental cues in species such as spadefoot toads (Pfennig et al. 1993), cane toads (Crossland et al. 2011), Hokkaido salamanders (Michimae and Wakahara 2001) and tiger salamanders (Lannoo and Bachmann 1984). These morphs are usually induced in starved conditions, where there are few resources (Michimae & Wakahara 2001), as well as in crowded conditions, where competition for resources is high (Collins and Cheek 1983; Hoffman and Pfennig 1999). Environmental conditions have also been shown to affect the frequency of cannibalistic

behaviour, such as invasive cane toads in Australia that show higher cannibalistic behaviour than indigenous cane toads from South America (DeVore et al. 2021). How cannibalistic morphs evolved mechanistically is still largely unknown. However, as a diverse array of species are induced by the same triggers of crowding and starvation, this suggests there may be conserved ancestral networks involved, or even convergent evolution taking place. The regulatory networks and mechanisms that lead to evolution of cannibalistic morphs therefore require further investigation.

### **1.3.3 Phenotypic plasticity as a driver of novelty**

Phenotypic plasticity is believed to be a driver of evolution and novelty in what is also termed the 'plasticity-first' hypothesis (West-Eberhard 2003; Moczek et al. 2011). In contrast to traditional Darwinian thought, the influence of environment on evolution is considered a central driving force of evolutionary events (West-Eberhard 2003). The mechanism of inheritance of plastic traits has been hypothesised to be due to accumulation of cryptic genetic variation. For instance, genetic variation underlying populations and/or individuals can accumulate over time in conditions where the phenotype is masked. A new environment can then expose the phenotype encoded by such cryptic variation. If such a phenotype is beneficial it will be selected for, which can lead to adaptive evolution (Pfennig et al. 2010; Levis et al. 2018). Novel phenotypes can therefore arise as the result of expression of accumulated variation in new conditions (Moczek et al. 2011).

Polyphenism in particular is said to promote rapid speciation. For instance, if alternative morphs are favoured in certain niches, they will accumulate differences from morphs in other niches. Mating will therefore produce offspring with low fitness, which favours evolution of reproductive isolation between divergent individuals and thus promotes speciation (Pfennig et al. 2010). Therefore, the emergence of phenotypic plasticity can lead to evolutionary events. In accordance with this, plasticity has been shown to facilitate morphological novelty in spadefoot toads (Levis et al. 2018) and is additionally associated with greater species richness in various fish and amphibians (Pfennig and McGee 2010). Notably, in nematodes the emergence of a polyphenism in adult mouth-form led to a substantial increase in morphological novelty and speciation (Susoy et al. 2015).

## 1.4. Nematodes as models for evolutionary biology

### 1.4.1 Why use nematodes?

Nematodes, also known as roundworms, contain some of the most convenient organisms for studying evolutionary processes. Short generation times, small genomes and self-fertilising hermaphroditism (androdiecy) are features of several nematodes, including *Caenorhabditis elegans*. Many of these hermaphroditic organisms also produce functional males, which enables genetic crosses (Avisé 2011). *C. elegans* is the most well-established nematode model and research within this nematode led to a number of noteworthy findings, including mapping of the somatic cell lineage (Sulston and Horvitz 1977), apoptosis (Ellis and Horvit 1986), small RNAs (Lee et al. 1993), RNA interference (Fire et al. 1998) and recently, single-cell mapping of the entire nervous system (Taylor et al. 2021). *C. elegans* alone has thus provided numerous advances in genetics, neurobiology and developmental biology.

It is important to note that nematodes are the most abundant animals on Earth (Bardgett and van der Putten 2014) and therefore, there are many other organisms to study that display diverse traits and thus allow contrasting investigations from what is performed in *C. elegans*. One notable family is that of the Diplogastridae, or diplogastrids. These nematodes belong to the same clade as *C. elegans*, based on the phylogeny by Blaxter et al. (1998), but are otherwise distantly related, with more than 200 million years of evolution separating the Diplogastridae and Rhabditidae (in which *C. elegans* belongs) (Dieterich et al. 2008). Diplogastrids have diverse diets, with species capable of feeding on bacteria, fungi, other nematodes or all of them combined (Hechler 1971; Von Lieven and Sudhaus 2000; Kanzaki 2016). Aside from having the same aforementioned advantages as *C. elegans* in terms of generation time and genome size, the diplogastrids are of particular interest for the study of evolutionary biology as they display an ancestral polyphenism that results in discrete adult mouth-forms based on environmental cues. Diplogastrids typically form one-of-two morphs in an irreversible switch mechanism (Von Lieven and Sudhaus 2000). Morphs with a narrow mouth are termed Stenostomatous (St), meaning ‘narrow-mouthed’ in Greek. Morphs with wider mouths are termed Eurystomatous (Eu), meaning ‘wide-mouthed’. Plasticity in mouth form is an ancestral trait within the diplogastrids, with the vast majority of species displaying this dimorphism as adults (Susoy et al. 2015).

Nevertheless, there are cases where some species evolved additional morphs. For example, *Pristionchus triformis* has three potential morphs (Ragsdale et al. 2013), while some fig-associated nematodes can form up to five morphs (Susoy et al. 2016). However, the mechanism underlying how these new morphs evolved is an ongoing puzzle that can be further investigated using diplogastrid model systems.

#### **1.4.2 The mouth-form regulatory network in *Pristionchus pacificus***

*Pristionchus pacificus* has been developed as a model for evolutionary developmental biology and the evolution of novel traits, particularly in the study of phenotypic plasticity (Sommer 2015; Schroeder 2021). In contrast to most diplogastrids, there is a difference in the number and shape of teeth in *Pristionchus* morphs. Only the Eu morph has both dorsal and subventral teeth and is thus able to predate, while the St is a strict bacterial feeder (Bento et al. 2010). This indicates the morphs are adapted to different environmental conditions. A critical window of morph decision making has been determined in *P. pacificus* using complementary studies in different environmental conditions. The mouth-form decision is plastic up until approximately the penultimate juvenile stage, known as J3. It is during this time that the environmental conditions determine the final decision (Werner et al. 2023). It remains to be seen whether this window of developmental decision making is shared among all diplogastrids; however, this finding provides valuable insight into the mouth-form decision making process.

Numerous genetic tools have been established in *P. pacificus*, including transgenesis (Schlager et al. 2009), *in situ* hybridisation (Tian et al. 2008) and, most notably, CRISPR (Witte et al. 2014). These tools have enabled various reverse genetics approaches to target genes of interest. Through these approaches, the gene regulatory network behind *P. pacificus* mouth-form plasticity has been well-elucidated (Sommer 2020) with various 'switch genes' identified. These switch genes act as master regulators, where a frameshift knock-out leads to constitutive expression of only one morph. Switch genes have been postulated as playing a key role in plastic processes (West-Eberhard 2003) as it enables easier switching of morphs to produce a polyphenic trait. One of the key switch genes in the *P. pacificus* mouth-form network encodes a sulfatase that is homologous to *sul-2* in *C. elegans* (Ragsdale et al. 2013). Knocking out this gene prevents the Eu morph from being

formed. Due to this, the gene was termed *eurystomatous form defective* (*eud-1*). The *eud-1* gene is the result of a recent gene duplication in *P. pacificus* and is expressed in somatic and pharyngeal neurons, suggesting it may play a role in sensory transduction. This would be in agreement with its involvement in a polyphenic trait that relies on environmental signal to make developmental decisions. Notably, *eud-1* is part of a so-called ‘supergene’ locus. Two sulfatases are part of this locus along with two other mouth-form genes, *nag-1* and *nag-2*, that are in tandem orientation (Sieriebriennikov et al. 2018). These four genes are all tightly linked and regulate mouth-form determination. Although *eud-1* has been well-studied, its ancestral role in diplogastrids is largely unknown. It is present in a broad range of diplogastrids (Biddle and Ragsdale 2020), but functional studies of its orthologs outside of *P. pacificus* are few and far between. Indeed, the emphasis on *Pristionchus* nematodes alone can lead to a bias in evaluating evolution within all diplogastrids. Therefore, in order to determine the importance of *eud-1* or other switch genes within diplogastrids as a whole, other members of the family need to be functionally evaluated.

### **1.4.3 *Allodiplogaster sudhausi*, an early-branching diplogastrid**

To examine the evolution of traits in a clade, organisms which diverged early on in the lineage (often called ‘basal’ although this term can be misleading as they can also be under selection) are useful for comparative analysis. We chose to use the nematode *Allodiplogaster sudhausi* to determine the ancestral function of mouth-form genes within the Diplogastridae. *A. sudhausi* is a free-living diplogastrid with one available strain that was isolated in Israel. *A. sudhausi* has been suggested as a potential biocontrol agent for agriculture as it predates on plant-parasitic nematodes (Bar-Eyal et al. 2008). However, little else was previously known about this nematode. For this dissertation, *A. sudhausi* was selected for comparative analysis as it is: 1) Early-diverging (Susoy et al. 2015), 2) Phenotypically plastic in adult mouth-form and 3) Androdiecious, producing self-fertilizing hermaphrodites and functional males (Fürst von Lieven 2008). It is therefore ideal to evaluate evolution of mouth-form plasticity within diplogastrids and determine how conserved the mouth-form genes are within the family.

The majority of knowledge on *A. sudhausi* originates from the descriptive paper by Fürst von Lieven (2008). There it is described as being dimorphic, meaning it is able to become

either St or Eu as an adult. In contrast to *P. pacificus*, *A. sudhausi* has two teeth in all morphs and at all life stages after egg-hatching; it is thus capable of predating even as juveniles. The taxonomic classification of *A. sudhausi* has undergone some changes. It was initially termed *Koerneria sudhausi* (Fürst von Lieven 2008); however, the genus was revised and split into *Koerneria* and *Allodiplogaster*, of which *A. sudhausi* now falls under the latter (Kanzaki et al. 2014).

In this dissertation, I will describe my work in characterising this species and establishing it as a model for comparative approaches. This resulted in a number of findings, including: 1) A recent whole genome duplication concomitant with TE expansion and increased body size, 2) A novel third morph not known in other species, 3) Cannibalistic behaviour of this novel morph and, most notably, 4) Regulation of the novel morph by two highly conserved WGD-derived homologs of *eud-1*. The final finding gives insight into evolution of plastic mouth-form related genes in diplogastrids. My results reveal the homologs of *P. pacificus eud-1*, termed *A. sudhausi sul-2-A* and *sul-2-B*, regulate two of the morphs in *A. sudhausi* and act as switch genes in both cases. Therefore, the mouth-form regulatory network is conserved over large evolutionary distances between *A. sudhausi* and *P. pacificus* and is likely ancestral within diplogastrids. Interestingly, the same *sul-2-A* and *sul-2-B* genes regulate both the conserved Eu morph and the novel cannibalistic morph in *A. sudhausi*, suggesting the novel morph may have evolved by co-option of the pre-existing machinery. Thus, the new trait formed from what was already present. Overall, these findings reveal *A. sudhausi* as a potential model for evaluating evolutionary novelty.



## 2. Research objectives

The overarching objective of this dissertation was to examine how conserved mouth-form related genes are within the Diplogastridae. To do this, I first had to establish *A. sudhausi* as a model. I thus sequenced the genome and implemented reproducible assays to identify morphs. I performed the following lines of research:

- 1) I sequenced and assembled the genome and found evidence of WGD through sequence analysis, which I confirmed via karyotyping. I further characterised the event by examining i) timing, ii) functional bias in gene retention iii) expression divergence, and iv) associations with body size and transposon proliferation.
- 2) I investigated the sex determination system in *A. sudhausi* and determined that it evolved an environmental sex determination system, as opposed to genetic sex determination that occurs in *P. pacificus* and *C. elegans*.
- 3) I developed assays where *A. sudhausi* was grown under various diets in order to induce different adult mouth morphs. I found the following three diets i) *Escherichia coli* bacteria, ii) *C. elegans* larvae, and iii) *Penicillium camemberti* fungus could induce three distinct mouth-forms in *A. sudhausi* adults.
- 4) I determined the third morph, termed Teratostomatous (Te) is morphologically novel and additionally showed it is induced by starvation and crowding, suggesting it is a stress response.
- 5) I evaluated the differences in cannibalism between the morphs and determined only the Te morph cannibalised on its isogenic kin.
- 6) I identified four homologs of *P. pacificus eud-1* in *A. sudhausi*, which were knocked out using CRISPR/Cas9 to generate single, double and quadruple mutant lines. I grew these mutants on the three diets and ascertained that two *eud-1* homologs (WGD-derived duplicates) act as switch genes in regulating both the Eu and Te morphs.

To summarise, the primary goal of identifying conservation among mouth-form genes resulted in numerous other surprising observations that I subsequently followed up on. The absence of prior knowledge regarding *A. sudhausi* provided me with opportunities to establish new findings, which have led to the following papers and manuscripts.

## 3. Results

### 3.1 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 DR, Han Z. 2019.

Scientific Reports 9:1–9. <https://doi.org/10.1038/s41598-019-55359-5>

#### 3.1.1 Synopsis

This paper demonstrates how manual gene annotation is able to improve the accuracy of predicted gene models. 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 *P. 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, mis-annotation of *C. elegans* orthologs was due to artificially fused gene predictions and completely missing gene models. The community-based curation of gene annotation in *P. pacificus* increased the number of annotated genes and improved the single copy ortholog completeness level drastically, from 86% to 97%. This pilot study demonstrates how crowdsourcing, even at a small scale, can drastically improve gene annotations.

#### 3.1.2 Own contribution

I generated data for the paper by manually curating a subset of genes. I estimate my contribution as 10%

## **3.2 Geometric morphometrics of microscopic animals as exemplified by model nematodes**

Theska T, Sieriebriennikov B, **Wighard S**, Werner MS, Sommer RJ. 2020.

*Nature Protocols* 15:2611–2644. <https://doi.org/10.1038/s41596-020-0347-z>

### **3.2.1 Synopsis**

This paper presents a toolkit for performing geometric morphometrics to examine shape differences between nematode mouth-forms. There are surprisingly few tools for quantitative evaluation of morphology in microscopic animals. In this study, we provided a standardised protocol for geometric morphometric analyses of 2D landmark data sets using a combination of the geomorph and Morpho R packages. We integrated clustering approaches to identify group structures within such datasets. Furthermore, we demonstrated our protocol by performing analyses on stomatal shapes in the model nematodes *Caenorhabditis* and *Pristionchus*, wherein we were able to identify significant differences between *Pristionchus* mouth-forms. In theory, this approach is adaptable to all microscopic model organisms to facilitate a thorough quantification of shape differences within and across species. This paper therefore added to the methodological arsenal of evo-devo studies on morphological evolution and novelty.

### **3.2.2 Own contribution**

I was involved in the discussions, data generation and writing. I estimate my contribution as 20%.

### **3.3 A New Hope: A Hermaphroditic Nematode Enables Analysis of a Recent Whole Genome Duplication Event**

**Wighard S**, Athanasouli M, Witte H, Rödelsperger C, Sommer RJ. 2022.

*Genome Biology and Evolution* 14:evac169. <https://doi.org/10.1093/gbe/evac169>

#### **3.3.1 Synopsis**

This paper reports on a whole genome duplication (WGD) event in *A. sudhausi*. WGD is considered a major driver of evolution and the formation of phenotypic novelties; however, the importance of WGD is still controversial because most documented WGD events occurred anciently and few experimental systems amenable to genetic analysis are available. In this paper, we documented WGD in *A. sudhausi* through karyotype analysis and whole genome sequencing, and found that the event was relatively recent, after the divergence of *A. sudhausi* and *A. seani*. We evaluated the genome and: 1) Identified functional bias in retention of protein domains and metabolic pathways, 2) Determined most duplicate genes are under evolutionary constraint and 3) Showed a link between sequence and expression divergence. We additionally showed WGD is associated with increased body size and an abundance of repeat elements (36% of the genome), including a recent expansion of the DNA-hAT/Ac transposon family. Finally, we demonstrated the use of CRISPR/Cas9 to generate mutant knockouts, whereby two WGD-derived duplicate genes display functional redundancy in that they both need to be knocked out to generate a phenotype. Overall, the paper shows *A. sudhausi* is a useful model for examining and characterizing WGD-derived genes both computationally and functionally.

#### **3.3.2 Own contribution**

I participated in all aspects of this paper; study design, data generation, computational analysis and writing. I estimate my contribution as 85%.

### **3.4. Conserved switch genes regulate a novel cannibalistic morph after whole genome duplication**

**Wighard S**, Witte H, Sommer R.J. 2023. *Submitted*.

bioRxiv <https://doi.org/10.1101/2023.08.22.554244>

#### **3.4.1 Synopsis**

This manuscript describes the evolution of an additional mouth-form morph in *A. sudhausi*, and demonstrates both its morphological and behavioural novelty, as well as the conserved genes that regulate it. Phenotypic plasticity facilitates morphological novelty, but associated regulatory mechanisms remain elusive. Nematodes are powerful models to study developmental plasticity and its evolution. We showed *A. sudhausi* evolved an additional third mouth morph, concomitant with whole genome duplication and increase body size. These three morphs are induced by different diets; bacteria, fungi and nematodes. CRISPR experiments indicated that regulation of the third morph involves co-option of a conserved developmental switch gene, which through WGD resulted in two mouth-form regulators. Gene dosage studies revealed a diverged role of these developmental switches, with functional redundancy and quantitative effects in the two mouth-form decisions, respectively. The third morph is cannibalistic and kills kin, whereas the other two morphs do not. Thus, the recent evolution of a new morph relies on pre-existing regulatory mechanisms and adds behavioural and social complexity.

#### **3.4.2 Own contribution**

I devised experiments, generated data, performed all analysis and participated in all writing. I estimate my contribution as 90%.

### **3.5 Multiple evidence for environmental sex determination in the hermaphroditic nematode *Allodiplogaster sudhausi***

**Wighard S, Rödelsperger C, Sommer RJ. *In preparation.***

#### **3.5.1 Synopsis**

This manuscript shows that *A. sudhausi* displays an environmental sex determination system. Sex determination, though ubiquitous, is poorly understood in the majority of organisms. Genetic sex determination is one of the most common mechanisms and was for a long time thought to be ancestral in nematodes; however, recent work has uncovered other systems may be more abundant than initially thought. Here, we show the androdioecious nematode *Allodiplogaster sudhausi* displays the same chromosome number between sexes as well as no clear genomic differences between male and hermaphrodite DNA, indicating it does not have a genetic sex determination system. In contrast to this, growth at different temperatures greatly affects the sex ratio of progeny, with induction of males at lower temperatures and the complete loss of males at the highest temperature. Additionally, high density also promotes male generation across different conditions. Therefore, *A. sudhausi* has environmental sex determination, which must have evolved from the ancestral genetic sex determination system. Overall, these findings provide further evidence of rapid evolution of sex determination systems within nematodes.

#### **3.5.2 Own contribution**

I devised experiments, generated data, performed analysis and wrote the paper. I estimate my contribution as 85%.

## 4. Discussion

I initially established *A. sudhausi* as a model for comparative approaches to *P. pacificus*; however, the process of establishing and characterising this nematode has led to a number of surprising yet fascinating findings. This work can be condensed into three major findings: 1) *A. sudhausi* underwent a recent species-specific WGD, 2) *A. sudhausi* evolved an environmental sex determination system, and 3) *A. sudhausi* gained a cannibalistic third mouth-form morph that is regulated by conserved sulfatase genes. These three discoveries were all unexpected and indicate that, although it was chosen as a so-called 'basal' lineage for comparative analysis, *A. sudhausi* has gained numerous unexpected new traits. Thus, it serves as a valuable model for studying the evolution of novel forms and functions in its own right.

The discovery of WGD in *A. sudhausi* was particularly interesting as these duplication events are believed to promote evolutionary novelty (Ohno, 1970). The WGD resulted in chromosome doubling, increased genome size and greater body size (Wighard et al. 2022). Importantly, it enabled me to present and characterise a very rare find: an animal that has undergone WGD recently and is amenable to functional analysis via CRISPR engineering. These traits make it unique and ideal for studying WGD in animals. Indeed, I was able to evaluate redundancy, determine functional bias and show there was substantial expansion of DNA transposons after duplication. Nevertheless, there are virtually limitless aspects of WGD that can still be studied, particularly with the establishment of CRISPR (Wighard et al. 2022). For instance, potentially neofunctionalised genes could be knocked out in order to determine any effects. The positive association between WGD and evolution is often theoretically proposed; thus the identification of neofunctionalised genes could be key to show a clear link between them.

Intriguingly, a number of neofunctionalised genes that have so far been identified are also key sex determination genes (Matsuda et al. 2002; Kruger et al. 2019; Hayashi et al. 2022). Through examining *A. sudhausi* I found that it does not display the XX-XO genetic sex determination system found in both *C. elegans* and *P. pacificus* (Pires-daSilva 2007). Analysis of the genome showed it has no signs of genetic differences between sexes. I then clearly showed the effect of both temperature and crowding in changing the sex ratio of

progeny, confirming *A. sudhausi* has an environmental sex determination system (Wighard et al. 2023 *In preparation*). As the ancestral system is believed to be genetic in diplogastrids (Yoshida et al. 2023 *In preparation*), this finding suggests *A. sudhausi* independently evolved a different sex determination system.

Strikingly, in the process of establishing growth conditions, I also found that *A. sudhausi* evolved an additional mouth-form morph, which we termed Teratostomatous or Te (Wighard et al. 2023). This morph can be consistently induced by a combination of crowding and *P. camemberti* conditions and is likely a stress response. Interestingly, the Te morph displays distinct novelty, both morphologically, in terms of mouth size, and behaviourally, in its cannibalism of isogenic kin. I compared *A. sudhausi* with its sister species *A. seani* and found that, although clear St and Eu equivalents could be formed, there was no Te-equivalent in *A. seani*, suggesting it evolved recently. As Te worms display cannibalistic behaviour not seen in St or Eu, we hypothesise that the Te morph evolved as an extreme stress response when nutrient availability is poor. The induction of cannibalistic morphs means they can feed on their kin, gather nutrients and then produce eggs enabling long-term survival of the species. In conditions with sufficient bacteria and nematodes, the progeny would not become cannibalistic, displaying the benefits of phenotypic plasticity in variable environments.

Lastly, I explored the genetic regulators of mouth-form plasticity in *A. sudhausi*. I identified the WGD-derived duplicates that encode sulfatases, *sul-2-A* and *sul-2-B*, which act as key players. These two genes are homologous to *P. pacificus eud-1* which itself acts as a switch gene and controls Eu formation, with knock-out mutants remaining St (Ragsdale et al. 2013). I showed that, in *A. sudhausi*, the *sul-2-A* and *sul-2-B* genes also act as switch genes, with double mutant knock-outs in *sul-2-A/B* becoming neither Eu nor Te. This leads to a number of conclusions; firstly, the regulation of Eu by these sulfatase genes is conserved over large evolutionary distances, implying the regulatory machinery may be ancestral within the diplogastrids. Secondly, as these genes regulate both morphs, it suggests the Te genetic network may have been co-opted from the existing Eu machinery. This is consistent with many studies that posit novelty arises from genetic material that is already present (Wagner 2011), however further work would be necessary to better



elucidate the genetic network regulating Te morphs. Lastly, I found evidence of potential dosage effects in Eu regulation by *sul-2-A* and *sul-2-B*. With decreasing wild type copy number, the mouth width gets smaller. This is in contrast to the single mutant knock-outs on *P. camemberti*, which have a wild type phenotype and thus functionally display redundancy, which is fairly common after recent duplication events (Lynch and Conery 2000). Therefore, it appears *sul-2-A* and *sul-2-B* display some disparity in their regulation of the different phenotypes. Future work could involve evaluating the quantitative expression of RNA in single and double mutants at the time of mouth-form change and determining if there are differences in expression that can be correlated with mouth width, supporting the claim of dosage effects.

Although research in *A. sudhausi* led to numerous findings, there are some limitations to working with it. Notably, it has no known additional strains or closely related species that share its novel traits (Te morph, large body size and WGD). This prevents the use of certain tools, such as genetic mapping using recombinant inbred lines. It also leads to limitations in determining the timescale of WGD and its direct impact on novelty. For instance, comparing different strains or closely related species can help in determining the type of WGD (i.e. autopolyploidisation or allopolyploidisation) that has taken place (Blanc-Mathieu et al. 2017). Furthermore, the availability of other strains could help determine the specificity of cannibalistic behaviour, as previously shown in *P. pacificus* (Lightfoot et al. 2021). Nevertheless, while these limitations seem significant in the context of nematode studies, where most species have multiple strains, there are few other animal models outside of nematodes that have their advantages (e.g. short generation times, small genomes and self-fertilising hermaphroditism). Therefore, the downsides of working with *A. sudhausi* are few in relation to the findings that have been revealed in this organism, particularly in terms of morphological and behavioural novelty.

Overall, the findings presented in this dissertation confirm the functional conservation of genes involved in mouth-form plasticity within the nematode diplogastrid family. In addition, through establishing this unusual nematode, numerous other surprising discoveries emerged. With the finding of WGD, environmental sex determination as well as a cannibalistic third morph, *A. sudhausi* has emerged as a potential model for the evaluation

of the evolution of novel traits, both morphologically and behaviourally. Thus far, all cases of newly evolved systems: WGD, sex determination and the third morph reflect trends seen in other organisms, while providing new insight. These results further provide support for the hypothesis that novel forms are produced by co-option of material that is already present. Furthermore, this work in *A. sudhausi* reveals how studying non-model organisms can produce surprising results and give deeper insight into the evolution of derived traits. Overall, these findings contribute to the wealth of knowledge with regards to the evolution of novel traits in all its different varieties, and hopefully lays the groundwork for future research.

## 5. References

- Amores A, Force A, Yan Y-L, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang Y-L, et al. 1998. Zebrafish hox Clusters and Vertebrate Genome Evolution. *Science* 282:1711–1714.
- Assis R, Bachtrog D. 2013. Neofunctionalization of young duplicate genes in *Drosophila*. *Proceedings of the National Academy of Sciences* 110:17409–17414.
- Aury J-M, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, Anthouard V, Aiach N, et al. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444:171–178.
- Avise JC. 2011. Hermaphroditism: A Primer on the Biology, Ecology, and Evolution of Dual Sexuality. New York: Columbia University Press
- Bardgett RD, van der Putten WH. 2014. Belowground biodiversity and ecosystem functioning. *Nature* 515:505–511.
- Bar-Eyal M, Sharon E, Spiegel Y, Oka Y. 2008. Laboratory studies on the biocontrol potential of the predatory nematode *Koerneria sudhausi* (Nematoda: Diplogastridae). *Nematology* 10:633–637.
- Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461:1243–1247.
- Bazazi S, Buhl J, Hale JJ, Anstey ML, Sword GA, Simpson SJ, Couzin ID. 2008. Collective Motion and Cannibalism in Locust Migratory Bands. *Current Biology* 18:735–739.
- 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.
- Biddle JF, Ragsdale EJ. 2020. Regulators of an ancient polyphenism evolved through episodic protein divergence and parallel gene radiations. *Proceedings of the Royal Society B: Biological Sciences* 287:20192595.
- Birchler JA, Riddle NC, Auger DL, Veitia RA. 2005. Dosage balance in gene regulation: Biological implications. *Trends in Genetics* 21:219–226.

- Blanc-Mathieu R, Perfus-Barbeoch L, Aury JM, Da Rocha M, Gouzy J, Sallet E, Martin-Jimenez C, Bailly-Bechet M, Castagnone-Sereno P, Flot JF, et al. 2017. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes.
- Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, et al. 1998. A molecular evolutionary framework for the phylum Nematoda. *Nature* 392:71–75.
- Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, Van de Peer Y. 2006. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biology* 7.
- Bradshaw AD. 1965. Evolutionary Significance of Phenotypic Plasticity in Plants. In: Caspari EW, Thoday JM, editors. *Advances in Genetics*. Vol. 13. Academic Press. p. 115–155.
- Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wüster W. 2011. Domain Loss Facilitates Accelerated Evolution and Neofunctionalization of Duplicate Snake Venom Metalloproteinase Toxin Genes. *Molecular Biology and Evolution* 28:2637–2649.
- Clark JW, Donoghue PCJ. 2018. Whole-Genome Duplication and Plant Macroevolution. *Trends in Plant Science* 23:933–945.
- Claverys J-P, Håvarstein LS. 2007. Cannibalism and fratricide: mechanisms and raisons d'être. *Nat Rev Microbiol* 5:219–229.
- Collins JP, Cheek JE. 1983. Effect of Food and Density on Development of Typical and Cannibalistic Salamander Larvae in *Ambystoma tigrinum nebulosum*. *American Zoologist* 23:77–84.
- Conant GC, Wolfe KH. 2008. Turning a hobby into a job: How duplicated genes find new functions. *Nat Rev Genet* 9:938–950.
- Crossland MR, Hearnden MN, Pizzatto L, Alford RA, Shine R. 2011. Why be a cannibal? The benefits to cane toad, *Rhinella marina* [= *Bufo marinus*], tadpoles of consuming conspecific eggs. *Animal Behaviour* 82:775–782.

- Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.
- Elgar MA & Crespi BJ. 1992. *Cannibalism: Ecology and Evolution among Diverse Taxa*. Oxford University Press, Oxford.
- Dehal P, Boore JL. 2005. Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLoS Biology* 3:e314.
- DeVore JL, Crossland MR, Shine R, Ducatez S. 2021. The evolution of targeted cannibalism and cannibal-induced defenses in invasive populations of cane toads. *Proceedings of the National Academy of Sciences* 118:e2100765118.
- 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.
- Elgar MA, Schneider JM. 2004. Evolutionary Significance of Sexual Cannibalism. In: *Advances in the Study of Behavior*. Vol. 34. Academic Press. p. 135–163.
- Ellis H, Horvit HR. 1986. Genetic control of programmed cell death in the nematode *C. elegans*. *Cell* 44:817–829.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806–811.
- Flemming AJ, Shen ZZ, Cunha A, Emmons SW, Leroi AM. 2000. Somatic polyploidization and cellular proliferation drive body size evolution in nematodes. *Proceedings of the National Academy of Sciences of the United States of America* 97:5285–5290.
- Fürst von Lieven A. 2008. *Koerneria sudhausi* n. sp. (Nematoda: Diplogastriidae); a hermaphroditic diplogastriid with an egg shell formed by zygote and uterine components. *Nematology* 10:27–45.
- Giese AC. 1938. Cannibalism and Gigantism in *Blepharisma*. *Transactions of the American Microscopical Society* 57:245–255.
- Gout JF, Lynch M. 2015. Maintenance and loss of duplicated genes by dosage subfunctionalization. *Molecular Biology and Evolution* 32:2141–2148.

- Gregory TR, Hebert PDN, Kolasa J. 2000. Evolutionary implications of the relationship between genome size and body size in flatworms and copepods. *Heredity* 84:201–208.
- Griffith TM, Sultan SE. 2005. Shade tolerance plasticity in response to neutral vs green shade cues in *Polygonum* species of contrasting ecological breadth. *New Phytologist* 166:141–148.
- Guttal V, Romanczuk P, Simpson SJ, Sword GA, Couzin ID. 2012. Cannibalism can drive the evolution of behavioural phase polyphenism in locusts. *Ecology Letters* 15:1158–1166.
- Hayashi S, Suda K, Fujimura F, Fujikawa M, Tamura K, Tsukamoto D, Evans BJ, Takamatsu N, Ito M. 2022. Neofunctionalization of a Noncoding Portion of a DNA Transposon in the Coding Region of the Chimerical Sex-Determining Gene dm-W in *Xenopus* Frogs. *Molecular Biology and Evolution* 39:1–6.
- Hechler HC. 1971. Redescriptions of *Neodiplogaster tropica* Cobb and *N. pinicola* Steiner (Nematoda: Diplogasteridae). *J Nematol* 3:341–348.
- Hoffman EA, Pfennig DW. 1999. Proximate causes of cannibalistic polyphenism in larval tiger salamanders. *Ecology* 80:1076–1080.
- Joly S, Schoen DJ. 2021. Repeated evolution of a reproductive polyphenism in plants is strongly associated with bilateral flower symmetry. *Current Biology* 31:1515–1520.e3.
- Kanzaki N. 2016. Stomatal Dimorphism of *Neodiplogaster acaloleptae* (Diplogastromorpha: Diplogastridae). *PLOS ONE* 11:e0155715.
- Kanzaki N, Ragsdale E, Giblin-Davis R. 2014. Revision of the paraphyletic genus *Koerneria* Meyl, 1960 and resurrection of two other genera of Diplogastridae (Nematoda). *ZooKeys* 442:17–30.
- Kruger AN, Brogley MA, Huizinga JL, Kidd JM, De Rooij DG, Hu Y-C, Mueller JL. 2019. A Neofunctionalized X-Linked Ampliconic Gene Family Is Essential for Male Fertility and Equal Sex Ratio in Mice. *Current Biology* 29:3699–3706.e5.
- Lannoo MJ, Bachmann MD. 1984. Aspects of cannibalistic morphs in a population of *Ambystoma t. tigrinum* larvae. *American Midland Naturalist* 112:103–109.

- Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75:843–854.
- Levis NA, Isdaner AJ, Pfennig DW. 2018. Morphological novelty emerges from pre-existing phenotypic plasticity. *Nature Ecology and Evolution* 2:1289–1297.
- Lightfoot JW, Dardiry M, Kalirad A, Giaimo S, Eberhardt G, Witte H, Wilecki M, Rödelsperger C, Traulsen A, Sommer RJ. 2021. Sex or cannibalism: Polyphenism and kin recognition control social action strategies in nematodes. *Science Advances* 7:eabg8042.
- Lüning J. 1992. Phenotypic plasticity of *Daphnia pulex* in the presence of invertebrate predators: morphological and life history responses. *Oecologia* 92:383–390.
- Lynch M, Conery JS. 2000. The Evolutionary Fate and Consequences of Duplicate Genes. *Science* 290:1151–1155.
- Madl JE, Herman RK. 1979. Polyploids and sex determination in *Caenorhabditis elegans*. *Genetics* 93:393–402.
- Marburger S, Alexandrou MA, Taggart JB, Creer S, Carvalho G, Oliveira C, Taylor MI. 2018. Whole genome duplication and transposable element proliferation drive genome expansion in corydoradinae catfishes. *Proceedings of the Royal Society B: Biological Sciences* 285.
- Marcet-Houben M, Gabaldón T. 2015. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. *PLOS Biology* 13:e1002220.
- Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, Morrey CE, Shibata N, Asakawa S, Shimizu N, et al. 2002. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417:559–563.
- Michimae H, Wakahara M. 2001. Factors which affect the occurrence of cannibalism and the broad-headed “cannibal” morph in larvae of the salamander *Hynobius retardatus*. *Behavioral Ecology and Sociobiology* 50:339–345.
- Moczek A. 1998. Horn polyphenism in the beetle *Onthophagus taurus*: larval diet quality and plasticity in parental investment determine adult body size and male horn morphology. *Behavioral Ecology* 9:636–641.

- Moczek AP, Sultan S, Foster S, Ledón-Rettig C, Dworkin I, Nijhout HF, Abouheif E, Pfennig DW. 2011. The role of developmental plasticity in evolutionary innovation. *Proceedings of the Royal Society B: Biological Sciences* 278:2705–2713.
- Monson RK. 2003. Gene Duplication, Neofunctionalization, and the Evolution of C4 Photosynthesis. *International Journal of Plant Sciences* 164:S43–S54.
- Moriyama Y, Ito F, Takeda H, Yano T, Okabe M, Kuraku S, Keeley FW, Koshiba-Takeuchi K. 2016. Evolution of the fish heart by sub/neofunctionalization of an elastin gene. *Nature Communications* 7:10397–10397.
- Nijhout HF. 2003. Development and evolution of adaptive polyphenisms. *Evolution and Development* 5:9–18.
- Ohno S. 1970. Evolution by Gene Duplication. London: Berlin: Springer-Verlag
- Otto SP. 2007. The Evolutionary Consequences of Polyploidy. *Cell* 131:452–462.
- Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34:401–437.
- Papp B, Pál C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424:194–197.
- Parisod C, Holderegger R, Brochmann C. 2010. Evolutionary consequences of autopolyploidy. *New Phytologist* 186:5–17.
- Parker BJ, Brisson JA. 2019. A Laterally Transferred Viral Gene Modifies Aphid Wing Plasticity. *Current Biology* 29:2098-2103.e5.
- Pener MP. 1991. Locust Phase Polymorphism and its Endocrine Relations. In: Evans PD, editor. *Advances in Insect Physiology*. Vol. 23. Academic Press. p. 1–79.
- Pfennig DW, McGee M. 2010. Resource polyphenism increases species richness: a test of the hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:577–591.
- Pfennig DW, Reeve HK, Sherman PW. 1993. Kin recognition and cannibalism in spadefoot toad tadpoles. *Animal Behaviour* 46:87–94.



- Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek AP. 2010. Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution* 25:459–467.
- Pires-daSilva A. 2007. Evolution of the control of sexual identity in nematodes. *Seminars in Cell & Developmental Biology* 18:362–370.
- 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.
- Richardson ML, Mitchell RF, Reagel PF, Hanks LM. 2010. Causes and Consequences of Cannibalism in Noncarnivorous Insects. *Annu. Rev. Entomol.* 55:39–53.
- 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.
- Schroeder NE. 2021. Introduction to *Pristionchus pacificus* anatomy. *Journal of Nematology* 53:1–9.
- Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, Fukui A, Hikosaka A, Suzuki A, Kondo M, et al. 2016. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* 538:336–343.
- 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 Reports* 23:2835-2843.e4.
- Simpson SJ, Sword GA, Lorch PD, Couzin ID. 2006. Cannibal crickets on a forced march for protein and salt. *Proc. Natl. Acad. Sci. U.S.A.* 103:4152–4156.
- Sommer RJ. 2015. Nematoda. In: *Evolutionary Developmental Biology of Invertebrates*. Vol. 3. Springer Berlin Heidelberg. p. 15–34.
- Sommer RJ. 2020. Phenotypic Plasticity: From Theory and Genetics to Current and Future Challenges. *Genetics* 215:1–13.
- Sulston JE, Horvitz HR. 1977. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Developmental Biology* 56:110–156.

- 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.
- Susoy V, Ragsdale EJ, Kanzaki N, Sommer RJ. 2015. Rapid diversification associated with a macroevolutionary pulse of developmental plasticity. *eLife* 4:1–39.
- Taylor SR, Santpere G, Weinreb A, Barrett A, Reilly MB, Xu C, Varol E, Oikonomou P, Glenwinkel L, McWhirter R, et al. 2021. Molecular topography of an entire nervous system. *Cell* 184:4329–4347.e23.
- Tian H, Schlager B, Xiao H, Sommer RJ. 2008. Wnt Signaling Induces Vulva Development in the Nematode *Pristionchus pacificus*. *Current Biology* 18:142–146.
- Triantaphyllou AC, Riggs RD. 1979. Polyploidy in an Amphimictic Population of *Heterodera glycines*. *Journal of nematology* 11:371–376.
- 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.
- Wagner A. 2011. The origins of evolutionary innovations: a theory of transformative change in living systems. OUP Oxford
- Wallace, A. R. 1858. On the Tendency of Varieties to Depart Indefinitely from the Original Type. *Journal of the Proceedings of the Linnean Society of London. Zoology* 3(9): 53–62.
- Walsh EJ, Zhang L. 1992. Polyploidy and body size variation in a natural population of the rotifer *Euchlanis dilatata*. *Journal of Evolutionary Biology* 5:345–353.
- Weider LJ. 1987. Life history variation among low-arctic clones of obligately parthenogenetic *Daphnia pulex*: a diploid-polyploid complex. *Oecologia* 73:251–256.
- Werner MS, Loschko T, King T, Reich S, Theska T, Franz-Wachtel M, Macek B, Sommer RJ. 2023. Histone 4 lysine 5/12 acetylation enables developmental plasticity of *Pristionchus* mouth form. *Nature Communications* 14:2095.
- West-Eberhard MJ. 2003. Developmental Plasticity and Evolution. Oxford University Press

- Wighard S, Athanasouli M, Witte H, Rödelsperger C, Sommer RJ. 2022. A New Hope: A Hermaphroditic Nematode Enables Analysis of a Recent Whole Genome Duplication Event. *Genome Biology and Evolution* 14:evac169.
- Wighard S, Witte H, Sommer RJ. 2023. Conserved switch genes regulate a novel cannibalistic morph after whole genome duplication. *bioRxiv:2023.08.22.554244*.
- Wighard S, Rödelsperger C, Sommer RJ. Multiple evidence for environmental sex determination in the hermaphroditic nematode *Allodiplogaster sudhausi*. *In preparation*
- Witte H, Moreno E, Rödelsperger C, Kim J, Kim JS, Streit A, Sommer RJ. 2014. Gene inactivation using the CRISPR/Cas9 system in the nematode *Pristionchus pacificus*. *Development Genes and Evolution* 225:55–62.
- Wolfe KH. 2001. Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics* 2:333–341.
- Wolfe KH. 2015. Origin of the yeast whole-genome duplication. *PLoS Biology* 13:1–7.
- Yoshida K, Witte H, Hatashima R, Sun S, Kikuchi T, Sommer RJ. Rapid chromosome evolution and acquisition of environmental sex determination in nematode androdioecious hermaphrodites. *In preparation*.
- Zakon HH, Lu Y, Zwickl DJ, Hillis DM. 2006. Sodium channel genes and the evolution of diversity in communication signals of electric fishes: Convergent molecular evolution. *Proceedings of the National Academy of Sciences* 103:3675–3680.
- Zhang J. 2003. Evolution by gene duplication: An update. *Trends in Ecology and Evolution* 18:292–298.

## 6. Appendix

OPEN

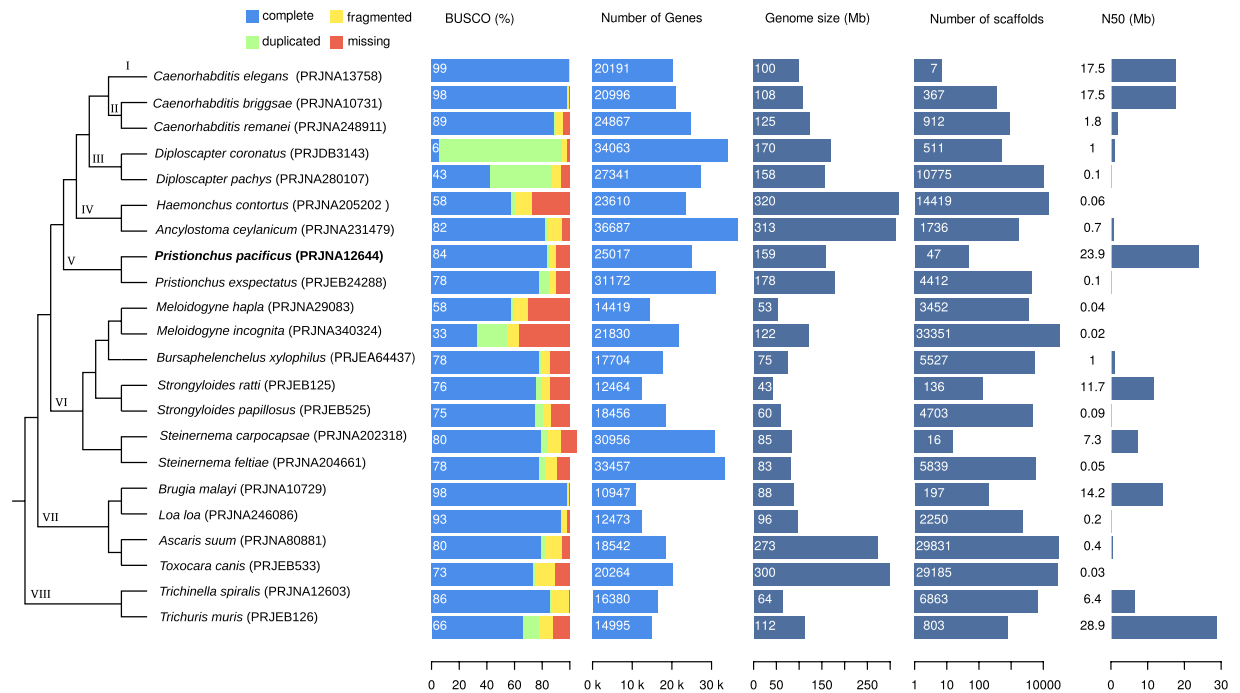
# Crowdsourcing and the feasibility of manual gene annotation: A pilot study in the nematode *Pristionchus pacificus*

Christian Rödelberger\*, 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*<sup>1,2</sup>. Since then, several comparative studies in developmental and ecological contexts have highlighted the importance of developmental system drift as a concept in evolution<sup>3</sup> and have demonstrated that the divergence between *Pristionchus* and *Caenorhabditis* was accompanied by extensive chemical<sup>4–6</sup>, genic<sup>7–9</sup>, and morphological<sup>10–12</sup> innovations. The establishment of multiple genetic<sup>13,14</sup> and genomic tools and resources<sup>15,16</sup> 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<sup>17–21</sup>. However, reverse genetic approaches based on candidate genes with known functions in *C. elegans*<sup>22,23</sup> have been hampered not only by the huge amount of lineage-specific duplications<sup>23–26</sup>, 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<sup>27–29</sup>. This can be complemented with evidence based approaches using transcriptomic and protein homology data<sup>30,31</sup>. While automated annotation pipelines perform reasonably well to be useful for genetic screens<sup>32–34</sup> and evolutionary genomic analyses<sup>35–37</sup>, their outcomes by far do not meet the standards of the gene annotations from classical model organisms such as *C. elegans*, *Drosophila melanogaster*, and *Mus musculus* that have been curated over decades by a large research community<sup>38</sup>. 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](mailto:christian.roedelsperger@tuebingen.mpg.de)



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

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

## Results

**The quality of nematode draft genomes is highly heterogeneous.** To obtain a general overview of the current status of nematode genome quality, we analyzed assemblies and gene annotations of 22 species (Fig. 1). The species were arbitrarily selected to span the diversity of the nematode phylum<sup>39</sup>. 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 (BUSCO<sup>40</sup>) reveal substantial amount of either missing or duplicated genes and it is not totally clear to what extent these differences are of biological or technical nature<sup>41</sup>. In the case of *Diploscapter coronatus*, the apparent high fraction of duplicated genes could either be explained by hybridization of two divergent lineages or a whole genome duplication<sup>42</sup>. The genome of *P. pacificus*, which was generated by assembly from single-molecule, long-read sequencing data and scaffolding with the help of a genetic linkage map<sup>15</sup>, shows one of the highest levels of contiguity (47 scaffolds, N50 = 24 Mb). Gene annotations were generated by the MAKER2 pipeline<sup>30,31</sup> which combined gene prediction algorithms, transcriptome data, and protein homology data from other *Pristionchus* species<sup>11,15,43</sup>. The completeness level of gene annotations (BUSCO completeness: 84%) is in the upper range when compared to most other nematode genomes (median 78%, interquartile range (IQR): 68–85%, Fig. 1). This demonstrates the relatively high quality of the current *P. pacificus* assembly and gene annotations.

**Complementary genome and transcriptomes reveal potentially missing gene models.** The completeness analysis as implemented in the software BUSCO<sup>40</sup> 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	<sup>15</sup>
El Paco annotation v1/WS268	84.0 (85.8)	1.8	4.3	9.9	<sup>15</sup>
<i>de novo</i> transcriptome assembly	59.1 (97.1)	38.0	2.6	0.3	<sup>16</sup>
Iso-Seq assembly	48.0 (73.3)	25.3	10.9	15.8	<sup>44</sup>
El Paco annotation v2	95.4 (97.1)	1.7	2.0	0.9	this study

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

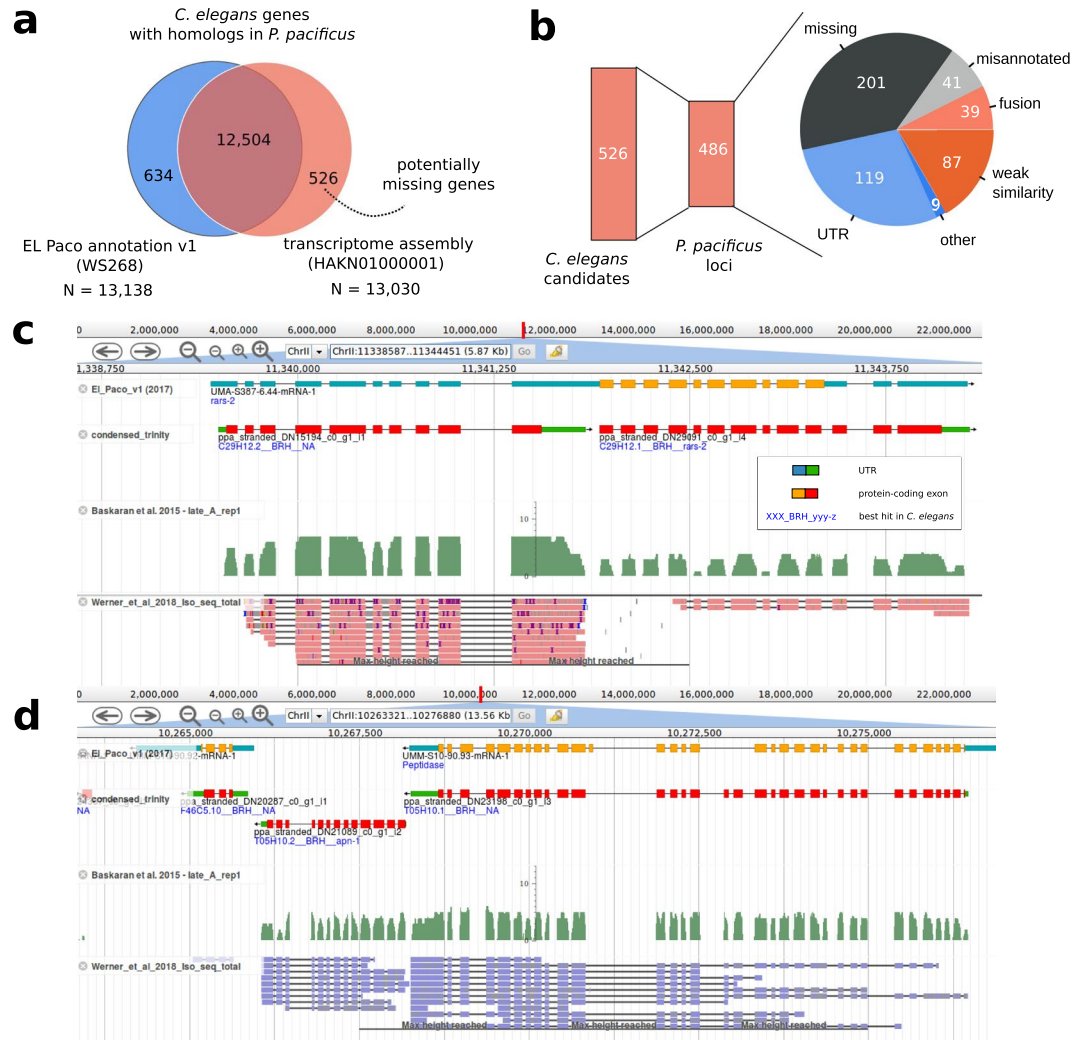
compared to 86% for the *P. pacificus* gene annotations and indicates towards the presence of incorrectly annotated or missing *C. elegans* orthologs in the genome of *P. pacificus*. Moreover, the fact that a recent *de novo* transcriptome assembly that was based on a strand-specific RNA-seq data set exhibited an even higher combined completeness level of 97% (Table 1) demonstrates even further room for improvement<sup>16</sup>. Finally, single-molecule, long-read transcriptome sequencing data were recently generated for *P. pacificus* which allows a much more accurate definition of gene structures from reference alignments of single reads<sup>44</sup>. 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<sup>-5</sup>) but having a matching open reading frame in the *de novo* transcriptome assembly (Fig. 2a). While 12,504 (62%) *C. elegans* genes had BLASTP hits in both data sets, 634 (3%) *C. elegans* genes showed only BLASTP hits against the current gene annotations suggesting that these genes are properly annotated but are expressed so weakly that they were not captured in the transcriptome assembly of mixed-stage cultures<sup>45</sup>. Similarly, we identified 526 (3%) *C. elegans* genes that were only found in the transcriptome assembly and therefore represent candidates for missing gene annotations.

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

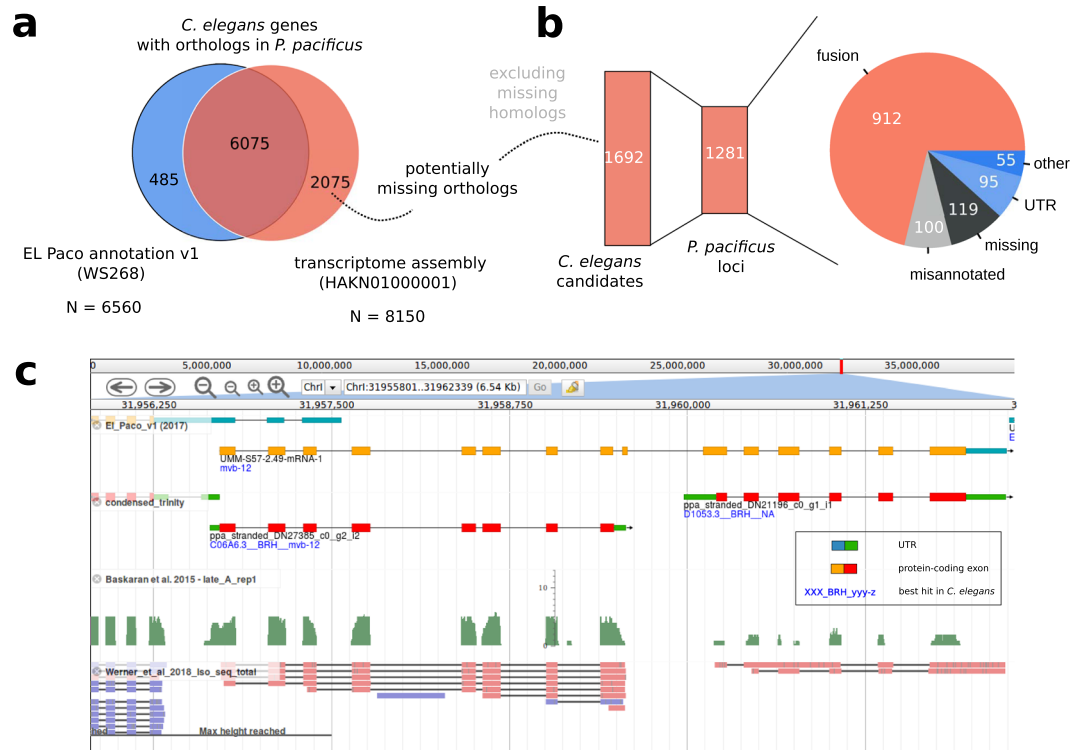
**Artificial gene fusions mask thousands of hidden orthologs.** A small number of *C. elegans* genes with missing homologs in the current gene annotations (version: El Paco v1/WS268) of *P. pacificus* were classified as located in fused gene models (Fig. 2b). One potential explanation could be that an 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 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 annotations<sup>15</sup> (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<sup>24,45,46</sup>, 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<sup>44,46</sup>. To further characterize *C. elegans* genes without orthologs in *P. pacificus*, we additionally carried out a phylostratigraphic analysis<sup>47</sup> to estimate their relative age. Basically, phylostratigraphy uses absence-presence patterns of a gene to map its origin to an internal branch in a species tree<sup>47</sup>. 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 orthAgo<sup>48</sup>. Computation of best reciprocal hits is a standard approach for predicting one-to-one orthologs across species<sup>49,50</sup>. In order to capture more complex orthology relationships (e.g. many-to-many), more general approaches such as Markov clustering have been widely applied<sup>48,51</sup>. Based on best reciprocal hits, we identified 8348 predicted one-to-one orthologs between both species (Supplemental Table 1) whereas the orthAgo 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 orthAgo could be explained by the fact that best reciprocal hits do not take inparalogs into account<sup>49</sup>. However, only 1049 (21%) of *C. elegans* genes that were not identified as one-to-one orthologs by orthAgo could be explained by the presence of lineage-specific inparalogs, suggesting that orthAgo with default settings might be too conservative for this analysis. This is further supported by the reanalysis of 57 one-to-one orthologous pairs that were previously confirmed by phylogenetic analysis<sup>46</sup>. 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<sup>2,3,52</sup>. For example, recent reverse genetic approaches in *P. pacificus* have revealed functional divergence of genes with known roles in *C. elegans* dauer formation<sup>22,23,53</sup>. 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<sup>33,54</sup>. Together with complementary studies of the functional importance of novel genes<sup>7,32,55</sup>, this makes nematodes an extremely powerful system to study genome evolution and gene function at a mechanistic level.

In order to facilitate fruitful functional studies across multiple model organisms, it is crucial to generate genomic resources (e.g. assemblies, annotations) and experimental genetic toolkits (e.g. forward and reverse genetics) of comparable quality. The chromosome-scale assembly of the *P. pacificus* genome<sup>15</sup> 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<sup>29,41</sup>, they cannot guarantee that gene annotations will only improve. In certain cases, new annotation pipelines will also cause new errors. In contrast, during manual inspection, each community curator has the choice to not propose any change of gene models in case of uncertainty. Thus, manual inspection should only lead to removal of errors and thus improve annotation quality without introducing biases elsewhere. While manual annotation is an incredibly tedious task that is probably not scalable to complete genomes<sup>38</sup>, 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<sup>16,44</sup> or previous rounds of gene prediction<sup>56,57</sup>. In our opinion, the most crucial aspect of this community project is a good training of new annotators. We achieved this by personal training sessions between experienced and new annotators and the possibility to always discuss cases of uncertainty with other curators. For larger projects, initial training could be achieved by comprehensive online tutorials and communication via email, but this will likely be less efficient. In the case of the *P. pacificus* gene annotations, our study raised the gene count from 25,517 to 28,036 and increased the single copy ortholog completeness level from 86% to 97%. In the *P. pacificus* genome, the greatest source of error was the artificial fusion of neighboring genes. This type of error might be more prevalent in nematodes where genomes are compact<sup>9</sup> and genes frequently overlap<sup>37,45</sup>. Consequently, manual annotation of restricted gene sets has been proposed and applied previously to circumvent this problem<sup>58</sup>. 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<sup>8,15</sup>, we hypothesize that errors due to artificial gene fusions should be much less pronounced at chromosome arms. To test this, an unbiased quantification of error rates across genomic segments would be needed. In future, we also plan to focus on large gene families and lineage-specific orphan genes<sup>55</sup> 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<sup>22,46</sup>, tissues<sup>10,46</sup>, environmental conditions<sup>59</sup>, sexes<sup>16</sup>, and genetic backgrounds<sup>60,61</sup> could be used to detect non-overlapping transcripts that exhibit anticorrelated expression within a single locus. Thus, while our study has demonstrated that community-based curation of gene annotations is feasible and can lead to substantial improvements, continued effort is needed to lift its quality to a level that would be similar to classical model organisms.

## Methods

**Comparative assessment of nematode genomes.** We downloaded 22 nematode genomes and corresponding protein sequences from WormBase ParaSite (release WBPS13). For *Steinernema carpocapsae*, the latest version at WBPS14 was used. In case of multiple isoforms, we selected the longest isoform for further analysis. We ran BUSCO (version 3.0.1) in protein mode (option: -m prot) against the nematode\_odb9 data set (N = 982 genes) to evaluate the completeness level of available protein sequences.

**Genome browser integration of transcriptomic resources.** To allow community annotators to propose alternative gene models, we integrated recent raw read alignments and reference guided transcript assemblies of Iso-seq data<sup>44</sup> and a *de novo* assembly of strand-specific RNA-seq data<sup>16</sup> into our genome browser (implemented in jbrowse<sup>62</sup>) on our webserver (<http://www.pristionchus.org>). Genomic coordinates for the *de novo* transcriptome assembly were generated by alignment to the *P. pacificus* reference genome (version El Paco) with the program exonerate<sup>63</sup> (version: 2.2.0, options: -m est2genome - dnowordlen 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<sup>-5</sup>) in

*C. elegans* was also incorporated into our jbrowse instance. In addition, predicted protein sequences of previous versions of *P. pacificus* annotations (Hybrid1<sup>56</sup> and TAU2011<sup>57</sup>) were mapped against the *P. pacificus* assembly by exonerate (version: 2.2.0, options: -m protein2genome -dwordlen 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<sup>-5</sup>) searches between *C. elegans* (version: WS260, longest isoform per gene) and two different *P. pacificus* data sets: the annotated proteins (version: El Paco v1, WS268) and the *de novo* transcriptome assembly<sup>16</sup>. 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<sup>-5</sup>) between *C. elegans* and *P. pacificus* and extracting best reciprocal hits from the BLAST output files. Simultaneously, the program orthAgogue was run with default setting on the same input files<sup>48</sup>.

### Data availability

The strand-specific *de novo* transcriptome was submitted to the European Nucleotide Archive under the accession HAKN01000001<sup>16</sup> and the Iso-seq data was submitted to the European Nucleotide Archive under the accessions ERX2315712 and ERX2315713<sup>44</sup>. 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., Seroby, V., Witte, H. & Baskaran, P. First insights into the nature and evolution of antisense transcription in nematodes. *BMC Evol. Biol.* **16**, 165 (2016).
46. Baskaran, P. *et al.* Ancient gene duplications have shaped developmental stage-specific expression in *Pristionchus pacificus*. *BMC Evol. Biol.* **15**, 185 (2015).
47. Domazet-Loso, T., Brajković, J. & Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* **23**, 533–539 (2007).
48. Ekseth, O. K., Kuiper, M. & Mironov, V. orthAgogue: an agile tool for the rapid prediction of orthology relations. *Bioinformatics* **30**, 734–736 (2014).
49. Remm, M., Storm, C. E. & Sonnhammer, E. L. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052 (2001).
50. Tatusov, R. L. A Genomic Perspective on Protein Families. *Science* **278**, 631–637 (1997).
51. Li, L., Stoeckert, C. J. Jr & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
52. Verster, A. J., Ramani, A. K., McKay, S. J. & Fraser, A. G. Comparative RNAi Screens in *C. elegans* and *C. briggsae* Reveal the Impact of Developmental System Drift on Gene Function. *PLoS Genetics* **10**, e1004077 (2014).
53. Sieriebriennikov, B., Markov, G. V., Witte, H. & Sommer, R. J. The Role of DAF-21/Hsp90 in Mouth-Form Plasticity in *Pristionchus pacificus*. *Mol. Biol. Evol.* **34**, 1644–1653 (2017).
54. Moreno, E. & Sommer, R. J. A cilia-mediated environmental input induces solitary behaviour in *Caenorhabditis elegans* and *Pristionchus pacificus* nematodes. *Nematology* **20**, 201–209 (2018).
55. Prabh, N. & Rödelsperger, C. Are orphan genes protein-coding, prediction artifacts, or non-coding RNAs? *BMC Bioinformatics* **17**, 226 (2016).
56. Borchert, N. *et al.* Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. *Genome Res.* **20**, 837–846 (2010).
57. Sinha, A., Sommer, R. J. & Dieterich, C. Divergent gene expression in the conserved dauer stage of the nematodes *Pristionchus pacificus* and *Caenorhabditis elegans*. *BMC Genomics* **13**, 254 (2012).
58. Stoltzfus, J. D., Minot, S., Berriman, M., Nolan, T. J. & Lok, J. B. RNAseq analysis of the parasitic nematode *Strongyloides stercoralis* reveals divergent regulation of canonical dauer pathways. *PLoS Negl. Trop. Dis.* **6**, e1854 (2012).
59. Sanghvi, G. V. *et al.* Life history responses and gene expression profiles of the nematode *Pristionchus pacificus* cultured on *Cryptococcus* yeasts. *PLoS One* **11**, e0164881 (2016).
60. Seroby, 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., McGaughan, 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](http://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



# Geometric morphometrics of microscopic animals as exemplified by model nematodes

Tobias Theska<sup>1</sup>, Bogdan Sieriebriennikov<sup>1,2</sup>, Sara S. Wighard<sup>1</sup>, Michael S. Werner<sup>1</sup>✉ and Ralf J. Sommer<sup>1</sup>✉

**While a host of molecular techniques are utilized by evolutionary developmental (evo-devo) biologists, tools for quantitative evaluation of morphology are still largely underappreciated, especially in studies on microscopic animals. Here, we provide a standardized protocol for geometric morphometric analyses of 2D landmark data sets using a combination of the geomorph and Morpho R packages. Furthermore, we integrate clustering approaches to identify group structures within such datasets. We demonstrate our protocol by performing exemplary analyses on stomatal shapes in the model nematodes *Caenorhabditis* and *Pristionchus*. Image acquisition for 80 worms takes 3–4 d, while the entire data analysis requires 10–30 min. In theory, this approach is adaptable to all microscopic model organisms to facilitate a thorough quantification of shape differences within and across species, adding to the methodological toolkit of evo-devo studies on morphological evolution and novelty.**

## Introduction

More than 150 years after *The Origin of Species*, identifying the mechanisms of morphological evolution remains a major focus in evolutionary developmental biology (evo-devo)<sup>1–6</sup>. In the modern era, investigations of morphological novelty primarily rely on three components: (i) morphometric quantification of evolutionary changes in the structure of interest, (ii) the usage of comparative developmental approaches to identify candidate genes and/or pathways promoting such morphological change and (iii) functional experiments aiming to demonstrate causative relationships between the observed morphological change and the candidate developmental mechanisms<sup>7–9</sup>. Dissecting the gene regulatory networks that underlie a given morphological trait has been a standard practice in the field, especially in studies on classical model organisms. However, efforts to quantify experimentally induced mutant phenotypes or anatomical differences resulting from divergent evolution have been scarce<sup>7,9</sup>. Here, we provide a protocol that facilitates the evaluation of differences in biological shape by combining geometric morphometrics with *k*-medoid and model-based clustering. This streamlined protocol can be used as a blueprint by all members of the evo-devo research community working on morphological evolution—particularly in microscopic animal taxa—to improve the ease of use and comparability of results. We hope that lowering the barrier to quantitative morphological techniques will stimulate future studies on morphological evolution in small animals to integrate all three components of contemporary evo-devo research.

## Development of the approach for nematodes

Several established approaches can be applied to nematodes to search for candidate genes underlying character development, including quantitative trait locus mapping, in situ hybridization and RNA-seq screens<sup>10–12</sup>. Functional experiments on candidate genes are now also feasible due to the recent establishment of precise genome editing tools like the CRISPR/Cas9 system<sup>13–15</sup>. In contrast, for the quantification of morphological changes—the first component of evo-devo research programs—most studies in nematodes still rely on traditional morphometry (i.e., linear measurements of structures<sup>16</sup>) and comparative qualitative descriptions of homologous anatomical elements. Only a handful of studies include more sophisticated methods of landmark-based geometric morphometrics to thoroughly quantify morphological data<sup>17–20</sup>. Thus, the power of geometric morphometrics has not

<sup>1</sup>Department for Integrative Evolutionary Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany. <sup>2</sup>Present address: Department of Biology, New York University, New York, NY, USA. ✉e-mail: [michael.werner@tuebingen.mpg.de](mailto:michael.werner@tuebingen.mpg.de); [ralf.sommer@tuebingen.mpg.de](mailto:ralf.sommer@tuebingen.mpg.de)

yet been fully realized in nematodes, despite it offering novel insights into the nature of morphological change and its implications for long-standing predictions of evolutionary theory.

One such prediction is that phenotypic plasticity, the ability of one genotype to produce different phenotypes based on environmental input, plays a significant role in the evolution of morphological novelty<sup>1,2</sup>. Recently, nematodes of the Diplogastridae family, specifically *Pristionchus pacificus* and its close relatives, have been developed as model systems to investigate the role of plasticity in development and evolution<sup>21–24</sup>. These species are characterized by the presence of prominent cuticular teeth as part of their stomata, which are absent in other nematodes like *Caenorhabditis elegans*<sup>25–27</sup>. In addition, *P. pacificus* displays a polyphenism in this morphological novelty, where adults irreversibly adopt either a wide eurystomatous (Eu) or narrow stenostomatous (St) mouth-form, based on a number of environmental cues. Mouth-form plasticity also has important consequences on feeding strategies; for example, in *P. pacificus*, only Eu animals are facultative predators on other nematodes and fungi, whereas St animals are strict bacterial feeders<sup>28</sup>.

The gene regulatory network controlling mouth-form plasticity in *P. pacificus*<sup>24</sup> has been revealed through unbiased genetic approaches and the subsequent functional manipulation of these genes<sup>29–32</sup>. In contrast, there is to date no standard protocol for quantifying morphological changes that accompany genetic perturbation. Similarly, the quantification of morphological changes through development and evolution has not been standardized in nematodes. Presumably, factors like their small body size and the triradial symmetry of their pharynx have prevented the previous establishment of sophisticated quantification methods, like geometric morphometrics, while such methods are regularly used in larger animals like vertebrates and arthropods<sup>33–35</sup>. Thus, while nematodes have traditionally been used as model systems in molecular cell biology, their vast diversity of form and the depth of taxonomic sampling renders them a largely untapped resource for fruitful evo-devo studies on morphological evolution. Here, we describe a protocol that can be applied to microscopic structures including nematode stomata to quantify morphological shape differences within and between species.

### Geometric morphometrics using R

The identification of patterns of shape variation in anatomical traits, as well as the covariation of shapes with environmental parameters, is a central goal of contemporary studies on ecology and evolution<sup>36–41</sup>. The most widely applied approach to quantitatively analyze shapes is landmark-based geometric morphometrics (GM)<sup>36,38</sup>. It utilizes 2D or 3D landmark coordinates containing information on the relative positions of homologous anatomical traits and boundary curves and surfaces, all of which serve as the basis for shape quantification<sup>36,38</sup>. Recently, many of the geometric morphometric techniques have been incorporated into various packages in the computational platform R/RStudio<sup>42–44</sup>. This platform has some inherent advantages: it is freely available for all common operating systems, an abundance of compatible software packages reduces the number of different applications needed to perform GM analyses and it is supported by a large active community of researchers from different fields<sup>43</sup>. Furthermore, several books, online forums and manuals and the interactive RStudio interface make it accessible to beginners. Two of the most widely used packages for GM analyses are geomorph<sup>45,46</sup> and Morpho<sup>47,48</sup>. While both of them are able to perform all fundamental steps of shape data analysis, each also comes with some useful functions that are unique to them. Therefore, a combination of both of these packages can be used to perform comprehensive geometric morphometric analyses.

### Overview of the procedure

The general procedure of this protocol follows the four steps of the Procrustes paradigm<sup>39</sup> plus an additional stage to identify group structures through unbiased clustering. The Procrustes paradigm is based on the acquisition of landmark configurations for all specimens under investigation (Steps 1–8). These data are then modified using general Procrustes analysis (GPA), so that all non-shape information is removed from the configurations (Step 20). The shape data obtained via GPA are subsequently used to (i) visualize and describe patterns of shape (co-)variation (Steps 21–25 and 28) and (ii) perform multivariate statistical testing of hypotheses regarding group differences (Step 26). These results can be complemented by unbiased clustering approaches to find and analyze potential group-structures in the shape data set (Step 27). The following sections explain each step in more detail, followed by a step-by-step protocol with exemplary data sets from *Caenorhabditis* and *Pristionchus* species. Nevertheless, the experimental design is applicable to all microscopic organisms or structures for which a set of reproducible landmark configurations can be designed.

### Landmark placement

First, the coordinates of defined sets of landmarks are acquired for each specimen<sup>36,39</sup> (Steps 1–8). These sets are called landmark configurations, and they contain information about the shape, size, orientation and location of homologous anatomical traits of interest. While landmarks always describe fixed homologous points defined by local anatomy, it is also possible to geometrically describe and quantify homologous curves or surfaces on which landmarks are sparse using ‘semilandmarks’<sup>49</sup>. A defined number of semilandmarks (Step 13) is placed along any given curve or surface in roughly corresponding positions across all specimens. The number of semilandmarks is dependent on the complexity of the curve or surface. Generally, one should use more semilandmarks with increasing structural complexity<sup>49</sup>. Unlike fixed landmarks, semilandmarks are allowed to slide along tangent lines to the respective curve (or tangent planes to the respective surface) that they describe to align their position, and thus establish a geometric correspondence of homologous curves or surfaces between individuals<sup>38,49</sup>. For semilandmarks located on a curve, it is generally desirable to start and end every curve with fixed landmarks, as this facilitates a reliable computation of tangent lines for the first and the last semilandmark along the curve<sup>49</sup>.

### GPA

In the second stage, GPA is performed to extract the shape information from all landmark configurations by removing the parameters of non-shape variation (i.e., differences in size, location and orientation) from the dataset<sup>36,38</sup> (Step 20). Information about location is removed by shifting the landmark configurations so that they share the same origin in a common coordinate system. Size differences are subsequently eliminated by scaling landmark configurations to an identical centroid size of 1.0. The centroid size describes the spread of landmarks around the centroid of a given landmark configuration and is calculated as the square root of the sum of squared distances between each landmark and the corresponding centroid<sup>36,38</sup>. All landmark configurations are subsequently rotated so that the sum of squared distances between single landmarks and their corresponding counterparts in the common target configuration is minimized<sup>36,38</sup>. Thus, orientation differences are eliminated. The resulting aligned landmark configurations therefore contain only shape information and are accordingly called ‘shapes’, with the landmarks representing a set of shape variables called ‘Procrustes coordinates’<sup>39</sup>. The GPA approach is iterative in nature and aligns all landmark configurations to a target configuration, which in the first iteration is one of the actual landmark configurations contained in the data set. After the first alignment, an average configuration is calculated, which becomes the target for the second iteration. With ongoing repetitions of this process, newly calculated average configurations become asymptotically more similar, and superimposition is halted after a threshold of similarity is reached (Step 20). The underlying convergence criterion is fixed for geomorph’s `gpgen()` function, but it can be manually set for Morpho’s `procSym()` function using the ‘`tol=`’ argument. Alternatively, it is possible to relax the convergence criterion in both functions by setting a maximum number of iterations (i.e., by adding the ‘`max.iter=`’ argument to `gpgen()` or ‘`iterations=`’ argument to `procSym()` in Step 20). Newly calculated average configurations are always rescaled to a centroid size of 1.0<sup>50</sup>. Thus, all shapes are optimally aligned to the calculated average and each other.

If landmark configurations contain semilandmarks, it is necessary to computationally slide them to establish a proper spacing between them and their bordering (semi-)landmarks<sup>49</sup> (Steps 13 and 20). Thus, their initial arbitrary placement on the curve or surface is aligned in reference to the average shape of the data set. In principle, semilandmark sliding can be achieved by minimizing either Procrustes distances (i.e., the same procedure that is used for landmark superimposition) or bending energy<sup>49</sup> (Step 20). The ‘minimize bending energy’ approach treats landmark configurations as ‘infinitely thin and infinitely large metal plates’<sup>38,50,51</sup> and describes the transformation from one 2D landmark configuration into another (i.e., the difference in shapes) by vertically bending this metal plate in the *z* axis. The amount of ‘work’ that is needed to bend the metal plate (i.e., a landmark configuration) has been coined the bending energy<sup>38</sup>. However, although bending energy is an inverse measure of a spatial scale, the metaphor of ‘bending metal planes’ is not to be equated with the actual biological processes that produce the observed differences in shape<sup>50</sup>.

Both available sliding approaches iteratively minimize the shape differences of curves or surfaces between each specimen and the average shape of the data set. In the bending energy approach, all semilandmarks slide together, and the sliding process is affected by fixed landmarks<sup>49</sup>. However, if Procrustes distances are minimized, each semilandmark slides separately, which can potentially result in semilandmarks passing each other or even sliding beyond one of the endpoints of a curve.



Therefore, minimizing bending energy is generally preferred for semilandmark sliding, even though it might consume more time for computation<sup>49</sup>.

Once the superimposition procedure (including semilandmark sliding) is finished, all of the obtained Procrustes coordinates describe the position of individual shapes in a curved (non-Euclidean) space, which is related to Kendall's shape space<sup>36,39,52</sup>. The non-Euclidean distances between landmarks in this curved shape space are called Procrustes distances<sup>36,46</sup>. However, the curvature of the shape space complicates the application of standard statistical procedures like multivariate analysis of variance (MANOVA) to such data sets (see 'Multivariate statistics' below). To avoid this problem, the Procrustes coordinates are projected into a linear tangent space (Step 20) that is defined at the level of the mean shape<sup>36,39,53</sup>. The Euclidean distances in the tangent space can be considered as approximations of the Procrustes distances and used for subsequent statistical analysis.

### Multivariate statistics

In the third stage of the Procrustes paradigm (Step 26), multivariate statistical analysis is performed on the aligned landmarks to identify biologically relevant differences in mean shape, with grouping factors usually including categorical variables like sex, age, culture conditions or species<sup>38</sup>. Traditional MANOVA is a well established parametric tool to compare the values of multiple variables between two or more experimental groups<sup>54</sup>. In geometric morphometrics, such variables are the X and Y (for 2D landmarks) or X, Y and Z (for 3D landmarks) coordinates of superimposed landmarks, which are projected into a linear tangent space. To facilitate statistical testing for differences in mean shapes using a MANOVA design, it is advisable to use the modified permutational MANOVA version 'PERMANOVA' (in geometric morphometrics 'Procrustes ANOVA'; Step 26)<sup>54–56</sup>. This test generates inferences based on permutations while no assumption of multivariate normality is made, which renders it more applicable to the type of biological data that are used in geometric morphometrics. Furthermore, the number of variables (i.e., Procrustes coordinates) can be higher than the number of specimens in the data set, as it is based on permutations rather than estimated degrees of freedom<sup>54</sup>. Thus, PERMANOVA has some inherent advantages that make it applicable to the Euclidean distances between corresponding landmarks in the linear tangent space<sup>45,53</sup>. Nonetheless, the use of permutation tests to obtain *P* values in Procrustes ANOVA designs does not circumvent the problem that this kind of statistical test is sensitive to heterogeneity of variances between groups<sup>54,55</sup>.

### Visualization

The last stage of the Procrustes paradigm involves visualization and description of shape (co-) variation patterns<sup>39,50</sup>. The first plots generated are typically scatterplots that show the position of superimposed shapes in a tangent space (Step 20; see Supplementary Figs. 3–5). These plots can be used to obtain a first informal impression of the data, but they do not show holistic patterns of shape variation, and therefore should not be relied upon for a thorough examination of the variation in the data set<sup>50</sup>. Instead, ordination methods like principal component (PC) analysis (PCA) should be utilized. PCAs generate scatterplots (Step 21) depicting the distribution of observations (i.e., specimens) along axes of major variation (i.e., PCs)<sup>57</sup>. Therefore, PCA can be used to visualize patterns of shape variation in tangent space and visually appreciate the importance of grouping factors on the axes themselves or on a combination of axes<sup>38,39</sup>. Extensive overlaps of groups in the PCA plot indicate that they share a common shape, while no overlap indicates morphological differences. To check for outliers, thin plate spline (TPS) interpolation<sup>51</sup> can be used, which warps shapes and computes deformation grids that depict the degree of stretching and compressing required to deform a starting into a target landmark configuration (Step 22). During this procedure, the TPS interpolation algorithm ensures that corresponding landmarks in the start and target configurations are located on exactly corresponding positions in the untransformed and transformed grids (or shapes)<sup>50</sup>.

Further useful information about shape differences can be gathered from the loadings and the PC scores obtained during PCA. The loadings describe the correlation between a PC and a variable (i.e., landmark coordinate)<sup>57</sup>, which can be used to estimate extreme shapes along PC axes (Step 24). Extreme shapes can be superimposed and then compared in deformation grids, lollipop plots or wireframes. These plots thus provide an impression of the shape variation that underlies the distribution of individuals in the morphospace. Wireframe plots are particularly useful to depict structural differences between extreme shapes because landmarks, which are placed on the same anatomical element, are connected by simple lines to give a simplified representation of the overall shape. Lollipop plots, on the other hand, are useful to visually emphasize the change in position that each landmark undergoes (relative to all other landmarks) when a shape is shifted along the

corresponding PC. However, if semilandmarks were slid by minimizing bending energy, their positional shifts cannot be interpreted easily, as only the shape of the curve they define has biological meaning. Thus, if bending energy was minimized, it is recommended to use wireframe plots that draw curve outlines.

It is also possible to estimate the relative contribution (%) of each landmark to the PCs<sup>58</sup> (Step 25). Landmarks that contribute more than expected can be explored and selectively displayed, for example in the context of a lollipop plot (Step 24), to emphasize which aspects of the anatomy most prominently influence the shape change along the respective PC. Taken together, PCA plots, deformation grids, wireframes and lollipop plots facilitate an evocative visualization and thorough interpretation of differences in mean shape.

### Identification of groups through clustering methods

Suspected group structure(s) in a data set can then be identified by clustering. A plethora of clustering software is available for R<sup>42</sup>, and we implement two commonly used methods, *k*-medoid clustering<sup>59,60</sup> (Step 27A) and model-based clustering<sup>61,62</sup> (Step 27B). In *k*-medoid clustering, a medoid is defined as the central-most data point in a cluster with the smallest sum of distances to all other observations within the cluster<sup>59,60</sup>. This is not to be confused with an alternative approach, called ‘*k*-means clustering’, in which data are partitioned around the mean (i.e., not around an actual data point) of a cluster<sup>59,60</sup>. First, a set of *k* representative medoids is chosen, and *k* corresponding clusters are generated around them. The algorithm then iteratively tests for better representative medoids by analyzing all possible combinations of representative medoids and non-representative observations for *k* clusters. The goal is achieved once a set of *k* medoids is identified that minimizes the sum of the dissimilarities of all observations to their closest medoids<sup>59,60</sup>. A heuristic and empirical way to determine the best number of medoids (*k*) for a given data set is to plot the average dissimilarity within the clusters against the number of clusters in which the data set can be arbitrarily partitioned (Step 27A(i)).

Alternatively, model-based clustering can utilize finite sets of Gaussian mixture models to identify group structures, which might underlie a data set<sup>61,62</sup>. Here, the best Gaussian mixture model for clustering, and therefore the best number of clusters, is selected according to the Bayesian information criterion<sup>61</sup>. In contrast to *k*-medoid clustering, model-based clustering independently identifies the best number of clusters and assort all observations accordingly<sup>62</sup> (Step 27B(ii)). Both clustering approaches can be used to identify the presence and the number of clusters in the morphospace. This might be of particular use if, for example, one wishes to identify the number of distinct morphs in polyphenic nematodes a priori (i.e., without assuming the number of morphs). Thus, the clustering approaches can be used to complement the results obtained from Procrustes ANOVA, as the latter tests for differences between groups that have been defined before the analysis. Alternatively, previously unknown group structures that were identified by clustering can be recursively tested with Procrustes ANOVA.

### Advantages and limitations of the method

Geometric morphometrics is a powerful technique for two explicit reasons. First, unlike traditional morphometrics, it captures information on spatial relations among landmarks, and it preserves these data throughout the entire analysis (statistical testing, PCA visualization, etc.). Second, it allows quantitative visualization of shape changes in an immediate anatomical context<sup>38,50</sup>. Thus, even the most complex morphological changes can be visualized and communicated in an effective and intuitive manner<sup>36,38,39,50</sup>. The set of landmarks depends on the microscopic animal of choice and particular anatomical question. In our exemplary data set, only a basic understanding of the nematode stoma is necessary to apply our proposed landmark configurations. However, users might need to practice landmark annotation before robust and repeatable data sets can be obtained for comparisons between different species and strains. In addition, only an elementary knowledge of the R language is necessary to apply the present protocol, as the entire code is included in the paper, and the computational workflows for geometric morphometrics and cluster analysis, which are utilized here, are well established and supported by active online communities<sup>45–48,59–62</sup>. Therefore, no particular expertise is needed to apply this protocol, rendering it feasible for all academic levels from undergraduates to postdoctoral researchers.

A current limitation of this protocol is that it is conceived for the analysis of 2D landmark data sets only, and that depending on the research question at hand, new sets of landmarks may need to be

defined. Furthermore, we have described semilandmarks to reconstruct curves, but a slightly modified procedure for surfaces may be required depending on the anatomical structure of interest. Finally, updates to the software packages used here might result in changes to the commands or in removal of deprecated functions, which might require adjustments to the code presented in this protocol in future applications.

### Applications and future developments

The protocol can be used and further developed in multiple circumstances in nematodes and other microscopic animals. As described here, it can be used for the quantification of both types of morphological variation, genetic and environmentally induced (phenotypic plasticity), in adult *Pristionchus* and *Caenorhabditis* mouth structures. This includes functional genetic experiments (i.e., CRISPR-induced mutation) and ecomorphological studies (i.e., multiple environmental conditions).

Second, although developed for *Caenorhabditis* or *Pristionchus*, the protocol can also be adapted to investigate the stomata of other nematode species. In such a phylogenetic context, the protocol can be applied to identify shape changes due to divergent evolution. In fact, recent studies demonstrated that geometric morphometrics can be useful in delimiting ‘cryptic’ centipede species<sup>35</sup>. Hence, this protocol could also be applied in future nematode species descriptions, where it might be of interest to visualize the shape differences between a reference dimorphism and any potential intermediate morphs, or additional stomatal morphs in highly plastic nematodes<sup>18</sup>. We also hope it will encourage members of the nematode research community to integrate geometric morphometrics into future studies focusing on morphological evolution and its underlying causal mechanisms.

Third, it can be adapted to quantify shape differences of other organs systems (e.g., spicules of male nematodes or imaginal wing discs in holometabolous insects like *Drosophila*). Fourth, one can define alternative landmarks sets, which can be applied to all larval stages to describe the ontogenetic trajectories along which shapes develop. Finally, our protocol can be easily adapted to other microscopic specimens, such as tardigrades, rotifers, copepods or mites (see ‘Experimental design’). Therefore, this simple, ready-to-use protocol fills a significant gap in the standard toolkit of evo-devo approaches on microscopic animals.

### Experimental design

#### Adapting the protocol for other microscopic animals

In the step-by-step protocol below, we demonstrate our approach by performing geometric morphometric analysis of nematode mouth structures. However, we anticipate that our protocol is broadly applicable to several other microscopic animals and structures. Our experimental design is based on image stacks acquired with differential interference contrast (DIC) on a standard upright light microscope. Thus, colleagues working on largely translucent animals such as tardigrades, rotifers, certain crustaceans (e.g., *Daphnia*), kinorhynch or loriciferans will profit most from our protocol, as they will be able to perform GM analysis on external and internal organs. Other microscopic animals like certain taxa of mites or copepods might be too opaque (due to stronger sclerotization or pigmentation) to clearly identify internal structures in standard DIC image stacks; however, GM on external body structures will still be possible.

The positioning of the animal will always play an important role for GM analysis. The orientation in which the animal is placed has to be as consistent as possible across specimens. Achieving this can be a challenge in animals that have rounded cross-sections, like nematodes or loriciferans. In such cases, it might be useful to prepare a larger number of individuals for imaging to ensure that enough samples in the proper orientation can be found (i.e., two to three times the number of desired samples). In the example described herein for nematodes, we used agarose pads as a substrate on which to place our animals and applied an anesthetic to keep them in position. This procedure works well in nematodes, and we suspect can similarly be applied to other microscopic animals. However, the diameter of some animals may be too large to place directly between an agarose pad and a cover slip. Instead, one can place specimens in a droplet of buffer and deposit small silicone pads in each corner of the cover slip as distance-holders between the glasses. For handling nematodes, we use self-made platinum sticks (see ‘Reagent setup’), which also allow handling of other small (but robust) animals such as mites or tardigrades. Very fragile specimens can be picked up with an eyelash glued to a tooth pick or directly placed in buffer and subsequently transferred to the microscope slide using a pipette. Apart from the specific requirements in maintaining and handling specimen cultures, the largest modification that will need to be made is to define a landmark configuration for the structure

of interest. The most important component to defining a configuration is to identify landmarks that can be reliably identified across all specimens. Thus, landmark configurations that were initially conceived to compare adults of a species might not be applicable to larval stages as well, since certain structures could still be absent (or rudimentary) in earlier developmental stages. These issues aside, only minimal changes need to be made to the code chunks given below to analyze and visualize the data, including: (i) defining sliding landmarks and curves on structures that lack options for fixed landmarks (Step 13), (ii) defining grouping factors of interest (e.g., species, culture conditions or larval stages) (Steps 14–19), (iii) defining the outlines for the wireframe plots (Step 24) and (iv) adjusting the model formula in the allometry assessment to visualize group-specific trends (Step 28).

### Example nematode data sets

The stomatal polyphenism of *Pristionchus* nematodes has become one of the most promising models in plasticity and novelty research. Therefore, we used this character complex to generate example data sets for this protocol (available as Supplementary Data 1–3). We obtained landmark data sets from the stomata of both morphs of *P. pacificus*, the secondarily monomorphic species *Pristionchus bucculentus* and *Pristionchus elegans*, as well as the stomata of *C. elegans* and *Caenorhabditis briggsae*, and performed exemplary geometric morphometric analyses to quantify shape differences between young adult hermaphrodites (or females). In addition, we performed exemplary clustering analyses on the PCA results obtained for the intraspecific comparison of *P. pacificus* morphs and for the interspecific comparison of stomata in several *Pristionchus* species. Lastly, we visualized the relationship of stomatal sizes and shapes in our data sets, to see whether shape differences on the species level or the polyphenism level are manifestations of allometry.

We followed the standard protocol for establishing and maintaining laboratory cultures of diplogastrid and rhabditid nematodes<sup>63</sup>. All species mentioned in this protocol were cultured on nematode growth medium agar plates for multiple generations using the uracil auxotroph OP50 strain of *Escherichia coli* as food source. Agar plates were stored in plastic boxes and kept in an incubator at a constant temperature of 20 °C. For microscopy, all specimens were mounted on slides with agarose pads containing sodium azide (NaN<sub>3</sub>) to sedate the animals (Steps 1–3). All specimens were examined using a Zeiss Axio Imager.Z1 microscope with a Zeiss Plan-Apochromat 100 × 1.4 DIC objective, together with immersion oil. All image stacks were taken in identical x-/y-dimensions (800 × 684 pixels) using a Zeiss AxioCam 506 mono charge-coupled device (CCD) camera (Steps 4 and 5). Landmark coordinates of 80 individuals per data set were acquired using Fiji's<sup>64</sup> multi-point measurement tool (Steps 6–8). All data analysis steps (GPA, PCA, multivariate statistics and clustering) were performed in RStudio, using the R statistical computing environment<sup>42</sup>. The geometric morphometric procedures were performed using the geomorph package<sup>45,46</sup>.

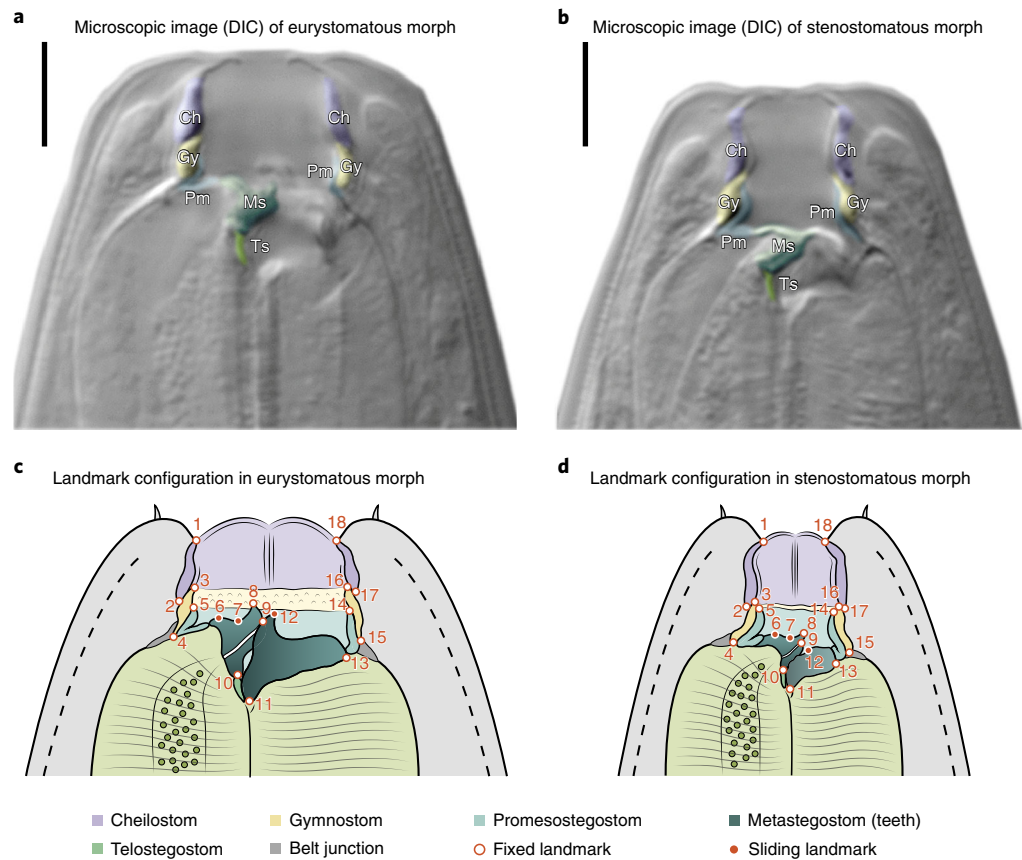
### Specific instructions for landmark configurations for model nematode stomata

The stomata of model nematodes, like *Pristionchus* and *Caenorhabditis*, can generally be divided into three cuticular compartments: cheilostom, gymnostom and stegostom (Fig. 1; Supplementary Fig. 1)<sup>25</sup>. The cheilostom is the anterior-most compartment of the stoma, which in *C. elegans* is secreted by hypodermal cells<sup>27</sup>. The gymnostom forms the second compartment of the stoma and is secreted by the arcade syncytium<sup>25,27</sup>. The stegostom is the posterior-most compartment of the stoma and is further subdivided into pro-, meso-, meta- and telostegostom<sup>25</sup>. The promesostegostom is secreted by pharyngeal epithelial cells, while meta- and telostegostom are secreted by pharyngeal muscle cells<sup>25,27</sup>. The metastegostom forms the characteristic teeth (i.e., the morphological novelty) in diplogastrids such as *P. pacificus*, whereas it forms simple triangular 'flaps' in other species, like *C. elegans* (Fig. 2a,b)<sup>26</sup>.

The two-dimensional landmark configuration proposed in Fig. 1 (also see Supplementary Fig. 1 and Supplementary Table 1) can likely be used for most species of the *Pristionchus* genus and many other nematodes of the Diplogastridae family, as the landmarks are placed on homologous structures, which are easily identifiable under a standard DIC microscope with a 100× oil objective. It is important to note that the proposed landmark configurations are suitable to quantify shape differences only in adult diplogastrids or *Caenorhabditis* nematodes, in which these landmarks can be appropriately identified. Thus, depending on the research question at hand—nematodes of different taxa, different developmental stages or other microscopic organisms—alternative or modified landmark configurations might need to be defined.

It is important to make sure that the orientation of all specimens is as similar as possible (Steps 3 and 4) to not introduce systematic error into the GM analysis. Since most nematodes tend to be lying on a side (either right or left), we applied the following criteria to make sure animals are facing right



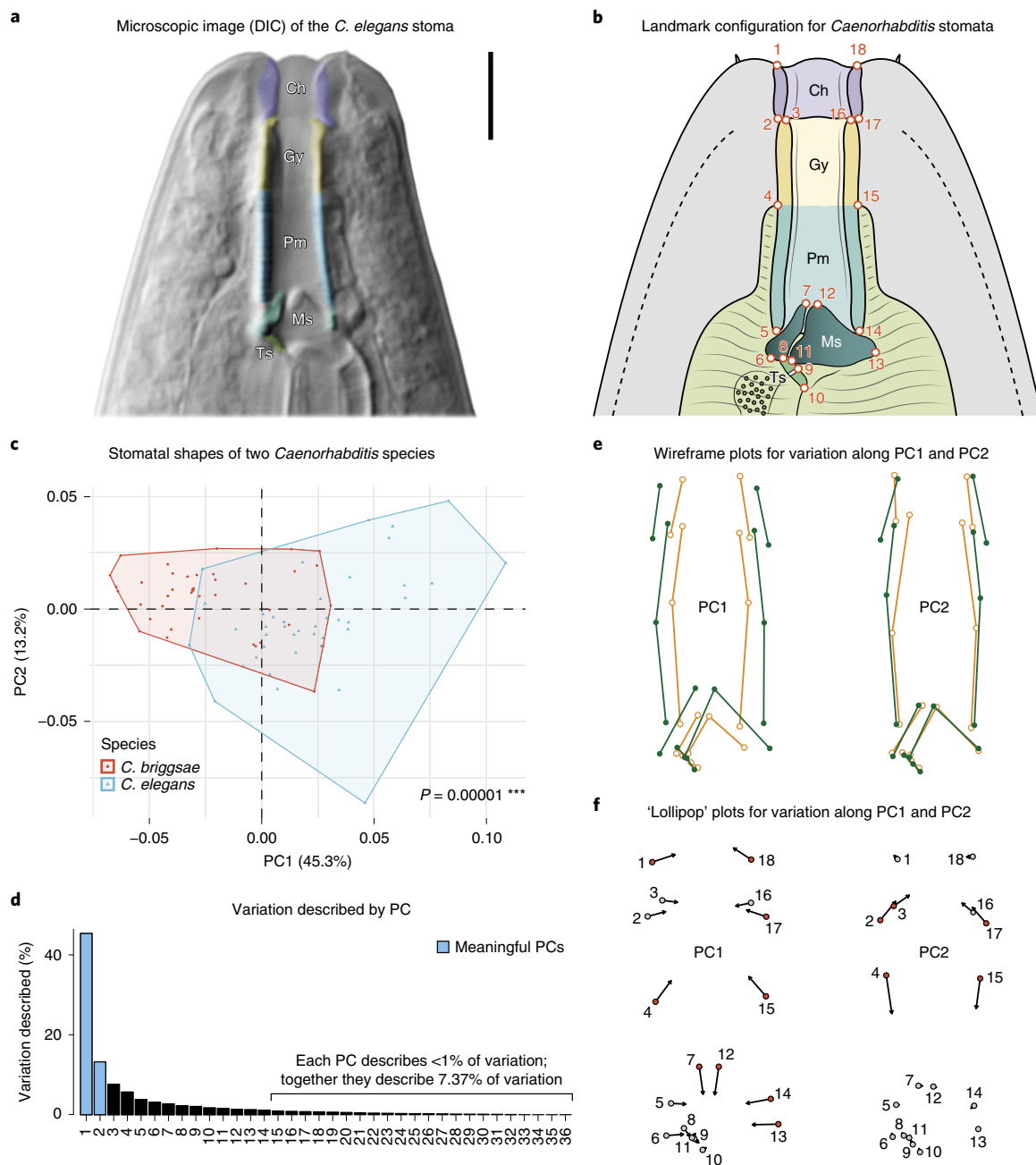


**Fig. 1 | The stomatal polyphenism in *Pristionchus pacificus*.** **a** and **b**, Overview of the stomatal cuticle elements in the eurystomatous (**a**) and stenostomatous (**b**) morph of *P. pacificus* (RS5205) under a 100× magnification using a DIC light microscope (only median plain; right lateral perspective). **c** and **d**, Schematic representation of the entire stomata of both morphs including the proposed landmark configurations (right lateral perspective; for a description of the landmarks, see Supplementary Table 1). Ch, cheilostom; Gy, gymnostom; Ms, metastegostom; Pm, promesostegostom; Ts, telostegostom. Scale bars in **a** and **b** are 5 μm.

side upwards (the right ventrosublateral tooth in diplogastrids is better visible in this position): (i) if the anterior end of the worm is on top of the field of view, and the posterior end of the worm is on the bottom field of view, the ventrally located vulva (or cloaca in males) must be found on the right side of the field of view, as the right lateral side of the animal is facing the viewer; (ii) confirm that the head of the worm is not rotated around its anterior-posterior axis by checking that the amphid opening is located close to the central axis of the animal, but usually with a slight offset toward the dorsal side of the body (i.e., opposite to the vulva or cloaca), and by checking that the right and left labial papillae align as close as possible with the central axis of the body; and (iii) confirm that the tips of the right and left labial papillae are approximately at the same level along the anterior-posterior axis to ensure that the head is not rotated around the dorsoventral body axis (i.e., tilted toward one of the lateral body sides). Note that rotation around the left-right lateral axis (i.e., the general orientation angle in which the worm appears in the image) can be ignored, as this kind of rotation is removed by GPA. Although not difficult, finding specimens that fit these criteria is the most time-consuming step in the protocol (Steps 1–5). In our experience, we typically found four to eight nicely oriented individuals per slide, which roughly equates to 10–20% of total specimens.

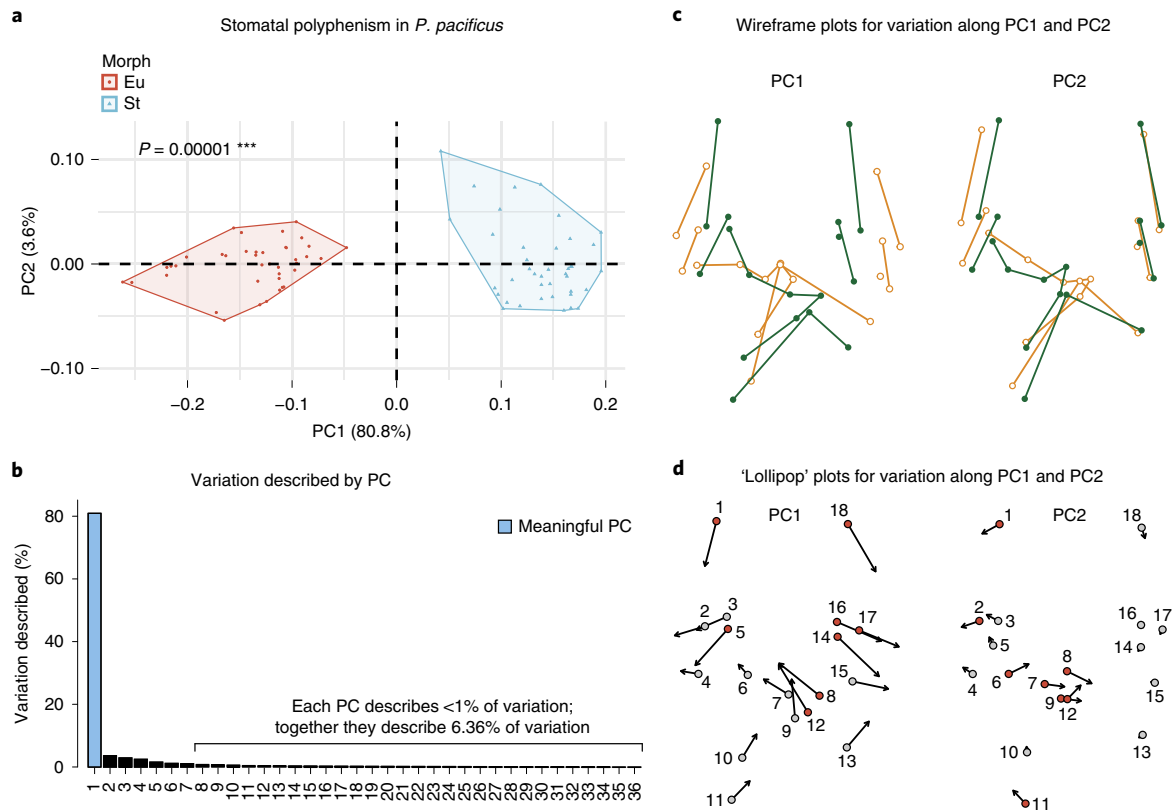
**Sample size**

It is important to have an appropriate sample size to perform geometric morphometric analyses and downstream multivariate statistical hypothesis testing. The number of specimens depends on the specific statistical test that is used for analysis of group differences. The permutation-based test designs, like PERMANOVA described above, do not assume any kind of distribution (e.g., Gaussian) of the data, and they are less sensitive to smaller sample sizes. Still, we recommend having at least two



**Fig. 2 | Analysis of shape differences between the stomata of *C. elegans* (N2) and *C. briggsae* (AF16).** **a**, 100× DIC image of the *C. elegans* stoma (only median plain; right lateral perspective). **b**, Proposed landmark configuration for the *Caenorhabditis* stoma (for a description of the landmarks, see Supplementary Table 2). Ch, cheilostom; Gy, gymnostom; Ms, metastegostom; Pm, promesostegostom; Ts, telostegostom. **c**, PCA plot showing the distribution of specimens (i.e., individual shapes) in a tangent space. *P* value obtained by Procrustes ANOVA. *F*-statistic: 19.928; effect size (*Z*): 5.5305; d.f.: 1 for species, 78 for residuals (total d.f.: 79). **d**, Barplot showing the variation described by each PC. Meaningful PCs are estimated by the `getMeaningfulPCs()` function of the `Morpho` package. **e**, Wireframe representations of superimposed stomatal extreme shapes along PC1 and PC2. Dark green, maximum (+) extreme; dark orange, minimum (−) extreme. **f**, Lollipop representations of superimposed stomatal extreme shapes along PC1 and PC2. Circles, maximum extreme; arrowtips, minimum (−) extreme. Landmark numbers are indicated in Arabic numerals. Red, landmarks that contribute more than average to the respective PC. Note that, when semilandmarks are slid by minimizing bending energy, their exact position on the curve cannot be interpreted easily. Thus, we recommend generating wireframe plots, rather than lollipop plots, in such cases. Scale bar in **a** is 5 μm.

to four times more individuals than shape variables<sup>36</sup> to obtain robust results with Procrustes ANOVA. In our examples, we used 18 landmarks each for *Pristionchus* or *Caenorhabditis* species (Fig. 1c,d; Supplementary Fig. 1; Fig. 2b), which means that there are 36 initial variables (i.e., *x* and *y* landmark coordinates) in every analysis. During Procrustes superimposition, four dimensions are lost due to removal of differences in location (two dimensions), orientation and size (one dimension

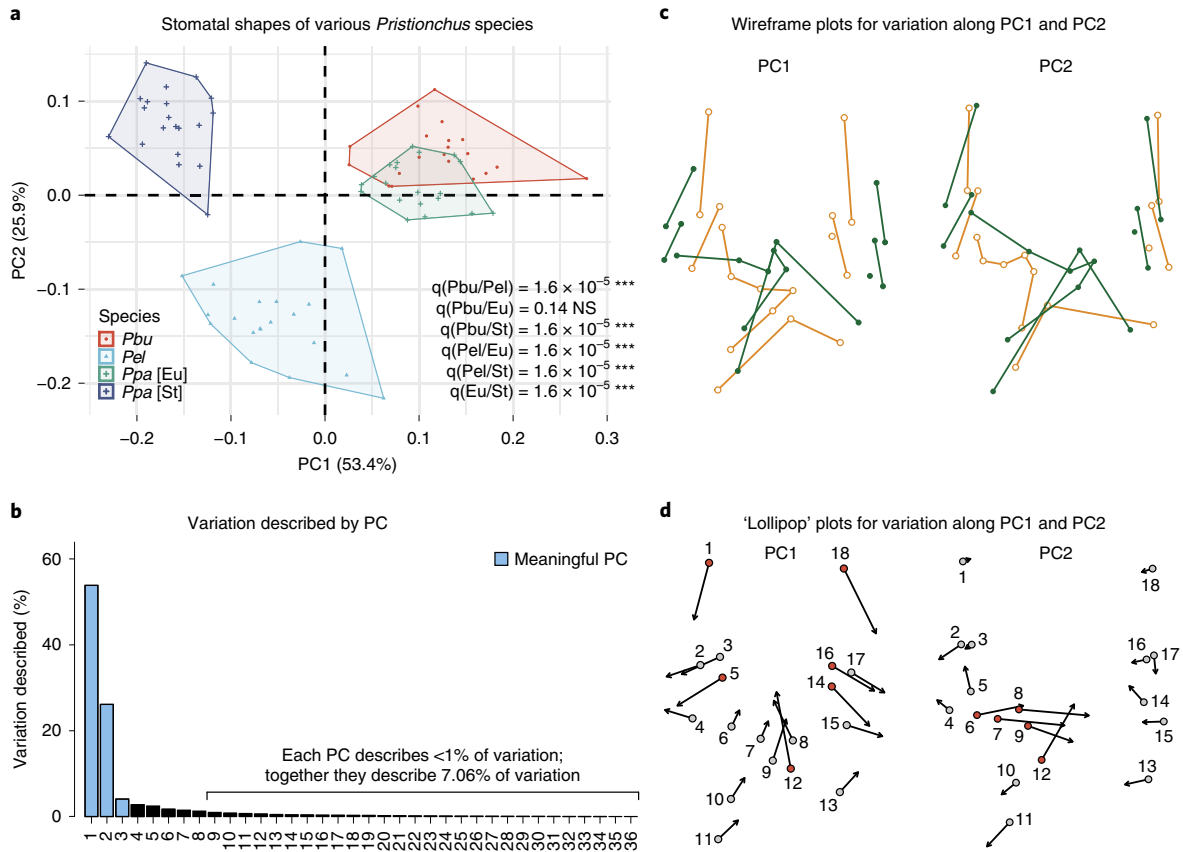


**Fig. 3 | Analysis of shape differences between the two stomatal morphs of *P. pacificus* (RS5205).** **a**, PCA plot showing the distribution of specimens (i.e., individual shapes) in a tangent space. *P* value obtained by Procrustes ANOVA. F-statistic: 218.31; effect size (*Z*): 6.0452; d.f.: 1 for morphs, 78 for residuals (total d.f.: 79). Eu, eurytomatous morph; St, stenotomatous morph. **b**, Barplot showing the variation described by each PC. Meaningful PCs are estimated by the `getMeaningfulPCs()` function of the `Morpho` package. **c**, Wireframe representations of superimposed stomatal extreme shapes along PC1 and PC2. Dark green, maximum (+) extreme; dark orange, minimum (−) extreme. **d**, Lollipop representations of superimposed stomatal extreme shapes along PC1 and PC2. Circles, maximum (+) extreme; arrowtips, minimum (−) extreme. Landmark numbers are indicated in Arabic numerals. Red, landmarks that contribute more than average to the respective PC. Note that, when semilandmarks are slid by minimizing bending energy, their exact position on the curve cannot be interpreted easily. Thus, we recommend generating wireframe plots, rather than lollipop plots, in such cases.

each)<sup>44</sup>. Therefore, we end up with  $36 - 4 = 32$  shape variables, and therefore  $\geq 64-128$  specimens should be contained in such datasets (here, we arbitrarily chose 80 animals for each example data set; Figs. 2c,3a and 4a). However, larger data sets might be necessary if the shape differences between groups are very small. In addition, we used equal numbers of individuals for each morph in the example analyses. However, it might be more useful to have different morphs represented in numbers that reflect morph frequency, especially if one aims to estimate the frequency from the results of the GM analysis. (Note that the *P. pacificus* strain RS5205 we use here shows a 1:1 morph ratio under standard laboratory conditions.) In summary, we recommend two to four times more individuals than shape variables with our pipeline for PERMANOVA statistical testing; yet, one has to make sure that the data set is large enough to address the specific question at hand.

**Replicability and error**

There are many sources of error in geometric morphometric analyses, including poor biological material or images, inherently variable anatomical structures and imprecise landmark labeling by the annotator. Here, it is crucial to assess whether a newly defined landmark configuration, such as the one proposed in Fig. 1 for model nematodes, can actually be used to produce clear and reproducible results. One can use several methods to assess the replicability of new landmark configurations for any microscopic animal. In this protocol, we provide two related tools to calculate replicability (Box 1): (i) estimation of the average repeatability (RPT)<sup>65,66</sup> of a shape by averaging the repeatability of Procrustes coordinates, and (ii) calculating the percent measurement error (%ME)<sup>43,44,67</sup> for shapes using PERMANOVA.



**Fig. 4 | Analysis of shape differences between the stomatal morphs of *P. pacificus* (RS5205), *P. bucculentus* (RS5596) and *P. elegans* (RS5698).** **a**, PCA plot showing the distribution of specimens (i.e., individual shapes) in a tangent space. *P* values obtained by Procrustes ANOVA and subsequently corrected for false discoveries (*q* values). F-statistic: 61.769; effect size (*Z*): 8.8245; d.f.: 3 for species, 76 for residuals (total d.f.: 79). Pairwise effect sizes (*Z*) for *Pbu*:*Pel* = 10.2778, *Pbu*:*Ppa*[Eu] = 1.3238, *Pbu*:*Ppa*[St] = 12.2343, *Pel*:*Ppa*[Eu] = 8.3855, *Pel*:*Ppa*[St] = 9.5263, *Ppa*[Eu]:*Ppa*[St] = 11.269. NS, not significant; Eu, eury stomatous morph; St, stenostomatous morph; *Pbu*, *P. bucculentus*; *Pel*, *P. elegans*; *Ppa*, *P. pacificus*. **b**, Barplot showing the variation described by each PC. Meaningful PCs are estimated by the `getMeaningfulPCs()` function of the `Morpho` package. **c**, Wireframe representations of superimposed stomatal extreme shapes along PC1 and PC2. Dark green, maximum (+) extreme; dark orange, minimum (–) extreme. **d**, Lollipop representations of superimposed stomatal extreme shapes along PC1 and PC2. Circles, minimum (–) extreme; arrowtips, maximum (+) extreme. Landmark numbers are indicated in Arabic numerals. Red, landmarks that contribute more than average to the respective PC. Note that, when semilandmarks are slid by minimizing bending energy, their exact position on the curve cannot be interpreted easily. Thus, we recommend generating wireframe plots, rather than lollipop plots, in such cases.

In this protocol, we replicated stomatal shapes for 20 individual *P. pacificus* nematodes (10 eury stomatous and 10 stenostomatous) using the configuration proposed in Fig. 1, and compared replicability for three annotators with varying levels of experience on stomatal morphology. Each of the annotators labelled all worms three times, with 20-min intervals between labelling specimens, on three different days spread over a 5-d period. Note that it was not possible to re-mount and re-image the same worm several times, as the animals do not survive such repeated stress. The 20 individuals were randomly selected from culture plates, and all specimen IDs had been encrypted during the replication sessions.

The RPT describes the fraction of the total phenotypic variance in a data set that can be attributed to variation among (biological replicates) or within (technical replicates) groups<sup>65,66</sup>. In our case, groups refer to the 20 individual worms. As an example, a repeatability of 80% would indicate that 80% of the total phenotypic variation is attributable to the variation that is naturally present between the 20 different specimens, while the remaining 20% of variation is attributable to differences between the three technical replicates obtained for each individual. In other words, 20% of the total phenotypic variance is due to the annotator’s imprecision of placing landmarks in the same specimen over several replicates. Here, we used the `rptR` package (version 0.9.22)<sup>66</sup> in R to estimate average repeatabilities obtained for each shape coordinate of all replicates (technical and biological). In addition, we calculated %ME for shapes based on a PERMANOVA design, following Claude<sup>43,44</sup>. This approach, similar to RPT, attempts to decompose the total phenotypic variation into within-group



**Box 1 | Testing the replicability of shapes by estimating RPT and %ME**

Before starting the replicability analysis, perform Steps 1–20 to obtain shape coordinates for a landmark data set containing biological and technical replicates (for details, see ‘Replicability and error’ in ‘Experimental design’). Expand the specimen ID (‘ID=’) so that it contains an additional part specifying the replicate number (e.g., ‘RS5205-Eu-ind-1-2’, where 1 indicates the number of the worm, and 2 indicates the technical replicate of that worm). Make sure to accordingly change the number of parts in Step 15 and to expand the code in Steps 16 and 17 so that you also extract the grouping factor information for ‘specimen’ and ‘replicate’ (e.g., we used ‘BR.part’ and ‘BR’ to get the information on the organism and ‘TR.part’ and ‘TR’ to extract the respective technical replicate) (see also Supplementary Data 4 and 5).

**▲ CRITICAL** When testing the code to estimate the RPT of your data, the number of parametric bootstraps and/or permutations should be changed from ‘1000’ to, e.g., ‘10’, as the full analysis will run for several minutes with both values set to ‘1000’.

**▲ CRITICAL** To calculate the %ME properly, the number of technical replicates (i.e., repeated measurements of the same animal) needs to be adjusted manually in the line that estimates ‘s2among’.

**Procedure**

To start the replicability analysis, run the following code:

```

```{r}
##Option-A: repeatability (RPT) estimation for shape coordinates
#load rptR package for repeatability estimation
library(rptR)
#generate input data frame for rptR package based on your superimposed landmarks:
#Input-Option-1 if superimposition was generated with *geomorph* use:
rptr_input <- sup.geo$data[grep("coord", colnames(sup.geo$data))]
#Input-Option-2 if superimposition was generated with *Morpho* use:
rptr_input <- data.frame(two.d.array(sup.mor$orpdata))
colnames(rptr_input) <- paste("coords", rep(seq(1, ncol(rptr_input)/2), each=2), rep(c("X", "Y"), ncol(
rptr_input)/2), sep=".")
#add a column with the names of the individuals
rptr_input <- cbind(rptr_input, Individual=paste(SP, MF, BR, sep="-"))
#rptR - Repeatability estimation for individual coordinates
rptr_RPT <- c()
rptr_SE <- c()
for (i in 1:(ncol(rptr_input)-1)) {
formula.i <- formula(paste(colnames(rptr_input)[i], "~
(1|Individual)", sep=""))
rptr_resul.i <- rptGaussian(formula = formula.i, gname =
"Individual", nboot = 1000, npermut = 1000, data = rptr_input)
rptr_RPT <- c(rptr_RPT, rptr_resul.i$R[1,1])
rptr_SE <- c(rptr_SE, rptr_resul.i$se[1,1])
}
# estimate average repeatability and standard error for whole shapes
mean(rptr_RPT)
mean(rptr_SE)
##Option-B: estimate percent measurement error (%ME)
#prepare a geomorph data frame for statistical analysis
ME_input <- geomorph.data.frame(shape = sup.geo$coords, individual = BR, replicate = TR)
#BR = biological replicate and TR = technical replicate
#Permutational Anova for replicate factor
replicate_factor <- summary(procD.lm(shape ~ TR, data = ME_input, iter = 1000, seed = 12345, RRPP = F, effect.
type = "F"))
#view influence of replicates on shape differences
replicate_factor
#Permutational Anova for specimen factor
individual_MS <- summary(procD.lm(shape ~ BR, data = ME_input, iter = 1000, seed = 12345, RRPP = F, effect.type
= "F"))
#view influence of individual differences on shape differences
individual_MS
# defining variation within individuals (across technical replicates)
MSwithin <- individual_MS[[1]] [2,3]
s2within <- MSwithin
# defining variation among individuals (across specimens)
MSamong <- individual_MS[[1]] [1,3]
s2among <- (MSamong - MSwithin)/3
#adjust number of technical replicates (here "3")
#calculate percentage of measurement error (ME) in the data set
ME <- s2within/(s2within + s2among) * 100
ME
```

```

and among-group variation components. However, %ME is based on the mean squares obtained via a PERMANOVA with individuals as sources of variation<sup>43</sup> and estimates the within-group variation (i.e., annotation imprecision)<sup>67</sup>. Meanwhile, RPT directly estimates the among-group variation (also in percentages)<sup>65</sup>. Thus, repeatability and %ME are complementary approaches for estimating replicability of landmark placement.

The estimated RPTs and %ME values in our exemplary data set clearly show differences in replicability of shapes (Supplementary Figs. 6 and 7), which related to the level of experience of the annotator. Interestingly, stenostomatous animals consistently show a smaller replicability across all users. Nevertheless, our replicability assessment passed two thresholds: (i) the amount of among-group variation was greater than the within-group variation (i.e., >50%, except for one case where  $100 - \%ME$  was 47.9%); and (ii) for %ME, the specimen factor was always significant for both *P. pacificus* morphs while the technical replicate factor was not (Supplementary Fig. 6 legend). This indicates that the shape variation across worms was stronger than the variation across replicates of the same individual<sup>44</sup>. Thus, our proposed landmark configuration (Fig. 1) can be used to produce robust shape data sets with high repeatability and a normal measurement error across different levels of experience. We advise checking your ability to replicate measurements using the same criteria before using a new landmark configuration for shape analysis. In our experience, practicing landmark placement for a few days before obtaining the first real data set will vastly increase replicability of results (Supplementary Figs. 6 and 7).

Another concern is that the anatomical character complex of interest contains moveable elements. Such structures can bias the entire GM analysis, since GPA aligns landmark configurations by translation, rotation and scaling without allowing for transformations like shearing. Thus, moveable elements oftentimes introduce systematic error, as their position relative to the rest of the landmark configuration can vary considerably. This caveat renders replicability assessments even more important. However, in addition to estimating shape repeatability or the percent measurement error of an annotator, we can also confirm the variance of each landmark across all specimens in a shape data set. This way, one can investigate whether landmarks that are placed on a moveable element collectively show increased variance in their position (after GPA), relative to the rest of the shape. This can be indicative of relative movements, which could lead the annotator to remove these specific landmarks from the analysis as they might introduce more systematic error than they provide morphometric information. For instance, we wondered if the metastegostomal teeth or flaps in our example nematode data set, which can be moved by muscular contractions, contribute excess variability compared to immovable elements. The  $\text{NaN}_3$  contained within the agarose pads sedates the worms and causes their musculature to relax. Thus, we expected that the teeth and flaps should always be in the same (non-protruded) position relative to the rest of the buccal cavity. Nevertheless, to validate that these movable elements do not introduce systematic errors, we performed a variance analysis of landmark position as mentioned above. Specifically, we performed Procrustes alignment (GPA) on 40 'Eu' and 40 'St' specimens of *P. pacificus*, as well as 40 *C. elegans* specimens. Afterwards, we estimated the variance of each landmark coordinate across each of the 40 individuals of the three different groups mentioned above. Then, we summed up the variances of the x and y coordinates of corresponding landmarks to obtain the total variance of each landmark in each of the three groups. This logic can be applied to any potentially moveable element in a microscopic organism. From our data, it is apparent that landmarks that are jointly placed on movable elements (teeth and flaps) do not show increased variances as compared to landmarks placed on non-movable elements (Supplementary Fig. 2), confirming that all of the proposed landmarks can be treated identically within both model species.

### Allometry

An important phenomenon one might want to consider when interpreting shape differences is allometry, the relationship of biological shape with size<sup>68–70</sup>. A common observation is that allometry is a major factor of shape variation in many morphometric data sets, as shapes of animals often change when they grow. To investigate ontogenetic allometry<sup>68,69</sup>, one can establish trajectories of shape change throughout development and subsequently compare their slopes across several groups of interest (typically species). However, it is also possible to check for static allometry (i.e., the size-shape covariation in animals within the same developmental stage), for example, in adults of different species or strains. Thus, static allometry can provide important insight into the coevolution of size and shape across species and facilitate studies on functional adaptation<sup>70</sup>.

Given its importance, several approaches have been applied to visualize allometry. A widely used class of plots that are based on linear models of shape change proved particularly informative when comparing allometries across different groups<sup>39,46</sup>. Here, we include one such approach (Step 28), in which a PCA is performed on predicted values obtained from a multivariate regression of shape on size, and the first PC of this PCA is plotted against the logarithm of the centroid size of each individual<sup>39,68,69</sup>. This way, one can obtain stylized graphs that show group-specific allometric trends in the form of straight lines. The slopes of these trend lines can then be compared to assess the presence of group allometry and to interpret differences in the degree of shape-size covariation between groups. In addition, statistical significance can be evaluated using, for example, the multivariate PERMANOVA approach we describe above.

## Materials

### Biological material

The approach described in this protocol can be applied to a broad range of microscopic animals (for details, see ‘Adapting the protocol for other microscopic animals’ in ‘Experimental design’). In the example shown in this protocol, we used the African strain RS5205 of *P. pacificus* (morph ratio of 1:1), the RS5596 strain of *P. bucculentus*, the RS5720 strain of *P. elegans*, the N2 reference strain of *C. elegans* and the AF16 reference strain of *C. briggsae*. Upon request, the Sommer laboratory can provide a plethora of *Pristionchus* strains, including the following: the reference strain PS312, numerous naturally isolated lines from multiple locations around the world, various transgenic and CRISPR-edited strains and a broad spectrum of other diplogastrid species and genera **!CAUTION** All national laws and institutional directives on animal welfare and the contained use of genetically modified organisms must be followed when working with such strains. The nematode strains/species used in this study were not genetically modified and are not included in the animal welfare act of Germany. Therefore, no ethical approval or guidance was required to work with any of the selected strains/species.

### Reagents

**▲ CRITICAL** The reagents specified in this section are specific for housing and sample preparation of nematodes. Different microscopic animals will require specific growth media and possibly different immobilization strategies (see ‘Experimental design’).

- Absolute ethanol (Merck, cat. no. 986) **!CAUTION** Ethanol is flammable and may cause serious eye irritation. Make sure to wear safety glasses while handling the substance and avoid using it close to open flames.
- Agarose (Sigma, cat. no. A9539)
- CaCl<sub>2</sub> (Sigma, cat. no. C5080)
- Cholesterol (Fluka, cat. no. 26740)
- KH<sub>2</sub>PO<sub>4</sub> (Sigma-Aldrich, cat. no. P3786)
- MgSO<sub>4</sub> (Fluka, cat. no. 63135-1KG-F)
- NaCl (Roth, cat. no. 3957.1)
- Na<sub>2</sub>HPO<sub>4</sub> (cat. no. S5136)
- NaN<sub>3</sub> (Sigma, cat. no. S8032) **!CAUTION** Sodium azide can be lethal if swallowed, absorbed through skin or inhaled; it may cause organ damage and is toxic to aquatic organisms. Make sure to protect yourself from skin contact by wearing Nitrile gloves.
- Tryptone (Fluka, cat. no. 61044)

### Equipment

#### Equipment for nematode handling and housing

**▲ CRITICAL** The equipment in this section is necessary for housing and handling of nematodes. Different microscopic animals may require specific handling strategies or housing conditions (see ‘Experimental design’).

- Glass alcohol burner (DWK Life Sciences, cat. no. 04-245-1)
- Glass pasteur pipettes (BRAND GmbH, cat. no. 7477 15)
- Incubator (Thermo Scientific Heraeus BK 6160, model no. 10759151)
- Platinum strips, 0.3 mm (Häberle, cat. no. 9.160 703)
- Petri dishes (Greiner Bio-One, cat. no. 628102)

### Imaging equipment

- Block heater (Techne, Dri-Block heater, model no. DB-3)
- CCD microscope camera (Zeiss Axiocam 506 mono, model no. 426557-0000-000)
- Coverslips, 18 × 18 mm (Carl Roth, cat. no. 0657)
- Flexible light guide (SCHOTT, cat. no. P/N 121.101)
- Gloves (e.g., KIMTECH Sterling nitrile, cat. no. 99212)
- Immersion oil (Zeiss Immersol 518F, cat. no. 444962-0000-000)
- Light source (SCHOTT KL 300 LED, cat. no. P/N 120 300)
- Microscope (Zeiss, model no. Axio Imager.Z1)
- Microscope lens wipes (Assistent, cat. no. 41019010)
- Microscope objective with DIC (Zeiss Plan-Apochromat 100 × 1.4 Oil, cat. no. 440780-9904)
- Microscope slides 76 × 26 mm (Carl Roth, cat. no. H868)
- Pipettes (2–20 µl, 20–200 µl, 100–1,000 ml) (Eppendorf, model Research @ plus, cat. nos. 3123000098, 3123000055 and 3123000063)
- Pipette tips (2–20 µl, 20–200 µl, 100–1,000 ml) (Greiner Bio-One, cat. nos. 765290, 739290 and 686295)
- Stereo microscope (Zeiss, Stemi 508, cat. no. 435064-9000-000)

### Software for image acquisition and data analysis

- Fiji<sup>64</sup> (ImageJ) (version 2.0.0 or later; <https://imagej.net/Fiji/Downloads>)
- Microsoft Excel (version 15.15 or later) **▲ CRITICAL** If you want to use Microsoft Excel, install version 15.15 or later. Alternatively, any other text editor of choice could be used to generate the landmark file.
- BBEdit (version 13 or later; <https://www.barebones.com/products/bbedit/download.html>)
- R (version 3.6.1; <https://cran.r-project.org/mirrors.html>) **▲ CRITICAL** Make sure to install the most recent version of R (3.6.1 or later; <https://cran.r-project.org/mirrors.html>) before you start analyzing your data.
- RStudio (version 1.2.5001 or later; <https://rstudio.com/products/rstudio/download/>) **▲ CRITICAL** Make sure to install the most recent version of RStudio (1.2.5001 or later; download available at <https://www.rstudio.com>) before you start analyzing your data.
- ZEN 2 Pro (Carl Zeiss Microscopy GmbH, 2011; version 2.0.14283.302; 64-bit)
- geomorph (R package version 3.2.1; <https://cran.r-project.org/package=geomorph>) **▲ CRITICAL** Version 3.2.1 of geomorph was just released, and the inbuilt plotAllometry() function currently seems to be unable to plot group-specific allometric trends (see Step 28). However, a corrected version of this function is provided by the package authors under the following link: <https://github.com/geomorphR/geomorph/blob/c027fbcf4c423a8e1b64b827966661d89194bcee/R/plotAllometry.r>. Please copy the function code from there and run it once in the console of R to overwrite the inbuilt version before you run the code chunk in Step 28.
- Morpho (R package version 2.8; <https://cran.r-project.org/web/packages/Morpho/index.html>)
- cluster (R package version 2.1.0; <https://cran.r-project.org/web/packages/cluster/index.html>)
- mclust (R package version 5.4.6; <https://cran.r-project.org/web/packages/mclust/index.html>)
- factoextra (R package version 1.0.7; <https://cran.r-project.org/web/packages/factoextra/index.html>)

### Reagent setup

**▲ CRITICAL** All items specified in Reagent setup are specific for housing and sample preparation of nematodes. Different microscopic animals likely require specific growth media and housing conditions (for details, see ‘Adapting the protocol for other microscopic animals’ in ‘Experimental design’).

- M9 buffer: Dissolve 3 g of KH<sub>2</sub>PO<sub>4</sub>, 6 g of Na<sub>2</sub>HPO<sub>4</sub> and 5 g of NaCl into 1 l of double deionized water and mix thoroughly. Autoclave the solution. Add 1 ml of 1 M MgSO<sub>4</sub> to the autoclaved solution. Store indefinitely at 4 °C.
- Nematode growth medium: Dissolve 17 g of agar, 2.9 g of NaCl and 2.5 g of tryptone in 1 l of double deionized water and mix thoroughly. Autoclave the solution. Cool to 55 °C and add the following (while swirling): 1 ml of cholesterol (5 mg/ml in ethanol), 1 ml of 1 M CaCl<sub>2</sub>, 1 ml of 1 M MgSO<sub>4</sub> and 25 ml of 1 M KH<sub>2</sub>PO<sub>4</sub> (pH 6.0). Store indefinitely at 4 °C.
- Platinum picks: Insert an ~4-cm-long platinum wire into the slender opening of a glass Pasteur pipette and shortly melt the glass tip over a glass ethanol burner to permanently attach the wire to the glass pipette. Let the glass cool down for 1–2 min. Slightly flatten and/or curve the tip of the wire with a

hammer to get a device that can be slid under the microscopic animal to pick it up. Picks can be stored at the workbench and should be disinfected using glass ethanol burners before handling specimens.

Procedure

**Digital microscopy ● Timing ~4 d for 80 animals (nematodes)**

**▲ CRITICAL** The procedures for sample preparation (Steps 1–3) and microscopy data acquisition (Step 4) need to be optimized for different animals (for details, see ‘Adapting the protocol for other microscopic animals’ in ‘Experimental design’). In this example, we specify how to generate a dataset for geometric morphometrics of nematode stomata.

- 1 Melt 4 ml of 5% (wt/vol) noble agarose in a microwave and add 40 µl of 1 M NaN<sub>3</sub> (10 mM final concentration). Use a block heater (set to ~70 °C) to keep the agarose solution liquid during the microscopy session.

**! CAUTION** NaN<sub>3</sub> may cause organ damage and can be lethal if swallowed, absorbed through skin or inhaled. Make sure to wear nitrile gloves to avoid skin contact with the substance while pipetting.

- 2 To get slides with agar pads, place three object slides next to each other (long sides contacting one another) and stick a line of common laboratory tape to the surface of the two outer slides to make them higher than the middle one. The strips of laboratory tape will act as spacers between the slides and ensure that the pad is of the right thickness. Use a pipette to place a drop of the agarose + NaN<sub>3</sub> solution on the object slide in the middle and flatten the drop into a pad by perpendicularly placing another object slide on top.
- 3 Place ~40 individual nematodes on each slide using 5–10 µl of M9 buffer as a medium between the agar pads and the coverslip.

**▲ CRITICAL STEP** Transfer nematodes (or other microscopic organisms) with self-made platinum picks and use glass burners filled with ethanol for sterilization of the pick.

- 4 Generate image stacks for the structure of interest in each specimen using a digital microscope with a CCD camera and a 100× DIC objective together with immersion oil.

**▲ CRITICAL STEP** All image stacks should be taken using the same settings. Keep the image dimensions constant for all image stacks that you obtain and use a roughly equal thickness (z-axis) for single images in all stacks. For nematode stomata, take ~40–50 images per z-stack. Note that image dimensions and the number of images per stack might need to be adjusted if the protocol is applied to anatomical structures other than stomata or to different microscopic animals.

**? TROUBLESHOOTING**

- 5 Save the image stacks as single files in either the native format for the microscopy software (e.g., ‘.czi’ for Zeiss’ ‘Zen 2 Pro’) or as single ‘.tif’ files, to retain scale and size information.

**▲ CRITICAL STEP** Make sure to save all image stacks in the same file format, to avoid potential differences in scaling.

**Obtaining landmark data for further analysis ● Timing ~6 h for 80 specimens**

- 6 Open each image stack in Fiji<sup>64</sup> and place all (semi-)landmarks on their respective locations by using the multi-point tool (Toolbar → Multi-point tool).

**▲ CRITICAL STEP** Always use the same total number of (semi-)landmarks for all of your specimens and place them in their exact numerical sequence.

- 7 Apply the measurement tool of Fiji<sup>64</sup> (Analyze → Measure) to obtain the x and y coordinates of each landmark and copy these coordinates to a spreadsheet in Microsoft Excel. Repeat this for all of your specimens.

**▲ CRITICAL STEP** Make sure to copy/keep only the landmark coordinates in the spreadsheet and not any other information from Fiji’s measurement panel.

- 8 Once all measurements are copied into a common spreadsheet, save this landmark file as ‘lands.txt’.

**▲ CRITICAL STEP** To use this landmark file for geometric morphometric analysis in RStudio, it has to have a format that is readable for the respective functions. Box 2 provides a step-by-step guide on how to format the text file manually. Alternatively, we provide a short R routine that can be used for automated formatting in Supplementary Data 6. Note that the functionality of the R routine can potentially be affected by the version of the text editor that was used to generate the original .txt file. The correctly formatted landmark files for all our example data sets are provided in Supplementary Data 1–3 and can be used for comparison.

**Box 2 | Manual formatting of the landmark file to fit the requirements of the readlands.tps() function**

The landmark file can be formatted using any text editor, as long as the following rules are followed:

- 1 In Microsoft Excel, arrange all landmark coordinates for separate specimens in subsequent rows of the same two columns, with columns one and two representing the x and y coordinates, respectively. Note that each subsequent specimen has to be separated by an empty row.
- 2 Give the total number of landmarks per specimen as initial information, by putting 'LM=x' (no quotation marks; x = number of landmarks) in the first row of the first column of every specimen. Lastly, give an ID to each specimen. In our example, we use the following ID code: 'Strain-mouthform-individual-number'. For a eurytomatous specimen of the *P. pacificus* strain RS5205, the ID for the third individual in the analysis would look like this: 'RS5205-Eu-ind-3'. Put 'ID=yourID' (no quotation marks; yourID = the chosen ID code) into the first column (i.e., x-coordinate) of the last row of every individual specimen in the spreadsheet.
- 3 Copy and paste the landmarks obtained with the measurement tool of Fiji into the text editor of choice (e.g., BBEdit). Make sure to show hidden symbols (view tab→text display→show invisibles). Remove all tabs by selecting 'detab' from the 'text' tab and replace them with a single space. Remove 'space-next line' with just 'next-line' by finding '\n' and replacing it with '\n' (without a preceding space). Finally, make sure there is a 'next line/return' on the last line, and replace all commas in coordinates with periods.
- 4 Save the output as a text file (.txt) and name it 'lands.txt' (e.g., Supplementary Data 1–3).

**Setting up R markdown for data analysis ● Timing 1–10 min**

- 9 Install and load the R packages necessary for data analysis and visualization. These packages include the following: geomorph and Morpho<sup>45–48</sup>, two alternative (but compatible) packages for GPA of points, curves and surfaces; cluster<sup>60</sup> for cluster analysis; mclust<sup>61,62</sup> for model-based clustering and factoextra<sup>58</sup> for extraction and visualization of the most important information generated by multivariate data analyses (e.g., PCA).
- 10 Open a new markdown file in RStudio and copy the code given below (including the code chunk delimiters ````{r}` and `````) into the upper left pane of RStudio (markdown pane), or load one of the ready-to-use markdown files we provide in Supplementary Data 4 and 5. Run the code chunks in the markdown pane separately by clicking on the 'play' button in its upper right corner. Packages can be found and downloaded via the 'packages' tab in the lower right pane of RStudio.

```
```{r}
library(geomorph)
library(Morpho)
library(mclust)
library(cluster)
library(factoextra)
```
```

**▲ CRITICAL STEP** The code chunks given in the main text are for a main data analysis based on the geomorph package<sup>45,46</sup>. If not noted otherwise, data containing figures depicted in this study were generated based on this code (see Supplementary Data 4 for the corresponding R markdown file). When semilandmarks were present, sliding was achieved by minimizing bending energy. An alternative code, which utilizes native functions of the Morpho package<sup>47,48</sup> for GPA and multivariate statistical analysis, is available as another ready-to-use R markdown file in Supplementary Data 5.

- 11 Set your working directory by pasting its path between the quotation marks in the setwd() function.

```
```{r}
setwd("D:\my\directory\morphometrics_folder")
```
```

**▲ CRITICAL STEP** Note that the example directory path we give below is for Microsoft Windows users. For MacOS users, folder names should be separated using slashes (e.g., /Users/my/directory/morphometrics\_folder).

- 12 Read the formatted landmark file 'lands.txt' using the readland.tps() function from geomorph and attribute it to a new object called 'lands'. Enter the path to your landmark file in between the quotation marks in the following code chunk.



```

```{r}
lands <- readland.tps (
"D:\my\directory\morphometrics_folder\lands.txt",
specID = "ID")
```

```

**▲ CRITICAL STEP** Note that the example directory path we give below is for PC users. For MacOS users, folder names should be separated using slashes (e.g., /Users/my/directory/morphometrics\_folder/lands.txt).

#### ? TROUBLESHOOTING

- Define the number of semilandmarks and their respective curves by using the `define.sliders()` function from `geomorph` and attribute each of the returned matrices to a new object called `'curveslide_partX'`, with 'X' being the number of the curve. After defining semilandmarks and curves, combine all of your curve objects into one matrix using the `rbind()` function. Assign the combined matrix to a single object called `'curveslide'`.

```

```{r}
curveslide_part1 <- define.sliders(c(5, 6, 7, 8), nsliders = 2)
curveslide_part2 <- define.sliders(c(11, 12, 13), nsliders = 1)
curveslide <- rbind(curveslide_part1, curveslide_part2)
curveslide
```

```

**▲ CRITICAL STEP** The `define.sliders()` function needs two inputs: (i) a vector containing a sequence of numbers that correspond to all (semi-)landmarks in the exact order in which they are placed along the curve and (ii) the number of sliding landmarks on that curve. Our proposed landmark configuration for *Pristionchus* nematodes contains two curves. The first curve runs between two fixed landmarks (landmarks 5 and 8), and two semilandmarks are allowed to slide along this curve (semilandmarks 6 and 7); the second curve runs between landmarks 11 and 13, with landmark 12 sliding between them (Fig. 1c, d). To define the sliding landmarks on a curve, create a vector with the `c()` function and list the curve landmarks in their correct sequence as an input (e.g., 5,6,7,8). Then, define the number of semilandmarks on the curve in the `'nsliders = 2'` argument. This way, the `define.sliders()` function treats the first and last values of the vector as fixed landmarks and all landmarks in between them as semilandmarks. Apply the same logic to all curves in your landmark configuration.

- Extract specimen IDs from the landmark file using the `dimnames()` function and assign it to a new object called `'lands.names'`. This object is a character that contains all IDs from all specimens in your dataset.

```

```{r}
lands.names <- dimnames(lands)[[3]]
```

```

- Define the number of parts in which the specimen ID can be split and save it to an object called `'n.parts'`.

```

```{r}
n.parts <- 4
```

```

**▲ CRITICAL STEP** In our case, the number of ID components is four (i.e., ID = 'RS5205-Eu-ind-01' can be split into 'RS5205', 'Eu', 'ind' and '01'). Adjust this number according to how you named your specimen.

- Define which of the four parts of the ID specify your grouping factors of interest.

```

```{r}
SP.part <- 1
MF.part <- 2
```

```

**▲ CRITICAL STEP** In our example, these factors are ‘strains of species (SP)’ and ‘mouth-form (MF)’, which are specified in the first and second parts of the specimen ID, respectively. Save this information in respective objects. In our example, we call them ‘SP.part’ and ‘MF.part’.

- 17 Extract the information on your grouping factors from the specimen ID and assign it to new objects.

```
```{r}
SP <- factor(matrix(unlist(strsplit(lands.names, "-")), length(lands.names), n.parts,
byrow=T)[,SP.part])
MF <- factor(matrix(unlist(strsplit(lands.names, "-")), length(lands.names), n.parts,
byrow=T)[,MF.part])
```
```

**▲ CRITICAL STEP** Here, we generate the two objects ‘SP’ and ‘MF’, which contain the information on species and mouth-form for each specimen, respectively. You may need to adjust the name of the object according to the grouping factor(s) of your own interest. Add the names of the objects you created in the previous step behind the comma in the square brackets at the end of both lines of code. This defines the part of the specimen ID that the factor() function has to use as an input. The strsplit() function splits the specimen IDs into parts, which are defined based on the hyphens in the ID name, and returns those parts as a list. The unlist() function will convert this list into a character (i.e., a simple vector containing all ID parts of all specimens). The matrix() function rearranges the output of the unlist() function into a matrix in which each column contains one part of the specimen ID and each row represents one specimen. The number of rows is extracted from the ‘lands.names’ object (see Step 14) and returned as an integer by the length() function. The number of columns is defined by the number of parts in the specimen ID (‘n.parts’). The matrix will be converted into a simple factor that contains only the information on your grouping factor(s) of interest for each specimen, as well as all levels of these factors. This is achieved by applying the factor() function. All of the functions in this code chunk come with the base R package.

- 18 Define a color code designating the levels of your first grouping factor. Use, for example, ‘darkgreen’ for the eurystomatous and ‘darkorange’ for stenostomatous (Fig. 1; Supplementary Fig. 3).

```
```{r}
g_color <- rep("", length(lands.names))
g_color[which(MF=="Eu")] <- "darkgreen"
g_color[which(MF=="St")] <- "darkorange"
```
```

**▲ CRITICAL STEP** Adjust the color coding according to your preferences and the number of grouping factor levels. If necessary, replace our first grouping factor (i.e., ‘MF’) and its levels (i.e., the two morphs ‘Eu’ and ‘St’) with your grouping factor and its levels in the following code chunk. More lines of code to assign various colors can be added as per your needs, if you have more or less than two grouping factor levels (e.g., Supplementary Fig. 4). The rep() function from base R is a generic function that replicates the action of designating a specimen with a color for all individuals in the data set (in our example, based on mouth-form). The which() function defines the subsets of specimens based on the level of the grouping factor of interest (e.g., grouping factor = ‘MF’ and level = ‘Eu’).

- 19 After defining a color code for the first grouping factor, assign different symbols to designate a second grouping factor of interest. Here, we use symbols to differentiate between species.

```
```{r}
g_pch<-rep(0, length(lands.names))
g_pch[which(SP=="Pel")] <- 21 #this number gives a circle
g_pch[which(SP=="Pbu")] <- 22 #this number gives a square
g_pch[which(SP=="RS5205E")] <- 23 #this number gives a diamond
g_pch[which(SP=="RS5205S")] <- 24 #this number gives a triangle
```
```



**▲ CRITICAL STEP** If necessary, replace our second grouping factor (i.e., ‘SP’) and its levels (e.g., ‘Pbu’ or ‘Pel’ for *P. bucculentus* and *P. elegans*, respectively) with your grouping factor of interest and its corresponding levels in the following code chunk. More species can be added with the same line of code we give below. The plotting symbols (e.g., triangles or squares) can be adjusted according to your preferences, by exchanging the pch numbers.

**Superimposition of landmark configurations ● Timing -1 min**

- 20 Perform the (semi-)landmark superimposition according to GPA. Use the landmark file ‘lands’ and, if present, the curveslide object as input data for the `gpagen()` function of `geomorph`<sup>45,46</sup>. Assign the GPA results to a new object called ‘sup.geo’ and visualize them with the R graphics function `plot()` (see Supplementary Figs. 3–5).

```

```{r}
sup.geo <- gpagen(lands, curves=curveslide, ProcD=F, Proj = T)
plot(sup.geo$coord[,1,], sup.geo$coord[,2,], Type="n",
xlim = c(-0.6,0.6), ylim = c(-0.6,0.6),
asp = 1, main = "geomorph / bending energy")
for (i in 1:dim(sup.geo$data)[1]) {
points(sup.geo$coord[,i], col = g_color[i])
}
```

```

**▲ CRITICAL STEP** Semilandmarks can be slid along a curve by minimizing either Procrustes distance or bending energy. To use the Procrustes Distance approach, the argument ‘ProcD=’ has to be set to ‘T’ (or ‘TRUE’), while setting it to ‘F’ (or ‘FALSE’) will make `gpagen()` use the bending energy approach. The last argument of the function ‘Proj =’ specifies whether the aligned Procrustes coordinates should be projected into a linear tangent space. Since this projection facilitates the downstream multivariate statistical analysis<sup>53</sup>, set this argument to ‘TRUE’ (or ‘T’). The arguments ‘sup.geo\$coord[,1,]’ and ‘sup.geo\$coord[,2,]’ represent the x- and y-coordinates of the superimposed landmarks of all specimens. Use ‘xlim’ and ‘ylim’ to set the limits of the x and y axes, respectively. Define the aspect ratio in the ‘asp=1’ argument and add a graph title with ‘main’ argument. The `for()` loop in the code chunk gives all of the superimposed landmarks a color based on the previously defined color code (see Step 18).

**▲ CRITICAL STEP** Before performing the final steps of shape analysis (Steps 21–28), it is recommended to test the replicability of shapes, especially if new landmark configurations are tested or if the protocol is adapted to microscopic animals other than nematodes. To test the replicability, a separate data set containing landmarks of technical replicates of your specimen has to be generated (for a detailed example, see ‘Replicability and error’ in ‘Experimental design’). If you wish to check your replicability, perform Steps 1–20 on this separate replicability data set and run the code that is provided in Box 1 on the obtained shapes, to estimate their RPT or %ME. If the replicability of shapes is acceptable, one can obtain the actual landmark data set, perform Steps 1–20, and then proceed beyond this step to analyze shape differences (Steps 21–28).

**? TROUBLESHOOTING**

**PCA ● Timing 1-15 min**

- 21 Run a PCA on the Procrustes coordinates obtained from the GPA in Step 20. Use the `prcomp()` function that comes with the R stats package to perform the analysis and save the result to an object called ‘pca.geo’. Visualize the results of the analysis by generating a ggplot2-based<sup>71</sup> PCA plot with the `fviz_pca_ind()` function of the `factoextra` package<sup>58</sup> and create a bar plot depicting the amount of variation described by each PC (in percentages). Use convex hulls in the PCA plot to visualize the portions of the morphospace that are occupied by our groups of interest.

```

```{r}
#Principal component analysis (PCA) using a base R function
pca.geo <- prcomp(two.d.array(sup.geo$coord))
#ggplot2-based plot of the PCA results using a factoextra function
fviz_pca_ind(pca.geo,

```

```
geom.ind = "point", # show points without labels
col.ind = SP, #adjust to your group
palette = "npg", #color palette, adjust to preferences
addEllipses = T, #adds ellipses or convex hulls
ellipse.type = "convex", #convex hulls
mean.point = F, #don't show average point
legend.title = "SP" #adjust accordingly) + coord_fixed(ratio = 1)
#adjustable for variation-weighted plots
# Barplot of variation described (in %) by each principal component
barplot(rev(100*pca.geo$sdev^2/sum(pca.geo$sdev^2)),horiz=T,
yaxt="n",xlab="variation described, %",xlim=c(0,100))
````
```

**▲ CRITICAL STEP** The geomorph function `two.d.array()`<sup>45,46</sup> converts the 3D array of Procrustes landmark coordinates into a 2D matrix, which is a readable input class for the `prcomp()` function. Set the `'ellipse.type ='` argument to `'convex'` so that the groups will be enclosed by convex hulls. Alternatively, setting the argument to `'confidence'` will add confidence ellipses for each group (default level = 0.95). If necessary, adjust grouping factors according to your input data. Keep the `'ratio='` argument in the `coord_fixed()` function set to 1 to obtain PCA plots with equal unit length along the x and y axes. You can adjust the ratio according to the relative difference in variation that is explained by each PC, to obtain variation-weighted PCA plots (see Supplementary Fig. 12). The amount of variation described by each PC is calculated by dividing the squared standard deviations of each PC (`'pca.geo$sdev^2'`) by the sum of squared standard deviations of all PCs (`'sum(pca.geo$sdev^2)'`) and multiplying it by 100. The base R function `rev()` produces a reversed sequence of this argument, ordering the PCs from highest 'variance explained' to lowest.

#### ? TROUBLESHOOTING

- 22 Before proceeding with the analysis, check whether there are potential outliers in the data set by using the `plotOutliers()` function<sup>45,46</sup> from the geomorph package and plot deformation grids for potential outliers using geomorph's `plotRefToTarget()` function<sup>45,46</sup> to visually check for errors (e.g., incorrect landmark sequence).

```
````{r}
plotOutliers(sup.geo$coords) #generate an outlier plot
#if a potential outlier is found, compare it to the mean shape of the data
set
shape1 <- sup.geo$consensus #mean shape
shape2 <- sup.geo$coords[, ,52]
#adjust the value so that it fits the number that corresponds to the
outlier in "lands.txt"
#generate deformation grid of mean shape and potential outlier to check
for errors
plotRefToTarget(shape1, shape2, method = "TPS", mag = 2)
````
```

**▲ CRITICAL STEP** This function plots the Procrustes distance of each specimen to the mean shape. Specimens that exceed the upper quantile of distances are marked in red as potential outliers. The `'method='` argument of the function is set to `'TPS'`. The `'mag='` argument allows the user to define the degree of magnification that is desired when visualizing the shape differences. Here, we use a twofold magnification (`'mag=2'`). Note that `plotOutliers()` identifies only 'potential' outliers. These are not necessarily 'wrong'; nor do they need to be removed in every case<sup>46</sup>. If, however, erroneous specimens can be found, one can remove them manually from the landmark file (`'lands.txt'`) and repeat the analysis. We intentionally did not remove potential outliers in our example data sets, so that readers can identify such cases by using the example landmark files provided in the supplements (Supplementary Data 1–3). If one uses, for example, the file containing landmarks of different *Pristionchus* species (Supplementary Data 2), one can identify a single potential outlier in specimen 'Pbu-Eu-ind-21'. Checking the deformation grid of this individual and the mean shape indicates a potentially incorrect sequence of landmarks, as the grids folds over in the middle of the

plain. Such outliers might need to be removed from the data set, and the analysis might need to be repeated, if the potential outlier drastically affects PCA or downstream statistical analysis (see Step 26). Note that in our example, the significance of the respective result does not change regardless of whether these potential outliers are removed.

- 23 Identify important or meaningful PCs from the PCA results. Calculate the variation described by each PC and arrange the returned values from highest to lowest. Use the eigenvalues obtained during the PCA (i.e., the squared standard deviations of the PCs obtained with `prcomp()` in Step 21) and the number of individuals (in our example,  $n = 80$ ) to determine whether a given PC is qualified to be considered meaningful with the `getMeaningfulPCs()` function of the Morpho package<sup>47,48</sup>.

```
```{r}
#view how much variation is described by each PC (ordered from PC1 to last PC)
100*pca.geo$sdev^2/sum(pca.geo$sdev^2)
#get meaningful PCs
getMeaningfulPCs(pca.geo$sdev^2, n = 80)
#adjust sample size (n=) accordingly
```
```

**▲ CRITICAL STEP** This method calculates a threshold between a PC and its successor based on a log-likelihood ratio and the sample size<sup>72</sup>. Both the meaningful PCs and the threshold value are returned as outputs of the function. In our comparison of different *Pristionchus* species (including 80 animals), a PC is considered meaningful if its value is >1.37 times the value of the subsequent PC. This criterion applied only to the first three PCs (Fig. 4b; compare to Figs. 3b and 2d).

- 24 Estimate extreme shapes for PC1 and PC2 and visualize underlying shape differences along these principal components by drawing superimposed wireframe and lollipop plots (Figs. 2e,f, 3c,d and 4c,d).

```
```{r}
#estimate extreme shapes along PC1 and PC2
meanshape <- as.vector(t(sup.geo$consensus))
min_PC1 <- t(matrix(meanshape+(min(pca.geo$x[,1])*pca.geo$rotation
[,1]),2,18))
max_PC1 <- t(matrix(meanshape+(max(pca.geo$x[,1])*pca.geo$rotation
[,1]),2,18))
min_PC2 <- t(matrix(meanshape+(min(pca.geo$x[,2])*pca.geo$rotation
[,2]),2,18))
max_PC2 <- t(matrix(meanshape+(max(pca.geo$x[,2])*pca.geo$rotation
[,2]),2,18))
#generate wireframe plot of superimposed PC1 extreme shapes
plotRefToTarget(max_PC1, min_PC1, method = "point", mag = 1)
jointline <- c(1:2, NA, 3:4, NA, 5:10, NA, 11:13, NA, 15:16, NA, 17:18)
lines(min_PC1[jointline,], lty = 1, col = "darkorange", lwd = 0.75)
lines(max_PC1[jointline,], lty = 1, col = "darkgreen", lwd = 0.75)
#generate wireframe plot of superimposed PC2 extreme shapes
plotRefToTarget(max_PC2, min_PC2, method = "point", mag = 1)
jointline <- c(1:2, NA, 3:4, NA, 5:10, NA, 11:13, NA, 15:16, NA, 17:18)
lines(min_PC2[jointline,], lty = 1, col = "darkorange", lwd = 0.75)
lines(max_PC2[jointline,], lty = 1, col = "darkgreen", lwd = 0.75)
#generate lollipop plots of superimposed extreme shapes for PC1 and PC2
plotRefToTarget(max_PC1, min_PC1, method = "vector", mag = 1)
plotRefToTarget(max_PC2, min_PC2, method = "vector", mag = 1)
```
```

**▲ CRITICAL STEP** Color coding of wireframe lines can be changed according to personal preferences. Depending on your landmark configuration, the wireframe line connections may need

to be redefined by changing the vectors *c()*, which give the input to the ‘jointline’ objects. Here, ‘NA’ indicates stops between lines, while numbers separated by colons (e.g., ‘5:10’) defines lines connecting the corresponding landmarks 5–10. If necessary, change the line definitions so that the wireframes represent the structures in your landmark configurations appropriately. Here, the ‘jointline’ vectors define proper wireframes for the stomata of *Pristionchus nematodes* (Fig. 1).

### ? TROUBLESHOOTING

- 25 Use the `fviz_contrib()` function to estimate the contribution (percentage) of every landmark on PC1/PC2. Extract and investigate the landmarks that contribute more to these PCs than expected.

```
```{r}
#estimate contribution (%) of landmarks on PC1
PC1_contrib <- fviz_contrib(pca.geo, choice = "var", axes = 1)
LM_contrib_PC1 <- data.frame(LM = rep(NA, nrow(PC1_contrib$data)/2),
PC1_contrib = rep(NA, nrow(PC1_contrib$data)/2))
for (i in 1:(nrow(PC1_contrib$data)/2)) {
LM_contrib_PC1$PC1_contrib[i] <- sum(PC1_contrib$data[(i*2-1):
(i*2),]$contrib)
LM_contrib_PC1$LM[i] <- as.character(i)
}
#estimate contribution (%) of landmarks on PC2
PC2_contrib <- fviz_contrib(pca.geo, choice = "var", axes = 2)
LM_contrib_PC2 <- data.frame(LM = rep(NA, nrow(PC2_contrib$data)/2),
PC2_contrib = rep(NA, nrow(PC2_contrib$data)/2))
for (i in 1:(nrow(PC2_contrib$data)/2)) {
LM_contrib_PC2$PC2_contrib[i] <- sum(PC2_contrib$data[(i*2-1):
(i*2),]$contrib)
LM_contrib_PC2$LM[i] <- as.character(i)
}
#examine landmarks with major contribution (%) to PC1 and PC2
important_LMs_PC1 <- LM_contrib_PC1[LM_contrib_PC1$PC1_contrib > 100/
(nrow(PC1_contrib$data)/2),]
important_LMs_PC1[order(decreasing = T, important_LMs_PC1$PC1_con-
trib),]
important_LMs_PC2 <- LM_contrib_PC2[LM_contrib_PC2$PC2_contrib > 100/
(nrow(PC2_contrib$data)/2),]
important_LMs_PC2[order(decreasing = T, important_LMs_PC2$PC2_con-
trib),]
```
```

**▲ CRITICAL STEP** To estimate which landmarks contribute more than expected, a threshold is set as 100 divided by the number of landmarks. The value obtained is the contribution (percentage) the landmarks are expected to have if all of them contribute equally to the shape differences. Important landmarks are ordered decreasingly by the `order()` function of base R. These landmarks can subsequently be indicated in the lollipop plot (obtained in Step 24) to give an impression of which landmarks differ profoundly between groups in the context of superimposed extreme shapes.

### Multivariate statistical testing using ‘Procrustes ANOVA’ ● Timing ~5 min

- 26 Perform a permutational MANOVA<sup>54–56</sup> (or ‘Procrustes ANOVA’ in geometric morphometrics<sup>46,56</sup>) and—if required—a pairwise permutational MANOVA (including FDR correction<sup>73</sup> of *P* values) to assess statistical hypotheses describing patterns of variation based on the Euclidean distances obtained from GPA in Step 20. Define a formula for the linear model in the `procD.lm()` function according to your needs.

```
```{r}
#prepare a geomorph data frame for your statistical analysis
```

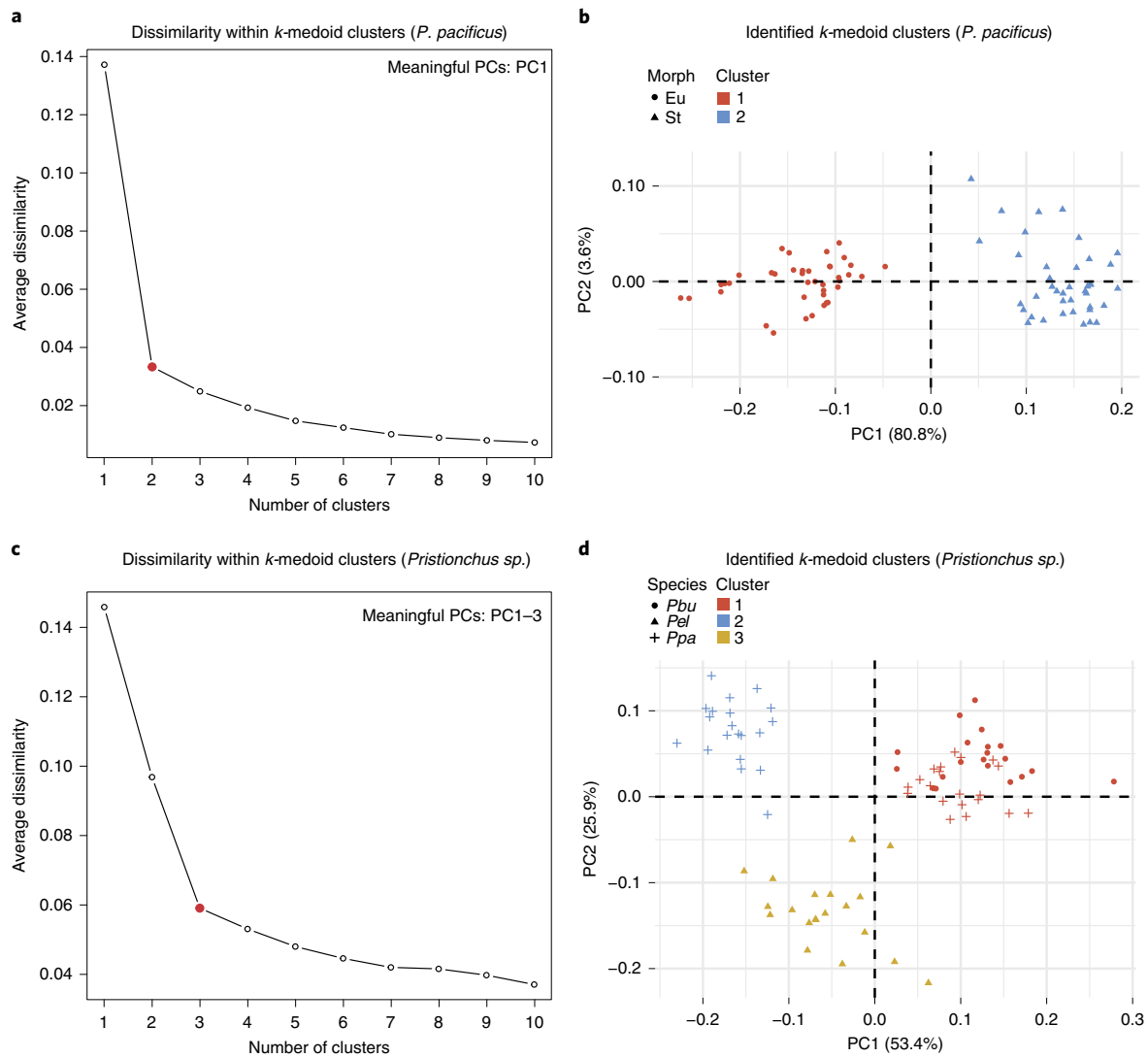
```

#adjust grouping factor according to interest
gdf <- geomorph.data.frame(shape = sup.geo$coords, species = SP,
mouth_form = MF)
#Procrustes Anova to test for shape differences between groups
#adjust grouping factor according to interest
a <- procD.lm(shape ~ species, data = gdf, iter = 100000, seed = 12345,
RRPP = F, effect.type = "F")
summary(a)
#pairwise Procrustes Anova
pw <- pairwise(a, groups = SP) #adjust grouping factor according to
interest
pw_summary <- summary(pw)
pw_summary
pw_p.values <- pw_summary$pairwise.tables$P
#transform the pairwise table into a data frame
#make sure to adjust row/column names and numbers accordingly
results_pw <- as.data.frame(matrix(pw_p.values,
dimnames = list(c("Pbu", "Pel", "Ppa_Eu", "Ppa_St"),
c("Pbu", "Pel", "Ppa_Eu", "Ppa_St")),
ncol = 4, nrow = 4))
#FDR-correction of p-values obtained from pairwise Procrustes Anova
adj.pw <- p.adjust(pw_summary$pairwise.tables$P, method = "fdr")
#transform the pairwise table into a data frame
#make sure to adjust row/column names and numbers accordingly
results_FDR.adj.pw <- as.data.frame(matrix(adj.pw,
dimnames = list(c("Pbu", "Pel", "Ppa_Eu", "Ppa_St"),
c("Pbu (FDR-adj.)", "Pel (FDR-adj.)", "Ppa_Eu (FDR-adj.)",
"Ppa_St (FDR-adj.)")),
ncol = 4, nrow = 4))
#view results of pairwise Procrustes Anova and FDR-corrected pairwise
Procrustes Anova
results_pw
results_FDR.adj.pw
...

```

**▲ CRITICAL STEP** To facilitate statistical analysis of the shape data, the array containing the Procrustes coordinates ('sup.geo\$coords') has to be converted into a list that can be used as a data frame. This is achieved by the `geomorph.data.frame()` function of `geomorph`<sup>45,46</sup>, if the array and the grouping factors are provided as arguments (e.g., 'shape = sup.geo\$coords' or 'species = SP'). Save the results to a new object called 'gdf'. This object serves as the input for `geomorph`'s `procD.lm()` function<sup>45,46</sup>, which performs the Procrustes ANOVA<sup>46</sup>. A formula for the linear model has to be provided as an argument (e.g., 'shape ~ species', which asks whether shapes differ between species), together with the `geomorph.data.frame` (i.e., 'data=gdf') and the number of permutations for significance testing (here: 100,000). The significance can be tested either by randomizing raw values (residuals of mean, 'RRPP=F') or by randomizing null model residuals ('RRPP=T')<sup>74,75</sup>. In our examples, we used raw value randomization. The results of the Procrustes ANOVA can be viewed by using the generic `summary()` function of base R. Subsequently, distributions of pairwise statistics for the linear model fit obtained from `procD.lm()` can be generated with the `pairwise()` function of the `RRPP` package<sup>74,75</sup>, after defining the grouping factor for the statistical test as an argument (e.g., 'groups = SP'). The results of the pairwise analysis are saved to a new object, and the *P* values are subsequently converted into and displayed as a data frame. Finally, the *P* values obtained by pairwise comparisons are corrected for false discoveries<sup>73</sup> using the `p.adjust()` function that is part of the R stats package. The obtained FDR-corrected *P* values (or 'q-values') are also formatted into a simple data frame. Results can be shown inside the corresponding PCA plot of the analysis (Figs. 2c, 3a and 4a). Make sure to adjust the names and the number of all rows and columns in the data frames according to your input data set.

#### ? TROUBLESHOOTING



**Fig. 5 | k-medoid clustering performed on the two *Pristionchus* datasets.** **a**, Saturation plot showing the change of average dissimilarity within each cluster in relation to the number of clusters ( $k$ ) for the *P. pacificus* dataset (including both morphs). The red dot indicates the best number of medoids for  $k$ -medoid clustering. **b**, Identified  $k$ -medoid clusters for the *P. pacificus* polyphenism dataset. Eu, eurytomatous morph; St, stenostomatous morph. **c**, Saturation plot showing the change of average dissimilarity within each cluster in relation to the number of clusters ( $k$ ) for the *Pristionchus* dataset (including different species). The red dot indicates the best number of medoids for  $k$ -medoid clustering. **d**, Identified  $k$ -medoid clusters for the *Pristionchus* polyphenism and species dataset. *Pbu*, *P. bucculentus*; *Pel*, *P. elegans*; *Ppa*, *P. pacificus*.

**Group identification by clustering**

27 Perform group identification by clustering. For partitional clustering around  $k$ -medoids, follow option A. For model-based clustering, follow option B.

**(A) Partitional clustering around  $k$ -medoids** ● **Timing ~2 min**

- (i) Identify the best number of medoids for the input data set by performing partitioning (i.e., clustering) around the medoids with the `pam()` function of the `cluster` package<sup>60</sup>. This calculates the average dissimilarity within the obtained clusters, depending on a number of clusters between 1 and 10. Generate a plot depicting the change in average dissimilarity of obtained clusters in relation to the number of medoids ( $k$ ) (Fig. 5a,c; for simplicity, clustering was performed only for *Pristionchus* data sets).

```

```{r}
saturation <- numeric()
for (i in 1:10) {

```

```
saturation[i] <- mean(pam(pca.geo$x[,1:(ncol(pca.geo$x)-33)],
i) [[6]][,3])
#remove non-meaningful PCs by adjusting the value behind the "-"
}
plot(1:10, saturation, xlab="number of clusters", ylab="average
dissimilarity", Type="b")
```

```

**▲ CRITICAL STEP** You may want to remove all non-meaningful PCs from your input matrix by adjusting the value behind the ‘-’. For example, we obtained only three meaningful PCs (i.e., PC1–3) from our stomatal shape comparison across several *Pristionchus* species (see Step 23). Since we had 36 PCs in total, we removed the 33 PCs that are considered non-meaningful (i.e., ‘pca.geo\$x[,1:(ncol(pca.geo\$x)-33)]’). This procedure is optional, and one can lose information by PC removal. If no PCs are to be removed, just delete the term ‘-XX’ with XX being the number of PCs.

- (ii) Choose the number of medoids/clusters based on the dissimilarity plot obtained previously and save this input to a new object called ‘best.n.kmedoid.clusters’. The best number of clusters corresponds to the point in the ‘saturation plot’ where the average dissimilarity curve starts to level out (i.e., the point where partitioning of the data set into more clusters barely changes the average dissimilarity within the clusters).

Afterwards, perform *k*-medoid clustering with the pam() function<sup>60</sup> for the chosen number of medoids.

```
```{r}
best.n.kmedoid.clusters <- 3 #set number of clusters here
pca.kmedoid.clusters <- pam(pca.geo$x[,1:(ncol(pca.geo$x)-33)],
best.n.kmedoid.clusters)
#adjust number of PCs you want to remove (exclude at least 4)
```
```

**▲ CRITICAL STEP** Again, non-meaningful PCs can be removed from your input matrix by adjusting the value behind the ‘-’. However, if all PCs should be taken into account, delete the term ‘-XX’ (with XX being the number of PCs).

- (iii) Generate a new PCA plot in which all observations (i.e., specimens) are colored according to the newly identified clusters (Fig. 5b,d).

```
```{r}
plot(pca.geo$x[,1:2], main = "k-medoid clustering",
pch = g_pch, bg = (1:best.n.kmedoid.clusters)[pca.kmedoid.
clusters$clustering], asp=1)
```
```

**(B) Model-based clustering ● Timing ~ 2 min**

- (i) Perform model-based clustering on the PCA data by using the Mclust() function that comes with the mclust package<sup>61,62</sup>.

```
```{r}
pca.model.clusters <- Mclust(pca.geo$x[,1:(ncol(pca.geo$x) -33)])
#adjust number of non-meaningful PCs to remove
```
```

**▲ CRITICAL STEP** A high number of variables could over-parameterize multivariate normal models. Thus, we recommend removing all non-meaningful PCs from your input matrix by adjusting the value behind the ‘-’, depending on the outcome of Step 23. Always exclude the worst four PCs at the minimum, because four degrees of freedom are lost during GPA.

- (ii) Extract the best number of clusters (i.e., the optimal number of mixture components) estimated in the previous step.



```

```{r}
best.n.model.clusters <- pca.model.clusters$parameters$variance$G
```

```

- (iii) Finally, plot the PCA with observations (i.e., specimens) colored according to newly identified clusters (see Supplementary Fig. 8 for clustering results based on the *Pristionchus* example data sets and two different rules for utilization of PCs).

```

```{r}
plot(pca.geo$x[,1:2], main="model-based clustering", asp = 1,
     pch = g_pch, bg = (1:best.n.model.clusters)[pca.model.clusters
     $classification])
```

```

### Visualize allometry ● Timing ~5 min

- 28 Visualize group-specific allometric trends in your data set. Prepare a list called ‘allometry’ that contains the original shape variables, your grouping factors and the centroid sizes of all specimens before GPA alignment with the help of the `geomorph.data.frame()` function. Use the `procD.lm()` function to fit a linear model for the correlation of shape with size (i.e., allometry) as well as other covariates (e.g., species) and assess the statistical significance with a distance-based PERMANOVA design including a residual randomization permutation procedure with  $\geq 10,000$  iterations. Generate a model-based allometry plot with `geomorph`’s `plotAllometry()` function.

```

```{r}
#prepare geomorph dataframe for permutation test
#adjust according to your grouping factor and grouping factor levels
allometry <- geomorph.data.frame(shape = sup.geo$coords, species = SP,
                                mouth_form = MF, size = sup.geo$Csize)
#permutation test based on linear model
allometry_model <- procD.lm(data=allometry, shape ~ size * SP,
                             iter=100000) #adjust the linear model according to your grouping
#summary(allometry_model)
#plot allometric trends (use "PredLine" or "RegScore")
#adjust symbols (pch=) and color (col=) according to preferences
plotAllometry(allometry_model, size = allometry$size, logsz = T,
              method = "PredLine", pch = g_pch, col = g_color)
```

```

▲ **CRITICAL STEP** Please note that, in the most recent `geomorph` release (v. 3.2.1), the `plotAllometry()` function seems to be unable to visualize allometric trends of multiple groups in the same data sets as separate lines in one plot. Copy the overhauled source code of the function that is currently provided by the package authors (see ‘Software for image acquisition and data analysis’ in ‘Materials’) and run it once in the R console to overwrite the initial version before running the code chunk above. We recommend using 100,000 rather than 10,000 iterations in the permutation procedure. The ‘size=’ argument in `plotAllometry()` defines the input that contains the centroid sizes of all individuals (before GPA alignment), while setting the ‘logsz=’ argument to ‘T’ (or ‘TRUE’) will result in plotting the logarithm of the centroid size on the x axis. The ‘method=’ argument specifies the plotting procedure, which has to be set to ‘PredLine’ to plot PC1 on the predicted values (obtained from the linear regression of shape versus size) on the y axis. This argument can be changed to ‘RegScore’ if one wishes to use a different model-based visualization of allometry.

### ? TROUBLESHOOTING



Troubleshooting

For troubleshooting guidance, see Table 1.

**Table 1 | Troubleshooting table**

| Step | Problem                                                                                                                                              | Possible reason                                                                                                             | Solution                                                                                                                                                                                                                                                                                                      |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 4    | Nematodes are still moving on the object slide                                                                                                       | The NaN <sub>3</sub> concentration in agar pads is too low                                                                  | 1 Simply wait another 5–10 min<br>2 Kill nematodes by placing the object slide on a heat block for a maximum of 3–4 s at -70 °C (note that the integrity of inner organs might be affected by this procedure)                                                                                                 |
| 12   | The readlands.tps() function gives an error message and reports that curve points are detected but that numbers vary among specimens                 | The number of landmarks defined by 'LM=' is not equal for all specimens in the data set, and/or coordinates contain commas  | Ensure that you have the same number of landmarks (in our example, 18) for all specimens in the landmark file and replace all commas in coordinate values with periods                                                                                                                                        |
| 20   | The gpagen() function gives an error message and reports that 'x' must be an array of at least two dimensions                                        | The number of semilandmarks or curves is too small                                                                          | Include at least two semilandmarks on a single curve or have at least two curves with one semilandmark each (more curves and/or semilandmarks are always possible)                                                                                                                                            |
| 21   | Semi-transparent filling of convex hulls is not included in the (exported) PCA plot generated by fviz_PCA_ind()                                      | The image was exported in file formats like .png or .jpg                                                                    | Export all plots as PDF and use a vector graphics editor like Adobe Illustrator for postediting                                                                                                                                                                                                               |
| 24   | Parts of the wireframe or lollipop plots are outside the plot                                                                                        | The graphical window is not set appropriately                                                                               | Change the sequence of the 'max_PC1/2' and 'min_PC1/2' arguments in the plotReftoTarget() functions that return the wireframe and lollipop plots                                                                                                                                                              |
| 26   | Pairwise comparisons yield results that clearly do not fit the observations in PCA                                                                   | The data frame that displays the results of the pairwise comparison is set up incorrectly                                   | Open the object 'pw_p.values' generated in the step that includes the pairwise() function and adopt the row and column names in their correct sequence into the functions that set up the data frames that contain P values from pairwise comparisons (normal and FDR adjusted)                               |
| 28   | Some members of a group are located far away from other members of the same group along the log(size) axis (see an example in Supplementary Fig. 9b) | Image stacks have been saved with an inconsistent scaling factor (e.g., some in micrometers versus some in inches)          | If you save all image stacks with identical x-/y-dimensions, you can manually calculate a re-scaling factor based on these dimensions and the scale unit. Multiply affected centroid sizes (see object 'sup.geo\$Czise') with that factor and create a new input dataframe that contains the corrected values |
|      | Group-specific allometries are not separated as individual trend lines in the allometry plot                                                         | 'bug' in the plotAllometry() function of the most recent geomorph release (v. 3.2.1)                                        | Overwrite the inbuilt plotAllometry() function with the updated version (see 'Software for image acquisition and data analysis')                                                                                                                                                                              |
|      | The procD.lm() function gives an error message and reports 'in fits\$full[[k]]: subscript out of bounds'                                             | The procD.lm() function of the most recent geomorph release (v. 3.2.1) potentially assumes a fully crossed factorial design | Write out the model formula using only '+' between factors and their interactions (instead of using '*'); try to identify factor interactions that equate to zero and delete them from the formula                                                                                                            |

Timing

Most of the time that goes into data analysis depends on the amount of troubleshooting and the number of adjustments that are needed to fit the analyses to your grouping factors of interest. The timeframe for replicability assessments (Box 1; or see the end of the R code in Supplementary Data 4 and 5) depends on the design of the study. If one follows the approach used herein (20 nematode specimens in three replicates on different days), then the according landmark files are obtained over a course of 5 d. Their analysis takes ~20–30 min per annotator.

Steps 1–5, acquisition of raw image stacks for 80 animals: ~4 d (for nematodes)

Steps 6–8, obtaining landmark data for 80 animals: ~6 h

Steps 9–19, setting-up the data in an R markdown file: 1–10 min

Step 20, generalized Procrustes analysis (i.e., superimposition): ~1 min

Steps 21–25, PCA: 1–15 min

Step 26, multivariate statistics (Procrustes ANOVA): ~5 min

Step 27A, *k*-medoid clustering: ~2 min  
Step 27B, model-based clustering: ~2 min  
Step 28, visualization of allometric trends: ~5 min  
Box 1, Replicability assessment: 5 d

## Anticipated results

The combination of geometric morphometric analysis and clustering methods allows quantification and statistical testing of morphological differences between groups of interest as well as an estimation of the number of separate groups that are present in the morphospace. In an appropriate context, these differences can be investigated as morphological shape changes due to, for example, divergent evolution (phylogeny), embryonic development (ontogeny), environmental inputs (ecomorphology) or mutagenesis (functional genetics). Therefore, the expected results from the application of the presented protocol will depend on the biological system under investigation, as well as on the context of the study.

Researchers who acquire landmark data (Steps 1–8) using one of the landmark configurations we proposed above (Figs. 1 and 2a,b) will be able to robustly quantify shape differences between adults of many species or strains within the nematode genera *Pristionchus* or *Caenorhabditis* in any of the respective study contexts mentioned above. With minor adjustment and a suitable landmark configuration, this protocol should allow similar shape quantification in any microscopic animal or structure of choice. However, it might be necessary to practice the correct placement of landmarks over a few days before obtaining the final landmark data set. Even when RPT and %ME are acceptable to obtain reproducible and significant results, it is clear that both estimates can be substantially affected by the level of experience of the annotator (Supplementary Figs. 6 and 7; compare annotators 1 and 2 with 3). Importantly, if the variation within technical replicates is higher than the variation across different individuals, it might be difficult to obtain meaningful results from geometric morphometric analyses. Besides practicing landmark placements, one could also label each specimen multiple times and perform GPA on the average landmark configurations of each individual, in order to reduce the measurement error in a data set<sup>67</sup>.

In the following sections, we provide a more detailed description of exemplary results in nematode mouth structures. First, we describe stomatal shape differences that were quantified for three species of *Pristionchus* nematodes using GPA, PCA, multivariate statistics and *k*-medoid clustering. We then proceed to describe the patterns of stomatal shape variation we identified for two species of *Caenorhabditis* nematodes. These conclusions are based on the lollipop and wireframe plots generated for (estimated) extreme shapes along axes of major variation. In addition, we explore which landmarks in the *Caenorhabditis* data set have a major contribution to PC1 and PC2 by assessing variable loadings obtained from PCA. We close by shortly discussing stomatal shape differences between the two adult morphs of *P. pacificus* in relation to changes in size (i.e., allometry).

### Quantification of shape differences in nematode stomata

As can be seen from our example data sets, our protocol can be used to quantify stomatal shape differences on the intraspecific level (Fig. 3) and the interspecific level (Figs. 2 and 4). The PCA analysis, together with the multivariate Procrustes ANOVA approach, allows statistical assessment of shape differences. Our analysis of different *Pristionchus* species, for example, reveals that the eury stomatous morph of *P. pacificus* overlaps with the monomorphic species *P. bucculentus* in the shape space (Fig. 4a). On the other hand, these two groups do not overlap either with the stenostomatous morph of *P. pacificus* or with *P. elegans*. This pattern of shape variation indicates that the shapes of *P. bucculentus* and eury stomatous *P. pacificus* are similar (if not identical), while stenostomatous *P. pacificus* and *P. elegans* occupy very different parts of the morphospace. This is particularly interesting in the context of canalization and morphological evolution, as *P. bucculentus* and *P. elegans* are described as monomorphic eury stomatous or monomorphic stenostomatous species, respectively<sup>76,77</sup>. The results of our exemplary analysis thus indicate that eury stomatous stomatal shapes can be similar across species, while stenostomatous morphs can indeed differ in terms of shape. This is also supported by the results of the Procrustes ANOVA on shape across morphs, as it estimates high effect sizes and low *P* values for all pairwise comparisons, except for the comparison of eury stomatous *P. pacificus* and *P. bucculentus* (Fig. 4a), which indicates that there is no significant shape difference between them. On top of that, *k*-medoid clustering provides supporting evidence for this assumption as well. We estimated the optimal number of clusters in this data set to be three

(Fig. 5c), and subsequent partitioning of data points around three medoids reveals a group structure in which all eury stomatous *P. pacificus* and *P. bucculentus* specimens form a common cluster (Fig. 5d). Furthermore, specimens of stenostomatous *P. pacificus* and those of *P. elegans* each form a distinct, non-overlapping cluster. Thus, the group structure revealed by *k*-medoid clustering (without an a priori hypothesis) perfectly complements the results obtained via multivariate Procrustes ANOVA performed on groups that were defined a priori (compare Fig. 4a with Fig. 5d).

### Visualizing and interpreting the quantified stomatal shape differences

Morphological variation can be described by depicting wireframe and lollipop plots generated for superimposed extreme shapes along PC1 and PC2 (Fig. 3c,d; Fig. 4c,d; and Fig. 2e,f). For example, if one considers the wireframe plots obtained for the stomatal shape differences between *C. elegans* and *C. briggsae* (Fig. 2e), it appears that along PC1 the entire buccal cavity is wider (on average) in *C. elegans* than in *C. briggsae*, while the relative length of the cheilo- and gymno/promesostegostom remains similar across species. Another surprising observation is that along PC1 the metastegostomal flaps are (on average) larger in *C. elegans* than in *C. briggsae*. The corresponding lollipop plots (Fig. 2f) can then be used to describe the change in landmark position (relative to all other landmarks) along PC1/2. In these plots, landmarks 2, 3, 16 and 17 move interiorly in a nearly perpendicular manner to the anteroposterior body axis in *C. briggsae*. Therefore, while the overall length of the buccal cavity is maintained, its lumen is constricted in *C. briggsae* as compared to *C. elegans* (Fig. 2f). It is conceivable that these differences have functional consequences related to bacterial diet, thus demonstrating the power of GM and our simplified protocol to provide new testable hypotheses for morphological evolution of microscopic organisms.

When interpreting such data, one might be tempted to deduce where the shape changes originate by looking at individual landmarks. However, it is important to properly interpret these results as follows. If we look at the loading-based estimation of relative contribution, we can see that landmarks 4 and 15, which are the main predictors of stomatal width, have a major influence on PC1 (Fig. 2f). This correlates with the separation of the two species along PC1 (Fig. 2c). On the other hand, the shape of the telostegostom (landmarks 8–10) is very similar across species along both PC1 and PC2 (Fig. 2e). Thus, landmarks 8–10 do not contribute substantially to the first two components of shape variation (Fig. 2f). However, we note that (due to the nature of GPA) differences in shapes always result from shifts of a given landmark relative to all other landmarks in the configuration. Furthermore, shape differences do not originate from landmarks themselves, but within the tissues on which these landmarks are placed<sup>50</sup>. Thus, landmarks that contribute more to the distribution of specimens in the tangent space than expected exclusively point to (i) the portion of the overall shape that differs most prominently between groups and (ii) the tissue(s) that might cause these morphological differences. They do not indicate contributions independent of other landmarks. Here, it appears that essentially all tissues that secrete the stoma are involved in causing the shape change, as shape differences are obvious in all stomatal structures (i.e., cheilostom, gymnostom and stegostom) of *C. briggsae* and *C. elegans*.

### Identification of size-related changes in stomatal shape

Lastly, we can visualize and describe group-specific allometries in our example data sets. By plotting predicted shapes against the logarithm of the centroid size, we can obtain stylized allometric trend lines for different groups. Allometry strongly influences shape differences in many morphometric data sets, and it often closely aligns with PC1 of the tangent space<sup>70</sup>. Thus, we might want to know whether the large shape differences between the two morphs in *P. pacificus* (Fig. 3a) are mere manifestations of allometry. In our case, allometry might be a good null hypothesis, especially since the respective data set is nearly one dimensional, with PC1 describing 80.8% variation (Fig. 3a,b). However, the allometric trend lines obtained for the two morphs indicate a contrasting situation. While there is slight allometry within each morph, they do not share a common allometric trajectory (Supplementary Fig. 9a). Simply put: it is impossible to obtain a eury stomatous mouth by just increasing the size of a stenostomatous one. Thus, the shape differences between them are not just allometric (Supplementary Fig. 9a). This observation has important implications regarding the genetic mechanisms regulating adult stomatal polyphenism. Assessing allometry might also be important when comparing wildtype and mutant phenotypes, because it could reveal whether mutant phenotypes represent different shapes or just mutation-induced allometric scaling effects.

### Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

All data generated and analyzed in this study are included in the paper or its Supplementary Data 1–3. Upon request, the raw data that were used to generate the example results are available from the corresponding authors.

### Code availability

The entire code for an analysis based on geomorph can be copied from the different steps of the procedure. In addition, two ready-to-use R markdown files that contain the code for both geometric morphometrics packages (geomorph and Morpho) and a R routine that allows the generation of landmark files can be downloaded from the supplementary material of this paper (Supplementary Data 4–6).

## References

1. Pigliucci, M. *Phenotypic Plasticity: Beyond Nature and Nurture* (Johns Hopkins University Press, 2001).
2. West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford University Press, 2003).
3. Shubin, N., Tabin, C. & Carroll, S. Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818 (2009).
4. Wagner, G. P. *Homology, Genes, and Evolutionary Innovation* (Princeton University Press, 2014).
5. Moczek, A. P. et al. The significance and scope of evolutionary developmental biology: a vision for the 21st century. *Evol. Dev.* **17**, 198–219 (2015).
6. Minelli, A. Grand challenges in evolutionary developmental biology. *Front. Eco. Evol.* **2**, 1–11 (2015).
7. Mallarino, R. & Abzhanov, A. Paths less traveled: evo-devo approaches to investigating animal morphological evolution. *Annu. Rev. Cell Dev. Biol.* **28**, 743–763 (2012).
8. Klingenberg, C. P. Evolution and development of shape: integrating quantitative approaches. *Nat. Rev. Genet.* **11**, 623–635 (2010).
9. Parsons, K. J. & Albertson, R. C. Unifying and generalizing the two strands of evo-devo. *Trends Ecol. Evol.* **28**, 584–591 (2013).
10. Andersen, E. C. et al. A powerful new quantitative genetics platform, combining *Caenorhabditis elegans* high-throughput fitness assays with a large collection of recombinant strains. *G3 (Bethesda)* **5**, 911–920 (2015).
11. Seydoux, G. & Fire, A. Whole-mount in situ hybridization for the detection of RNA in *Caenorhabditis elegans* embryos. *Methods Cell Biol.* **48**, 323–337 (2015).
12. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
13. Friedland, A. E. et al. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat. Methods* **10**, 741–743 (2013).
14. Witte, H. et al. Gene inactivation using the CRISPR/Cas9 system in the nematode *Pristionchus pacificus*. *Dev. Genes Evol.* **225**, 55–62 (2015).
15. Au, V. et al. CRISPR/Cas9 methodology for the generation of knockout deletions in *Caenorhabditis elegans*. *G3 (Bethesda)* **9**, 135–144 (2019).
16. Yoshida, K. et al. Two new species of *Pristionchus* (Nematoda: Diplogastridae) from Taiwan and the definition of the *pacificus* species-complex sensu stricto. *J. Nematol.* **50**, 355–368 (2018).
17. Susoy, V., Ragsdale, E. J., Kanzaki, N. & Sommer, R. J. Rapid diversification associated with a macro-evolutionary pulse of developmental plasticity. *Elife* **4**, e05463 (2015).
18. Susoy, V. et al. Large-scale diversification without genetic isolation in nematode symbionts of figs. *Sci. Adv.* **2**, e1501031 (2016).
19. 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).
20. Sieriebriennikov, B. et al. Conserved nuclear hormone receptors controlling a novel trait target fast-evolving genes expressed in a single cell. *PLoS Genet* **16**, e1008687 (2020).
21. Hong, R. L. & Sommer, R. J. *Pristionchus pacificus*: a well-rounded nematode. *Bioessays* **28**, 651–659 (2006).
22. Sommer, R. J. & McGaughan, A. The nematode *Pristionchus pacificus* as a model system for integrative studies in evolutionary biology. *Mol. Ecol.* **22**, 2380–2393 (2013).
23. Sommer, R. J. *Pristionchus pacificus: A Nematode Model for Comparative and Evolutionary Biology* (Brill, 2015).
24. Sommer, R. J. et al. The genetics of phenotypic plasticity in nematode feeding structures. *Open Biol.* **7**, 160332 (2017).
25. De Ley, P., Van De Velde, M. C., 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* **41**, 153–182 (1995).

26. Von Lieven, A. F. & Sudhaus, W. Comparative and functional morphology of the buccal cavity of Diplogastrina (Nematoda) and a first outline of the phylogeny of this taxon. *J. Zool. Syst. Evol. Res.* **38**, 37–63 (2000).
27. Jay Burr, A. & Baldwin, J. G. The nematode stoma: homology of cell architecture with improved understanding by confocal microscopy of labeled cell boundaries. *J. Morphol.* **277**, 1168–1186 (2016).
28. Wilecki, M., Lightfoot, J. W., Susoy, V. & Sommer, R. J. Predatory feeding behaviour in *Pristionchus* nematodes is dependent on phenotypic plasticity and induced by serotonin. *J. Exp. Biol.* **218**, 1306–1313 (2015).
29. Ragsdale, E. J., Müller, M. R., Rödelsperger, C. & Sommer, R. J. A developmental switch coupled to the evolution of plasticity acts through a sulfatase. *Cell* **155**, 922–933 (2013).
30. 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).
31. Namdeo, S. et al. Two independent sulfation processes regulate mouth-form plasticity in the nematode *Pristionchus pacificus*. *Development* **145**, dev166272 (2018).
32. Moreno, E., Lightfoot, J. W., Lenuzzi, M. & Sommer, R. J. Cilia drive developmental plasticity and are essential for efficient prey detection in predatory nematodes. *Proc. Biol. Sci.* **286**, 20191089 (2019).
33. Bardua, C., Wilkinson, M., Gower, D. J., Sherratt, E. & Goswami, A. Morphological evolution and modularity of the caecilian skull. *BMC Evol. Biol.* **19**, 30 (2019).
34. Tatsuta, H., Takahashi, K. H. & Sakamaki, Y. Geometric morphometrics in entomology: basics and applications. *Entomol. Sci.* **21**, 164–184 (2018).
35. Siritwut, W., Edgecombe, G. D., Sutcharit, C. & Panha, S. The centipede genus *Scolopendra* in mainland Southeast Asia: molecular phylogenetics, geometric morphometrics and external morphology as tools for species delimitation. *PLoS One* **10**, e0135355 (2015).
36. Zelditch, M. L., Swiderski, D. L. & Sheets, H. D. *Geometric Morphometrics for Biologists: A Primer* (Academic Press, 2004).
37. Adams, D. C., Rohlf, F. J. & Slice, D. E. Geometric morphometrics: ten years of progress following the ‘revolution’. *Ital. J. Zool.* **71**, 5–16 (2004).
38. Webster, M. & Sheets, H. D. A practical introduction to landmark-based geometric morphometrics. *Paleontological Soc. Pap.* **16**, 163–188 (2010).
39. Adams, D. C., Rohlf, F. J. & Slice, D. E. A field comes of age: geometric morphometrics in the 21st century. *Hystrix It. J. Mamm.* **24**, 7–14 (2013).
40. Collyer, M. L. & Adams, D. C. Phenotypic trajectory analysis: comparison of shape change patterns in evolution and ecology. *Hystrix It. J. Mamm.* **24**, 75–83 (2013).
41. Feilich, K. L. & López-Fernández, H. When does form reflect function? Acknowledging and supporting ecomorphological assumptions. *Integr. Comp. Biol.* **59**, 358–370 (2019).
42. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
43. Claude, J. *Morphometrics with R* (Springer, 2008).
44. Claude, J. Log-shape ratios, Procrustes superimposition, elliptic Fourier analysis: three worked examples in *R*. *Hystrix It. J. Mamm.* **24**, 94–102 (2013).
45. Adams, D. C. & Otárola-Castillo, E. geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods Ecol. Evol.* **4**, 393–399 (2013).
46. Adams, D. C., Collyer, M. & Kaliontzopoulou, A. Geomorph: Software for geometric morphometric analyses. R package version 3.2.1. <https://cran.r-project.org/package=geomorph> (2020).
47. Schlager, S. 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. & Székely, G.) 217–256 (Academic Press, 2017).
48. Schlager, S. Morpho: calculations and visualisations related to geometric morphometrics. R package version 2.8. <https://rdrr.io/cran/Morpho/> (2020).
49. Gunz, P. & Mitteroecker, P. Semilandmarks: a method for quantifying curves and surfaces. *Hystrix It. J. Mamm.* **24**, 103–109 (2013).
50. Klingenberg, C. P. Visualizations in geometric morphometrics: how to read and how to make graphs showing shape changes. *Hystrix It. J. Mamm.* **24**, 15–24 (2013).
51. Bookstein, F. L. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 567–585 (1989).
52. Rohlf, F. J. Shape statistics: Procrustes superimpositions and tangent spaces. *J. Classif.* **16**, 197–223 (1999).
53. Dryden, I. L. & Mardia, K. V. Multivariate shape analysis. *Sankhya* **55**, 460–480 (1993).
54. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).
55. Anderson, M. J. 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> (2014).
56. Goodall, C. Procrustes methods in the statistical analysis of shape. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **53**, 285–321 (1991).
57. Abdi, H. & Williams, L. J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2**, 433–459 (2010).



58. Kassambara, A. & Mundt, F. Factoextra: extract and visualize the results of multivariate data analyses. R package version 1.0.7. <https://cran.r-project.org/web/packages/factoextra/index.html> (2020).
59. Jin, X. & Han, J. K-medoids clustering. in *Encyclopedia of Machine Learning and Data Mining* (eds. Sammut, C. & Webb, G. I.) 697–700 (Springer, 2017).
60. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. Cluster: cluster analysis basics and extensions. R package version 2.1.0. <https://cran.r-project.org/web/packages/cluster/index.html> (2019).
61. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R. J.* **8**, 289–317 (2016).
62. Fraley, C., Raftery, A. & Scrucca, L. mclust: normal mixture modeling for model-based clustering, classification, and density estimation. R package version 5.4.6. <https://cran.r-project.org/web/packages/mclust/index.html> (2020).
63. Stiernagle, T. Maintenance of *C. elegans*. In *WormBook* (eds. The *C. elegans* research community, WormBook). <http://www.wormbook.org> (2006).
64. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676 (2012).
65. Stoffel, M. A., Nakagawa, S. & Schielzeth, H. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol. Evol.* **8**, 1639–1644 (2017).
66. Schielzeth, H. & Nakagawa, S. rptR: Repeatability for Gaussian and non-Gaussian data. R package version 0.9.22. <https://cran.r-project.org/web/packages/rptR/index.html> (2019).
67. Yezerinac, S. M., Loughheed, S. C. & Handford, P. Measurement error and morphometric studies: statistical power and observer experience. *Syst. Biol.* **41**, 471–482 (1992).
68. Adams, D. C. & Nistri, A. Ontogenetic convergence and evolution of foot morphology in European cave salamanders (Family: Plethodontidae). *BMC Evol. Biol.* **10**, 216 (2010).
69. Esquerré, D., Sherratt, E. & Keogh, J. S. Evolution of extreme ontogenetic allometric diversity and heterochrony in pythons, a clade of giant and dwarf snakes. *Evolution* **71**, 2829–2844 (2017).
70. Mitteroecker, P. et al. A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. *Hystrix It. J. Mamm.* **24**, 59–66 (2013).
71. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
72. Bookstein, F. L. *Measuring and Reasoning: Numerical Inference in the Sciences* (Cambridge University Press, 2014).
73. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289–300 (1995).
74. Collyer, M. L. & Adams, D. C. RRPP: An r package for fitting linear models to high-dimensional data using residual randomization. *Methods Ecol. Evol.* **9**, 1772–1779 (2018).
75. Collyer, M. & Adams, D. RRPP: linear model evaluation with randomized residuals in a permutation procedure. R package version 0.5.0. <https://cran.r-project.org/web/packages/RRPP/index.html> (2019).
76. Kanzaki, N. et al. *Pristionchus bucculentus* n. sp. (Rhabditida: Diplogastriidae) isolated from a shining mushroom beetle (Coleoptera: Scaphidiidae) in Hokkaido, Japan. *J. Nematol.* **45**, 78–86 (2013).
77. Kanzaki, N. et al. Two new species of *Pristionchus* (Rhabditida: Diplogastriidae): *P. fissidentatus* n. sp. from Nepal and La Réunion Island and *P. elegans* n. sp. from Japan. *J. Nematol.* **44**, 80–91 (2012).

## Acknowledgements

J. Claude (Institut des Sciences de l'Évolution-Montpellier, Montpellier, France) is gratefully acknowledged for his advice on geometric morphometrics using geomorph and Morpho in RStudio during a workshop given in Calgary (attended by B.S. in 2017). Earlier drafts of this manuscript were improved by advice on many steps of data analysis by F. Chan (Friedrich Miescher Laboratory, Tübingen, Germany) and critical comments on the general methodology of geometric morphometrics by P. Arnold (University of Potsdam, Potsdam, Germany). Furthermore, we would like to thank A. Rohrlach (Max Planck Institute for the Science of Human History, Jena, Germany) for his comments on statistical procedures and multivariate hypothesis testing using R. N. Kanzaki (Forestry and Forest Products Research Institute, Tsukuba, Japan) provided expertise on nematode stomatal morphology in an early phase of this project. Further thanks go to M. Riebesell for providing the TEM image of the *P. pacificus* stoma (Supplementary Fig. 1a). D. Sharma and J. de la Cuesta are to be thanked for general discussion. This work was funded by the Max Planck Society. T.T. was supported by the IMPRS 'From Molecules to Organisms'.

## Author contributions

Conceptualization: T.T., B.S. and M.S.W. Coding: B.S. and T.T. Data acquisition: T.T., S.S.W. and M.S.W. Data analysis: T.T. Writing the original draft: T.T. and S.S.W. Reviewing and editing the initial draft: B.S., M.S.W. and R.J.S. Revision of the original manuscript: T.T., B.S., S.S.W., M.S.W. and R.J.S. Supervision: M.S.W. and R.J.S.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41596-020-0347-z>.

**Correspondence and requests for materials** should be addressed to M.S.W. or R.J.S.

**Peer review information** *Nature Protocols* thanks Ming Bai and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

Received: 24 January 2020; Accepted: 27 April 2020;

Published online: 6 July 2020

#### Related links

##### Key references using earlier versions of this protocol

Sieriebriennikov, B. et al. *Mol. Biol. Evol.* **34**, 1644–1653 (2017): <https://academic.oup.com/mbe/article/34/7/1644/3067497>

Sieriebriennikov, B. et al. *PLoS Genetics*. **16**, e1008687 (2020): <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008687>

##### Early paper from our lab using a similar approach but a different protocol

Susoy, V. et al. *Elife* **4**, e05463 (2015): <https://elifesciences.org/articles/05463>

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- |                                     |                                     |                                                                                                                                                                                                                                                            |
|-------------------------------------|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement                                                                                                                                    |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly                                                                                                                                    |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>                                                               |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of all covariates tested                                                                                                                                                                                                                     |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons                                                                                                                                        |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings                                                                                                                                                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes                                                                                                                                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated                                                                                                                                                         |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

- ZEN 2 Pro (Carl Zeiss Microscopy GmbH, 2011; version: 2.0.14283.302; 64-bit)
- Fiji (ImageJ) (version 2.0.0 or later, <https://imagej.net/Fiji/Downloads>)
- Microsoft Excel (version 15.15 or later)
- BBEdit (version: 13 or later, <https://www.barebones.com/products/bbedit/download.html>)

Data analysis

- R (version 3.6.1 or later, <https://cran.r-project.org/mirrors.html>)
- RStudio (version 1.2.5001 or later, <https://rstudio.com/products/rstudio/download/>)
- geomorph (R package version 3.2.1, <https://cran.r-project.org/package=geomorph>)
- Morpho (R package version 2.8, <https://cran.r-project.org/web/packages/Morpho/index.html>)
- cluster (R package version 2.1.0, <https://cran.r-project.org/web/packages/cluster/index.html>)
- mclust (R package version 5.4.6, <https://cran.r-project.org/web/packages/mclust/index.html>)
- factoextra (R package version 1.0.7, <https://cran.r-project.org/web/packages/factoextra/index.html>)
- rptR (R package version 0.9.22, <https://cran.r-project.org/web/packages/rptR/index.html>)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.



## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated and analyzed in this study is included in the paper or its supplements. Upon reasonable request, the raw data that was used to generate the example results is available from the corresponding authors. The supplementary data files 1-3 contain raw data (i.e. landmark coordinates of example data sets).

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

### Study description

This manuscript provides a standardized and ready-to-use protocol for geometric morphometric and cluster analysis in microscopic organisms. GM utilizes two- or three-dimensional landmark coordinates to quantify morphological shapes. First, a set of landmark configurations containing information about the shape, size, orientation and location of homologous anatomical traits is acquired for all specimens. In the second step, general Procrustes analysis (GPA) is performed to remove all parameters of non-shape variation from the dataset. Afterwards, multivariate statistical analysis is performed on the shape coordinates to identify biologically relevant differences between categorical variables like sex, age, culture conditions or species. Finally, patterns of shape (co-)variation are described and visualized using various tools like PCA analysis or deformation grids. Additionally, we use clustering approaches in order to identify group structures in the data set and we describe group-specific allometric trends. We use various nematode species (including well establish model systems like *C. elegans* and *P. pacificus*) to generate example data sets for the proposed analysis (see below).

### Research sample

Our example data sets contain (i) the two different adult morphs of the diplogastrid nematode *Pristionchus pacificus*, (ii) adults of the monomorphic diplogastrid species *Pristionchus bucculentus* and *Pristionchus elegans* and (iii) the two monomorphic rhabditid species *Caenorhabditis elegans* and *Caenorhabditis briggsae*. All biological specimen can be cultured in the lab, and came from laboratory stocks, which are available upon request. Generally only young adult hermaphrodites were considered in this study. We chose *P. pacificus* and *C. elegans* as both of these nematodes represent well-established model organisms in evo-devo studies. We chose both morphs of *P. pacificus* to show that our protocol can reliably differentiate shape differences within the same species. We chose to compare these two morphs to two monomorphic species of *Pristionchus* to show that this protocol for geometric morphometrics can reliably identify identical morphs across species (here: *P. bucculentus* and the eurytomatous morph of *P. pacificus*), as well as intermediate morphs of other species (here: *P. elegans* vs. the two morphs of *P. pacificus*). Additionally we chose *C. elegans* and compared it to *C. briggsae* as both of these model systems are described as superficially similar in morphology. This shows that, by defining other landmark configurations, the proposed protocol for geometric morphometrics and cluster analysis can easily be applied to other nematodes (or microscopic animals in general).

### Sampling strategy

Each of the three example data sets was based on 80 specimens and all groups in these data sets were represented equally (e.g. 80 animals in the *Pristionchus pacificus* data set with 40 animals per morph). A general rule in geometric morphometrics is that for proper statistical analysis (using distance-based PERMANOVA on shape) one should have two to four times as many individuals as variables. Our proposed landmark sets contain 18 landmarks and therefore 36 landmark coordinates (i.e. variables). Four of these variables are lost during the GPA procedure. Thus we have 32 variables. We chose 80 individuals per example data set as this number is 2.5 times as large as the amount of variables, and is easily divisible by the number of groups in each data set (e.g. strains or species), allowing equality in group sizes.

### Data collection

For microscopic imaging, all worms were mounted on object slides with 5% noble agarose pads, which contained 10 mM sodium azide to sedate the animals. All specimens were examined using a 100x DIC objective together with immersion oil. All image stacks were taken using the same settings (8.33 x 7.12 inches; 800 x 684 px; each image being 0.27  $\mu\text{m}$  thick in z direction) and saved in ".czi" or ".tif" format. All (semi-)landmarks were obtained using the multi-point and measurement tools of Fiji and saved in an Excel spreadsheet. Finally this spreadsheet was edited and converted into ".txt" file using BBEdit or a R routine (Suppl. data 6). Sampling and data acquisition was performed by TT, SSW and MSW.

### Timing and spatial scale

The raw data (i.e. the DIC images) were taken in a period of time from June 2019 to August 2019. No temporal gaps occurred during sampling. The time scale is largely explained by the number of different strains or species used in the study, as we subsequently thawed frozen strains/species after finishing sampling of one strain/species. Establishment of new laboratory cultures took up to 2 weeks. Landmark data was collected from Juli to September (in parallel to ongoing sampling).

### Data exclusions

No data was excluded from the analysis.

### Reproducibility

The reproducibility of landmark coordinates is evident from our high biological replicate number, data visualization (PCA), and

Reproducibility  statistical inference (permanova). Nevertheless, to verify the reproducibility of our data sets, we calculated the variances of each landmark across specimen of the same species (Supp. Fig. 2). Furthermore, we estimated the repeatability and percent measurement error for shapes of *P. pacificus* using 10 randomly picked biological and 3 technical replicates per morph (generated with 20 minute breaks in between labeling worms), over a course of several days. This was done by three annotators with varying degree of training in landmark annotation (Supp. fig. 6 and 7).

Randomization  All specimens have been picked randomly from agar plates. Once all image stacks of the desired number of properly orientated individuals was reached, no more individuals were considered.

Blinding  Specimen IDs have been encrypted during the replication process.

Did the study involve field work?  Yes  No

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a                                 | Involvement                                                     |
|-------------------------------------|-----------------------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                             |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology                          |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                          |

### Methods

| n/a                                 | Involvement                                     |
|-------------------------------------|-------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals  We used the African strain (RS5205) of *Pristionchus pacificus*, the RS5596 strain of *P. bucculentus*, the RS5720 strain of *P. elegans*, the N2 reference strain of *C. elegans* and the AF16 reference strain of *C. briggsae*. All specimens used in the study are hermaphrodites and only young adults have been considered.

Wild animals  This study did not involve wild animals.






Field-collected samples  This study did not involve animals which were collected during fieldwork.

Ethics oversight  Nematodes are not included in the animal welfare act of Germany. Therefore, no ethical approval or guidance is required to work with any of the selected strains/species.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

In the format provided by the authors and unedited.

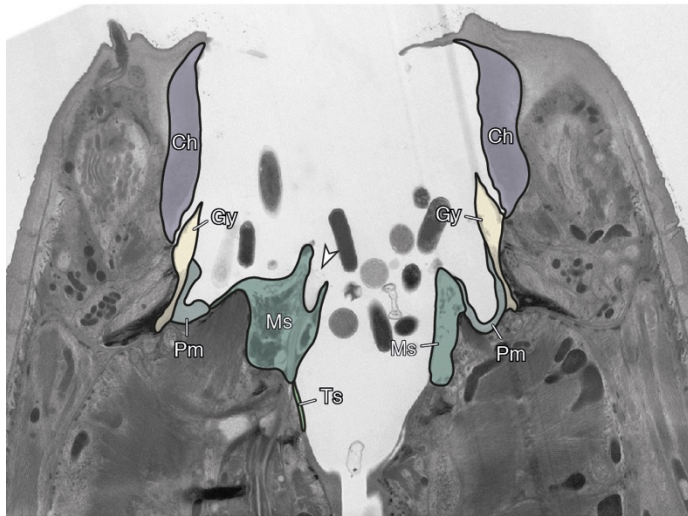
# Geometric morphometrics of microscopic animals as exemplified by model nematodes

Tobias Theska <sup>1</sup>, Bogdan Sieriebriennikov <sup>1,2</sup>, Sara S. Wighard <sup>1</sup>, Michael S. Werner <sup>1</sup>✉ and Ralf J. Sommer <sup>1</sup>✉

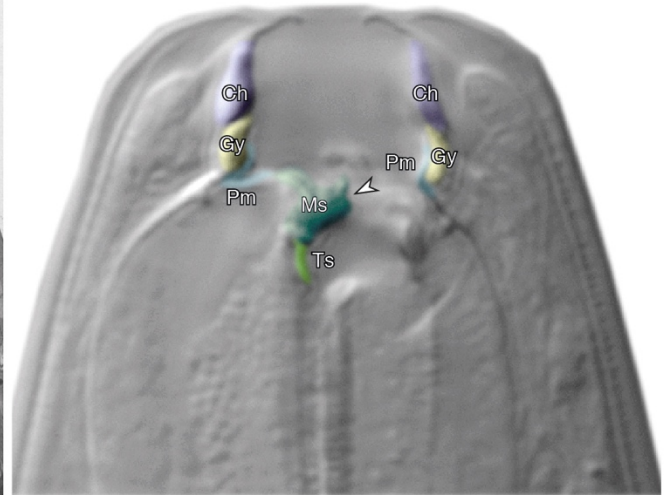
---

<sup>1</sup>Department for Integrative Evolutionary Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany. <sup>2</sup>Present address: Department of Biology, New York University, New York, NY, USA. ✉e-mail: [michael.werner@tuebingen.mpg.de](mailto:michael.werner@tuebingen.mpg.de); [ralf.sommer@tuebingen.mpg.de](mailto:ralf.sommer@tuebingen.mpg.de)

**a** TEM section (sagittal) of eurystomatous morph

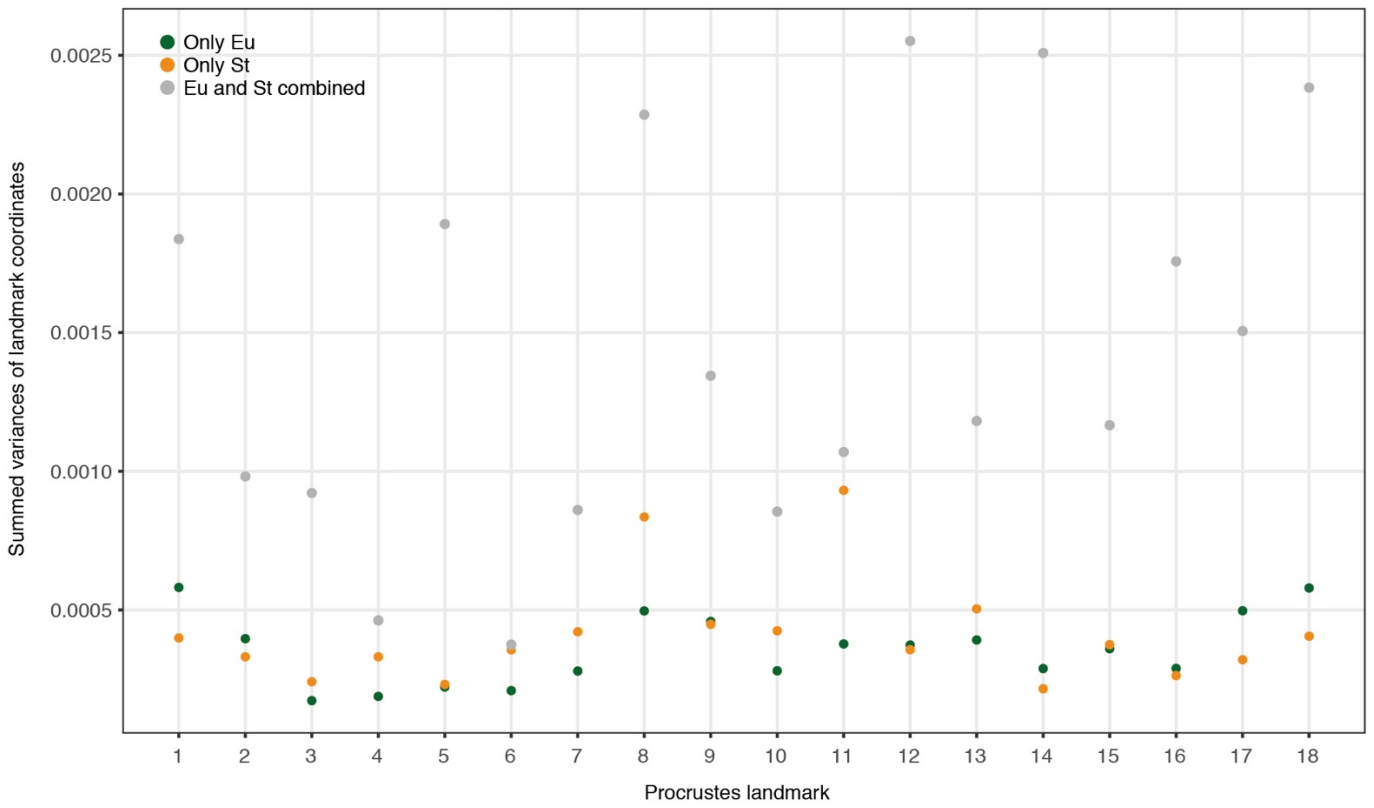


**b** Microscopic image (DIC) of eurystomatous morph

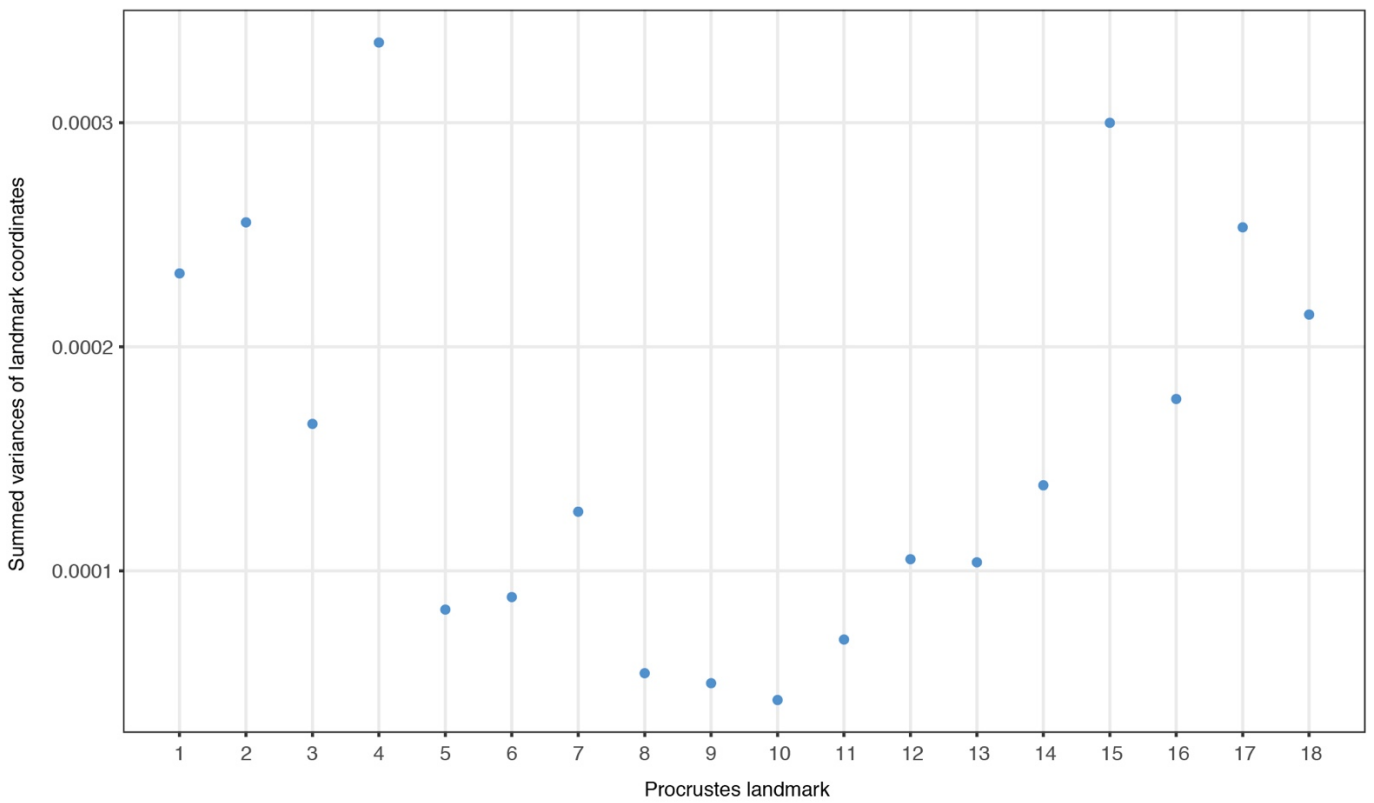


**Supplementary figure 1.** Comparison of the eurystomatous morph of *P. pacificus* under TEM and DIC microscope. **a**, TEM image of the stoma. **b**, Light microscope (DIC) image of the stoma (100x). Both images show the sagittal plane in right lateral view. Arrowheads indicate the position of the dorsal gland opening. ch, cheilostom; gy, gymnostom; ms, metastegostom; pm, promesostegostom; ts, telostegostom.

**a** Summed variances for each landmark in *P. pacificus*

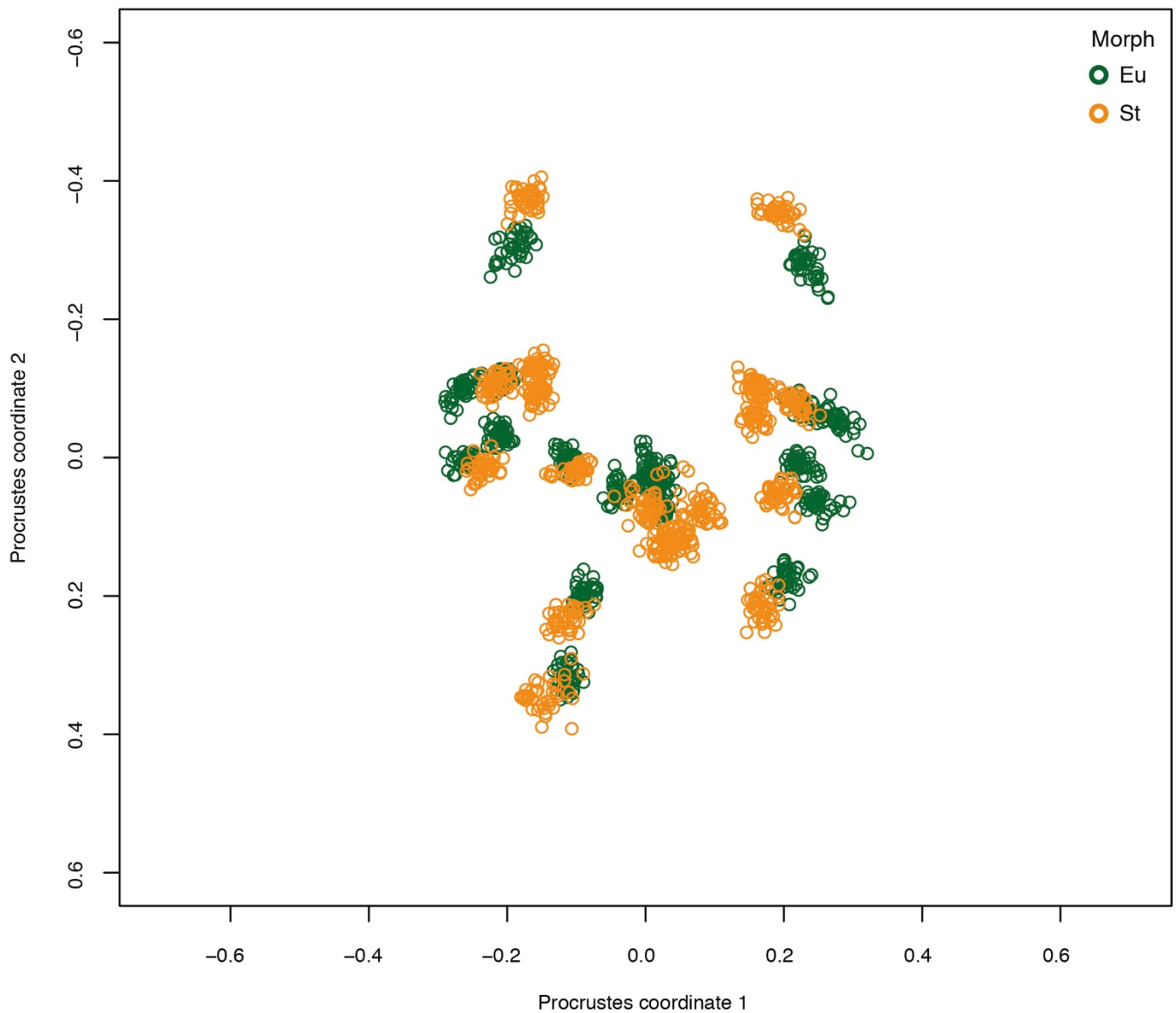


**b** Summed variances for each landmark in *C. elegans*



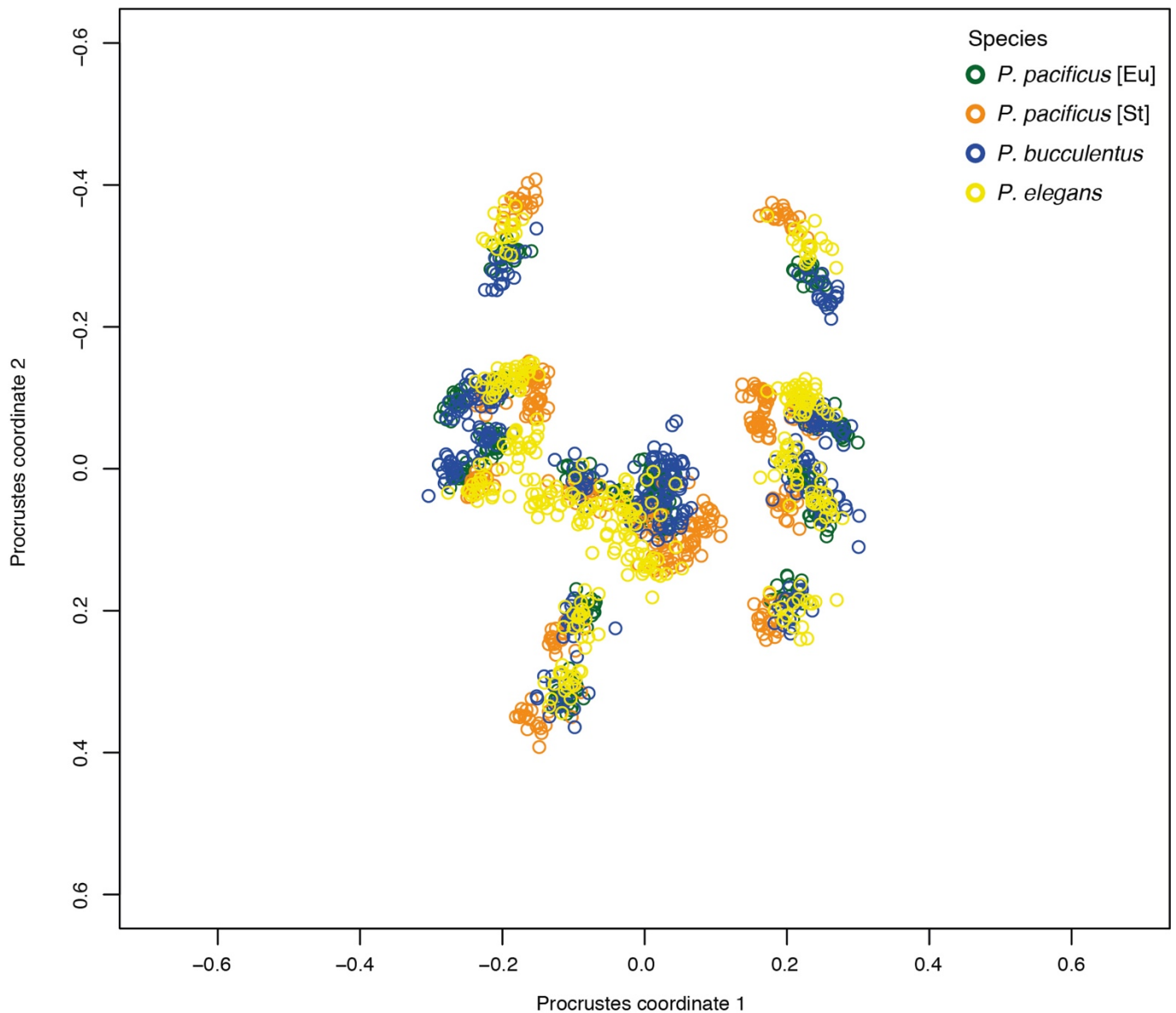
**Supplementary figure 2.** Variances of single landmarks for *P. pacificus* and *C. elegans*. Total variances for each landmark were calculated by summing up the variance in both coordinates of a respective landmark. Single coordinate variances were calculated using the **var()** function of the R statistics package. Note that the *Procrustes coordinates*, and thus also their variances, are dimensionless. **a**, Total variance of single landmarks for each morph (Eu and St) of *P. pacificus*, as well as total variance of single landmarks across all specimens regardless of the morph. **b**, Total variance of single landmarks for all *C. elegans* specimens. **a, b**, Smaller variances indicate a robust landmark positioning. Larger variances can indicate either inherently noisy landmarks (i.e. a high degree of naturally occurring variation in the structure), or biologically meaningful differences between groups. For example: Landmark “5” shows very small variances within each of the two morphs of *P. pacificus*, which indicates that it can be found reliably within each morph. However, the variance in landmark position is very high if all individuals (i.e. Eu and St combined) are used for estimation. This indicates that the position of landmark “5”, relative to all other landmarks of the configuration, is very different between the morphs and thus potentially relevant in explaining biological shape differences. On the other hand, the high variance of landmark “11” across all *P. pacificus* specimens is due to the slightly increased variance of this landmark in stenostomatous animals (St) as compared to the low variance observed for euryostomatous animals (Eu). Thus, this landmark might be less reliable for the identification of shape differences between morphs. Note that landmarks which are jointly placed on movable elements (e.g. landmark 6, 7, 8, 9 & 10 in *P. pacificus*) do not show systematically increased variances as opposed to landmarks which are placed on non-movable elements (e.g. 1-4 and 15-18 in *P. pacificus*). This indicates that landmarks on movable and non-movable elements can be treated equally for downstream analysis.

Aligned Procrustes landmarks of all individuals of *P. pacificus* (stomatal polyphenism)



**Supplementary figure 3.** Scatterplot of Procrustes aligned coordinates obtained for the *P. pacificus* dataset (including both morphs). Generalized Procrustes analysis (GPA) was carried out using the **gpagen()** function of the *geomorph* package (see step 20 of the Procedure). *Procrustes coordinates* are projected into a linear (euclidean) tangent space. Semilandmarks were slid by minimizing bending energy. Output image was flipped horizontally, to depict aligned shapes in the original (right lateral) orientation.

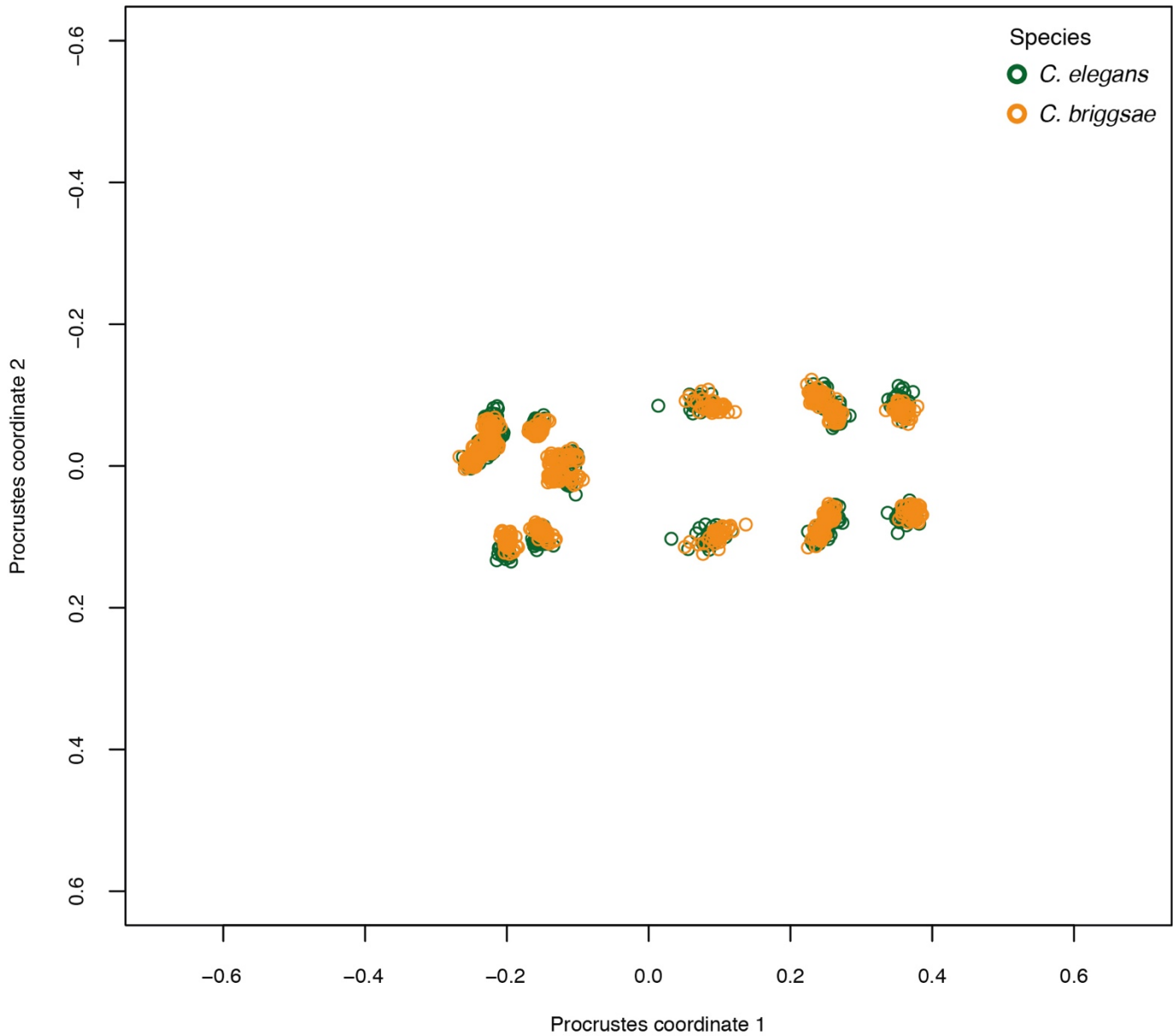
Aligned Procrustes landmarks of all individuals in the interspecific comparison (*Pristionchus sp.*)



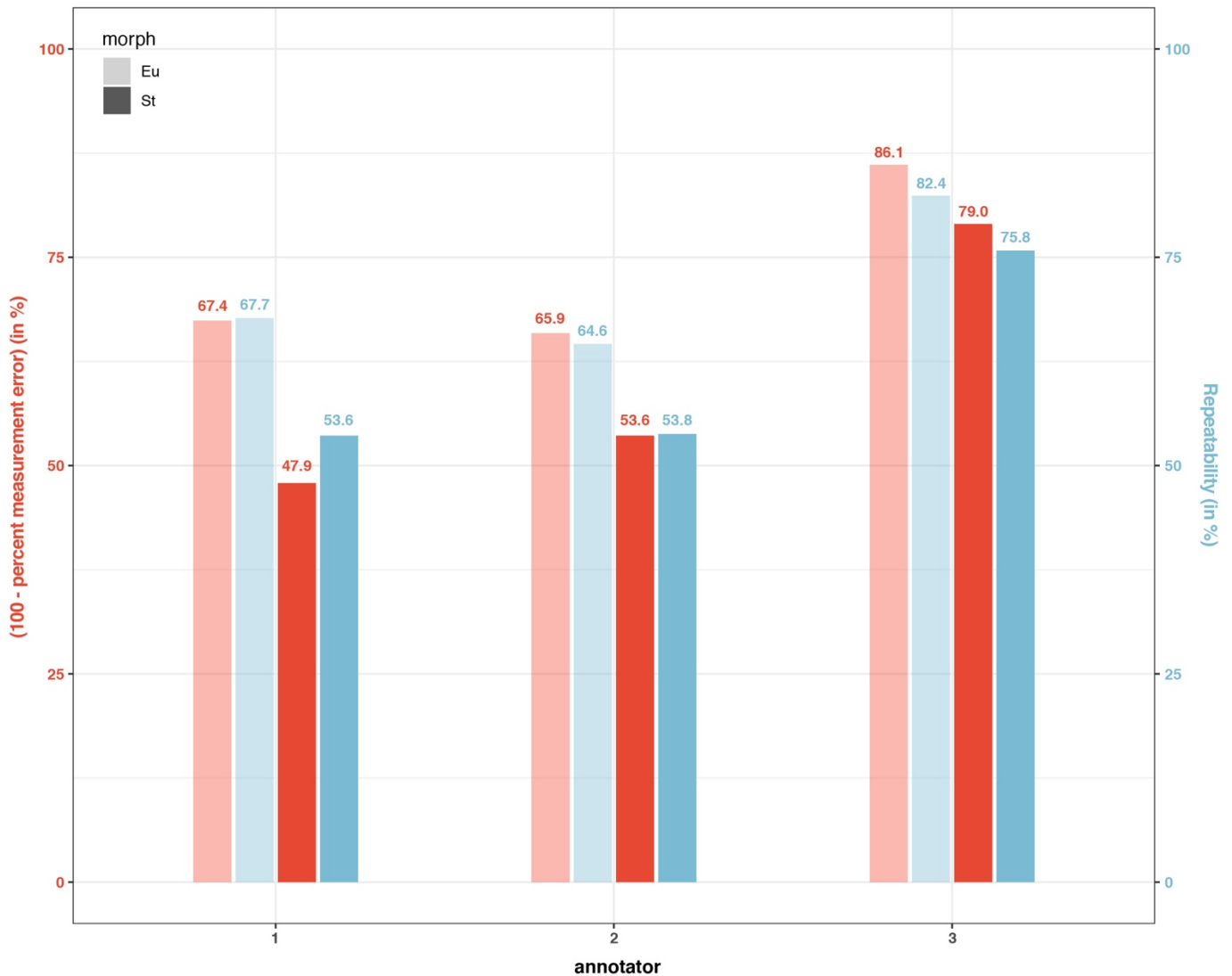
**Supplementary figure 4.** Scatterplot of Procrustes aligned coordinates obtained for the *Pristionchus* dataset (including *P. pacificus*, *P. bucculentus* and *P. elegans*). Generalized Procrustes analysis (GPA) was carried out using the **gpagen()** function of the *geomorph* package (see step 20 of the Procedure). *Procrustes coordinates* are projected into a linear (euclidean) tangent space. Semilandmarks were slid by minimizing bending energy. Output image was flipped horizontally, to depict aligned shapes in the original (right lateral) orientation.



Aligned Procrustes landmarks of all individuals of *C. elegans* and *C. briggsae*



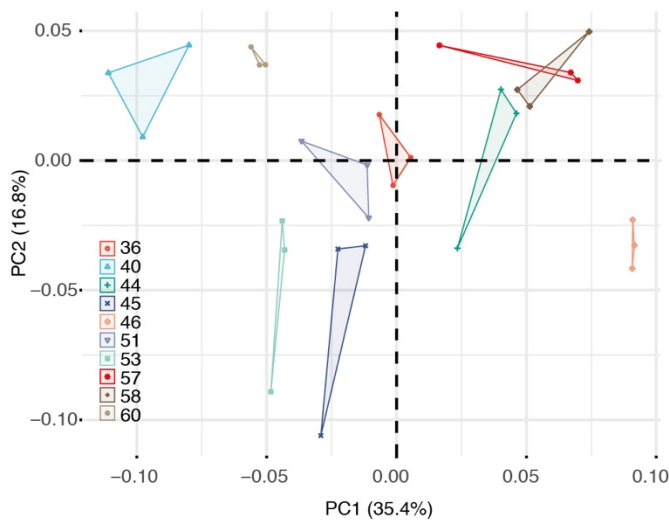
**Supplementary figure 5.** Scatterplot of Procrustes aligned coordinates obtained for the *Caenorhabditis* dataset (including *C. elegans* and *C. briggsae*). Generalized Procrustes analysis (GPA) was carried out using the **gpagen()** function of the *geomorph* package (see step 20 of the Procedure). *Procrustes coordinates* are projected into a linear (euclidean) tangent space. Output image was flipped horizontally, to depict aligned shapes in the original (right lateral) orientation.



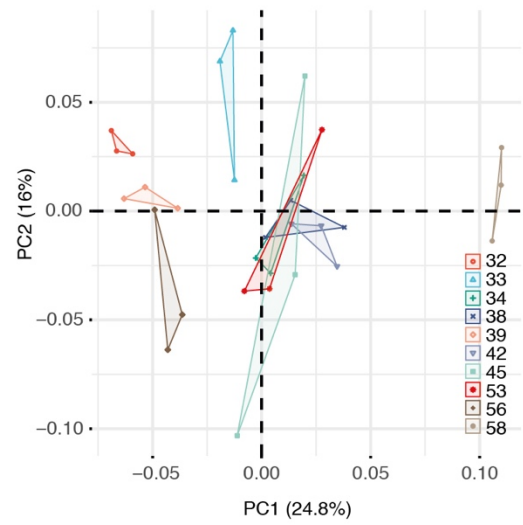
**Supplementary figure 6.** Replicability performance of annotators with different levels of training. Repeatability (i.e. the amount of variation between individuals) and percent measurement error (i.e. the amount of variation within individuals/among replicates) were calculated for shapes using the code in Box 2 of the main text. Annotators 1 and 2 are less trained in the procedure of landmark placement, while annotator 3 is familiar with the procedure. Color coding in y-axis corresponds to color coding of bars. Red, percent measurement error (%ME); blue, repeatability (RPT). Light bars indicate the values obtained for Eu, and dark bars indicate those obtained for St animals. Abbreviations: Eu, eurystomatous and St, stenostomatous morph of *P. pacificus*. Superimposition was performed using the **gpagen()** function of *geomorph* (see step 20 of the Procedure). Semilandmarks were slid by minimizing bending energy. For repeatability estimation, the number of parametric bootstraps and permutations was set to 1,000. In all cases (across morphs and annotator), when estimating the %ME of shapes, the specimen factor was significant while the replicate factor was not. This means that the variation between worms was always

stronger than the variation across technical replicates of the same individual. Statistics (PERMANOVA on shape) for annotator-1: ME[Eu] = 32.6%, ME[St] = 52.1%. Specimen factor [Eu],  $F = 7.2156$ , effect size ( $Z$ ) = 8.7559,  $P < 0.001$ . Replicate factor [Eu],  $F = 0.4828$ , effect size ( $Z$ ) = -1.6918,  $P = 0.96$ . Specimen factor [St],  $F = 3.7629$ , effect size ( $Z$ ) = 7.5127,  $P < 0.001$ . Replicate factor [St],  $F = 0.7633$ , effect size ( $Z$ ) = -0.7179,  $P = 0.76$ . Statistics (PERMANOVA on shape) for annotator-2: ME[Eu] = 34.1%, ME[St] = 46.4%. Specimen factor [Eu],  $F = 6.8098$ , effect size ( $Z$ ) = 7.4885,  $P < 0.001$ . Replicate factor [Eu],  $F = 0.7233$ , effect size ( $Z$ ) = -0.63636,  $P = 0.72$ . Specimen factor [St],  $F = 4.465$ , effect size ( $Z$ ) = 7.4693,  $P < 0.001$ ; Replicate factor [St],  $F = 0.6583$ , effect size ( $Z$ ) = -1.0927,  $P = 0.88$ . Statistics (PERMANOVA on shape) for annotator-3: ME[Eu] = 13.9%, ME[St] = 21.0%. Specimen factor [Eu],  $F = 19.647$ , effect size ( $Z$ ) = 9.2233,  $P < 0.001$ ; Replicate factor [Eu],  $F = 0.1969$ , effect size ( $Z$ ) = -2.9198,  $P = 0.999$ . Specimen factor [St],  $F = 12.286$ , effect size ( $Z$ ) = 11.612,  $P < 0.001$ ; Replicate factor [St],  $F = 0.3595$ , effect size ( $Z$ ) = -2.5964,  $P = 0.995$ .

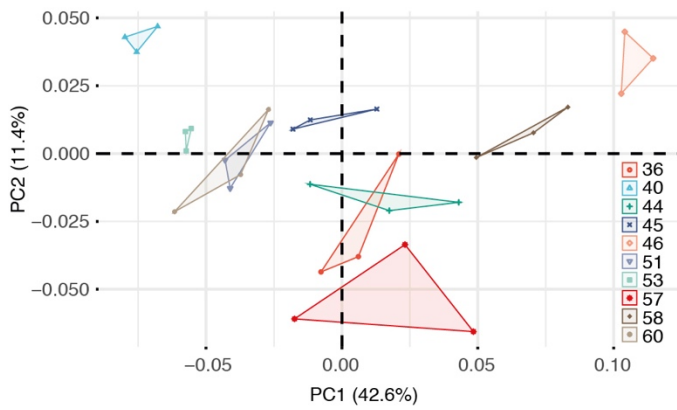
**a** PCA of *P. pacificus* [Eu] replicates - annotator 1



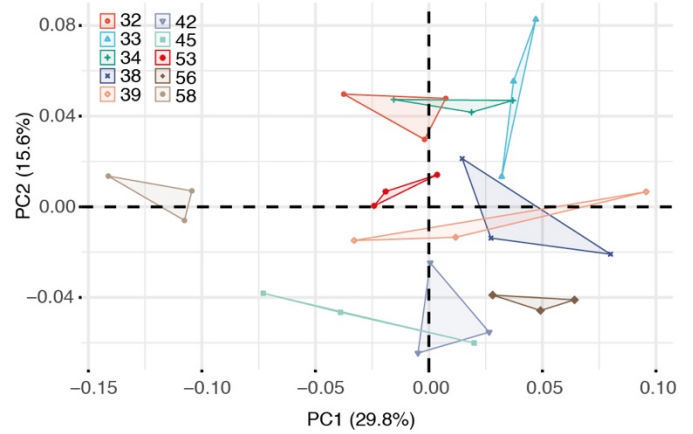
**d** PCA of *P. pacificus* [St] replicates - annotator 1



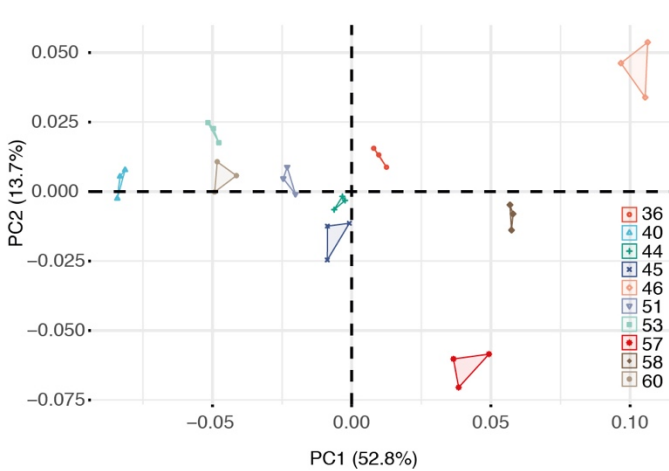
**b** PCA of *P. pacificus* [Eu] replicates - annotator 2



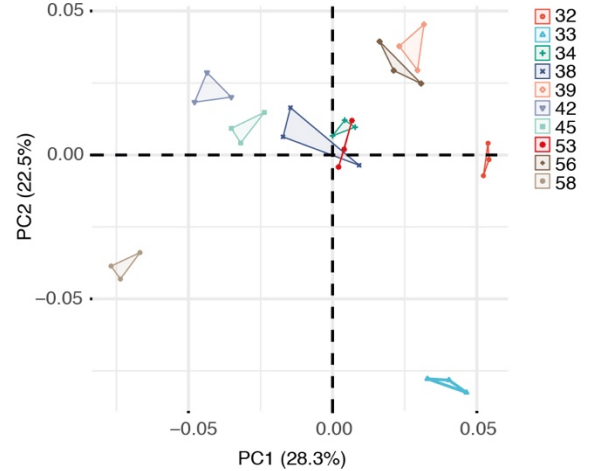
**e** PCA of *P. pacificus* [St] replicates - annotator 2



**c** PCA of *P. pacificus* [Eu] replicates - annotator 3



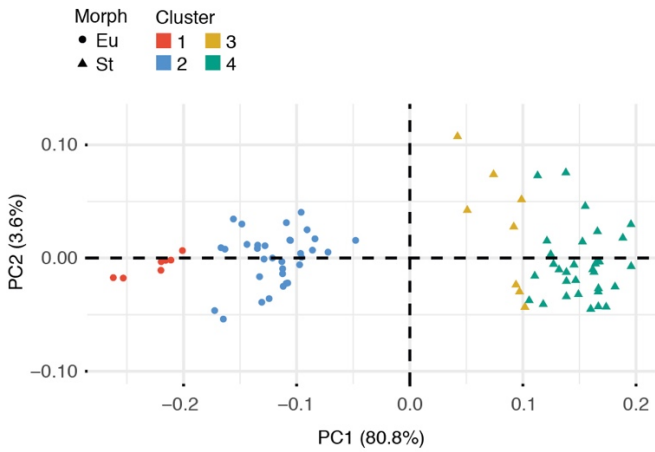
**f** PCA of *P. pacificus* [St] replicates - annotator 3



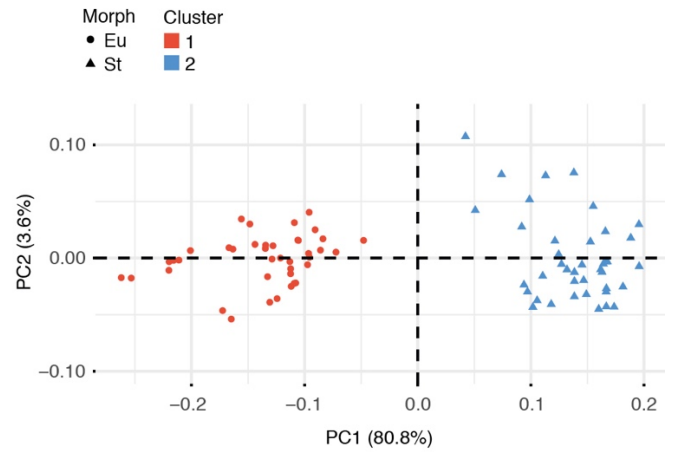
**Supplementary figure 7.** Annotator-specific distribution of technical replicates (stomatal shapes) in the shape tangent space. Each individual was labeled three times independently by three annotators of

different levels of training in landmark placement. Numbers in the legends refer to identifiers of the ten eurystomatous (a-c) or ten stenostomatous (d-f) specimens (i.e. the biological replicates of each morph). Generalized Procrustes analysis (GPA) was carried out using the **gpagen()** function of the *geomorph* package (see step 20 of the Procedure). *Procrustes coordinates* are projected into a linear (euclidean) tangent space. Semilandmarks were slid by minimizing bending energy. PCA was performed according to step 21 of the Procedure. Note that specimen ID numbers are *not* identical between eurystomatous (a-c) and stenostomatous (d-f) individuals (e.g. '58' refers to two different animals in a-c versus d-f).

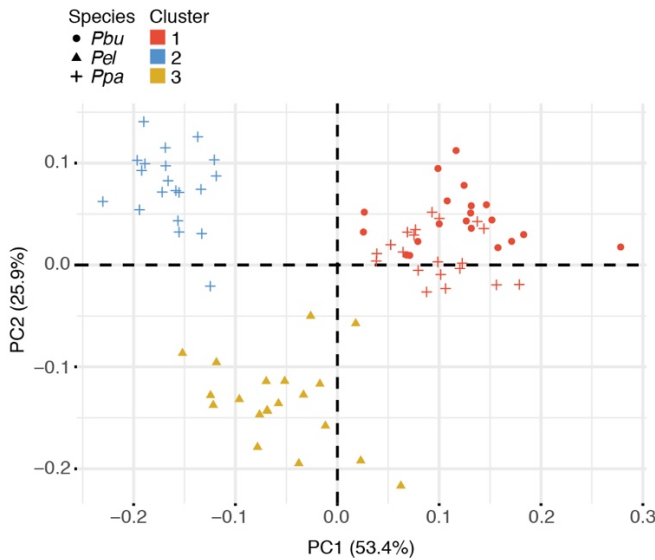
**a** Stomatal polyphenism in *P. pacificus* (mPCs)



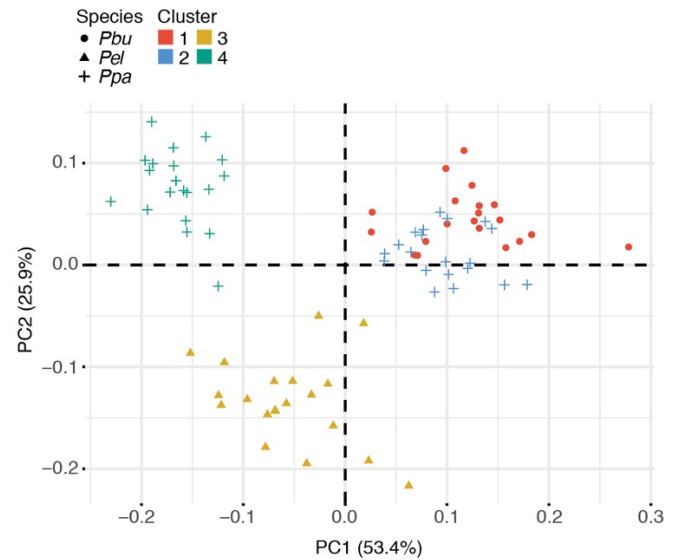
**b** Stomatal polyphenism in *P. pacificus* (PCs > 1%)



**c** Stomata of different *Pristionchus* species (mPCs)

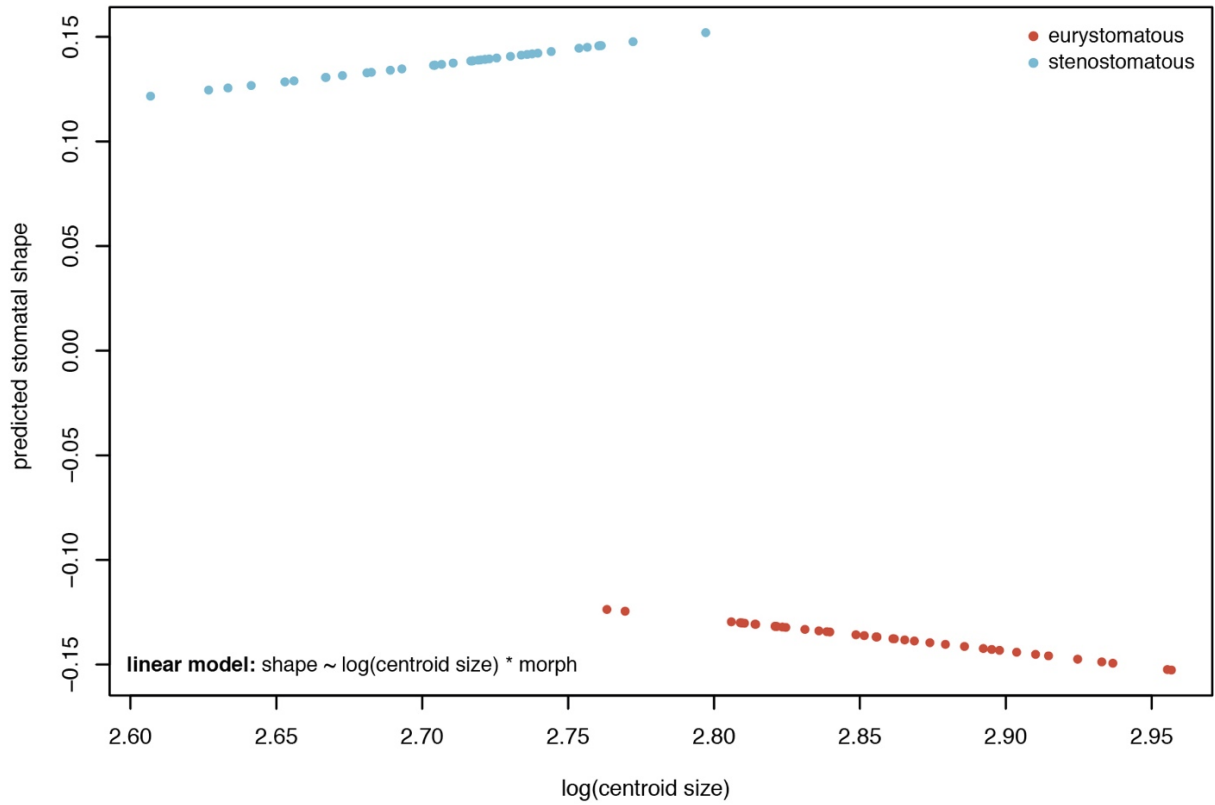


**d** Stomata of different *Pristionchus* species (PCs > 1%)

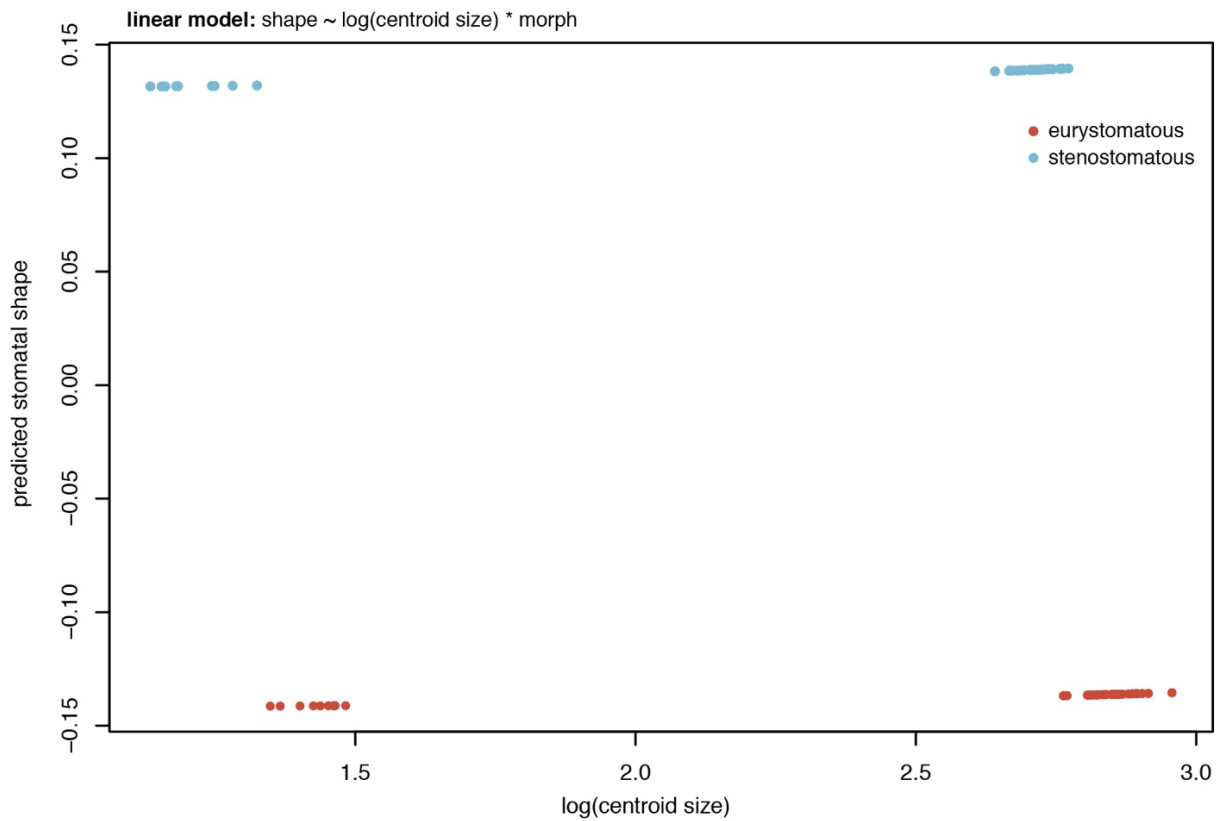


**Supplementary figure 8.** model-based clustering performed on the two *Pristionchus* datasets using two different approaches for PC inclusion. **a**, Clusters identified for the *P. pacificus* dataset by using only meaningful PCs (mPCs) as an input. **b**, Clusters identified for the *P. pacificus* dataset by including all PCs which describe at least 1% variation (PCs>1%) as an input. **c**, Clusters identified for the *Pristionchus* species dataset by using only meaningful PCs (mPCs) as an input. **d**, Clusters identified for the *Pristionchus* species dataset by including all PCs which describe at least 1% variation (PCs>1%) as an input. Abbreviations: Eu, eury stomatous morph; St, stenostomatous morph; *Pbu*, *P. bucculentus*; *Pel*, *P. elegans*; *Ppa*, *P. pacificus*. Meaningful PCs are estimated by the `getMeaningfulPCs()` function of the *Morpho* package.

**a** Allometric trajectories of the two different stomatal morphs in *P. pacificus*



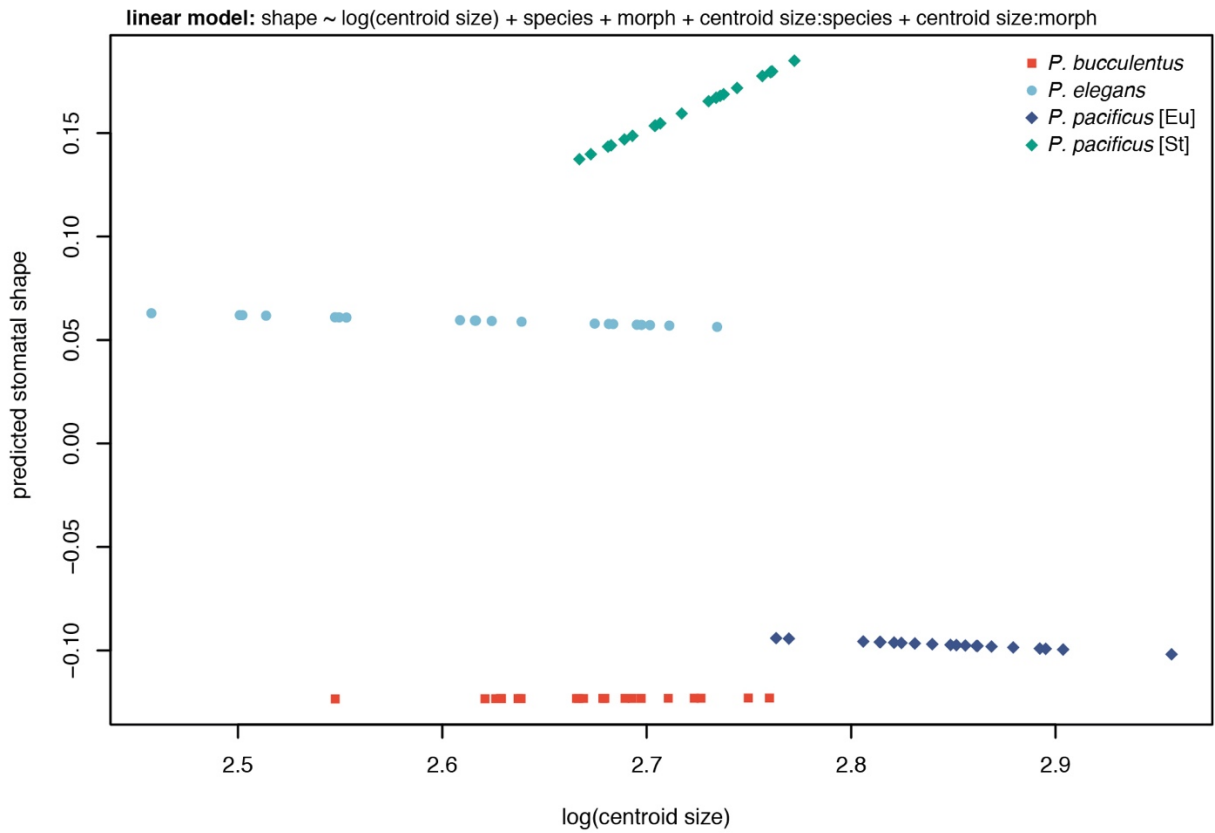
**b** Allometric trajectories for the two morphs of *P. pacificus* based on uncorrected centroid sizes



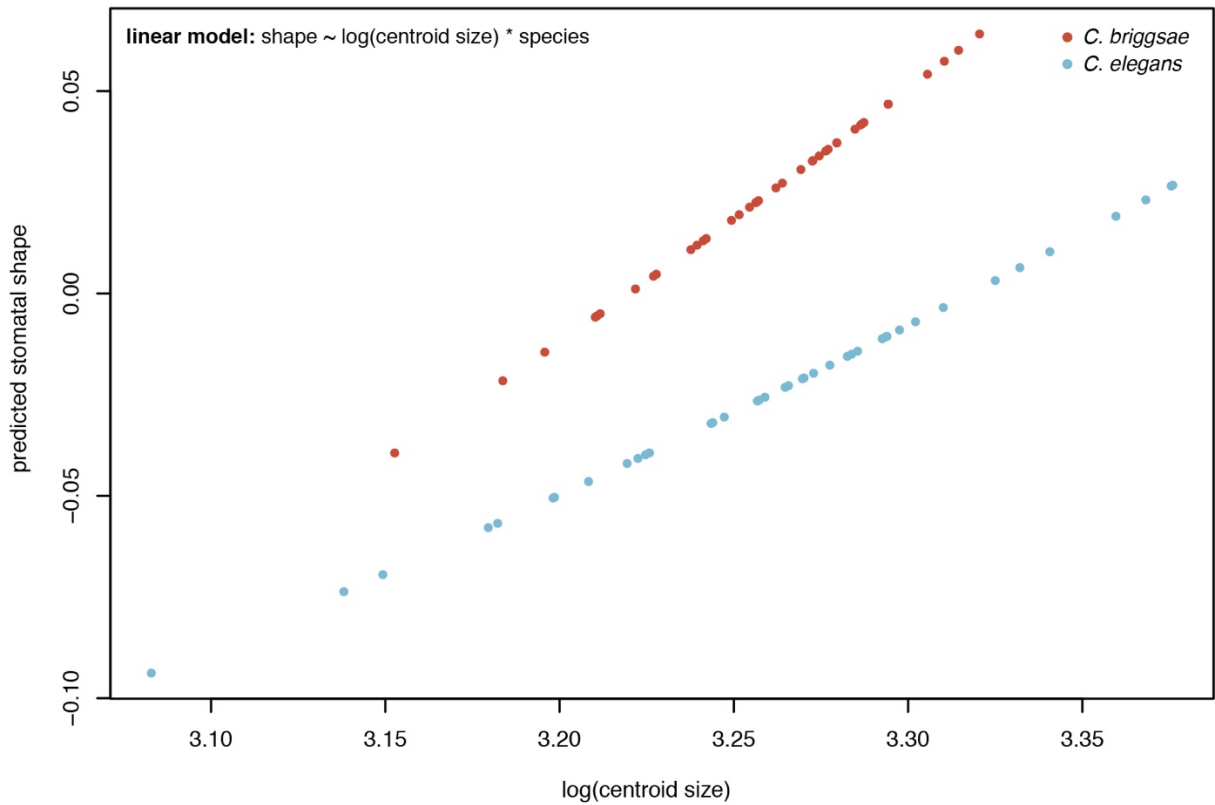
**Supplementary figure 9.** Allometry plots for young adults of the two *P. pacificus* morphs. **a**, Morph-specific (static) allometric trajectories after troubleshooting (see main text) differences in centroid size due to inconsistent image stack scaling (inches versus microns). **b**, Morph-specific (static) allometric trajectories before troubleshooting centroid sizes (note that individuals whose centroid size was in inches separated from their kin towards the left of the x-axis). Predicted stomatal shapes on the y-axis represent the PC1 scores of the PCA that was performed on the predicted values from multivariate regression on shape versus log(centroid size). Superimposition was performed using the **gpagen()** function of *geomorph* (see step 20 of the Procedure). Note that, while each morph shows static allometry, the differences between morphs are not manifestations of allometric scaling as they do not share a common trajectory. Linear models are indicated in the panels. Statistics on shape (PERMANCOVA): size,  $F = 165.8679$ , effect size ( $Z$ ) = 5.7801,  $P < 0.0001$ . morph,  $F = 59.6372$ , effect size ( $Z$ ) = 6.1827,  $P < 0.0001$ . size x morph,  $F = 1.7444$ , effect size ( $Z$ ) = 1.356,  $P = 0.097$ .



**a** Allometric trajectories of the stomata in various *Pristionchus* species

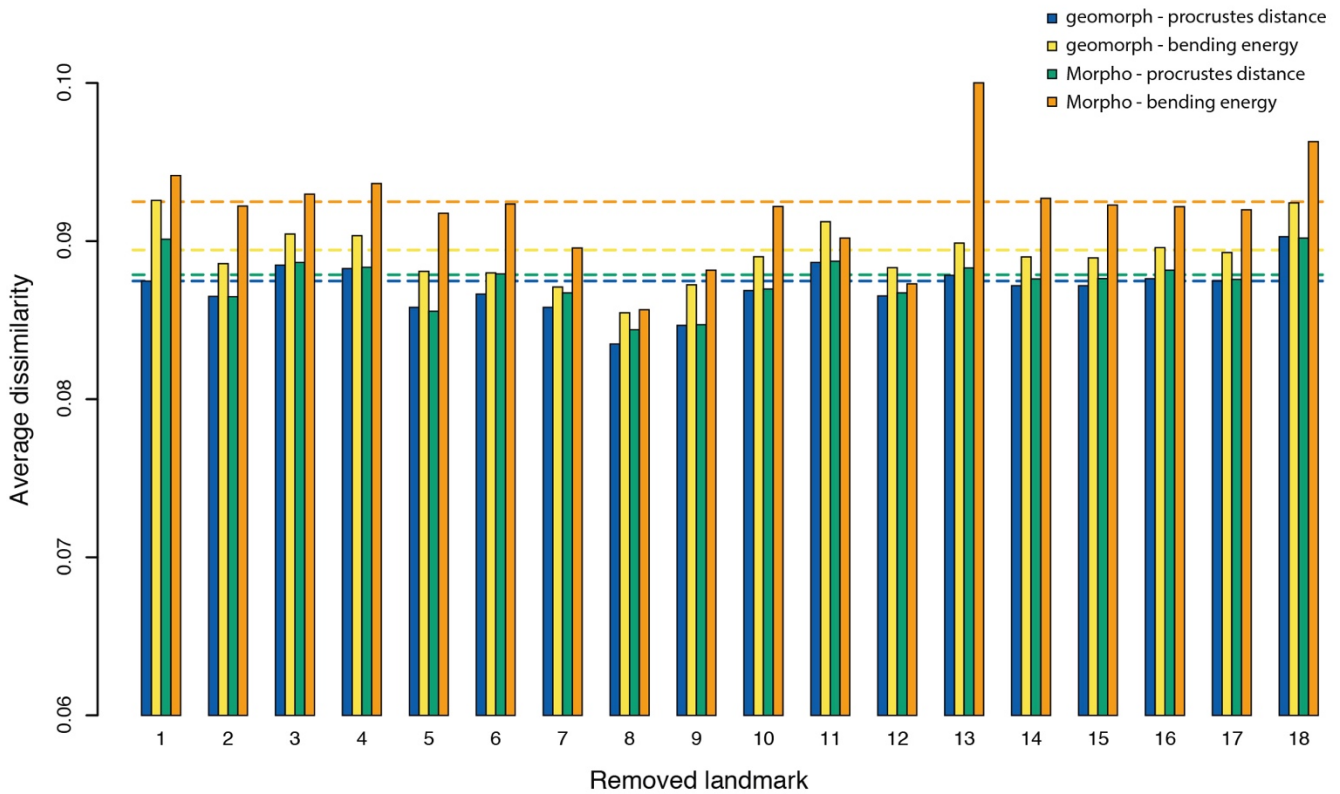


**b** Allometric trajectories of the stomata in *C. briggsae* and *C. elegans*



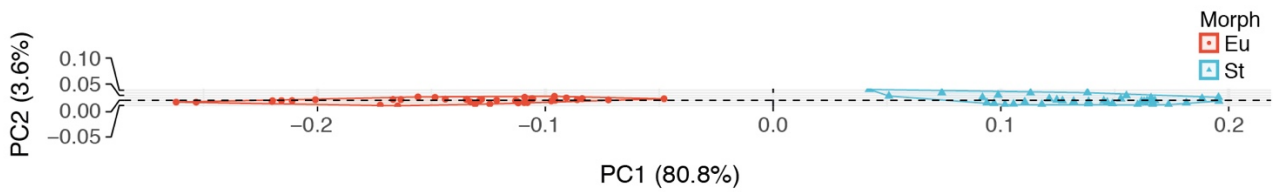
**Supplementary figure 10.** Static allometry trajectories estimated for young adults of **(a)** three *Pristionchus* species and **(b)** two *Caenorhabditis* species. Note that differences in centroid size due to inconsistent image stack scaling (inches versus microns) were corrected before plotting panel **a** (see ‘troubleshooting’ in the main text). Predicted stomatal shapes on the y-axis represent the PC1 scores of the PCA that was performed on the predicted values from multivariate regression on shape versus log(centroid size). Superimposition was performed using the **gpagen()** function of *geomorph* (see step 20 of the Procedure). Note that *P. elegans*, eury stomatous *P. pacificus* and *P. bucculentus* (**a**) show essentially isometric trendlines (i.e. shapes do not change with increasing body size), while the stenostomatous morph of *P. pacificus* shows a specific allometric trajectory. Both *Caenorhabditis* species (**b**) show strong allometric relationships in stomatal shape, indicative of increasing morphological divergence between species with growth. Abbreviations: Eu, eury stomatous and St, stenostomatous morph. Linear models are indicated in the panels. Statistics on shape (PERMANCOVA) for the *Pristionchus* data set **(a)**: size,  $F = 20.2303$ , effect size ( $Z$ ) = 4.3164,  $P < 0.0001$ . species,  $F = 62.7531$ , effect size ( $Z$ ) = 7.4549,  $P < 0.0001$ . morph,  $F = 48.0781$ , effect size ( $Z$ ) = 6.2528,  $P < 0.0001$ . size x species,  $F = 1.2129$ , effect size ( $Z$ ) = 0.7009,  $P = 0.23831$ . size x morph,  $F = 2.1124$ , effect size ( $Z$ ) = 1.758,  $P < 0.05$ . Statistics on shape (PERMANCOVA) for the *Caenorhabditis* data set **(b)**: size,  $F = 27.1836$ , effect size ( $Z$ ) = 5.1015,  $P < 0.0001$ . species,  $F = 30.6839$ , effect size ( $Z$ ) = 6.4582,  $P < 0.0001$ . size x species,  $F = 4.3322$ , effect size ( $Z$ ) = 3.7579,  $P < 0.001$ .

Average dissimilarity within the two best  $k$ -medoid clusters (*Ppa* polyphenism)

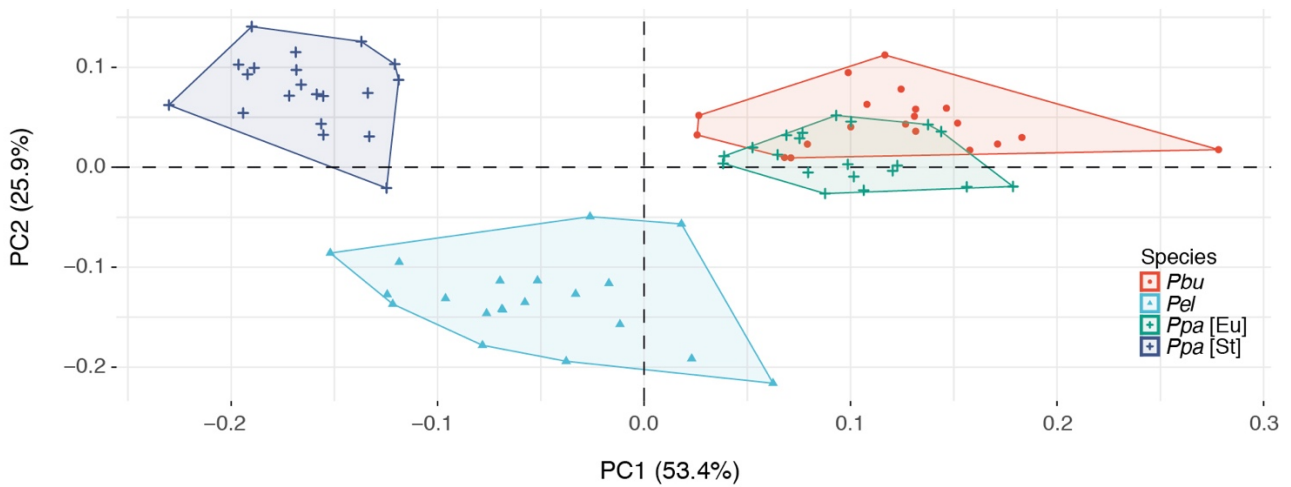


**Supplementary figure 11.** Change of average dissimilarity within the two-best  $k$ -medoid clusters upon single landmark removal in the *P. pacificus* dataset (including both morphs). Dotted lines represent the average dissimilarity of the best two  $k$ -medoid clusters in the data set, when no landmarks are removed (colors correspond to the package and sliding approach used to perform GPA). Note that the detected change for each landmark removal shows a similar direction, regardless of the package which was used for GPA (*geomorph* and *Morpho*) and regardless of the sliding approach that was chosen for GPA (minimization of Procrustes distances or bending energy). Changes of dissimilarity are generally subtle, indicating that the proposed landmark configuration is robustly able to quantify shape differences, even after single landmarks are removed.

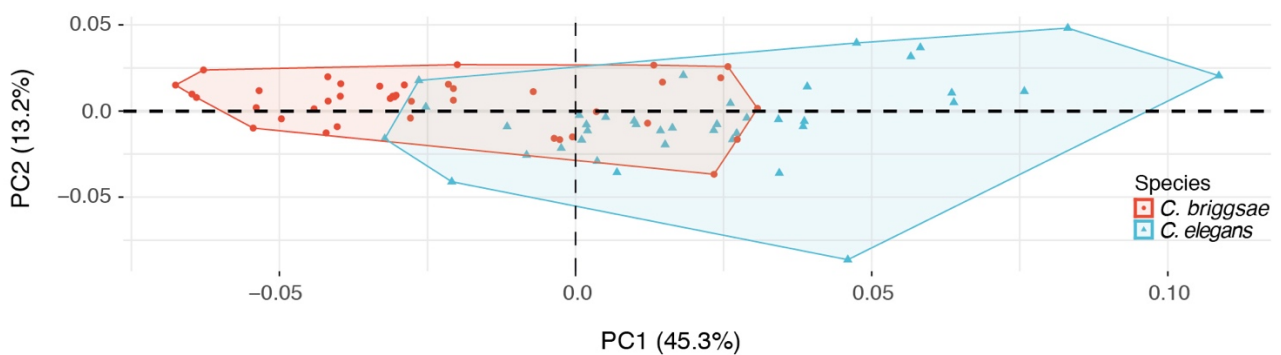
**a** Variation-weighted PCA plot showing the two different mouth-forms of *P. pacificus*



**b** Variation-weighted PCA plot showing stomatal shapes of various *Pristionchus* species



**c** Variation-weighted PCA plot showing stomatal shapes of two *Caenorhabditis* species



**Supplementary figure 12.** Variation-weighted PCA plots. Length of the y-axis is scaled relative to the x-axis, based on the variation described by PC1 and PC2. **a**, Comparison of two stomatal morphs in *Pristionchus pacificus*; **b**, Comparison of stomatal shapes across various *Pristionchus* species; **c**, Comparison of two *Caenorhabditis* species. Abbreviations: Eu, eurystomatous; St, stenostomatous; *Pbu*, *Pristionchus bucculentus*; *Pel*, *Pristionchus elegans*; *Ppa*, *Pristionchus pacificus*.






| landmark | type    | description                                                                         |
|----------|---------|-------------------------------------------------------------------------------------|
| 1        | fixed   | anteriormost point of cheilostom on the dorsal side                                 |
| 2        | fixed   | posteriormost point of cheilostom on the dorsal side                                |
| 3        | fixed   | anteriormost point of gymnostom on the dorsal side                                  |
| 4        | fixed   | posteriormost point of gymnostom on the dorsal side                                 |
| 5        | fixed   | anteriormost point of promesostegostom on the dorsal side                           |
| 6        | sliding | most convex point on the dorsal outline of the dorsal tooth                         |
| 7        | sliding | most concave point on the dorsal outline of the dorsal tooth                        |
| 8        | fixed   | anterior tip of the dorsal tooth                                                    |
| 9        | fixed   | opening of the dorsal pharyngeal gland                                              |
| 10       | fixed   | posteriormost point of the metastegostom on dorsal side (i.e. base of dorsal tooth) |
| 11       | fixed   | lateral base point of the right ventrosublateral tooth or ridge                     |
| 12       | sliding | anterior tip of the right ventrosublateral tooth or ridge                           |
| 13       | fixed   | ventral base point of the right ventrosublateral tooth or ridge                     |
| 14       | fixed   | anteriormost point of promesostegostom on the ventral side                          |
| 15       | fixed   | posteriormost point of cheilostom on the ventral side                               |
| 16       | fixed   | anteriormost point of gymnostom on the ventral side                                 |
| 17       | fixed   | posteriormost point of cheilostom on the ventral side                               |
| 18       | fixed   | anteriormost point of cheilostom on the ventral side                                |

**Supplementary table 1.** Description of landmarks for *Pristionchus*.

| <b>landmark</b> | <b>type</b> | <b>description</b>                                                                                           |
|-----------------|-------------|--------------------------------------------------------------------------------------------------------------|
| 1               | fixed       | anteriormost point of cheilostom on the dorsal side                                                          |
| 2               | fixed       | posteriormost point of cheilostom on the dorsal side (deepest point within the cheilostomal ridge)           |
| 3               | fixed       | anteriormost point of gymnostom on the dorsal side                                                           |
| 4               | fixed       | border point of cheilo- and promesostegostom on the dorsal side (at the level of the pharyngeal sleeve tip)  |
| 5               | fixed       | posteriormost point of promesostegostom on the dorsal side                                                   |
| 6               | fixed       | posterior base point of the dorsal flap                                                                      |
| 7               | fixed       | anterior tip of the dorsal flap                                                                              |
| 8               | fixed       | anteriormost point of the telostegostom on the dorsal side                                                   |
| 9               | fixed       | opening of the dorsal pharyngeal gland                                                                       |
| 10              | fixed       | posteriormost point of the telostegostom on the dorsal side                                                  |
| 11              | fixed       | lateral base point of the right ventrosublateral flap                                                        |
| 12              | fixed       | anterior tip of the right ventrosublateral flap                                                              |
| 13              | fixed       | ventral base point of the right ventrosublateral flap                                                        |
| 14              | fixed       | posteriormost point of promesostegostom on the ventral side                                                  |
| 15              | fixed       | border point of cheilo- and promesostegostom on the ventral side (at the level of the pharyngeal sleeve tip) |
| 16              | fixed       | anteriormost point of gymnostom on the ventral side                                                          |
| 17              | fixed       | posteriormost point of cheilostom on the ventral side (deepest point within the cheilostomal ridge)          |
| 18              | fixed       | anteriormost point of cheilostom on the ventral side                                                         |

**Supplementary table 2.** Description of landmarks for *Caenorhabditis*.

# A New Hope: A Hermaphroditic Nematode Enables Analysis of a Recent Whole Genome Duplication Event

Sara S. Wighard , Marina Athanasouli , Hanh Witte , Christian Rödelsperger , and Ralf J. Sommer \*

Department for Integrative Evolutionary Biology, Max Planck Institute for Biology Tübingen, Max Planck Ring 9, 72076 Tübingen, Germany

\*Corresponding author: E-mail: ralf.sommer@tuebingen.mpg.de.

Accepted: 22 November 2022

## Abstract

Whole genome duplication (WGD) is often considered a major driver of evolution that leads to phenotypic novelties. However, the importance of WGD for evolution is still controversial because most documented WGD events occurred anciently and few experimental systems amenable to genetic analysis are available. Here, we report a recent WGD event in the hermaphroditic nematode *Allodiplogaster sudhausi* and present a comparison with a gonochoristic (male/female) sister species that did not undergo WGD. Self-fertilizing reproduction of *A. sudhausi* makes it amenable to functional analysis and an ideal system to study WGD events. We document WGD in *A. sudhausi* through karyotype analysis and whole genome sequencing, the latter of which allowed us to 1) identify functional bias in retention of protein domains and metabolic pathways, 2) show most duplicate genes are under evolutionary constraint, 3) show a link between sequence and expression divergence, and 4) characterize differentially expressed duplicates. We additionally show WGD is associated with increased body size and an abundance of repeat elements (36% of the genome), including a recent expansion of the DNA-hAT/Ac transposon family. Finally, we demonstrate the use of CRISPR/Cas9 to generate mutant knockouts, whereby two WGD-derived duplicate genes display functional redundancy in that they both need to be knocked out to generate a phenotype. Together, we present a novel experimental system that is convenient for examining and characterizing WGD-derived genes both computationally and functionally.

**Key words:** whole genome duplication, polyploidization, *Allodiplogaster sudhausi*, *Pristionchus pacificus*, Diplogastridae, transposable elements, ohnologs, body size.

## Significance

Whole genome duplication (WGD) has been proposed as a major factor for evolution as it results in doubling of the genetic material of an organism. However, its role in evolution is still controversial as all documented cases have occurred in ancient history. Also, no study systems are available for experimental manipulation of WGD in animals. Here, we report that the hermaphroditic nematode *Allodiplogaster sudhausi* has recently undergone WGD. We document WGD by karyotype analysis and whole genome sequencing, which allowed studying several associated features. Finally, we establish CRISPR-mediated gene knockout, which allows functional manipulation in this organism, a useful tool for investigating the consequence of WGD.

## Introduction

Whole genome duplication (WGD), also known as polyploidization, is when the full genome, including the chromosomes and regulatory elements, is doubled. The

important role of duplication events was hypothesized as far back as the 1930s (Haldane 1932; Bridges 1936), but it is Susumu Ohno's seminal work (Ohno 1970) that popularized the notion of duplication, and WGD in particular,

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

driving evolution and novelty. WGD-derived duplicate genes are referred to as “ohnologs” in reference to his work. Duplication is believed to be a driver of evolution as it produces additional redundant genetic material that can potentially diverge and gain a new function. WGD is considered a greater evolutionary driver than local gene duplications as the genes are not constrained by dosage compensation (Birchler et al. 2005) and indeed, retention of duplicates derived from WGD is greater than those derived from local gene duplication events (Blomme et al. 2006).

In principle, WGD can come about from one of two ways: 1) Autopolyploidization, where multiple chromosome sets derive from a single taxon—usually due to an error in meiosis or 2) Allopolyploidization where two closely related species hybridize to form a new one. However, it is usually difficult to tease apart the exact origin (Parisod et al. 2010). After duplication, ohnologs are initially redundant with the same expression pattern and role, but after some time the fates of ohnologs tend to diverge. The degeneration or silencing of one ohnolog (nonfunctionalization) is the most common fate, while the addition of a novel function (neofunctionalization)—which could drive evolution—is the rarest (Lynch and Conery 2000; Maere et al. 2005).

It is thought that ancient WGD occurred in most eukaryotic lineages (Wolfe 2015). Indeed, nearly all documented cases of WGD have occurred ancestrally and often the exact timing of the WGD event is poorly understood. For example, WGD is characteristic of most land plants, but its timing remains elusive so that the role of WGD for morphological and functional diversity of land plants is constrained (Clark and Donoghue 2018). At least two rounds of ancient WGD took place in vertebrates (Dehal and Boore 2005), while additional WGD events are also reported in teleost fish (Amores et al. 1998) and *Xenopus laevis* (Session et al. 2016). Aside from multicellular organisms, WGD events have been well-documented in unicellular eukaryotes such as protozoans and yeast (Aury et al. 2006; Marcet-Houben and Gabaldón 2015).

WGD is associated with a number of novelties, including increases in body size (Walsh and Zhang 1992; Otto and Whitton 2000) and an expansion of transposable elements (TEs) after WGD has occurred (Marburger et al. 2018). However, despite the advances in genomics since the effects of WGD were first hypothesized, the importance of WGD in driving evolution is still controversial, with some believing it leads to an evolutionary dead-end. Evaluating the impact of WGD events on genes and evolution is difficult for a number of reasons. For instance, the most reliable indication of WGD driving evolution is the identification of neofunctionalized genes, which have been recorded in some fishes (Zakon et al. 2006; Moriyama et al. 2016). However, neofunctionalization is hard to identify as, after

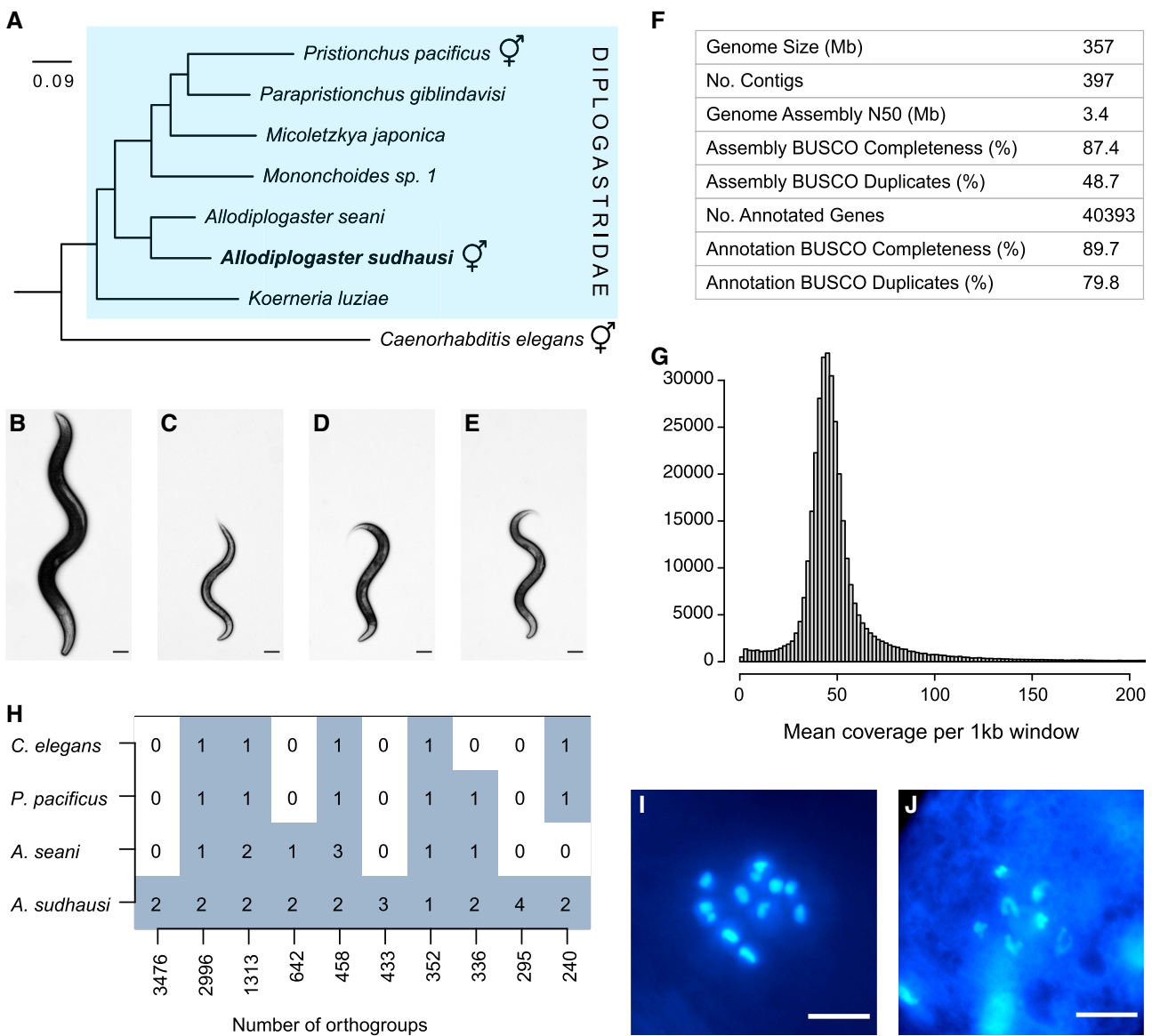
much divergence, the gene may no longer be similar enough to be identified as ohnologs derived from WGD. Thus, the rarity and antiquity of recorded WGD events have prevented full documentation of their impact. Additionally, it is difficult to evaluate WGD in organisms with large complex genomes and limited methods of experimental manipulation.

Nematodes are a useful group of organisms for characterizing genome biology. With their small genome sizes, easy maintenance (in the case of free-living nematodes) and available genetic tools, nematodes are an ideal system to characterize evolutionary and genetic processes. Additionally, there have been multiple evolutionary transitions toward self-fertilizing hermaphroditism, which have created isogenic study systems. Many of these hermaphroditic organisms also produce functional males, which enables genetic crosses (Avisé 2011). *Caenorhabditis elegans* is already a well-established model system, while *Pristionchus pacificus* has recently been developed as a model for evolutionary developmental biology (evo-devo) and the evolution of novel traits (Sommer 2015; Schroeder 2021).

Here, we present the nematode *Allodiplogaster sudhausi*, a hermaphrodite in the same family (Diplogastridae) as *P. pacificus* that displays phenotypic plasticity in the form of a mouth-form polyphenism (Fürst von Lieven 2008; Kanzaki et al. 2014; Susoy et al. 2015). It diverged early within its family, is one of the few hermaphroditic species outside of the genus *Pristionchus*, produces functional males enabling genetic crosses, and it has a sister species, the gonochoristic (male/female) *Allodiplogaster seani* (Kanzaki et al. 2015), as another more closely related point of comparison (fig. 1A). *Allodiplogaster sudhausi* is strikingly large compared to its relatives (fig. 1B), with hermaphrodites and males having body lengths one and a half times those measured in *C. elegans* (Wood 1988) and *P. pacificus* (Sommer et al. 1996) (fig. 1C and D), and is also much larger than males and females in its sister species *A. seani* (fig. 1E).

We show through karyotype analysis and whole-genome sequencing that a WGD event occurred in the lineage leading to *A. sudhausi*, which is absent in *A. seani*. Subsequent analysis uncovered a number of findings. First, we characterized the retained ohnologs and those that likely underwent nonfunctionalization to identify protein domains and metabolic pathways that are more likely to be retained after WGD. Second, we showed a link between sequence and expression divergence. Third, we identified a vast abundance of repeat elements, including the very recent expansion of the DNA transposon hAT-Ac family. Lastly, we demonstrated the use of CRISPR/Cas9 in *A. sudhausi* for the first time. Specifically, we generated mutant knockouts of a common nematode marker gene and show that both ohnologs need to be knocked out to





**FIG. 1**—Whole genome duplication in *A. sudhausi*. (A) Phylogenetic relationship of selected Diplogastridae, with *C. elegans* as an outgroup. Hermaphroditic (androdioecious) species are indicated by ♂. The hermaphroditic *A. sudhausi* is an early diverging lineage. The phylogeny represents a subtree from Susoy et al. (2015). (B–E) Young adult hermaphrodites and female (*A. seani*): (B) *A. sudhausi*, (C) *C. elegans*, (D) *P. pacificus*, (E) *A. seani*. Scale bar: 100 μM. (F) Evaluation of the *A. sudhausi* genome assembly and the resulting size, contig number, N50 as well as the number of predicted gene annotations. BUSCO analysis of the genome assembly and annotations is shown, with duplication rates very high. (G) Coverage analysis of the *A. sudhausi* genome (see Methods) shows a single peak, suggesting the high duplication values are not due to allelism. (H) Orthology clustering using OrthoFinder based on *A. sudhausi*, *P. pacificus*, and *C. elegans* annotations and the *A. seani* transcriptome. The heatmap shows the 10 most abundant orthogroups that include *A. sudhausi* orthologs. *Allodiplogaster sudhausi* has two copies in most orthogroups, with many orthogroups specific to *A. sudhausi* alone. (I) Hoechst staining of the *A. sudhausi* hermaphrodite oocyte shows 12 chromosomes, double the number previously reported. Due to the chromosomes being in different planes, a maximal intensity measurement was used to show all of them at once. Two of the chromosomes are stacked on top of another (shared z axis). Scale bar: 10 μM. (J) Hoechst staining of the *A. seani* female oocyte shows seven chromosomes.

generate a phenotype; that is, they display genetic redundancy. Overall, we present a novel nematode system with a relatively small manageable genome that can be functionally evaluated using CRISPR technology. These features allow us to examine a recent WGD duplication and DNA transposon expansion.

## Results

### The *A. sudhausi* Genome Contains an Exceptionally High Number of Duplicated Genes

We sequenced an inbred line of *A. sudhausi* (SB413B) using PacBio long-read sequencing, resulting in 2 × 11 Gb of

sequencing data from two SMRT cells. We assembled the genome de novo with the Canu assembler (Koren et al. 2017) and obtained approximately 60x coverage. We estimated a genome size of 357 Mb, more than double the 159 Mb genome of *P. pacificus* (Rödelsperger et al. 2017) and larger than the reported genome sizes of other Diplogastriidae, which range from 143 to 297 Mb (Prabh et al. 2018). We obtained a BUSCO (Simão et al. 2015) completeness value of 87.4%, in line with previous de novo diplogastriid assemblies (Prabh et al. 2018). However, we observed an extremely high duplication rate of 48.7% (fig. 1F), which dwarfs the 1.3% duplicate estimate in *P. pacificus* (Rödelsperger et al. 2019). A high duplication rate is also seen in the Illumina sequencing data generated by Sieriebriennikov et al. (2018) (supplementary Table S1, Supplementary Material online). High duplication values can sometimes be due to heterozygosity; however, this is unlikely in hermaphrodites (Barriere et al. 2009) and the SB413B strain had been extensively inbred to limit such effects. In addition, we examined the coverage of the contigs and found only a single peak (fig. 1G), indicating the regions are not allelic.

Next, we annotated the *A. sudhausi* genome and obtained 40,393 evidence-based gene models that are either supported by transcriptomic data or *P. pacificus* protein homology (Rödelsperger 2021). When we applied the OrthoFinder software to group orthologs in *A. sudhausi*, *A. seani*, *P. pacificus*, and *C. elegans* based on sequence similarity, we found that *A. sudhausi* had two gene copies in the majority of orthologous clusters (7 out of 10) (fig. 1H). Altogether, these findings provide evidence for a large-scale duplication event in *A. sudhausi*, which is in agreement with previously reported gene duplication events (Sieriebriennikov et al. 2018; Biddle and Ragsdale 2020). We, therefore, wanted to determine if WGD had taken place.

### Karyotype Analysis Confirms a Whole Genome Duplication Event in *A. sudhausi*

A WGD (polyploidization) event leads to an instant doubling of the chromosomal number, meaning a higher chromosome number in *A. sudhausi* would reliably indicate that WGD had occurred. Previous work suggested that *A. sudhausi* has six chromosomes (Fürst von Lieven 2008), the same as in *C. elegans* (Wood 1988) and *P. pacificus* (Sommer et al. 1996). However, since our analyses suggested a potential WGD event, we repeated karyotype analysis in *A. sudhausi* by staining the gonads of hermaphrodites using Hoechst 33342 dye. Strikingly, we counted 12 chromosomes in the hermaphrodite oocytes during diakinesis (fig. 1I), double the amount seen in *C. elegans* and *P. pacificus*. A recent catalog of chromosome numbers in nematodes revealed that the majority of

investigated species have six or seven chromosomes, with some species exhibiting even smaller numbers after chromosome fusions (Gonzalez de la Rosa et al. 2021; Carlton et al. 2022). Thus, the observation of 12 chromosomes in *A. sudhausi* shows that WGD has taken place in the lineage leading to this species.

### The Whole Genome Duplication is Specific to *A. sudhausi*

To ascertain when in evolution the WGD occurred, we examined the sister species of *A. sudhausi* that was recently described as *A. seani* (Kanzaki et al. 2015) (fig. 1A and E). Staining of female oocytes showed there are seven chromosomes in *A. seani* (fig. 1J). This chromosome count is similar to the numbers observed in many nematode species (Carlton et al. 2022), thereby suggesting that the WGD event occurred after the divergence of these two *Allodiplogaster* species. Ortholog clustering analyses further support the notion that the WGD occurred after the split of *A. sudhausi* and *A. seani*. Specifically, orthology clustering revealed that for 7 out of 10 orthogroups *A. sudhausi* has two-gene clusters (fig. 1H). The two biggest clusters, which have 3,476 and 2,996 orthogroups, are two-gene clusters for *A. sudhausi*, in which *A. seani* has zero and one corresponding orthologs, respectively. Finally, the BUSCO duplication values are also lower in *A. seani* than in *A. sudhausi* (supplementary Table S1, Supplementary Material online). Overall, these results indicate a WGD that is specific to *A. sudhausi*. This finding is compelling as it presents us with a hermaphroditic organism in which to characterize various processes related to WGD, such as the fates of WGD-derived duplicate genes (henceforth referred to as ohnologs). Note that there is unfortunately only one isolate of *A. sudhausi* available to date. Thus, it remains unknown if the WGD is fixed in this species, and if other, more closely related species (if they exist) would share the WGD event.

### Comparative Analysis Shows Expansions of GPCR and Ribosomal Domains in *P. pacificus* Relative to *A. sudhausi*

We next examined the predicted Pfam protein domains in *A. sudhausi*, *A. seani*, *P. pacificus*, and *C. elegans* (the latter three have no indication of a recent WGD). The frequency of unique domain predictions for each gene was compared between species. We found an increase in the protein binding domains ankyrin (ANK) and Broad-Complex, Tramtrack and Bric-a-brac (BTB) in *A. seani* compared to *A. sudhausi* (fig. 2A). BTB domain-containing proteins are adaptors involved in protein degradation, which show signatures of positive selection in *C. elegans* (Thomas 2006). They have been found to be overrepresented of gonochorists in both *Caenorhabditis* and *Pristionchus* nematodes due to

gene loss in hermaphroditic species (Rödelsperger et al. 2018; Yin et al. 2018). Thus, the overrepresentation in gonochoristic *A. seani* compared to *A. sudhausi* is consistent with previous results.

Surprisingly, we found domains in seven-transmembrane G-protein-coupled receptor class (7TM GPCRs) are disproportionately under-represented in *A. sudhausi* relative to *P. pacificus* (fig. 2B). A comparison with *C. elegans* again shows that 7TM GPCRs are also highly under-represented in *A. sudhausi* (supplementary fig. S1, Supplementary Material online). 7TM GPCRs mediate chemoreception in *C. elegans* where they play important roles in sense and stimuli (Troemel et al. 1997). Due to the abundance of these domains in *C. elegans* and *P. pacificus*, which are separated by large evolutionary distances (fig. 1A), we assumed they would be abundant in most free-living nematodes. However, a comparison between the number of domains with 7TM GPCRs across many nematodes shows 7TM GPCRs are only highly overrepresented in *C. elegans* and *Pristionchus* species (supplementary fig. S2, Supplementary Material online). Thus, the high abundance of 7TM GPCRs is not evolutionarily conserved across all free-living nematodes, and rather evolved independently in *Caenorhabditis* and *Pristionchus* species.

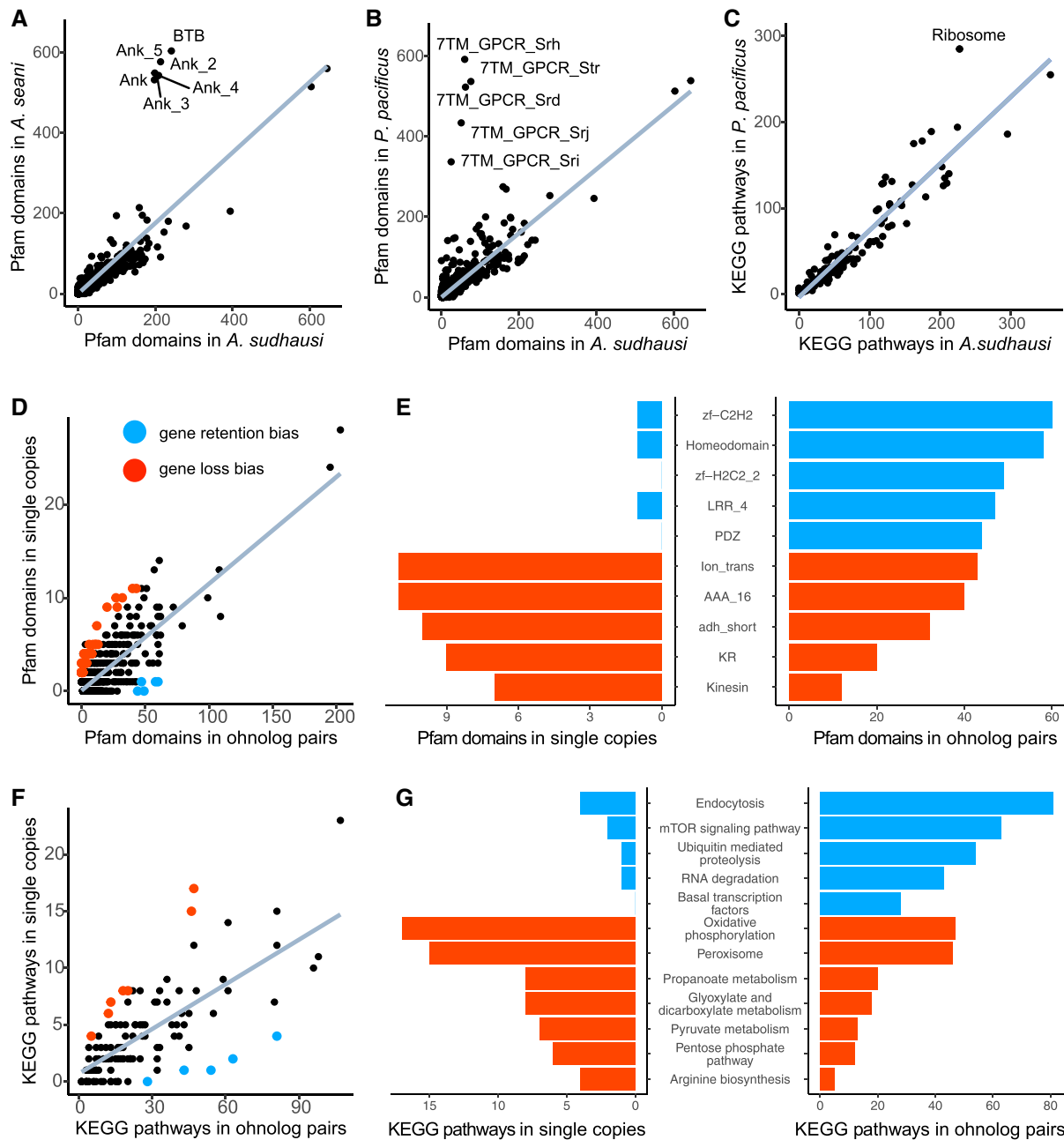
Next, we predicted KEGG (Kanehisa and Goto 2000) metabolic pathways and compared the abundance between species. There appears to have been an expansion of genes involved in the Mitogen-activated protein kinase (MAPK) pathway (04010) in *A. seani* (supplementary fig. S3, Supplementary Material online). MAPK is involved in cellular processes such as proliferation, differentiation, and development (Seeger and Krebs 1995), as well as oxidative stress response in *C. elegans* (Inoue et al. 2005). Interestingly, we found an expansion of ribosomes (03010) in *P. pacificus* relative to *A. sudhausi* (fig. 2C). A greater abundance of ribosomal pathways in *P. pacificus* relative to *C. elegans* had previously been observed (Dieterich et al. 2008). By additionally examining the Pfam domain predictions across many Diplogastridae, we saw that ribosomal domains are highly over-represented in *P. pacificus* relative to others, including fellow *Pristionchus* species (supplementary fig. S4, Supplementary Material online). This finding suggests a recent expansion of ribosomal pathways and domains may have taken place in *P. pacificus*.

#### Ohnologs With Homeodomains and Those Involved in the mTOR Signaling Pathway are More Likely to be Retained

The young nature of the WGD event allowed us to examine potential functional biases between ohnologs pairs that had been lost or retained in *A. sudhausi*. We identified

ohnologs that were retained and those that lost their duplicate based on the orthology clustering analysis described above (fig. 1H). For the purposes of this analysis, we deem orthogroups with two copies ohnolog pairs, and orthogroups with one single-copy genes. We hypothesize that these single-copy genes denote orthogroups that lost their ohnologous duplicate; therefore, a case of nonfunctionalization (the loss of one duplicate). We then predicted their respective Pfam domains and determined which domain-containing genes were more likely to either be retained as ohnologs or lose their duplicate. We found only five domains that were more likely to be retained (Fisher's exact test, FDR adjusted  $P < 0.05$ ) (fig. 2D and E and supplementary Table S2, Supplementary Material online), including zf-C2H2 (PF00096) and Homeodomain (PF00046), potentially suggesting these domains are more dosage sensitive or haploinsufficient. The retention of homeodomains is noteworthy as they are transcribed by *Homeobox* genes, which have remained after WGD in ray-finned fishes and play important roles (Amores et al. 1998; Blomme et al. 2006). Notably, only one homeodomain-containing gene lost its duplicate in *A. sudhausi* (fig. 2E). We took this gene, which also has a predicted LIM domain, and identified its *C. elegans* ortholog using the orthology clustering data (fig. 1H). The *C. elegans* ortholog is *lim-4*, a LIM homeobox gene that has been found to specify olfactory neurons (Sagasti et al. 1999). This finding might suggest that this process may not be as conserved in *A. sudhausi*, reflecting the previous results that showed surprisingly few genes with GPCR domains in *A. sudhausi* (supplementary fig. S1, Supplementary Material online). In contrast to the small number of significantly retained domains, we found 82 domains that were significantly more likely to lose their ohnologous duplicate (fig. 2D and E and supplementary Table S3, Supplementary Material online), with an abundance of domains that act as enzymes, specifically ATPases, and those involved in transportation and microtubule binding. We hypothesize there is low dosage sensitivity in genes that contain these domains.

We repeated this analysis using KEGG pathway predictions and found five pathways significantly more likely to be retained as ohnologs (Fisher's exact test, FDR adjusted  $P < 0.05$ ), including endocytosis (04144) and the mTOR signaling pathway (04150) (fig. 2F and G and supplementary Table S4, Supplementary Material online). The mTOR pathway is a regulator of cell growth and proliferation that is associated with cancers (Sarbasov et al. 2005). There were seven pathways significantly more likely to lose their ohnologous duplicate, with most involved in metabolism, including the oxidative phosphorylation pathway (00190) which is involved in ATP synthesis (Wilson 2017) (fig. 2F and G, supplementary Table S5, Supplementary Material online). Interestingly, ohnologs more likely to be retained consisted of many pathways with regulatory roles, and previous



**Fig. 2**—Functional bias in ohnolog loss and retention. (A) A comparison of Pfam protein domain predictions in *A. sudhausi* and *A. seani* indicate a cluster of protein binding domains, ANKs and BTB, are underrepresented in *A. sudhausi* after the species diverged. A linear regression trendline is shown. (B) A comparison of Pfam protein domain predictions shows an abundance of 7TM\_GPCR chemoreceptors in *P. pacificus* compared to *A. sudhausi*. (C) A comparison of KEGG metabolic pathway predictions in *A. sudhausi* and *P. pacificus* indicate there has been a vast increase in ribosomes (pathway 03010) in *P. pacificus*. (D) The plot shows the frequency of predicted Pfam domains between orthogroups that either have two copies (putative ohnologs) or 1 copy (candidates that lost their ohnolog copy). Significant domains (Fisher’s exact test, FDR adjusted  $P < 0.05$ ) are highlighted. More domains are significantly more likely to lose than retain a copy (counts of 82 and 5, respectively). (E) The barplot shows a subset of significant Pfam domains and their abundance in single copy genes (left) and putative ohnologs (right). Domains most likely to lose an ohnolog copy are involved in processes such as transport and binding. The five domains predicted to remain as ohnologs, such as zf-C2H2 and homeodomains, may be dosage sensitive. (F) The plot shows the frequency of KEGG pathway predictions between orthogroups that either have two copies (putative ohnologs) or one copy (candidates that lost their ohnolog copy). Significant domains (Fisher’s exact test, FDR adjusted  $P < 0.05$ ) are highlighted. (G) The bar plot shows the significant KEGG pathways and their abundance in single copy genes (left) and putative ohnologs (right). Genes with pathways involved in metabolism appear more likely to lose an ohnologous copy, while genes with pathways involved in processing are more likely to remain as ohnologs.



studies have suggested regulatory genes are more dosage sensitive (Birchler et al. 2005). Conversely, single copy genes were mostly involved in metabolism. This reflects work done by Shiu et al. (2006) where they showed genes involved in metabolism were less likely to be retained after duplication in mammals. Overall, we show which protein domains and metabolic pathways show functional bias in being lost or retained after WGD.

### Ohnologs in *A. sudhausi* are Evolutionarily Constrained and Show Low Sequence Divergence

In order to date the *A. sudhausi* WGD and characterize the evolutionary distance to *A. seani*, we calculated divergence measures based on the orthogroup data. dN and dS estimates were calculated for ohnologs, as well as for orthologous genes that have copies in *A. seani* (as no second strain of *A. sudhausi* is currently available). The median dS for ohnologs is a relatively low value of 0.12 (interquartile range: 0.08 to 0.18), while the median dS for orthologs is 3.17 (fig. 3A). These findings imply that *A. sudhausi* and *A. seani* diverged a long time ago. We estimated the timing of the WGD by referring to Cutter (2008), where dS values were used to time events. The interquartile range of 0.08 to 0.18 would roughly correspond to an event that happened somewhere between 1.3 to 3.3 million years ago. Note that this timing of the WGD would only be in the case of autopolyploidization (where the entire chromosome set is duplicated). If allopolyploidization (species hybridization) had taken place, the dS values would not help in timing the event.

Finally, we calculated the dN/dS values for ohnologs and found the vast majority are below 1, with a median dN/dS value of 0.16. This indicates that most ohnologs are evolutionarily constrained (fig. 3B). We obtained the ohnolog dN/dS values in the top 5%ile. This subset had values of 1.35 at minimum, suggesting that they might underly positive selection. Interestingly, the majority of these ohnologs did not have associated Pfam predictions for them. Specifically, of the 397 ohnolog pairs, only 9 unique domains were predicted (supplementary Table S6, Supplementary Material online). We then looked at the ohnolog pairs with the highest dN/dS values and found no matching *C. elegans* orthologs (supplementary Table S7, Supplementary Material online), indicating strong sequence divergence in the ohnologs that are positively selected for.

### A Third of Ohnologs are Differentially Expressed and Overlap With Sequentially Diverged Ohnologs

Next, we characterized the expression of ohnologs by calculating the fragments per kilobase of transcript per million mapped reads (FPKM) based on transcription analysis. First, we found that for most ohnologs both copies were indeed

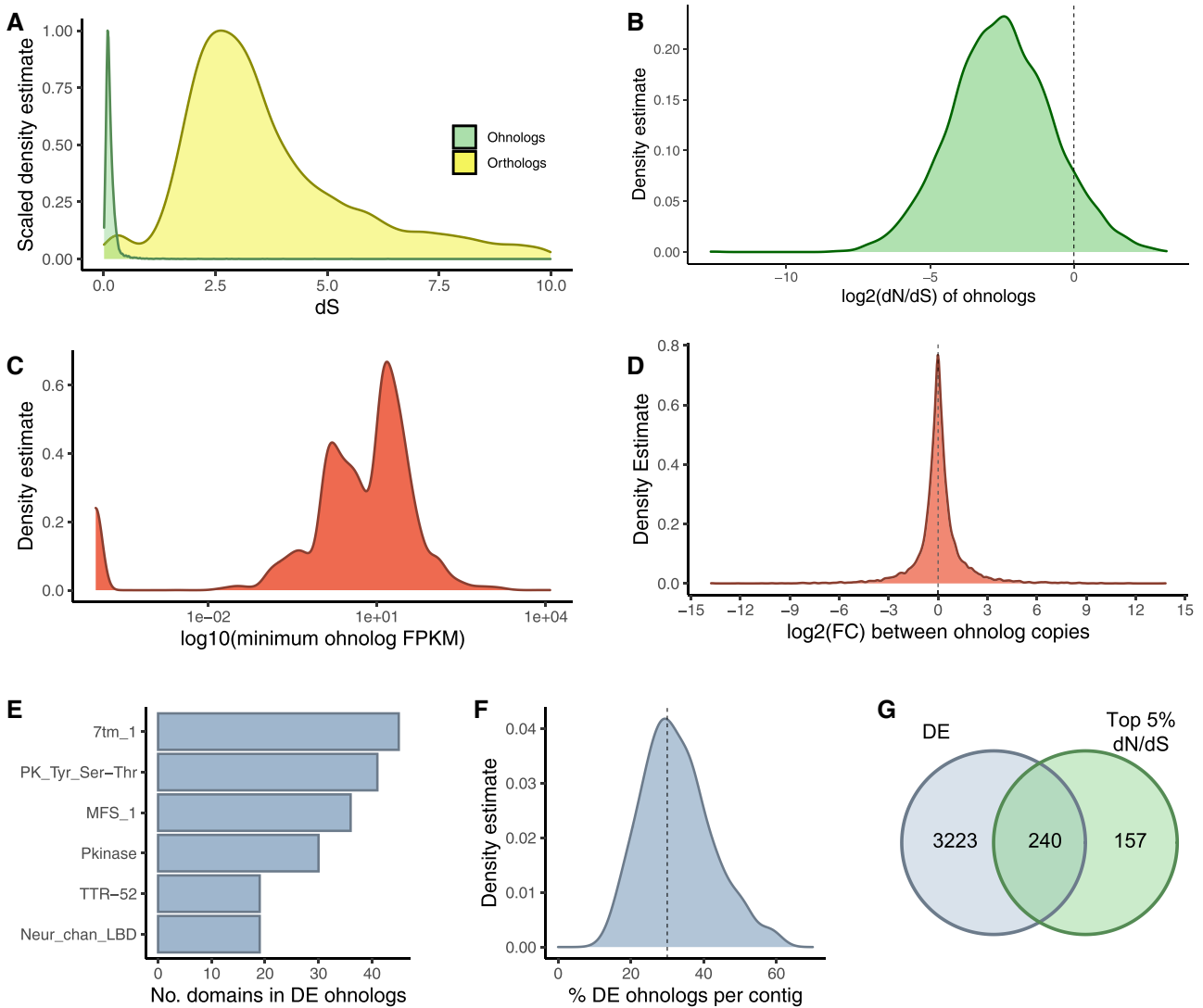
expressed (fig. 3C). Second, when we examined fold change (FC) between ohnolog pairs we found that 30.4% of ohnologs had expression that had a more than two-fold difference in transcript abundance (absolute  $\log_2$ -FC)  $\geq 1$ ). However, the majority of ohnolog pairs have little to no difference in expression, with a peak at 0 (fig. 3D). Thus, there is little expression difference for ohnologs overall, a finding that is consistent with the multiple lines of evidence suggesting that the WGD occurred recently and that there has not been enough time for most ohnologs to diverge. Finally, we predicted the Pfam domains of these differentially expressed (DE) ohnologs, with the 7tm\_1 domain showing the highest abundance (fig. 3E). We additionally looked at the ohnologs with the highest DE values and found the greatest difference was from ohnologs orthologous to the *C. elegans* ribosomal protein *rps-30* (supplementary Table S8, Supplementary Material online).

Expression divergence between ohnologs could potentially be explained by gene-specific regulatory evolution or by silencing of large chromosomal segments, as seen in dosage compensation (Pala et al. 2008). We, therefore, examined if there was positional bias by comparing the percentage of DE ohnologs for each contig. If there had been no bias, we would expect the distribution of DE ohnologs per contig to fall around 30.4%, which is the DE value calculated genome-wide. Indeed, the actual values were fairly evenly distributed around that figure (fig. 3F), with a median of 31.3% calculated. Thus, there is no evidence of ohnolog positional bias, suggesting expression divergence is more likely to be caused by gene-specific evolution.

Finally, we determined if there was overlap between the DE ohnologs and the ohnologs in the top 5 percentile dN/dS. Indeed, we found that the majority of ohnologs with high dN/dS values were also differentially expressed (fig. 3G). Specifically, there were more high dN/dS ohnologs that were differentially expressed than not (240 out of 397). This translates to 60.5% of the top dN/dS ohnologs being differentially expressed, which is significantly higher than the overall ohnolog median of 30.4% (Fisher's exact test, FDR adjusted  $P < 0.001$ ). Altogether, this analysis reveals a link between expression and sequence divergence in ohnologs. This could be explained by positive selection on function and gene dosage but could also represent degeneration and silencing of one copy.

### Repeat Elements, Particularly DNA Transposons, are Abundant in *A. sudhausi*

WGD is often succeeded by the expansion of TEs (Lien et al. 2016; Blanc-Mathieu et al. 2017; Marburger et al. 2018). We, therefore, analyzed repeat elements in *A. sudhausi* and compared it to those found in *P. pacificus* and *C. elegans*. Libraries were generated for *A. sudhausi*,



**FIG. 3**—Sequence and expression divergence in *A. sudhausi*. (A) The density plot shows the divergence (dS) of orthologs (between *A. sudhausi* and *A. seani*) and ohnologs. Orthologs exhibit high divergence, with a median value of 3.17 indicating that *A. seani* is distantly related. *Allodiplogaster sudhausi* ohnologs show much lower divergence (median of 0.12), suggesting the WGD is recent. (B) The distribution of dN/dS in ohnologs is relatively low, with the majority below 1 (indicated by the dashed line at 0 on the x axis log<sub>2</sub> scale) and a median dN/dS value of 0.16, indicating most ohnologs are evolutionarily constrained. (C) The distribution of the lowest FPKM value of each ohnolog pair shows the majority of ohnologs exhibit expression of both copies. Those ohnologs with no expression are shown on the left of the plot, with a pseudocount of log<sub>10</sub> 10<sup>-4</sup>. (D) Distribution of expression fold changes between ohnolog copies indicates the majority have no difference in expression. Only approximately 30.4% of ohnologs have an absolute log<sub>2</sub>(fold change) ≥ 1, which would indicate 1 copy has double or more expression than the other. (E) The bars show the most abundant Pfam domain predictions in the DE ohnologs (those with absolute log<sub>2</sub>(fold change) ≥ 1). (F) The distribution of the DE ohnologs on the contigs (only including contigs containing ≥ 10 ohnologs) shows no evidence of positional bias. The expected peak of 30.4% (indicated by the dashed line), correlates with the overall distribution of the DE ohnologs. (G) The Venn diagram displays the overlap between the top 5% highest dN/dS ohnologs and the DE ohnologs (absolute log<sub>2</sub>(FC) ≥ 1). The majority of the high dN/dS ohnologs are also differentially expressed (240 out of 397).

*P. pacificus*, and *C. elegans* using RepeatModeler based on the protocol of Athanasouli and Rödelsperger (2022). Strikingly, we found a far higher proportion of the *A. sudhausi* genome was covered by spans of repeats compared to the other two species. Specifically, 36.3% of the *A. sudhausi* genome is covered by repeat elements, which is far higher than the values of 21.3% in *P. pacificus* and

13.7% in *C. elegans* (fig. 4A). The high repeat rate may thus be linked to the WGD.

Further analysis revealed that DNA transposons are by far the most abundant type of repeats in *A. sudhausi* (fig. 4B). There are nearly 40Mb of DNA transposons in *A. sudhausi*, which is almost half the total classified repeats (87.5 Mb) (fig. 4A). DNA transposons, therefore, drive the

large repeat abundance in *A. sudhausi*. They are also the largest classification of repeats in *C. elegans* (fig. 4B); in contrast, the amount of DNA transposons is small in *P. pacificus*, despite them belonging to the same nematode family as *A. sudhausi*. This finding is not altogether unsurprising as there has been shown to be large diversity in transposon superfamilies even between strains in *C. elegans* (Laricchia et al. 2017).

We further broke down the families of the classified transposon types in each species and discovered that DNA/hAT-Ac elements take up the vast majority of repeat sequences in *A. sudhausi* (fig. 4B). The extensive amount of DNA transposons can be inferred to largely be due to hAT-Ac. The hAT superfamily, under which the Ac family belongs, is ancient and found in plants, animals, and fungi (Rubin et al. 2001; Wicker et al. 2007). It includes one of the first transposons ever discovered, the Ac element (McClintock 1950). DNA/hAT-Ac covers 3% of the *A. sudhausi* genome alone (fig. 4D); however, its numbers are relatively low in *P. pacificus* and *C. elegans*, indicating the expansion of hAT-Ac is specific to *A. sudhausi*. We performed pairwise comparisons of the *A. sudhausi* DNA/hAT-Ac elements to better determine when expansion of this family occurred. We found a peak in the percent identity for the comparisons at 100% (fig. 4E), suggesting the expansion is extremely recent. Taken together, our analysis of repeat elements revealed a high abundance in *A. sudhausi*, with a strong overrepresentation of DNA transposons.

### Two *A. sudhausi dpy-1* Ohnologs Formed After WGD

WGD as observed for *A. sudhausi* can provide important insight into genome evolution; however, WGD can also limit functional investigation through forward and reverse genetic approaches. CRISPR/Cas9 is now a well-established molecular technique that enables the introduction of mutations into targeted loci (Jinek et al. 2012). We selected the *dpy-1* gene in order to establish CRISPR technology in *A. sudhausi* because the *dpy-1* gene is highly conserved and has an easy-to-score mutant phenotype that is shared in both *P. pacificus* and *C. elegans* (Kenning et al. 2004; Witte et al. 2014). In general, *Dumpy* (*Dpy*) mutants are shorter than wild type and more than 30 genes of *C. elegans* have been described that result in a *Dpy* phenotype when mutated (Brenner 1974).

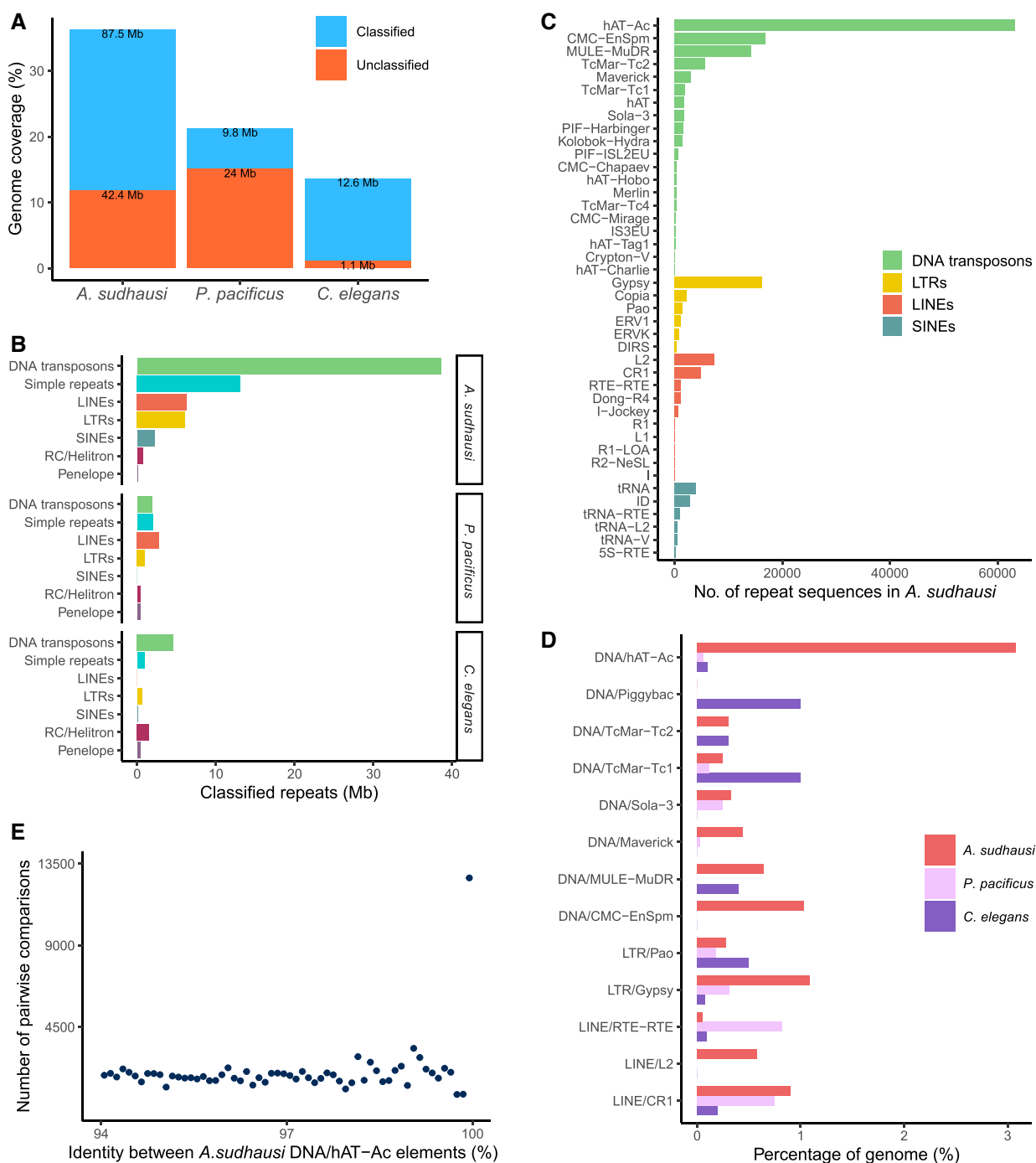
BLAST searches revealed the conservation of *dpy-1* in many species, including members of the Rhabditidae (containing *C. elegans*) and Diplogastridae (containing *A. sudhausi* and *P. pacificus*) (fig. 5A). Based on the phylogeny of orthologous *dpy-1* sequences, it is clear there is one *dpy-1* copy ancestrally. Notably, and unsurprisingly, *A. sudhausi* has two *dpy-1* genes which diverged very recently, presumably ohnologs that resulted from the WGD. We

annotated the gene structure of these ohnologs based on the transcriptome analysis (which was indexed against the genome) and termed them *Asu-dpy-1-A* and *Asu-dpy-1-B* in agreement with the previous nomenclature. The make-up of these ohnologous genes is very similar, with both containing 15 putative exons (fig. 5B). Overall, the low phylogenetic evolutionary distance and similar genetic structure suggest a very recent duplication event, consistent with the above analysis.

### CRISPR/Cas9-induced *A. sudhausi dpy-1* Knock-out Mutants Demonstrate Genetic Redundancy

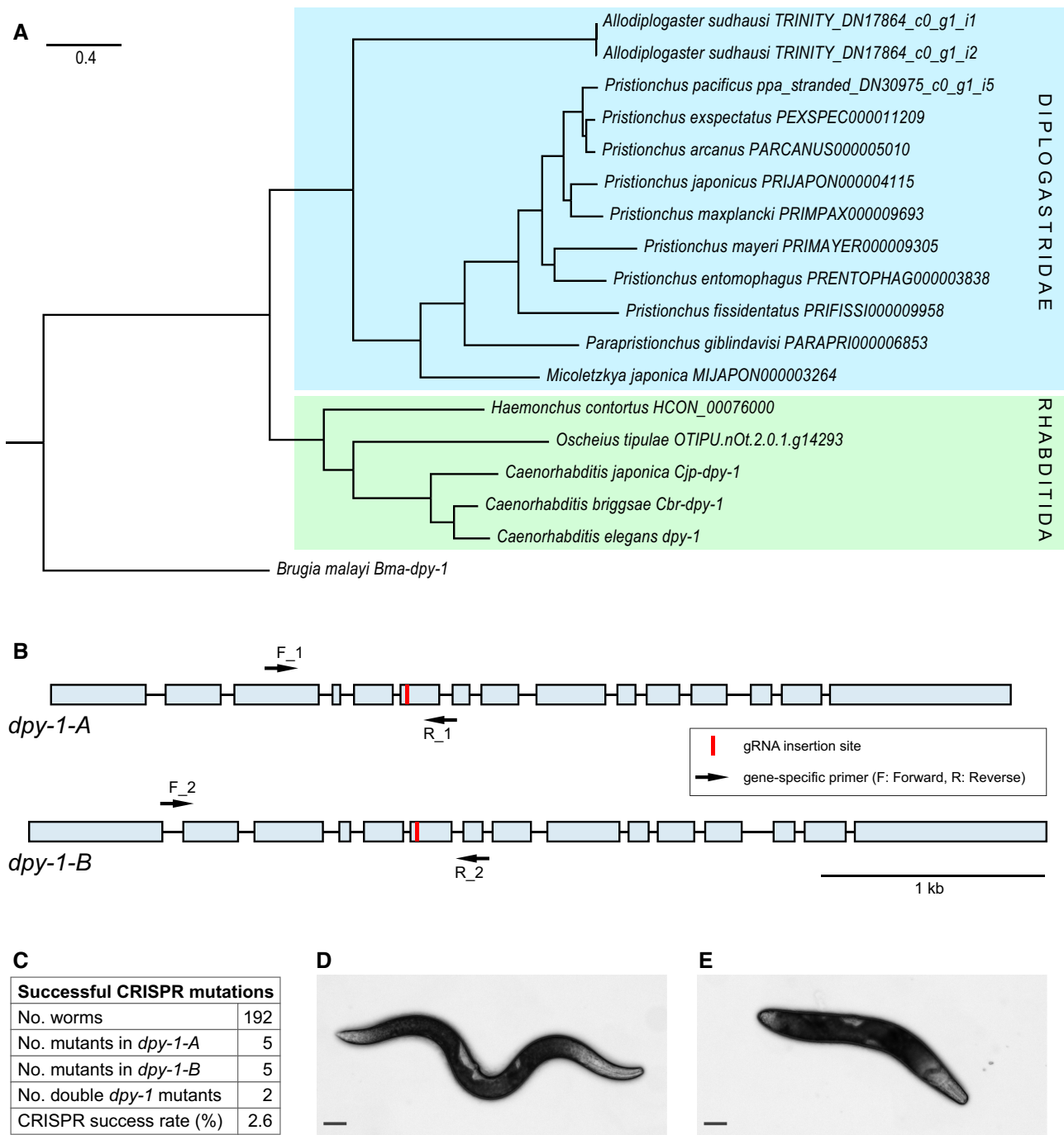
CRISPR/Cas9 technology was recently optimized for *P. pacificus* (Han et al. 2020; Hiraga et al. 2021), but nothing has previously been shown in *A. sudhausi*. We attempted to acquire knock-out mutations in *A. sudhausi* following the CRISPR/Cas9 protocol in *P. pacificus* (Witte et al. 2014), and targeted *dpy-1* ohnologs. We chose a gRNA target sequence that was conserved between the *Asu-dpy-1* ohnologs to try and target both genes together (fig. 5B). Initially, 30 young adult hermaphrodites were injected with the gRNA. We sequenced the resulting F1 progeny using gene-specific primer pairs to determine what mutations may have resulted. In total, 192 F1 worms were sequenced with two separate primer pairs, resulting in 384 sequences. Of these, 10 resulted in mutations in the targeted genes, with a mutation success rate of 2.6% (fig. 5C), where the majority of mutants come from just one injected adult (supplementary Table S9, Supplementary Material online). We also managed to acquire double mutants from a single injection, indicating that both loci can be targeted simultaneously. Together, we demonstrate the first successful example of CRISPR/Cas9 knockouts in a potential new nematode model system.

We found single *dpy-1* mutants displayed no obvious difference in length compared to wild-type worms (figs. 1B and 5D). However, the two mutant lines with knockouts in both *dpy-1* ohnologs displayed the characteristic dumpy phenotype with a short body length (fig. 5D). Interestingly, one double mutant had only in-frame mutations but this was still sufficient to generate a *Dpy* phenotype (supplementary Table S9, Supplementary Material online). Overall, our results shows expression of both *dpy-1* ohnologs needs to be disrupted in order to generate a morphological phenotype, suggesting one gene alone is sufficient for the wild-type phenotype to be produced. This is an example of genetic redundancy, which is common in recent duplication events (Lynch and Conery 2000) and is in line with our expression analysis that showed similar ohnolog expression levels (fig. 3D). Indeed, we confirmed there was no significant difference in expression between the *dpy-1* ohnologs (supplementary fig. S5, Supplementary Material online). With this work we provide a model to



**Fig. 4**—Increased transposon activity in *A. sudhausi*. (A) The comparison of classified and unclassified spans of repeats in *A. sudhausi*, *P. pacificus*, and *C. elegans* reveals that the percentage of genome coverage is greatest in *A. sudhausi*, which also has the greatest overall repeat content (42.4 Mb unclassified and 87.5 Mb classified repeats). (B) The comparison of the types of classified repeat elements between *A. sudhausi*, *P. pacificus*, and *C. elegans* demonstrates a strikingly high amount of DNA transposons in *A. sudhausi*. (C) The number of repeat sequences in transposon families (as classified by RepeatModeler) in *A. sudhausi* is dominated by the hAT-Ac family (a DNA transposon element). DNA/CMC-EnSpm, DNA/MULE-MuDR, LINE/L2 and LTR/Gypsy are also overrepresented. (D) A comparison of transposon families between *A. sudhausi*, *P. pacificus*, and *C. elegans* (only looking at families that spanned  $\geq 1$  Mb in one species) shows that LINE/L2 and DNA/CMC-EnSpm appear unique to *A. sudhausi* while Piggybac is not found in either *P. pacificus* or *A. sudhausi*. DNA/hAT-Ac is found in all three species but is highly overrepresented in *A. sudhausi*. (E) The plot shows the percentage identity distribution of pairwise comparisons (blastn searches) between DNA/hAT-Ac elements. The peak toward 100% suggests that this transposon burst happened very recently and most likely after the WGD (in comparison with the dS value of 0.12 which translates to an expected identity of 88%).





**FIG. 5**—CRISPR knockouts of *dpy-1* ohnologs demonstrate genetic redundancy in *A. sudhausi*. (A) The phylogenetic tree shows DPY-1 orthologues in the order Rhabditida. Protein sequences were obtained by BLASTing *C. elegans* DPY-1 against other nematodes. The majority of species have a single DPY-1 ortholog. *Allodiplogaster sudhausi* has two DPY-1 copies that only very recently duplicated, suggesting they are ohnologs that resulted from the WGD. (B) The figure shows the gene structure of *A. sudhausi* *dpy-1* ohnologs based on the transcriptome. Exons are shown as blocks and introns as connecting lines. The gRNA sequence is shared by both ohnologs and located in the sixth exon. The gene structure and number of exons (15) is similar for both ohnologs, showing the similarities that remain after whole genome duplication. (C) Table showing the output of CRISPR knock-out attempts. Out of 192 initial F1 (where P0 has been injected with the CRISPR/Cas9 gRNA), there were five knock-out mutations in *dpy-1-A* and five in *dpy-1-B*. In two cases, there were double mutants wherein both ohnologs had mutations. The overall successful CRISPR rate was 2.6%. (D) Single mutant knock-out *dpy-1-A* adult hermaphrodite. The nematode has a wild-type phenotype. Scale bar: 100  $\mu$ M. (E) Double mutant knock-out *dpy-1-A* and *dpy-1-B* adult hermaphrodite. The length of the worm is noticeably smaller and dumper than the single mutant. Scale bar: 100  $\mu$ M.

examine the phenotypes of recent duplication events, particularly in WGD, and to further functionally examine the fates of these duplicate genes using CRISPR/Cas9.

## Discussion

The role and consequences of WGD for driving evolution are still considered controversial, largely due to two important limitations. First, there are a limited number of organisms available for analysis, and second, most WGD events are of distant age. In the nematode *A. sudhausi*, we discovered WGD took place recently as did an expansion of the DNA transposon hAT-Ac family. With these findings, we were able to make several inferences about the early effects of WGD.

First, *A. sudhausi* differs from its relatives not just by the presence of a WGD, but by its (relatively) gigantic body size (fig. 1B–E). We hypothesize the large body size is due to the WGD that is specific to *A. sudhausi*. WGD is known to have an effect on body size although this differs depending on the lineage (Otto and Whitton 2000). In plants, it sometimes has an effect, but given the ancestral nature of most WGD events uncertainties remain (Clark and Donoghue 2018). In animals, the consensus is that WGD drives an increased body size in invertebrates, but not in vertebrates. Indeed, the evidence of WGD driving increased invertebrate body size is from studies long ago, where positive correlations between WGD and body size were seen in rotifers (Walsh and Zhang 1992), water fleas (Weider 1987), and even in nematodes (Madl and Herman 1979; Triantaphyllou and Riggs 1979). Flemming et al. (2000) previously showed endoreduplication drives an increased body size in nematodes. We, therefore, theorize the increased body size of *A. sudhausi* (fig. 1B–E) is due to the WGD.

Second, a benefit of evaluating a recent WGD is that the majority of genes have not yet diverged. This enabled us to determine if there was functional bias in the fate of ohnologs by examining which Pfam domains and KEGG pathways are more likely to be lost. We deemed two-copy orthogroups from *A. sudhausi* to be retained ohnologs and one-copy orthogroups to be nonfunctionalized genes. It can be argued that the lack of an identified duplicate could also be due to neofunctionalization. However, neofunctionalization is far rarer than nonfunctionalization (Moriyama and Koshiba-Takeuchi 2018). Thus, as we examined a large dataset, we feel confident the trend reflects nonfunctionalized genes. We found genes involved in metabolism, transportation, or those that encoded enzymes were significantly more likely to lose their duplicate after WGD (fig. 2E and G), which is consistent with results from other organisms. For instance, duplicate genes involved in metabolism and transport were also more likely to be lost in both humans and mice (Shiu et al. 2006). Additionally, ohnologs encoding enzymes were more likely

to lose their duplicate in *Arabidopsis thaliana* (Seoighe and Gehring 2004). Thus, there appears to be some conserved functional bias across different kingdoms, potentially due to a shared underlying molecular basis. Studies have suggested that housekeeping genes are less likely to be retained while those with regulatory functions are more likely to be retained after duplication (Birchler et al. 2005; Shiu et al. 2006). This is believed to be due to differences in haplosufficiency or dosage sensitivity (Papp et al. 2003; Kondrashov and Koonin 2004). To support this, Kondrashov and Koonin (2004) determined genes encoding enzymes are highly haplosufficient, meaning they are far less dosage sensitive and better able to cope with losing a duplicate. We, therefore, suggest that dosage sensitivity is the main driver in the functional bias of gene fate that appears to be shared across different kingdoms.

Third, WGD is often followed by an expansion of TEs in both plants (Vicent and Casacuberta 2017) and animals (Lien et al. 2016; Blanc-Mathieu et al. 2017). Along these lines, we found a striking abundance of TEs in *A. sudhausi* (fig. 4A). TEs are subdivided into two classes: Class I retrotransposons and Class II DNA transposons (Makalowski et al. 2019), with the latter being most prevalent in *A. sudhausi* (fig. 1B). Studies have shown that TEs can play important roles in regulation, with their ability to produce novel networks and genes via mobility and insertion, or by generating new splice sites (Cosby et al. 2021). They are also involved in regulatory processes (Chénais et al. 2012), including the regulation of duplicate genes (Lisch 2013; Tan et al. 2021), meaning they may play important roles after WGD. The abundance of DNA transposons in *A. sudhausi* can largely be put down to the hAT-Ac family (fig. 4C and D), which falls under the ancient hAT superfamily (Arensburger et al. 2011). DNA/hAT is found in plants, animals, and fungi (Rubin et al. 2001), and has been shown to be a driver of evolutionary events. For example, DNA/hAT contributed highly to exon shuffling to generate novel genes in tetrapods (Cosby et al. 2021). DNA/hAT is active and recently expanded in the bat genus *Myotis*, which has high plasticity and diversification (Ray et al. 2006). Additionally, a DNA/hAT element is involved in neofunctionalization in *X. laevis* (Hayashi et al. 2022). In *A. sudhausi*, we determined the DNA/hAT-Ac family expanded very recently based on pairwise analysis (fig. 4E) and infer that, if autopolyploidization led to WGD, the expansion only happened afterward. As the DNA/hAT superfamily has been shown to contribute to various evolutionary processes, they may also be modulating *A. sudhausi* after the WGD event. Interestingly, a recent study revealed that the Homeobox and zf-C2H2 domains sometimes fuse with transposons to drive novelty and evolution in tetrapods (Cosby et al. 2021). As these domains show high retention in *A. sudhausi* ohnologs (fig. 2E), they may potentially work together with DNA/hAT-Ac to drive novelty. However,

further work is necessary to investigate this and the overall role of DNA transposons in regulating WGD and driving evolutionary processes.

Finally, the *A. sudhausi* system has many advantages for genetic analysis. Firstly, it is a self-fertilizing hermaphroditic nematode system. This means it has the benefits of short generation times and easy maintenance in an isogenic study system. Secondly, it behaves like a diploid with two alleles needing to be knocked out in both genes. Diploidization, the reversion of a polyploid system back to a diploid one, is a well-known, although not well understood, phenomenon that commonly occurs following WGD (Wolfe 2001). Lastly, we have CRISPR/Cas9 available to generate mutants. With this tool, it is possible to examine the phenotypic effects of genes. This is particularly useful for examining potential differences in gene fate. For example, neofunctionalized genes could be identified using this approach.

WGD has also been recorded in other nematodes, although it does not seem to be a common phenomenon. It has been well-studied in the plant parasitic genus *Meloidiogyne*, especially in the triploid *M. incognita*, which arose via allopolyploidization (Blanc-Mathieu et al. 2017; Szitenberg et al. 2017). Interestingly there has also been an expansion of TEs following WGD in these species (Blanc-Mathieu et al. 2017). In particular, a high abundance of DNA transposons was found in *M. incognita*, which is thought to drive genomic plasticity (Kozłowski et al. 2021). Additionally, WGD has been identified in the tetraploid plant parasite *Heterodera glycines* (Triantaphyllou and Riggs 1979). Only one other free-living nematode genus has been shown to undergo WGD to our knowledge, the panagrolaimid nematodes that underwent allopolyploidization (Schiffer et al 2019). The study of WGD in all these nematodes has added to our repertoire of knowledge and provides a useful comparative approach. However, of the known nematodes that underwent WGD, *A. sudhausi* is the only diploid and the only one in which CRISPR/Cas9 tools have been optimized, making it ideal for genetic analysis.

One unanswered question is how the WGD took place; whether it was a case of auto- or allopolyploidization. In *Meloidiogyne* (Blanc-Mathieu et al. 2017) and panagrolaimid (Schiffer et al 2019) nematodes, the duplication was shown to be due to allopolyploidization. This was determined by comparing the genomes of a number of closely related species. Unfortunately, we only have one strain of *A. sudhausi* available and no other closely related sister species other than *A. seani* at hand, although we and others did multiple sampling trips to the type locality and related regions to find more strains. Additionally, *A. seani* and *A. sudhausi* diverged long ago based on the ortholog analysis (fig. 3A), meaning it is also not the best comparison. Although other species of *Allodiplogaster* have been

reported, many are hard to keep in laboratory cultures and most are not available as a living material. Without more closely related species we cannot in full confidence make inferences about the origin of the WGD. Thus, the lack of other strains and more closely related species is the biggest limitation in using *A. sudhausi*. Comparative approaches would benefit from having more closely related systems at hand in order to better characterize the effects and origin of WGD. While this is a downside, few other animal systems have such closely related species.

There are a number of interesting avenues of research that could be followed in the future. The generation of reporter lines in *A. sudhausi* would enable us to see if ohnologs show differences in where they are expressed, as sometimes happens in subfunctionalization (Force et al. 1999). Another possibility is to examine the impact of dosage sensitivity, potentially by knocking out only one ohnolog and evaluating fitness consequences. An interesting avenue of research is to determine if there is any novelty in *A. sudhausi* besides the aforementioned increase in body size. WGD potentially drives evolution (Ohno 1970), while TEs drive genomic plasticity and phenotypic change (Faino et al. 2016; Kozłowski et al. 2021). It would, therefore, be interesting to examine if any phenotypic changes have been driven by these processes, particularly as *A. sudhausi* already displays phenotypic plasticity as a mouth-form polyphenism (Fürst von Lieven 2008).

In conclusion, we show a recent WGD and DNA transposon expansion occurred in a free-living hermaphroditic nematode. We, therefore, provide another organism to join the small number of recorded animals that underwent WGD. As most recorded WGD events are ancient, the recency of this event allows analysis into genes that are still mostly redundant. With this study, we contributed further evidence to the body of work that examines WGD events and their impacts, particularly in animals.

## Materials and Methods

### Nematode Maintenance and Inbreeding

The following nematodes and laboratory strains were used in this study: *C. elegans* N2 (*C. elegans* Genetics Center); *P. pacificus* PS312 (Sommer et al. 1996), *A. sudhausi* SB413 (Bar-Eyal et al. 2008; Fürst von Lieven 2008), and *A. seani* RS1982 (Kanzaki et al. 2015). Note that *A. sudhausi* was originally described as *Koerneria sudhausi*. The taxonomic status of the genus *Koerneria* was recently revised, resulting in a split into the genera *Koerneria* and *Allodiplogaster* (Kanzaki et al. 2014). All strains were maintained on nematode growth medium (NGM) agar plates (Sommer et al. 1996). For inbreeding, a single late J4 stage hermaphrodite was moved to a fresh new plate to lay eggs. This step was repeated for 10 generations to eventually

isolate an inbred isogenic line that was used for subsequent downstream applications under the strain designation SB413B.

### DNA Extraction and Sequencing

*Allodiplogaster sudhausi* nematodes were washed off of 100 NGM agar plates using M9 buffer and pelleted by centrifugation at  $1,300\times g$  for 1 min. The pellet was washed twice in M9 before worms were frozen in liquid nitrogen and ground to a fine powder using a mortar and pestle. The powder was directly transferred into the lysis buffer from the QIAGEN genomic DNA extraction kit, which was used in combination with QIAGEN genomic tip columns (500/G) (QIAGEN, Hamburg, Germany). The protocol was performed following the manufacturer's instructions. All steps involving vortexing of the sample were replaced by inversion to limit unwanted DNA shearing. DNA quality and quantity were determined with a NanoDrop ND 1000 spectrometer (PepLab, Erlangen, Germany), a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, USA), and by a Femto pulse system (Agilent, CA, USA). A total of 20  $\mu\text{g}$  *A. sudhausi* genomic DNA was sheared to a target fragment size of 45 kb using a needle. A 45-kb template library was prepared using the BluePippin size-selection system according to the manufacturer's protocol (P/N 100-286-000-07, Pacific Biosciences, California, USA). The final library was sequenced on a Pacific Biosciences Sequel instrument following the Magbead loading protocol and version 1.2.1 sequencing kits. A total of two SMRT cells (version 1.2.1) generated 40 Gb (100-fold coverage).

### Genome Assembly and Evaluation

Raw long-read data were assembled using Canu version 1.8 (Koren et al. 2017). The completeness of the genome assembly was evaluated using the BUSCO software (version 3.0.1, with the -m genome option against the nematode odb9 dataset) (Simão et al. 2015). To investigate the sequencing coverage distribution, we downloaded previously generated Illumina sequencing data from the European Nucleotide Archive (Accessions: ERR2208557, ERR2208558, SRR12424054, and SRR12424056) (Sieriebriennikov et al. 2018; Casasa et al. 2021) and aligned these reads against the *A. sudhausi* genome with BWA mem program (version 0.7.17) (Li and Durbin 2009). The coverage profiles were generated from the resulting alignment files by the samtools depth program (version 0.1.18) (Li et al. 2009).

### Gene Annotations

We generated a transcriptome assembly of mixed-stage RNA-seq data from *A. sudhausi* with the help of the Trinity program (version 2.2.0, with the -normalize\_reads option) (Grabherr et al. 2011). This transcriptome assembly

was combined with the community-curated gene annotations of *P. pacificus* (El Paco gene annotations version 3) (Athanasouli et al. 2020) to generate evidence-based gene annotations for *A. sudhausi*. Specifically, both datasets were aligned against the *A. sudhausi* assembly by the exonerate program (version 2.2.0, with -bestn 2 -dnawordlen 20 -maxintron 20,000 options) (Slater and Birney 2005). Subsequently, the alignments were processed by the PPCAC software (version 1.0) to select one representative gene model with the longest open reading frame per locus (Rödelsperger 2021). Final gene annotations were assessed by the BUSCO software (version 3.0.1, with the -m prot option against the nematode odb9 dataset) (Simão et al. 2015; Rödelsperger 2021).

### Chromosome Staining

We made a 1:100 dilution of 20 mM Hoechst 33342 dye (Chazotte 2011) in sperm salts (50 mM PIPES, 25 mM KCl, 1 mM  $\text{MgSO}_4$ , 45 mM NaCl, 2 mM  $\text{CaCl}_2$ , pH 7). 10  $\mu\text{l}$  of this solution was then pipetted onto a microscope slide. Ten adult hermaphrodites were moved onto the solution. A surgical blade was used to decapitate the heads of the worms (cutting right below the pharynx). This resulted in the gonadal arms being pushed out due to the internal pressure, freeing them from the worm body. A cover slip was then placed on top and the worms were imaged using a Nomarsky DIC microscope and the chromosomes were subsequently counted.

### Comparative Genomic Analysis, KEGG, and Pfam Annotations

For comparative genomic analyses, we compiled additional protein data for *A. seani*, *P. pacificus* (El Paco gene annotations version 3) (Athanasouli et al. 2020), and *C. elegans* (WormBase ParaSite version WBPS16) (Howe et al. 2016). The dataset for *A. seani* was generated from mixed-stage RNA-seq data and was assembled with the Trinity program (version 2.2.0, with the -normalize\_reads option) (Grabherr et al. 2011). To reduce isoform information, we selected the assembled transcript with the longest open reading frame per Trinity gene and further clustered protein sequences with the cd-hit program (version 4.3) (Li and Godzik 2006). Protein domains were annotated by the hmmsearch program (version 3.3, with -E 0.001 option) using the Pfam-A data (version 3.1b2) as target database (Bateman et al. 1999). Clusters of orthologous genes were generated by the OrthoFinder software (version 2.5.2). From the resulting orthogroups, we extracted orthogroups with ohnologs (orthogroups with two *A. sudhausi* copies and at most one copy in *A. seani*) and orthogroups with *A. seani* orthologs (either one or two copies in *A. sudhausi* and one copy in *A. seani*). dN and dS values were computed from intraspecies pairs of the ohnolog



orthogroups and cross-species pairs of the orthogroups with *A. seani* orthologs. This was done by aligning protein sequences with MUSCLE (version 3.8.31) (Edgar 2004), conversion into codon alignment with PAL2NAL (version 14) (Suyama et al. 2006), and divergence estimation with the codeml program of PAML (version 4.9) (Yang 2007). Metabolic pathway annotations were generated by identification of orthologs in the KEGG database using the blastkoala web application (with the “eukaryotes” and “family\_eukaryotes” values for taxonomic group and database level, respectively) (Kanehisa et al. 2016). Genes with orthologs in the KEGG database were then annotated with the corresponding KEGG accessions.

### Ohnolog Expression

The FPKM values were calculated from the *A. sudhausi* transcriptome. This was done by 1) Summing up the total number of reads and dividing by  $1 \times 10^6$  to get the scaling factor (per million), 2) Dividing each read count by the scaling factor, 3) Dividing these values by the length of each gene (in kb). This was calculated separately for four biological replicates. The mean value for each gene was then calculated to get the FPKM. Fold change was calculated by dividing one ohnolog by its duplicate. To examine position bias of ohnologs on contigs, we first filtered for contigs that had  $\geq 10$  ohnologs. We then examined the distribution of ohnologs on the contigs (by referring to the assembly) and calculated the percentage of DE ohnologs (absolute  $\log_2(\text{FC}) \geq 1$ ) for each contig.

### Repeat Annotation

We used RepeatModeler2 (version 2.0.1, parameters: -LTRstruct) (Flynn et al. 2020) for de novo repeat detection in *A. sudhausi* and compared the TE content with available TE data for *P. pacificus* (Athanasouli and Rödelsperger 2022) and *C. elegans*. The *C. elegans* TE dataset was created using RepeatMasker's incorporated libraries (parameters: -species worm) (A.F.A. Smit, R. Hubley, and P. Green, <http://repeatmasker.org/>). Based on the RepeatModeler2 classification, we calculated the percentage of the genome coverage by repeat elements as well as the overall length of the TEs and simple repeats across the three species. We investigated the span of the superfamilies present in *A. sudhausi* for the four major TEs orders (DNA transposons, LINEs, LTRs and SINEs). Furthermore, we identified the superfamilies spanning at least 1 Mb in any of the 3 species and evaluated the percentage of the genome covered by each superfamily in the organisms being compared. To time the DNA/hAT-Ac expansion, we extracted DNA/hAT-Ac sequences from the RepeatModeler/RepeatMasker output files and ran all-against-all blastn searches (version 2.10.1, with -dust no, -evalue 0.001, -qcov \_hspperc 80, and -perc\_identity 80 options).

Pairwise percentage identities were extracted from blastn hits between different DNA/hAT-Ac elements (exclusion of self-hits) that span at least 100 nucleotides.

### CRISPR Injection and *dpy-1* Mutant Identification

CRISPR knockouts were generated following the *P. pacificus* protocol (Witte et al. 2014). CRISPR RNAs (crRNAs) and trans-activating crRNA (tracrRNA) (Cat. No. 1072534) were synthesized by Integrated DNA Technologies (IDT), while the Cas9 endonuclease (Cat. No. 1081058) was purchased from IDT. The CRISPR/Cas9 complex was prepared by mixing 0.5 mg/ml Cas9 nuclease, 0.1 mg/ml tracrRNA, and 0.056 mg/ml crRNA in the TE buffer followed by a 10-min incubation at 37° C. Microinjections were performed in late-stage J4 hermaphrodites following standard practice using an Eppendorf microinjection system. The gRNA (CTCAAAGAGAACTCCAGCTG) sequence was designed just before an NGG PAM site and targeted exon six of both *dpy-1* genes. Gene specific primers were designed for both. Transcriptomic reads were mapped against the *A. sudhausi* genome using IGV (Integrative Genomics Viewer, version 2.8.9) to see where the coding regions were. For *dpy-1-A*, the forward primer F\_1 (5'-CTTCAGGCACCCCTCTAGGCA-3') was designed in exon 3 and the reverse R\_1 (5'-GCAACATGCTCGGCAAGGCT-3') in exon 6 (amplicon size: 650 bp). For *dpy-1-B*, the forward primer F\_2 (5'-CCCAAATCATTCGTTGCC-3') was designed in exon 2 and the reverse R\_2 (5'-CACTTAATCCACGCTCTTC-3') in the intron between exons 6 and 7 (amplicon size: 1200 bp). Polymerase chain reactions (PCRs) were run using both primer pairs on F1 young adults of injected worms. Heterozygotes were identified via Sanger sequencing and homozygous mutants were then obtained by self-fertilizing heterozygous F1 to eventually obtain homozygous knock-out mutants.

### Phylogeny Generation

A subset of species from Susoy et al. (2015) was edited using figTree software v.1.4.4 ([tree.bio.ed.ac.uk/software/figtree](http://tree.bio.ed.ac.uk/software/figtree)) to obtain the species tree. For the DPY-1 phylogeny, the protein sequence *C. elegans* DPY-1 isoform a (the longest) was obtained from wormbase.org (version WS284). This sequence was used to BLAST (Altschul et al. 1990) (query type: protein) against selected nematodes on parasite.wormbase.org (version WBPS16). For species belonging to the *Pristionchus* genus, pristonchus.org (version 2.0.0.rc8) was used to BLAST DPY-1 against the protein databases. It should be noted that we obtained two copies each in *P. expectatus* and *P. arcanus*, but believe the copies are due to heterozygosity which is common in out-crossing species. Thus we only took one copy of each for the

phylogeny. The *A. sudhausi* DPY-1 proteins were translated from the best matches in the *A. sudhausi* transcriptome. The DPY-1 protein phylogeny was then generated using RAxML version 8.2.12 (raxmlHPC -f a -m PROTGAMMAAUTO -p 12345 -x 12345 -N 100) (Stamatakis 2014).

### Statistical Analysis

Each unique domain or pathway prediction per annotation was counted and compared both within and between species. A Fisher's exact test was run using the 2X2 matrix, whereby each gene with a given domain/pathway was compared against the whole dataset to identify those that were significantly different. The *P* values were adjusted using the false discovery rate (FDR). All analyses were performed using RStudio Statistical Software (v1.4.1717; R Core Team 2021).

### Supplementary material

Supplementary data are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

### Acknowledgments

We thank Waltraud Röseler for help with the generation of the PacBio library and all members of the Sommer Lab for discussion. This study was funded by the Max-Planck Society.

### Author Contributions

S.S.W. and R.J.S. designed the study. S.S.W., M.A., and CR conducted genome analysis and H.W. performed the CRISPR experiments. S.S.W. and R.J.S. wrote the manuscript with input from others.

### Data availability

Raw reads, genome, and transcriptome assemblies have been submitted to the European Nucleotide archive under the study accession PRJEB48369.

### Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.  
 Amores A, et al. 1998. Zebrafish hox clusters and vertebrate genome evolution. *Science* 282:1711–1714.  
 Arensburg P, et al. 2011. Phylogenetic and functional characterization of the hAT transposon superfamily. *Genetics* 188:45–57.  
 Athanasouli M, et al. 2020. Comparative genomics and community curation further improve gene annotations in the nematode *Pristionchus pacificus*. *BMC Genomics* 21:708.  
 Athanasouli M, Rödelsperger C. 2022. Analysis of repeat elements in the *Pristionchus pacificus* genome reveals an ancient invasion by horizontally transferred transposons. *BMC Genomics* 23:523.

Aury J-M, et al. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444:171–178.  
 Avise JC. 2011. Hermaphroditism: A primer on the biology, ecology, and evolution of dual sexuality. New York: Columbia University Press.  
 Bar-Eyal M, Sharon E, Spiegel Y, Oka Y. 2008. Laboratory studies on the biocontrol potential of the predatory nematode *Koerneria sudhausi* (Nematoda: Diplogastridae). *Nematology* 10:633–637.  
 Barrière A, et al. 2009. Detecting heterozygosity in shotgun genome assemblies: lessons from obligately outcrossing nematodes. *Genome Res* 19:470–480.  
 Bateman A, et al. 1999. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res* 27:260–262.  
 Biddle JF, Ragsdale EJ. 2020. Regulators of an ancient polyphenism evolved through episodic protein divergence and parallel gene radiations. *Proc R Soc B: Biol Sci* 287:20192595.  
 Birchler JA, Riddle NC, Auger DL, Veitia RA. 2005. Dosage balance in gene regulation: biological implications. *Trends Genet* 21:219–226.  
 Blanc-Mathieu R, Perfus-Barbeoch L, Aury JM, Da Rocha M, Gouzy J, Sallet E, Martin-Jimenez C, Bailly-Bechet M, Castagnone-Sereno P, Flot JF, et al. 2017. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS Genet* 13:e1006777.  
 Blomme T, et al. 2006. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol* 7:R43.  
 Brenner S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics* 77:71–94.  
 Bridges CB. 1936. The bar “gene” a duplication. *Science* 83:210–211.  
 Carlton PM, Davis RE, Ahmed S. 2022. Nematode chromosomes. *Genetics* 221:1.  
 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.  
 Chazotte B. 2011. Labeling nuclear DNA with hoechst 33342. *Cold Spring Harb Protoc* 2011:pdb.prot5557.  
 Chénais B, Caruso A, Hiard S, Casse N. 2012. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene* 509:7–15.  
 Clark JW, Donoghue PCJ. 2018. Whole-genome duplication and plant macroevolution. *Trends Plant Sci* 23:933–945.  
 Cosby RL, et al. 2021. Recurrent evolution of vertebrate transcription factors by transposase capture. *Science* 371:6531.  
 Cutter AD. 2008. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Mol Biol Evol* 25:778–786.  
 Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 3:e314.  
 Dieterich C, et al. 2008. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat Genet* 40:1193–1198.  
 Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:1–19.  
 Faino L, et al. 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res* 26:1091–1100.  
 Flemming AJ, Shen ZZ, Cunha A, Emmons SW, Leroi AM. 2000. Somatic polyploidization and cellular proliferation drive body size evolution in nematodes. *Proc Natl Acad Sci USA* 97:5285–5290.  
 Flynn JM, et al. 2020. Repeatmodeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA* 117:9451–9457.

- Force A, et al. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
- Fürst von Lieven A. 2008. *Koerneria sudhausi* n. sp. (Nematoda: Diplogastridae); a hermaphroditic diplogastrid with an egg shell formed by zygote and uterine components. *Nematology* 10: 27–45.
- Gonzalez de la Rosa PM, et al. 2021. A telomere-to-telomere assembly of *Oscheius tipulae* and the evolution of rhabditid nematode chromosomes. *G3* 11(1):jkaa020.
- Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29: 644–652.
- Haldane JBS. 1932. *The causes of evolution*. London: Longmans, Green and Co.
- Han Z, et al. 2020. Improving transgenesis efficiency and CRISPR-associated tools through codon optimization and native intron addition in *Pristionchus* nematodes. *Genetics* 216:947–956.
- Hayashi S, et al. 2022. Neofunctionalization of a noncoding portion of a DNA transposon in the coding region of the chimerical sex-determining gene *dm-W* in *Xenopus* frogs. *Mol Biol Evol* 39:1–6.
- Hiraga H, Ishita Y, Chihara T, Okumura M. 2021. Efficient visual screening of CRISPR/Cas9 genome editing in the nematode *Pristionchus pacificus*. *Dev Growth Differ* 63:488–500.
- Howe KL, et al. 2016. Wormbase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res* 44:D774–D780.
- Inoue H, et al. 2005. The *C. elegans* p38 MAPK pathway regulates nuclear localization of the transcription factor SKN-1 in oxidative stress response. *Genes Dev* 19:2278–2283.
- Jinek M, et al. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821.
- Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30.
- Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731.
- Kanzaki N, Giblin-Davis RM, Ragsdale EJ. 2015. *Allodiplogaster josephi* n. sp. and *A. seani* n. sp. (Nematoda: Diplogastridae), associates of soil-dwelling bees in the eastern USA. *Nematology* 17:831–863.
- Kanzaki N, Ragsdale E, Giblin-Davis R. 2014. Revision of the paraphyletic genus *Koerneria* Meyl, 1960 and resurrection of two other genera of Diplogastridae (Nematoda). *ZooKeys* 442:17–30.
- Kenning C, Kipping I, Sommer RJ. 2004. Isolation of mutations with dumpy-like phenotypes and of collagen genes in the nematode *Pristionchus pacificus*. *Genesis* 40:176–183.
- Kondrashov FA, Koonin EV. 2004. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet* 20:287–290.
- Koren S, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27:722–736.
- Kozlowski DKL, et al. 2021. Movements of transposable elements contribute to the genomic plasticity and species diversification in an asexually reproducing nematode pest. *Evol Appl* 14:1844–1866.
- Laricchia KM, Zdraljevic S, Cook DE, Andersen EC. 2017. Natural variation in the distribution and abundance of transposable elements across the *Caenorhabditis elegans* species. *Mol Biol Evol* 34: 2187–2202.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li H, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659.
- Lien S, et al. 2016. The Atlantic salmon genome provides insights into rediploidization. *Nature* 533:200–205.
- Lisch D. 2013. How important are transposons for plant evolution? *Nat Rev Genet* 14:49–61.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
- Madl JE, Herman RK. 1979. Polyploids and sex determination in *Caenorhabditis elegans*. *Genetics* 93:393–402.
- Maere S, et al. 2005. Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci USA* 102:5454–5459.
- Makalowski W, Gotea V, Pande A, Makalowska I. 2019. Transposable elements: classification, identification, and their use as a tool for comparative genomics. *Methods Mol Biol* 1910:177–207.
- Marburger S, et al. 2018. Whole genome duplication and transposable element proliferation drive genome expansion in Corydoradinae catfishes. *Proc R Soc B Biol Sci* 285:20172732.
- Marcet-Houben M, Gabaldón T. 2015. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the Baker's yeast lineage. *PLoS Biol* 13:e1002220.
- McClintock B. 1950. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA* 36:344–355.
- Moriyama Y, et al. 2016. Evolution of the fish heart by sub/neofunctionalization of an elastin gene. *Nat Commun* 7:10397.
- Moriyama Y, Koshiba-Takeuchi K. 2018. Significance of whole-genome duplications on the emergence of evolutionary novelties. *Brief Funct Genomics* 17:329–338.
- Ohno S. 1970. *Evolution by gene duplication*. London: Springer-Verlag.
- Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu Rev Genet* 34:401–437.
- Pala I, Coelho MM, Scharl M. 2008. Dosage compensation by gene-copy silencing in a triploid hybrid fish. *Curr Biol* 18:1344–1348.
- Papp B, Pál C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424:194–197.
- Parisod C, Holderegger R, Brochmann C. 2010. Evolutionary consequences of autopolyploidy. *New Phytol* 186:5–17.
- Prabh N, et al. 2018. Deep taxon sampling reveals the evolutionary dynamics of novel gene families in *Pristionchus nematodes*. *Genome Res* 28:1664–1674.
- R Core Team. 2021. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ray DA, Pagan HJT, Thompson ML, Stevens RD. 2006. Bats with hATs: evidence for recent DNA transposon activity in genus *Myotis*. *Mol Biol Evol* 24:632–639.
- Rödelsperger C, et al. 2017. Single-molecule sequencing reveals the chromosome-scale genomic architecture of the nematode model organism *Pristionchus pacificus*. *Cell Rep* 21:834–844.
- Rödelsperger C, et al. 2018. Phylotranscriptomics of *Pristionchus nematodes* reveals parallel gene loss in six hermaphroditic lineages. *Curr Biol* 28:3123–3127.e5.
- Rödelsperger C, et al. 2019. Crowdsourcing and the feasibility of manual gene annotation: a pilot study in the nematode *Pristionchus pacificus*. *Sci Rep* 9:1–9.
- Rödelsperger C. 2021. The community-curated *Pristionchus pacificus* genome facilitates automated gene annotation improvement in related nematodes. *BMC Genomics* 22:216.
- Rubin E, Lithwick G, Levy AA. 2001. Structure and evolution of the hAT transposon superfamily. *Genetics* 158:949–957.
- Sagasti A, Hobert O, Troemel ER, Ruvkun G, Bargmann CI. 1999. Alternative olfactory neuron fates are specified by the LIM homeobox gene *lim-4*. *Genes Dev* 13(14):1794–1806.
- Sarbassov DD, Ali SM, Sabatini DM. 2005. Growing roles for the mTOR pathway. *Curr Opin Cell Biol* 17:596–603.

- Schiffer PH, et al. 2019. Signatures of the evolution of parthenogenesis and cryptobiosis in the genomes of Panagrolaimid nematodes. *iScience* 21:587–602.
- Schroeder NE. 2021. Introduction to *Pristionchus pacificus* anatomy. *J Nematol* 53:1–9.
- Seger R, Krebs EG. 1995. The MAPK signaling cascade. *FASEB J* 9:726–735.
- Seoighe C, Gehring C. 2004. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends Genet* 20:461–464.
- Session AM, et al. 2016. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* 538:336–343.
- Shiu SH, Byrnes JK, Pan R, Zhang P, Li WH. 2006. Role of positive selection in the retention of duplicate genes in mammalian genomes. *Proc Natl Acad Sci USA* 103:2232–2236.
- Sieriebriennikov B, et al. 2018. A developmental switch generating phenotypic plasticity is part of a conserved multi-gene locus. *Cell Rep* 23:2835–2843.e4.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
- Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:1–11.
- Sommer RJ, Carta LK, Kim SY, Sternberg PW. 1996. Morphological, genetic and molecular description of *Pristionchus pacificus* sp. n. (Nematoda: Neodiplogastridae). *Fundam Appl Nematol* 19:511–521.
- Sommer RJ. 2015. Nematoda. In: *Evolutionary developmental biology of invertebrates*. Vol. 3. Berlin: Springer. p. 15–34.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Susoy V, Ragsdale EJ, Kanzaki N, Sommer RJ. 2015. Rapid diversification associated with a macroevolutionary pulse of developmental plasticity. *eLife* 4:1–39.
- Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609–W612.
- Szitenberg A, et al. 2017. Comparative genomics of apomictic root-knot nematodes: hybridization, ploidy, and dynamic genome change. *Genome Biol Evol* 9:2844–2861.
- Tan S, et al. 2021. DNA Transposons mediate duplications via transposition-independent and -dependent mechanisms in metazoans. *Nat Commun* 12:1–14.
- Thomas JH. 2006. Adaptive evolution in two large families of ubiquitin-ligase adapters in nematodes and plants. *Genome Res* 16:1017–1030.
- Triantaphyllou AC, Riggs RD. 1979. Polyploidy in an amphimictic population of *Heterodera glycines*. *J Nematol* 11:371–376.
- Troemel ER, Kimmel BE, Bargmann CI. 1997. Reprogramming chemotaxis responses: sensory neurons define olfactory preferences in *C. elegans*. *Cell* 91:161–169.
- Vicient CM, Casacuberta JM. 2017. Impact of transposable elements on polyploid plant genomes. *Ann Bot* 120:195–207.
- Walsh EJ, Zhang L. 1992. Polyploidy and body size variation in a natural population of the rotifer *euchlanis dilatata*. *J Evol Biol* 5:345–353.
- Weider LJ. 1987. Life history variation among low-Arctic clones of obligately parthenogenetic *Daphnia pulex*: a diploid-polyploid complex. *Oecologia* 73:251–256.
- Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8:973–982.
- Wilson DF. 2017. Oxidative phosphorylation: regulation and role in cellular and tissue metabolism. *J Physiol (Lond)* 595:7023–7038.
- Witte H, et al. 2014. Gene inactivation using the CRISPR/Cas9 system in the nematode *Pristionchus pacificus*. *Dev Genes Evol* 225:55–62.
- Wolfe KH. 2001. Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet* 2:333–341.
- Wolfe KH. 2015. Origin of the yeast whole-genome duplication. *PLoS Biol* 13:e1002221.
- Wood WB. 1988. *The nematode Caenorhabditis elegans*. New York: Cold Spring Harbor Laboratory.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
- Yin D, et al. 2018. Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. *Science* 359:55–61.
- Zakon HH, Lu Y, Zwickl DJ, Hillis DM. 2006. Sodium channel genes and the evolution of diversity in communication signals of electric fishes: convergent molecular evolution. *Proc Natl Acad Sci* 103:3675–3680.

Associate editor: Prof. Kenneth Wolfe



**A New Hope: A hermaphroditic nematode enables analysis of a recent whole genome duplication event**

**Sara S. Wighard, Marina Athanasouli, Hanh Witte, Christian Rödelsperger & Ralf J. Sommer\***

**Supplementary Online Material**

Content

Supplementary Figures 2

Supplementary Tables 5

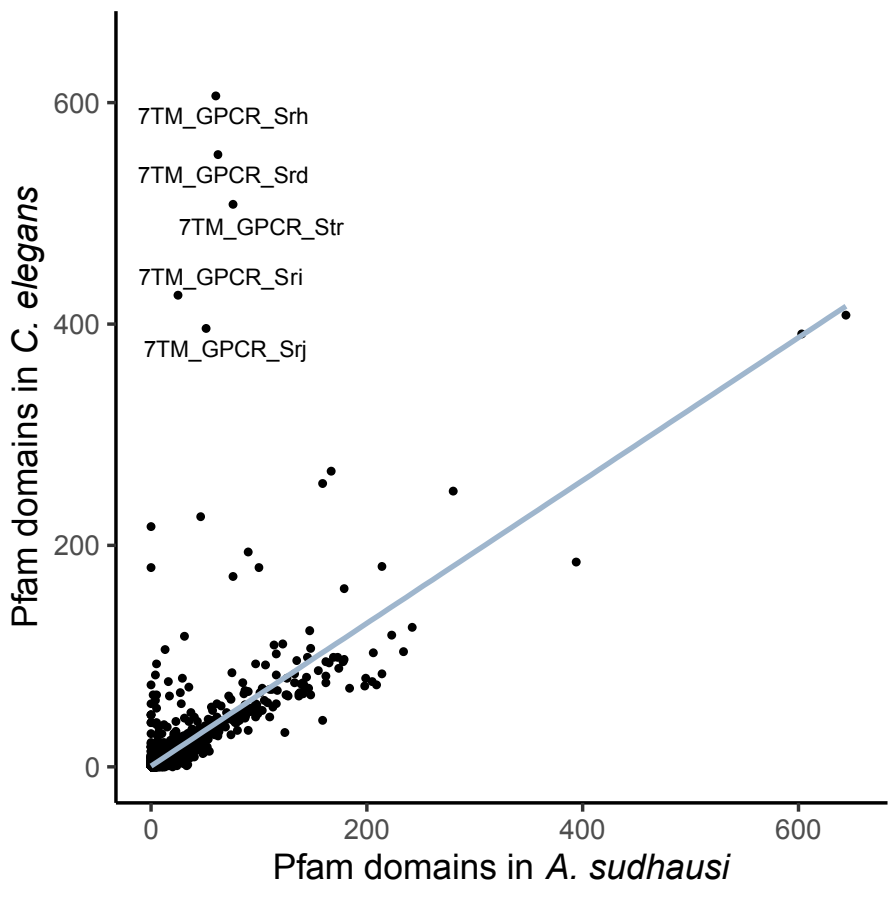


Fig. S1. A comparison between *A. sudhausi* and *C. elegans* predicted Pfam domains

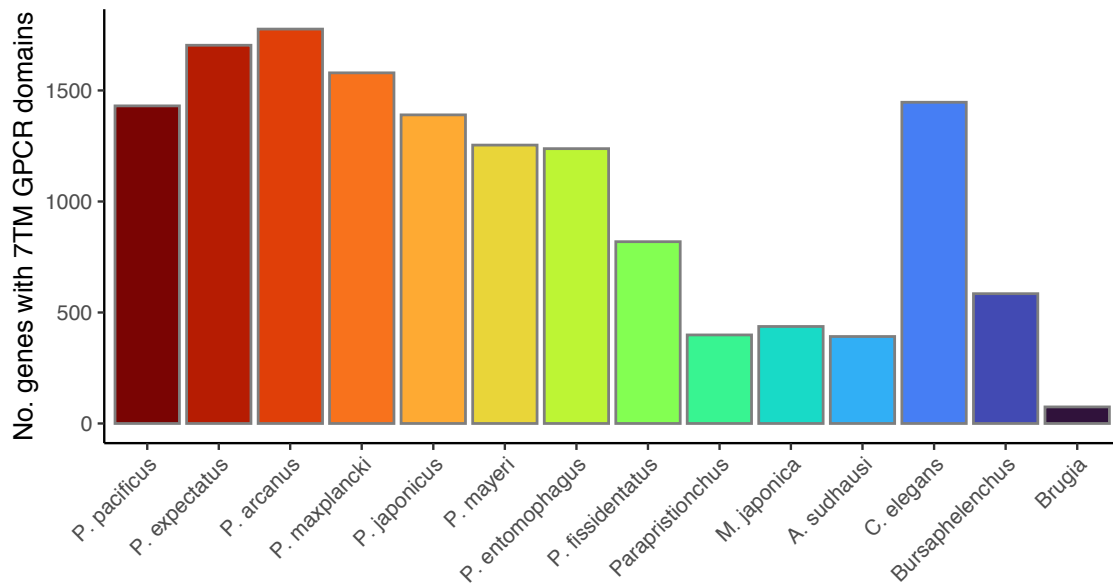


Fig. S2. A comparison between the number of domains with 7TM-GPCRs across various nematode species shows 7TM-GPCRs are only highly overrepresented in *C. elegans* and *Pristionchus* species

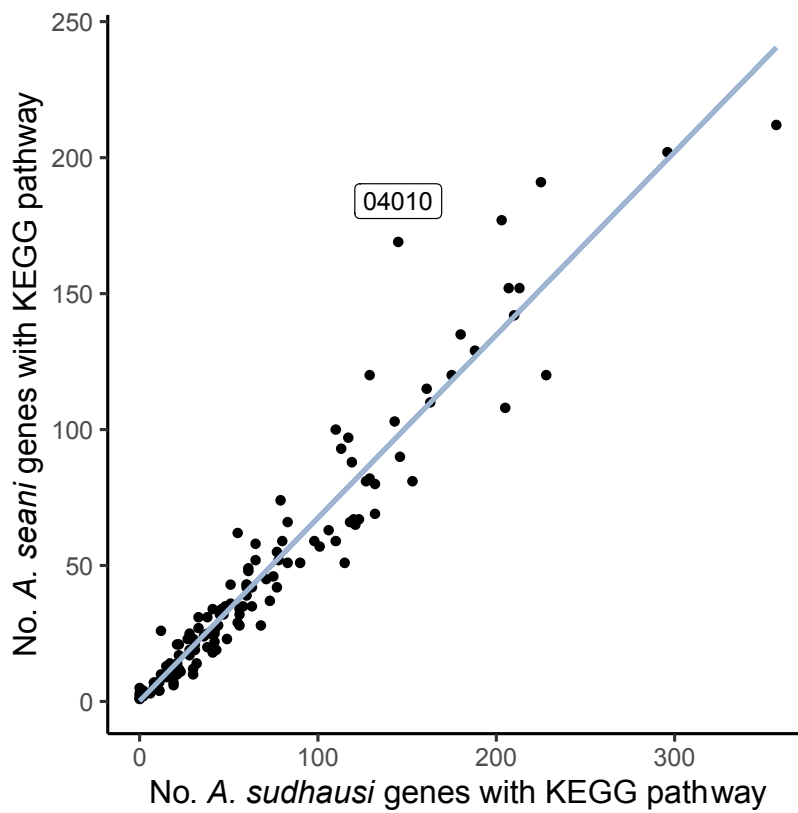


Fig. S3. A KEGG comparison showed an upregulation of the MAPK pathway (04010) in *A. seani* compared to *A. sudhausi*

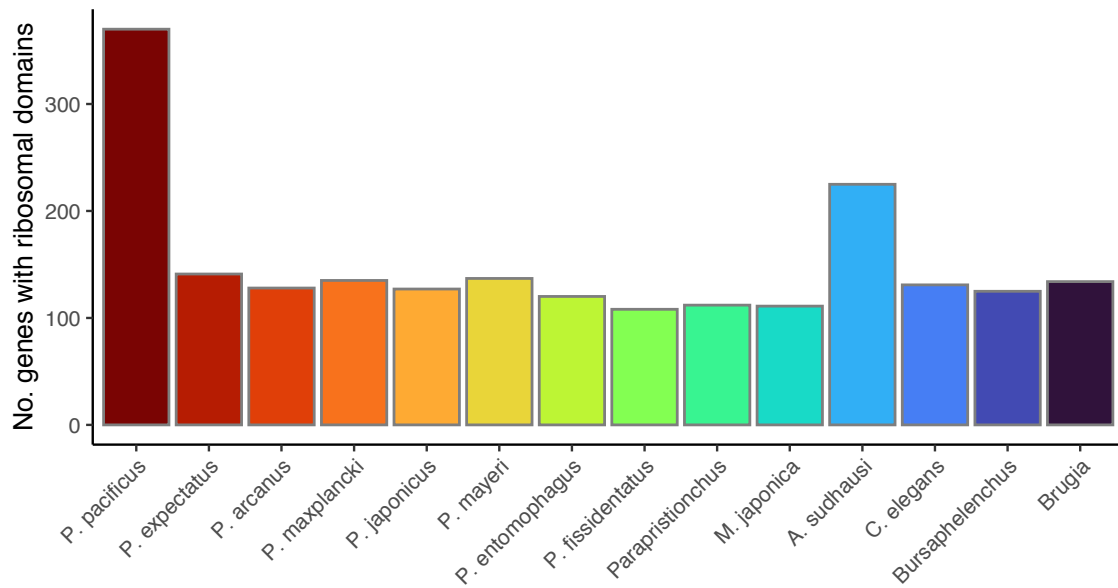


Fig. S4. The frequency of ribosomal Pfam domains across various nematode species, with *P. pacificus* showing particularly high representation.

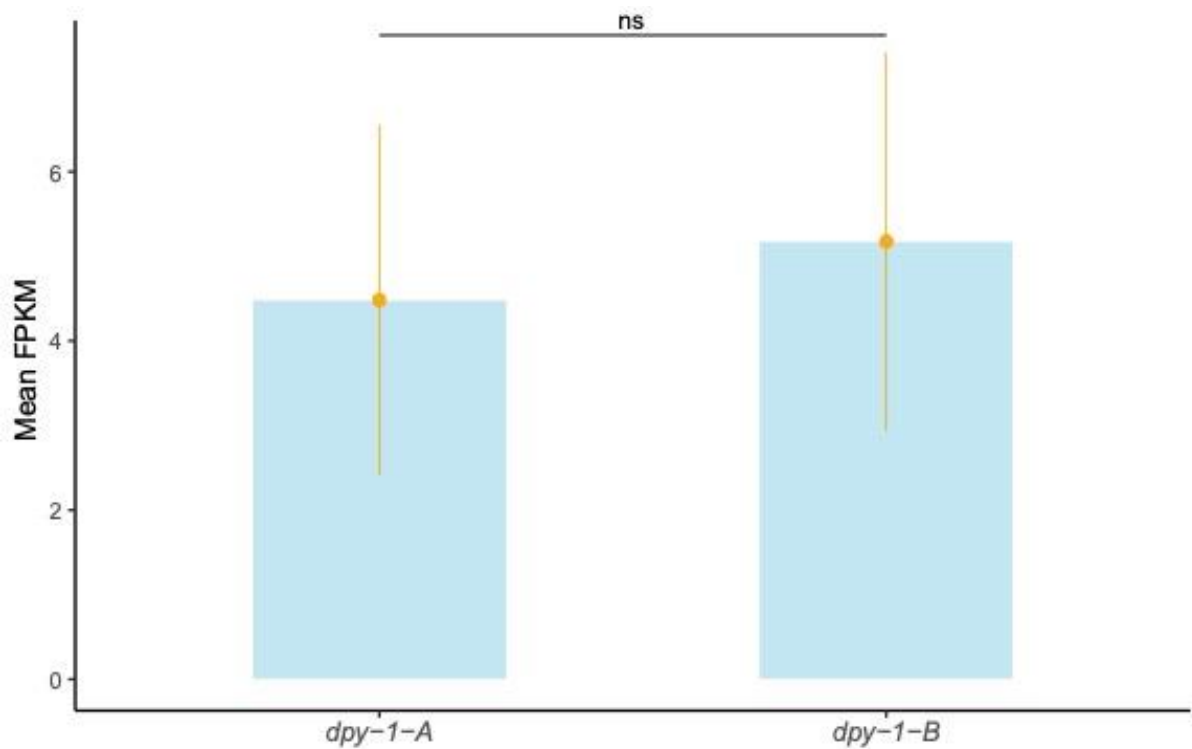


Fig. S5. The expression values of the *A. sudhausi dpy-1* ohnologs display no significant difference from one another (Welch Two Sample t-test,  $p < 0.05$ ). Error bars indicate standard deviation.

Table S1. BUSCO analysis of the *A. sudhausi* and *A. seani* assemblies.

|                                     | <i>A. sudhausi</i><br>PacBio Genome | <i>A. sudhausi</i><br>Illumina Genome | <i>A. sudhausi</i><br>Annotations | <i>A. sudhausi</i><br>Transcriptome | <i>A. seani</i><br>Transcriptome |
|-------------------------------------|-------------------------------------|---------------------------------------|-----------------------------------|-------------------------------------|----------------------------------|
| Complete BUSCOs (C)                 | 87.4%                               | 82.7%                                 | 89.7%                             | 90.1%                               | 84.2%                            |
| Complete and single-copy BUSCOs (S) | 38.7%                               | 46.1%                                 | 9.9%                              | 29.0%                               | 36.2%                            |
| Complete and duplicated BUSCOs (D)  | 48.7%                               | 36.6%                                 | 79.8%                             | 61.1%                               | 48.0%                            |
| Fragmented BUSCOs (F)               | 6.5%                                | 9.0%                                  | 5.9%                              | 5.9%                                | 10.0%                            |
| Missing BUSCOs (M)                  | 6.1%                                | 8.3%                                  | 4.4%                              | 4.0%                                | 5.8%                             |
| Total BUSCO groups searched (n)     | 982                                 | 982                                   | 982                               | 982                                 | 982                              |

Table S2. Pfam Domains significantly more likely to be retained (i.e. found in ohnologs). Determined using the Fisher's 2x2 matrix.

| <b>Domain</b> | <b>Accession</b> | <b>p value</b> | <b>FDR adj. p value</b> |
|---------------|------------------|----------------|-------------------------|
| Homeodomain   | PF00046.31       | 0.012          | 0.0263023255813953      |
| LRR_4         | PF12799.9        | 0.039          | 0.0403928571428571      |
| PDZ           | PF00595.26       | 0.008          | 0.0263023255813953      |
| zf-C2H2       | PF00096.28       | 0.013          | 0.0263023255813953      |
| zf-H2C2_2     | PF13465.8        | 0.005          | 0.0263023255813953      |

Table S3. Pfam domains significantly more likely to be lost (i.e. found in single copies). Determined using the Fisher's 2x2 matrix

| Domain          | Accession  | p value | FDR adj. p value   |
|-----------------|------------|---------|--------------------|
| AA_permease     | PF00324.23 | 0.038   | 0.0398313253012048 |
| AAA_16          | PF13191.8  | 0.041   | 0.0419647058823529 |
| AAA_5           | PF07728.16 | 0.031   | 0.0396455696202532 |
| ACBP            | PF00887.21 | 0.035   | 0.0396455696202532 |
| Acyl-CoA_dh_M   | PF02770.21 | 0.038   | 0.0398313253012048 |
| adh_short       | PF00106.27 | 0.023   | 0.0396455696202532 |
| adh_short_C2    | PF13561.8  | 0.007   | 0.0263023255813953 |
| Aminotran_1_2   | PF00155.23 | 0.005   | 0.0263023255813953 |
| An_peroxidase   | PF03098.17 | 0.028   | 0.0396455696202532 |
| ATPgrasp_ST     | PF14397.8  | 0.035   | 0.0396455696202532 |
| ATPgrasp_YheCD  | PF14398.8  | 0.035   | 0.0396455696202532 |
| Beta_helix      | PF13229.8  | 0.035   | 0.0396455696202532 |
| BTB_2           | PF02214.24 | 0.045   | 0.0455232558139535 |
| Calreticulin    | PF00262.20 | 0.035   | 0.0396455696202532 |
| Calsequestrin   | PF01216.19 | 0.035   | 0.0396455696202532 |
| Carn_acyltransf | PF00755.22 | 0.008   | 0.0263023255813953 |
| Cation_ATPase_N | PF00690.28 | 0.022   | 0.0396455696202532 |
| Coatomer_WDAD   | PF04053.16 | 0.036   | 0.0396455696202532 |
| Diphthamide_syn | PF01866.19 | 0.035   | 0.0396455696202532 |
| DO-GTPase2      | PF19993.1  | 0.035   | 0.0396455696202532 |
| DUF815          | PF05673.15 | 0.022   | 0.0396455696202532 |
| E1-E2_ATPase    | PF00122.22 | 0.038   | 0.0398313253012048 |
| EFG_C           | PF00679.26 | 0.012   | 0.0263023255813953 |
| EFG_III         | PF14492.8  | 0.012   | 0.0263023255813953 |
| EFG_IV          | PF03764.20 | 0.005   | 0.0263023255813953 |
| Ephrin_rec_like | PF07699.15 | 0.035   | 0.0396455696202532 |
| Fer4_7          | PF12838.9  | 0.035   | 0.0396455696202532 |
| G-patch         | PF01585.25 | 0.028   | 0.0396455696202532 |
| GIDA            | PF01134.24 | 0.002   | 0.0263023255813953 |
| Glyco_transf_7C | PF02709.16 | 0.038   | 0.0398313253012048 |
| HSP90           | PF00183.20 | 0.035   | 0.0396455696202532 |
| Hydrolase       | PF00702.28 | 0.028   | 0.0396455696202532 |
| Hydrolase_3     | PF08282.14 | 0.035   | 0.0396455696202532 |
| IMPDH           | PF00478.27 | 0.035   | 0.0396455696202532 |
| Ion_trans       | PF00520.33 | 0.049   | 0.049              |
| KIF1B           | PF12423.10 | 0.035   | 0.0396455696202532 |
| Kinesin         | PF00225.25 | 0.003   | 0.0263023255813953 |
| Kinesin_assoc   | PF16183.7  | 0.035   | 0.0396455696202532 |

|                 |            |       |                    |
|-----------------|------------|-------|--------------------|
| KR              | PF08659.12 | 0.004 | 0.0263023255813953 |
| MCM             | PF00493.25 | 0.013 | 0.0263023255813953 |
| MCM_lid         | PF17855.3  | 0.002 | 0.0263023255813953 |
| MCM_N           | PF14551.8  | 0.002 | 0.0263023255813953 |
| MCM_OB          | PF17207.5  | 0.002 | 0.0263023255813953 |
| Mg_chelatase    | PF01078.23 | 0.028 | 0.0396455696202532 |
| Microtub_bd     | PF16796.7  | 0.015 | 0.0296590909090909 |
| P_proprotein    | PF01483.22 | 0.035 | 0.0396455696202532 |
| Peptidase_M13   | PF01431.23 | 0.028 | 0.0396455696202532 |
| Peptidase_S8    | PF00082.24 | 0.012 | 0.0263023255813953 |
| PKD_channel     | PF08016.14 | 0.036 | 0.0396455696202532 |
| PRELI           | PF04707.16 | 0.035 | 0.0396455696202532 |
| Prot_ATP_ID_OB  | PF16450.7  | 0.012 | 0.0263023255813953 |
| Ricin_B_lectin  | PF00652.24 | 0.036 | 0.0396455696202532 |
| RVT_1           | PF00078.29 | 0.035 | 0.0396455696202532 |
| SCP2            | PF02036.19 | 0.012 | 0.0263023255813953 |
| SLC12           | PF03522.17 | 0.035 | 0.0396455696202532 |
| TatD_DNase      | PF01026.23 | 0.035 | 0.0396455696202532 |
| TPP_enzyme_C    | PF02775.23 | 0.005 | 0.0263023255813953 |
| TPP_enzyme_M    | PF00205.24 | 0.035 | 0.0396455696202532 |
| XPG_I           | PF00867.20 | 0.035 | 0.0396455696202532 |
| Zeta_toxin      | PF06414.14 | 0.035 | 0.0396455696202532 |
| 4F5             | PF04419.16 | 0.001 | 0.0263023255813953 |
| Bravo_FIGEY     | PF13882.8  | 0.013 | 0.0263023255813953 |
| CERK_C          | PF19280.1  | 0.013 | 0.0263023255813953 |
| COX6B           | PF02297.19 | 0.013 | 0.0263023255813953 |
| DHO_dh          | PF01180.23 | 0.013 | 0.0263023255813953 |
| DUF4139         | PF13598.8  | 0.013 | 0.0263023255813953 |
| DUF4140         | PF13600.8  | 0.013 | 0.0263023255813953 |
| EF_TS           | PF00889.21 | 0.013 | 0.0263023255813953 |
| FAD_binding_4   | PF01565.25 | 0.001 | 0.0263023255813953 |
| Fer4_21         | PF14697.8  | 0.013 | 0.0263023255813953 |
| Gln-synt_C      | PF00120.26 | 0.013 | 0.0263023255813953 |
| Glyco_hydro_15  | PF00723.23 | 0.013 | 0.0263023255813953 |
| Indigoidine_A   | PF04227.14 | 0.013 | 0.0263023255813953 |
| KASH_CCD        | PF14662.8  | 0.013 | 0.0263023255813953 |
| KPBB_C          | PF19292.1  | 0.013 | 0.0263023255813953 |
| Lyase_1         | PF00206.22 | 0.013 | 0.0263023255813953 |
| P-mevalo_kinase | PF04275.16 | 0.013 | 0.0263023255813953 |
| PTPlike_phytase | PF14566.8  | 0.013 | 0.0263023255813953 |
| Ribosomal_L13   | PF00572.20 | 0.013 | 0.0263023255813953 |
| SRCR            | PF00530.20 | 0.013 | 0.0263023255813953 |

|              |            |       |                    |
|--------------|------------|-------|--------------------|
| TPP_enzyme_N | PF02776.20 | 0.013 | 0.0263023255813953 |
| XPG_N        | PF00752.19 | 0.013 | 0.0263023255813953 |

Table S4. KEGG pathways significantly more likely to be retained (i.e. found in ohnologs). Determined using the Fisher's 2x2 matrix.

| KEGG Pathway | KEGG Description               | p value | FDR adj. p value |
|--------------|--------------------------------|---------|------------------|
| 03018        | RNA degradation                | 0.025   | 0.0312           |
| 03022        | Basal transcription factors    | 0.026   | 0.0312           |
| 04120        | Ubiquitin mediated proteolysis | 0.005   | 0.024            |
| 04144        | Endocytosis                    | 0.010   | 0.03             |
| 04150        | mTOR signaling pathway         | 0.006   | 0.024            |

Table S5. KEGG pathways significantly more likely to be lost (i.e. found in single copies). Determined using the Fisher's 2x2 matrix.

| KEGG Pathway | KEGG Description                        | p value | FDR adj. p value   |
|--------------|-----------------------------------------|---------|--------------------|
| 00030        | Pentose phosphate pathway               | 0.029   | 0.0316363636363636 |
| 00190        | Oxidative phosphorylation               | 0.006   | 0.024              |
| 00220        | Arginine biosynthesis                   | 0.026   | 0.0312             |
| 00620        | Pyruvate metabolism                     | 0.014   | 0.0312             |
| 00630        | Glyoxylate and dicarboxylate metabolism | 0.020   | 0.0312             |
| 00640        | Propanoate metabolism                   | 0.046   | 0.046              |
| 04146        | Peroxisome                              | 0.023   | 0.0312             |

Table S6. List of Pfam domains predicted for ohnologs with the top 5% highest dN/dS values

| Domain  | Accession  | Frequency |
|---------|------------|-----------|
| BBIP10  | PF14777.8  | 1         |
| CAP     | PF00188.28 | 1         |
| Chromo  | PF00385.26 | 1         |
| DUF4256 | PF14066.8  | 1         |
| ECH_1   | PF00378.22 | 1         |
| GBP     | PF02263.21 | 1         |
| MLANA   | PF14991.8  | 1         |
| Sod_Cu  | PF00080.22 | 1         |
| TTR-52  | PF01060.25 | 1         |



Table S7: Table showing the top 20 ohnolog pairs with the highest dN/dS values (minimum dS  $\geq 0.01$ ). The corresponding expression values and absolute log<sub>2</sub>(FC) values are also shown. No *C. elegans* orthologs were found for these ohnologs.

| Orthogroup | Ohnolog 1             | Ohnolog 2             | dN   | dS   | dN/dS | Mean FPKM Ohnolog 1 | Mean FPKM Ohnolog 2 | Absolute log <sub>2</sub> (FC) | <i>C.elegans</i> ortholog |
|------------|-----------------------|-----------------------|------|------|-------|---------------------|---------------------|--------------------------------|---------------------------|
| OG0017566  | ALDISUDHAUS000007547  | ALDISUDHAUS000007599  | 0,10 | 0,01 | 9,32  | 2,14                | 4,44                | 1,05                           | NA                        |
| OG0019875  | ALDISUDHAUS0000031914 | ALDISUDHAUS0000031976 | 0,19 | 0,02 | 7,88  | 0,00                | 0,11                | Inf                            | NA                        |
| OG0020086  | ALDISUDHAUS0000032896 | ALDISUDHAUS0000039664 | 0,08 | 0,01 | 7,79  | 1,99                | 1,96                | 0,03                           | NA                        |
| OG0019055  | ALDISUDHAUS0000026520 | ALDISUDHAUS0000030560 | 0,11 | 0,01 | 7,23  | 0,52                | 0,10                | 2,43                           | NA                        |
| OG0018831  | ALDISUDHAUS0000024096 | ALDISUDHAUS0000027749 | 0,09 | 0,01 | 7,02  | 0,77                | 1,38                | 0,85                           | NA                        |
| OG0020112  | ALDISUDHAUS0000033025 | ALDISUDHAUS0000033311 | 0,09 | 0,01 | 6,96  | 2,18                | 1,57                | 0,48                           | NA                        |
| OG0020269  | ALDISUDHAUS0000033887 | ALDISUDHAUS0000035022 | 0,15 | 0,02 | 6,85  | 2,33                | 0,31                | 2,91                           | NA                        |
| OG0018255  | ALDISUDHAUS0000019364 | ALDISUDHAUS0000036621 | 0,07 | 0,01 | 6,71  | 19,63               | 19,26               | 0,03                           | NA                        |
| OG0018894  | ALDISUDHAUS0000024662 | ALDISUDHAUS0000024950 | 0,09 | 0,01 | 6,34  | 2,39                | 1,04                | 1,20                           | NA                        |
| OG0019818  | ALDISUDHAUS0000031678 | ALDISUDHAUS0000039109 | 0,26 | 0,04 | 6,23  | 7,27                | 0,26                | 4,81                           | NA                        |
| OG0020291  | ALDISUDHAUS0000034016 | ALDISUDHAUS0000039894 | 0,08 | 0,01 | 6,22  | 0,13                | 0,46                | 1,78                           | NA                        |
| OG0014260  | ALDISUDHAUS0000022716 | ALDISUDHAUS0000027390 | 0,06 | 0,01 | 6,08  | 3562,42             | 3119,11             | 0,19                           | NA                        |
| OG0014326  | ALDISUDHAUS0000024525 | ALDISUDHAUS0000025343 | 0,08 | 0,01 | 5,95  | 105,12              | 119,19              | 0,18                           | NA                        |
| OG0019209  | ALDISUDHAUS0000028500 | ALDISUDHAUS0000028726 | 0,06 | 0,01 | 5,81  | 0,99                | 0,89                | 0,16                           | NA                        |
| OG0020125  | ALDISUDHAUS0000033105 | ALDISUDHAUS0000037123 | 0,07 | 0,01 | 5,76  | 0,50                | 0,59                | 0,25                           | NA                        |
| OG0017584  | ALDISUDHAUS0000007902 | ALDISUDHAUS0000010320 | 0,07 | 0,01 | 5,71  | 7,58                | 5,80                | 0,38                           | NA                        |
| OG0014673  | ALDISUDHAUS0000034782 | ALDISUDHAUS0000038967 | 0,12 | 0,02 | 5,54  | 1,00                | 0,58                | 0,79                           | NA                        |
| OG0019969  | ALDISUDHAUS0000032319 | ALDISUDHAUS0000033823 | 0,07 | 0,01 | 5,44  | 3,50                | 1,12                | 1,64                           | NA                        |
| OG0020076  | ALDISUDHAUS0000032839 | ALDISUDHAUS0000036366 | 0,06 | 0,01 | 5,38  | 0,00                | 0,82                | Inf                            | NA                        |
| OG0018895  | ALDISUDHAUS0000024683 | ALDISUDHAUS0000028406 | 0,06 | 0,01 | 5,31  | 1,57                | 1,58                | 0,01                           | NA                        |

Table S8: Table showing the top 20 ohnolog pairs with the highest differential expression differences. The top log<sub>2</sub>(FC) values are all infinite due to one ohnolog copy having zero expression. The dN, dS and dN/dS values are also shown, as well as the *C. elegans* orthologs.

| Orthogroup | Ohnolog 1             | Ohnolog 2            | dN   | dS   | dN/dS | Mean FPKM Ohnolog 1 | Mean FPKM Ohnolog 2 | Absolute log <sub>2</sub> (FC) | <i>C. elegans</i> ortholog |
|------------|-----------------------|----------------------|------|------|-------|---------------------|---------------------|--------------------------------|----------------------------|
| OG0010977  | ALDISUDHAUS000025105  | ALDISUDHAUS000027375 | 0,00 | 0,10 | 0,03  | 1003,90             | 0                   | Inf                            | <i>C26F1,4 (rps-30)</i>    |
| OG0014240  | ALDISUDHAUS000022219  | ALDISUDHAUS000026032 | 0,04 | 0,01 | 3,99  | 308,97              | 0                   | Inf                            |                            |
| OG0018098  | ALDISUDHAUS000017139  | ALDISUDHAUS000018536 | 0,01 | 0,25 | 0,06  | 298,47              | 0                   | Inf                            |                            |
| OG0020253  | ALDISUDHAUS0000033770 | ALDISUDHAUS000037970 | 0,03 | 0,17 | 0,18  | 164,75              | 0                   | Inf                            |                            |
| OG0017709  | ALDISUDHAUS000010158  | ALDISUDHAUS000022172 | 0,02 | 0,12 | 0,19  | 66,55               | 0                   | Inf                            |                            |
| OG0006834  | ALDISUDHAUS000021463  | ALDISUDHAUS000021814 | 0,01 | 0,12 | 0,05  | 59,86               | 0                   | Inf                            | <i>K10C3,2 (ensa-1)</i>    |
| OG0004688  | ALDISUDHAUS000006992  | ALDISUDHAUS000008228 | 0,00 | 0,09 | 0,00  | 57,63               | 0                   | Inf                            | <i>R13A5,8 (rpl-9)</i>     |
| OG0017689  | ALDISUDHAUS000009667  | ALDISUDHAUS000023377 | 0,06 | 0,07 | 0,79  | 43,59               | 0                   | Inf                            |                            |
| OG0009384  | ALDISUDHAUS000029847  | ALDISUDHAUS000031626 | 0,00 | 0,17 | 0,00  | 38,64               | 0                   | Inf                            | <i>Y41E3,8a</i>            |
| OG0007269  | ALDISUDHAUS0000033948 | ALDISUDHAUS000039075 | 0,00 | 0,14 | 0,03  | 32,29               | 0                   | Inf                            | <i>Y37D8A,14 (cox-5A)</i>  |
| OG0017562  | ALDISUDHAUS000007512  | ALDISUDHAUS000019487 | 0,59 | 1,29 | 0,46  | 29,01               | 0                   | Inf                            |                            |
| OG0019745  | ALDISUDHAUS0000031316 | ALDISUDHAUS000033468 | 0,03 | 0,08 | 0,35  | 27,88               | 0                   | Inf                            |                            |
| OG0019571  | ALDISUDHAUS0000030628 | ALDISUDHAUS000033858 | 0,21 | 0,55 | 0,39  | 26,00               | 0                   | Inf                            |                            |
| OG0020406  | ALDISUDHAUS0000034697 | ALDISUDHAUS000036548 | 0,02 | 0,16 | 0,11  | 23,04               | 0                   | Inf                            |                            |
| OG0007582  | ALDISUDHAUS0000032431 | ALDISUDHAUS000037466 | 0,00 | 0,23 | 0,02  | 20,99               | 0                   | Inf                            | <i>C32E8,3 (tppp-1)</i>    |
| OG0020192  | ALDISUDHAUS0000033511 | ALDISUDHAUS000040026 | 0,02 | 0,03 | 0,68  | 20,75               | 0                   | Inf                            |                            |
| OG0020603  | ALDISUDHAUS0000036095 | ALDISUDHAUS000038387 | 0,07 | 0,04 | 1,66  | 20,30               | 0                   | Inf                            |                            |
| OG0019197  | ALDISUDHAUS000028356  | ALDISUDHAUS000038200 | 0,02 | 0,10 | 0,25  | 19,51               | 0                   | Inf                            |                            |
| OG0006166  | ALDISUDHAUS000007646  | ALDISUDHAUS000008647 | 0,00 | 0,16 | 0,00  | 18,34               | 0                   | Inf                            | <i>F39B2,2 (uev-1)</i>     |
| OG0017753  | ALDISUDHAUS000011072  | ALDISUDHAUS000025136 | 0,32 | 0,15 | 2,17  | 18,17               | 0                   | Inf                            |                            |

Table S9. Successful CRISPR mutants with their exact mutations listed. Most mutants came from the same injected P0 worm A. Two double mutants were obtained from one injection. Both these mutants had a dumpy phenotype, despite being in-frame. The mutations in their sequence are listed, with the parts containing the gRNA sequence underlined. The majority of mutations were deletions. The strain and allele names of the lines that were frozen is listed.

| No. | Injected P0 | Mutation                        |                | Dumpy phenotype? | Mutation in Sequence                                    |                                        | Strain | Allele |
|-----|-------------|---------------------------------|----------------|------------------|---------------------------------------------------------|----------------------------------------|--------|--------|
|     |             | <i>dpy-1-A</i>                  | <i>dpy-1-B</i> |                  | <i>dpy-1-A</i>                                          | <i>dpy-1-B</i>                         |        |        |
| 1   |             |                                 | 6bp del        | X                |                                                         | <u>GAGAACT</u> -----GTGG               |        |        |
| 2   |             | 6bp indel<br>(14bp ins+8bp del) |                | X                |                                                         |                                        | RS3704 | tu1470 |
| 3   |             |                                 | 3bp del        | ✓                | <u>AAGAG</u> ----- <u>GGAATCATCAGATT</u> <u>CT</u> GTGG | <u>GAGAACTC</u> -- <u>CT</u> GTGG      |        |        |
| 4   | A           | 5bp del                         | 7bp ins        | ✓                | <u>GAACTCCAG</u> -----GCTCATCCG                         | <u>GAACTCCAG</u> <u>AAGAGAACT</u> GTGG | RS3702 | tu1468 |
| 5   |             |                                 | 7bp ins        | X                |                                                         |                                        | RS3705 | tu1471 |
| 6   |             | 5bp del                         |                | X                | <u>GAACTCCAG</u> -----GCTCATCCG                         |                                        | RS3703 | tu1469 |
| 7   |             |                                 | 12bp del       | X                |                                                         | <u>GAGAACTC</u> -----ATCCGAGT          |        |        |
| 8   | B           | 6bp del                         |                | X                | <u>GAGAACT</u> -----GTGG                                |                                        |        |        |
| 9   | C           | 6bp del                         |                | X                | <u>GAGAA</u> ----- <u>CT</u> GTGG                       |                                        |        |        |

# **Title: Conserved switch genes regulate a novel cannibalistic morph after whole genome duplication**

**Authors:** Sara Wighard<sup>1</sup>, Hanh Witte<sup>1</sup> and Ralf J. Sommer<sup>1\*</sup>

**Affiliation:** <sup>1</sup>Max Planck institute for Biology, Tübingen, 72076, Germany

\*Corresponding author. Email: [ralf.sommer@tuebingen.mpg.de](mailto:ralf.sommer@tuebingen.mpg.de)

## **Abstract**

Developmental plasticity facilitates morphological and behavioural novelty, but associated regulatory mechanisms remain elusive. Nematodes have emerged as a powerful model to study developmental plasticity and its evolution. Here, we show the predatory nematode *Allodiplogaster sudhausi* evolved an additional third mouth morph, concomitant with whole genome duplication (WGD) and a strong increase in body size. The three mouth morphs are induced by different diets; bacteria, fungi and nematodes. CRISPR experiments indicate that regulation of the third morph involves co-option of a conserved developmental switch gene, which through WGD resulted in two mouth-form regulators. Gene dosage studies revealed a diverged role of these developmental switches, with functional redundancy and quantitative effects in the two mouth-form decisions, respectively. The third morph is cannibalistic and kills kin, whereas the other two morphs do not. Thus, the recent evolution of a new morph relies on pre-existing regulatory mechanisms and adds behavioural and social complexity.

**One-Sentence Summary:** Experimental genetics in a nematode reveals a key role for developmental plasticity in the evolution of nutritional diversity

## Main Text

Novel morphological structures and behaviors are often observed in developmentally plastic systems, including ant worker castes and predatory spadefoot toads (1, 2). However, the regulatory processes and molecular mechanisms that lead to novel traits remain elusive (3). Does the evolution of novelty require large-scale genomic alterations? Is it controlled by pre-existing or novel regulatory genes and mechanisms? And what type of environmental variations, e.g. changes in diet, are involved in such evolutionary diversifications? Nematodes can help address such questions given their easy husbandry, simple diets, and rapid genetic and experimental manipulation, in particular in isogenic self-fertilizing hermaphrodites.

The developmentally plastic soil nematode *Allodiplogaster sudhausi* has several unique features, including its extreme body size (Fig. 1A, B) and predatory activity as juveniles and adults (4). As a distant relative of the model organism *Pristionchus pacificus* (5), it is one of the few self-fertilizing hermaphrodites outside *Pristionchus* with available CRISPR technology tools (6). *A. sudhausi* forms teeth-like denticles that were previously described to exhibit developmental plasticity with two discrete morphs, the narrow-mouthed ‘stenostomatous’ (St) and the wide-mouthed ‘eurystomatous’ (Eu) (4). In contrast to other predatory species, both mouth forms are able to kill other nematodes, including *C. elegans*. We recently showed *A. sudhausi* underwent a whole genome duplication (WGD) that is not seen in its closest relative *Allodiplogaster seani* (6). Additionally, *A. seani* is similar in size to *C. elegans* and *P. pacificus* (Fig. 1B), indicating that the extreme increase in body size of *A. sudhausi* correlates with the WGD event.

### **A. sudhausi exhibits three distinct mouth-forms**

WGD is often associated with morphological novelties; however, it is unknown if it also affects developmentally plastic traits. To observe potential changes in mouth-form plasticity of *A. sudhausi*, we cultured these worms on a collection of bacterial, fungal and nematode diets. Remarkably, we found three distinct adult mouth-forms when *A. sudhausi* was grown on *Escherichia coli* OP50 bacteria, *C. elegans* N2 worms and *Penicillium camemberti* fungi (Fig. 1 C-H). Specifically, on *E. coli*, the standard maintenance method for nematodes, *A. sudhausi* formed the narrow St morph (Fig. 1C, F; fig. S1). On *C. elegans*, an inducer of Eu in other nematodes (7),

*A. sudhausi* also became Eu (Fig. 1D, G). In contrast, *A. sudhausi* was able to form a third morph on *P. camemberti* (Fig. 1E, H). While the *E. coli* and *C. elegans* diets caused 100% St and Eu morph formation, respectively, the fungal diet resulted in both the St and third morph. This new morph was also seen at low frequency on starved *E. coli* plates (fig. S2). Measurements of the stoma (mouth width) and geometric morphometrics ( $\delta$ ) of the mouth revealed significant differences between the three morphs in size and shape (Fig. 1I, J). As the third morph has the largest mouth, we termed it ‘teratostomatous’ (Te), from the Greek word *teras* for monster. Note that previous studies in *A. sudhausi* only observed two distinct morphs, but they did not test different dietary conditions (4, 9). Taken together, we have identified a third discrete mouth-form in *A. sudhausi*.

Next, we investigated whether the Te morph is restricted to *A. sudhausi* or if it is also found in other nematodes. We focused on the sister species *A. seani* and grew it under the same three dietary conditions. *A. seani* became St on *E. coli* and *P. camemberti*, and Eu on *C. elegans*; however, it did not form the Te morph (fig. S3). Additionally, we tested a range of different bacteria, fungi and other worms, but were unable to produce any Te equivalent on *A. seani*. Starved plates also did not contain the Te morph. Similarly, no other species of *Allodiplogaster* or the closely related genus *Koerneria* have been described to form three mouth morphs. Thus, the Te morph is only found in *A. sudhausi* and likely evolved *de novo* after the two *Allodiplogaster* species diverged. Together, these findings indicate that *A. sudhausi* shows three novel traits that are not present in its closest relative; WGD, increased body size and a third mouth-form.

### **The Te morph is induced by poor nutrition, crowding and displays cannibalism**

The induction of a third morph by *P. camemberti* was unexpected. Why would a fungus better known for the cheese it makes induce a new morph? We tested other fungi, including *Penicillium rubens*, *Penicillium digitatum*, *Emericella nidulans*, *Aspergillus* spp., and *Botrytis cinerea*, which all induced the Te morph (table S1). Therefore, Te morph induction is due to the fungal diet and not *P. camemberti* in particular. One potential reason might be the low nutritional value of these fungi, which could initiate a stress response in *A. sudhausi*. This would be consistent with our observation that the Te morph is seen on starved *E. coli* plates. Besides starvation, crowding also results in stressful environments in nematodes. Indeed, we

found that crowding increased Te occurrences on *P. camemberti* (Fig. 2A). When few worms were present, the majority became St; however, higher density led to increased Te incidence. This trend was seen across differently sized plates, with smaller plates requiring less worms to reach the same Te percentage as larger plates (Fig. 2B). In contrast, crowding did not result in the formation of the Te morph on *E. coli* (fig. S4). These results suggest that stressful environmental conditions, combining crowding and low nutrition, induce the Te morph. Consistent with this, we show that Te animals have a smaller brood size and shorter body length than St and Eu animals (fig. S5). *A. sudhausi* therefore appears to have evolved a new phenotype in response to stressful conditions. However, what is the ecological significance of being Te?

Stressful and low nutritional conditions induce cannibalism in a number of vertebrates and invertebrates (2, 10). In nematodes, previous studies in *P. pacificus* have shown morph-specific cannibalism (11). These studies also revealed that the likelihood of cannibalism negatively correlates with genetic relatedness of *P. pacificus* isolates and isogenic kin do not kill each other at all (12). In contrast, we observed regular cannibalism of the Te morph on their kin (Fig. 2C). Specifically, systematic studies using well-established predation assays (13) revealed that only the Te morph exhibits cannibalism against kin, whereas St and Eu animals did not (Fig. 2D,E; fig. S6). Although adult Te worms usually prey on juveniles (Movie S1), cannibalism against adults can also occur including on other adult Te animals. Cannibalism occurred in Te from both *P. camemberti* and starved plates (Fig. 2D,E). Movie S2 shows an adult Te that quickly digests another Te adult (Fig. 2F,G). Thus, cannibalism in *A. sudhausi* appears Te-specific and occurs despite the worms being isogenic. Therefore, our findings represent a novel type of cannibalism.

To determine if the new behaviour of Te adult morphs is associated with differential gene expression, we generated transcriptomes of adult worms of all three morphs. Specifically, we compared four distinct conditions; Eu animals from *C. elegans* cultures, St animals from *E. coli* and *P. camemberti* cultures and Te animals also from *P. camemberti* cultures. We found that Te morphs had markedly different expression patterns (Fig. 2H). In total, we identified 1841 differentially expressed genes that are shared in Te vs. the other three conditions ( $\log_2FC > 2$ , adj.  $p < 0.5$ ) (Fig. 2I). Notably, genes containing a heat shock protein (HSP20) domain were significantly upregulated in Te morphs. As HSPs are commonly activated during

stress response (14, 15), this finding further supports the idea that the Te morph is a response to low nutritional conditions resulting from starvation and crowding.

### **Conserved sulfatase switch genes control morph-form regulation**

Next, we wanted to know how mouth-form plasticity in *A. sudhausi* is genetically determined. In principle, the regulation of the Te morph may involve a novel regulatory network or the co-option of pre-existing machinery. To examine this, we looked at the *P. pacificus* mouth-form gene regulatory network, which has been well characterized (16) (Fig. 3A). A supergene locus plays a crucial role in *P. pacificus* mouth-form determination (17, 18). Specifically, the *eud-1*/sulfatase gene acts as a developmental switch with *eud-1* mutants constitutively St under all conditions (19). The *eud-1* switch gene is homologous to *C. elegans sul-2* and has two more paralogs in *P. pacificus*. However, only *Ppa-eud-1* is involved in mouth-form regulation, whereas mutants in the other two genes display wild type mouth forms. In *A. sudhausi*, four homologs of *Ppa-eud-1* have been identified, which are not clustered as in *P. pacificus* (17, 20, 21). These four genes represent two pairs resulting from the recent WGD. We refer to the one pair as *Asu-sul-2-A* and *Asu-sul-2-B*, and the other pair as *Asu-sul-2-D* and *Asu-sul-2-E* (Fig. 3B, table S2) (20).

We targeted all four *A. sudhausi sul-2* homologs using CRISPR technology. Given the high sequence similarity of the WGD-derived gene pairs, we used identical sgRNAs for each and were able to generate single and double knockouts per pair (Fig. 3C). This strategy allowed us to obtain quadruple, double and single sulfatase gene knockouts. The quadruple mutant displayed the St morph under both *C. elegans* and *E. coli* diets, indicating that sulfatases regulate the Eu phenotype in *A. sudhausi*, identical to *P. pacificus* (Fig. 3D). Strikingly, when we grew the quadruple mutant on *P. camemberti*, all animals were St as well. Thus, both mouth-form decisions are regulated by *sul-2*/sulfatases.

### **The Te morph is regulated by the same genes that control Eu formation**

Next, we wanted to know if all sulfatases are involved in mouth-form regulation or if only one pair of sulfatase genes is functionally homologous to *Ppa-eud-1*. We found that the *Asu-sul-2-D/E* double mutant knockout displayed a wild type phenotype under all three dietary conditions, indicating this pair of genes has no role in mouth-form determination (Fig. 3D; fig S7). In contrast, the *Asu-sul-2-A/B* double



knockout did not form either the Eu or Te morphs (Fig. 3D; table S3). Thus, the *Asu-sul-2-A/B* gene pair is functionally homologous to *Ppa-eud-1* in term of Eu formation, and additionally controls the *A. sudhausi*-specific Te morph.

These findings culminate in the question as to whether *Asu-sul-2-A* and *Asu-sul-2-B* are redundant in mouth-form regulation. If both genes are fully redundant, single gene knockouts would be phenotypically wild type. In contrast, if *Asu-sul-2-A* and *Asu-sul-2-B* acquired different functions one would expect distinct phenotypes of single mutant animals. We found that single mutants displayed a wild type phenotype when grown on *P. camemberti* (Fig. 3E). This finding indicates that both genes are functionally redundant for Te morph formation.

In contrast, single mutants grown on *C. elegans* suggest more complex regulatory interactions involving dosage effects in Eu morph formation. Specifically, we found an intermediate mouth width of *Asu-sul-2-A* and *Asu-sul-2-B* single mutants when compared with wild type and double mutant animals (Fig. 3F). This surprising finding might indicate a quantitative role of *Asu-sul-2-A* and *Asu-sul-2-B* for the formation of the Eu phenotype. To test this hypothesis, we crossed wild type and mutant lines to generate progeny with four, three, two, one and zero wild type copies of *Asu-sul-2-A/B*. Phenotypic characterization of these individuals indicates a gradual decrease of mouth width from four wild type copies to one (Fig. 3G). Note the strong drop from one to zero wild type copies, suggesting that one functional copy of either gene is sufficient to recapitulate Eu-type features. These results support a dosage-dependent function of *Asu-sul-2-A/B* in the regulation of the Eu morph. Together, our findings indicate that *Asu-sul-2-A/B* control both Eu and Te formation with different regulatory mechanisms in both mouth-form decisions.

### **Developmental plasticity and WGD facilitate morphological and behavioural diversity**

In summary, the discovery of an additional mouth morph associated with cannibalistic behaviour supports the importance of developmental plasticity as a major driver of morphological and behavioural diversification. Our study identifies a pre-existing developmental switch gene that controls the Eu vs. St decision to then be co-opted for the regulation of the Te morph. The co-occurrence of three novel traits; WGD, doubling in body size and the evolution of a third mouth-form, are intriguing. We are currently unable to determine the exact order in which these

characters have been gained, due to the absence of additional wild isolates and/or more closely related species. However, the WGD-derived *Asu-sul-2-A/B* already show functional divergence in Eu morph formation. Such molecular processes provide ecological opportunities highlighting the importance of plasticity. The cannibalistic Te morph, induced by crowding and starvation, enables greater chance of long-term survival in stressful conditions. This phenomenon relates to examples of cannibalism in salamanders, toads and fish (2, 10). However, *A. sudhausi* represents, to the best of our knowledge, the first animal system with regular cannibalism against genetically identical kin. Together, nematode mouth-form plasticity and predation provide a powerful system to investigate evolutionary innovation, its underlying molecular mechanisms, and behavioural and ecological consequences.

## References

1. M. J. West-Eberhard, *Developmental Plasticity and Evolution* (Oxford University Press, 2003).
2. D. W. Pfennig, P. W. Sherman, J. P. Collins, Kin recognition and cannibalism in polyphenic salamanders. *Behavioral Ecology*. **5**, 225–232 (1994).
3. R. J. Sommer, Phenotypic Plasticity: From Theory and Genetics to Current and Future Challenges. *Genetics*. **215**, 1–13 (2020).
4. A. Fürst von Lieven, *Koerneria sudhausi* n. sp. (Nematoda: Diplogastridae); a hermaphroditic diplogastrid with an egg shell formed by zygote and uterine components. *Nematology*. **10**, 27–45 (2008).
5. N. E. Schroeder, Introduction to *Pristionchus pacificus* anatomy. *Journal of Nematology*. **53**, 1–9 (2021).
6. S. S. Wighard, M. Athanasouli, H. Witte, C. Rödelberger, R. J. Sommer, A New Hope: A Hermaphroditic Nematode Enables Analysis of a Recent Whole Genome Duplication Event. *Genome Biology and Evolution*. **14**, evac169 (2022).
7. M. Wilecki, J. W. Lightfoot, V. Susoy, R. J. Sommer, Predatory feeding behaviour in *Pristionchus* nematodes is dependent on phenotypic plasticity and induced by serotonin. *Journal of Experimental Biology*. **218**, 1306–1313 (2015).
8. T. Theska, B. Sieriebriennikov, S. S. Wighard, M. S. Werner, R. J. Sommer, Geometric morphometrics of microscopic animals as exemplified by model nematodes. *Nature Protocols*. **15**, 2611–2644 (2020).
9. V. Susoy, E. J. Ragsdale, N. Kanzaki, R. J. Sommer, Rapid diversification associated with a macroevolutionary pulse of developmental plasticity. *eLife*. **4**, 1–39 (2015).
10. L. R. Fox, Cannibalism in Natural Populations. *Annual Review of Ecology and Systematics*. **6**, 87–106 (1975).
11. J. W. Lightfoot, M. Wilecki, C. Rödelberger, E. Moreno, V. Susoy, H. Witte, R. J. Sommer, Small peptide-mediated self-recognition prevents cannibalism in predatory nematodes. *Science*. **364**, 86–89 (2019).
12. J. W. Lightfoot, M. Dardiry, A. Kalirad, S. Giaimo, G. Eberhardt, H. Witte, M. Wilecki, C. Rödelberger, A. Traulsen, R. J. Sommer, Sex or cannibalism: Polyphenism and kin recognition control social action strategies in nematodes. *Science Advances*. **7**, eabg8042 (2021).
13. J. W. Lightfoot, M. Wilecki, M. Okumura, R. J. Sommer, Assaying Predatory Feeding Behaviors in *Pristionchus* and Other Nematodes. *JoVE (Journal of Visualized Experiments)*, e54404 (2016).

14. D.-C. Li, F. Yang, B. Lu, D.-F. Chen, W.-J. Yang, Thermotolerance and molecular chaperone function of the small heat shock protein HSP20 from hyperthermophilic archaeon, *Sulfolobus solfataricus* P2. *Cell Stress Chaperones*. **17**, 103–108 (2012).
15. A. Butov, T. Johnson, J. Cypser, I. Sannikov, M. Volkov, M. Sehl, A. Yashin, Hormesis and debilitation effects in stress experiments using the nematode worm *Caenorhabditis elegans*: the model of balance between cell damage and HSP levels. *Experimental Gerontology*. **37**, 57–66 (2001).
16. M. S. Werner, T. Loschko, T. King, S. Reich, T. Theska, M. Franz-Wachtel, B. Macek, R. J. Sommer, Histone 4 lysine 5/12 acetylation enables developmental plasticity of *Pristionchus* mouth form. *Nature Communications*. **14**, 2095 (2023).
17. B. Sieriebriennikov, N. Prabh, M. Dardiry, H. Witte, W. Röseler, M. R. Kieninger, C. Rödelsperger, R. J. Sommer, A Developmental Switch Generating Phenotypic Plasticity Is Part of a Conserved Multi-gene Locus. *Cell Reports*. **23**, 2835-2843.e4 (2018).
18. M. Dardiry, G. Eberhardt, H. Witte, C. Rödelsperger, J. W. Lightfoot, R. J. Sommer, Divergent combinations of cis-regulatory elements control the evolution of phenotypic plasticity. *PLOS Biology* (2023).
19. E. J. Ragsdale, M. R. Müller, C. Rödelsperger, R. J. Sommer, A Developmental Switch Coupled to the Evolution of Plasticity Acts through a Sulfatase. *Cell*. **155**, 922–933 (2013).
20. J. F. Biddle, E. J. Ragsdale, Regulators of an ancient polyphenism evolved through episodic protein divergence and parallel gene radiations. *Proceedings of the Royal Society B: Biological Sciences*. **287**, 20192595–20192595 (2020).
21. E. J. Ragsdale, N. A. Ivers, Specialization of a polyphenism switch gene following serial duplications in *Pristionchus* nematodes. *Evolution*. **70**, 2155–2166 (2016).
22. S. Namdeo, E. Moreno, C. Rödelsperger, P. Baskaran, H. Witte, R. J. Sommer, Two independent sulfation processes regulate mouth-form plasticity in the nematode *Pristionchus pacificus*. *Development*. **145**, dev166272 (2018).
23. L. T. Bui, N. A. Ivers, E. J. Ragsdale, A sulfotransferase dosage-dependently regulates mouthpart polyphenism in the nematode *Pristionchus pacificus*. *Nat Commun*. **9**, 4119 (2018).
24. I. A. Hope, *C. elegans: A Practical Approach* (OUP Oxford, 1999).
25. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat Methods*. **9**, 676–682 (2012).
26. B. T. Moore, J. M. Jordan, L. R. Baugh, WormSizer: High-throughput Analysis of Nematode Size and Shape. *PLOS ONE*. **8**, e57142 (2013).

27. T. Theska, B. Sieriebriennikov, S. S. Wighard, M. S. Werner, R. J. Sommer, Geometric morphometrics of microscopic animals as exemplified by model nematodes. *Nature Protocols*. **15**, 2611–2644 (2020).
28. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. (2021).
29. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).
30. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. **30**, 923–930 (2014).
31. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. **15**, 550 (2014).
32. A. Bateman, E. Birney, R. Durbin, S. R. Eddy, R. D. Finn, E. L. L. Sonnhammer, Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Research*. **27**, 260–262 (1999).

## Acknowledgments

We thank Eva Briem for capturing the movies. We thank Jürgen Berger for the SEM pictures. We also thank the Sommer Lab for advice and discussions, particularly Christian Rödelsperger.

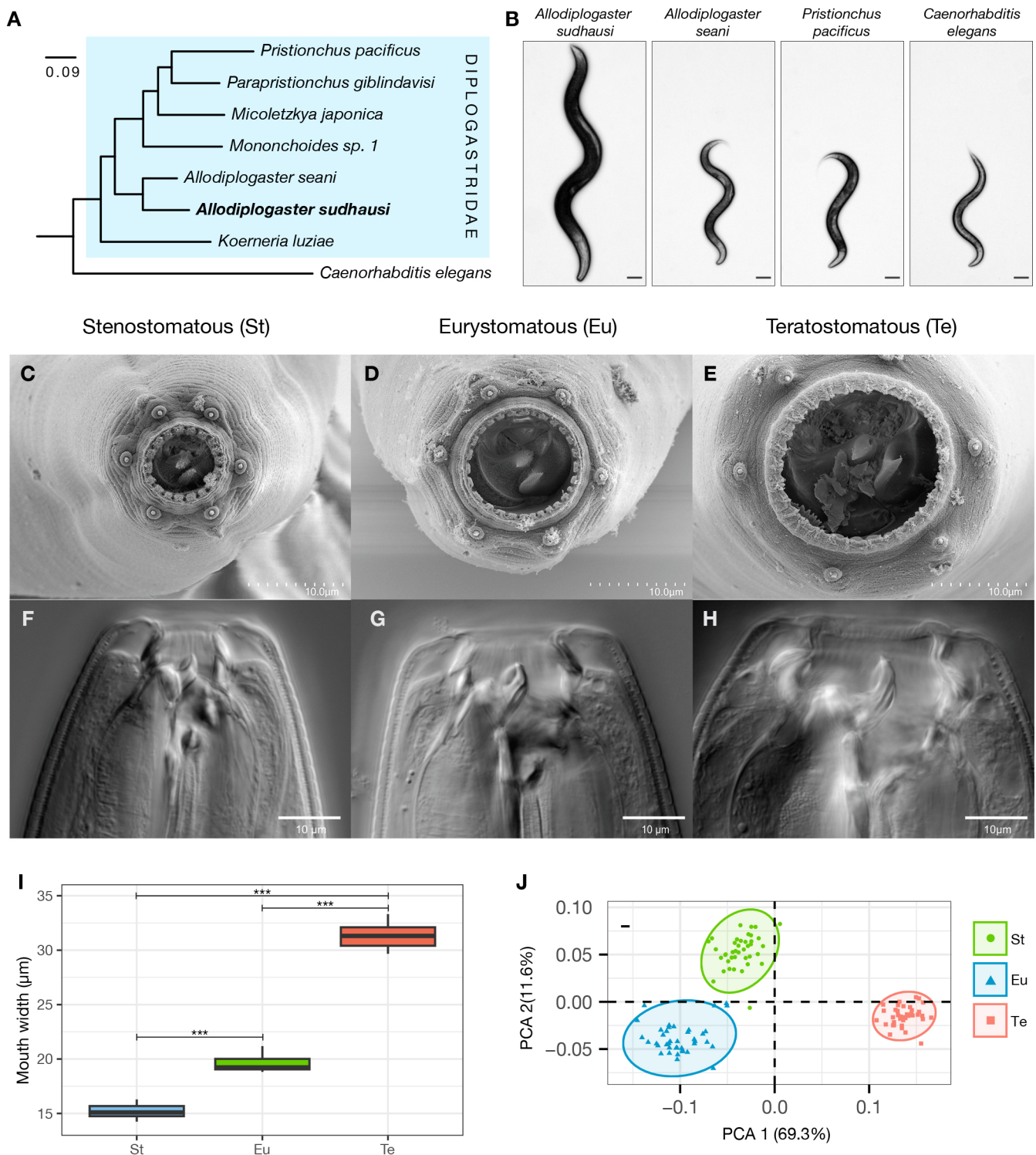
**Funding:** This work was funded by the Max Planck Society.

**Author contributions:** S.W. performed all experiments, with assistance from H.W. Data analysis was performed by S.W. Experiments were designed by S.W. and R.J.S. The manuscript was written by S.W. and R.J.S.

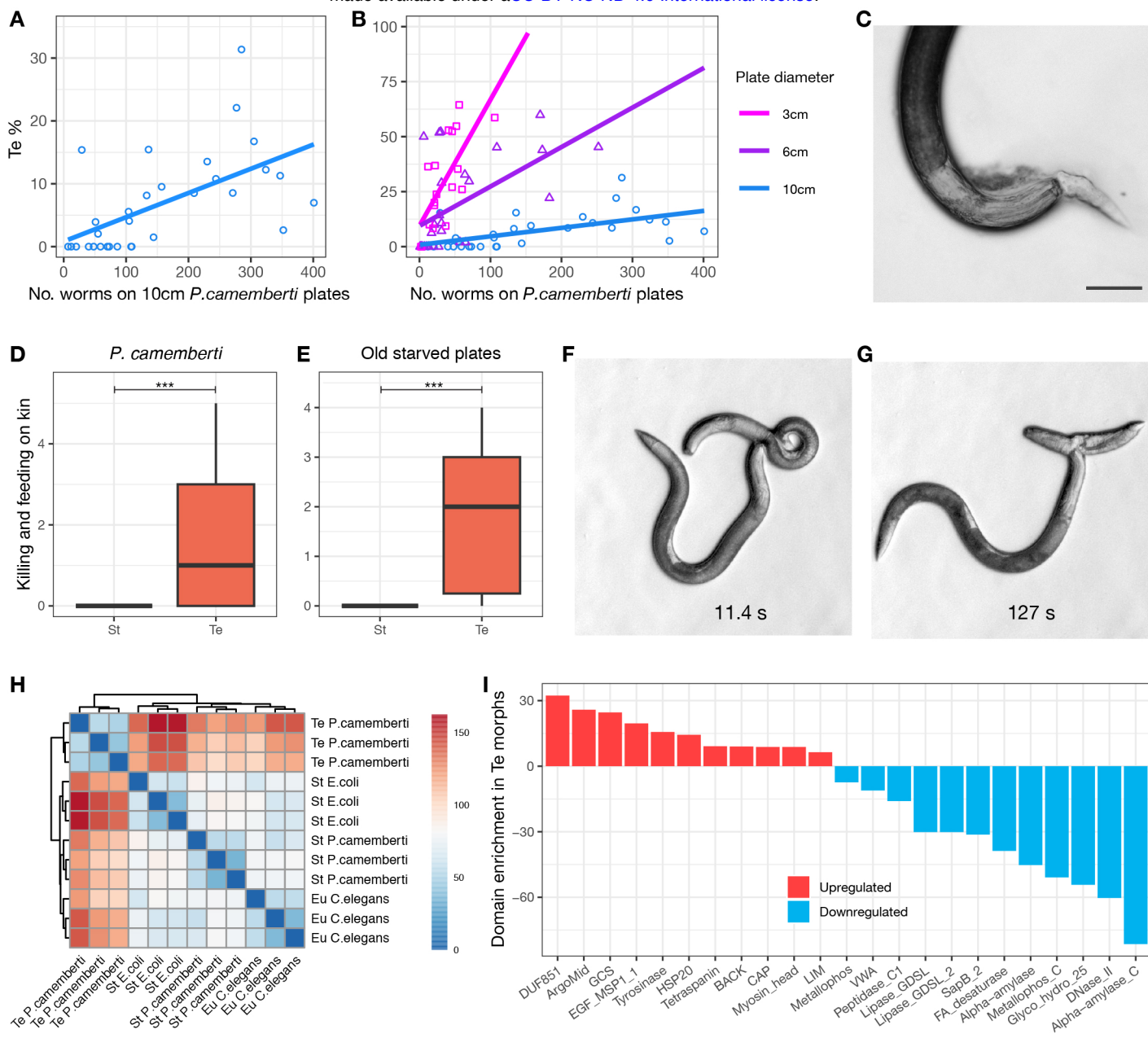
**Competing interests:** Authors declare that they have no competing interests.

**Data and materials availability:** Raw reads containing gene expression of adult morphs have been submitted to the European Nucleotide archive under the project accession number PRJEB65236. Remaining data is available in the supplementary materials.

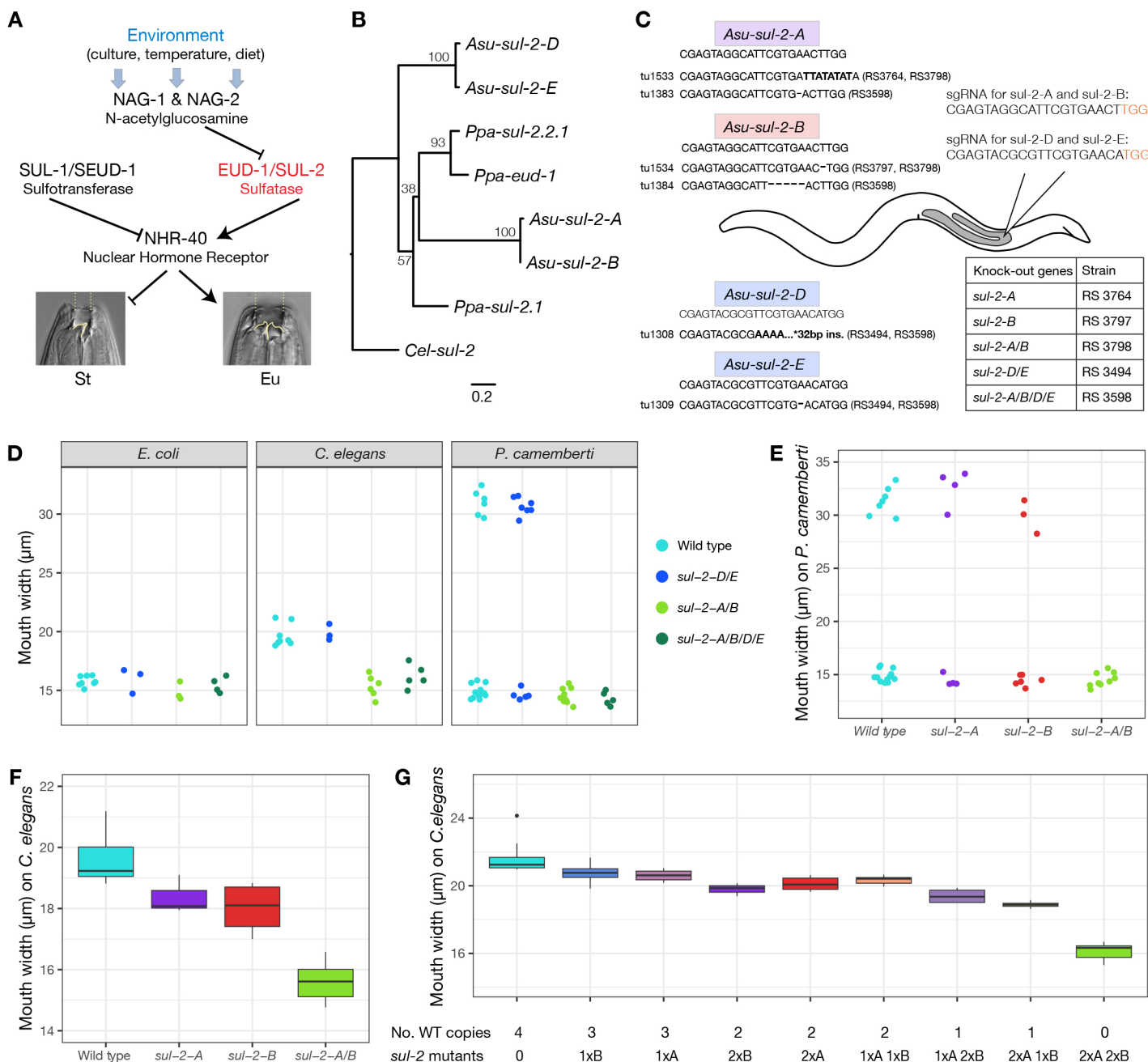




**Fig. 1. *A. sudhausi* forms an additional third mouth morph.** (A) The phylogeny shows *A. sudhausi* diverged early within the Diplogastridae. (B) The size of different hermaphrodite/female free-living nematodes is shown, with *A. sudhausi* much larger than others. (C-E) Scanning electron microscopy (SEM) pictures of *A. sudhausi* (C) St, (D) Eu and (E) Te mouths, with the dorsal tooth on top. Differential interference contrast (DIC) microscopy pictures of *A. sudhausi* (F) St, (G) Eu and (H) Te mouths, with the dorsal tooth on the left. (I) Mouth width measurements show significant size differences between St, Eu and Te mouths (Kruskal-Wallis,  $p < 0.001$ ; Pairwise Wilcox test,  $p < 0.01$  for each comparison)  $n = 8 - 12$  biological replicates). (J) Geometric morphometrics of the three morphs show significant differences in their shape dimensions (pairwise PERMANOVA,  $p < 0.001$  for each comparison)  $n = 40$  for each morph.



**Fig. 2. The Te morph cannibalizes, seemingly as a stress response.** (A-B) Crowding is positively correlated with Te morph induction. The number of adult *A. sudhausi* on (A) 10cm *P. camemberti* plates is associated with greater Te percentage (pearson's product-moment correlation,  $p < 0.001$ ) while (B) increased density increases Te induction as smaller plates show similar trends but at higher rates (pearson's product-moment correlation,  $p < 0.01$  for 6cm and  $p < 0.001$  for 3cm). Linear regression lines are plotted for all. (C) An image shows Te feeding on younger kin (the tail of the juvenile is still seen), as commonly occurs. (D) From *P. camemberti* plates, Te morphs cannibalize on their living kin, unlike St which don't kill and feed on kin at all (Mann-Whitney U test,  $p < 0.01$ ).  $n = 20$  each. (E) From old starved plates, only Te cannibalize (Mann-Whitney U test,  $p < 0.01$ ).  $n = 14$  each. (F-G) Te feed on other Te morphs and are quick at devouring prey, with pictures taken (F) 11.4 and (G) 127 seconds after feeding started. The full video can be found online (Movie S2). (H) A heat map of the adult transcriptome (measuring Euclidean distances between samples) shows that the Te morphs exhibit markedly distinct expression compared to the Eu and St samples, even from the same *P. camemberti* fungus. (I) Genes were identified that show significantly differential expression in Te versus other morphs (adj.  $p < 0.5$ ,  $\log_2(\text{FC}) > 2$ ). The Pfam domains of these genes was predicted and the significantly enriched domains that are upregulated and downregulated in Te samples is depicted.



**Fig. 3. Genome editing indicates two functionally conserved sulfatases regulate both Eu and Te morphs.** (A) Key genes that regulate mouth-form in *P. pacificus* are shown (17, 19, 22, 23). One of the major players, *eud-1*, encodes a sulfatase that regulates the Eu morph. (B) The phylogenetic relationship of the *eud-1* homologs in *P. pacificus* and *A. sudhausi* is shown, with *C. elegans sul-2* as the outgroup and bootstrap values indicating node support. There are four homologs in *A. sudhausi*, with two of each clustering, likely resulting from WGD. The relationship of the *A. sudhausi* sulfatase genes to *Ppa-eud-1* is not clear due to low bootstrap support. (C) The genotype of the *sul-2* mutant alleles is shown as well as the corresponding strains. One sgRNA was designed for each pair of genes. The PAM site next to the sgRNAs is indicated in red. (D) The mouth width of double and quadruple mutant lines was examined after being grown on the three conditions. Mutants with frameshift knock-outs in both *sul-2-A* and *sul-2-B* remain constitutively St under the three assays, in contrast to the *sul-2-D/E* mutant that displays wild type Eu and Te phenotypes on *C. elegans* and *P. camemberti*, respectively.  $n \geq 14$  biological replicates for each line. (E) The mouth width of knock-out mutants grown on *P. camemberti* is shown. The single mutants, *sul-2-A* and *sul-2-B*, are capable of becoming Te in contrast to the *sul-2-A/B* double and *sul-2-A/B/D/E* quadruple mutant lines.  $n \geq 8$  biological replicates for each



line. **(F)** The mouth width of knock-out mutants grown on *C. elegans* is shown. Interestingly, the single mutants, *sul-2-A* and *sul-2-B* have an intermediate mouth width in comparison to wild type and *sul-2-A/B* mutants. **(G)** Crosses between wild type and mutants produced progeny with varying amount of wild type copies in *sul-2-A/B* when grown on *C. elegans*. There is a gradual decrease in mouth width as wild type copy number decreases and a stark drop in mouth width when there are no wild type copies.  $n \geq 3$  biological replicates for each line.

## Materials and Methods

### Nematode maintenance and inbreeding

The *A. sudhausi* inbred strain SB413B/RS6132B (6) was used for wild type analyses. *A. seani* (RS1982) and *C. elegans* (N2) were also used. Mutant strains are frozen and kept at the Max Planck Institute for Biology. All nematode strains were maintained on nematode growth medium (NGM) agar plates with *E. coli* OP50 bacteria.

### Nematode culturing conditions

Freshly starved NGM *E. coli* plates containing many eggs were bleached (24) onto separate NGM containing three different food sources: 1) *E. coli*, 2) *C. elegans* and 3) *P. camemberti*. For *C. elegans*, plates with many larvae were washed with M9 and passed through a 20  $\mu$ m nylon net filter (Merck) onto unseeded 6 cm NGM plates, so that only larvae passed through (7). *P. camemberti* was obtained from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH (<https://www.dsmz.de>) and maintained weekly on fresh unseeded NGM plates using sterile tips to transfer spores. Note that the other tested fungi were also obtained from DSMZ, except for *Aspergillus spp.* which was a lab contaminant that was naturally isolated. After bleaching of *A. sudhausi* onto these three conditions, all assays were maintained at 20°C until the eggs grew into adults. For all experiments, only young hermaphrodites were selected (containing less than five eggs).

### Count assays

The worms were counted and phenotyped under a Zeiss Stemi 508 light microscope when they reached adulthood. The mouth-form of adult worms was determined from all three assays. This data was then analyzed to compare the overall number of adult worms and the Te percentage.

### Standard Electron Microscopy (SEM)

Young adult worms were individually picked from *P. camemberti* and *C. elegans* culture conditions. St and Te adult hermaphrodites were taken from *P. camemberti* while Eu hermaphrodite were picked from *C. elegans*. The worms were moved into 1.5ml Eppendorf tubes containing 1ml PBS (Phosphate Buffered Saline) buffer. The

SEM protocol was performed at the Electron Microscopy facility at the Max Planck Institute for Biology. The nematodes were fixed in 2.5% glutaraldehyde in PBS and post-fixed with 1% osmium tetroxide. Subsequently, samples were dehydrated in a graded ethanol series followed by critical point drying (Polaron) with CO<sub>2</sub>. Finally, the cells were sputter-coated with a 5 nm thick layer of platinum (CCU-010, Safematic) and examined with a field emission scanning electron microscope (Regulus 8230, Hitachi High Technologies).

### **Microscopy images and measurements**

Young adult hermaphrodites were fixed onto a solution with 5% Noble Agar and 0.3% NaN<sub>3</sub>, with M9 buffer for resuspension. They were imaged at 100x magnification using the Differential Interference Contrast (DIC) setting. Only worms fixed in the correct orientation (on the lateral side) were picked. Z-stacks were taken of the head region and stored as raw data czi files. Fiji/ImageJ (25) was used to measure the width of the mouth (stoma), using the base of the cheilostom and the start of the gymnostom the markers. All worms taken from the same assay plate were counted as one biological replicate.

### **Worm size measurements**

Young adult St and Te hermaphrodites from *P. camemberti* were moved to NGM plates without bacteria. Brightfield images were taken at 20x magnification using the Axio Zoom V16 microscope. Analysis was done on Fiji/ImageJ using the Wormsizer plugin for Image J/Fiji (26) to detect worm length. All worms from one assay plate were counted as one biological replicate.

### **Geometric morphometrics**

For each of the three morphs, young adult hermaphrodites were picked directly from their assay plates and image stacks were obtained using DIC microscopy, with the dorsal tooth always located on the left. Geometric morphometrics was performed using 18 fixed landmark coordinates, following the previous protocol (27). We generated landmark data for 40 individuals for each of the three morphs. Analysis was performed in RStudio (2021.09.2) (28) using the geomorph package (v. 4.0.1). Shape and form differences were visualized using PCA. To test for significant differences between morphs, we performed Procrustes ANOVA (PERMANOVA)

using procD.lm. Pairwise PERMANOVA was then performed to determine differences between groups.

### **Brood size analysis**

Young virgin *A. sudhausi* hermaphrodites that had not yet laid eggs were moved from well-fed *E. coli* assay to NGM plates containing 30  $\mu$ l litres of *E. coli*. They were transferred to fresh plates ever two days for a total of eight days. The number of live offspring that each hermaphrodite produced was counted. All plates were stored at 20°C. Progeny from adults belonging to the same plate were counted as one biological replicate.

### **RNA sequencing and analysis**

Young adult hermaphrodites were picked, moved into 20 $\mu$ l PBS buffer in 1.5ml Eppendorf tubes, and frozen at -80°C. RNA was extracted using Zymo Research Direct-zol RNA MiniPrep Kit following the manufacturer's protocol, with the addition of flash freezing in liquid nitrogen three times after adding TRIzol, to ensure successful lysis. RNA was quality controlled using the NanoDrop One Spectrophotometer and sent to Novogene Co. for library prep and Illumina sequencing (paired end 150bp, 3G per sample). Over 20 million raw reads were generated for each sample. STAR (version 2.7.9a) (29) was used to align the reads against the genome (6), which were then summarized using featureCounts (v. 2.0.2) (30) ("-p" "-T" "8"). Subsequent analysis was performed using DesSeq2 (v. 1.38.3) (31). Low count data was filtered out ( $\geq 3$  samples had a count of  $\geq 10$ ). For the heat map, data was log transformed ('blind' setting) and Euclidean distances were measured between samples. The R package pheatmap (v. 1.0.1) was used to generate a matrix showing sample distances. To examine the genes differentially expressed (DE) in Te morphs from *P. camemberti*, Te was compared to other morphs to determine DE genes that were upregulated (adj.  $p < 0.05$ ,  $\log_2(\text{FC}) > 2$ ) and downregulated (adj.  $p < 0.05$ ,  $\log_2(\text{FC}) < -2$ ) in Te versus other conditions. The common DE genes upregulated in Te versus other morphs were then selected, as were the common DE genes downregulated in Te. Protein domains were annotated for all the genes using the Pfam database (v. 3.1b2) (32) as previously described (6). A Fisher's 2x2 matrix was run, whereby each DE gene with a given domain was compared to the overall dataset to identify those that were significantly different (adj.

$p < 0.05$ , FDR). Enrichment values reflect occurrence of the significant DE domain-containing genes compared to overall genome-wide abundance.

### **Feeding on kin assays**

A feeding assay was adapted from that previously described (7): NGM was added to 12-well cell culture plates (Sarstedt). *A. sudhausi* juveniles were filtered into the wells using 41  $\mu\text{m}$  nylon net filters (Merck) so that only J3 and J2 larvae could pass through. Individual adult morphs were then added to each separate well containing younger kin. An Axio Zoom V16 microscope rotated between each well and took pictures every 20 seconds over a 90 minute period. Afterwards, the feeding events were counted. Videos S1 and S2 showing Te cannibalising was taken using an Axiocam attached to a Zeiss Stereo Discovery V20 microscope, set at 10 frames per second. The Axio zoom V16 microscope was used (Brightfield setting) to take the image of the adult Te feeding on a juvenile.

### **Genetic crosses**

*A. sudhausi* hermaphrodites at the late J4 stage were moved to separate plates containing *E. coli*. They were left for eight days on these plates, at which point they would have run out of sperm. They were then moved onto individual plates containing *E. coli* for 24 hours. Note that these plates were later checked to see if any viable progeny had in fact been produced. The old sperm-depleted hermaphrodites were then moved to separate plates containing 5  $\mu\text{l}$  of *E. coli* as well as 10 males of whichever line needed to be crossed. These worms were kept together for 24 hours to mate. They were then moved onto NGM plates containing *C. elegans* juveniles (as previously described). The resulting progeny would have one allele from the hermaphrodite and one from the male. Their adult mouth width was then measured. Note that these assays were performed for all crosses shown, even for hermaphrodites and males of the same line, so as to keep conditions consistent for comparisons.

### **Phylogenetic trees**

The species tree was previously used (6, 9). For the EUD-1/SUL-2 phylogeny, the protein sequence of *C. elegans* SUL-2 was obtained from wormbase.org (version WS284). The corresponding sequences for *P. pacificus* and *A. sudhausi* were found

using pristonchus.org. For *P. pacificus*, the annotation (El Paco V3 2020) for each protein is as follows, EUD-1: PPA43535, SUL2.2.1: PPA06135, SUL2.1: PPA21290. *A. sudhausi* gene models homologous to SUL-2 were also identified (Table S2). The phylogeny was generated using RAxML version 8.2.12 (raxmlHPC -f a -m PROTGAMMAAUTO -p 12345 -x 12345 -N 100), with maximum likelihood bootstrap values included (raxmlHPC -f b -m PROTGAMMAILG) (13). The phylogeny was displayed using FigTree software (v.1.4.4) (tree.bio.ed.ac.uk/software/figtree).

### CRISPR Engineering and Mutant Identification

CRISPR knockouts were generated as previously described (6). CRISPR RNAs (crRNAs) and trans-activating crRNA (tracrRNA) (Cat. No. 1072534) were synthesized by Integrated DNA Technologies (IDT), while the Cas9 endonuclease (Cat. No. 1081058) was purchased from IDT. The CRISPR/Cas9 complex was prepared by mixing 0.5 mg/ml Cas9 nuclease, 0.1 mg/ml tracrRNA, and 0.056 mg/ml crRNA in the TE buffer followed by a 10-min incubation at 37° C. Microinjections were performed in late-stage J4 hermaphrodites using an Eppendorf microinjection system. The sgRNA sequence was designed just before an NGG PAM site and targeted the putative exon two of all *A. sudhausi sul-2* homologs (Fig. 3C). Specific primers were designed for all four of the genes. The primers for *sul-2-A* are: 5'-AGAATGTAGCCAGGCAAGC-3' and 5'-CTCAGTCGACATGGAAAAGC-3' (445 bp amplicon). For *sul-2-B*: 5'-ATTGCAGGATACGGCGACC-3' and 5'-CTAATCTCGTCTATGCCGACG-3' (442 bp amplicon). For *sul-2-D*: 5'-ATGGTTCGACGATCTTGGTAGCG-3' and 5'-CCTGGAATGGAAGTCACACATGG-3' (745bp amplicon). For *sul-2-E*: 5'-GGTTTGATGACGCGTCATTGC-3' and 5'-CGACCATGCCCGTGATATAGC-3' (862bp amplicon). Polymerase chain reactions (PCRs) were run using these primers on the F1 of injected worms. Heterozygotes were identified via Sanger sequencing (GENEWIZ Germany GmbH) and homozygous mutants were obtained by self-fertilizing heterozygotes to generate homozygous knock-out mutants. One sgRNA was designed for each pair of similar genes. Frameshift mutants in *sul-2-A* (RS3764), *sul-2-B* (RS3797) and double frameshift mutants in *sul-2-A/B* (RS3798) were obtained after single injection of the *sul-2-A/B* sgRNA. The *sul-2-D/E* (RS3494) double mutant was obtained after a single injection as the gRNA knocked out both genes. In order to generate the quadruple mutant, the *sul-2-A/B* sgRNA was injected into the *sul-2-D/E* double

mutant line. The resulting injection resulted in triple mutants in *sul-2-A/B/D* and *sul-2-A/B/E*. These triple mutants were crossed to each other to eventually obtain the *sul-2-A/B/D/E* quadruple mutant knockout line (RS3598).

### **Statistical analysis**

The mean of technical replicates was determined for each biological replicate, which was subsequently analysed. For mouth width and worm size measurements the number of technical replicates varied, as not all worms were in the correct orientation to get accurate measurements. To test for significant differences between groups, the Shapiro-Wilk test was first performed to determine whether groups displayed normal distribution. If so, a student's t test was performed to compare the two groups. If the assumption of normality was violated, nonparametric tests were performed: Mann-Whitney U test for two comparisons and Kruskal-Wallis for group comparisons between groups, followed by the Pairwise Wilcox test. To analyse the association between independent and dependent variables, Pearson's correlation coefficient was used. All statistics and analysis was performed using RStudio Statistical Software (2021.09.2).

## **Supplementary Materials**

Figs. S1 to S7

Tables S1 to S3

References (24–32)

Movies S1 to S2

Data S1 to S5



**Supplementary Materials for**  
**Conserved switch genes regulate a novel cannibalistic morph after**  
**whole genome duplication**

Sara Wighard, Hanh Witte and Ralf J. Sommer

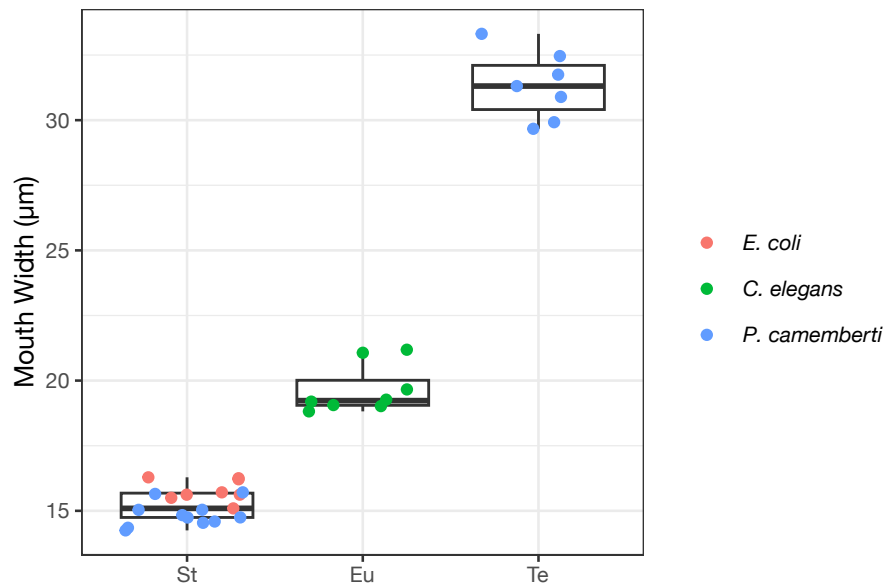
Corresponding author: [ralf.sommer@tuebingen.mpg.de](mailto:ralf.sommer@tuebingen.mpg.de)

**The PDF file includes:**

Figs. S1 to S7  
Tables S1 to S3

**Other Supplementary Materials include the following:**

Movie S1 to S2  
Data S1 to S5

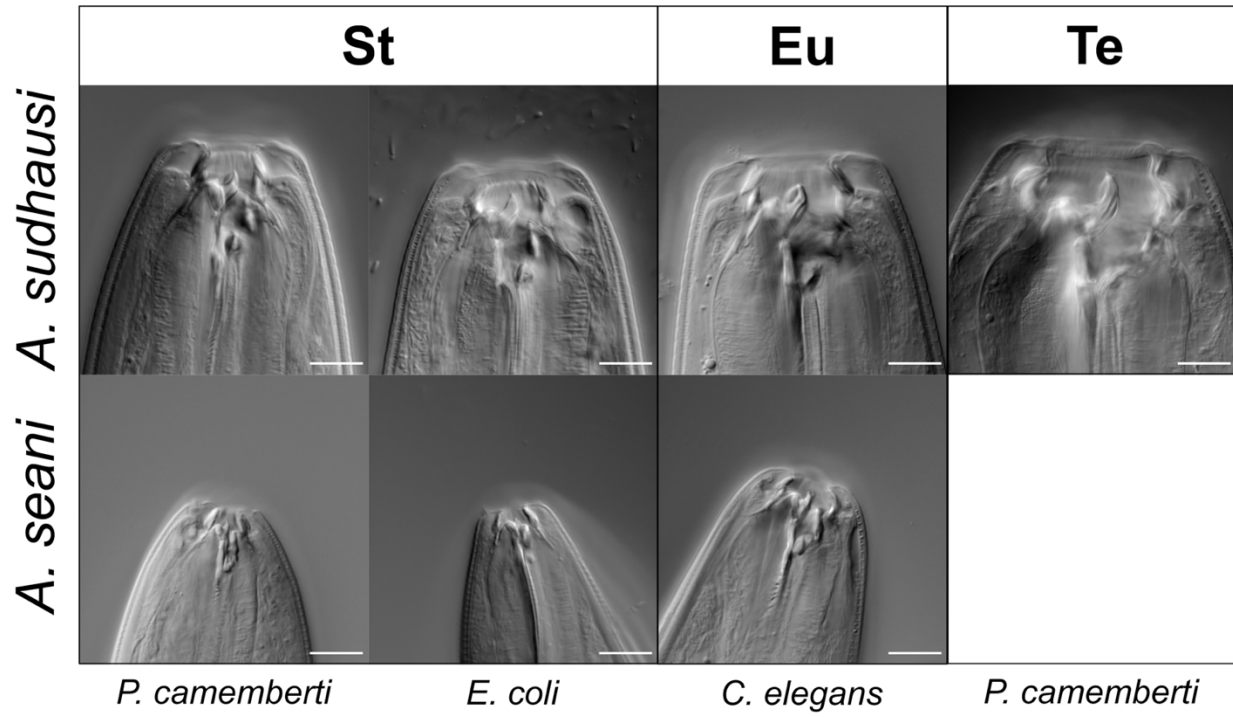


**Fig. S1.** The mouth width measurements of all three different assays is shown, with *E.coli* and *P.camemberti* constituting the St morph, *C.elegans* diets constituting the Eu and *P. camemberti* constituting the Te morph.

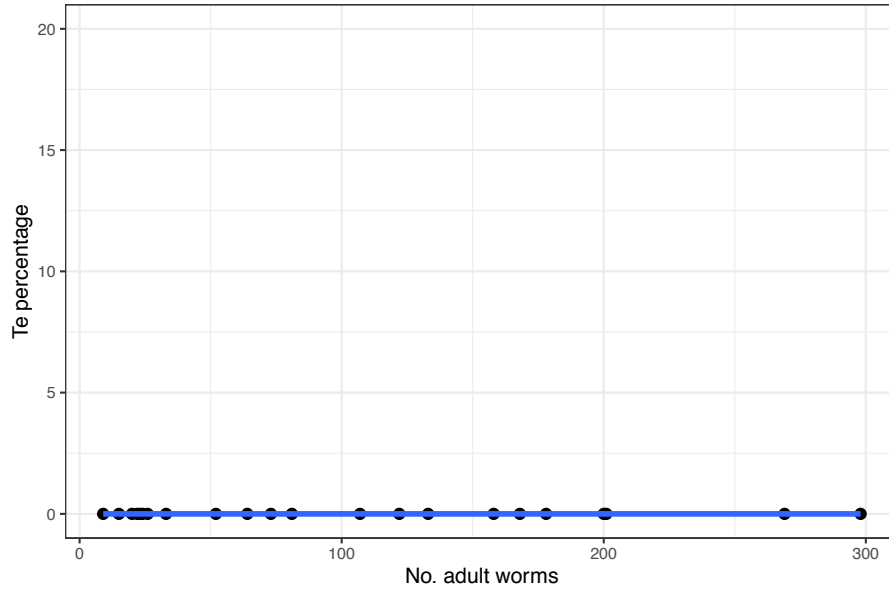


**Fig. S2.**

The Te morph can also be seen on old plates that have been starved of *E.coli*. Scale bar: 10  $\mu\text{m}$ .

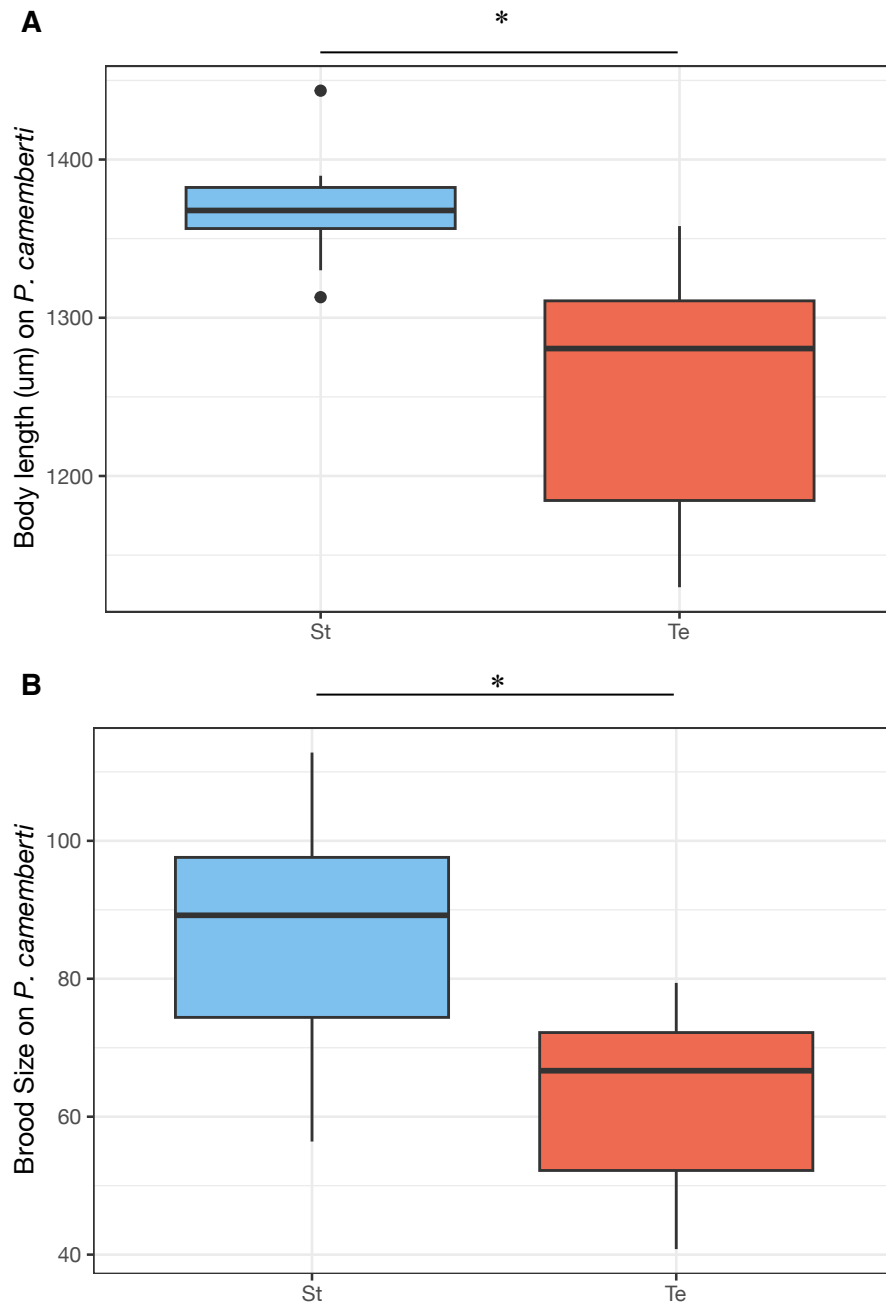


**Fig. S3.** *A. seani* can form the St (from *E.coli* and *P.camemberti*) and Eu (from *C.elegans*) morphs. No Te morph is seen from *P. camemberti*. Scale bar: 10 $\mu$ m.

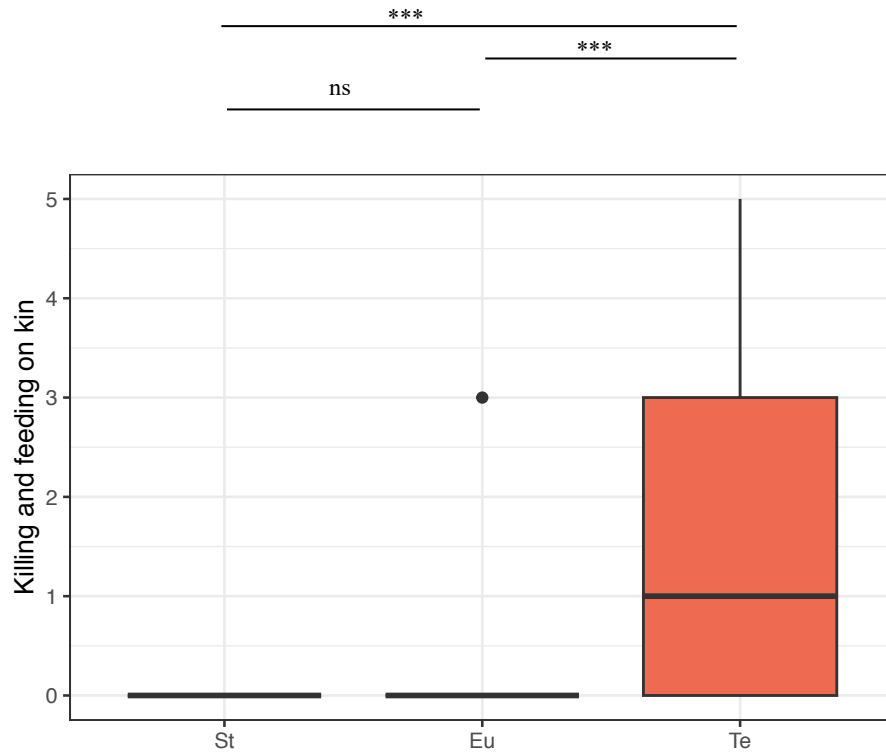


**Fig. S4.**

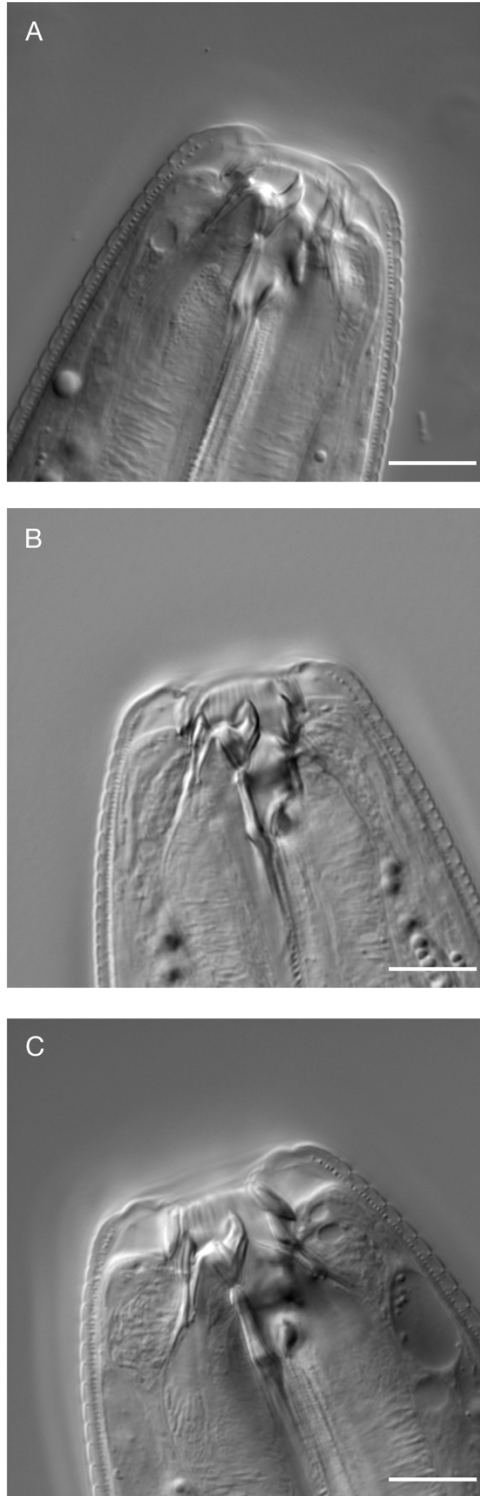
There is no correlation between the number of *A.sudhausi* adults and Te percentage. On *E.coli* plates, no worms form the Te morph. A linear regression plot is shown.



**Fig. S5.** Te morphs from *P. camemberti* assay show fitness costs compared to St morphs. **(A)** Te worms are significantly shorter than St (t test,  $p < 0.05$ ).  $n = 10$  biological replicates each. **(B)** Te worms also display decreased brood size (t test,  $p < 0.05$ ).  $n = 7$  biological replicates each.



**Fig. S6.** Feeding of *A. sudhausi* on its younger kin was compared for all three mouth-form morphs. This includes St grown on *E. coli*, *P. camemberti* and old starved plates; Eu grown on *C. elegans*, and Te from *P. camemberti* and old starved plates. There were significant differences between Te and the other morphs (Kruskal-Wallis test,  $p < 0.001$ ; Pairwise Wilcox,  $p < 0.001$ ) but not between St and Eu ( $p > 0.05$ ). The full dataset is shown in Data S4.



**Fig. S7.** DIC images of adult mouth-form in *sul-2-A/B* double mutants show they remain St when grown on (A) *E.coli*, (B) *C. elegans* and (C) *P.camemberti*. Scale bar: 10  $\mu$ m.



**Table S1.** Table showing the different fungal strains *A. sudhausi* was initially tested on. All strains were able to produce the Te morph at varying percentages.

| <b>Fungus</b>                 | <b>Replicate</b> | <b>Te Percentage (%)</b> | <b>No. Hermaphrodites</b> |
|-------------------------------|------------------|--------------------------|---------------------------|
| <i>Penicillium rubens</i>     | 1                | 4,69                     | 64                        |
| <i>Emericella nidulans</i>    | 1                | 2,27                     | 88                        |
| <i>Botrytis cinerea</i>       | 1                | 52,38                    | 21                        |
| <i>Aspergillus spp.</i>       | 1                | 54,55                    | 22                        |
| <i>Aspergillus spp.</i>       | 2                | 57,14                    | 28                        |
| <i>Penicillium digitatum</i>  | 1                | 7,14                     | 42                        |
| <i>Penicillium digitatum</i>  | 2                | 5,26                     | 38                        |
| <i>Penicillium digitatum</i>  | 3                | 17,39                    | 23                        |
| <i>Penicillium digitatum</i>  | 4                | 12,90                    | 31                        |
| <i>Penicillium camemberti</i> | 1                | 20,31                    | 64                        |
| <i>Penicillium camemberti</i> | 2                | 8,14                     | 86                        |
| <i>Penicillium camemberti</i> | 3                | 4,08                     | 49                        |
| <i>Penicillium camemberti</i> | 4                | 9,52                     | 84                        |
| <i>Penicillium camemberti</i> | 5                | 10,68                    | 103                       |
| <i>Penicillium camemberti</i> | 6                | 8,51                     | 94                        |

**Table S2.** Table showing the gene names and their corresponding gene models. The matching gene and contig names in previous literature is also shown. Note that, based on our analysis, the previously annotated *sul-2-C* and *sul-2-D* are in fact one gene, which we refer to as *Asu-sul-2-D*.

| <b>Gene Name</b>   | <b>Predicted Gene Model (Wighard et al. 2022)</b> | <b>Biddle &amp; Ragsdale (2020)</b> | <b>Sieriebriennikov et al. (2018)</b> |
|--------------------|---------------------------------------------------|-------------------------------------|---------------------------------------|
| <i>Asu-sul-2-A</i> | ALDISUDHAUS000034192                              | <i>sul-2-A</i>                      | Contig 2291                           |
| <i>Asu-sul-2-B</i> | ALDISUDHAUS000032677                              | <i>sul-2-B</i>                      | Contig 5979                           |
| <i>Asu-sul-2-D</i> | ALDISUDHAUS000002417                              | <i>sul-2-C &amp; sul-2-D</i>        | Contig 474                            |
| <i>Asu-sul-2-E</i> | ALDISUDHAUS000002604                              | <i>sul-2-E</i>                      | Contig 8694                           |

**Table S3.** Table showing the mouth-form of *sul-2A/B* and *sul-2-A/B/DE* adult mutants grown on *P. camemberti*. No Te morph was ever formed.

| Gene knock-out       | Strain | Replicate | No. St | No. Te |
|----------------------|--------|-----------|--------|--------|
| <i>sul-2-A/B</i>     | RS3798 | 1         | 56     | 0      |
| <i>sul-2-A/B</i>     | RS3798 | 2         | 65     | 0      |
| <i>sul-2-A/B</i>     | RS3798 | 3         | 6      | 0      |
| <i>sul-2-A/B</i>     | RS3798 | 4         | 101    | 0      |
| <i>sul-2-A/B</i>     | RS3798 | 5         | 133    | 0      |
| <i>sul-2-A/B</i>     | RS3798 | 6         | 6      | 0      |
| <i>sul-2-A/B</i>     | RS3798 | 7         | 73     | 0      |
| <i>sul-2-A/B</i>     | RS3798 | 8         | 6      | 0      |
| <i>sul-2-A/B</i>     | RS3798 | 9         | 165    | 0      |
| <i>sul-2-A/B</i>     | RS3798 | 10        | 82     | 0      |
| <i>sul-2-A/B</i>     | RS3798 | 11        | 27     | 0      |
| <i>sul-2-A/B/D/E</i> | RS3598 | 1         | 35     | 0      |
| <i>sul-2-A/B/D/E</i> | RS3598 | 2         | 8      | 0      |
| <i>sul-2-A/B/D/E</i> | RS3598 | 3         | 53     | 0      |
| <i>sul-2-A/B/D/E</i> | RS3598 | 3         | 75     | 0      |
| <i>sul-2-A/B/D/E</i> | RS3598 | 4         | 36     | 0      |
| <i>sul-2-A/B/D/E</i> | RS3598 | 5         | 89     | 0      |
| <i>sul-2-A/B/D/E</i> | RS3598 | 6         | 157    | 0      |
| <i>sul-2-A/B/D/E</i> | RS3598 | 7         | 186    | 0      |

**Movie S1. Te morph cannibalizes on younger kin.**

An adult Te hermaphrodite kills and feeds upon juvenile kin. Video set at 2x speed.

**Movie S2. Adult Te cannibalizes on another Te morph.**

An adult Te hermaphrodite latches to, kills and feeds upon another adult Te hermaphrodite. Video set at 2x speed.

**Data S1. (mouth\_width\_measurements.xlsx)**

Excel file containing the mouth width measurements of the wild type and mutant lines that were grown on *E. coli*, *C. elegans* and *P. camemberti*.

**Data S2. (worm\_sizing.xlsx)**

Excel table showing the body measurements of wild type worms grown on *P. camemberti* using the WormSizer plugin. Measurements of young adult St and Te hermaphrodites are shown. The file displays volume, length, middle width, mean width and surface area (all in  $\mu\text{m}$ ) of the worm body. The length measurements were compared in fig. S5A.

**Data S3. (brood\_size\_pcmemberti.xlsx)**

Excel table showing the number of viable progeny produced by wild type St and Te morphs that had grown up on *P. camemberti*.

**Data S4. (cannibalism.xlsx)**

Excel table displaying the number of juvenile kin fed on by different adult *A. sudhausi* morphs.

**Data S5. (GM\_landmarks.txt)**

Text file containing the 18 landmark co-ordinates for the geometric morphometrics comparison between wild type morphs.

**Title: Multiple evidence for environmental sex determination in the hermaphroditic nematode *Allodiplogaster sudhausi***

**Authors:** Sara Wighard<sup>1</sup>, Christian Rödelsperger<sup>1</sup> and Ralf J. Sommer<sup>1</sup>

**Affiliation:** Max Planck Institute for Biology, Tübingen, 72076, Germany

**Abstract**

Sex determination, though widespread, is poorly understood in the majority of organisms. Genetic sex determination is one of the most common mechanisms and was for a long time thought to be predominant in nematodes; however, recent work has suggested other systems, such as environmental sex determination, may be more abundant than initially thought. *Pristionchus* nematodes from the Diplogastridae family have recently been shown to display rapid evolution of sex determination systems within their genus. Here, we show an early branching nematode from the same family, *Allodiplogaster sudhausi*, displays similar genome coverage between males and hermaphrodites, shared chromosome number and no sex-linked SNPs, indicating it is not a genetic sex determination system. In contrast to this, growth at different temperatures affects the number of male progeny produced, with induction of males at lower temperature and the complete loss of males at the highest. Additionally, crowding also strongly induces males across multiple conditions. Overall, these findings provide further evidence of rapid evolution of sex determination systems within nematodes.

## Introduction

Sex determination is one of the fundamental processes of life, as reproduction is necessary for species to survive for subsequent generations. Despite this, sex determination is poorly understood in the vast majority of organisms. Here, we describe the evolution of environmental sex determination (ESD) in an early branching nematode species from a family that ancestrally displays genetic sex determination (GSD).

Sex determination systems have historically been broken up into GSD or ESD; while a third, stochastic sex determination (SSD) has recently been suggested (Perrin 2016). GSD is the most well-known mechanism; in this case sex chromosomes usually determine whether the progeny are male or female. Conversely, ESD occurs when environmental cues determine sex ratios, with temperature effects being the most well-studied (Bachtrog et al. 2014). This system is thus vulnerable to fluctuations in the environment. Additionally, SSD occurs when a random process determines sex (Shinya et al. 2022). To further complicate matters, organisms do not necessarily display each mechanism exclusively. In certain cases, it is a combination of some or all of the three sex determination systems that can result in sex differentiation (Bradley et al. 2011; Perrin 2016). However, our understanding of the mechanisms behind these systems and how they evolved is poor.

Nematodes have been proposed as ideal organisms to investigate the evolution of sex determination. As nematodes have a broad range of reproduction in large taxa groups, it allows researchers to easily examine and pinpoint evolutionary changes (Van Goor et al. 2021). For a long time, nematodes were thought to have ancestral GSD systems abundant in all taxa (Van Goor et al. 2021; Carlton et al. 2022). The model worm, *C. elegans* is androdioecious, producing either hermaphrodites or males, and displays a XX-XO GSD system where hermaphrodites have two copies of the X chromosome and males only have one (Nigon 1949; Madl and Herman 1979). This same mechanism is present in another nematode model, *Pristionchus pacificus* (Pires-daSilva 2007), which belongs to the Diplogastridae family and diverged from *C. elegans* roughly 200 millions years ago (Pires-daSilva and Sommer 2004). Therefore,

XX-XO GSD was thought to be ancestral in nematodes as it was detected in most well-studied species (Pires-daSilva 2007).

However, contrary to what was initially thought, diverse mechanisms of sex determination are rapidly evolving (Bachtrog et al. 2014). Recent work has uncovered previously unknown ESD and SSD systems in nematodes (Shinya et al. 2022; Yoshida et al. *In preparation*), suggesting mechanisms other than GSD may be more prevalent than initially thought. Comprehensive analyses of sex determination in the *Pristionchus* genus showed that species underwent extensive chromosomal rearrangements and, in certain cases, rapidly evolved new sex determination systems (Yoshida et al. 2023 *In preparation*). Notably, ESD evolved independently in two androdioecious *Pristionchus* species, *P. mayeri* and *P. entomophagous*. These unexpected findings led to the question of whether these traits are specific to the genus *Pristionchus*, or reflective of a wider trend.

We wanted to investigate if different sex determination systems were present in species outside of the *Pristionchus* genus. We therefore examined *Allodiplogaster sudhausi*, which is one of the earliest diverging lineages in the Diplogastridae (Susoy et al. 2015) (Fig. 1A). *A. sudhausi* is an androdioecious hermaphrodite that can outcross to males, making it ideal for genetic analysis (Fürst von Lieven 2008). *A. sudhausi* has already evolved certain novel traits, including increased body size (Fig. 1B) and an additional mouth-form morph (Fig 1C) (Wighard et al. 2023). Notably, different morphs can be formed on different diets of bacteria, worms and fungi, indicating it is adapted to different environments. Further, *A. sudhausi* recently underwent whole genome duplication that resulted in twelve chromosomes (Wighard et al. 2022), suggesting it had six pre-duplication. Intriguingly, WGD has often led to the formation of sex-determining genes from duplicates (Matsuda et al. 2002; Kruger et al. 2019; Hayashi et al. 2022). We therefore wanted to determine whether *A. sudhausi* displays the ancestral GSD system, or whether it evolved a new sex determination system.

In this study, we show *A. sudhausi* does not display ancestral GSD and instead evolved ESD, with male generation strongly influenced by temperature and crowding. Thus, independent evolution of ESD appears to be commonplace throughout the nematode Diplogastridae family.

## Results

### **Karyotype analysis and crossing experiments show there is no difference in chromosome number between sexes**

We first karyotyped *A. sudhausi* to determine the number of chromosomes. In XX-XO GSD, males lack an X chromosome, meaning their sperm would contain one less chromosome than what is seen in hermaphrodite oocytes. We counted twelve chromosomes in hermaphrodite oocytes (Fig. 2A), as previously shown (Wighard et al. 2022). Interestingly, we also counted twelve chromosomes in male spermatocytes (Fig. 2B; Fig S1). This number was also observed in sperm (Fig. S2). Thus, males and hermaphrodites appear to inherit the same number of chromosomes, strongly suggesting *A. sudhausi* does not have an XX-XO system.

Next, we compared the percentage of male progeny after selfing vs mate-crossing. In chromosomal GSD, such as XO, XY or ZW, you would expect a cross-progeny sex ratio of 1:1 as they would get one chromosomal copy each from the male and hermaphrodite parent. This 1:1 sex ratio after crossing has previously been observed in nematodes with chromosomal GSD systems (Nigon 1949; Pires-daSilva 2007). For selfing in *A. sudhausi*, the male progeny of virgin hermaphrodites was counted. A median value of 7% spontaneous males was obtained for selfing *A. sudhausi*, consistent with previous observations (Fürst von Lieven 2008) and far higher than the rates seen in *P. pacificus* (Sommer et al. 1996) and *C. elegans* (Brenner 1974).

For crossing, sperm-depleted old hermaphrodites were mated with males, as has been successfully performed in *P. pacificus* (Yoshida et al. 2023). Note that CRISPR knock-out mutants have been used to confirm that *A. sudhausi* hermaphrodite-male mating produces cross progeny (Wighard et al. 2023 *Submitted*). We found that



crossing hermaphrodites and males in *A. sudhausi* produced male ratios far below the expected 1:1 ratio. In fact, there was no significant difference compared to self-fertilised progeny (Fig 2C). Therefore, the crossing experiments again indicated that *A. sudhausi* does not have a chromosomal GSD system, such as XX-XO found in *P. pacificus*.

### **No genomic differences between males and hermaphrodites**

In GSD, the genome coverage between males and hermaphrodites is different. For example, if one sex receives a chromosome from one parent and not the other it would result in abnormal peak shapes such as shoulder peaks. This is because there would be less coverage in the sex-specific region of the genome compared to the rest of the genome. This has previously been shown in nematodes that have GSD (Wang et al. 2022). We therefore examined genome coverage in hermaphrodites (Fig. 2D) and males (Fig. 2E). In *A. sudhausi*, the genome coverage between sexes is almost identical, suggesting there are no sex-specific differences in the genomes.

Although sex chromosome differences are most common in GSD systems, there could potentially be other smaller genomic differences that affect sex ratios. Therefore, we identified potential sex-specific SNPs that may be biased to one sex. We tested for sex-specific differences in the allele-frequencies between hermaphrodites and males and found only 62 candidate positions that showed significantly different allele frequencies (Fig. 1 F). Furthermore, the magnitude of change in allele frequencies was small, with the candidate positions scattered across the whole genome assembly (Fig. 2G). In addition, candidate positions exhibit much higher coverage than non-candidate positions (Figure S3), suggesting that these candidate positions are more likely a result of assembly problems rather than genetic differences between sexes. Altogether, there are no signs of genomic differences between sexes despite thorough analysis. Therefore, our findings show that *A. sudhausi* does not have a GSD system.

### **Temperatures influences male ratios in environmental sex determination system**

Next, we looked for possible environmental cues that may affect sex ratio in *A. sudhausi*; as has been observed in *Pristionchus* species that independently evolved ESD (Yoshida et al. 2023 *In preparation*). Temperature is one of the most common features of ESD in nematodes (Triantaphyllou 1973; Harvey et al. 2000). We thus examined the sex ratio of *A. sudhausi* progeny reared under different temperatures. Strikingly, we found an obvious effect of temperature on progeny, with a high percentage of *A. sudhausi* males at the lowest temperature of 14°C and no male generation whatsoever at 30°C (Fig. 3A). Furthermore, there was a clear and gradual decrease in male progeny ratio from lower to higher temperatures. This strongly indicates the sex ratio in *A. sudhausi* is environmentally influenced, with temperature a major factor.

### **Crowding consistently promotes male production in different conditions**

We wanted to see if factors other than temperature also affect sex ratio. Increased density has been shown to induce males in nematodes with ESD (Triantaphyllou 1973; Harvey et al. 2000). We therefore examined the sex ratio with different numbers of worms on a plate. We found a strong correlation between male progeny and crowding on standard *E. coli* plates, indicating increased density promotes male production. To determine if the density-dependent effect was due to the growth conditions, *A. sudhausi* was also examined on worm (*C. elegans*) (Fig. 3C) and fungal (*P. camemberti*) (Fig.3D) diets (Wighard et al. 2023). In all cases, crowding strongly affected the male sex ratio and, importantly, the effect was in the same direction, with increased density elevating the number of male progeny.

In summary, we present strong evidence to show that 1) *A. sudhausi* does not have the ancestral GSD system, and 2) the sex ratio in *A. sudhausi* is instead heavily influenced by the environment.

## Discussion

GSD was for a long time thought to be ancestral in nematodes (Pires-daSilva 2007); however, an increasing body of research is demonstrating that other sex determination systems are more prevalent than initially thought. Indeed, the perceived commonality of GSD is likely due to bias in nematode species studied and small sample sizes. Our findings reinforce the importance of studying a broad range of taxa to get a more realistic idea of prevailing mechanisms, as opposed to making broad assumptions based on a few. With increased sequencing efforts worldwide, we are starting to get a clearer picture of the different sex determination systems present. Here, we show that an early-branching free-living nematode species displays ESD, which likely evolved from an ancestral GSD system (Yoshida et al. 2023 *In preparation*).

There are cases where organisms can have either GSD or ESD systems. That is, they display genomic differences, but environment can still choose sex (Sarre et al. 2004). However, we present strong evidence that *A. sudhausi* is not influenced by any genetic components as there is consistency in karyotype number and genome coverage between sexes, with no clear evidence for sex-specific SNPs. Furthermore, it is logical that *A. sudhausi* displays no difference in sex chromosomes as that would have been a strong barrier for WGD to successfully take place (Ohno 1970).

Our results conclusively show that environmental conditions affect the *A. sudhausi* sex ratio. Temperature in particular had a marked effect on male production. Many males were produced at lower temperatures, with a consistent decrease at subsequently higher temperatures, culminating in no males being produced at 30°C (Fig. 3A). Intriguingly, temperature has also been shown to have a small effect on male ratios in both *C. elegans* and *P. pacificus*, despite these nematodes having GSD systems. Specifically, male generation increases at higher percentages due to spontaneous meiotic non-disjunction. It has been theorised that male production in nematodes is elevated under temperatures that are not commonly expressed in their natural habitats, suggesting it is an adaptive response to stress (Morran et al. 2009). Specifically, in *P. pacificus* certain isolates from cold locations produce more males at higher temperatures compared to isolates from warmer locations that produce more

males at lower temperatures (Morgan et al. 2017). Strikingly, this *A. sudhausi* strain SB413B was isolated in Israel, which has a warm climate (average temperature of 21°C at Tel Aviv Sede Dov station, Israel Meteorological Service), and shows an increase of males at lower temperatures. Our results thus reflect previous studies that suggest male production is a stress response to new environmental conditions.

Crowding is another notable stressor for nematodes, with previous work having demonstrated that it increases male generation in nematodes (Triantaphyllou 1973; Harvey et al. 2000). In accordance with this, our findings also show increased density leads to greater male generation (Fig. 3B-D). Furthermore, the direction of male generation in response to crowding is consistent across three different diets, strongly suggesting the change in sex ratio is due to crowding and not an effect of different food sources. Therefore, these results are consistent with increased male production being a stress response. Overall, *A. sudhausi* clearly displays an ESD system.

ESD has been proposed to have evolved from GSD systems (Bull 1981) as an adaptation to environmental conditions that favour one sex over the other (Charnov and Bull 1977). This seems to also be the case for *A. sudhausi* as diplogastrid nematodes are predicted have ancestral GSD systems (Yoshida et al. 2023 *In preparation*). Therefore, ESD must have evolved in *A. sudhausi*. This means that *A. sudhausi* gained a new sex determination system; yet another novelty found in this nematode along with WGD (Wighard et al. 2022), increased body size and evolution of a novel mouth morph (Wighard et al. 2023).

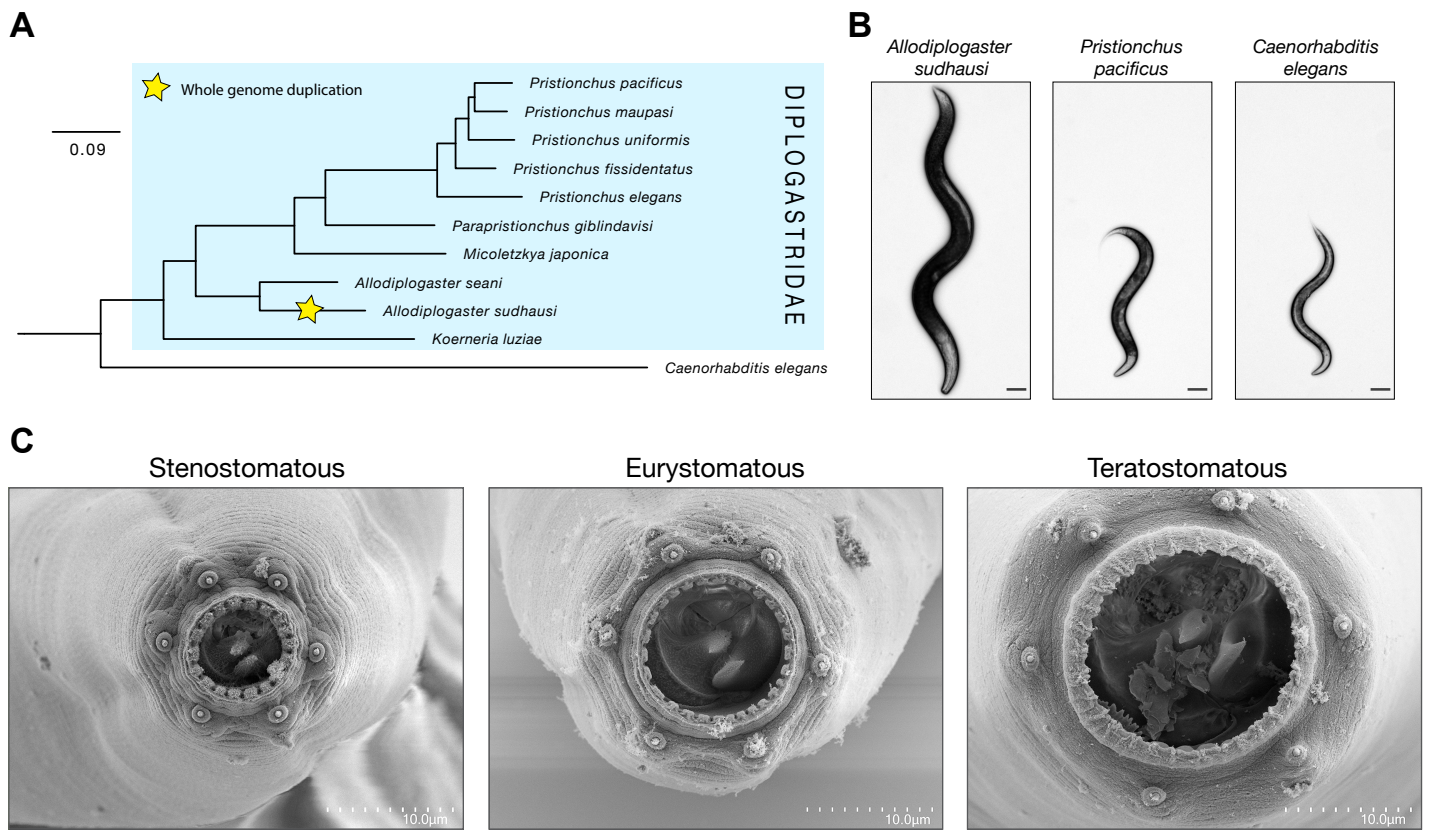
Overall, these findings reflect recent work examining sex determination in the *Pristionchus* genus (Yoshida et al. 2023 *In preparation*). Specifically, the authors found rapid evolution of sex determination systems, as well as the independent gain of ESD in some species. Our findings build on this work and further suggest that rapid evolution of sex determination systems is not necessarily *Pristionchus*-specific. In fact, sex determination systems appear to be rapidly evolving with the Diplogastridae, and potentially, within nematodes as a whole. It suggests, for those interested in evolution of sex determination mechanisms, nematodes may be an ideal group of organisms to

study due to the abundance of different sex determination systems that rapidly evolve. There are likely more unexpected sex determination systems that have not yet been revealed. Indeed, most plant-parasitic nematodes do not have GSD systems either, with ESD seemingly predominant (Triantaphyllou 1973; Charnov and Bull 1977; Castagnone-Sereno and Danchin 2014; Anjam et al. 2020) and SSD also proposed (Shinya et al. 2022). In the ESD systems of plant-parasites, male generation is also greater in adverse conditions (Triantaphyllou 1973; Anjam et al. 2020). Altogether, this suggests that ESD is more common than presumed and that increased male generation is often a stress response in ESD nematodes. It has been suggested that increased male generation helps prevent overpopulation in unfavourable conditions (Triantaphyllou 1973).

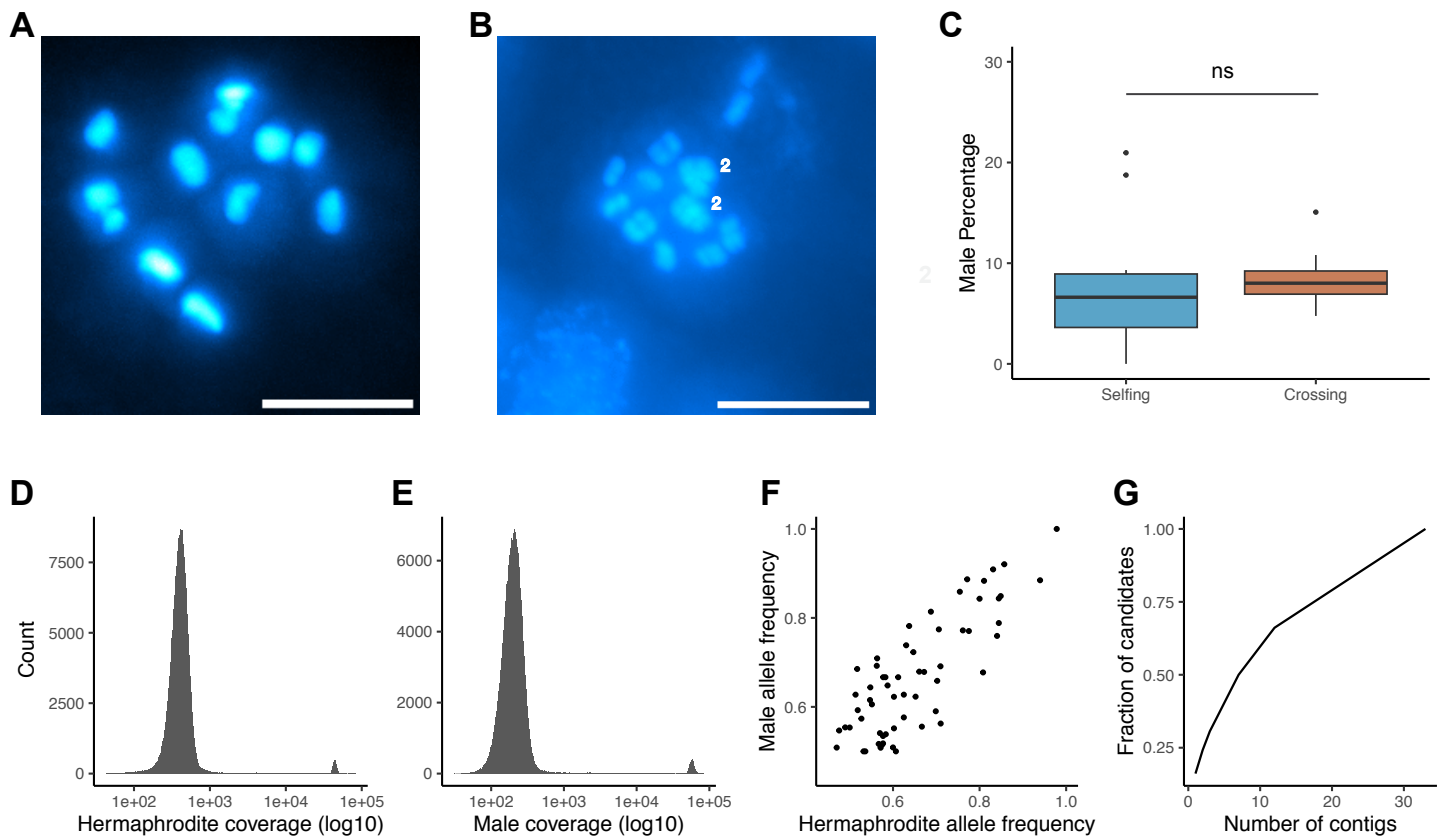
Lastly, it would be remiss not to mention the broader implications of different sex systems, particularly ESD, with variable climate conditions. It is particularly important to study mechanisms affected by the environment as organisms impacted by this are highly vulnerable to environmental changes, with negative impacts due to climate change having already been shown (Mitchell and Janzen 2010). Most studies in ESD have been performed in fish and reptiles (Bachtrog et al. 2014); however, ESD may be commonplace in many nematodes, particularly plant-parasitic ones as mentioned. As nematodes play key roles in biodiversity, it is important to better understand their sex determination systems and evaluate their vulnerability to environmental change.

### **Acknowledgements**

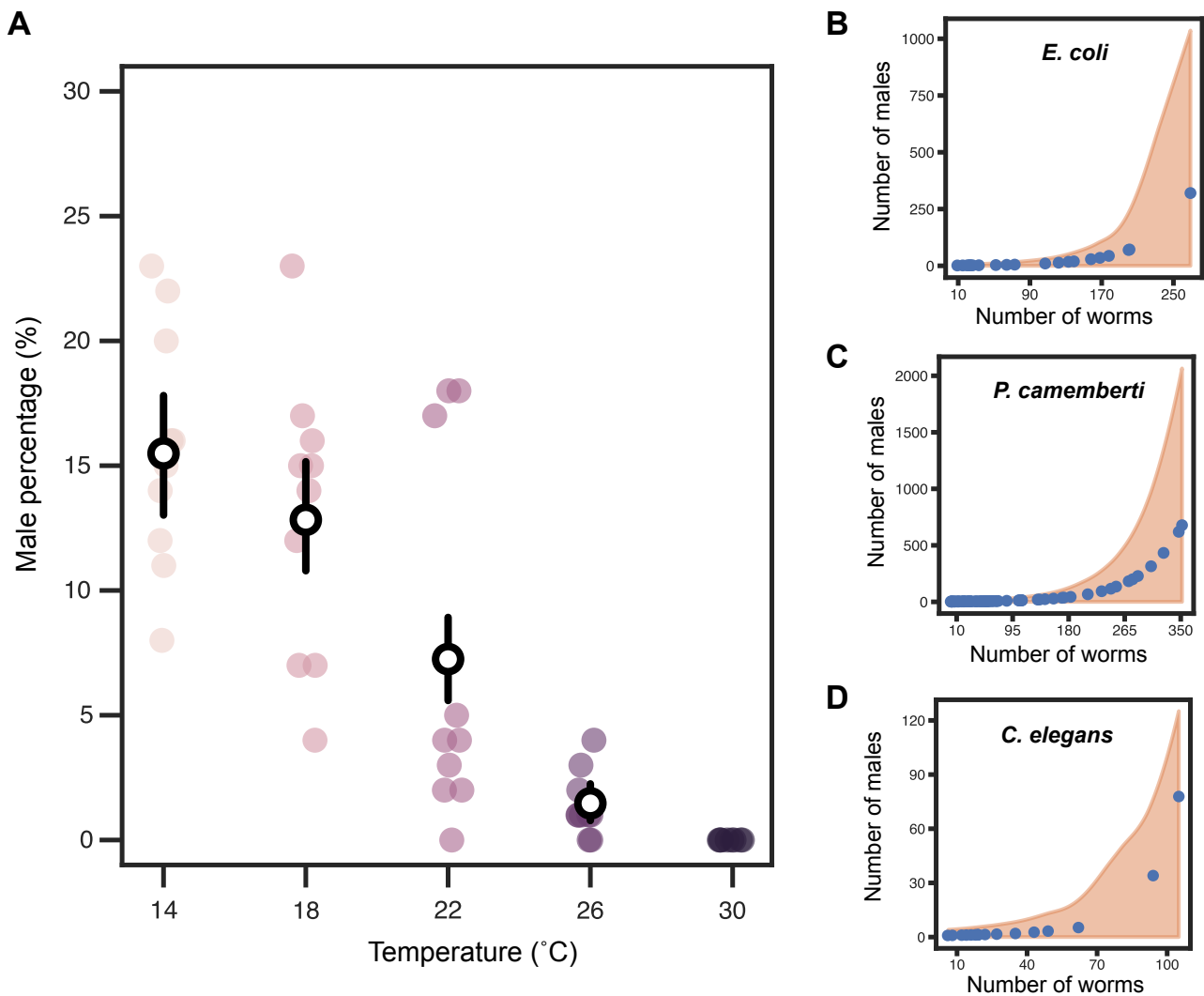
We'd like to thank Ata Kalirad for performing the statistical analyses.



**Figure 1: *Allodiplogaster sudhausi* recently underwent whole genome duplication and displays morphological novelty. A)** *A. sudhausi* is an early-branching species within the Diplogastridae nematode family that underwent whole genome duplication after diverging from its sister species *Allodiplogaster seani*. The phylogeny represents a subtree from previous data (Susoy et al. 2015). **B)** *A. sudhausi* is much larger than *P. pacificus* and *C. elegans*, as these pictures of young adult hermaphrodites shows. Scale bar: 100  $\mu$ M (Wighard et al. 2022). **C)** *A. sudhausi* has three potential adult mouth-form morphs, including a novel third morph (Teratostomatous) that is not seen in other diplogastrids (Wighard et al. 2023).



**Figure 2: *A. sudhausi* displays no genetic sex differences. A,B)** Twelve chromosomes are found in *A. sudhausi* **A)** hermaphrodite oocytes and **B)** male spermatocytes. A maximal intensity measurement was used to detect chromosomes in different planes. The ‘2’ symbol indicates where two chromosomes are located on the same Z plane. Scale bars: 10  $\mu$ M. **C)** Crossing sperm-depleted hermaphrodites with males does not give the 1:1 sex ratio expected in XX-XO GSD systems. There is no significant difference in male progeny percentage between selfing hermaphrodites (median = 7%) and crossed hermaphrodite-males (median = 8%) (t test, ns). **D)** Normalised coverage of the genome in hermaphrodites show one large peak with high coverage and a larger smaller peak likely showing mitochondrial DNA. **E)** Normalised coverage of the male genome also shows similar clean peaks, indicating there are no genomic sex-specific differences. **F)** Candidate positions with significantly different allele frequencies between sexes were identified (Bonferroni corrected  $p < 0.05$ , Binomial test), however the change in allele frequencies between sexes is minor. **G)** The candidate positions were plotted against the assembled genome and show no signs of positional bias.



**Figure 3: *A. sudhausi* has environmental sex determination. A)** Temperature affects the sex ratio, with more males produced at lower temperatures and less at higher temperatures, with no males formed at 30°C. Bars indicate the 95% Highest Density Interval (HDI), estimated using a Poisson model. The circles within the bar indicate the mean value, and the coloured dots represent experimental data values. **B)** Crowding increases the incidence of male progeny on *E. coli*. **C,D)** Crowding also increased male production on other diets, including **C)** *C. elegans* worms and **D)** *P. camemberti* fungus. A Bayesian binomial regression model was used to generate the plots, with the mean predicted value shown as circles and the shaded gradient representing the 95% HDI based on experimental data.



## **Materials and methods**

### **Nematode husbandry**

Strain SB413B was used for all analyses. The worms were maintained on nematode growth medium (NGM) agar plates and spotted with *Escherichia coli* OP50.

### **Phylogeny**

Data from Susoy et al. (2015) was used to generate a species subtree, which was edited using figTree software v.1.4.4 ([tree.bio.ed.ac.uk/software/figtree](http://tree.bio.ed.ac.uk/software/figtree)).

### **Karyotyping**

Chromosome staining was performed, as previously described (Wighard et al. 2022), on both *A. sudhausi* hermaphrodites and males. Briefly, we performed a 1:100 dilution of 20 mM Hoechst 33342 dye in sperm salts (50 mM PIPES, 25 mM KCl, 1 mM MgSO<sub>4</sub>, 45 mM NaCl, 2 mM CaCl<sub>2</sub>, pH 7). 10µl of this solution was then pipetted onto a microscope slide, and adults were moved onto the solution. A surgical blade was used to cut right below the pharynx, which resulted in the gonadal arms being pushed out of the worm body, as well as the release of sperm. A cover slip was placed on top and the worms were imaged using the Zeiss Axio Imager Z1 microscope.

### **Male progeny counts**

For selfing experiments, 10 virgin hermaphrodites were moved onto fresh NGM plates containing 30µl of *E. coli*. This assay was performed for six different temperatures: 14°C, 18°C, 20°C, 22°C, 26°C and 30°C. The hermaphrodite mothers were then transferred to new plates every few days until they stopped laying viable eggs. Once the progeny became adults, the number of total males was counted for each hermaphrodite. The assays were stored in incubators at the set temperatures.

### **Mate-crossing**

Individual sperm-depleted hermaphrodites were crossed with multiple males, as previously described (Yoshida et al. 2023; Wighard et al. 2023). To clarify, *A. sudhausi* hermaphrodites at the late J4 stage were moved to separate plates containing *E. coli*. They were left for nine days, at which point they would have run out of sperm. These

sperm-depleted hermaphrodites were then moved to separate plates containing 5 $\mu$ l of *E.coli* and 10 wild type males. These worms were kept together for 48 hours to mate and the males were then killed. The percentage of male progeny was then calculated. All assays were kept at 20°C.

### **DNA extraction and sequencing**

Approximately 50 adult hermaphrodites and males each were picked from standard *E. coli* maintenance plates and moved into separate 1.5ml tubes Eppendorf containing M9 buffer. The Epicentre MasterPure Complete DNA & RNA Purification Kit was used to extract DNA, following the manufacturer's protocol for tissue samples. The resulting DNA was quality controlled using the NanoDrop One Spectrophotometer and sent to Novogene Co. for Illumina sequencing (paired end 150bp, 10G per sample).

### **Coverage analysis**

The genome assembly (Wighard et al. 2022) was first indexed (bwa index). The hermaphrodite and male reads were then separately aligned to the indexed genome (bwa mem) and sorted (samtools sort). The read coverage at each position was then generated (samtools depth) for both sexes. Following this, the data was normalised for both sexes (individual read depth divided by overall mean depth). The sum per 1kb was then calculated for both males and hermaphrodites, and coverage count was plotted to generate peaks. For the detection of potential sex-specific SNPs, the allele frequencies were compared between hermaphrodites and males.

### **Crowding on male generation**

*E. coli* plates containing many eggs were bleached (Hope 1999) onto separate 6cm NGM plates with three different diets: 1) *E.coli* OP50, 2) *C. elegans* larvae and 3) *P. camemberti*, as previously described (Wighard et al. 2023). Only eggs survive after bleaching, allowing them to hatch and grow under the various conditions. Once they became adults, the number of hermaphrodites and males were counted.

## Statistical analysis

To test for mean differences between the male progeny of crossed and selfed hermaphrodites, the Shapiro-Wilk test was first performed to determine normality. Then, the student's t test was performed to compare the two groups. This analysis was performed using RStudio Statistical Software (2021.09.2). To estimate the rate of male production for temperature treatments, a Poisson model was fitted to data:

$$\lambda \sim \text{Exponential}(0.2)$$

$$y_i \sim \text{Poisson}(\lambda),$$

where  $y_i$  is the number of males produced in treatment  $i$ .

To illustrate the relationship between crowding ( $x$ ) on the number of males observed ( $y$ ), the following negative binomial regression model was fitted to the crowding data:

$$\beta_0 \sim \mathcal{N}(0, 2.5)$$

$$\beta_1 \sim \mathcal{N}(0, 2.5)$$

$$\alpha \sim \text{HalfNormal}(2.5)$$

$$\eta = \beta_0 + \beta_1 x$$

$$y \sim \text{NB}(e^\eta, \alpha).$$

The models were fitted to experimental data using PyMC(version 5.6.0) (Abril-Pla et al. 2023).

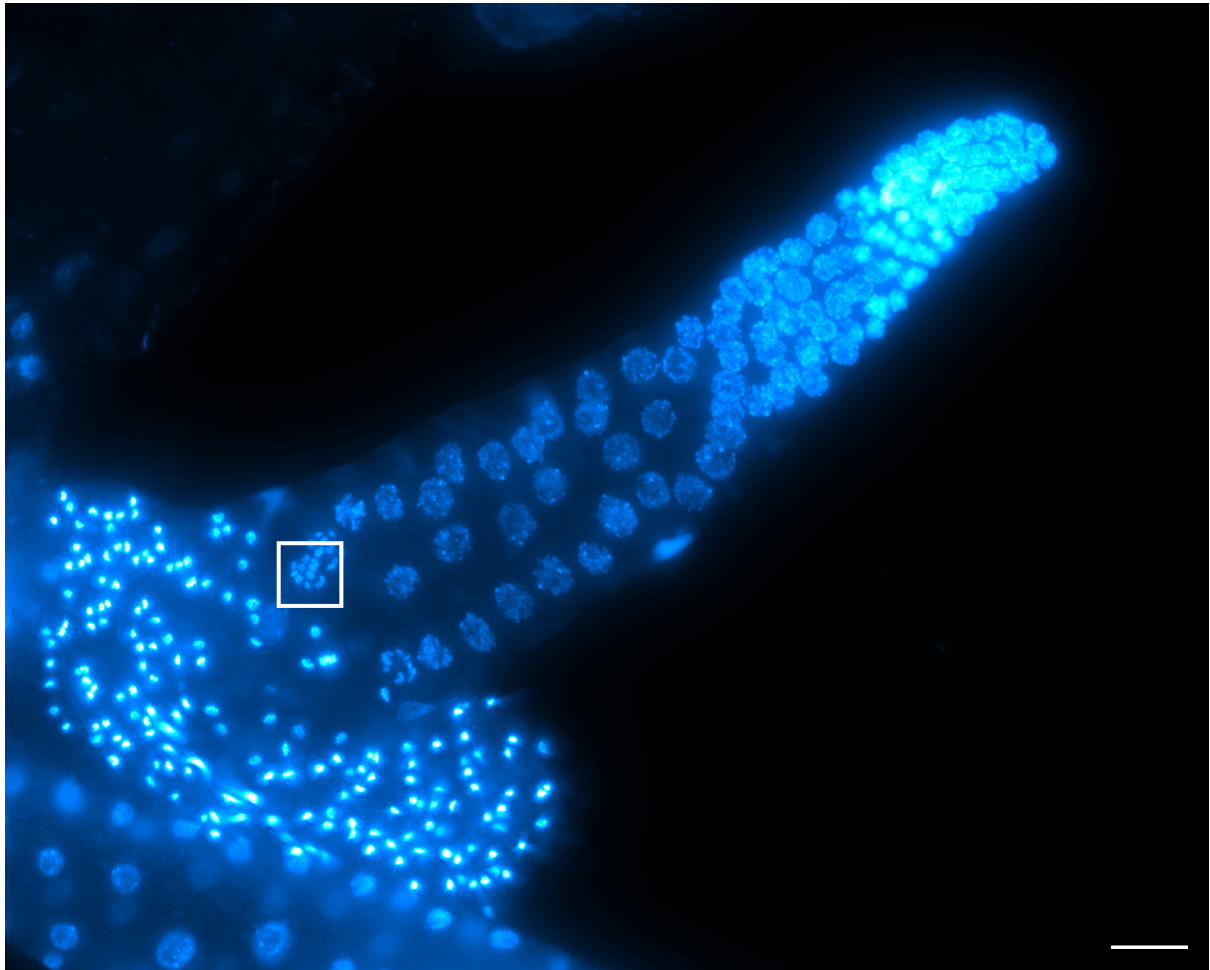
## References

- Abril-Pla O, Andreani V, Carroll C, Dong L, Fannesbeck C, Kochurov M, Kumar R, Lao J, Luhmann C, Martin O, et al. 2023. A Modern and Comprehensive Probabilistic Programming Framework in Python. *PeerJ* 9:1516.
- Anjam MS, Shah SJ, Matera C, Róžańska E, Sobczak M, Siddique S, Grundler FMW. 2020. Host factors influence the sex of nematodes parasitizing roots of *Arabidopsis thaliana*. *Plant, Cell & Environment* 43:1160–1174.
- Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I, Ming R, et al. 2014. Sex Determination: Why So Many Ways of Doing It? *PLOS Biology* 12:e1001899.
- Bradley KM, Breyer JP, Melville DB, Broman KW, Knapik EW, Smith JR. 2011. An SNP-Based Linkage Map for Zebrafish Reveals Sex Determination Loci. *G3 Genes/Genomes/Genetics* 1:3–9.
- Brenner S. 1974. The genetics of *Caenorhabditis elegans*. *genetics* 77:71–94.
- Bull JJ. 1981. Evolution of environmental sex determination from genotypic sex determination. *Heredity* 47:173–184.
- Carlton PM, Davis RE, Ahmed S. 2022. Nematode chromosomes. Strome S, editor. *Genetics* [Internet] 221. Available from: <https://academic.oup.com/genetics/article/doi/10.1093/genetics/iyac014/6551978>
- Castagnone-Sereno P, Danchin EGJ. 2014. Parasitic success without sex – the nematode experience. *Journal of Evolutionary Biology*:1323–1333.
- Charnov EL, Bull J. 1977. When is sex environmentally determined? *Nature* 266:828–830.
- Fürst von Lieven A. 2008. *Koerneria sudhausi* n. sp. (Nematoda: Diplogastridae); a hermaphroditic diplogastrid with an egg shell formed by zygote and uterine components. *Nematology* 10:27–45.
- Harvey SC, Gemmill AW, Read AF, Viney ME. 2000. The control of morph development in the parasitic nematode *Strongyloides ratti*. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267:2057–2063.
- Hayashi S, Suda K, Fujimura F, Fujikawa M, Tamura K, Tsukamoto D, Evans BJ, Takamatsu N, Ito M. 2022. Neofunctionalization of a Noncoding Portion of a DNA Transposon in the Coding Region of the Chimerical Sex-Determining Gene *dm-W* in *Xenopus* Frogs. Takahashi A, editor. *Molecular Biology and Evolution* 39:1–6.
- Hope IA. 1999. *C. elegans: A Practical Approach*. OUP Oxford

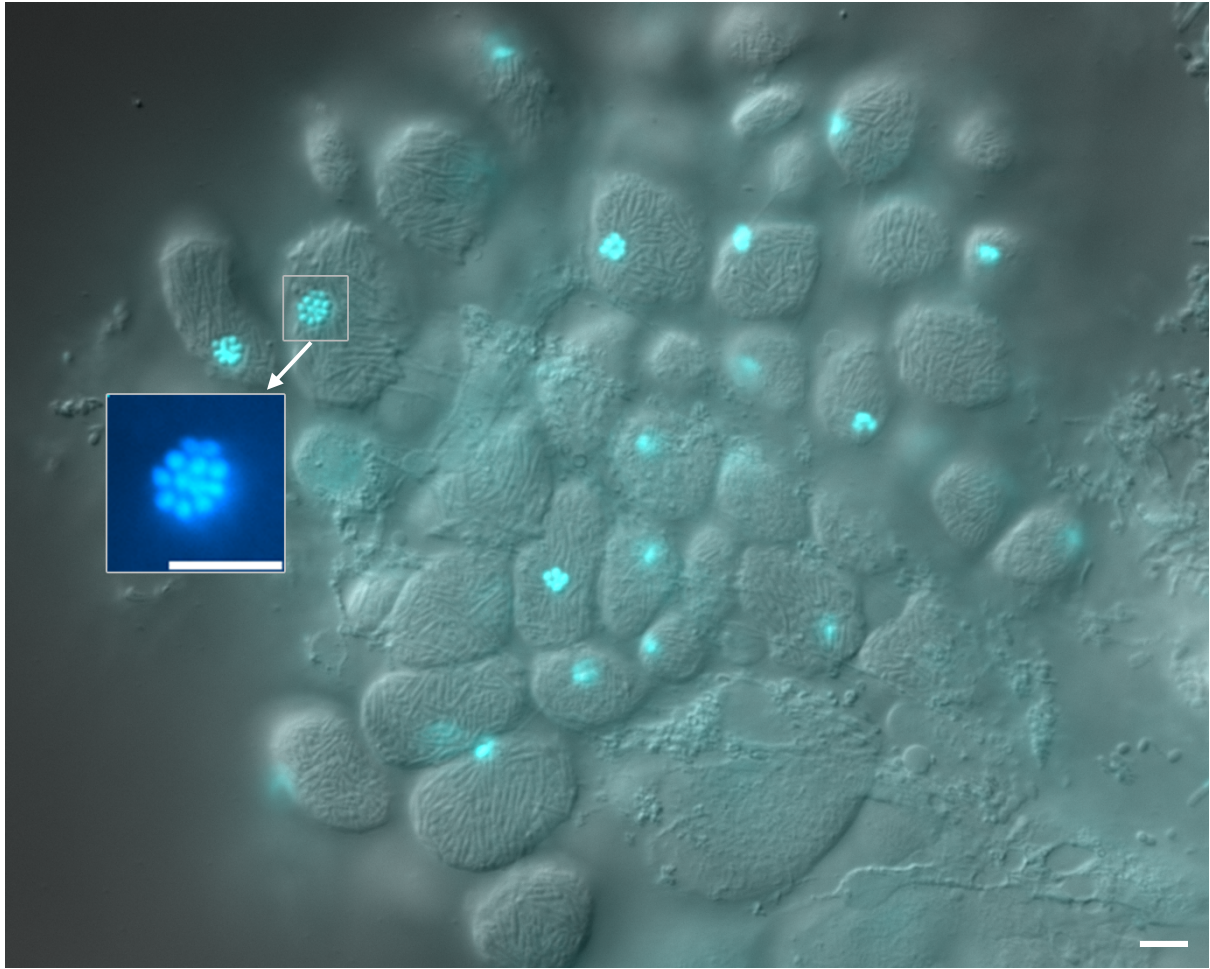
- Kruger AN, Brogley MA, Huizinga JL, Kidd JM, De Rooij DG, Hu Y-C, Mueller JL. 2019. A Neofunctionalized X-Linked Ampliconic Gene Family Is Essential for Male Fertility and Equal Sex Ratio in Mice. *Current Biology* 29:3699-3706.e5.
- Madl JE, Herman RK. 1979. POLYPLOIDS AND SEX DETERMINATION IN CAENORHABDITIS ELEGANS. *Genetics* 93:393–402.
- Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, Morrey CE, Shibata N, Asakawa S, Shimizu N, et al. 2002. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417:559–563.
- Mitchell NJ, Janzen FJ. 2010. Temperature-Dependent Sex Determination and Contemporary Climate Change. *Sexual Development* 4:129–140.
- Morgan K, McGaughran A, Rödelsperger C, Sommer RJ. 2017. Variation in rates of spontaneous male production within the nematode species *Pristionchus pacificus* supports an adaptive role for males and outcrossing. *BMC Evolutionary Biology* 17:57.
- Morran LT, Cappy BJ, Anderson JL, Phillips PC. 2009. SEXUAL PARTNERS FOR THE STRESSED: FACULTATIVE OUTCROSSING IN THE SELF-FERTILIZING NEMATODE CAENORHABDITIS ELEGANS. *Evolution* 63:1473–1482.
- Nigon, V. (1949). Les modalités de la reproduction et le déterminisme due sexe chez quelques nematodes libres. *Ann. Sci. Nat.*, 11, 1-132.
- Ohno S. 1970. Evolution by Gene Duplication. Springer Berlin, Heidelberg
- Perrin N. 2016. Random sex determination: When developmental noise tips the sex balance. *BioEssays* 38:1218–1226.
- Pires-daSilva A. 2007. Evolution of the control of sexual identity in nematodes. *Seminars in Cell & Developmental Biology* 18:362–370.
- Pires-daSilva A, Sommer RJ. 2004. Conservation of the global sex determination gene *tra-1* in distantly related nematodes. *Genes & Development* 18:1198–1208.
- Sarre SD, Georges A, Quinn A. 2004. The ends of a continuum: genetic and temperature-dependent sex determination in reptiles. *BioEssays* 26:639–645.
- Shinya R, Sun S, Dayi M, Tsai IJ, Miyama A, Chen AF, Hasegawa K, Antoshechkin I, Kikuchi T, Sternberg PW. 2022. Possible stochastic sex determination in *Bursaphelenchus* nematodes. *Nat Commun* 13:2574.
- Sommer RJ, Carta LK, Kim SY, Sternberg PW. 1996. Morphological, genetic and molecular description of *pristionchus pacificus* sp. n. (nematoda : neodiplogastridae). *Fundamental and Applied Nematology* 19:511–521.

- Susoy V, Ragsdale EJ, Kanzaki N, Sommer RJ. 2015. Rapid diversification associated with a macroevolutionary pulse of developmental plasticity. *eLife* 4:1–39.
- Triantaphyllou AC. 1973. Environmental Sex Differentiation of Nematodes in Relation to Pest Management. *Annual Review of Phytopathology* 11:441–462.
- Van Goor J, Shakes DC, Haag ES. 2021. Fisher vs. the Worms: Extraordinary Sex Ratios in Nematodes and the Mechanisms that Produce Them. *Cells* 10:1793.
- Wang Y, Gasser RB, Charlesworth D, Zhou Q. 2022. Evolution of sexual systems, sex chromosomes and sex-linked gene transcription in flatworms and roundworms. *Nat Commun* 13:3239.
- Wighard S, Athanasouli M, Witte H, Rödelsperger C, Sommer RJ. 2022. A New Hope: A Hermaphroditic Nematode Enables Analysis of a Recent Whole Genome Duplication Event. *Genome Biology and Evolution* 14:evac169.
- Wighard S, Witte H, Sommer RJ. 2023. Conserved switch genes regulate a novel cannibalistic morph after whole genome duplication. *bioRxiv:2023.08.22.554244*.
- Yoshida K, Rödelsperger C, Röseler W, Riebesell M, Sun S, Kikuchi T, Sommer RJ. 2023. Chromosome fusions repatterned recombination rate and facilitated reproductive isolation during *Pristionchus* nematode speciation. *Nat Ecol Evol* 7:424–439.
- Yoshida K, Witte H, Hatashima R, Sun S, Kikuchi T, Sommer RJ. Rapid chromosome evolution and acquisition of environmental sex determination in nematode androdioecious hermaphrodites. *In preparation*.

## Supplementary figures

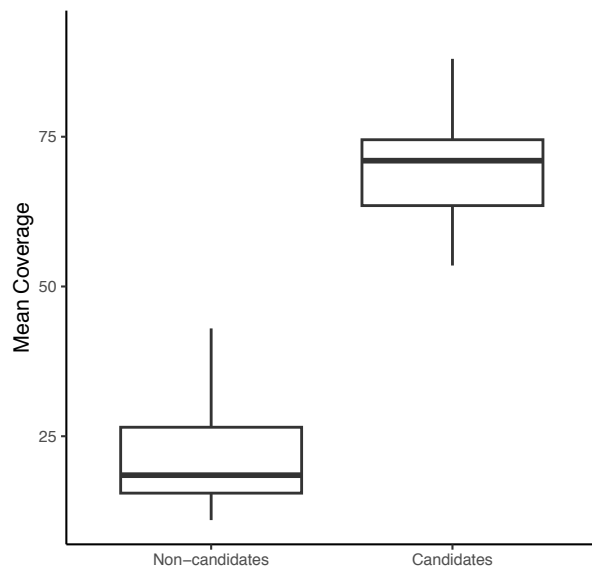


**Figure S1: The male gonadal arm is shown after Hoechst staining.** A primary spermatocyte containing twelve chromosomes is highlighted in a square (Fig. 2B). A maximal intensity image was taken of several stacks. Scale bar: 20  $\mu\text{m}$ .



**Figure S2: Image of *A. sudhausi* sperm.** A DIC image is merged with Hoechst dye (indicated in cyan). One of the sperm cells is highlighted and magnified. Twelve chromosomes can be counted. Scale bars: 5  $\mu\text{m}$ .





**Figure S3: Coverage is far higher for candidate SNPs compared to non-candidates.** The 62 candidate positions identified after examining sex-specific allele-frequencies show surprisingly high coverage, indicating they may be a result of assembly problems.