

# 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).

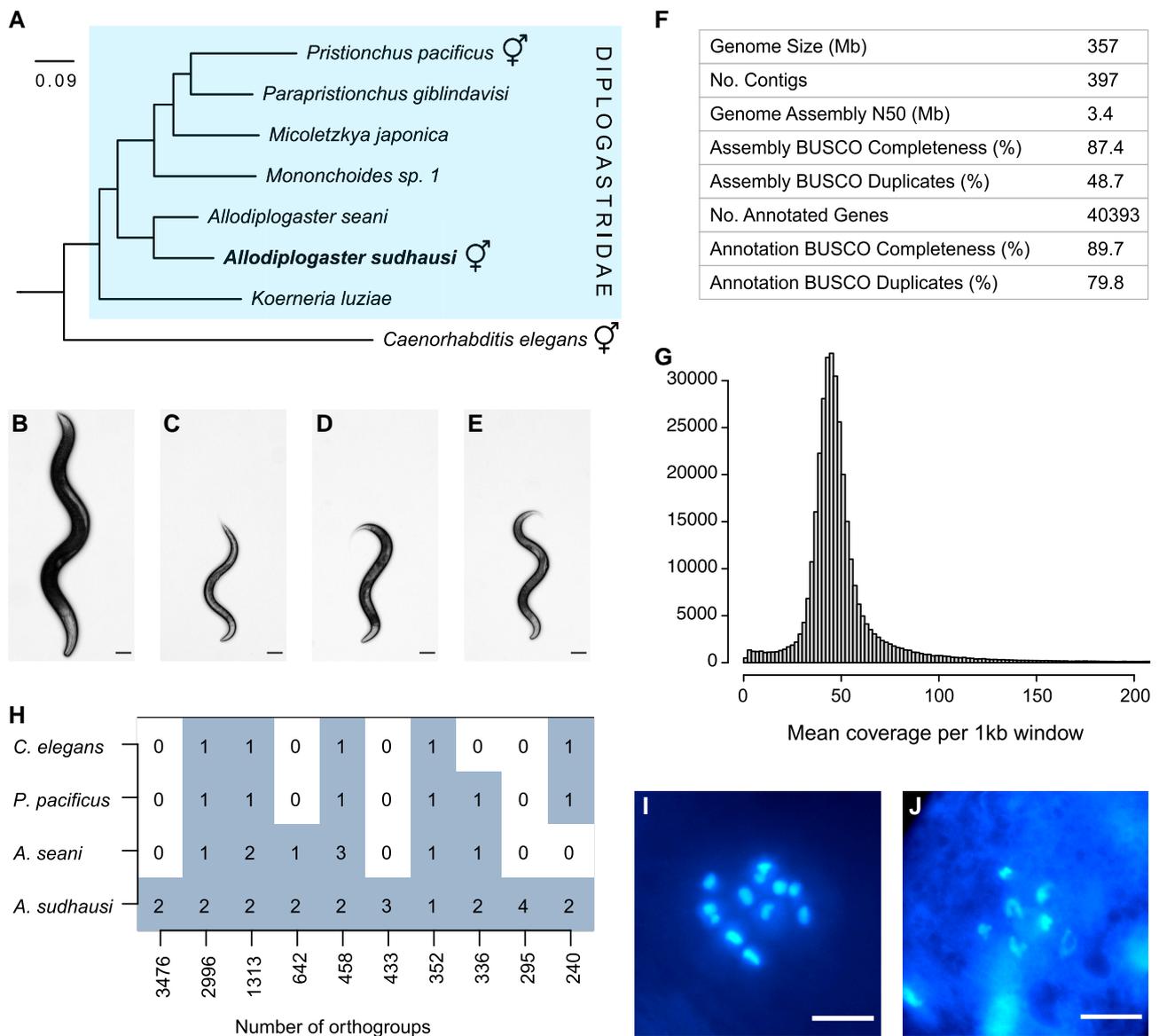
WGD is associated with a number of novelties, including increases in body size (Walsh and Zhang 1992; Otto and Whitton 2000) and an expansion of transposable elements (TEs) after WGD has occurred (Marburger et al. 2018). However, despite the advances in genomics since the effects of WGD were first hypothesized, 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 Diplogastridae, 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 diplogastrid 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.

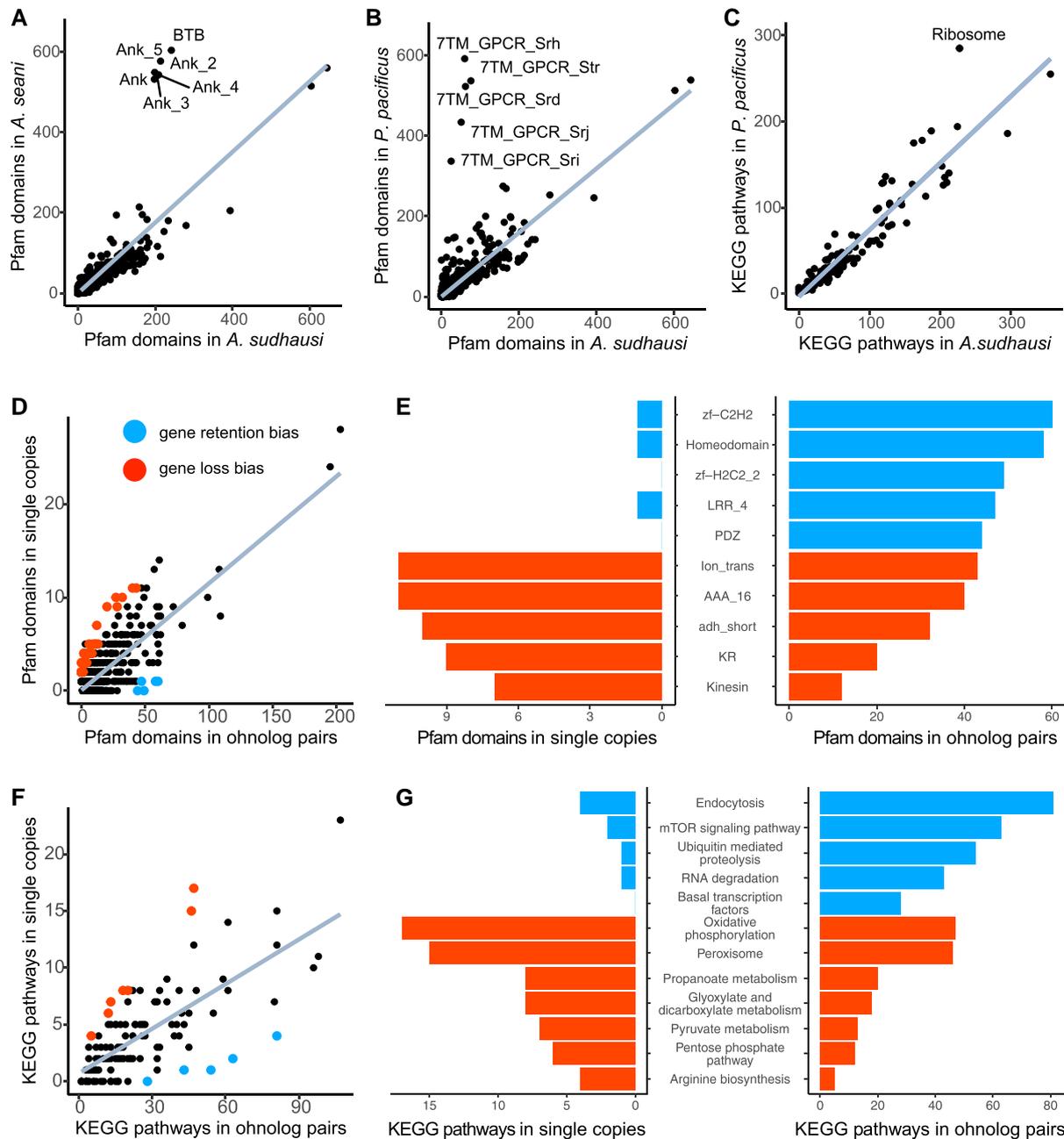
Next, we predicted KEGG (Kanehisa and Goto 2000) metabolic pathways and compared the abundance between species. There appears to have been an expansion of genes involved in the Mitogen-activated protein kinase (MAPK) pathway (04010) in *A. seani* (supplementary fig. S3, Supplementary Material online). 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

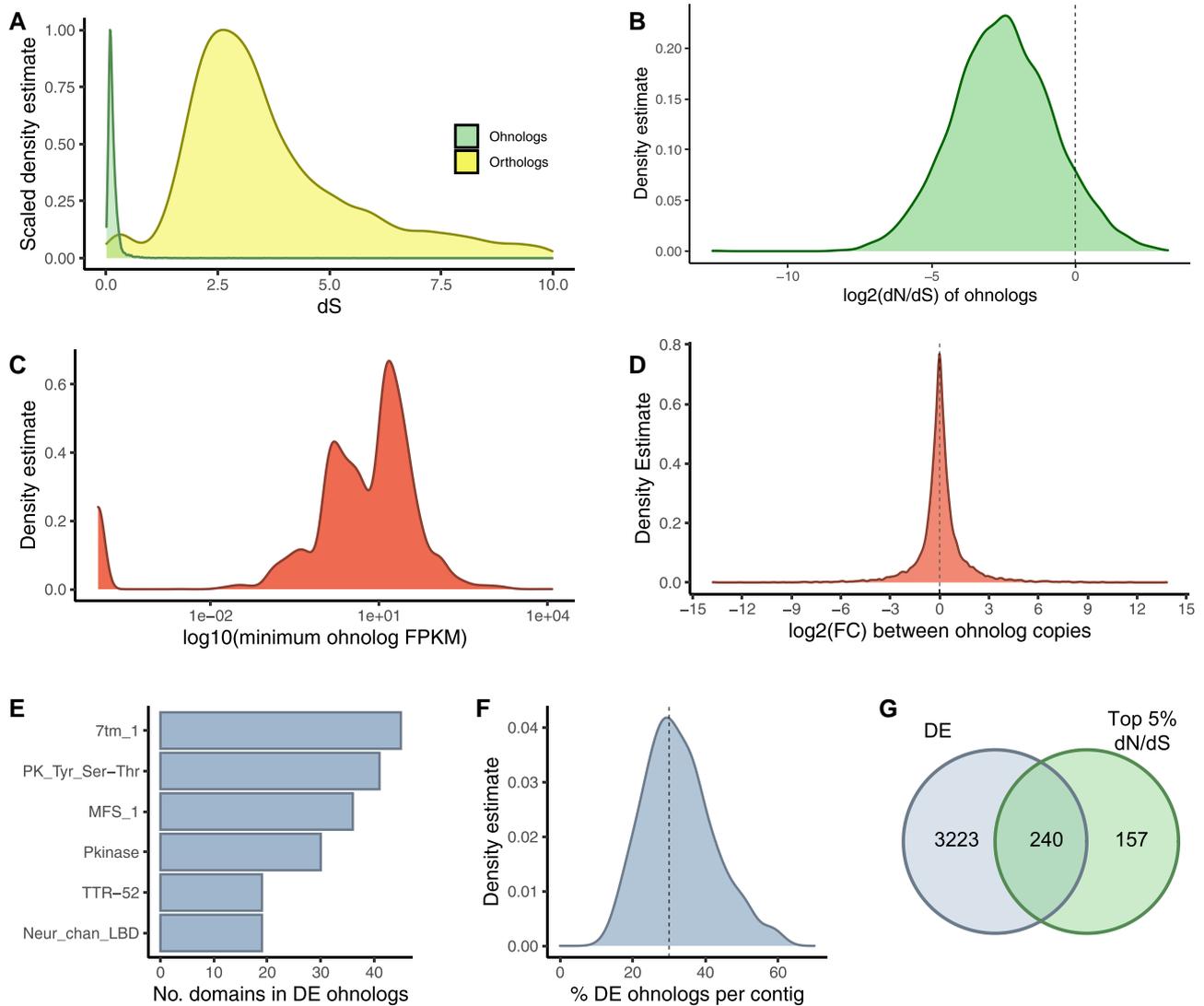
expressed (fig. 3C). Second, when we examined fold change (FC) between ohnolog pairs we found that 30.4% of ohnologs had expression that had a more than two-fold difference in transcript abundance (absolute  $\log_2$ -FC  $\geq 1$ ). However, the majority of ohnolog pairs have little to no difference 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).

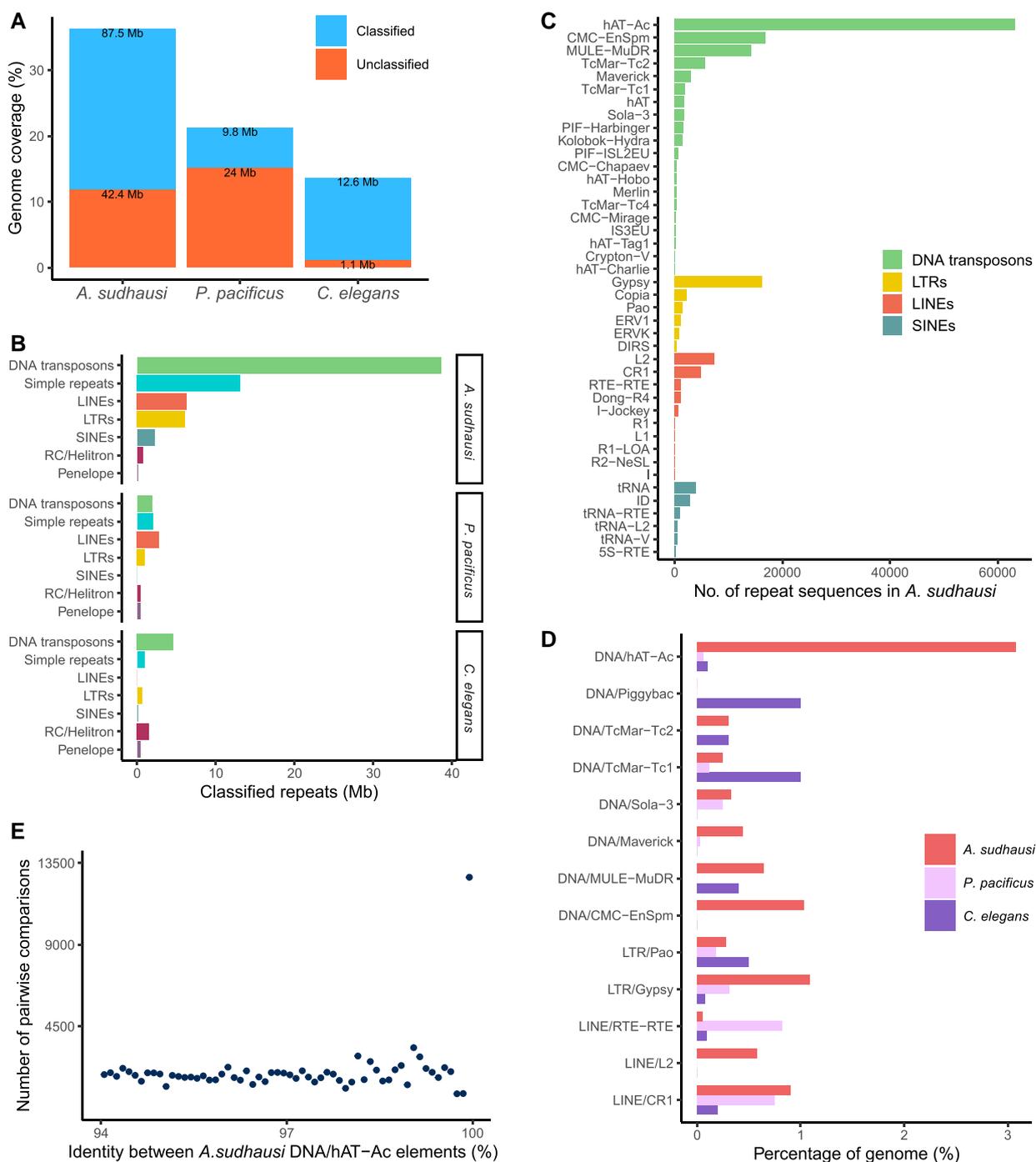
BLAST searches revealed the conservation of *dpy-1* in many species, including members of the Rhabditidae (containing *C. elegans*) and Diplogastridae (containing *A. sudhausi* and *P. pacificus*) (fig. 5A). Based on the phylogeny of orthologous *dpy-1* sequences, it is clear there is one *dpy-1* copy 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.

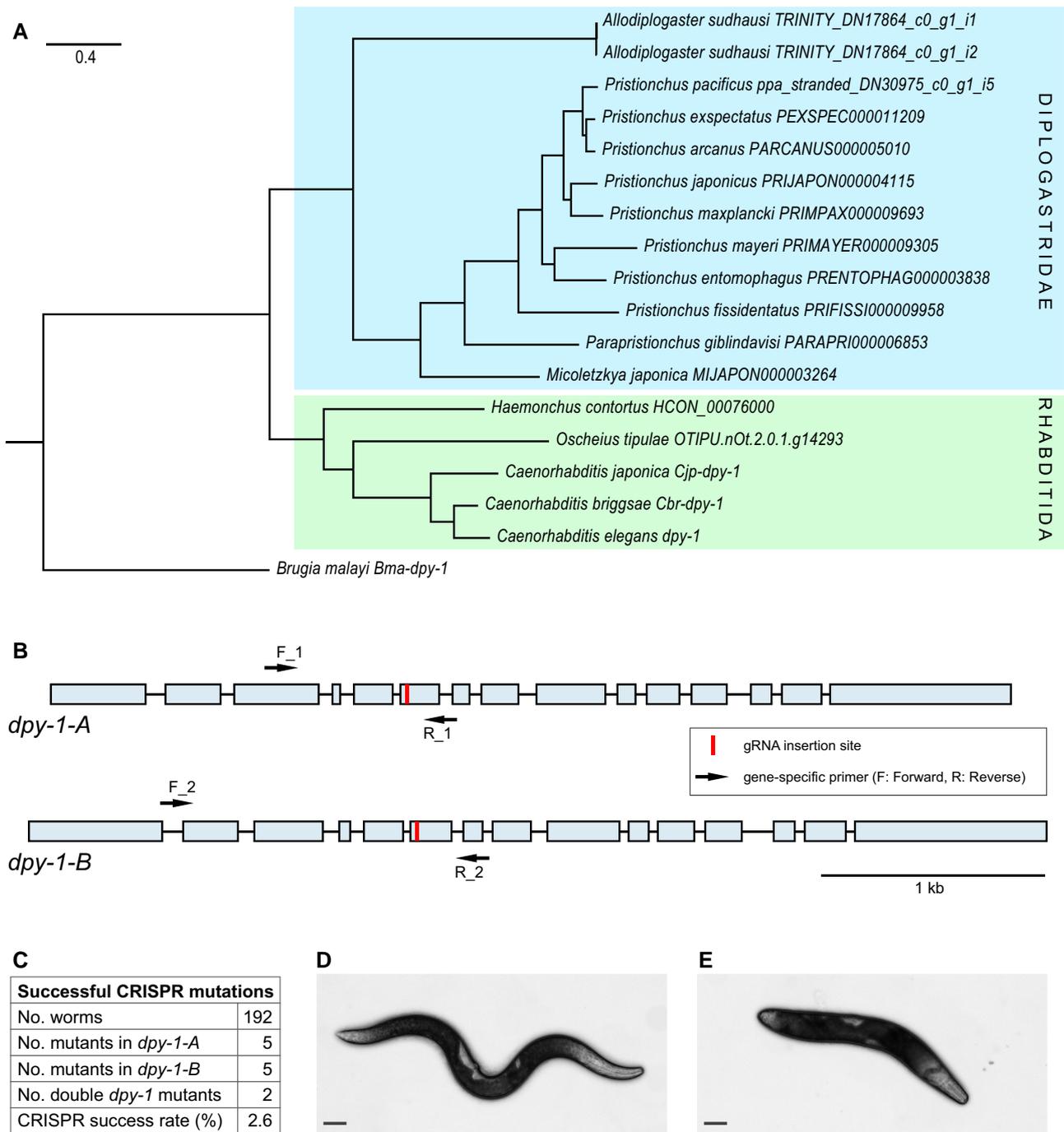
### CRISPR/Cas9-induced *A. sudhausi* *dpy-1* Knock-out Mutants Demonstrate Genetic Redundancy

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 autoployploidization 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 (PiqLab, 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 -dnaworden 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 crispr RNA (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.  
 Arensburger 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, Scharf 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