A maternal-effect toxin-antidote element causes larval arrest in *C. elegans*

Authors:
Stefan Zdraljevic[1,2,3 † #], Laura Walter-McNeill[1,2,3#], Giancarlo Bruni[1,2,3], Joshua S. Bloom[1,2,3], Daniel H.W. Leighton[1,2,3], Heriberto Marquez[1,2,3], and Leonid Kruglyak[1,2,3 †]

Affiliations:

[1] Department of Human Genetics, University of California, Los Angeles, CA, USA
[2] Department of Biological Chemistry, University of California, Los Angeles, CA, USA
[3] Howard Hughes Medical Institute, Chevy Chase, MD, USA

[#]Co-first authors
[†] Corresponding author:
Stefan Zdraljevic, szdralje@g.ucla.edu
Leonid Kruglyak, lkruglyak@mednet.ucla.edu

ORCID
Stefan Zdraljevic: 0000-0003-2883-4616
Laura Walter-McNeill: 0000-0002-8238-5025
Giancarlo Bruni: 0000-0003-2850-4633
Joshua S. Bloom: 0000-0002-7241-1648
Daniel H.W. Leighton: 0000-0002-1379-0078
Leonid Kruglyak: 0000-0002-8065-3057

Abstract

Toxin-antidote (TA) elements are selfish DNA sequences that bias their transmission to the next generation. TAs typically consist of two linked genes: a toxin and an antidote. The toxin kills progeny that do not inherit the TA, while the antidote counteracts the toxin in progeny that inherit the TA. Many TAs have been found in bacteria and archaea, but only a few in eukaryotes. We previously discovered two TAs in *C. elegans*: *peel-1/zeel-1* and *sup-35/pha-1*. Here, we report a third TA in C. *elegans* which has several distinctive features. Unlike the toxins PEEL-1 and SUP-35, the new maternally deposited toxin causes larval arrest rather than embryonic lethality, raising the question of how the toxicity is delayed to this stage. We identified the toxin and antidote genes, which encode proteins of unknown function. We detected toxin mRNA transcripts in Z2/Z3 cells in larvae, which suggests that the transcript is sequestered from translation and degradation in germ cells. At the population level, previously described TAs are characterized by presence or absence of both the toxin and the antidote in a given individual. In contrast, we identified three clades with distinct haplotypes at the new TA locus. In addition to strains with an intact TA and a second set of strains that lost the toxin, many strains, including the standard laboratory strain N2, carry an intact but highly divergent version of the toxin that appears to be active. The antidote is pseudogenized in these strains, presenting the puzzle of how they inactivate the toxin. Current evidence suggests that this is accomplished by MUT-16- and PRG-1-dependent production of small interfering RNAs that target the toxin mRNA.

Introduction

Selfish genetic elements are widespread in nature because they employ diverse strategies to promote their own transmission, even at the expense of organismal fitness [1]. Transposons represent a classic example of selfish genetic elements that are capable of lowering host fitness by disrupting genes after transposition [2–4]. When fitness is compromised, host genomes that suppress transposition of these selfish elements are evolutionarily favored. This conflict initiates an internal Red Queen dynamic in which the host genome and selfish genetic elements constantly evolve to evade each other [5,6].

Toxin-antitoxin or toxin-antidote (TA) elements are extreme examples of selfish genetic elements that typically consist of two linked genes encoding a toxin and a cognate antidote. The toxin kills individuals that don't inherit the element and hence lack the antidote to counteract the effects of the toxin [7–11]. TA elements are ubiquitous in bacteria and likely function as defense mechanisms against bacteriophages, either by directly inhibiting the infection cycle of a phage or by targeting host factors to prevent the spread of mature virions [12]. Like other immune genes involved in pathogen recognition, TA components are poorly conserved across bacteria because they evolve rapidly to maintain a competitive edge against their target phages [13,14].

While the maintenance of extremely toxic genes makes sense in the context of microbial phage-defense systems, a beneficial function of TA elements in metazoans remains elusive despite multiple recent reports of their prevalence. TA elements appear to be common in hermaphroditic *Caenorhabditis* nematodes, even though primarily hermaphroditic reproduction is not conducive to their spread [8–10,15,16]. Each of the known *Caenorhabditis* TA elements resides in a hyper-divergent genomic region, which suggests that these elements predate the evolution of selfing [17]. TA elements are expected to drive to fixation in outcrossing populations. Once it is fixed or nearly fixed, the TA can no longer increase in frequency, which generally causes the element to degrade over time unless it provides a fitness advantage, as has been observed in bacteria. A recent report suggests that *peel-1*, the toxin component of the first TA element discovered in *C. elegans*, increases host fitness in laboratory conditions, raising the possibility that toxic genes can take on new roles that allow them to be maintained at high frequencies in primarily selfing nematode populations [18].

Here, we describe a novel maternally inherited TA element in *C. elegans* with distinctive features. The maternally deposited toxin causes larval arrest rather than embryonic lethality, raising the question of how the toxicity is delayed to this late developmental stage. At the population level, we identified three clades with distinct haplotypes at the new TA locus, the most common of which appears to possess a functional toxin without an antidote.

Results

Identification of a novel *C. elegans* toxin-antidote element

To study the phenotypic effects of genetic variation in *C. elegans*, we generated large cross populations between highly divergent strains—XZ1516 x QX1211 and XZ1516 x DL238. We chose these strains because they are compatible at the two incompatibility loci we previously discovered: *peel-1/zeel-1* and *sup-35/pha-1*[8,9]. We introduced a *fog-2* loss-of-function allele, which feminizes hermaphrodites and prevents them from selfing, into each genetic background to facilitate the construction of large cross populations and intercrossed each population for 10 generations, with minimal selection. Despite minimal selection across each generation, whole-genome sequencing across generations revealed multiple genomic loci with allele frequency distortions, indicating that genetic differences at these loci influenced relative fitness in standard laboratory growth conditions (Supplemental Figure 1A). We observed that by generation four of the XZ1516 x QX1211 cross, the XZ1516 allele frequency rose to 75% on the right arm of chromosome V (Figure 1A). We also observed allele frequency distortion at this region in later generations of the XZ1516 x DL238 cross, which suggested that the same underlying genetic difference was being selected in both crosses. Based on previous studies, we hypothesized that this strong depletion of the QX1211 genotype by generation four is caused by a toxin-antidote (TA) element at this locus [19]. To test this hypothesis, we performed new crosses between QX1211 and XZ1516 and tracked the phenotypes and genotypes of F2 progeny. We observed that ~27% of the F2 self-progeny of heterozygous QX1211/XZ1516 F1 hermaphrodites arrested as L1 larvae (Figure 1B). The observed larval arrest phenotype is reminiscent of the rod-like larval lethal (*rod*) phenotype [20]. When we crossed QX1211/XZ1516 F1 hermaphrodites to QX1211 males, ~47% of the progeny exhibited the *rod* phenotype, while we observed no *rod* progeny in the reciprocal cross between QX1211/XZ1516 F1 males and QX1211 hermaphrodites (Figure 1B). The *rod* progeny were all homozygous for QX1211 alleles at the locus on the right arm of chromosome V that displayed the allele frequency distortion in the mapping populations. This inheritance pattern suggests that the XZ1516 genome encodes a maternally inherited toxin and a linked zygotically expressed antidote that form a novel TA element responsible for the observed allele frequency distortions on the right arm of chromosome V (Figure 1C). We observed the same F2 phenotypes in crosses between DL238 and XZ1516, indicating that DL238 is a noncarrier of the TA element (Supplemental Figure 1B).

Identifying the components of the XZ1516 toxin-antidote element

To isolate the XZ1516 TA element, we introgressed the right arm of chromosome V from XZ1516 into QX1211. We confirmed the identity of the resulting near-isogenic line (NIL) by whole-genome sequencing and verified the presence of the TA element with crosses (Figure 2A). We were unable to further localize the TA location with standard fine-mapping approaches, likely because of low recombination rates near the ends of *C. elegans* chromosomes [21]. To overcome the limited natural recombination in this region, we developed a method to induce targeted recombination at double-stranded DNA breaks generated by Cas9 [22]. This approach

enabled us to localize the TA element to a 50 kb region containing 10 candidate genes. We tested these genes for potential toxin or antidote activity by systematically knocking them out in the XZ1516 genetic background (Figure 2A).

We isolated three deletion strains that did not induce larval lethality when crossed to QX1211, suggesting that these strains lacked the toxin (Figure 2B). The computationally predicted gene FUN_019829 is deleted in all three of these strains, and in one of the strains only this gene is deleted, confirming that this gene encodes the toxin. We hereafter refer to FUN_019829 as *mll-1* (Maternal Larval Lethal). We were unable to generate homozygous deletion lines of gene FUN_019825, which suggested that this gene is either essential for survival or encodes the antidote. We successfully isolated homozygous deletion lines of FUN_019825 in a Δ*mll-1* genetic background, indicating that this gene encodes the antidote. We hereafter refer to FUN_019825 as *smll-1* (Suppressor of Maternal Larval Lethal). We showed that a strain with deletions of both *mll-1* and *smll-1* behaves like susceptible strains in crosses (Figure 2B).

To determine whether *smll-1* is sufficient to suppress *mll-1*-induced larval lethality, we constructed a rescue plasmid to drive *smll-1* expression with a constitutive promoter. We injected the rescue plasmid into XZ1516, crossed individuals harboring the rescue array to a TA-susceptible strain, and selfed the F1 progeny that inherited the array. We observed a dramatic reduction in larval arrest from 25% to 3.5% in F2 progeny, and all F2 progeny that inherited the rescue array survived. These results confirm that *smll-1* is sufficient to suppress *mll-1* toxicity (Figure 2B).

Long-read RNA sequencing revealed two distinct *mll-1* isoforms, a short isoform with three predicted exons and a long isoform with eight predicted exons (Supplemental Figure 2A). We constructed plasmids with inducible versions of each *mll-1* isoform. When we injected susceptible strains with the short *mll-1* isoform array, every F1 individual carrying the array died, with 64% of larvae exhibiting the *rod* phenotype, indicating that uninduced expression levels of the short *mll-1* isoform are sufficient to induce lethality. By contrast, we were able to isolate susceptible strains that maintained the long *mll-1* isoform array or a short *mll-1* isoform array with a premature stop codon in *mll-1*. We observed no *rod* progeny upon induction of these arrays, indicating that the short isoform encodes the functional toxin, and that the toxin acts as a protein.

Because lethality only occurs at the L1 stage, we reasoned that *mll-1* might be deposited in embryos as a transcript and sequestered from translation. We performed fluorescence *in situ* hybridization (FISH) on developing XZ1516 embryos and larvae with RNA probes that target the *mll-1* mRNA. We observed *mll-1* puncta as early as the 2-cell embryo stage (Supplemental figure 3A), indicating that *mll-1* transcripts are maternally deposited because zygotic transcription does not initiate prior to the 4-cell stage [23]. At later embryonic and L1 development stages, the *mll-1* transcript is localized to two cells that likely correspond to the primordial germ cells (Supplemental figure 3B-C).

Genomic and population features of the *mll-1*/*smll-1* TA element

The XZ1516 genomic region surrounding the *mll-1*/*smll-1* TA element is hyper-divergent from the reference (N2) genome [17]. We characterized the genetic variation at this region in the *C. elegans* population by calculating the relatedness of 550 wild isolates [24]. This analysis separated the population into three distinct clades: an XZ1516-like TA clade which contains 29 strains, a 10-strain clade, and an N2-like susceptible clade composed of 511 strains, including QX1211 and DL238 (Figure 3A). We verified that the 28 additional isolates with the XZ1516-like haplotype have intact *mll-1/smll-1* genes by aligning sequencing reads from these isolates to the XZ1516 genome assembly. All but four of the isolates in the XZ1516-like clade were collected within three miles of each other on the island of Kauai, two were isolated on Oahu, and one each on Maui and Moloka'i. Four of the isolates from the 10-strain clade were isolated on Maui and the remaining six are globally distributed, while strains with the susceptible haplotype, which represent the majority of the known *C. elegans* isolates, are globally distributed and present on all the Hawaiian islands with the exception of Moloka'i (Figure 3B).

The 10-strain clade carries a haplotype that does not contain a gene resembling the toxin. However, this haplotype does carry a divergent *smll-1* allele that is predicted to contain a full-length coding sequence. We therefore asked whether this *smll-1* allele is capable of suppressing the toxic effects of *mll-1*. We observed the *rod* phenotype in only 3% of F2 progeny derived from crosses between XZ1516 and a representative strain with this haplotype, NIC195, (Figure 3C), indicating that this antidote is at least partially functional. When we knocked out the *smll-1* allele in NIC195, 22.5% of F2 progeny were *rod*, confirming that this divergent allele confers reduced susceptibility to the effects of *mll-1* (Figure 3C).

While the previously described *C. elegans* TA elements are characterized by their absence in susceptible strains [8,9], the N2 genome harbors a divergent allele of *mll-1* with an intact coding sequence, as well as a pseudogenized version of *smll-1*. The *mll-1/smll-1* genomic region contains several genomic rearrangements between XZ1516 and N2, including likely inversion events that occurred between *smll-1* and its corresponding divergent N2 allele, *B0250.4*; these inversions may have contributed to its pseudogenization (Figure 3D). While synteny is maintained between *mll-1* and the corresponding divergent N2 allele, *B0250.8*, many of the surrounding N2 genes are predicted to be pseudogenized. The divergence between *mll-1* and *B0250.8* is the highest among one-to-one orthologs in the XZ1516 and N2 genomes (nucleotide identity: 63%; protein identity: 47%) (Figure 3E). We estimated the divergence time for these two alleles under the assumption of neutrality to be between 160 and 325 million generations based on estimates of divergence at synonymous sites (dS)[25,26]. This implausibly old estimate suggests that positive selection has been driving the diversification of this gene. The fact that *B0250.8* has an intact coding sequence raises the question of whether this gene has maintained its function as a toxin, and if so, how individuals with this haplotype can exist without a functional antidote.

To determine whether *B0250.8* acts as a toxin, we used a tetracycline-inducible system to drive the expression of *B0250.8* in XZ1516, DL238, and N2. We hatched worms carrying the inducible array on doxycycline plates to induce *B0250.8* expression and recorded their phenotypes 48

hours after hatching. The majority of worms expressing *B0250.8* displayed a variety of abnormal phenotypes (91% affected N2 (n=58); 100% affected DL238 (n=42); 60% affected XZ1516 (n=61)). Notably, we observed the stereotypical *mll-1*-dependent *rod* phenotype at low frequencies in all strains (2/58 N2, 5/42 DL238, 4/61 XZ1516). Furthermore, the abnormal phenotypes we observed upon induction of *B0250.8* were also seen upon induction of *mll-1* (Supplemental Figure 4), which suggests that *B0250.8* has retained its function as a toxin. The presence of a functional toxin and a pseudogenized antidote in N2-like strains suggests that a different mechanism suppresses the toxicity associated with *B0250.8* and that this suppression mechanism does not affect the XZ1516 *mll-1* toxin (Figure 1B).

Small-RNA-mediated suppression of the *B0250.8* toxin

A potential mechanism that N2-like strains could employ to suppress the activity of *B0250.8* is RNA interference (RNAi). RNAi pathways are evolutionarily conserved and can act to silence the expression of potentially deleterious genes [27]. In these pathways, argonaute proteins interact with small RNAs (sRNAs) to transcriptionally and post-transcriptionally regulate gene expression. In *C. elegans*, primary sRNAs initiate the amplification of secondary small inhibitory RNAs (siRNAs) in perinuclear granules known as *Mutator* foci [28,29]. MUT-16 is a glutamine/asparagine (Q/N)-rich protein that is required for the formation of *Mutator* foci at the nuclear periphery of germline nuclei [29]. Previous work has shown that *B0250.8* is heavily targeted by secondary 22G siRNAs, the production of which is dependent on MUT-16 and other *Mutator* foci components [29]. Animals in which *mut-16* is disrupted with a *mut-16*(*pk170*) mutation show a 137.7-fold decrease in 22G siRNAs that target *B0250.8* and a corresponding 23.9-fold increase in the expression level of the gene [30]. Furthermore, high levels of larval arrest occur in mutant strains where *Mutator* foci formation is disrupted, including in *mut-16*(*pk170*) strains [31]. Consistent with this report, we observed that ~15% of Δ*mut-16* progeny arrest at various larval stages, and 2% of progeny are *rod*, which is suggestive of derepression of *B0250.8*.

Amplification of 22G siRNAs can be initiated by different primary sRNAs, including ERGO-1- and ALG-3/4-dependent 26G siRNAs and PRG-1/2-dependent 21U piRNAs. Given that production of MUT-16-dependent 22G siRNAs can be initiated by multiple independent pathways, we queried published sequencing data for sRNAs that are complementary to *B0250.8* [32]. This search identified multiple sRNAs that bind throughout the length of the *B0250.8* transcript. All but one of these sRNAs were not dependent on the argonautes in the queried datasets (NRDE-4, WAGO-9, PRG-1). We identified one PRG-1-dependent sRNA with a binding site just downstream of two predicted piRNAs, 21ur-8336 and 21ur-14170, suggesting that piRNA recognition of the *B0250.8* transcript might be involved in its regulation. Furthermore, small RNA sequencing following immunoprecipitation of PRG-1 identified several piRNAs that are bound to *B0250.8* (21ur-2794, 21ur-2025, 21ur-9583, 21ur-5840, 21ur-4143) [33]. Interestingly, the 22G siRNAs bound to the *B0250.8* transcript surrounding these piRNA binding sites did not depend on the PARN-1 RNase that trims piRNAs [33]. In line with these observations, 22G siRNAs that target *B0250.8* are significantly downregulated in *prg-1*(*n4357*) gonads as compared to wild type (fold change -17.1; adjusted *p*-value = 2.2e-16) [30]. The

depletion of these PRG-1-dependent siRNAs coincides with a 10.3-fold increase in expression of *B0250.8* in *prg-1*(*n4357*) gonads [30]. PRG-1-dependent 22G siRNAs produced in the *Mutator* foci interact with the WAGO-1 argonaute in P-granules to silence transcripts [34]. Recent work has shown that the *B0250.8* transcript is co-immunoprecipitated with WAGO-1, providing additional evidence that this transcript is regulated by the endogenous RNAi machinery [35]. Taken together, these observations suggest that strains with the N2-like haplotype suppress toxic activity of *B0250.8* through post-transcriptional silencing mediated by MUT-16-dependent 22G siRNAs that are partially dependent on PRG-1 activity.

Discussion

We identified a novel toxin-antidote element in *C. elegans* that consists of two genes, *mll-1* and *smll-1*, that encode a maternally deposited toxin and a zygotically expressed antidote, respectively. Unlike the previously characterized toxins in *C. elegans*, PEEL-1 and SUP-35, which induce embryonic lethality in susceptible strains, MLL-1 induces *rod*-like larval lethality. The delayed onset of lethality suggests that the *mll-1* transcript is sequestered from translation and degradation throughout embryogenesis and into the early larval stages. This is supported by our observations that *mll-1* mRNA is distributed across all cells in early embryonic development but is only present in the Z2/Z3 germ cells in older embryos and L1 larvae. While *mll-1* has no detectable homologous sequences across all sequence databases and a very low-confidence protein structure prediction, the induction of the *rod* phenotype by MLL-1 in susceptible strains suggests that it disrupts osmoregulation in the absence of SMLL-1. The *rod* phenotype is caused by fluid filling of the *C. elegans* pseudocoelom and has been observed after laser and genetic ablation of the excretory canal cell, duct cell, pore cell, or CAN neurons [36–38], which suggests that these cells are affected by MLL-1.

A unique feature of the *mll-1/smll-1* element is that three distinct haplotypes exist across the *C. elegans* population. The XZ1516-like haplotype that we originally identified in two crosses represents a “standard” toxin-antidote element that encodes toxin and antidote proteins. The NIC195-like haplotype represents a snapshot of an expected evolutionary trajectory for a toxin-antidote element, in which the toxin is lost through mutation and the antidote is no longer needed to counteract the toxin. This is supported by the accumulation of mutations in the NIC195 antidote that have reduced its ability to counteract the MLL-1 toxin. These two haplotypes are present in 7% of the known *C. elegans* strains, while the remaining 93% of strains have the N2-like haplotype.

The N2-like haplotype contains an allele of *mll-1* (*B0250.8*) that is the most divergent one-to-one ortholog between the N2 and XZ1516 genomes. It is important to note that the two *mll-1* alleles are hyperdivergent at the nucleotide level, as indicated by extremely high dN (0.56) and dS (1.77) values, and an average dN/dS ratio of 0.32. This value of dN/dS is indicative of purifying selection on the protein sequence, which is in line with our results showing that this divergent *mll-1* allele has retained its toxicity, and raises the question of how strains with this haplotype neutralize the toxin. The elevated dN and dS values give implausibly long estimates for the

divergence time between these two alleles and suggest that positive selection has been driving the diversification of this gene at the nucleotide level. Previous work has shown that B0250.8 is one of the five protein-coding genes that are most heavily targeted by 22G siRNAs, and that the production of these 22G siRNAs is dependent on both PRG-1 and MUT-16 [29,30]. It has been proposed that after mRNA transcripts exit the nucleus, they are marked for siRNA-mediated silencing in perinuclear P granules by various argonauts, including PRG-1. After being marked for silencing, these transcripts are routed to the *Mutator* foci, where 22G siRNAs are produced by *Mutator* class genes and the transcripts are recognized by silencing argonautes like WAGO-1. We propose that positive selection for piRNA binding sites in the *mll-1* transcript drove the diversification of this gene toward the N2 version. Once the transcript was recognized by genomically encoded piRNAs, it became marked for silencing and shuttled to the *Mutator* foci for 22G siRNA amplification, and the *smll-1* antidote gene was pseudogenized because it was no longer needed to counteract the toxin.

Materials and methods

Strain maintenance

Unless otherwise specified, all strains were propagated at 20 ºC on a modified nematode growth medium (NGMA) containing 1% agar and 0.7% agarose and fed *Escherichia coli* strain OP50 [39]. Strain names and genotypes can be found in Supplementary Table 1.

Plasmids

Plasmid descriptions can be found in Supplementary Table 2. All plasmids generated in this study (with the exception of pWM17) were assembled using gene sequences ordered from IDT (gBlocks) and plasmid backbones (ordered from Addgene or derived from an existing Addgene plasmid) using the NEB Gibson Assembly Master Mix (#E2611). Gibson assemblies were transformed the assembly mix into DH5α competent cells (NEB #C2987H). Plasmid backbones were digested with restriction enzymes and purified using Qiagen QIAquick gel extraction kit (#28706). The gBlocks ordered from IDT, the restriction enzymes used, and the primers used to add homology arms to backbones can be found in Supplementary Table 3. pWM17 was generated using the NEB Q5 site-directed mutagenesis kit (#E0554S) using pWM11 as the template and primers oZ288 & oZ289. All plasmids were purified using the Invitrogen PureLink HQ Mini Plasmid DNA Purification Kit (#K210001) and verified using whole plasmid sequencing with Primordium.

Multi-generational cross

QX2538 (*fog-2*(qq212[P17stop]) in XZ1516), QX2539 (*fog-2*(qq212[P17stop]) in QX1211), and QX2327 (qqIr39[fog-2(q71), N2 > DL238] V) were used to generate the cross populations. QX2538 males were used to start the QX2538 x QX2539 cross population and QX2327 males were used to start the QX2327 x QX2538 cross population. Crosses were amplified on 10cm NGMA plates until approximately 20,000 worms were obtained, at which point the population size was maintained. For each generation, worms were washed off the plates using M9, bleach synchronized, and arrested overnight as L1s. The following day, 20,000 L1s were plated across ten 10cm plates at ~2000 worms/plate. The crosses were propagated until the 10th generation of intercrossing. Multiple timepoints were saved for sequencing by freezing a pellet of several thousand worms. For generation ten, DNA libraries were generated using the Illumina Nextera XT DNA Library Preparation Kit (#FC-131-1024) and whole genome sequenced using Illumina NextSeq2000 P1 reagents (20074933). For generation four, allele frequencies were inferred as previously described [40].

Crosses

Descriptions of the crosses and their phenotypes can be found in the Supplementary Table 4. Workflow for crosses varied slightly depending on the use of fluorescent males. For all crosses,

males and L4 hermaphrodites were allowed to mate 24-36 hours before potentially mated hermaphrodites were singled out onto fresh plates. For crosses set up with wild type males, only progeny from plates with a 50:50 male to hermaphrodite ratio were used for subsequent crosses. For crosses set up with fluorescent males, only progeny with the fluorescent marker were used for subsequent crosses. For self crosses, ~10 hermaphrodite cross progeny were transferred to fresh agar plates and allowed to lay embryos for ~24 hours. For parental backcrosses, ~5 L4 cross progeny were allowed to mate with the designated parent for 24-36 hours before potentially mated hermaphrodites were singled out onto fresh plates and laid embryos for ~24 hours. For all crosses, a defined number of embryos were transferred to fresh plates and allowed to hatch overnight. The following day, the number of arrested larvae were counted.

NIL construction

The QX2500 and QX2501 NILs were constructed as previously described [22]. Briefly, QX2500 was constructed by genotyping F2 progeny of a QX1211 x XZ1516 cross. DNA from F2/F3 progeny was amplified using primer pairs: oZ50-51 and oZ52-53, which enabled us to identify recombinants within a defined genomic region. Recombinant progeny identified with these primer pairs were recovered and backcrossed to QX1211 for 6 generations. We generated QX2501 using Cas9-induced non-homologous recombination to induce strand exchange [22]. Young adult QX2500/QX1211 heterozygotes were injected with four target gRNAs (gSZ63, gSZ65, gSZ68, and gSZ70), the *dpy-10* gRNA and repair template (see Cas9 injections). We transferred F2 *rol* animals to 96-well plates, allowed them to self, and identified recombinant individuals using oZ66-67 and oZ64-65. We identified and isolated a recombinant that we named QX2501, which contains the QX1211 genotype from V:1-21,536,657, followed by a deletion spanning V:21,536,657-21,547,827, and the XZ1516 genotype from V:21,547,828-22,058,188. All genotyping primers can be found in Supplementary Table 5.

Fine-mapping the element

We previously described the construction of the 10 gene CINR-generated NIL [22]. Briefly, young adult QX1211/QX2501 were injected with gSZ71 and progeny were genotyped with oZ80-82-86 and oZ64-65 to identify individuals that underwent a loss of heterozygosity event.

Injection workflow

The following workflow was used for all injections (candidate gene KOs, antidote rescue, and tet-inducible system). The day before injections, L4 animals were transferred to fresh 6 cm NGMA plates and allowed to develop overnight. The following day, young adults were injected and transferred to a fresh 6 cm NGMA plate to recover. The injected animals were singled to fresh 6 cm NGMA plates approximately 16 hours after injections and monitored for the co-injection phenotype.

Candidate gene knockouts with CRISPR/Cas9

All guide RNAs were designed to target the XZ1516 genome using the *multicrispr* R package [22,41,42]. Guide RNAs were purchased from Synthego as synthetic spacer-scaffold fusions. Guide RNAs were resuspended in 30 µl of water (50 µM) and stored at -20ºC. Cas9 was purchased from IDT (cat #1081059) and stored in single-use aliquots (0.5 µl ~ 5 µg Cas9) at -80ºC. Injection mixtures were made on ice immediately before use. For each mixture, gRNAs (including oZ30; *dpy-10*) were added to a Cas9 aliquot and incubated at 37ºC for 10 minutes. Then the *dpy-10* single-stranded oligodeoxynucleotide (ssODN) repair template (oZ31) and water were added to a final volume of 20 µl. The mixture was spun down in a table-top centrifuge at maximum speed for five minutes and 10µl was taken off of the top to use in injections. The final concentrations in the injection mixtures were: 1.5 µM of Cas9, 4.45 µM gRNA (with each gRNA represented equally), and 0.5 µM of the ssODN repair template. gRNA and repair template sequences can be found in Supplementary Table 6.

Antidote rescue

The antidote rescue injection mixture was made on the same day as injections. The mixture was made at room temperature, spun down at maximum speed for five minutes, and 10µl was taken off of the top to use in injections. The final concentrations were: 30 µM pWM4, 5 µM pCFJ104, and 65 µM GeneRuler 1kb DNA ladder (CAT#SM0311). Both DL238 and XZ1516 were injected. Progeny of injected animals that expressed the co-injection marker pCFJ104 were isolated and propagated. DL238 co-inj (+) animals were crossed to XZ1516 WT and XZ1516 co-inj (+) animals were crossed to DL238 WT to assess lethality. In addition to the previously described cross workflow, we also took note of whether not dead animals inherited the array (as indicated by co-inj (+)).

Long read direct RNA sequencing

We extracted and purified RNA (MasterPure #MC85200) from a mixed stage culture of XZ1516. A TapeStation confirmed the high quality of the sample (RINe = 9.3). We prepared long read RNA libraries using Oxford Nanopore library prep kit #SQK-RNA002 on the purified XZ1516 RNA. Libraries were run on a MinION Flow Cell (R9.4.1), bases were called using *guppy*, and the long reads were aligned to the XZ1516 genome using *minimap2 [43]*.

Tet-inducible system

We used a tet-inducible system to test the toxicity of several constructs [44]. The system consists of three plasmids: a tet promoter driving the CDS of interest, a tet activator (TC374), and a tet GFP (TC358). Injection mixtures were made at room temperature, spun down at maximum speed for five minutes, and 10µl was taken off of the top to use in injections. The final concentrations were: 5 µM expression construct (pWM8, 11, 12, 17, or 21), 5 µM TC374, 5µM TC358, 5 µM pCFJ104, and 80 µM GeneRuler 1kb DNA ladder (CAT#SM0311). DL238 was injected with pWM11, pWM12, and pWM17 to test *mll-1* toxicity. DL238, XZ1516, and N2 were

injected with pWM21 to test *B0250.8* toxicity and with pWM8 as a negative control for all tet-inducible injections. Progeny of injected animals that expressed the co-injection marker pCFJ104 were isolated and propagated. We assessed toxicity of the expression construct of interest through induction on NGM plates with ~0.17% doxycycline hyclate (Sigma # D9891), seeded with HT115. Most often pCFJ104 (+) gravid worms were bleached onto dox and control plates and phenotyped 48 hours later.

Microscopy

RNA FISH: The sequence corresponding to the *mll-1* transcript was uploaded to Stellaris's probe design feature, which yielded a mixture of 25 RNA probes, each 19 nucleotides long (Biosearch Technologies). Probes were labeled with CAL Fluor Red 610. XZ1516, DL238, and QX2513 were grown in liquid cultures to amplify the populations. Gravid adults were bleached from liquid cultures and embryos were fixed immediately or allowed to arrest in liquid overnight. Embryos and arrested L1s were prepared according to the Stellaris RNA FISH Protocol for *C. elegans* with the following modifications. Rather than using chambered coverglass, the hybridization and subsequent wash steps were conducted on fixed embryos or larvae in a microcentrifuge tube. ProLong™ Glass Antifade Mountant (ThermoFisher #P36982), rather than Vectashield mounting medium, was added to prepared samples before mounting on slides. Images were taken using a NIKON Eclipse Ti2 widefield microscope with a 100X/1.45 plan apochromat lambda D oil objective, a Photometrics Prime 95B large field of view monochrome fluorescence camera, and a SpectraIII/Celesta/Ziva, MultiLaser(SpectraIII/Lida) lightsource. The following filters were used: UV excitation 365nm, emission 435nm, and mCherry excitation 514nm, emission 543nm. z-stack images were taken with a 0.2μm step size. Image analysis was performed in NIS Elements AR5.4.

Population analysis of the TA element

To determine the strain relatedness of the *C. elegans* population, we extracted variants in the region surrounding the B0250.4 and B0250.8 (V:20454811-20473950) from the CeNDR VCF (version - 20220216). After subsetting the VCF, we used the *vcf2dist* and *dist2tree* functions in the *fastreeR* R package to generate the relatedness dendrogram.

We used the R package *orthologr* to calculate nucleotide and amino acid identity and dS [45]. We calculated dS across all orthologs between the N2 and XZ1516 genomes using all of the methods available in *orthologr*. We used the following formula to calculate divergence time: $T = (dS - \pi anc)/(2\mu)$ [25], as previously described [26].

Figures

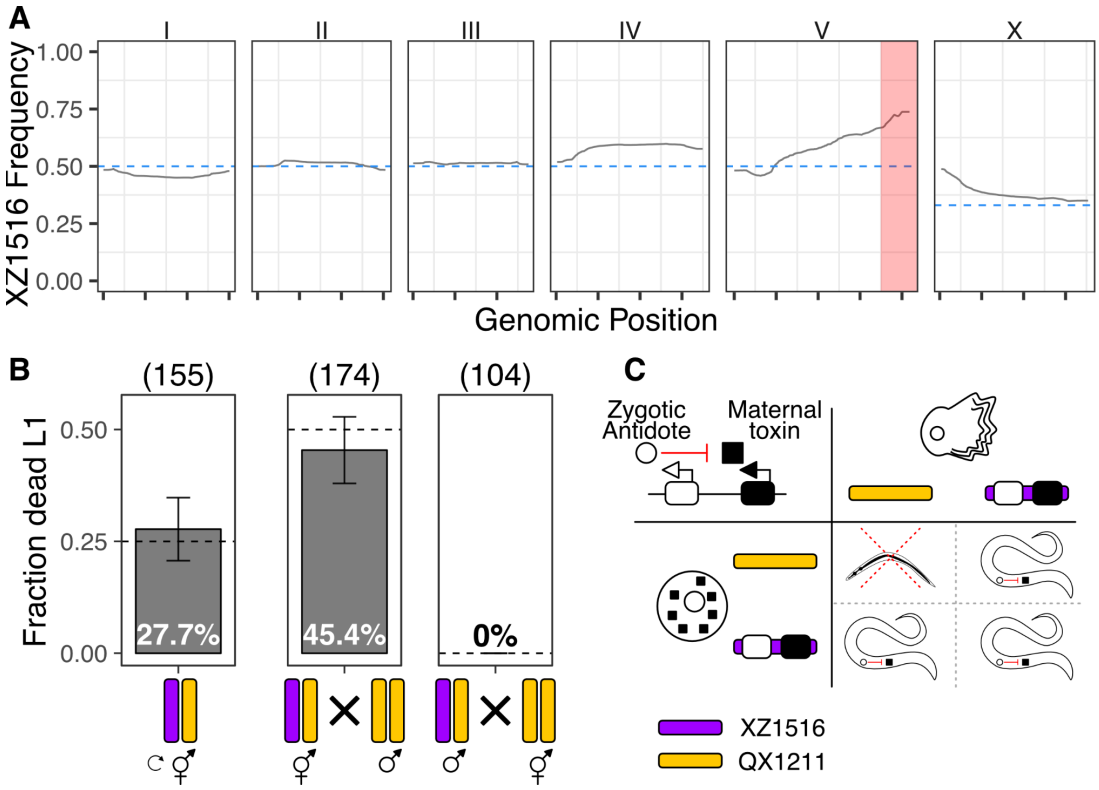


Figure 1. A) Frequency of XZ1516 alleles across the genome after four generations of intercrossing with QX1211. Each panel corresponds to a *C. elegans* chromosome and each x-axis tick indicates 5 Mb. The dotted blue line represents the expected allele frequency for each chromosome. The region highlighted in red on the right side of chromosome V shows the greatest allele frequency deviation from expectation. B) Crosses between XZ1516 (purple) and QX1211 (yellow) establish the inheritance pattern of the TA element. Bar plots show the fraction of dead L1s observed in each cross. Error bars indicate 95% binomial confidence intervals calculated using the normal approximation method. Crosses from left to right: selfing of XZ1516/QX1211 heterozygous hermaphrodites; XZ1516/QX1211 heterozygous hermaphrodites crossed to QX1211 males; XZ1516/QX1211 heterozygous males crossed to QX1211 hermaphrodites. C) Model of the TA inheritance. Punnett square shows the lethality pattern expected in progeny from selfing of XZ1516/QX1211 heterozygous hermaphrodites. A maternally deposited toxin (black square) is present in all progeny and causes L1 lethality unless a zygotically expressed antidote (white circle) is also present.

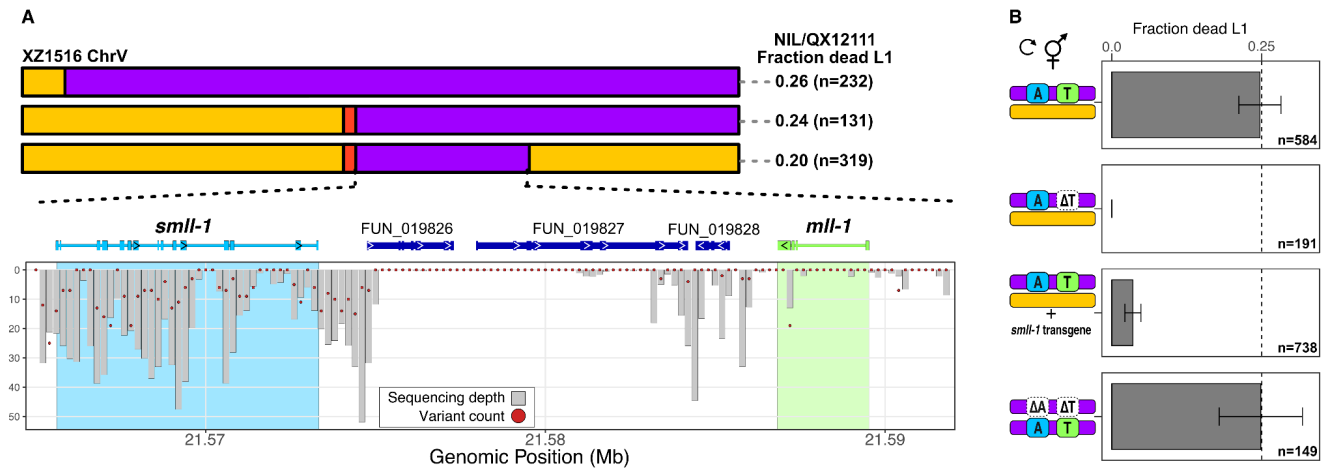


Figure 2. A) Localization of the TA element genes in XZ1516. Top panel: Strain genotypes of near-isogenic lines are displayed as colored rectangles (XZ1516 in purple; QX1211 in yellow; Cas9-induced deletion in red) for chromosome V. The fraction of L1 lethality after selfing of NIL/QX1211 hermaphrodites is shown to the right of each NIL. Bottom panel depicts a summary of QX1211 sequencing reads aligned to the XZ1516 genome. Gray bars denote coverage depth in 200 bp windows and red dots denote the number of variants detected between QX1211 and XZ1516 in each window. The toxin and antidote genes are highlighted in green and light blue, respectively. B) Knockout and transgenic rescue experiments define the TA components. Bar plots denote the fraction of dead L1s derived from selfing F1 heterozygous individuals. Error bars indicate 95% binomial confidence intervals calculated using the normal approximation method. Blue and green boxes with “A” and “T” indicate intact antidote and toxin genes, respectively; white boxes indicate deletions of these genes. XZ1516 genotypes are depicted in purple and QX1211 genotypes are depicted in yellow. Panels from top to bottom: XZ1516/QX1211 control cross; toxin knockout cross to QX1211; antidote transgenic rescue cross; toxin and antidote double knockout cross to XZ1516.

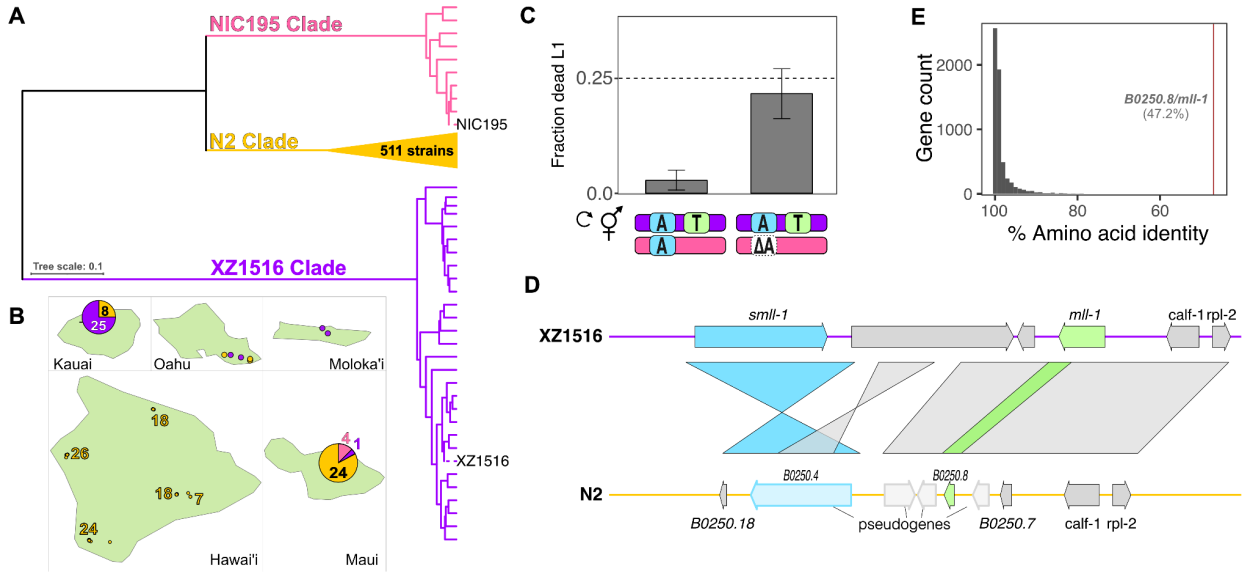
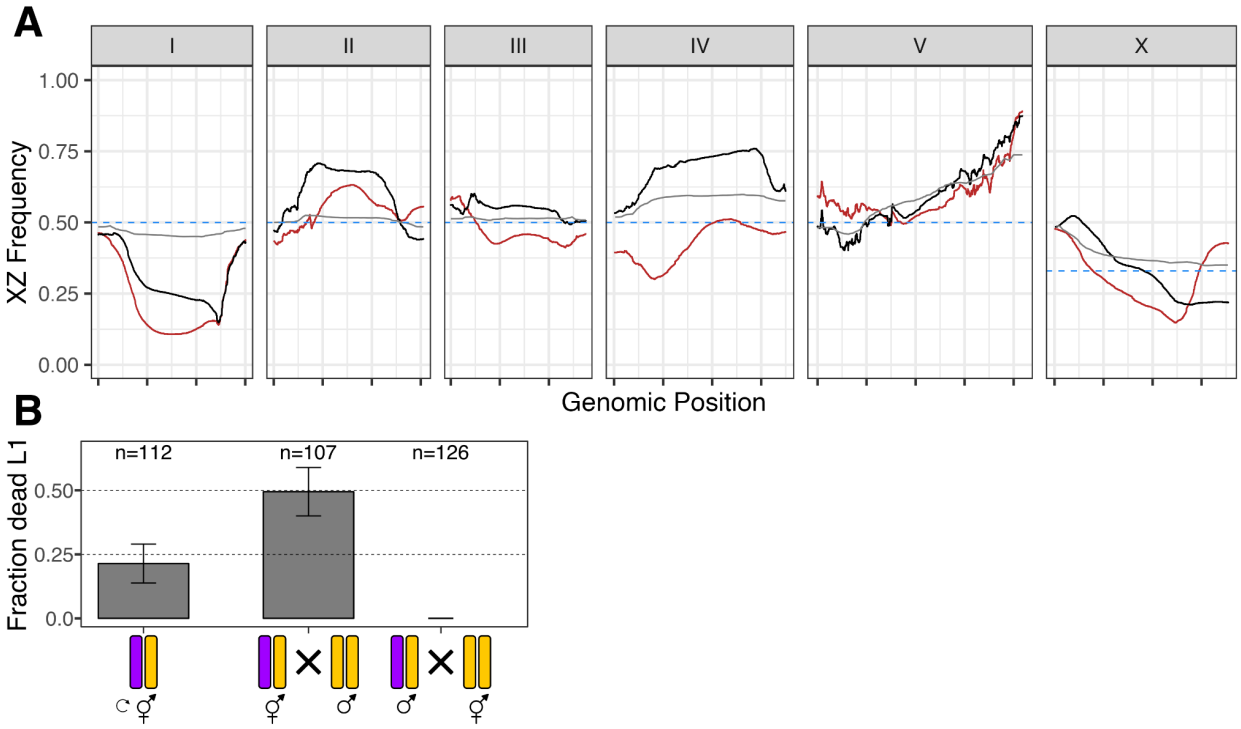
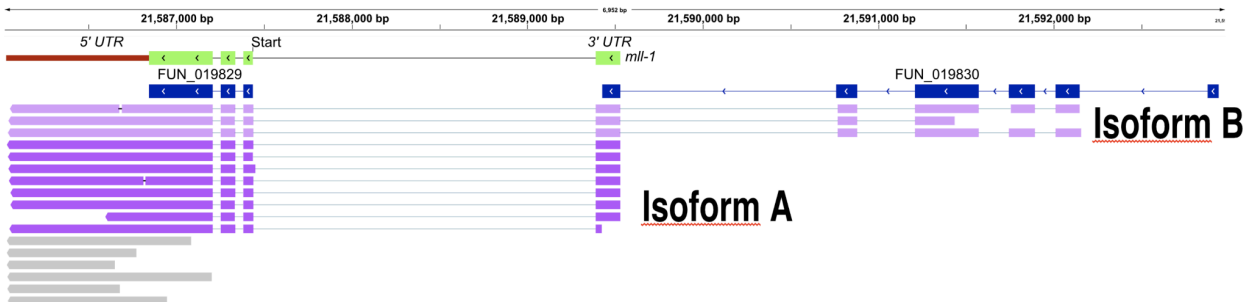


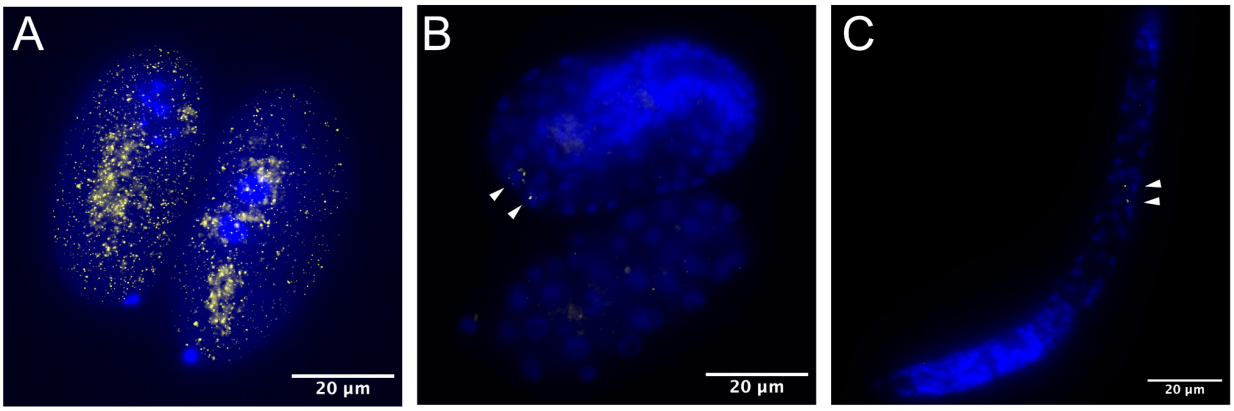
Figure 3. A) A dendrogram showing the relatedness of 550 wild *C. elegans* strains at the TA locus. Branches are colored to represent the three distinct clades, where purple denotes the XZ1516-like clade, yellow denotes the N2-like clade, and pink denotes the NIC195-like clade. B) Isolation location of strains collected in Hawaii. Pie charts show the number of isolates from each clade when multiple strains were collected at one location, with colors as in A. C) Bar plots show the fraction of dead L1s in crosses between XZ1516 and NIC195 (left) and between XZ1516 and NIC195 with its antidote allele knocked out (right), indicating that this antidote is active against the XZ1516 toxin. Error bars indicate 95% binomial confidence intervals calculated using the normal approximation method. D) Synteny plot of the TA region between the XZ1516 (top) and N2 (bottom) genomes. The TA components *mll-1* and *smll-1* are colored green and blue, respectively. E) Percent amino acid identity of ~5500 one-to-one orthologs identified between the XZ1516 and N2 genomes. Amino acid identity for *mll-1* is indicated with a red line.



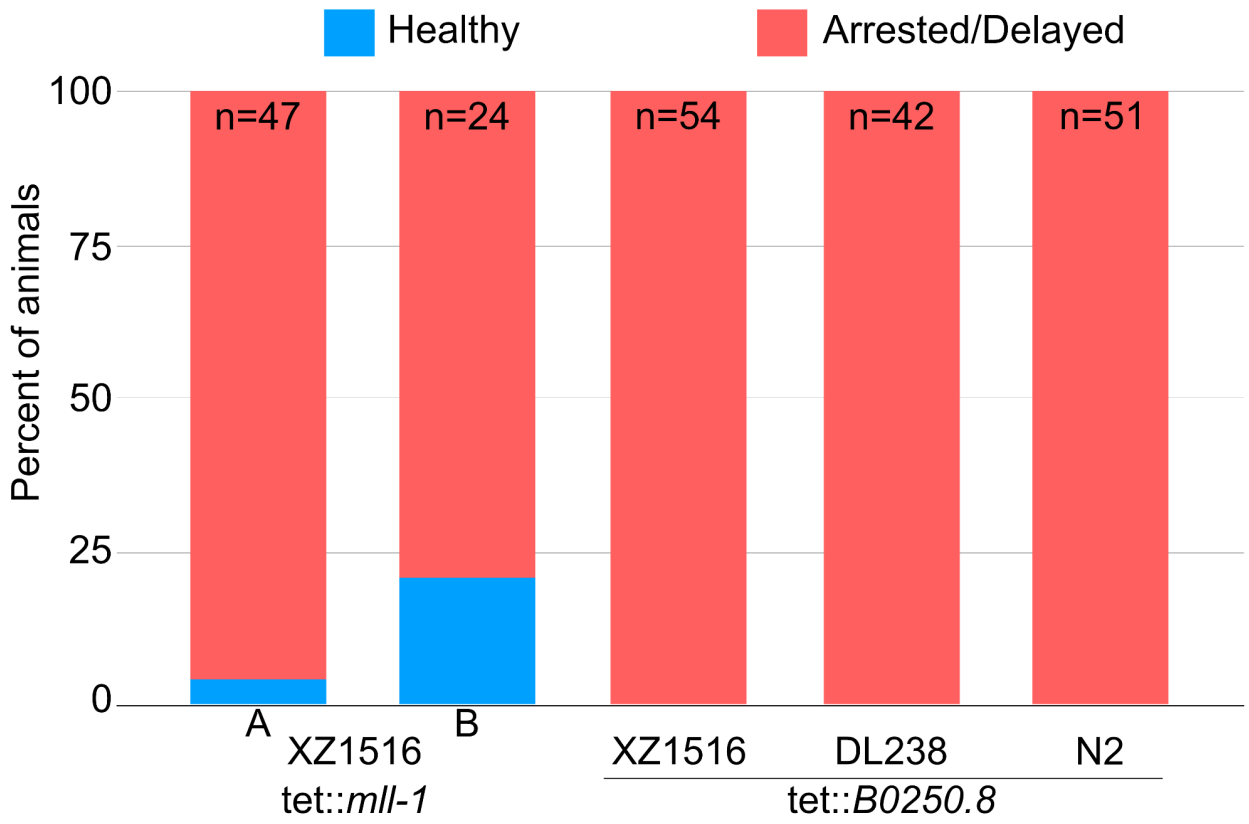
Supplemental Figure 1. A) Estimates of the XZ1516 allele frequency at each position in the genome after four (gray) or ten generations of intercrossing with QX1211 (black, gray) or DL238 (red). Each panel corresponds to a *C. elegans* chromosome and each x-axis tick represents 5 Mb. The dotted blue line represents the expected allele frequency for each chromosome. B) Crosses between XZ1516 (purple) and DL238 (yellow) that establish the inheritance pattern of the TA element. The y-axis represents the fraction of dead L1s observed for each cross depicted on the x-axis. Crosses from left to right: selfing of XZ1516/DL238 heterozygous; XZ1516/DL238 heterozygous hermaphrodites crossed to DL238 males; XZ1516/DL238 heterozygous males crossed to DL238 hermaphrodites.



Supplemental Figure 2. Long-read RNA sequencing reads of the different *mll-1* isoforms aligned to the XZ1516 genome visualized in the Integrative Genomics Viewer (IGV). Light purple reads represent isoform B, the dark purple reads represent isoform A, and gray reads represent the 5'UTR of *mll-1*. The functional gene model of *mll-1* is depicted in green above the computationally predicted "FUN" gene models.



Supplemental Figure 3. Representative *mll-1* FISH images in XZ1516. Yellow corresponds to *mll-1* transcript puncta and nuclei staining with DAPI is in blue. A) Max projection of 1- and 2-cell embryos. B) Post-gastrulation embryos showing *mll-1* localization to the Z2/Z3 cells. C) L1 larvae showing *mll-1* localization to the Z2/Z3 cells.



Supplemental Figure 4. Phenotypes induced by *mll-1* and *B0250.8*. The percent of healthy (blue) or arrested or delayed (red) individuals is shown on the y-axis. The first two bars represent technical replicates of tetracycline-induced *mll-1* in the XZ1516 genetic background. The remaining three bars correspond to tetracycline-induced *B0250.8* in XZ1516, DL238, and N2, respectively.

Supplemental tables:

Supplemental Table 1: Strains

Supplemental Table 2: Plasmids

Supplemental Table 3: Plasmid Construction

Supplemental Table 4: Crosses

Supplemental Table 5: Oligos

Supplemental Table 6: gRNAs and repair

Supplemental Table 7: tet-induced *mll-1* and *B0250.8* phenotypes

References:

1. Arvid Ågren J, Clark AG. Selfish genetic elements. PLoS Genet. 2018;14: e1007700. doi:10.1371/journal.pgen.1007700

2. Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. Nature. 1980;284: 604–607. doi:10.1038/284604a0

3. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. Nature. 1980;284: 601–603. doi:10.1038/284601a0

4. Stitzer MC, Khaipho-Burch MB, Hudson AI, Song B, Valdez-Franco JA, Ramstein G, et al. Transposable element abundance subtly contributes to lower fitness in maize. bioRxiv. 2023. p. 2023.09.18.557618. doi:10.1101/2023.09.18.557618

5. Valen L. A new evolutionary law. 1973;1: 1–30. Available: https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=PASCALGEODEBRGM7720012907

6. Brockhurst Michael A., Chapman Tracey, King Kayla C., Mank Judith E., Paterson Steve, Hurst Gregory D. D. Running with the Red Queen: the role of biotic conflicts in evolution. Proceedings of the Royal Society B: Biological Sciences. 2014;281: 20141382. doi:10.1098/rspb.2014.1382

7. Beeman RW, Friesen KS, Denell RE. Maternal-effect selfish genes in flour beetles. Science. 1992;256: 89–92. doi:10.1126/science.1566060

8. Ben-David E, Burga A, Kruglyak L. A maternal-effect selfish genetic element in Caenorhabditis elegans. Science. 2017;356: 1051–1055. doi:10.1126/science.aan0621

9. Seidel HS, Rockman MV, Kruglyak L. Widespread Genetic Incompatibility in \emphC. Elegans Maintained by Balancing Selection. Science. 2008;319: 589–594. doi:10.1126/science.1151107

10. Ben-David E, Pliota P, Widen SA, Koreshova A, Lemus-Vergara T, Verpukhovskiy P, et al. Ubiquitous Selfish Toxin-Antidote Elements in Caenorhabditis Species. Curr Biol. 2021;31: 990–1001.e5. doi:10.1016/j.cub.2020.12.013

11. Jurėnas D, Fraikin N, Goormaghtigh F, Van Melderen L. Biology and evolution of bacterial toxin–antitoxin systems. Nat Rev Microbiol. 2022;20: 335–350. doi:10.1038/s41579-021-00661-1

12. LeRoux M, Laub MT. Toxin-antitoxin systems as phage defense elements. Annu Rev Microbiol. 2022;76: 21–43. doi:10.1146/annurev-micro-020722-013730

13. Shultz AJ, Sackton TB. Immune genes are hotspots of shared positive selection across birds and mammals. Elife. 2019;8. doi:10.7554/eLife.41815

14. Daugherty MD, Malik HS. Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet. 2012;46: 677–700. doi:10.1146/annurev-genet-110711-155522

15. Seidel HS, Ailion M, Li J, van Oudenaarden A, Rockman MV, Kruglyak L. A novel

sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in C. elegans. PLoS Biol. 2011;9: e1001115. doi:10.1371/journal.pbio.1001115

16. Noble LM, Yuen J, Stevens L, Moya N, Persaud R, Moscatelli M, et al. Selfing is the safest sex for Caenorhabditis tropicalis. Elife. 2021;10. doi:10.7554/eLife.62587

17. Lee D, Zdraljevic S, Stevens L, Wang Y, Tanny RE, Crombie TA, et al. Balancing selection maintains hyper-divergent haplotypes in Caenorhabditis elegans. Nat Ecol Evol. 2021;5: 794–807. doi:10.1038/s41559-021-01435-x

18. Long L, Xu W, Valencia F, Paaby AB, McGrath PT. A toxin-antidote selfish element increases fitness of its host. Elife. 2023;12. doi:10.7554/eLife.81640

19. Burga A, Ben-David E, Lemus Vergara T, Boocock J, Kruglyak L. Fast genetic mapping of complex traits in C. elegans using millions of individuals in bulk. Nat Commun. 2019;10: 2680. doi:10.1038/s41467-019-10636-9

20. Rocheleau CE, Howard RM, Goldman AP, Volk ML, Girard LJ, Sundaram MV. A lin-45 raf enhancer screen identifies eor-1, eor-2 and unusual alleles of Ras pathway genes in Caenorhabditis elegans. Genetics. 2002;161: 121–131. doi:10.1093/genetics/161.1.121

21. Rockman MV, Kruglyak L. Recombinational landscape and population genomics of \emphCaenorhabditis elegans. PLoS Genet. 2009;5: e1000419. doi:10.1371/journal.pgen.1000419

22. Zdraljevic S, Walter-McNeill L, Marquez H, Kruglyak L. Heritable Cas9-induced nonhomologous recombination in C. elegans. microPublication Biology. 2023;2023. doi:10.17912/micropub.biology.000775

23. Robertson S, Lin R. Chapter One - The Maternal-to-Zygotic Transition in C. elegans. In: Lipshitz HD, editor. Current Topics in Developmental Biology. Academic Press; 2015. pp. 1–42. doi:10.1016/bs.ctdb.2015.06.001

24. Cook DE, Zdraljevic S, Roberts JP, Andersen EC. CeNDR, the Caenorhabditis elegans natural diversity resource. Nucleic Acids Res. 2017;45: D650–D657. doi:10.1093/nar/gkw893

25. Gillespie JH, Langley CH. Are evolutionary rates really variable? J Mol Evol. 1979;13: 27–34. doi:10.1007/BF01732751

26. Thomas CG, Wang W, Jovelin R, Ghosh R, Lomasko T, Trinh Q, et al. Full-genome evolutionary histories of selfing, splitting, and selection in Caenorhabditis. Genome Res. 2015;25: 667–678. doi:10.1101/gr.187237.114

27. Rogers AK, Phillips CM. A Small-RNA-Mediated Feedback Loop Maintains Proper Levels of 22G-RNAs in C. elegans. Cell Rep. 2020;33: 108279. doi:10.1016/j.celrep.2020.108279

28. Uebel CJ, Anderson DC, Mandarino LM, Manage KI, Aynaszyan S, Phillips CM. Distinct regions of the intrinsically disordered protein MUT-16 mediate assembly of a small RNA amplification complex and promote phase separation of Mutator foci. PLoS Genet. 2018;14: e1007542. doi:10.1371/journal.pgen.1007542

29. Phillips CM, Montgomery TA, Breen PC, Ruvkun G. MUT-16 promotes formation of

perinuclear mutator foci required for RNA silencing in the C. elegans germline. Genes Dev. 2012;26: 1433–1444. doi:10.1101/gad.193904.112

30. Reed KJ, Svendsen JM, Brown KC, Montgomery BE, Marks TN, Vijayasarathy T, et al. Widespread roles for piRNAs and WAGO-class siRNAs in shaping the germline transcriptome of Caenorhabditis elegans. Nucleic Acids Res. 2020;48: 1811–1827. doi:10.1093/nar/gkz1178

31. Rogers AK, Phillips CM. Disruption of the mutator complex triggers a low penetrance larval arrest phenotype. microPublication Biology. 2020 [cited 10 Jan 2024]. doi:10.17912/micropub.biology.000252

32. Makeyeva YV, Shirayama M, Mello CC. Cues from mRNA splicing prevent default Argonaute silencing in C. elegans. Dev Cell. 2021;56: 2636–2648.e4. doi:10.1016/j.devcel.2021.08.022

33. Tang W, Tu S, Lee H-C, Weng Z, Mello CC. The RNase PARN-1 Trims piRNA 3' Ends to Promote Transcriptome Surveillance in C. elegans. Cell. 2016;164: 974–984. doi:10.1016/j.cell.2016.02.008

34. Gu W, Shirayama M, Conte D Jr, Vasale J, Batista PJ, Claycomb JM, et al. Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. Mol Cell. 2009;36: 231–244. doi:10.1016/j.molcel.2009.09.020

35. Seroussi U, Lugowski A, Wadi L, Lao RX, Willis AR, Zhao W, et al. A comprehensive survey of C. elegans argonaute proteins reveals organism-wide gene regulatory networks and functions. Elife. 2023;12. doi:10.7554/eLife.83853

36. Nelson FK, Riddle DL. Functional study of the Caenorhabditis elegans secretory-excretory system using laser microsurgery. J Exp Zool. 1984;231: 45–56. doi:10.1002/jez.1402310107

37. Forrester WC, Garriga G. Genes necessary for C. elegans cell and growth cone migrations. Development. 1997;124: 1831–1843. doi:10.1242/dev.124.9.1831

38. Liégeois S, Benedetto A, Michaux G, Belliard G, Labouesse M. Genes required for osmoregulation and apical secretion in Caenorhabditis elegans. Genetics. 2007;175: 709–724. doi:10.1534/genetics.106.066035

39. Andersen EC, Bloom JS, Gerke JP, Kruglyak L. A variant in the neuropeptide receptor npr-1 is a major determinant of \emphCaenorhabditis elegans growth and physiology. 2014;10: e1004156. doi:10.1371/journal.pgen.1004156

40. Ben-David E, Boocock J, Guo L, Zdraljevic S, Bloom JS, Kruglyak L. Whole-organism eQTL mapping at cellular resolution with single-cell sequencing. Elife. 2021;10. doi:10.7554/eLife.65857

41. Bhagwat AM, Graumann J, Wiegandt R, Bentsen M, Welker J, Kuenne C, et al. multicrispr: gRNA design for prime editing and parallel targeting of thousands of targets. Life Sci Alliance. 2020;3. doi:10.26508/lsa.202000757

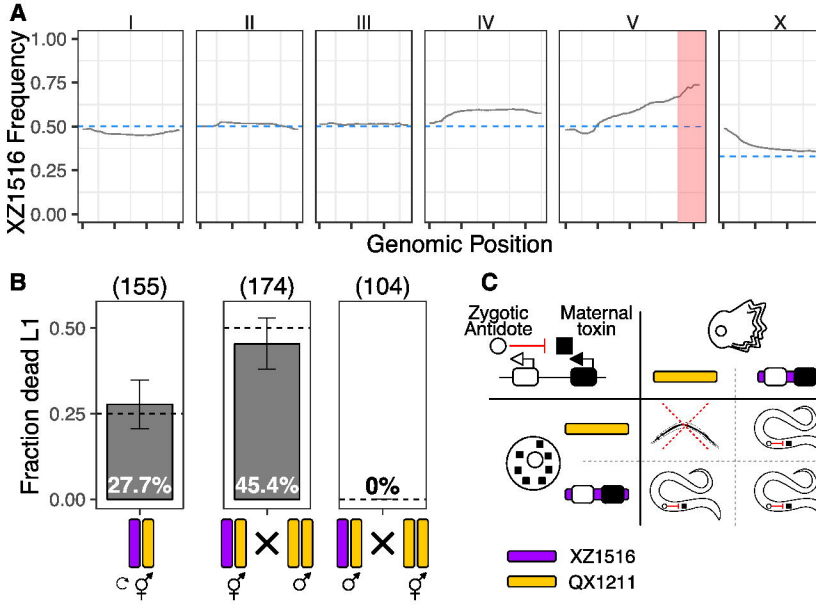
42. Core R. TEAM, 2017. R: A language and environment for statistical computing. R

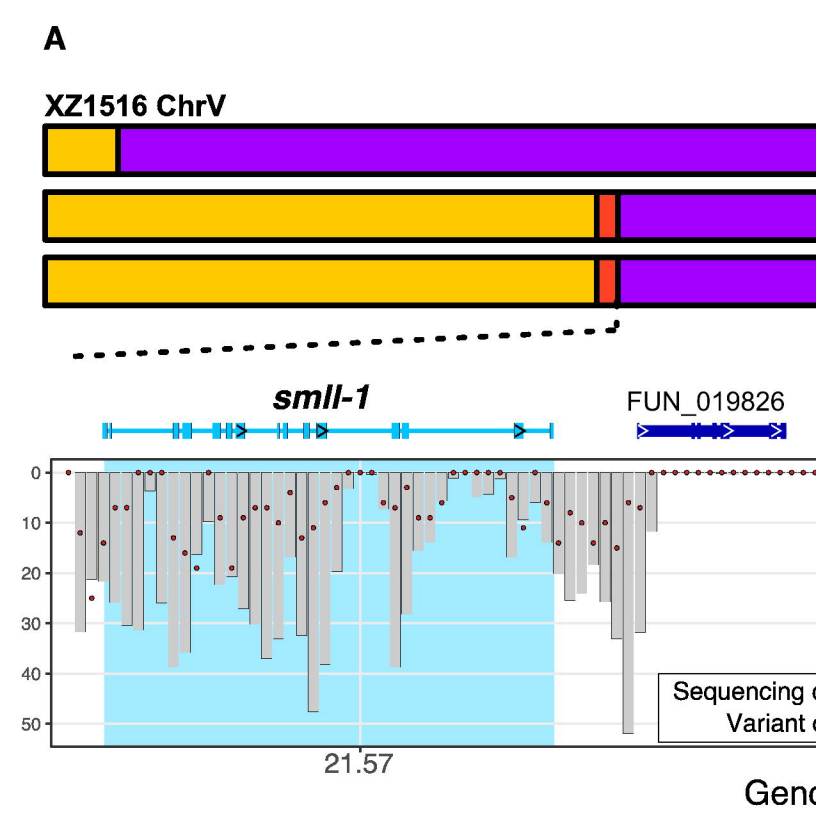
Foundation for Statistical Computing, Vienna, Austria. Online: https://www r-project org. 2022. Available: https://scholar.google.ca/scholar?cluster=10024430617971346517&hl=en&as_sdt=0,5&sciodt=0,5

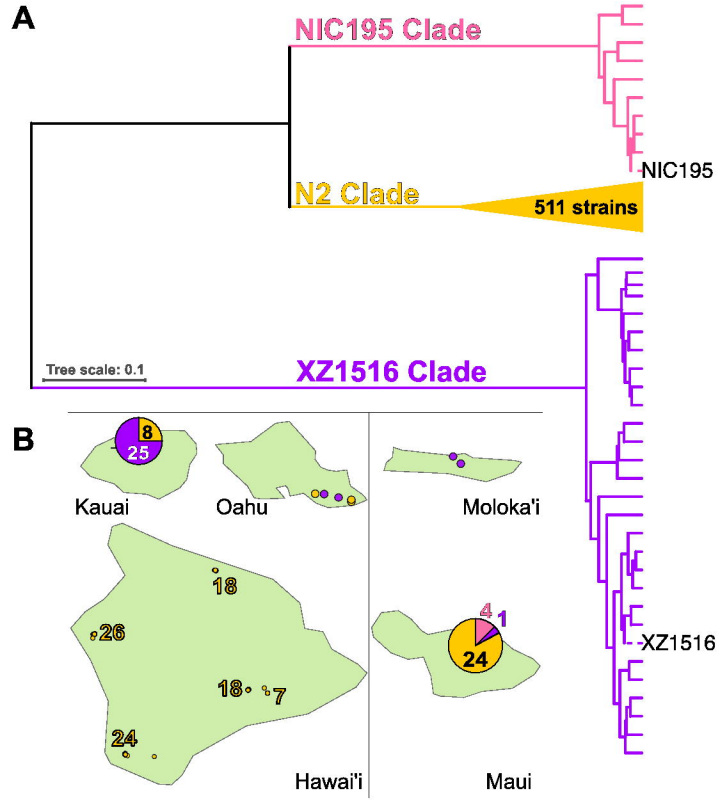
43. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34: 3094–3100. doi:10.1093/bioinformatics/bty191

44. Mao S, Qi Y, Zhu H, Huang X, Zou Y, Chi T. A Tet/Q Hybrid System for Robust and Versatile Control of Transgene Expression in C. elegans. iScience. 2019;11: 224–237. doi:10.1016/j.isci.2018.12.023

45. Drost H-G, Gabel A, Grosse I, Quint M. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. Mol Biol Evol. 2015;32: 1221–1231. doi:10.1093/molbev/msv012


A
I
II
III
IV
V
X
XZ1516 Frequency
1.00
0.75
0.50
0.25
0.00
Genomic Position
B
(155)
(174)
(104)
Fraction dead L1
0.50
0.25
0.00
27.7%
45.4%
0%
C
Zygotic Antidote
Maternal toxin
XZ1516
QX1211


A
XZ1516 ChrV
smll-1
FUN_019826
0
10
20
30
40
50
Sequencing
Variant
21.57
Genc


A
NIC195 Clade
N2 Clade
511 strains
NIC195
XZ1516 Clade
XZ1516
Tree scale: 0.1
B
8
25
Kauai
Oahu
Moloka'i
18
26
18
7
24
Hawai'i
4
1
24
Maui
C
Fraction dead L1
D
XZ15