GENETICS

Transposon-mediated genic rearrangements underlie variation in small RNA pathways

Gaotian Zhang[1]*, **Marie-Anne Félix**[1]*, **Erik C. Andersen**[2]*

Transposable elements (TEs) can alter host gene structure and expression, whereas host organisms develop mechanisms to repress TE activities. In the nematode *Caenorhabditis elegans*, a small interfering RNA pathway dependent on the helicase ERI-6/7 primarily silences retrotransposons and recent genes of likely viral origin. By studying gene expression variation among wild *C. elegans* strains, we found that structural variants and transposon remnants likely underlie expression variation in *eri-6/7* and the pathway targets. We further found that multiple insertions of the DNA transposons, *Polintons,* reshuffled the *eri-6/7* locus and induced inversion of *eri-6* in some wild strains. In the inverted configuration, gene function was previously shown to be repaired by unusual trans-splicing mediated by direct repeats. We identified that these direct repeats originated from terminal inverted repeats of *Polintons*. Our findings highlight the role of host-transposon interactions in driving rapid host genome diversification among natural populations and shed light on evolutionary novelty in genes and splicing mechanisms.

INTRODUCTION

Transposable elements (TEs) are ubiquitous mobile DNA sequences. With their parasite-like nature and the invasive mechanisms of transposition, these selfish genetic elements propagate in host genomes and cause diverse mutations, ranging from point mutations to genome rearrangements and expansions (*1*–*3*). They can even transfer horizontally across individuals and species, leading to movement of genetic material between widely diverged taxa (*4*, *5*). To the hosts, recent TE insertions are mostly deleterious. Various pathways have evolved in hosts to repress expression and transposition of TEs (*6*–*9*). By contrast, hosts can benefit from TEs because TE sequences can serve as building blocks for the emergence of protein-coding genes, noncoding RNAs, centromeres, and cis-regulatory elements (*10*–*12*).

Small RNAs are widely used to repress expression of TEs and other genes (*6*, *7*, *9*). In the nematode *Caenorhabditis elegans*, the helicase ERI-6/7–dependent small interfering RNAs (siRNAs) primarily target long terminal repeat (LTR) retrotransposons and pairs or groups of nonconserved genes and pseudogenes that show extensive homology and have likely viral origins (*9*, *13*). The closest known species of *C. elegans*, *Caenorhabditis inopinata*, lost the *eri-6/7*–related small RNA pathway, which was suggested to have caused the expansion of transposons in its genome compared to *C. elegans* and another related species, *Caenorhabditis briggsae* (*9*, *14*). In *C. elegans*, ERI-6/7 is required for the biogenesis of the Argonaute ERGO-1–associated endogenous siRNAs (Fig. 1A) (*13*). Likely because endogenous and exogenous siRNA pathways share and compete for downstream resources (*15*), mutants of *eri-6/7* display enhanced RNA interference (RNAi) responses to exogenous double-stranded RNAs (dsRNAs) (*16*). Competition also exists among different endogenous siRNA pathways. Within the *eri-6/7* locus, three other local open reading frames (*eri-6[e]*, *eri-6[f]*, and *sosi-1*) act independently of one another in a feedback loop to modulate the expression of ERI-6/7 and maintain a balance between different endogenous siRNAs (Fig. 1, A and B) (*17*).

In addition to the vital role of ERI-6/7 in RNAi pathways, its discovery (*16*) revealed a highly unusual expression mechanism. Fischer and Ruvkun (*16*) showed that *eri-6* and *eri-7*, two adjacent genes oriented in opposing genomic directions in the *C. elegans* reference strain N2, use a trans-splicing mechanism to generate fused *eri-6/7* mRNAs encoding the helicase ERI-6/7 (Fig. 1A). They further demonstrated that a direct repeat flanking *eri-6* facilitated the trans-splicing process (Fig. 1A). They also noticed variation of the locus within and between species: A single contiguous gene structure at the *eri-6/7* locus was found in some wild *C. elegans* strains and the *C. briggsae* reference strain AF16. However, the evolutionary history and consequence of the polymorphic variation remained unknown.

Expression quantitative trait loci (eQTL) are genomic loci that explain variation in gene expression across a species (*18*). We recently conducted a genome-wide eQTL analysis among 207 wild *C. elegans* strains using single-nucleotide variants (SNVs) as markers (fig. S1A) (*19*). Here, we show that the cis-acting eQTL of the *eri-6/7* locus is associated with a genomic hotspot enriched for trans-acting eQTL of nonconserved genes and pseudogenes, including known ERI-6/7–dependent siRNA targets. We identify structural variation within the *eri-6/7* locus, including a distinct gene structure and multiple TE remnants. Our results further demonstrate that the insertion of multiple copies of the virus-like DNA transposon, *Polinton* (*20*, *21*), might have caused gene inversion and fission of a single ancestral *eri-6-7* gene. Although some wild strains still have the single *eri-6-7* gene, other strains such as N2 evolved the *eri-6/7* trans-splicing mechanism to compensate for the *eri-6* inversion. The direct repeats used for trans-splicing originated from the terminal inverted repeats (TIRs) of *Polintons*. The neighboring putative genes *eri-6[e]*, *eri-6[f]*, and *sosi-1* are affected by other *Polinton*-induced structural variants and could have acquired their regulatory functions because of the inversions. Together, the *eri-6/7* gene structure polymorphisms and further structural variants at the locus impart sophisticated regulatory effects on the biogenesis of the ERI-6/7 helicase, downstream siRNAs, and the expression of their gene targets.

[1]Institut de Biologie de l'École Normale Supérieure, CNRS, INSERM, Paris, France.
[2]Biology Department, Johns Hopkins University, Baltimore, MD, USA.
*Corresponding author. Email: gzhang@bio.ens.psl.eu (G.Z.); felix@bio.ens.psl.eu (M.-A.F.); erik.andersen@gmail.com (E.C.A.)

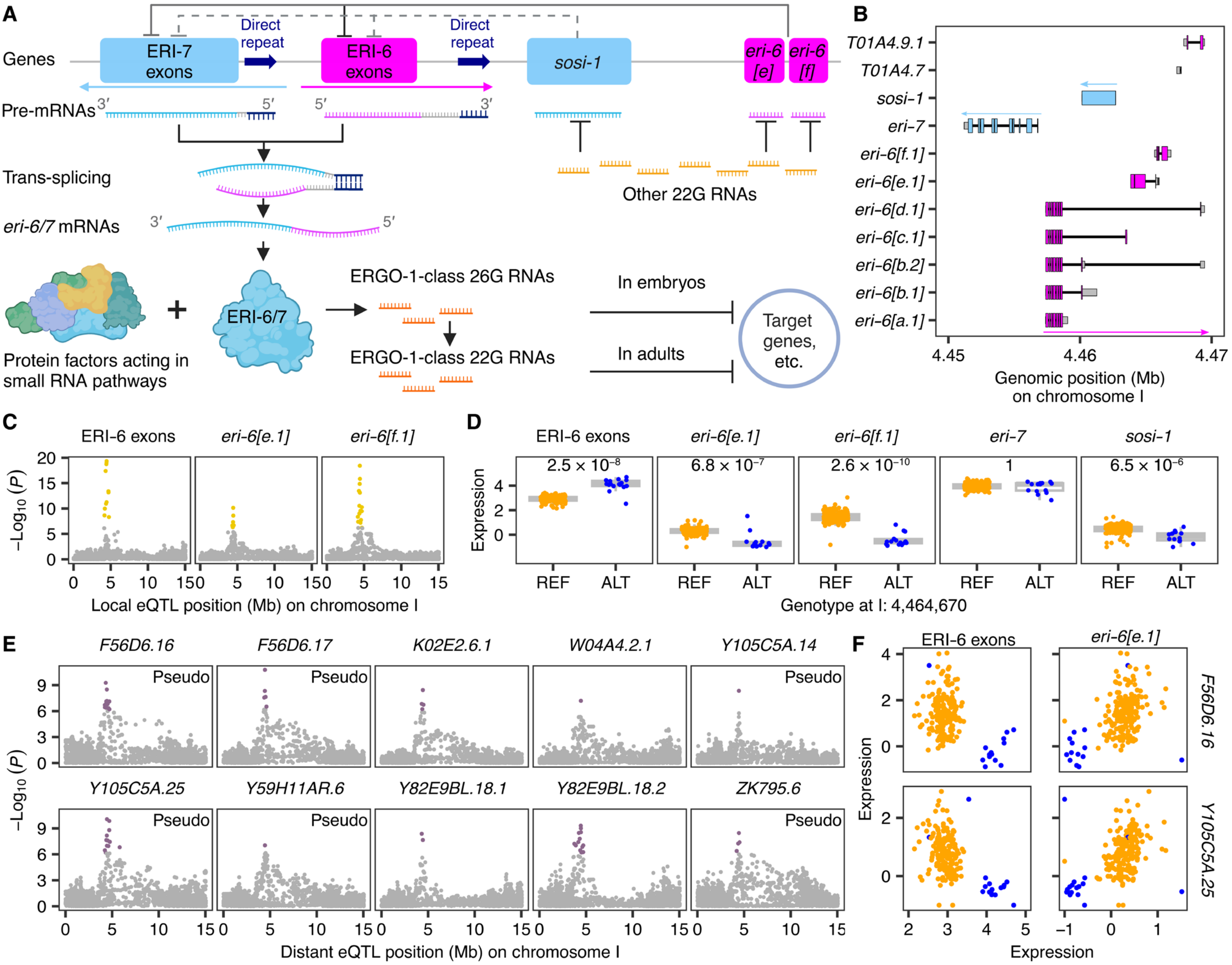

Fig. 1. Expression variation in *eri-6* potentially mediates a trans-acting eQTL hotspot. (**A**) Graphic illustration of the ERI-6/7–dependent siRNA pathways and the feedback loop. Dark blue arrows indicate direct repeats. Pink and blue rectangles indicate exons on the plus and minus strand, respectively (The same color scheme is used in the following figures). Created with BioRender.com. (**B**) Structures of genes and isoforms at the *eri-6/7* locus in the reference genome. (**C** and **E**) Manhattan plots indicating the GWAS mapping results of (C) transcript expression traits on chromosome I for ERI-6 exons, *eri-6[e]*, and *eri-6[f]* and (E) 10 transcripts across the genome. Each point represents a SNV that is plotted with its genomic position (*x* axis) against its $-\log_{10}$ (*P*) value (*y* axis) in mappings. SNVs that pass the 5% FDR threshold are colored gold and purple for local and distant eQTL, respectively. Transcripts of pseudogenes are indicated. (**D**) Tukey box plots showing expression [$-\log_2$ (normalized TPM + 0.5)] variation of five transcripts at the *eri-6/7* locus between 207 strains with different alleles at the top candidate SNV (I: 4,464,670). Statistical significance of each comparison is shown above and was calculated using the two-sided Wilcoxon test and was corrected for multiple comparisons using the Bonferroni method. (**F**) Correlations of expression variation of two transcripts to expression variation of ERI-6 exons and *eri-6[e]*. Each point [(D) and (F)] represents a strain and is colored orange and blue for strains with the reference (REF) or the alternative (ALT) allele at the SNV, respectively.

RESULTS

Natural variation in *eri-6* correlates with differential expression of small RNA targets

The genes *eri-6* and *eri-7* are next to each other in an opposite head-to-head orientation at 4.45 to 4.47 Mb on chromosome I in the N2 reference genome (WS283) (*22*) (Fig. 1B). The *eri-6* gene has had a changing transcript annotation in WormBase (*22*) because of a variety of rare splicing events. Presently, it includes six isoforms [*a-f*] that do not all share exons: *eri-6[a-d]* share their first seven exons (hereafter "ERI-6 exons," which encode the ERI-6 portion of ERI-6/7) and short downstream exons, some of them quite distant; *eri-6[e]* and *eri-6[f]* do not share ERI-6 exons but are transcribed from distinct downstream exons (Fig. 1B). Because the small downstream exons of *eri-6[a-d]* do not contribute many RNA sequencing (RNA-seq) reads, we used the combined expression of *eri-6[a-d]* as a proxy for the total expression of ERI-6 exons (fig. S1B). We investigated the genetic basis of expression variation (eQTL) for ERI-6 exons, *eri-6[e]*, *eri-6[f]*, and other protein-coding genes in *C. elegans* (see Materials and Methods). Here, we focused on eQTL related to the *eri-6/7* locus (tables S1 and S2).

We classified eQTL into local and distant eQTL based on the location of the QTL in the genome relative to its expression targets (fig. S1A and table S2) (*19*). At the threshold used (see Materials and Methods), we detected local eQTL for expression variation in ERI-6 exons, *eri-6[e]* and *eri-6[f]* (Fig. 1C and table S2). Computational fine mappings of these local eQTL identified the top candidate variant (I: 4,464,670) (fig. S2 and table S3), a missense mutation (D259Y) in the coding region of *eri-6[e]*. Strains with the alternative allele at this site showed significantly lower *eri-6[e]* and *eri-6[f]* expression than strains with the reference allele but higher expression in ERI-6 exons (Fig. 1D).

Expression variation in ERI-6 exons could further affect the production of the ERI-6/7 helicase, the biogenesis of siRNAs in the ERGO-1 pathway, and lastly the expression of target genes (Fig. 1A). We found that 12 genes across the genome, including 6 and 11 known targets of ERI-6/7 (*13*) and ERGO-1 (*23*), respectively, have their distant eQTL (I: 4.3 to 4.7 Mb) located nearby the *eri-6/7* locus (Fig. 1E and tables S2 and S4). Computational fine mappings of these distant eQTL also identified the I: 4,464,670 *eri-6[e]* variant as the top candidate (fig. S3 and table S3). These transcripts showed significantly lower expression in strains with the alternative allele than strains with the reference allele (fig. S4). Their expression also exhibited negative correlations with ERI-6 exons but positive correlations with *eri-6[e]* expression (Fig. 1F). As mentioned above, pseudogenes and nonconserved genes are among the primary targets of the ERI-6/7–dependent siRNAs (*9*, *13*). Nine of 12 genes are pseudogenes, and seven of them lack known orthologs in other species (table S4) (*22*). Together, all of these 12 genes are potential targets of ERI-6/7–dependent siRNAs. Genetic variation at the *eri-6/7* locus functions as a potential trans-acting hotspot to regulate expression of target genes across the genome using the siRNA pathways.

In addition to the top variant, our candidate prioritization identified a second candidate variant (I: 4,464,857, R321Q), which is likely a more conservative amino acid substitution than the top candidate (I: 4,464,670) (see Materials and Methods), for local eQTL of *eri-6* and distant eQTL of the 12 genes above. The two top candidate variants are in perfect linkage disequilibrium (LD) among the 207 wild strains used in the eQTL mapping. We used CRISPR-Cas9 genome editing to generate single and double mutants for the two candidates in different genetic backgrounds (see Materials and Methods and table S5) and showed that the two variants did not underlie the local eQTL of *eri-6* (fig. S5) nor the distant eQTL of potential targets.

Two of the strains (CB4856 and MY18) in our expression dataset with an alternative allele at the *eri-6[e]* variants were previously found to have *eri-6* and *eri-7* on the same (Crick) strand, similar to the *eri-7* ortholog in the reference genomes of the species *C. briggsae* and *Caenorhabditis brenneri* (Fig. 2) (*16*, *22*). We thus focused on structural variants, which were not included in the eQTL mapping because of the difficulty in characterizing them. We first studied them at the genomic level to uncover the diversity of structural variants, then found their transposon origin, and lastly demonstrated the association of these structural polymorphisms with a diversity of gene expression phenotypes.

High diversity of structural variants and TE insertions throughout the *eri-6/7* locus

Long-read genome assemblies of 17 wild *C. elegans* strains are presently available (*24*–*27*), in addition to the reference strain N2. We first performed a multiple pairwise alignment of the *eri-6/7* region among these strains (Fig. 2 and fig. S6A) (*24*–*28*). Nine of the 17 strains are approximately identical to the reference strain N2 in this region, with *eri-6* on the Watson strand (pink in figures) and *eri-7* on the Crick strand (blue in figures). Hereafter, the first seven exons of *eri-6* in the N2 reference orientation are called "Watson ERI-6 exons." The strain JU1400 has a 2.8-kb duplication that includes the Watson ERI-6 exons and one copy of the direct repeats that flank ERI-6 exons (Fig. 2).

The other seven strains harbor a large diversity of deletions, insertions, and inversions compared to the reference genome. The two strains ECA396 and JU2526 have a largely inverted *sosi-1* gene compared to the N2 strain, two different *sosi-1* fragments, and several other insertions (Fig. 2 and figs. S6A and S7A). The remaining five strains show inversion of ERI-6 exons compared to the N2 strain (hereafter "Crick ERI-6 exons" when in the same orientation as *eri-7*): The strains XZ1516, ECA36, and NIC526 also lack the direct repeats that flank ERI-6 exons and include a ~1.7-kb insertion between their Crick ERI-6 exons and *sosi-1*; the strains CB4856 and DL238 have retained most of the direct repeat sequences and show multiple large insertions with sizes up to ~8 kb within *eri-7* and surrounding the Crick ERI-6 exons (Fig. 2 and fig. S6A). The Crick orientation of the ERI-6 exons in these five strains likely represents the ancestral genetic structure at the *eri-6/7* locus based on the following: (i) *eri-6-7* orthologs in *C. briggsae*, *C. brenneri*, and at least another eight *Caenorhabditis* species show a simple continuous structure on a single strand (Fig. 2 and table S6); (ii) the XZ1516, ECA36, CB4856, and DL238 strains were found to have patterns of ancestral genetic diversity in the *C. elegans* species (fig. S8) (*29*–*31*).

This structural diversity corresponds to an astonishing diversity of polymorphic TEs within the 18 kb locus (Fig. 2 and fig. S6A). First, a 435–base pair (bp) fragment of *CELETC2* (a nonautonomous *Tc2*-related DNA transposon) (*22*) resides in the ~1.7-kb insertion on the right of Crick ERI-6 exons in the strains XZ1516, ECA36, and NIC526. Second, two different fragments (354 and 299 bp) of the unclassified transposon *Ce000179* (*22*) constitute most of the 838-bp insertion within *eri-7* in the strains CB4856, DL238, and ECA396. Third, a full-length *CEREP1A* (a putative nonautonomous 3.4-kb DNA transposon likely using *HAT*-related transposase for propagation) (*22*) was found in both the CB4856 and DL238 strains, and the CB4856 strain has two other *CEREP1A* fragments immediately upstream in the opposite orientation. Fourth, the strain ECA396 has a full-length *Tc4v* (a variant class of the DNA transposon *Tc4*) (*22*, *32*) within the first exon of *eri-6[f]*. Last, we found multiple TE insertions from a family of autonomous double-stranded DNA transposons derived from viruses, called *Polintons* (*20*, *21*). Four different sizes of *Polinton* remnants were identified at this locus in the strains CB4856, DL238, ECA396, and JU2526 (Fig. 2 and fig. S7B).

The direct repeats allowing *eri-6/7* trans-splicing originate from *Polintons*

Polintons (a.k.a. *Mavericks*) were identified across unicellular and multicellular eukaryotes and proposed to transpose through protein-primed self-synthesis (*5*, *20*, *33*). They code numerous proteins, including two core components, a protein-primed DNA polymerase B (pPolB1) and a retroviral-like integrase (INT), and different capsid proteins (*20*, *21*). The different *Polinton* remnants that we found at the *eri-6/7* locus in wild strains are all likely from the pPolB1 end of the

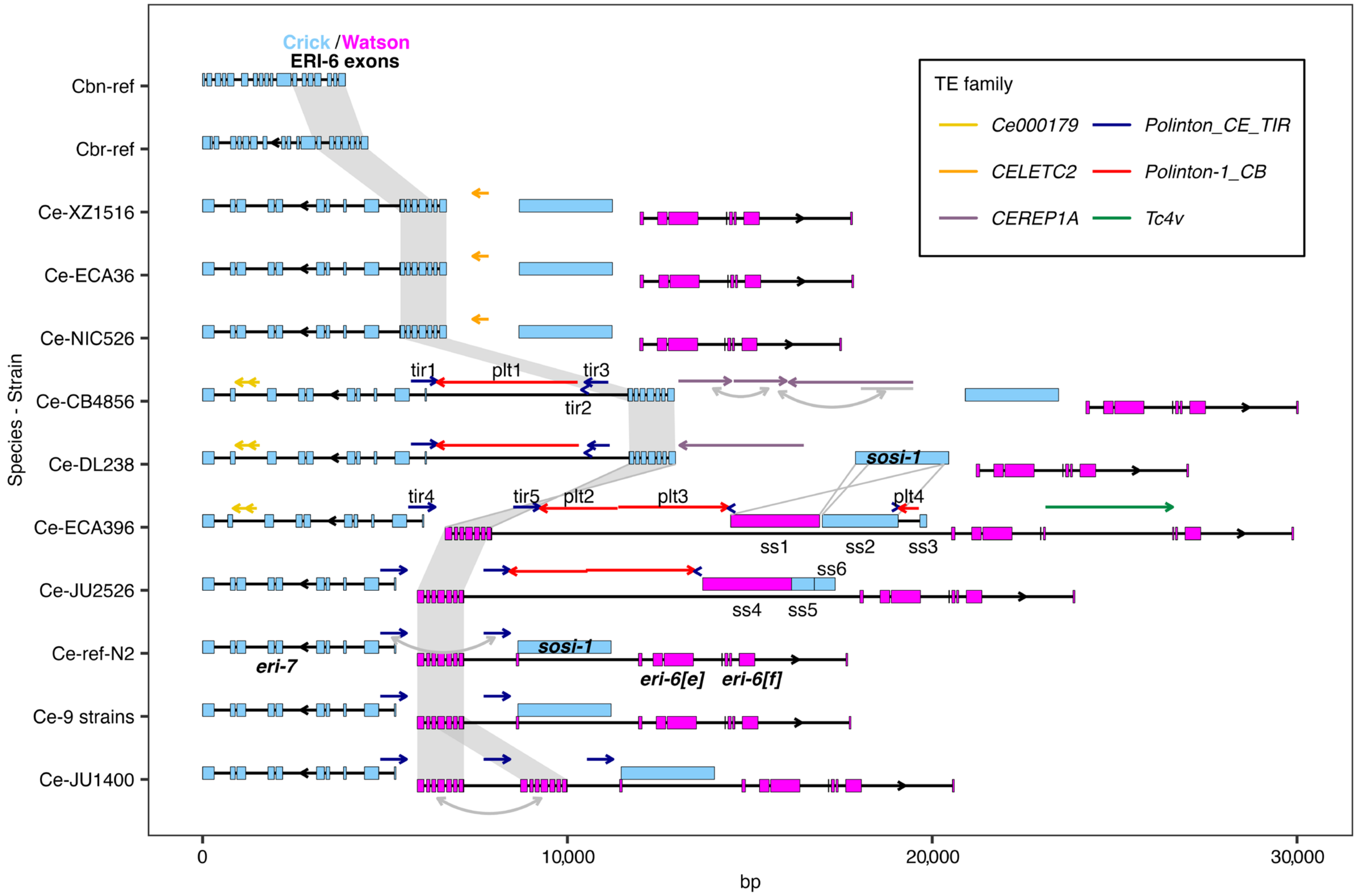

Fig. 2. Hypervariable structural variants and TEs at the *eri-6/7* locus. Graphic illustration of DNA sequence alignment at the *eri-6/7* locus in the 18 *C. elegans* (Ce) strains with genome assemblies. The gene structures of the *C. briggsae* reference (Cbr-ref) *eri-7* and its best match homolog in *C. brenneri* reference (Cbn-ref) are shown on top. The exon structures of the *C. elegans* strains are shown based on the reference N2 genome. Regions with a potential transposon origin are indicated as colored single-headed arrows, with the color indicating the family of transposon and the arrow direction representing their potential coding orientation when inserted. Double-headed arrows indicate duplications. ERI-6 exons are shaded gray. Detailed alignment to the reference of regions with labels "tir1-5" (for TIRs), "plt1-4" (for *Polintons*), and "ss1-6" (for *sosi-1*) are shown in fig. S7.

Polinton-1_CB (named after the *Polintons* in *C. briggsae*) (fig. S7B) (*22*). In the reference genome of *C. elegans*, three partial copies of *Polinton-1_CB* have been identified at 10.30 to 10.32 Mb (WBTransposon00000738) and 13.08 to 13.10 Mb (WBTransposon00000637) on chromosome I and at 17.25 to 17.27 Mb (WBTransposon00000739) on chromosome X, with lengths ranging from 13.4 to 15.4 kb (*22*). We found 744-bp inverted repeats perfectly flanking WBTransposon00000738 (fig. S7B and table S7) and partially flanking the other two *Polintons* in the genome of the reference strain N2. We hypothesized that these inverted repeats were specific TIRs of *Polintons* in *C. elegans*. They were previously not regarded as *Polintons* because *C. briggsae Polinton* consensus sequences were used to identify *Polintons* in *C. elegans*. To examine the validity and species specificity of the TIRs, we first identified potential *Polintons* by searching colocalization (within 20 kb) of pPolB1 and INT in the genomes of 18 *C. elegans* and 3 *C. briggsae* strains (fig. S9). We identified three to nine potential *Polintons* in each *C. elegans* strain and 13 to 15 in each *C. briggsae* strain. Complete or partial sequences of the 744-bp TIRs were flanking 63 of the total 107 *Polintons* in the 18 *C. elegans* strains but none in the three *C. briggsae* strains (fig. S9). We also found colocalization of pPolB1 and the TIR but not INT at 10 loci, including but not limited to the *eri-6/7* locus in *C. elegans* genomes of both N2-like strains and the divergent strains (fig. S9A). Furthermore, all significant National Center for Biotechnology Information (NCBI) BLAST (*34*) results in the query of the TIR sequence are from *C. elegans*. Together, the 744-bp TIRs are components of *Polintons* specifically in *C. elegans*, termed *Polinton_CE_TIR*. We distinguish them from the annotated *Caenorhabditis Polinton-1_CB*.

The *Polinton_CE_TIR* sequences are present as direct repeats instead of inverted repeats exclusively at the *eri-6/7* locus in the reference N2, the nine N2-like strains, JU1400, JU2526, and ECA396 (Fig. 2 and fig. S9A). Approximately 700 bp of the ~930-bp direct repeats that facilitate trans-splicing are exactly *Polinton_CE_TIR* (fig. S6B and table S7). The repeat sequences also include additional putative transcription factor binding sites for transcriptional regulation (fig. S6C). Therefore, strains such as the reference N2 use components of *Polintons* to compensate for the disruptive gene inversion that was likely caused by the *Polintons* themselves.

Multiple *Polinton* copies likely mediated inversions and other structural rearrangements

To evaluate the diversity of this locus using a larger set of strains, we obtained short-read whole-genome sequencing (WGS) data of 550 isotype strains, aligned to the reference N2, representing 1384 wild strains from the *Caenorhabditis* Natural Diversity Resource (CaeNDR, 20220216 release) (*35*, *36*). We aimed to detect inversions and other structural variants in the species using information of split reads and mapping coverages (see Materials and Methods) and relate them to the SNV haplotypes in the region.

We identified diverse structural variants within the *eri-6/7* locus among the 550 wild strains (figs. S6D and S10 and table S8): (i) inversions, 93 strains have Crick ERI-6 exons and 34 strains have partial inversions of *sosi-1* (INV*sosi-1*) (fig. S10A); (ii) *Polinton* insertions, 48 strains likely have partial remnants of the pPolB1 end of the *Polinton-1_CB* (fig. S10A); (iii) lack of reference genes (which might result from deletion or maybe an ancestral lack of insertion), 14 strains lack the reference *sosi-1*, *eri-6[e]*, and *eri-6[f]*, whereas two strains only lack *eri-6[e]* and *eri-6[f]* (fig. S10, B and D); (iv) deletions, 13 strains showed a ~250-bp deletion mostly spanning the 3’UTR of *eri-6[f]*; (v) duplications, the strain JU1896 might have duplications of *eri-6[e]* and *eri-6[f]*; (vi) high heterozygosity in *sosi-1*, 80 strains with the reference *sosi-1* might have a second copy of *sosi-1* beyond the locus, which was also had by three of the 14 strains lacking the reference *sosi-1* (table S8).

The short-read data are limited in their ability to detect the full extent of structural variants. However, we observed *Polintons* (*Polinton_CE_TIR* and *Polinton-1_CB*) at multiple sites throughout the *eri-6/7* locus (fig. S10A), especially at flanking regions of ERI-6 exons and *sosi-1*. TEs have been associated with chromosomal rearrangements since their first discoveries (*1*). Ectopic recombination between TE copies or alternative transposition mechanisms could cause structural variants such as inversions, duplications, or deletions (*2*). We reasoned that the inversions of ERI-6 exons and *sosi-1* were possibly induced by homologous recombination between the flanking *Polintons* or simply the TIRs.

To understand the evolutionary relationships of the 550 strains at *eri-6/7* and group them, we performed a haplotype network analysis using the 95 SNVs within the locus (Fig. 3). We observed and defined two major groups, “Single *eri-6-7*” and “Reverse-oriented *eri-6,7*,” with 112 and 438 strains, respectively (Fig. 3). As expected, a Crick orientation of ERI-6 exons was detected for all strains in the Single *eri-6-7* group, except 17 strains that were clustered with CB4856 and DL238, using short-read WGS data. We hypothesized that all these 17 strains also have the ancestral Crick orientation of ERI-6 exons but with large *Polinton* remnants between them and *eri-7*: We defined them as “CB4856-like” strains together with the strains DL238 and ECA1186 (Fig. 3). The Reverse-oriented *eri-6,7* group of strains includes the reference strain N2 and likely all have Watson ERI-6 exons and the direct repeats for trans-splicing (figs. S6D and S10, B and C, and table S8). Most strains in this group are clustered with N2, whereas the strain ECA396 and 19 other strains formed a second cluster based on SNVs and likely all have INV*sosi-1* (Fig. 3). Remnants of *Polinton-1_CB* were found in both groups but mostly in CB4856-like strains and strains with INV*sosi-1 (Fig. 3*). Strains with deletion polymorphisms in *eri-6[e]*, *eri-6[f]*, and *sosi-1* formed two clusters exclusively in the Single *eri-6-7* group (Fig. 3). It is challenging to associate these structural variants with *Polintons* or other TEs. Nevertheless, these deletions and duplications might also affect expression of *eri-6/7* and siRNA pathways.

Cis- and trans-effects of *Polinton*-induced structural variants on gene expression

Among the 550 wild *C. elegans* strains, ~20% likely have a single “classical” *eri-6-7* gene to encode the ERI-6/7 protein, as in *C. briggsae* and *C. brenneri*. The remaining ~80% strains make a fused *eri-6/7* mRNA by some amount of trans-splicing between the pre-mRNAs of the Watson ERI-6 exons and *eri-7* as in the reference strain N2. Although trans-splicing compensates inversion of ERI-6 exons to continue ERI-6/7 production, previous work (*16*) could not consider whether the reverse-oriented *eri-6/7* gene structure might represent a hypomorphic form of the locus compared to the ancestral, compact gene. We thus turned our focus back to gene expression consequences of structural variants, which could affect expression at two levels: the expression abundances of different exons and their splicing.

We first examined local regulatory effects at the *eri-6/7* and *sosi-1* locus, starting with diversity among strains having the Crick ERI-6 exons and *eri-7*. The strains with a potential compact *eri-6-7* gene (green box color in Fig. 4A) expressed both parts of the gene at similar levels, as expected, and expressed low levels of *eri-6[e]*, *eri-6[f]*, and *sosi-1*. The exception in this group is the two strains ECA703 and ECA812, which do not have *eri-6[e]*, *eri-6[f]*, and *sosi-1* and showed low expression in ERI-6 exons and ref-*eri-7* (mRNA sequences of *eri-6[a-d]* and *eri-7* in the N2 reference, respectively) (Figs. 3 and 4A and table S8). Mutants of *sosi-1*Δ, *eri-6[e-f]*Δ, and *sosi-1 eri-6[e-f]*Δ in the background of the reference strain N2 were previously found to show reduced expression in ERI-6 exons and ref-*eri-7* compared to wild-type animals (*17*). The explanation for these observations is unknown. In addition, the strain JU1896, which likely has a duplication in *eri-6[e]* and *eri-6[f]*, showed higher expression in both (Figs. 3 and 4A and fig. S10D). The subgroup of CB4856-like strains (blue color), with large *Polinton* remnants between ERI-6 exons and the downstream ERI-7 exons (Fig. 2), exhibited significantly elevated expression in ERI-6 exons and significantly decreased expression in ref-*eri-7* (Fig. 4A and table S9): The large intronic insertion likely affects transcription of the downstream exons, i.e., *eri-7*.

The second large group of Reverse-oriented *eri-6,7* strains (orange and purple colors) showed significantly lower expression in ERI-6 exons and significantly higher expression in *eri-6[e]*, *eri-6[f]*, and *sosi-1* than strains in the Single *eri-6-7* group (Fig. 4A and table S9). The lower expression ERI-6 exons might be the result of either enhancer/promoter rearrangement or deficiencies in splicing or polyadenylate tail formation, making the mRNA less stable. By contrast, these strains exhibited a similar level of expression of the ref-*eri-7* to the Single *eri-6-7* group. The subgroup of strains with INV*sosi-1* (purple color) showed significantly lower expression in both *sosi-1* and ERI-6 exons than other strains in the Reverse-oriented *eri-6,7* group. Those strains with genome assemblies show *Polinton* insertions upstream of and within *sosi-1* as well as partial inversions and deletions of *sosi-1* (Fig. 2 and fig. S7A), which could explain the lower expression of *sosi-1*. In summary, the diverse structural variations correlate with their expected effect on the *eri-6/7* locus between and within the two large structural variant groups.

Different splicing mechanisms between the two groups further alter the efficiency of the ERI-6/7-dependent siRNA pathways. In

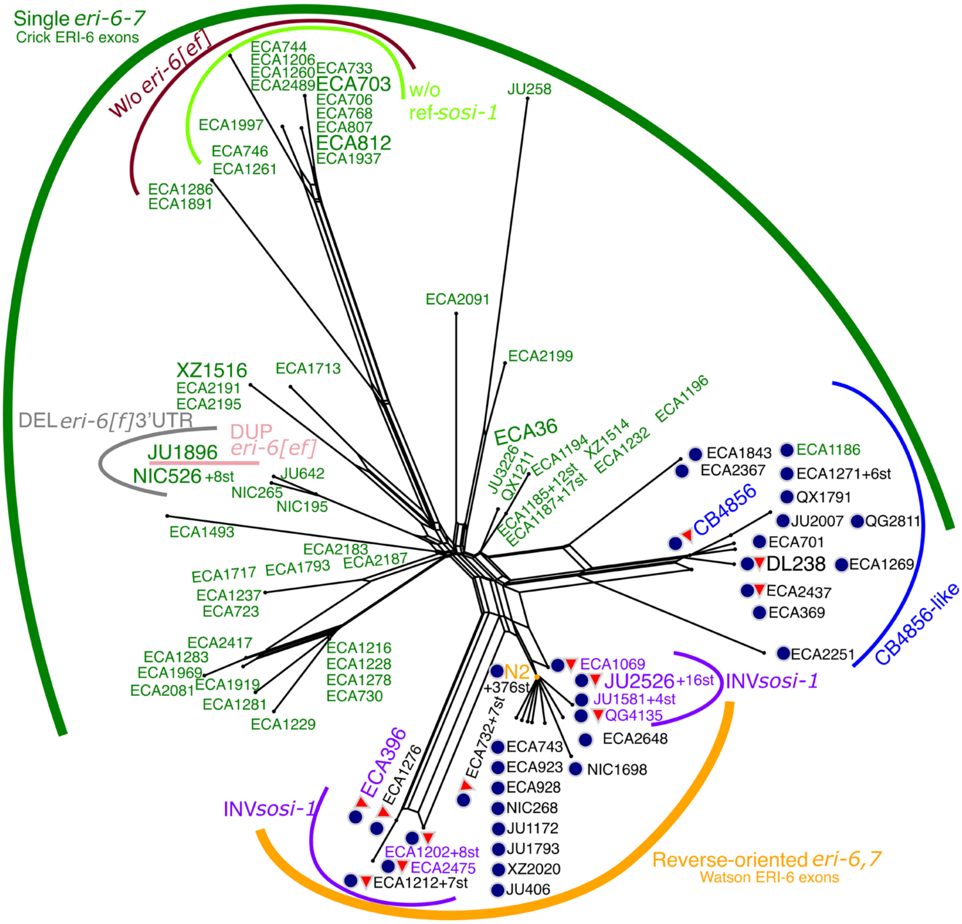

Fig. 3. Haplotype network with clustered strains sharing structural variation. Neighbor-joining net depicting 550 strains based on 95 SNVs within the *eri-6/7* locus. Two major groups, Single *eri-6-7* and Reverse-oriented *eri-6,7*, were defined on the basis of orientation of ERI-6 exons and denoted with dark green and orange curves. Subgroups with other structural variations were indicated using thin curves and labels ("w/o" for deletions or no insertions, "DEL" for deletions, and "DUP" for duplications). Strain names are colored in green and purple for detection of Crick ERI-6 exons and inversion of *sosi-1* (INV*sosi-1*), respectively, using short-read WGS data in fig. S10A. Dark blue circles and red triangles next to strain names represent strains with *Polinton_CE_TIR* (TIRs only) and *Polinton-1_CB* (TIRs excluded) insertions, respectively, based on figs. S6D and S10 and manual inspection of genome alignments. Some strains (st) share all alleles of the 95 SNVs, and all detected structural variations are collapsed to only show a representative strain followed by the number of strains with this *eri-6/7* haplotype (e.g., "N2 + 376st"). Trapezoidal junctions indicate that some recombination occurred within the locus.

strains with a single *eri-6-7* gene, the ERI-6/7 protein is produced through standard transcription and translation. In contrast, strains with reverse-oriented *eri-6,7* perform separate transcription of pre-mRNAs in opposite orientation and trans-splicing (*16*), which could reduce the efficiency of ERI-6/7 production. We analyzed spanning reads in the RNA-seq data of 207 strains to compare their splicing efficiency between the seventh exon of *eri-6* and the first of *eri-7* (Fig. 4B). In strains with a single *eri-6-7* gene, most split RNA-seq reads at the end of the Crick ERI-6 exons should have their chimeric alignment to ERI-7 exons through cis-splicing. In strains with the Watson ERI-6 exons, split RNA-seq reads at the end of ERI-6 exons could splice to downstream exons for *eri-6[b-d]* or partially map to ERI-7 exons because of trans-spliced *eri-6/7* mRNAs (*16*) (Fig. 4B). Among the 207 strains in our RNA-seq dataset, all 16 strains with a single *eri-6-7* gene showed higher than 90% and mostly 100% splicing between ERI-6 and ERI-7 exons. Instead, the 183 strains with reverse-oriented *eri-6,7* but not INV*sosi-1* showed a median of 10% and a maximum of 32% trans-splicing (Fig. 4B). In conclusion, the evolutionary inversion of *eri-6* does affect the synthesis of full-length *eri-6/7* mRNA.

Together, the expression level of ERI-6 and ERI-7 exons and their splicing rate alter the biogenesis of the helicase ERI-6/7. Strains with a single *eri-6-7* gene but no extra insertions or deletions might generate the most abundant ERI-6/7 because of their high expression in ERI-6/7 exons and mostly 100% cis-splicing (Fig. 4, A and B). The reverse-oriented *eri-6/7* gene structure represents a hypomorphic form of the locus because strains in this group showed reduced expression of ERI-6 exons and low splicing rate between ERI-6/7 exons (Fig. 4, A and B), which likely led to reduced abundance of ERI-6/7 protein.

The structural variants showed various local effects on gene expression, but their influences likely extended beyond the locus because of the pivotal role of ERI-6/7 in *C. elegans* endogenous siRNA pathways (Fig. 1A). Differences in ERI-6/7 abundances will affect

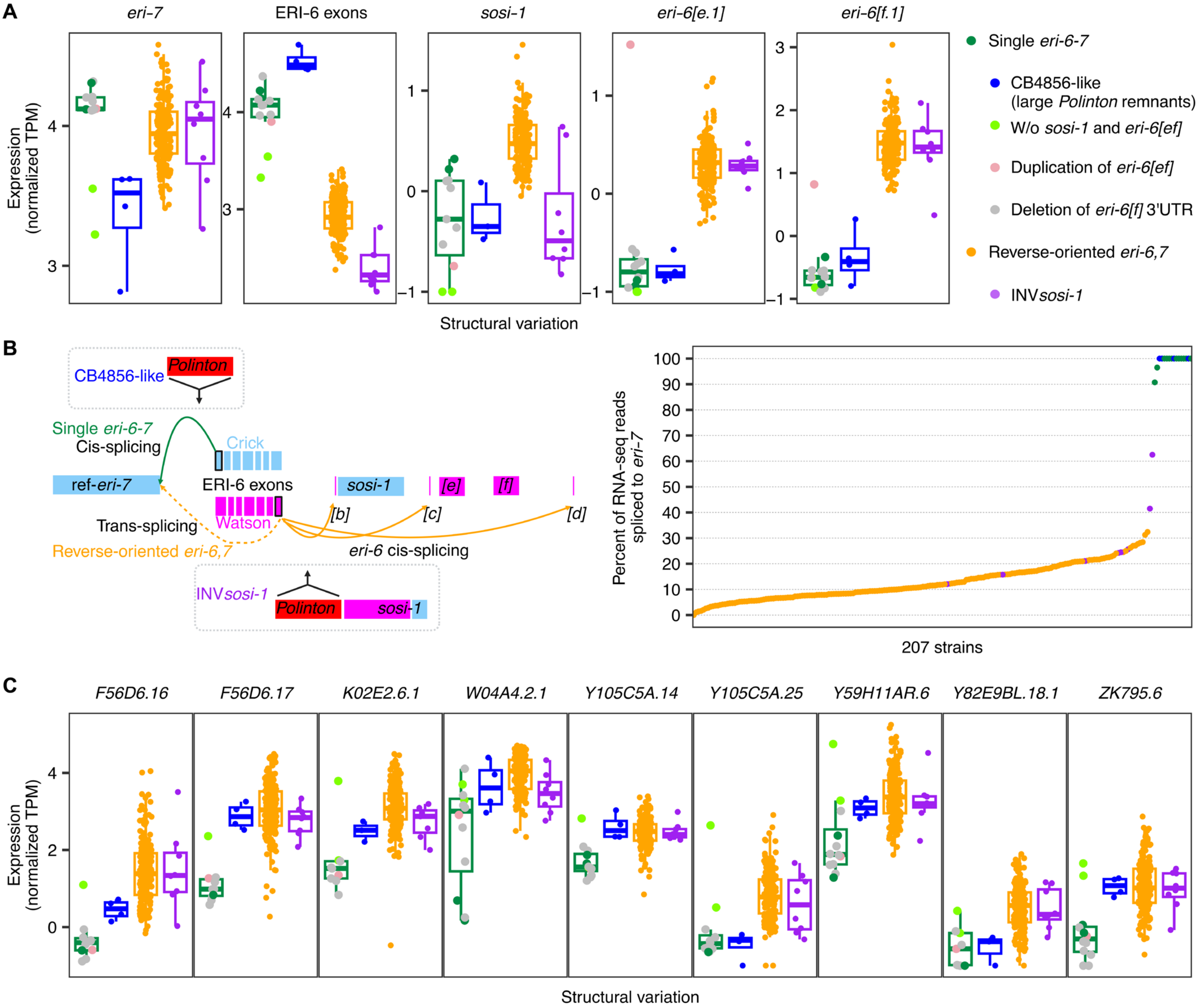

Fig. 4. Structural variations at the *eri-6/7* locus regulate genes in cis and trans. (**A** and **C**) Tukey box plots showing expression [$-\log_2$ (normalized TPM + 0.5)] variation of (A) five transcripts at the *eri-6/7* locus and (C) nine transcripts across the genome that include known targets of siRNAs requiring the ERI-6/7 helicase, among strains with major and minor structural variants within the locus. Each box is colored by major structural variants. Box edges denote the 25th and 75th quantiles of the data, and whiskers represent 1.5× the interquartile range. Statistical pairwise comparison results using two-sided Wilcoxon tests and Bonferroni corrections were presented in table S9. (**B**) Percent of spanning RNA-seq reads at the end of the last (seventh) ERI-6 exon that was spliced to *eri-7* when mapped to the reference genome for 207 strains. Graphic illustration of structural variation within the *eri-6/7* locus was created with BioRender.com. Each data point represents a strain, color-coded by structural variants. 3′UTR, 3′ untranslated region.

the generation of ERGO-1–dependent siRNAs and repression of their target genes. Among the putative targets of ERI-6/7–dependent siRNAs from our eQTL analysis, we observed significantly lower expression in strains in the Single *eri-6-7* group than strains in the Reverse-oriented *eri-6,7* group (Fig. 4C and table S9). We also found potential effects of structural variants in the CB4856-like strains on target expression variation within the Single *eri-6-7* group. Together, these results demonstrate that diverse structural variants at the *eri-6/7* locus probably altered *C. elegans* endogenous siRNA pathways from the production of the ERI-6/7 helicase to the expression of target genes.

DISCUSSION

Evolutionary genomic history of the *eri-6/7* locus driven by *Polintons*

Most strains with a single *eri-6-7* gene were isolated from the Hawaiian Islands or the Pacific region, where the highest known genetic diversity in the *C. elegans* species is found (Fig. 3 and figs. S8 and S11), which likely reflects the retention of ancestral diversity (*29*–*31*). Strains with an inversion of ERI-6 exons, however, are more widely distributed over the world and predominant in Europe. This set of strains show reduced genetic diversity at the locus, in agreement with an evolutionary-derived inversion of ERI-6 exons from the Crick to the Watson strand within the species (*31*) (Fig. 3 and figs. S8 and S11).

We thus favor the following scenario of evolution at the *eri-6/7* locus (Fig. 5). The *eri-6/7* gene was ancestrally coded as a single gene as in *C. briggsae* and at least nine other *Caenorhabditis* species (table S6) without *Polinton* insertions. The lack of *eri-6/7* homolog in *C. inopinata* (*14*) prevents us from using it as a closer outgroup. The ancestor of all *C. elegans* strains likely conserved the compact single *eri-6-7* gene structure as in *C. briggsae* and *C. brenneri*. Some strains, such as XZ1516, likely kept this ancestral single *eri-6-7* gene with no trace of *Polintons* (Figs. 2 and 5). Alternatively, in these strains, the *Polintons* were fully eliminated from the *eri-6/7* locus, yet the parsimonious explanation is that *Polintons* invaded the locus after the speciation of *C. elegans*.

We found *Polinton* remnants in the genome of every *C. elegans* strain with available WGS data in CaeNDR (*35*, *36*). At some time during the evolutionary history of *C. elegans*, a *Polinton* copy transposed, likely from another location in the genome or through horizontal transfer, and interrupted the *eri-6/7* gene with a large insertion on the left side of ERI-6 exons. No strain in our dataset retains a full *Polinton* at the locus; thus, this *Polinton* was either a partial copy when it transposed or subsequently became largely deleted. In strains such as CB4856, the still large *Polinton* remnants (~5 kb in CB4856) appear to impair *eri-7* transcription (Figs. 2 and 4A and 5).

Further *Polintons* insertions occurred in the vicinity, including perhaps to the right side of ERI-6 exons (Figs. 2 and 5). The occurrence of several *Polinton* copies at the same locus may have favored

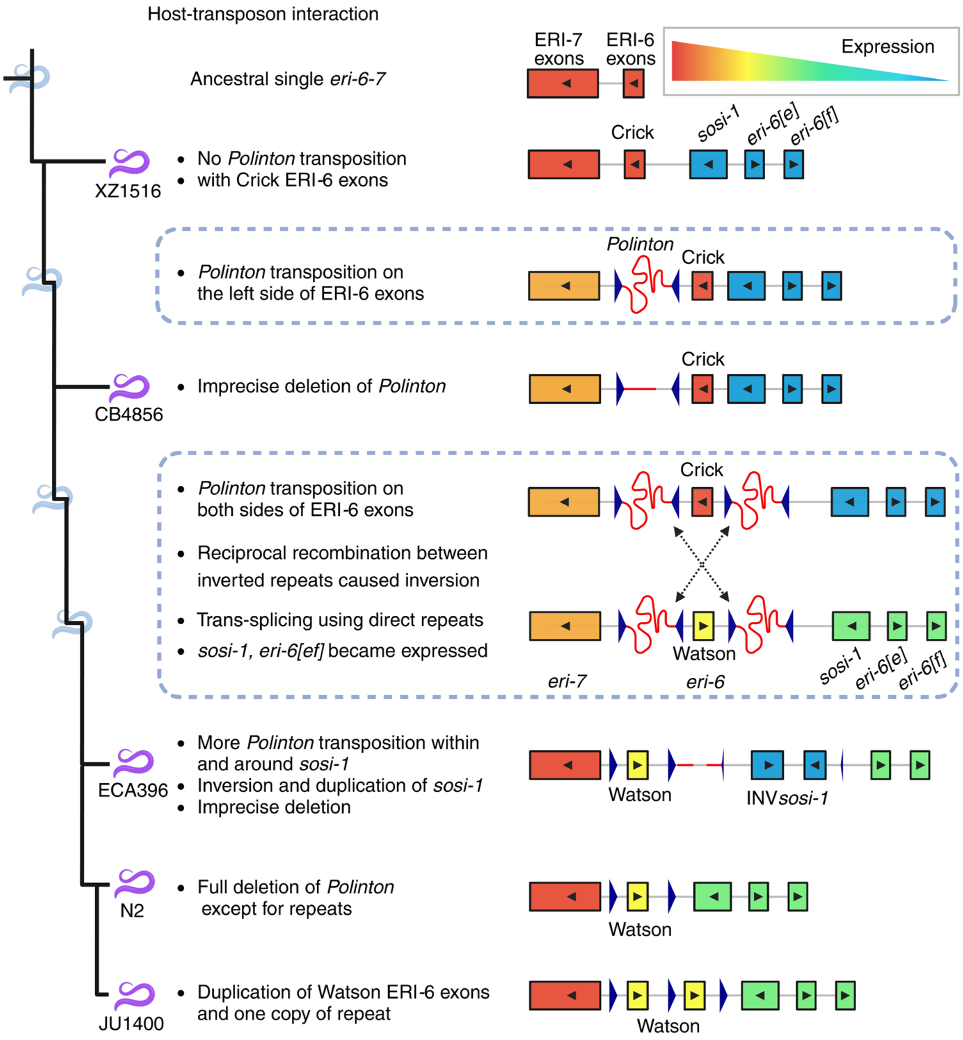
Host-transposon interaction
ERI-7 exons
ERI-6 exons
Expression
Ancestral single *eri-6-7*
Crick
sosi-1
eri-6[e]
eri-6[f]
XZ1516
• No *Polinton* transposition
• with Crick ERI-6 exons
Polinton
Crick
• *Polinton* transposition on the left side of ERI-6 exons
Crick
CB4856
• Imprecise deletion of *Polinton*
Crick
• *Polinton* transposition on both sides of ERI-6 exons
• Reciprocal recombination between inverted repeats caused inversion
• Trans-splicing using direct repeats
• *sosi-1, eri-6[ef]* became expressed
Watson
eri-7
eri-6
sosi-1
eri-6[e]
eri-6[f]
ECA396
• More *Polinton* transposition within and around *sosi-1*
• Inversion and duplication of *sosi-1*
• Imprecise deletion
Watson
INV*sosi-1*
N2
• Full deletion of *Polinton* except for repeats
Watson
JU1400
• Duplication of Watson ERI-6 exons and one copy of repeat
Watson

Fig. 5. Possible scenario for evolution at the *eri-6/7* locus with *Polintons*. Purple and light blue worms on the tree represent nodes with or without actual strains, respectively, to the best of our knowledge. Rectangles for different segments of *eri-6/7* were filled with gradient colors to indicate expression level across segments and branches on the tree. Black triangles inside rectangles represent orientation of gene segments. Dark blue triangles represent repeats. Red curved lines indicate *Polintons* other than the repeats. Created with BioRender.com.

ectopic recombination between inverted sequences and the ERI-6 exon inversion (Fig. 5). Surviving descendants of this inversion, such as the ECA396 and N2 strains, use repeats from *Polintons* for trans-splicing and thus maintain a hypomorphic *eri-6/7* function (Figs. 4C and 5). Meanwhile, the inversion activated *sosi-1*, *eri-6[e]*, and *eri-6[f]*, which were barely expressed in most strains with a single *eri-6-7* gene, at least in the tested conditions (Figs. 4A and 5). Ancestors of the reference strain N2 eliminated other *Polinton* fragments from the locus, except for the direct repeats that are necessary for trans-splicing. Strains such as JU1400 evolved a duplication of the Watson ERI-6 exons and one copy of the direct repeat, which could increase the number of correctly spliced *eri-6/7* transcripts (Figs. 2 and 5). *Polintons* might have caused more structural variations such as INV*sosi-1 (Figs. 2* and 5 and fig. S10).

The actual evolutionary process within this locus must be more complex than the model proposed above. Continuous insertions of *Polintons* might gradually weaken the ERI-6/7–dependent small RNA pathway, which could act in silencing of TEs. Reduced silencing might facilitate further transposition of *Polintons* and other TEs at the locus (Fig. 2). Alternatively, the *Polinton* insertions could have occurred through sudden bursts of transposition instead of gradually. Sudden environmental stress might have caused the high transposition rate of *Polintons* and other TEs (Fig. 2). Overall, the numerous TE insertions and genic rearrangements at this locus, which regulates small RNA pools and thereby TEs, support the hypothesis of a presumed battle against TEs by *C. elegans* hosts to preserve ERI-6/7 function and combat further TE activity. Only through further investigations of gene expression and TE positions in de novo assemblies will we learn more about the broad evolutionary significance of this type of battle.

Phenotypic effect of the structural variation at *eri-6/7* on siRNA pathways and their targets

With the ERGO-1 Argonaute, the ERI-6/7 helicase is required for production of endogenous primary 26G siRNAs by noncanonical Dicer processing of target mRNAs (*13*). Secondary siRNAs are produced by an amplification machinery, for which different pools of primary siRNAs compete (*15*, *37*), including endo-siRNAs dependent on Argonautes ERGO-1 and ALG-3/4, the genomically encoded Piwi-interacting RNAs (piRNAs), and the siRNAs derived from exogenous dsRNAs (*13*, *16*, *38*–*40*). Depending on the genomic and environmental contexts, genetic variation favoring one or the other primary siRNA pathway could have been selected (*41*–*44*). Research in mammals has shown the importance of dosage of the orthologous MOV10 helicase on retrovirus silencing (*45*). We showed here that natural structural variants at the *eri-6/7* locus were likely a major driver of variation in ERGO-1 pathway activity and mRNA levels of its downstream regulated targets. Two events, likely driven by *Polintons*, reduced ERI-6/7 pathway activity and increased piRNA-dependent and exogenous RNAi pathways: (i) the initial insertion of a *Polinton* within the *eri-6/7* gene and (ii) the inversion of ERI-6 exons. Other events might have acted in the reverse direction: the deletion of most of the intervening *Polintons*, the retention of direct repeats used in trans-splicing, and, in the strain JU1400, the duplication of the inverted ERI-6 exons. Because ERI-6/7–dependent siRNAs primarily target retrotransposons and unconserved, duplicated genes, with few introns, potentially of viral origins (*9*, *13*), the insertion of the *Polintons* and the resulting inversion could have at least transiently increased expression of target genes and retrotransposons while enhancing exogenous RNAi. We investigated potential correlation between ERI-6/7 production and response to exogenous RNAi (*46*) among wild *C. elegans* strains. We observed weak to mild negative correlation (Spearman correlation coefficient of −0.006 to −0.32 and *P* values of 0.057 to 0.97) between *eri-6/7* mRNA splicing rate (Fig. 4B) and embryonic lethality in 20 of the 29 maternal-effect genes individually targeted by RNAi among wild strains (*46*).

However, it is unclear what the effect might have been on *Polintons* themselves. Since their recent discovery in *C. elegans*, their possible regulation by small RNAs remains to be studied. The DNA polymerase of *Polintons* might be an ancient target of ERI-6/7–dependent siRNAs because the gene *E01G4.5*, a known target of ERI-6/7–dependent siRNAs in *C. elegans*, encodes a protein that has homology to viral DNA polymerases (*9*, *13*). *Polintons* might also bring foreign genes within them (*5*), which are potential targets of the ERGO-1 or piRNA pathways. The genes *sosi-1, eri-6[e]*, and *eri-6[f]* are absent at the *eri-6/7* locus in a subset of Hawaiian strains showing the most divergent *eri-6/7* region based on SNVs (Fig. 3). It is tempting to suggest that they appeared at this locus during the evolution of the species. The *eri-6[f]* exons are highly similar to another gene, *K09B11.4*, in the genome (*17*). Protein BLAST (*34*) suggested that both genes might originate from the *gypsy* retrotransposon *Cer1* (*47*). The gene *sosi-1* keeps additional copies in some wild strains and is a distant paralog of *eri-7* and other helicases in its C-terminal part. Further research can test whether *sosi-1* and *eri-6[e]* have been carried by a *Polinton* transposon. Similarly, the mode of duplication of the ERI-6/7 targets remains to be investigated.

Detailed genetic studies in the N2 reference strain have uncovered intricate regulatory interactions at the *eri-6/7/sosi-1* locus and between this locus and the splicing machinery. First, in the N2 strain, partly through matching piRNAs, *eri-6[e]*, *eri-6[f]*, and *sosi-1* are strong ERI-6/7–independent siRNA targets (*17*). Their downregulation by MUT-16–dependent siRNAs enables *eri-6/7* expression, perhaps by spreading chromatin marks (*17*). This regulation has been proposed to act as a negative feedback loop balancing ERGO-1–dependent secondary siRNAs and other secondary siRNA classes. Second, the use of the *Polinton* repeats as trans-splicing signal partially rescues the production of ERI-6/7. This peculiar mechanism of *eri-6/7* trans-splicing was proposed to act as a compensatory sensor of the splicing machinery, enabling more exogenous RNAi when an overwhelmed splicing machinery increases endo-siRNA production on poorly spliced genes (*48*). It remains unclear whether these seemingly intricate effects on siRNA pools in the N2 reference strain are an evolutionary leftover of transposon-driven structural variation at the locus. We hypothesize that across the evolutionary history of *C. elegans*, different siRNA pools may have been successively favored by natural selection. Alternatively, successive structural variants could have endowed the *eri-6/7* locus with physiological regulatory loops used in balancing the different siRNA classes downstream of environmental and organismal inputs.

The genetic diversity within the *eri-6/7* locus and its potential effects on small RNA pathways in *C. elegans* and target expression raise several intriguing questions for future studies. First, do other factors acting in the ERGO-1 pathway harbor this genetic diversity? We did observe multiple intronic structural variants in the Argonaute gene *ergo-1*, especially in the strain XZ1516 of the 17 *C. elegans* strains with long-read genome assemblies (table S10). We also found insertion variation likely related to the rolling-circle TEs, *Helitrons*

(*49*), in the strains JU2526, JU2600, and QX1794. In addition, 20 Hawaiian strains have the *ergo-1* locus in hyperdivergent regions (*26*) in CaeNDR (*35*, *36*), indicating great genetic diversity at the locus. It will be interesting to systematically identify structural variants in key small RNA pathway factors, in genes that act in TE silencing, and across the genome among wild *C. elegans* and examine roles of TEs in inducing these variants. The effects of structural variants in key small RNA pathway factors can be explored with a combination of mRNA and small RNA-seq. Second, what is the spectrum of target genes of the ERGO-1/ERI-6/7–dependent siRNA pathway among wild *C. elegans* strains. In addition to the interspecifically unconserved feature of the known ERI-6/7–dependent siRNA target genes (*13*), 36 wild *C. elegans* strains, mostly with the single *eri-6-7* gene, likely lack 15 to 20 of the 83 known target genes (fig. S12) (*13*). Mutants of *eri-6/7* or *ergo-1* in the background of wild strains can help to identify the diversity of the target genes of the pathway. Last, what are the necessary features of repeats for the trans-splicing mechanisms? Will part of the ~930-bp direct repeats support trans-splicing and show the same efficiency as in the full length? Will direct repeats with different sequences facilitate trans-splicing? Furthermore, how prevalent is the mechanism among wild *C. elegans* strains or other species? Experimental and computational approaches are needed to elaborate the mechanism and answer these important questions.

To conclude, our work dissected a distant eQTL hotspot and identified diverse TEs and structural variants within the *eri-6/7* locus potentially underlying variation in *C. elegans* endogenous siRNA pathways. This locus appears to have been the target of a large number of TE insertions including multiple copies of the otherwise rare *Polinton* transposon, which induced high genetic diversity at the locus through genome rearrangements. Some *C. elegans* strains evolved an odd trans-splicing mechanism to maintain hypomorphic function of the locus using *Polinton* TIRs that came to form direct repeats. The remarkable interactions between hosts and TEs play a major role in genome rearrangements and the regulation of gene expression.

MATERIALS AND METHODS

Genomic and transcriptomic data

We obtained the reference genomes of *C. elegans* (N2) and *C. briggsae* (AF16), the Gene Transfer Format (GTF) files of *C. elegans*, *C. briggsae*, and *C. brenneri* from WormBase (WS283) (*22*); the *de novo* assemblies of 17 wild *C. elegans* strains (CB4856, DL226, DL238, ECA36, ECA396, EG4725, JU310, JU1395, JU1400, JU2526, JU2600, MY2147, MY2693, NIC2, NIC526, QX1794, and XZ1516) and two wild *C. briggsae* strains (QX1410, VX34) from the NCBI Sequence Read Archive (SRA projects PRJNA523481, PRJNA622250, PRJNA692613, PRJNA784955, and PRJNA819174) (*24*–*28*); the alignment of whole-genome sequence data in the BAM format of 550 wild *C. elegans* strains, the soft-filtered, hard-filtered, and imputed isotype reference strain Variant Call Format (VCF) files from CaeNDR (20220216 release) (*35*, *36*); and the Illumina RNA-seq FASTQ files of 608 samples of 207 wild *C. elegans* strains from the NCBI SRA (projects PRJNA669810) (*19*).

RNA-seq mapping and eQTL analysis

To put transcriptomic data on the same page with the genomic data, we remapped RNA-seq reads using the *C. elegans* reference genome (WS283), the GTF file (WS283), and the pipeline *PEmRNA-seq-nf* (v1.0) (https://github.com/AndersenLab/PEmRNA-seq-nf) (*50*). Then, we selected reliably expressed transcripts, filtered outlier samples, and normalized expression abundance across samples using the R scripts *counts5strains10.R*, *nonDivergent_clustered.R*, and *norm_transcript_gwas.R* (https://github.com/AndersenLab/WI-Ce-eQTL/tree/main/scripts) (*51*), respectively, as previously described (*19*). In summary, we collected reliable expression abundance for 23,349 transcripts of 16,172 genes (15,449 protein-coding genes and 723 pseudogenes) from 560 samples of 207 strains. We also used *STAR* (v2.7.5) (*52*) to identify chimeric RNA-seq reads in the 560 samples.

We further used our recently developed genome-wide association study (GWAS) mapping pipeline, *Nemascan* (*53*), to identify eQTL for the 23,349 transcript expression traits (*54*), following the steps outlined previously (*19*). Briefly, we extracted SNVs of the 207 strains from the hard-filtered isotype reference strain VCF and filtered out variants that had any missing genotype calls and variants that were below the 5% minor allele frequency using *BCFtools* (v.1.9) (*55*). We further pruned variants with a LD threshold of $r^2 \geq 0.8$ using *-indep-pairwise 50 10 0.8* in *PLINK* (v1.9) (*56*, *57*) to generate the genotype matrix containing 27,854 markers. We randomly selected 200 traits and permuted each of them 200 times. For each of the 40,000 permuted traits, we used the leave-one-chromosome-out (LOCO) approach and the INBRED approach in the *GCTA* software (v1.93.2) (*58*, *59*) and calculated the eigen-decomposition significance (EIGEN) threshold as $-\log_{10}(0.05/N_{\text{test}})$ to identify QTL.

We determined the 5% false discovery rate (FDR) significance threshold for LOCO and INBRED, respectively, by calculating the 95th percentile of the significance of all detected QTL above using each approach. The LOCO and INBRED 5% FDR thresholds were 5.81 and 6.18, respectively. We then performed GWAS mapping on all 23,349 traits using LOCO and INBRED approaches and identified eQTL that passed their respective 5% FDR thresholds. Overall, we detected 10,291 eQTL for 5668 transcript expression traits, with 4899 eQTL for 4254 traits in LOCO and 5392 eQTL for 4700 traits in INBRED (table S2).

We classified eQTL as local (within 2 Mb surrounding the transcript) or distant (nonlocal) (fig. S1A and table S2). For distant eQTL located outside of the common hyperdivergent regions among the 207 strains (*19*, *26*), we identified hotspot regions enriched with distant eQTL for LOCO and INBRED results, respectively (table S2) (*19*).

The genomic region harboring the *eri-6/7* locus at 21 cM on chromosome I was identified as a distant eQTL hotspot in both LOCO and INBRED with 18 and 12 distant eQTL, respectively, for 19 different transcript expression traits (table S2). This hotspot was also identified in our previous study using a different GWAS mapping pipeline (*19*).

Computational fine mappings to search for common SNV candidates

Computational fine mappings were further performed for eQTL related to the *eri-6/7* locus. Briefly, for each eQTL (the peak marker with the highest significance), we defined a QTL of interest as ±100 SNVs from the rightmost and leftmost markers above the 5% FDR significance threshold surrounding the eQTL. Then, using genotype data from the imputed VCF, we generated a QTL of interest genotype matrix that was filtered as described above, with the one exception that we did not perform LD pruning. We used the LOCO and the INBRED approaches as above to perform fine mappings (figs. S2 and S3 and table S3). To prioritize top candidates among markers used in fine mappings, we applied the following per QTL per trait filters: top 5% most significant markers, LD with the peak marker higher than 0.6, out of common hyperdivergent genomic regions (*26*, *35*), and high-impact variants as predicted by BCFtools (v.1.9) (*55*) and annotated in CaeNDR (*35*, *36*).

Among top candidates of all distant eQTL in the hotspot at 21 cM on chromosome I, we searched for common candidates likely affecting the most eQTL. The most common candidates are SNVs I: 4,464,670 (D259Y) and I: 4,464,857 (R321Q), both of which were top candidates for 11 of the 18 eQTL in LOCO and 8 of the 12 eQTL in INBRED in the hotspot at 21 cM on chromosome I (tables S3 and S4). We further found two other transcript expression traits, *W04B5.1* and *Y82E9BL.18.1*, also have the above two variants as top candidates for their distant eQTL at 20.5 cM on chromosome I. The two transcripts are also known targets of ERGO-1/ERI-6/7–dependent siRNAs (*13*, *23*). Furthermore, the two variants were top candidates for local eQTL of *eri-6[c]*, *eri-6[e]*, *eri-6[f]*, and ERI-6 exons (combined expression of *eri-6[a-d]*) (fig. S2 and table S3).

Both of the two top candidate variants are located in the second exon of the isoform *eri-6[e]* in the gene *eri-6*. The two SNVs are also in perfect LD among our collection of wild *C. elegans* strains. Although both SNVs are missense mutations, the variant I: 4,464,670 has a negative and lower BLOSUM (*36*, *60*) score of "−3" (annotated in CaeNDR) compared to a score of "1" of the variant I: 4,464,857, indicating a more radical amino acid substitution of the former than the latter in comparison of the alternative allele to the reference allele at each variant. Therefore, the variants I: 4,464,670 and I: 4,464,857 were the first- and second-best candidate variants, respectively.

CRISPR-Cas9 genome editing

We used CRISPR-Cas9 genome editing to test effects of different alleles of the top two fine mapping candidate variants for the local eQTL of *eri-6[e]* in different wild *C. elegans* strains. Among the 207 wild strains in our RNA-seq dataset, 191 and 16 strains have reference (REF) and alternative (ALT) alleles, respectively, at the two top candidates. In addition to this local eQTL, we also identified one distant eQTL (IV: 16,045,665) in this study (table S2) and two distant eQTL (IV: 17,072,978 and V: 2,792,989) in our previous study (*19*) for *eri-6[e]*. To avoid possible confounding effects from these distant eQTL, we selected strains with reference alleles at all three distant eQTL for editing. Among the 207 wild strains, 165 of the 191 REF strains and 4 of the 16 ALT strains above have reference alleles at all three distant eQTL. We randomly chose two REF strains (JU2141 and JU3144) and two ALT strains (JU642 and JU2106) for genome editing. We used CRISPR-Cas9 genome editing to individually introduce the reference and alternative alleles of each candidate variant into ALT and REF strains, respectively. We generated single edits in the four strains for the top candidate variant (I: 4,464,670) and in the two strains, JU3144 and JU2106, for the second-best candidate variant (I: 4,464,857). We also generated double edits of both candidate variants in strains JU3144 and JU2106.

Genome editing was performed using a co-CRISPR approach with the coconversion marker *dpy-10* as previously described (*61*, *62*). Single-strand guide RNAs (sgRNAs) for the two top candidate variants were designed using the online analysis platform Benchling (www.benchling.com). All sgRNAs were ordered from Synthego (Redwood City, CA). Single-stranded oligodeoxynucleotides (ssODNs) templates used for homology-directed repair were ordered as ultramers from IDT (Coralville, IA). Mixed reagents—including sgRNAs for *dpy-10* at 1 μM and a single target variant at 6 μM, ssODN templates for *dpy-10* at 0.5 μM and the target variant at 5 μM, and purified Cas9 protein (QB3 Macrolab, Berkeley, CA) at 5 μM—were used in injection for young adult hermaphrodites. Injected animals were singled to fresh plates and allowed to lay until F1s developed to the L4 stage. F1s were screened for the *dpy-10* mutant "roller" phenotype and singled to fresh plates. F1s were allowed to lay eggs before single-animal lysis and polymerase chain reaction, and the products were sequenced using Sanger sequencing by MCLab Molecular Cloning Laboratories (South San Francisco, CA). F2 non-Roller offspring of successfully edited parents were singled to fresh plates. All alleles were confirmed by sequencing singled offspring for at least two additional generations to confirm accuracy and homozygosity of the edited sequence. Two independent edits of each allele in each genetic background were generated to control for off-target effects. All oligonucleotides and genome-edited strains are listed in table S5.

DNA alignment

We aligned each of the 17 *de novo* PacBio assemblies of wild *C. elegans* strains to the N2 reference genome using *MUMmer* (v3.1) (*63*) and extracted sequences that were aligned to the N2 *eri-6/7* locus using BEDTools (v2.29.2) (*64*). Then, we performed pairwise alignments among these sequences and to the *eri-6/7* N2 reference sequence using Unipro UGENE (v.47.0) (*65*). Large insertions (>50 bp) in the wild strains to the reference were blasted in WormBase (*22*) to identify potential transposon origins.

Scan for *Polinton* and TIRs in genome assemblies

We obtained the amino acid sequences of pPolB1 and INT in *C. briggsae Polinton-1* (WBTransposon00000832) (*22*) using *ORFfinder* (www.ncbi.nlm.nih.gov/orffinder/) and the 744-bp DNA sequence for the TIRs from 10,302,516 to 10,303,259 bp on chromosome I in the *C. elegans* (N2) reference genome. We searched for the *Polinton* and TIRs sequences in the 21 genome assemblies using tblastn and blastn in BLAST (v2.14.0) (*66*), respectively. We filtered the results by a maximum *e* value of 0.001 and a minimum bitscore of 50 (*33*). We merged pPolB1, INT, and TIR hits within 4, 2, and 2 kb, respectively, with consideration of strandedness. *Polinton* insertions were identified by the presence of both pPolB1 and INT within 20 kb.

We also searched for *sosi-1* outside of the *eri-6/7* locus in the genome assemblies using DNA sequence of *sosi-1* in the reference and found an additional copy in the strains JU2526, ECA396, XZ1516, and JU1400, and two additional copies in the strains ECA36 and QX1794 in their PacBio genome assemblies. Genomic locations surrounding these additional copies in the six strains correspond to ~0.31 Mb on the chromosome III in the reference N2 genome. The additional copies of *sosi-1* outside the *eri-6/7* locus in the six strains share most alleles compared to the *sosi-1* within the *eri-6/7* locus.

Identification of structural variants using short-read WGS data

We extracted information of split reads mapped to the reference *eri-6/7* locus (I: 4,451,194 to 4,469,460 bp) and with a minimum quality score equal of 20 from the BAM files of the 550 wild *C. elegans* strains. (i): To identify potential inversions in the *eri-6/7* locus, we first selected split reads with both the primary and chimeric alignments mapped to this region but to different strands. We assigned the primary and chimeric alignment positions of each split read into 200-bp bins and required at least four reads that had the primary and chimeric alignments in the same pair of bins for a relatively reliable inversion event in each strain. We focused on inversions spanning at least three bins and found in more than 10 strains. (ii): To identify potential sites of *Polinton* remnants, we selected the split reads outside of the direct repeats at the *eri-6/7* locus and with the chimeric alignment mapped to *Polinton* (*Polinton-1_CB*, WB-Transposon00000738) and its surrounding *Polinton_CE_TIR* on chromosome I from 10,302,516 to 10,319,657 bp. At least two reads were required. The primary alignment of these reads indicated the potential sites of *Polinton* remnants in the *eri-6/7* locus in wild strains.

Furthermore, we counted the coverage per bp in the *eri-6/7* locus for each short-read WGS BAM file using BEDTools (v2.29.2) (*64*). We calculated the percentage of the coverage at each base pair to the mean coverage within the *eri-6/7* locus in each strain. Then, we performed a sliding window analysis with a 200-bp window size and a 100-bp step size for each strain. A 173-bp tandem repeat region from 4,465,414 to 4,465,586 bp on chromosome I was masked in the results.

To identify additional copies and haplotypes of *sosi-1* among the 550 wild strains, we focused on 93 variants of the 101 SNVs tagged "high heterozygosity" within the *sosi-1* region in the soft-filtered isotype VCF. We used the following threshold to define *sosi-1* haplotype and copy numbers among the 550 strains: 449 strains show homozygous reference alleles at all 93 SNVs (except one strain at 92 SNVs), indicating that they only have the reference haplotype *sosi-1*; 80 strains show heterozygous alleles at more than 60 SNVs, indicating two copies of *sosi-1* with divergent haplotypes; three strains have homozygous alternative alleles at more than 90 SNVs, indicating missing of the reference *sosi-1* in the *eri-6/7* locus and the existence of the alternative *sosi-1* copy; and 11 strains show undetected genotype at 60 to 93 SNVs and extreme low coverages in *sosi-1* (fig. S10D), indicating they may lack *sosi-1* in the genomes; the *sosi-1* haplotype and copy number of the remaining seven strains are unclear as they have numbers of homozygous and homozygous alleles in between the above threshold (table S8).

Genetic relatedness

Genetic variation data across the genome among the 550 *C. elegans* strains were extracted from the hard-filtered VCF above using BCFtools (v.1.9) (*55*). These variants were pruned to the 1,199,944 biallelic SNVs without missing genotypes. We converted this pruned VCF file to a PHYLIP file using the *vcf2phylip.py* script (*67*). The unrooted neighbor-joining tree was made using the R packages phangorn (v2.5.5) (*68*) and ggtree (v1.14.6) (*69*).

A second PHYLIP file was built by the same method above but only with 95 SNVs within the *eri-6/7* locus. A haplotype network was generated using this PHYLIP file and SplitsTree CE (v6.1.16) (*70*).

Supplementary Materials

The PDF file includes:
Figs. S1 to S12
Legends for tables S1 to S10

Other Supplementary Material for this manuscript includes the following:
Tables S1 to S10

REFERENCES AND NOTES

1. B. McClintock, The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U.S.A.* **36**, 344–355 (1950).
2. Y. H. Gray, It takes two transposons to tango: Transposable-element-mediated chromosomal rearrangements. *Trends Genet.* **16**, 461–468 (2000).
3. G. Bourque, K. H. Burns, M. Gehring, V. Gorbunova, A. Seluanov, M. Hammell, M. Imbeault, Z. Izsvák, H. L. Levin, T. S. Macfarlan, D. L. Mager, C. Feschotte, Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).
4. C. Gilbert, C. Feschotte, Horizontal acquisition of transposable elements and viral sequences: Patterns and consequences. *Curr. Opin. Genet. Dev.* **49**, 15–24 (2018).
5. S. A. Widen, I. C. Bes, A. Koreshova, P. Pliota, D. Krogull, A. Burga, Virus-like transposons cross the species barrier and drive the evolution of genetic incompatibilities. *Science* **380**, eade0705 (2023).
6. J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, G. J. Hannon, Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**, 1089–1103 (2007).
7. H. Ito, Small RNAs and transposon silencing in plants. *Dev. Growth Differ.* **54**, 100–107 (2012).
8. Ö. Deniz, J. M. Frost, M. R. Branco, Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* **20**, 417–431 (2019).
9. S. E. J. Fischer, G. Ruvkun, Caenorhabditis elegans ADAR editing and the ERI-6/7/MOV10 RNAi pathway silence endogenous viral elements and LTR retrotransposons. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5987–5996 (2020).
10. R. Rebollo, M. T. Romanish, D. L. Mager, Transposable elements: An abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* **46**, 21–42 (2012).
11. D. Gao, N. Jiang, R. A. Wing, J. Jiang, S. A. Jackson, Transposons play an important role in the evolution and diversification of centromeres among closely related species. *Front. Plant Sci.* **6**, 216 (2015).
12. E. B. Chuong, N. C. Elde, C. Feschotte, Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 (2017).
13. S. E. J. Fischer, T. A. Montgomery, C. Zhang, N. Fahlgren, P. C. Breen, A. Hwang, C. M. Sullivan, J. C. Carrington, G. Ruvkun, The ERI-6/7 helicase acts at the first stage of an siRNA amplification pathway that targets recent gene duplications. *PLOS Genet.* **7**, e1002369 (2011).
14. N. Kanzaki, I. J. Tsai, R. Tanaka, V. L. Hunt, D. Liu, K. Tsuyama, Y. Maeda, S. Namai, R. Kumagai, A. Tracey, N. Holroyd, S. R. Doyle, G. C. Woodruff, K. Murase, H. Kitazume, C. Chai, A. Akagi, O. Panda, H.-M. Ke, F. C. Schroeder, J. Wang, M. Berriman, P. W. Sternberg, A. Sugimoto, T. Kikuchi, Biology and genome of a newly discovered sibling species of Caenorhabditis elegans. *Nat. Commun.* **9**, 3216 (2018).
15. R. C. Lee, C. M. Hammell, V. Ambros, Interacting endogenous and exogenous RNAi pathways in Caenorhabditis elegans. *RNA* **12**, 589–597 (2006).
16. S. E. J. Fischer, M. D. Butler, Q. Pan, G. Ruvkun, *Trans*-splicing in *C. elegans* generates the negative RNAi regulator ERI-6/7. *Nature* **455**, 491–496 (2008).
17. A. K. Rogers, C. M. Phillips, A small-RNA-mediated feedback loop maintains proper levels of 22G-RNAs in *C. elegans*. *Cell Rep.* **33**, 108279 (2020).
18. F. W. Albert, L. Kruglyak, The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16**, 197–212 (2015).
19. G. Zhang, N. M. Roberto, D. Lee, S. R. Hahnel, E. C. Andersen, The impact of species-wide gene expression variation on *Caenorhabditis elegans* complex traits. *Nat. Commun.* **13**, 3462 (2022).
20. V. V. Kapitonov, J. Jurka, Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4540–4545 (2006).
21. E. J. Pritham, T. Putliwala, C. Feschotte, Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* **390**, 3–17 (2007).
22. T. W. Harris, V. Arnaboldi, S. Cain, J. Chan, W. J. Chen, J. Cho, P. Davis, S. Gao, C. A. Grove, R. Kishore, R. Y. N. Lee, H.-M. Muller, C. Nakamura, P. Nuin, M. Paulini, D. Raciti, F. H. Rodgers, M. Russell, G. Schindelman, K. V. Auken, Q. Wang, G. Williams, A. J. Wright, K. Yook, K. L. Howe, T. Schedl, L. Stein, P. W. Sternberg, WormBase: A modern Model Organism Information Resource. *Nucleic Acids Res.* **48**, D762–D767 (2020).
23. U. Seroussi, A. Lugowski, L. Wadi, R. X. Lao, A. R. Willis, W. Zhao, A. E. Sundby, A. G. Charlesworth, A. W. Reinke, J. M. Claycomb, A comprehensive survey of *C. elegans* argonaute proteins reveals organism-wide gene regulatory networks and functions. *eLife* **12**, e83853 (2023).
24. C. Kim, J. Kim, S. Kim, D. E. Cook, K. S. Evans, E. C. Andersen, J. Lee, Long-read sequencing reveals intra-species tolerance of substantial structural variations and new subtelomere formation in C. elegans. *Genome Res.* **29**, 1023–1035 (2019).
25. L. T. Bubrig, J. M. Sutton, J. L. Fierst, Caenorhabditis elegans dauers vary recovery in response to bacteria from natural habitat. *Ecol. Evol.* **10**, 9886–9895 (2020).
26. D. Lee, S. Zdraljevic, L. Stevens, Y. Wang, R. E. Tanny, T. A. Crombie, D. E. Cook, A. K. Webster, R. Chirakar, L. R. Baugh, M. G. Sterken, C. Braendle, M.-A. Félix, M. V. Rockman, E. C. Andersen, Balancing selection maintains hyper-divergent haplotypes in *Caenorhabditis elegans*. *Nat. Ecol. Evol.* **5**, 794–807 (2021).

27. B. Y. Lee, J. Kim, J. Lee, Long-read sequencing infers a mechanism for copy number variation of template for alternative lengthening of telomeres in a wild *C. elegans* strain. *MicroPubl. Biol.* **2022**, 10.17912/micropub.biology.000563 (2022).
28. L. Stevens, N. D. Moya, R. E. Tanny, S. B. Gibson, A. Tracey, H. Na, R. Chitrakar, J. Dekker, A. J. M. Walhout, L. R. Baugh, E. C. Andersen, Chromosome-level reference genomes for two strains of *Caenorhabditis briggsae*: An improved platform for comparative genomics. *Genome Biol. Evol.* **14**, evac042 (2022).
29. T. A. Crombie, S. Zdraljevic, D. E. Cook, R. E. Tanny, S. C. Brady, Y. Wang, K. S. Evans, S. Hahnel, D. Lee, B. C. Rodriguez, G. Zhang, J. van der Zwagg, K. Kiontke, E. C. Andersen, Deep sampling of Hawaiian *Caenorhabditis elegans* reveals high genetic diversity and admixture with global populations. *eLife* **8**, e50465 (2019).
30. T. A. Crombie, P. Battlay, R. E. Tanny, K. S. Evans, C. M. Buchanan, D. E. Cook, C. M. Dilks, L. A. Stinson, S. Zdraljevic, G. Zhang, N. M. Roberto, D. Lee, M. Ailion, K. A. Hodgins, E. C. Andersen, Local adaptation and spatiotemporal patterns of genetic diversity revealed by repeated sampling of Caenorhabditis elegans across the Hawaiian Islands. *Mol. Ecol.* **31**, 2327–2347 (2022).
31. E. C. Andersen, J. P. Gerke, J. A. Shapiro, J. R. Crissman, R. Ghosh, J. S. Bloom, M.-A. Félix, L. Kruglyak, Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. *Nat. Genet.* **44**, 285–290 (2012).
32. W. Li, J. E. Shaw, A variant Tc4 transposable element in the nematode C. elegans could encode a novel protein. *Nucleic Acids Res.* **21**, 59–67 (1993).
33. D.-E. Jeong, S. Sundrani, R. N. Hall, M. Krupovic, E. V. Koonin, A. Z. Fire, DNA polymerase diversity reveals multiple incursions of polintons during nematode evolution. *Mol. Biol. Evol.* **40**, msad274 (2023).
34. E. W. Sayers, E. E. Bolton, J. R. Brister, K. Canese, J. Chan, D. C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B. W. Trawick, K. D. Pruitt, S. T. Sherry, Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **50**, D20–D26 (2022).
35. D. E. Cook, S. Zdraljevic, J. P. Roberts, E. C. Andersen, CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Res.* **45**, D650–D657 (2017).
36. T. A. Crombie, R. McKeown, N. D. Moya, K. S. Evans, S. J. Widmayer, V. LaGrassa, N. Roman, O. Tursunova, G. Zhang, S. B. Gibson, C. M. Buchanan, N. M. Roberto, R. Vieira, R. E. Tanny, E. C. Andersen, CaeNDR, the *Caenorhabditis* natural diversity resource. *Nucleic Acids Res.* **52**, D850–D858 (2024).
37. T. F. Duchaine, J. A. Wohlschlegel, S. Kennedy, Y. Bei, D. Conte Jr., K. Pang, D. R. Brownell, S. Harding, S. Mitani, G. Ruvkun, J. R. Yates III, C. C. Mello, Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell* **124**, 343–354 (2006).
38. J. M. Claycomb, P. J. Batista, K. M. Pang, W. Gu, J. J. Vasale, J. C. van Wolfswinkel, D. A. Chaves, M. Shirayama, S. Mitani, R. F. Ketting, D. Conte Jr., C. C. Mello, The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* **139**, 123–134 (2009).
39. W. Gu, M. Shirayama, D. Conte Jr., J. Vasale, P. J. Batista, J. M. Claycomb, J. J. Moresco, E. M. Youngman, J. Keys, M. J. Stoltz, C.-C. G. Chen, D. A. Chaves, S. Duan, K. D. Kasschau, N. Fahlgren, J. R. Yates III, S. Mitani, J. C. Carrington, C. C. Mello, Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol. Cell* **36**, 231–244 (2009).
40. C. C. Conine, P. J. Batista, W. Gu, J. M. Claycomb, D. A. Chaves, M. Shirayama, C. C. Mello, Argonautes ALG-3 and ALG-4 are required for spermatogenesis-specific 26G-RNAs and thermotolerant sperm in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3588–3593 (2010).
41. M. Tijsterman, K. L. Okihara, K. Thijssen, R. H. A. Plasterk, PPW-1, a PAZ/PIWI protein required for efficient germline RNAi, is defective in a natural isolate of *C. elegans*. *Curr. Biol.* **12**, 1535–1540 (2002).
42. D. A. Pollard, M. V. Rockman, Resistance to germline RNA interference in a *Caenorhabditis elegans* wild isolate exhibits complexity and nonadditivity. *G3* **3**, 941–947 (2013).
43. A. Ashe, T. Bélicard, J. Le Pen, P. Sarkies, L. Frézal, N. J. Lehrbach, M.-A. Félix, E. A. Miska, A deletion polymorphism in the *Caenorhabditis elegans* RIG-I homolog disables viral RNA dicing and antiviral immunity. *eLife* **2**, e00994 (2013).
44. H. T. Chou, F. Valencia, J. C. Alexander, A. D. Bell, D. Deb, D. A. Pollard, A. B. Paaby, Diversification of small RNA pathways underlies germline RNAi incompetence in wild *Caenorhabditis elegans* strains. *Genetics* **226**, iyad191 (2024).
45. Y. Guan, H. Gao, N. A. Leu, A. Vourekas, P. Alexiou, M. Maragkakis, Z. Kang, Z. Mourelatos, G. Liang, P. J. Wang, The MOV10 RNA helicase is a dosage-dependent host restriction factor for LINE1 retrotransposition in mice. *PLOS Genet.* **19**, e1010566 (2023).
46. A. B. Paaby, A. G. White, D. D. Riccardi, K. C. Gunsalus, F. Piano, M. V. Rockman, Wild worm embryogenesis harbors ubiquitous polygenic modifier variation. *eLife* **4**, e09178 (2015).
47. R. J. Britten, Active gypsy/Ty3 retrotransposons or retroviruses in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 599–601 (1995).
48. M. A. Newman, F. Ji, S. E. J. Fischer, A. Anselmo, R. I. Sadreyev, G. Ruvkun, The surveillance of pre-mRNA splicing is an early step in C. elegans RNAi of endogenous genes. *Genes Dev.* **32**, 670–681 (2018).
49. V. V. Kapitonov, J. Jurka, Rolling-circle transposons in eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8714–8719 (2001).
50. G. Zhang, *PEmRNA-seq-nf* (Zenodo, 2022); https://doi.org/10.5281/zenodo.6595320.
51. G. Zhang, *WI-Ce-eQTL* (Zenodo, 2022); https://doi.org/10.5281/zenodo.6595353.
52. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
53. S. J. Widmayer, K. S. Evans, S. Zdraljevic, E. C. Andersen, Evaluating the power and limitations of genome-wide association studies in *Caenorhabditis elegans*. *G3* **12**, jkac114 (2022).
54. G. Zhang, *Ce-eri-67* (Zenodo, 2024); https://doi.org/10.5281/zenodo.11450742.
55. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
56. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
57. C. C. Chang, C. C. Chow, L. C. A. M. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
58. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
59. L. Jiang, Z. Zheng, T. Qi, K. E. Kemper, N. R. Wray, P. M. Visscher, J. Yang, A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.* **51**, 1749–1755 (2019).
60. S. Henikoff, J. G. Henikoff, Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10915–10919 (1992).
61. A. Paix, Y. Wang, H. E. Smith, C.-Y. S. Lee, D. Calidas, T. Lu, J. Smith, H. Schmidt, M. W. Krause, G. Seydoux, Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 Sites in *Caenorhabditis elegans*. *Genetics* **198**, 1347–1356 (2014).
62. C. M. Dilks, S. R. Hahnel, Q. Sheng, L. Long, P. T. McGrath, E. C. Andersen, Quantitative benzimidazole resistance and fitness effects of parasitic nematode beta-tubulin alleles. *Int. J. Parasitol. Drugs Drug Resist.* **14**, 28–36 (2020).
63. S. Kurtz, A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S. L. Salzberg, Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
64. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
65. K. Okonechnikov, O. Golosova, M. Fursov, UGENE team, Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167 (2012).
66. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
67. E. M. Ortiz, *vcf2phylip v2.0: Convert a VCF Matrix into Several Matrix Formats for Phylogenetic Analysis* (Zenodo, 2019); https://doi.org/10.5281/zenodo.2540861.
68. K. P. Schliep, phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
69. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, GGTREE: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
70. D. H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).

Acknowledgments: We like to thank E. Koury for microinjection in CRISPR-Cas9 genome editing. We like to thank N. D. Moya for suggestions on useful bioinformatics tools. We would like to thank members of the E.C.A.'s Laboratory for helpful comments on the manuscript. We also like to thank CaeNDR and WormBase for which these analyses would not have been possible. Figures 1, 4, and 5 were created with BioRender.com. **Funding:** This research was supported by the National Science Foundation (1764421) and Simons Foundation (597491) Research Center for Mathematics of Complex Biological Systems and a Human Frontiers Research Program grant (RGP0001/2019) to E.C.A. G.Z. is supported by a grant from the Fondation pour la Recherche Médicale (ARF202209015859). M.-A.F. is supported by the Centre National de la Recherche Scientifique and a grant from the Agence Nationale de la Recherche (ANR-19-CE12-0025). CaeNDR is supported by NSF Capacity Grant (2224885) to E.C.A. **Author contributions:** G.Z., M.-A.F., and E.C.A. conceived the study. G.Z. analyzed the data. G.Z. wrote the manuscript. M.-A.F. and E.C.A. edited the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Codes for analysis of figures are available at Zenodo (DOI: https://doi.org/10.5281/zenodo.11450742).

Submitted 28 February 2024
Accepted 13 August 2024
Published 20 September 2024
10.1126/sciadv.ado9461