

Telomeric small RNAs in the genus *Caenorhabditis*

STEPHEN FRENK,^{1,2} EVAN H. LISTER-SHIMAUCHI,^{1,2} and SHAWN AHMED^{1,2,3}

¹Department of Genetics, ²Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599-3280, USA

³Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599-3280, USA

ABSTRACT

Telomeric DNA is composed of simple tandem repeat sequences and has a G-rich strand that runs 5' to 3' toward the chromosome terminus. Small RNAs with homology to telomeres have been observed in several organisms and could originate from telomeres or from interstitial telomere sequences (ITSs), which are composites of degenerate and perfect telomere repeat sequences found on chromosome arms. We identified *Caenorhabditis elegans* small RNAs composed of the *Caenorhabditis* telomere sequence (TTAGGC)_n with up to three mismatches, which might interact with telomeres. We rigorously defined ITSs for genomes of *C. elegans* and for two closely related nematodes, *Caenorhabditis briggsae* and *Caenorhabditis remanei*. Most telomeric small RNAs with mismatches originated from ITSs, which were depleted from mRNAs but were enriched in introns whose genes often displayed hallmarks of genomic silencing. *C. elegans* small RNAs composed of perfect telomere repeats were very rare but their levels increased by several orders of magnitude in *C. briggsae* and *C. remanei*. Major small RNA species in *C. elegans* begin with a 5' guanine nucleotide, which was strongly depleted from perfect telomeric small RNAs of all three *Caenorhabditis* species. Perfect G-rich or C-rich telomeric small RNAs commonly began with 5' UAGGCU and 5' UUAGGC or 5' CUAAGC, respectively. In contrast, telomeric small RNAs with mismatches had a mixture of all four 5' nucleotides. We suggest that perfect telomeric small RNAs have a mechanism of biogenesis that is distinct from known classes of small RNAs and that a dramatic change in their regulation occurred during recent *Caenorhabditis* evolution.

Keywords: small RNA; telomere; interstitial telomere sequence; ITS

INTRODUCTION

Telomeres are repetitive sequences at the end of linear eukaryotic chromosomes that possess a G-rich strand that runs 5' to 3' toward chromosome termini (Smogorzewska and de Lange 2004). These sequences play an essential role in the maintenance of genome stability by protecting chromosome ends from recognition by DNA damage response pathways, thereby preventing chromosome fusions (Palm and de Lange 2008). Due to the end replication problem, telomeres are eroded with each cell division. Persistent telomere erosion triggers DNA damage response activation and results in senescence. To counteract telomere erosion, the reverse transcriptase telomerase uses an RNA template to add de novo telomeric repeats to chromosome termini. Although telomerase is essential for maintenance of proper germline function across generations, this enzyme is absent from most human somatic cells. Somatic silencing of telomerase and associated telomere shortening represents a barrier to tumorigenesis in large mammals (Seluanov et al. 2007; Gomes et al. 2011).

Telomere maintenance in around 90% of cancers is achieved through activation of telomerase, while the remaining 10% utilize a telomerase-independent mechanism that is less well understood known as Alternative Lengthening of Telomeres (ALT) (Pickett and Reddel 2015).

The protective function of telomeres depends in part on binding of a protein complex known as Shelterin, composed of POT1, TRF1, TRF2, TIN2, TPP1, and RAP1, that interacts with single- (POT1) and double-stranded (TRF1 and TRF2) telomeric DNA to promote telomere capping as well as the maintenance of a repressive chromatin environment (Dyer et al. 2017; de Lange 2018). Heterochromatin associated histone modifications such as H3K9me3, H3k20me3 and hypoacetylation of H3 and H4 are a common feature of telomeres (Schoeftner and Blasco 2010). Somewhat paradoxically, transcription by RNA polymerase II has also been observed at telomeres, leading to formation of the G-rich telomere repeat containing long noncoding RNA

© 2019 Frenk et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://majournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Corresponding author: shawn@med.unc.edu

Article is online at <http://www.majournal.org/cgi/doi/10.1261/rna.071324.119>.

TERRA, which is composed of subtelomeric DNA sequences and some perfect telomere repeats, as well as the less abundant C-rich antisense telomeric transcript ARRET (Azzalin et al. 2007; Bah et al. 2012). TERRA expression is up-regulated at critically short and damaged telomeres, where it may recruit telomerase and chromatin modifiers in order to promote telomere elongation and protect telomeres (Cusanelli et al. 2013; Azzalin and Lingner 2015). It has been observed that TERRA prominently coats telomeres of the silent X chromosome in mammals and may promote X chromosome inactivation, but also localizes to autosomal telomeres and to nontelomeric regions of the genome (Schoeftner and Blasco 2008; Zhang et al. 2009; Chu et al. 2017).

Silencing of certain repetitive regions of the genome such as pericentromeres has been shown to require transcription of long noncoding transcripts that interact with homologous small interfering RNAs (siRNAs) that are bound to Argonaute proteins (Reinhart and Bartel 2002; Holoch and Moazed 2015). siRNAs can also promote mRNA destruction or translational repression in the cytoplasm.

Endogenous small RNA pathways have been well characterized in the nematode worm *Caenorhabditis elegans* (*C. elegans*) (Billi et al. 2014). Primary siRNAs can be created by DCR-1, the *C. elegans* homolog of the highly conserved RNase class III enzyme Dicer, which cleaves double-stranded RNA transcripts to create 26 nt long siRNAs beginning with a 5' guanine nucleotide (26G RNAs) (Bernstein et al. 2001; Ketting et al. 2001; Knight and Bass 2001). These primary siRNAs are loaded onto Argonaute proteins that bind to a target transcript through complementary base-pairing, and then recruit RNA dependent RNA polymerases to promote formation of 22 nt long secondary RNAs with a 5' guanine (22G RNAs), which bear perfect complementarity to their targets. A distinct class of germline primary siRNAs termed piRNAs are 21 nt small RNAs with a 5' Uracil that are created from thousands of loci in the genome. piRNAs contribute to genome defense by silencing non-self transcripts such as transposons (Batista et al. 2008; Das et al. 2008; Wang and Reinke 2008; Shi et al. 2013). piRNAs interact with the conserved Argonaute PIWI and can bind target RNAs with up to three mismatches, thereby triggering by recruitment of RDRPs and formation of 22G RNAs that are perfectly homologous to their targets and promote genomic silencing, similar to secondary siRNA formation in response to Dicer-dependent 26G primary siRNAs (Ruby et al. 2006; Shen et al. 2018; Zhang et al. 2018).

Consistent with a potential role of TERRA in telomere maintenance, siRNAs with homology to telomeres have been observed in a variety of organisms, where they may contribute to telomere maintenance. Small RNAs produced from telomere repeat-containing transcripts were found to promote DNA methylation at *Arabidopsis* telo-

meres (Vrbsky et al. 2010). Telomeric small RNAs have also been detected in mouse embryonic stem cells (Cao et al. 2009). Telomere dysfunction can induce production of telomeric siRNAs with perfect homology to the vertebrate telomere repeat sequence (TTAGGG)_n, which are created by dsRNAs that promote the DNA damage response (Rossiello et al. 2017). However, the origin of small RNAs with homology to telomeres that are present under standard growth conditions of wild-type metazoan cells is not well understood.

Interstitial telomere sequences (ITSs) are degenerate telomeric sequences that are scattered along chromosome arms and are distinct from chromosome termini. A pioneering study by Meyne and coworkers identified cytologically visible ITSs in 55 different vertebrate species using fluorescence in situ hybridization (FISH) (Meyne et al. 1990; Lin and Yan 2008). Azzalin et al. (2001) identified a number of ITS sites in the human genome by cloning methods and found that ITSs could be divided into three categories representing different putative mechanisms of origin: short, subtelomeric, and fusion. Short ITSs are 24–130 bp in length and display few mismatches with perfect telomere sequence. The authors suggested that these ITSs may be formed during DNA double-strand break repair. Fusion ITSs are likely a consequence of ancestral head-to-head telomere fusion events. Notably, only a single fusion-type ITS was identified in the human genome at 2q13, which led to the creation of chromosome 2 from two ancestral Hominid chromosomes (IJdo et al. 1991).

A number of studies have suggested a link between ITS sites and genome instability (Ashley and Ward 1993; Kilburn et al. 2001; Lin and Yan 2008; Bosco and de Lange 2012). High mutation rates were observed at a region of telomeric sequence that was inserted into a yeast chromosome close to the left telomere (Aksenova et al. 2013; Moore et al. 2018). Intriguingly, the types of mutation observed depended on the orientation of the telomeric sequence, which can lead to gross chromosomal translocations and inversions or repeat loss or gain (Aksenova et al. 2015). Although ITSs can promote genomic instability, Shelterin components in mammals bind ITSs where they may act to stabilize these sequences by preventing the formation of secondary structures that interfere with DNA replication (Yang et al. 2011).

Inactivation of telomerase in *C. elegans* causes progressive sterility in the vast majority of animals. However, a small number manage to reproduce indefinitely in the absence of telomerase due to activation of ALT. In an analysis of stable *C. elegans* ALT strains, several strains were shown to have amplified a region of DNA between a telomere and an ITS sequence, which had then been trans-duplicated to the ends of the other chromosomes, where it acted as a surrogate telomere (Seo et al. 2015). This event is believed to have initiated from recombination between

a critically shortened telomere and the ITS site, indicating that the high degree of similarity of ITS sites and telomeres has the potential to contribute to tumorigenesis and genome evolution by promoting a form of ALT-mediated telomere maintenance. However, despite the presence of ITSs in diverse species, the evolutionary origin and biological significance of these sequences remains mysterious.

Due to the combination of a wide availability of sequencing data and a genome that has been fully sequenced and assembled end-to-end, *C. elegans* provides a highly convenient model system in which to study small RNAs and genome biology. In this study, we analyzed published *C. elegans* small RNA-seq data sets and found numerous examples of small RNA species that map to telomere repeats with up to three mismatches. We found that the majority of these small RNAs map to ITS sites, which are enriched for repressive chromatin marks and found mainly in the introns of protein-coding genes. We show that the closely related nematodes *Caenorhabditis briggsae* and *Caenorhabditis remanei* possess less abundant and generally more degenerate ITS sites than those in *C. elegans*, but display a substantially higher abundance of small RNAs that map perfectly to telomeres. We speculate that telomeric small RNAs may represent part of a functional pathway that has been significantly altered in *C. elegans* but retained in closely related species.

RESULTS

Identification of small RNAs with high levels of telomere homology

We interrogated 75 publicly available small RNA sequencing libraries created from RNA isolated from various *C. elegans* strains and found rare reads that mapped perfectly to the *C. elegans* telomere repeat (TTAGGC)_n, occurring at a frequency of one per 10,000,000 reads (Supplemental Table S9; Gu et al. 2009; Gent et al. 2010; Fischer et al. 2011; Seth et al. 2013; Phillips et al. 2014, 2015; Sapetschnig et al. 2015). As piRNAs with up to three mismatches can target regions of the *C. elegans* genome for silencing (Bagijn et al. 2012), we counted telomeric siRNAs with up to three mismatches and found that these were substantially more abundant than those composed of perfect telomere repeat sequence (1 per 100,000) (Fig. 1A). In total, 2352 distinct telomeric small RNA species

were detected. Telomeric siRNAs with zero to two mismatches were significantly more likely to be composed of the telomeric G-strand sequence than would be expected by chance ($P < 2.2 \times 10^{-16}$, binomial test) (Fig. 1D). The proportion of telomeric siRNAs with one to three mismatches beginning with a G nucleotide was significantly greater than that for all siRNAs that could be mapped to the genome ($P < 1 \times 10^{-15}$, Fisher's exact test for all comparisons) and the median length for telomeric siRNAs with 1–3 mismatches RNAs was 22 nt (Fig. 1B, C). Therefore, telomeric siRNAs with 1–3 mismatches are enriched for the effector 22G RNA species, which has been established to promote either posttranscriptional or genomic silencing depending on the Argonaute they associate with (Billi et al. 2014). In contrast, rare telomeric siRNAs with 0 mismatches (perfect telomeric siRNAs) were not enriched for 22G RNAs but displayed peak lengths of 18 and 23 nt.

Although the frequency of perfect telomeric siRNAs that we observed in *C. elegans* was rare, only one read in 10 million, the number of possible sequences for a 20mer is 4^{20} , or approximately 1×10^{12} . Given that there are 12 possible perfect telomeric 20mers (six possible starting bases, multiplied by two for sense/antisense), the expected frequency of a 20mer being a perfect telomere sequence comes to approximately 1 in 1×10^{11} , which is considerably lower than the frequency of 1 in 10^7 that we observed. To further investigate the prevalence of perfect telomeric small RNAs, we ran 1000 simulations in which 10,000,000 random

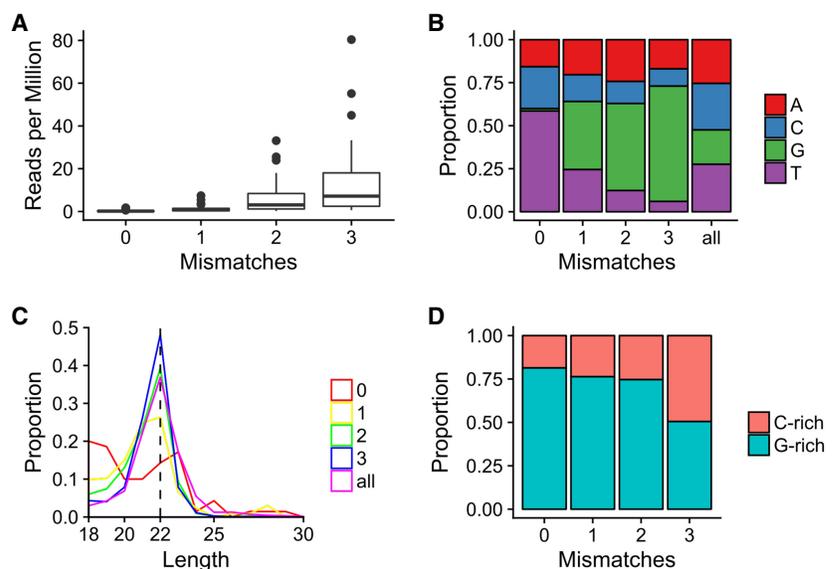


FIGURE 1. Identification of telomeric small RNAs in *C. elegans*. (A) Quantification of small RNA species in published data sets that map to telomere sequence with different numbers of mismatches. (B) Proportion of telomere-mapping small RNA reads in published data beginning with a particular nucleotide. (C) Distribution of telomeric small RNA length. (D) Proportion of telomeric small RNA reads mapping to the G-rich (TTAGGC) or C-rich (GCCTAA) telomeric strand.

nucleotide sequences were generated. The length of each sequence was determined probabilistically based on the length distribution of all the reads in the libraries we analyzed. We did not detect any perfect telomere sequences in any of the simulations. We therefore conclude that the low numbers of perfect telomeric reads we observed are not due to chance.

Regarding telomeric siRNAs with mismatches, there are 1140 combinations of three positions in a 20mer. At each mismatch, there are four possible bases, so we multiplied 1140 by 4 to get 4560. Multiplying that value by our probability of $1/(4^{20})$ for finding a specific 20mer comes to $1/2.41 \times 10^8$. Dividing this probability by 12, as above, resulted in a probability of finding a telomeric siRNA with a mismatch to be approximately 1 in 20,000,000. We ran a similar simulation to the one above but allowed up to three mismatches and only simulated 100,000 reads each time. Out of 1000 simulations, three showed a frequency of at least one telomeric sequence per 100,000 sequences, giving a *P*-value of 0.003. We therefore conclude that the observed number of telomeric small RNAs with up to three mismatches is greater than expected by chance. Given the high number of possible telomeric sequences with up to three mismatches allowed (656,640), it is interesting that we only found 2352 distinct species. This is however consistent with the hypothesis that telomeric small RNAs are largely produced from a relatively small number of genomic loci (see below).

ITS sites explain the origin of most telomeric small RNAs

Telomeric small RNAs could be divided into those that map to unique locations in the genome, which amounted to 10% and 57% of telomeric small RNAs with two or three mismatches, respectively, and those that map to multiple locations, which amounted to 100% of telomeric small RNAs with zero or one mismatches and 90% and 33% of telomeric small RNAs with two or three mismatches, respectively. The overall mapping frequency of telomeric small RNAs compares favorably with mapping frequency of all reads in a small RNA library, which generally amounts to 75%–80% in our experience.

As unique telomeric small RNA reads mapped to loci that were scattered along chromosome arms, we reasoned that these loci may represent ITS sites that were previously shown to be concentrated on *C. elegans* chromosomes arms, based on the Ce000094 repeat sites that were automatically detected using the RECON algorithm (Bao and Eddy 2002; Lowden et al. 2011) (<https://wormbase.org/>). Although the Ce000094 motif corresponds to telomeric DNA, we later noticed that there were many ITS sites that were covered only partially by this motif or not at all. This is likely a consequence of the degeneracy of ITS sites, which makes them challenging to detect de novo using au-

tomated algorithms. In order to more thoroughly map ITS sites genome-wide, we located all instances of the telomeric hexanucleotide “TTAGGC” in the genome. We sought to capture perfect runs of telomere repeats as well as more degenerate sequences consisting of clusters of perfect repeats interspersed with nontelomeric sequence. TTAGGC hexanucleotides located on the same strand and within 100 nt of each other were merged to form ITS contigs, and contigs that contained at least four perfect telomeric hexanucleotides were retained, resulting in 1229 ITS sites (Supplemental File 1).

ITS lengths ranged from 24 to 1204 nt with a median length of 134 nt (Fig. 2F). ITS sites were found to be largely localized to the autosome arms, which have been shown to be largely heterochromatic and enriched for repetitive sequences (Fig. 2A; Garrigues et al. 2015). There was a five-fold depletion of ITS sites on the X chromosome (Fig. 2B), consistent with lower levels of constitutive heterochromatin on this chromosome and an enrichment for Polycomb marks that silence the X chromosome in germ cells (Gaydos et al. 2012).

Although the mechanism by which ITS sites are created is unclear, these could be created by translocation of a telomere to an internal segment of the genome or by telomerase activity that acts at a double-strand break. Therefore, ITS sites are likely to be initially composed of perfect telomere repeats that degenerate over evolutionary time scales, such that they accumulate mutations in the absence of selective pressure. To obtain an estimate of the relative ages of ITS sites, we assigned a “homology score” to each site by performing local alignment between the ITS site and perfect telomere sequence of the same strand and length. The homology score was defined as the percent sequence identity between the two sequences. Homology scores ranged from 28.6 to 100, with a median score of 84.2 (Fig. 2G; Supplemental File 1). Just over half (650) of ITSs had two or fewer consecutive perfect telomere repeats, although 25 ITS sites contained 7–15 consecutive perfect telomere repeats (Table 1). We found no obvious genome localization patterns in terms of homology score or sequence length. ITS sites therefore represent abundant repetitive sequences displaying a broad range of degeneracy (and therefore age) and genome localization patterns consistent with other types of repetitive sequence (Ikegami et al. 2010; Garrigues et al. 2015).

The origin of ITS sites is poorly understood. One proposed mechanism for ITS formation is the fusion of two chromosomes, although there currently exists evidence for only one such event in the human genome (IJdo et al. 1991). We sought to determine whether ancestral chromosome fusion events could account for the presence of ITSs in the *C. elegans* genome. However, we could find no example of an ITS site arranged head-to-head, and therefore conclude that ITS sites in the *C. elegans* genome are unlikely to arise as a result of chromosome fusions.

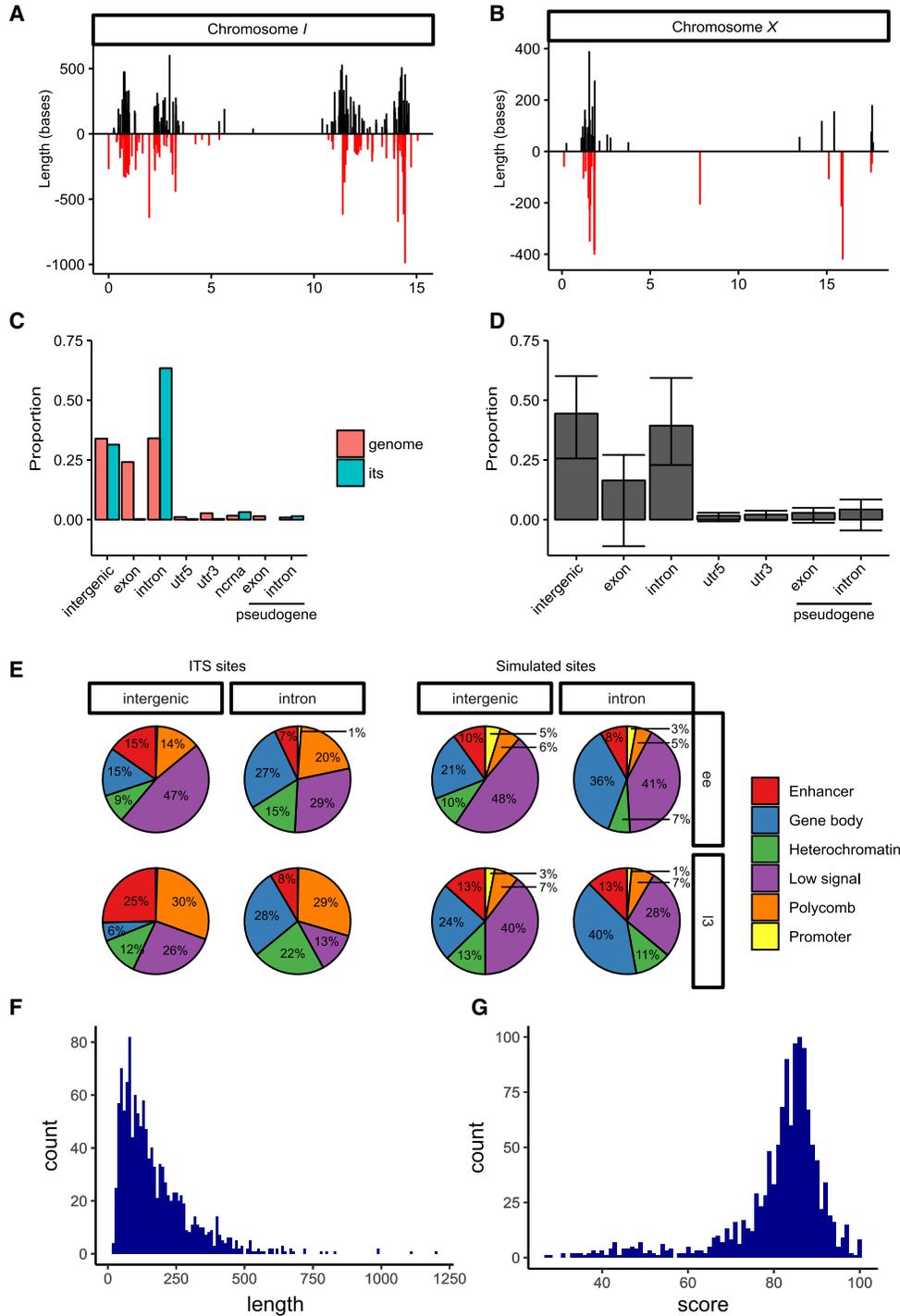


FIGURE 2. Characterization of ITSs in *C. elegans*. (A) Localization of ITS sites on chromosome I regions in the sense orientation (TTAGGC) are shown in black, antisense in red. (B) Localization of ITS sites on the X chromosome. (C) Proportion of ITS sites overlapping different genome features (blue bars) and amount of genomic space taken up by genomic features (pink). (D) Average proportion of features overlapped by hexanucleotide repeat sequences from 1000 simulations. Error bars show standard deviation. (E) Percentage of real or permuted (“simulation”) ITS sites that are covered by a particular chromatin state. Distributions of ITS length (F) and homology score (G).

To determine whether ITS sites are a source of telomeric small RNAs, we mapped all small RNAs to the genome using a pipeline that assigns multimapping reads probabilistically based on the local density of uniquely mapped

reads (Axtell 2014). 76%, 91%, 83%, and 68% of small RNAs mapping to telomeres with zero, one, two, and three mismatches, respectively were assigned to ITS sites (Table 2), whereas 92% and 75% of uniquely mapping small RNA

TABLE 1. Tally of the maximum number of consecutive perfect telomere repeats in ITS sites

Species	Maximum consecutive perfect repeats																				
	1	2	3	4	5	6	7	8	9	10	12	13	14	15	16	17	18	21	28	29	38
<i>C. elegans</i>	214	436	344	140	53	17	6	5	3	4	1	2	3	1	0	0	0	0	0	0	0
<i>C. briggsae</i>	344	15	0	0	2	2	1	1	0	2	1	0	1	0	0	1	1	0	0	0	0
<i>C. remanei</i>	164	99	7	0	1	2	2	1	1	1	1	0	1	1	1	1	1	1	1	1	1

reads with two or three mismatches, respectively, could be assigned to ITS sites. Although it is impossible to determine whether perfect telomeric small RNAs originate from telomeres or internal regions with certainty, these results suggest that such small RNAs are more likely to be produced from internal genomic regions. This is because many ITS sites are in heterochromatic segments of the genome and are surrounded by high densities of siRNAs, whereas few siRNAs map closely to subtelomeres. Furthermore, we looked for uniquely mapping reads that overlapped with regions of perfect telomere repeat sequence of at least 30 bases in length, of which there were 95 in ITSs, and found 18 reads within internal regions, while only three reads could be found that overlapped telomere–subtelomere boundaries. As each ITS tract has two borders, this means that 18 small RNAs were present at these 190 ITS tract borders (1%), whereas three siRNAs were present at 12 telomere–subtelomere boundaries (25%). These results suggest that perfect telomeric small RNAs could originate from either telomeres or ITS sites.

We asked whether telomeric small RNA abundance varied across ITS sites. Due to the uncertainties associated with multimapping reads, we decided to focus on the uniquely mapping subset of telomeric small RNAs with two or three mismatches. We could find no strong correlation between telomere homology score and small RNA mapping frequency for ITS (Spearman's $\rho = 0.068$), although the ITS sites with the highest telomeric small RNA abundance had homology scores of around 80% and no telomeric small RNAs were found at sites with scores <53.8% (Fig. 3A). For example, all ITS tracts with >10 telomeric small RNAs had homology scores of >76% (Table 3). The abundance of uniquely mapping telomeric

small RNAs correlated well with total small RNA abundance at an ITS tract (Spearman's $\rho = 0.70$), indicating that ITS sites that produce high numbers of telomeric small RNAs are likely to be sites of high small RNA production in general. Of the top 10 ITSs with the highest overall small RNA abundance, seven also displayed high telomeric small RNA abundance. The remaining three ITSs that had no telomeric small RNAs assigned had homology scores of <56%. We found one ITS site (ITS_60) which displayed an unusually high frequency of small RNA mapping that was five- to 13-fold greater than other ITS tracts with abundant siRNAs and accounted for 43% and 18% of total telomeric small RNAs with two and three mismatches, respectively (Table 3; Fig. 3). This ITS is intergenic and fully overlaps two annotated converging noncoding RNAs, whose transcripts could anneal to produce dsRNA that can be processed into abundant ITS siRNAs. (Fig. 3B). As telomeric small RNAs did not map exclusively to ITS sites, we also looked at the genome-wide distribution of telomeric small RNAs (Fig. 3C). We found three peaks of high coverage that did not overlap any ITS sites. Two of these, located on the left arms of chromosomes III and V, contained two consecutive perfect telomeric repeats and were assigned a number of reads with two or three mismatches to telomeres, but none with zero or one. The other peak, located on the right arm of chromosome I, contained three consecutive perfect telomere repeats next to a degenerate one, narrowly missing the criteria for being detected as an ITS site. Therefore, telomeric small RNAs that do not originate from ITS sites are likely to originate from non-ITS regions that happen to have short stretches of telomeric DNA.

ITS sites are mainly found in introns and are enriched for repressive chromatin marks

We looked for overlap between ITS sites and annotated genomic features and found that they were enriched in introns and in intergenic sequences. Surprisingly, 796 ITS sites (63% of total) were located in introns of protein-coding genes (Fig. 2C). In terms of sequence coverage, this represents almost a twofold enrichment for ITS sequence at introns based on the proportion of the genome that is comprised of intronic sequence. In contrast, only two ITS sites (0.16%) were found in protein-coding exons,

TABLE 2. Summary of unique telomeric small RNA reads

Mismatches	Reads	ITS	% ITS assignment
0	29	23	79.3
1	182	167	91.8
2	769	640	83.2
3	1372	933	68.0

Percentage of reads assigned mapping to ITS sites is shown.

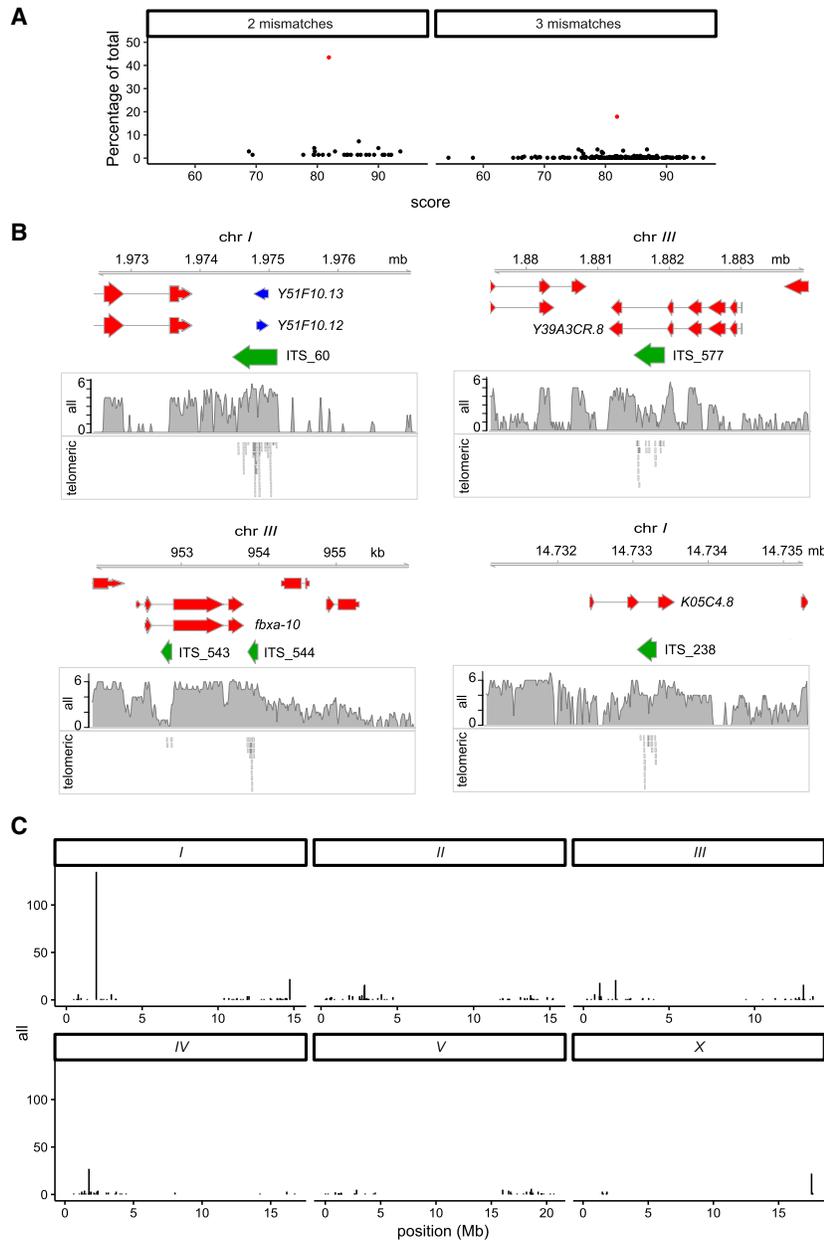


FIGURE 3. Telomeric small RNAs map to ITS sites. (A) Scatter plots showing ITS score versus uniquely mapping small telomeric small RNAs count as a percentage of total abundance for each category for ITS sites with at least one small RNA read. ITS_60 is shown in red. (B) Genome track showing the top four ITSs with the highest uniquely mapping telomeric small RNA count. Total small RNA coverage and telomeric small RNA reads are shown as separate tracks (C) Genome-wide distribution of ITS sites. Y-axis shows uniquely binding telomeric small RNA count.

indicating that ITS sites are depleted from exome, possibly because they are deleterious for mRNA expression. Consistently, ITS sites were also significantly depleted from both 5'- and 3'-UTRs. Such a pattern could be a general feature of blocks of DNA comprised largely of degenerate hexanucleotide repeats. To test this, we repeated the ITS detection pipeline with 100 random hexanucleotide combinations. For each hexamer, we defined degenerate

repeat sites using the same method used to detect ITS sites and looked for overlap between these sites and annotated genomic features. The simulated repeat sites did not show an overall enrichment at introns or depletion at exons (Fig. 2D), indicating that this pattern of localization is specific to ITS sites. Genes containing intronic ITS sites were not significantly enriched for any GO terms.

Features of eukaryotic chromatin such as histone modifications and nucleosome occupancy drive spatio-temporal regulation of genomic regions. Ho et al. (2014) used a hidden Markov model based algorithm to classify chromatin into a series of "states" based on the integration of various CHIP-seq and microarray data sets. The 16 states were grouped into six categories: promoter, enhancer, gene body, Polycomb-repressed, heterochromatin, and weak or low signal. We assigned chromatin states to each ITS sequence using the *C. elegans* data set. To determine statistically significant enrichment for chromatin states, we performed 1000 simulations in which a new set of 1219 sequences with the same length distribution as the ITS sequences was randomly selected from the genome and assigned chromatin states. Both intronic and intergenic ITS sites were enriched two- to fourfold for Polycomb-repressed chromatin states in both the early embryo and at the L3 larval stage of development relative to the corresponding simulated sites ($P < 0.0003$ for all classes), with the highest enrichment in the L3 stage (Fig. 2E; Supplemental Table S1). Intronic but not intergenic ITS sites were enriched twofold for heterochromatin chromatin states ($P = 0.008$ and 0.02 for early embryo and L3, respectively) and were depleted for enhancer chromatin states. The chromatin state of intronic ITSs matched that of the nearest exon in over 99% of cases for both early embryo and L3, indicating that the chromatin environment of intronic ITSs is almost always shared by the entire gene. Intergenic ITS sites were enriched 1.5-fold and twofold for enhancer states in early embryo and L3 larvae, respectively ($P < 0.00048$, $P = 0.0003$ for L3 larvae and early embryo, respectively). We

TABLE 3. ITS sites with more than 10 telomeric small RNAs

ID	Chromosome	Start	End	Length	Score	Type	Telomeric small RNAs		
							2	3	All
ITS_60	I	1,974,471	1,975,116	645	81.9	Intergenic	30	105	135
ITS_813	IV	1,752,185	1,752,299	114	86.8	Intergenic	5	22	27
ITS_238	I	14,733,054	14,733,310	256	78.7	Intron	0	22	22
ITS_1226	X	17,527,653	17,527,730	77	75.6	3'-UTR	0	22	22
ITS_576	III	1,881,503	1,881,935	432	82.9	Intron	2	19	21
ITS_544	III	953,856	953,988	132	76.1	Intergenic	0	18	18
ITS_316	II	2,859,211	2,859,720	509	79.5	Intergenic	3	13	16
ITS_739	III	12,856,138	12,856,366	228	79.4	Intron	0	16	16
ITS_315	II	2,847,764	2,848,131	367	79.6	Intron	2	13	15
ITS_314	II	2,846,566	2,846,973	407	76.4	Intron	0	11	11

looked at the genes most proximal to intergenic ITSs with enhancer chromatin states but found no significantly enriched GO terms among these genes. Both intronic and intergenic ITS sites were strongly depleted for promoter chromatin states at both stages ($P < 0.001$ in each case). We found no significant relationship between chromatin state and proportion of ITSs that produce at least five small RNAs ($P = 0.1$, χ^2 test for both early embryo and L3 stages). We conclude that ITS sites are generally associated with repressive chromatin, that intronic ITS sites may promote repressive chromatin formation via heterochromatin or Polycomb pathways, but that some intergenic ITS sites may possess enhancer function.

C. briggsae* and *C. remanei* ITS sites are less common and more degenerate than those of *C. elegans

We sought to gain insights into the evolutionary conservation of ITS sites and the relationship between ITS sites and small RNA production in nematodes. *C. briggsae* is one of the closest known living relatives of *C. elegans*, and diverged from *C. elegans* between 18–100 million years ago (Stein et al. 2003; Cutter 2008). It is physically virtually indistinguishable from *C. elegans*. *C. remanei* is another close relative of *C. elegans*. Although *C. briggsae* and *C. remanei* may have shared a recent common ancestor that diverged from *C. elegans*, *C. remanei* reproduces by gonochoristic male–female sexual reproduction, whereas *C. elegans* and *C. briggsae* are both hermaphroditic selfers (Fierst et al. 2015).

We applied the same method for detecting ITS sites in *C. briggsae* and *C. remanei* and found only 372 and 292 sites, respectively, less than one third of the number found in *C. elegans* in each case. ITS sites in *C. briggsae* and *C. remanei* were significantly more divergent than those from *C. elegans*, with median homology scores of 49.5 and 57.1 ($P < 2.2 \times 10^{-16}$, Mann–Whitney *U*-test for both

comparisons) (Fig. 4A), while the median length for ITS sites was slightly higher for *C. briggsae* and *C. remanei* at 206 ($P < 2.2 \times 10^{-16}$, Mann–Whitney *U*-test) and 178 nt ($P = 6.058 \times 10^{-09}$, Mann–Whitney *U*-test), respectively (Fig. 4B). Interestingly, although 93% and 57% of ITSs in *C. briggsae* and *C. remanei* contained no consecutive telomere repeats, both species possessed a handful of 36–228 nt ITSs with perfect telomere homology (Table 1; Supplemental File 1). Although *C. remanei* had the longest perfect telomere tracts of the three species, *C. remanei* and *C. briggsae* possessed few ITS tracts with long stretches of perfect telomere repeats in comparison to *C. elegans* (Table 1). All three nematode genomes showed an intron bias for ITS sites, which in comparison to *C. elegans* was stronger for *C. briggsae* and weaker for *C. remanei*. (Fig. 4C,D). It was not possible to determine an enhancer bias, because the other genomes have not been annotated for this feature.

To gain insights into the evolutionary history of ITSs across different species, we looked for intronic ITSs that are shared within orthologous genes. We could not identify genes for which orthologues between all three species contained ITS sites. We performed pairwise comparisons between the three nematode species and found a handful of genes in each case for which orthologous genes between the two species contained ITS sites in both species, which we termed “shared” ITS sites as opposed to “unique” sites (Fig. 4E; Supplemental Table S2). We found that the median score was slightly but significantly lower for shared sites compared with unique sites in *C. elegans* (78.7 versus 84, $P = 0.008443$, Mann–Whitney *U*-test), although no significant difference was noted for *C. briggsae* or *C. remanei* (Supplemental Fig. S1). Notably, there were no shared ITS sites with scores $>94.4\%$, consistent with unique ITS sites having been formed more recently in evolution. We performed GO term analysis on *C. elegans* genes orthologous to *C. briggsae* and *C. remanei* genes

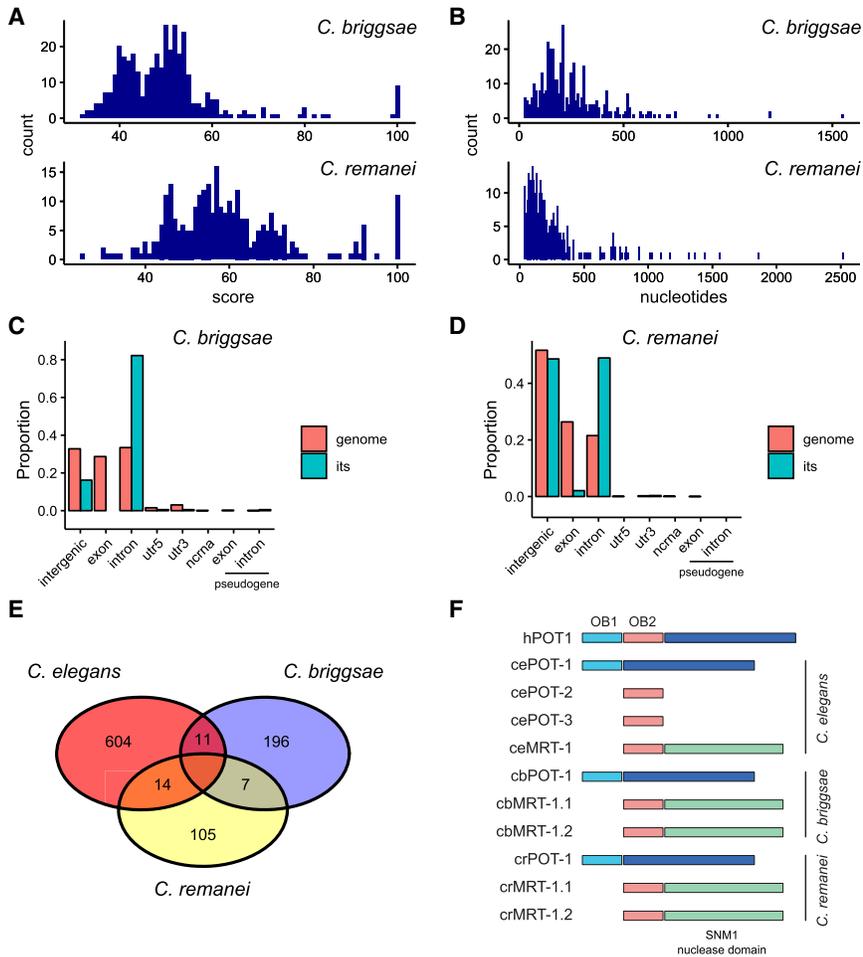


FIGURE 4. ITS sites in *C. briggsae* and *C. remanei*. (A) Distribution of ITS scores. (B) Distribution of ITS lengths. (C) Proportion of ITS sites overlapping different genome features (blue bars) and amount of genomic space taken up by genomic features (pink). (D) As (C) but for *C. remanei*. (E) Number of genes for which orthologous genes in two species containing intronic ITS sites in both species. (F) Domain structures of *C. elegans*, *C. briggsae*, and *C. remanei* human POT1 homologs.

that contain ITSs. The *C. briggsae* gene set was enriched for GO terms related to reproduction cell cycle and organelle organization. No significant GO terms were enriched in the *C. remanei* gene set. Therefore, the mechanism of ITS formation is ancient in origin but ITSs themselves are divergent among nematode species, suggesting that a possible function of ITS tracts in regulation of chromatin if present diverges rapidly between species.

Telomeric small RNAs are more abundant in *C. elegans* relatives

We then interrogated small RNA libraries in *C. briggsae* and *C. elegans* (Shi et al. 2013) and looked for small RNAs that map to telomeres. The same data set contained a *C. elegans* library, which we also analyzed as a control. Consistent with our previous analysis, reads mapping perfectly to telomeres were exceedingly rare in the *C. elegans*

libraries; three were detected in hermaphrodites, one in embryos and none in males, corresponding to a total of around one read per 10,000,000 small RNA reads. Strikingly, libraries from *C. briggsae* and *C. remanei* contained several orders of magnitude more perfect telomeric small RNAs compared with *C. elegans* (Fig. 5A). Perfect telomeric small RNAs were detected at an abundance of more than 22 per million in *C. briggsae* embryos and mixed male/female *C. remanei* adults. The *C. elegans* perfect telomeric small RNAs detected in control libraries from the *briggsae* and *remanei* data sets showed the same G-rich strand bias as found in the libraries described earlier. However, the *C. briggsae* and *C. remanei* libraries showed a strong bias for the C-rich strand bias for siRNAs composed of perfect telomere repeats (Fig. 5B). Despite this strong distinction in strand bias, perfect telomeric small RNAs from all three species showed a striking depletion of perfect telomeric small RNAs species that begin with a 5' G, whereas abundant levels of small RNAs with a 5' guanine were observed for telomeric siRNAs with mismatches.

If the perfect telomeric siRNAs are split into G- and C-strand categories, the G-rich RNAs from all three species were enriched for RNAs beginning with 5' UAGGCU or 5' UUAGGC, ranging in size from 18–29 nt. The most abundant sRNAs from all three species were 23 nt RNA UUAGGCUUAGGCUUAGGCUUAGG and 22 nt RNA UAGGCUUAGGCUUAGGCUUAGG (Fig. 6). On the other hand, C-rich telomeric RNAs of *C. briggsae* and *C. remanei* species showed enrichment for 22 nt perfect telomeric small RNAs beginning with a C (Fig. 5C,D) of 30% and 32%, respectively. Most perfect C-rich telomeric RNAs from both *C. briggsae* and *C. remanei* began with 5' CUAAGC and ranged from 18–24 nt in length, with the most abundant C-rich RNA being the 22 nt sequence CUAAGCCUAAGCCUAAGCCUAA, which comprised 22% and 29% of all sequences for *C. briggsae* and *C. remanei*, respectively (Supplemental Table S2). This sequence reflects the majority of 22 nt perfect telomeric small RNAs for each species (52% and 57%, respectively). Given that there are only six permutations of a perfect telomeric sequence of a given length, each permutation would be expected to occur at a probability of 1/6 if

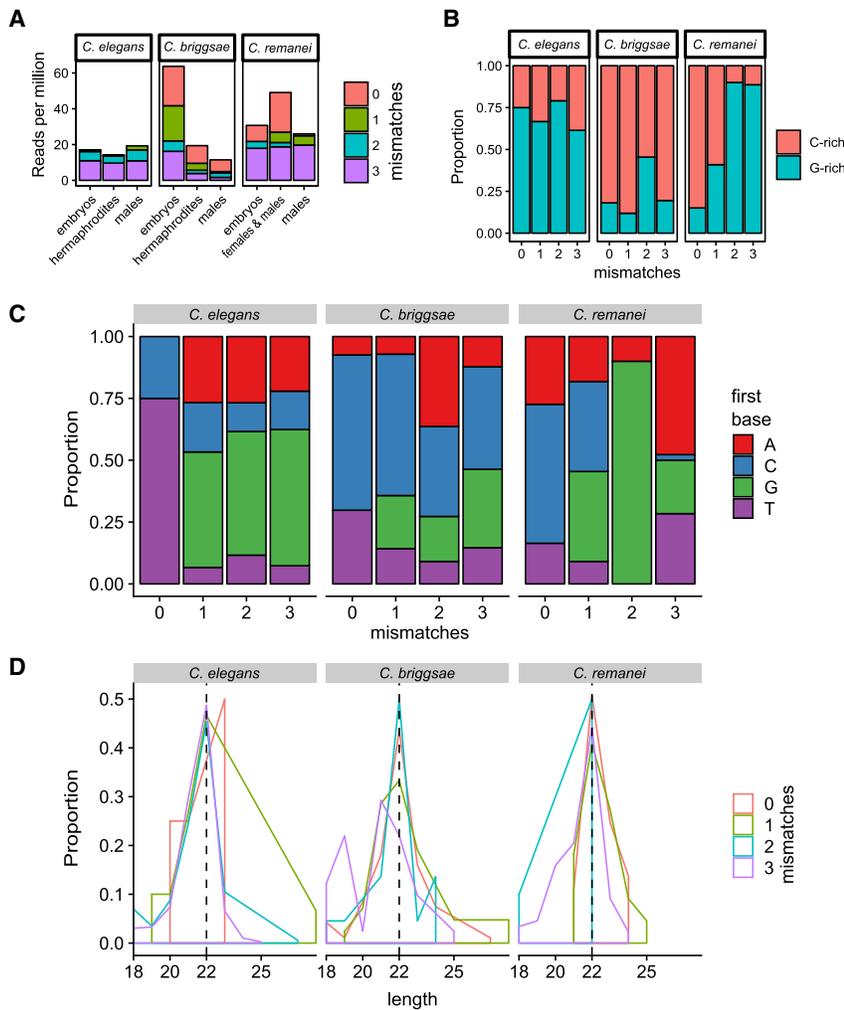


FIGURE 5. Comparison of telomeric small RNAs between three *Caenorhabditis* species. (A) Telomeric small RNA abundance in small RNA sequencing libraries (Shi et al. 2013) from three *Caenorhabditis* species. (B) Proportion of telomeric small RNA reads mapping to the G-rich (TTAGGC) or C-rich (GCCTAA) telomeric strand. (C) Proportion of telomere-mapping small RNA reads in published data beginning with a particular nucleotide. (D) Distribution of telomeric small RNA length.

the sequence distribution was random. We conclude that the general depletion of 5'G nucleotides from perfect telomeric siRNAs can be largely explained by the fact that G-rich RNAs commonly begin with 5' UAGGCU and 5' UUAGGC, whereas C-rich RNAs commonly begin with 5' CUAAGC (Fig. 6; Supplemental Table S2).

Due to the relatively small number of available small RNA libraries for *C. briggsae* and *C. remanei* and the fact that telomeres have not been assembled in the reference genomes for these species, we could not make inferences about whether telomeric small RNAs were more likely to originate from telomeres or ITSs based on siRNAs that map to borders of perfect telomere repeat tracts. If perfect telomeric small RNAs originate from ITS sites, an increase in abundance for these RNAs could be due to an increase in the amount of ITS sequence that is capable of producing

them. However, the *C. elegans* genome contains a combined total of 3666 bases of perfect telomere existing in stretches of at least 30 nt in length, compared with 1596 for *C. briggsae* and 708 for *C. remanei*, indicating that increased levels of perfect telomeric ITS sequence cannot account for the increase in perfect telomeric small RNA abundance. We conclude that the composition of telomeric small RNAs in *C. elegans* has diverged substantially from its close relatives, but that both classes of small RNAs are dramatically depleted for 5' G RNA species, which could reflect either a toxicity of this species or its mechanism of biogenesis.

Argonaute proteins can promote the biogenesis of specific classes of small RNAs. For example, ALG-3 and ALG-4 associate with 26G RNAs and PRG-1/Piwi associate with 21U RNAs, whose levels are dramatically reduced in response to deficiency for *prg-1* (Ruby et al. 2006; Batista et al. 2008; Das et al. 2008; Han et al. 2009; Vasale et al. 2010). We studied small RNAs that immunoprecipitate with Argonaute proteins PRG-1, WAGO-1, HRDE-1, and CSR-1 (Batista et al. 2008; Das et al. 2008; Claycomb et al. 2009; Gu et al. 2009; Ashe et al. 2012; Buckley et al. 2012). Perfect telomeric small RNAs were exceedingly rare in the input RNA data sets, as expected, and we found no enrichment for C- or G-strand telomeric RNAs in any IP data set

(Supplemental Table S3). However, these data sets revealed the only three *C. elegans* examples that we observed of the major C-strand RNA species found in *C. briggsae* and *C. remanei*, 23 nt 5' CUAAGC RNA. The IP data sets brought the total number of *C. elegans* C-rich telomeric RNAs to 14 (Fig. 6; Supplemental Tables S2, S3). Significant numbers of telomeric small RNAs with mismatches were associated with PRG-1, WAGO-1, HRDE-1, and CSR-1 Argonaute proteins.

Evolution of telomeric small RNAs

Perfect telomeric siRNAs have been previously reported to associate with the *Tetrahymena* Argonaute protein Twi10, were of lengths 23–24 nt and 33–36 nt and all possessed the sequence 5' UGGGGU (Couvillion et al. 2009),

A	perfect telomeric G-strand sRNAs			B	perfect telomeric C-strand sRNAs		
	<i>elegans</i>	<i>briggsae</i>	<i>remanei</i>		<i>elegans</i>	<i>briggsae</i>	<i>remanei</i>
29 nucleotides							
UAGGCUUAGGCUUAGGCUUAGGCUUAGGC	1	0	0				
28 nucleotides							
UAGGCUUAGGCUUAGGCUUAGGCUUAGG	1	0	0				
27 nucleotides							
UAGGCUUAGGCUUAGGCUUAGGCUUAG	1	1	0				
25 nucleotides							
UAGGCUUAGGCUUAGGCUUAGGCUU	2	0	0				
CUUAGGCUUAGGCUUAGGCUUAGGC	1	0	0				
24 nucleotides							
UUAGGCUUAGGCUUAGGCUUAGGC	1	0	0	24 nucleotides			
				CUAAGCCUAAGCCUAAGCCUAAGC	0	6	10
				UAAGCCUAAGCCUAAGCCUAAGCC	0	1	0
23 nucleotides				23 nucleotides			
UUAGGCUUAGGCUUAGGCUUAGG	5	4	8	CUAAGCCUAAGCCUAAGCCUAAG	3	5	3
UAGGCUUAGGCUUAGGCUUAGGC	1	1	1	AAGCCUAAGCCUAAGCCUAAGCC	0	0	6
				UAAGCCUAAGCCUAAGCCUAAGC	1	5	0
				CCUAAGCCUAAGCCUAAGCCUA	2	0	0
22 nucleotides				22 nucleotides			
UAGGCUUAGGCUUAGGCUUAGG	5	7	1	CUAAGCCUAAGCCUAAGCCUA	1	21	21
UUAGGCUUAGGCUUAGGCUUAG	1	3	0	AAGCCUAAGCCUAAGCCUAAGC	1	6	9
AGGCUUAGGCUUAGGCUUAGGC	1	0	0	UAAGCCUAAGCCUAAGCCUAAG	0	3	2
CUUAGGCUUAGGCUUAGGCUUA	1	0	0	CCUAAGCCUAAGCCUAAGCCUA	0	1	2
				AGCCUAAGCCUAAGCCUAAGCC	0	0	2
21 nucleotides				21 nucleotides			
AGGCUUAGGCUUAGGCUUAGG	5	0	1	CUAAGCCUAAGCCUAAGCCUA	0	16	5
CUUAGGCUUAGGCUUAGGCUU	1	0	0	AAGCCUAAGCCUAAGCCUAAG	1	0	2
UUAGGCUUAGGCUUAGGCUUA	0	1	0				
20 nucleotides				20 nucleotides			
UUAGGCUUAGGCUUAGGCUU	5	0	0	CUAAGCCUAAGCCUAAGCCU	0	7	0
CUUAGGCUUAGGCUUAGGCU	1	0	0	UAAGCCUAAGCCUAAGCCUA	0	1	0
UAGGCUUAGGCUUAGGCUUA	1	0	0				
19 nucleotides				19 nucleotides			
UAGGCUUAGGCUUAGGCUU	6	0	0	CCUAAGCCUAAGCCUAAGC	1	0	0
CUUAGGCUUAGGCUUAGGC	4	0	0	CUAAGCCUAAGCCUAAGCC	0	1	0
UUAGGCUUAGGCUUAGGCU	1	0	0	UAAGCCUAAGCCUAAGCCU	1	0	0
18 nucleotides				18 nucleotides			
UAGGCUUAGGCUUAGGCU	5	0	0	CUAAGCCUAAGCCUAAGC	2	2	0
AGGCUUAGGCUUAGGCUU	3	0	0	AAGCCUAAGCCUAAGCCU	0	1	0
UUAGGCUUAGGCUUAGGC	3	0	0	UAAGCCUAAGCCUAAGCC	0	1	0
GCUUAGGCUUAGGCUUAG	1	0	0				

C	G-strand 23 nt	G-strand 22 nt
	5' UUAGGCUUAGGCUUAGGCUUAGG 3'	5' UAGGCUUAGGCUUAGGCUUAGG 3'
	3' GAAUCCGAAUCCGAAUCCGAAUC 5'	3' AAUCCGAAUCCGAAUCCGAAUC 5'
	C-strand 23 nt	C-strand 22 nt

FIGURE 6. Perfect telomeric small RNAs of the genus *Caenorhabditis*. Total observed G-rich (A) and C-rich (B) telomeric small RNAs from RNA-seq and IP data sets. (C) Abundant telomeric sRNAs can anneal to form duplexes.

potentially similar to the 5' UAGGCU G-rich RNAs that were present in *Caenorhabditis* species. Telomeric siRNAs have been previously reported to be rare in mammals, where they may promote the response to DNA damage at telomeres. We therefore examined ~650,000,000 reads in 77 small RNA libraries from mouse and found 156, 56, 88, and 860 telomeric reads with 0, 1, 2, and 3 mismatches, respectively (Supplemental Tables S5, S6, S9; Pandey et al. 2013; Manakov et al. 2015; Watanabe et al. 2015; Wenda et al. 2017). Perfect telomeric siRNAs occurred at a frequency of ~1 in 4,170,000 reads and were therefore almost as rare as *C. elegans* perfect telomeric siRNAs. G-rich perfect telomeric siRNA were also more common in mammalian telomeric siRNAs, the most abundant of which began with 5' GUUAGG and were 19 and 23 nt in length (Supplemental Tables S7). The most common C-rich telomeric siRNA species from mammals began with

5' UAACCC and was 24, 28, and 30 nt in length (Supplemental Table S8). Other permutations of 5' ends were represented for both G- and C-rich perfect telomeric RNAs from mammals.

We speculated that the striking difference in telomeric small RNA composition between *C. elegans* and its relatives could occur in response to a functional difference in telomere biology between these species. The *C. elegans* shelterin complex contains the OB-fold containing proteins POT-1 and POT-2, and mutation of either *pot-1* or *pot-2* results in successive telomere elongation over multiple generations (Raices et al. 2008; Shtessel et al. 2013). Variants in the *pot-2* gene have also been found to have a prominent effect on telomere length in wild *C. elegans* strains, suggesting that this protein may affect fitness in the wild (Cook et al. 2016). A third *C. elegans* homolog of human Pot1, MRT-1, is required for telomerase activity

(Meier et al. 2009). We performed tBLASTX searches for orthologues of the POT-2 protein in *C. briggsae* and *C. remanei* genomes, respectively. Surprisingly, we found that the reference genomes of *C. briggsae* and *C. remanei* do not contain *pot-2* homologs. Instead, these species each possess two copies of MRT-1, which has an SNM1 interstrand crosslink repair nuclease fused to the POT-2 OB-fold protein (Fig. 4F; Meier et al. 2009). It is possible that the absence of *pot-2* or presence of an extra *mrt-1* paralog in *C. briggsae* and *C. remanei* could represent a fundamental difference in telomere biology that is related to the increased abundance of perfect telomeric small RNAs in these species.

We examined telomeres of several wild isolates of *C. briggsae* and *C. remanei* and found that the AF16 strain that has high levels of perfect telomeric siRNAs had elongated telomeres, whereas the other strains examined had generally short telomeres, similar in size to those of *C. elegans*. However, several *C. briggsae* and *C. remanei* strains possessed weak telomeric DNA signals that failed to migrate into the gel and that ran at limit mobility, which are not observed in the genomic DNA from the Bristol N2 wild-type *C. elegans* strain, but commonly occur in genomic DNA from *C. elegans* strains of *trt-1* mutants whose telomeres are maintained by the telomerase-independent telomere maintenance pathway ALT (Fig. 7; Cheng et al. 2012).

DISCUSSION

Small RNAs composed of telomeric repeats have been shown to occur in several different organisms (Couvillion et al. 2009; Vrbsky et al. 2010; Rossiello et al. 2017). We interrogated a number of published *C. elegans* small RNA sequencing data sets and found rare examples of small RNAs mapping perfectly to telomeres as well as more abundant species mapping to telomeres with up to three mismatches. Telomeric small RNAs with an increasing number of mismatches correlated with an increasing enrichment for 22G small RNA species, which represent secondary “effector” siRNA molecules that are produced de novo by RNA-dependent RNA polymerases during both exogenous and endogenous RNA interference. All telomeric small RNAs mapping to telomeres with zero or one mismatches and the majority of species with two or three mismatches could be mapped to ITS sites using a method that randomly distributes nonunique reads, suggesting that ITS sites could be responsible for the formation of perfect telomeric small RNAs. However, mapping of siRNAs to borders of perfect telomere repeats throughout the genome suggested that while long stretches of perfect telomere repeats at either ITS tracts or at telomeres can give rise to siRNAs, telomeres could represent the more abundant source of perfect telomeric small RNAs. Perfect telomeric small RNAs that are created from ITS

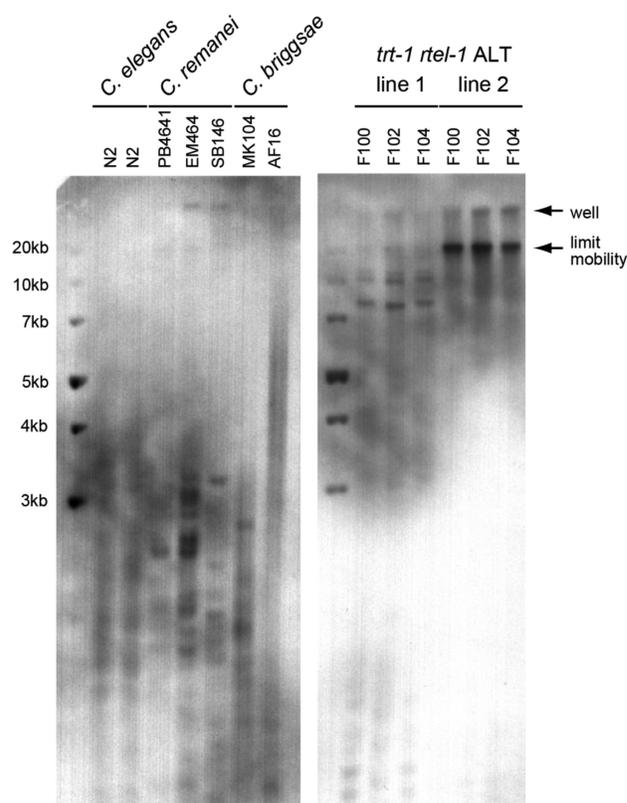


FIGURE 7. Telomere length of *C. elegans*, *C. briggsae*, and *C. remanei* strains. (Left) Southern blot analysis of Hinf1-digested genomic DNA isolated from *C. elegans*, *C. briggsae* and *C. remanei* strains using a (TTAGGC)_n probe. (Right) Southern blot analysis of two *C. elegans* *trt-1* telomerase mutant strains whose telomeres are maintained via ALT. Positions of telomeric bands located in the well or at limit mobility are noted by arrows.

tracts might possess characteristics of small RNAs derived from ITS tracts, such as beginning with any 5′ nucleotide.

Strikingly, perfect telomeric small RNAs were substantially more abundant in *C. briggsae* and *C. remanei*, whereas telomeric small RNAs with mismatches were generally of comparable abundance to *C. elegans*. However, perfect telomere repeat sequences within ITSs were depleted in *C. briggsae* and *C. remanei* relative to *C. elegans* (Table 1), suggesting that telomeres rather than perfect telomere sequence within ITS tracts are likely to be the source of the majority of perfect telomeric small RNAs. In parallel with their altered level of abundance, perfect telomeric small RNAs of *C. briggsae* and *C. remanei* displayed a substantial bias for being composed of the C-rich strand, which contrasts from the G-rich bias of small RNAs from *C. elegans*. The overrepresentation of C-rich telomeric small RNAs in *C. briggsae* and *C. remanei* is puzzling, possibly reflecting a change to the mechanism of telomeric small RNA biogenesis. In this regard, the C-rich telomeric small RNA species would be complementary to the telomeric noncoding RNA TERRA, which is considered to be

the most abundant telomeric RNA (Azzalin and Lingner 2015). We note that mammalian telomeric small RNAs were more frequently composed of the G-rich telomeric sequence and that their levels were very low, as observed for *C. elegans*. Together, our results argue that small RNA classes that are very rare may be functionally significant, consistent with a proposed role for telomeric small RNAs in responding to DNA damage at mammalian telomeres (Rossiello et al. 2017).

Perfect telomeric G-rich RNAs from all three *Caenorhabditis* species commonly began with 5' UAGGCU and 5' UUAGGC. C-rich telomeric small RNAs of *C. briggsae* and *C. remanei* were enriched for 5' ends that began with 5'CUAAGC (Fig. 6). Despite having prominent 5' end sequences, both G- and C-rich telomeric RNAs displayed a wide range of lengths, suggesting that they may have a common mechanism of biogenesis that may be distinct from other known *C. elegans* pathways that often display pronounced enrichment for a discrete RNA length. Perhaps consistently, *Tetrahymena* possesses telomeric small RNAs with 5' UGGGGU ends that also displayed a range of sizes (Couvillion et al. 2009). We did not observe consistent 5' sequences for mammalian perfect telomeric small RNAs, but found that 3' ends for major species of G-rich small RNAs for *Caenorhabditis* species and mammals often aligned perfectly with the 5' end of the C-rich DNA base at the chromosome terminus (Fig. 8).

Given that *C. briggsae* and *C. remanei* are more closely related to each other than to *C. elegans*, we examined their genomes and found consistent loss of the POT-2 protein that represses telomerase activity and ALT in *C. elegans*. Some wild isolates of *C. elegans* have long telomeres (Raices et al. 2005), which can be due to loss-of-function *pot-2* mutations (Cook et al. 2016). We examined telomeric DNA from several wild isolates of *C. briggsae* and *C. remanei* and found it to be distinct from Bristol N2 *C. elegans* DNA in that a minor fraction of their telomeres were long or unusually structured, similar to what is observed in telomerase-negative strains where ALT is active (Pickett and Reddel 2015). Thus, loss of *pot-2* or high levels of telomeric small RNAs may create ALT-like telomeric DNA in *C. briggsae* and *C. remanei*. It is possible that *C. elegans* POT-2 promotes turnover of telomeric small RNAs or that the second MRT-1 homolog of *C. briggsae* and *C. remanei* promotes biogenesis of telomeric siRNAs. Both POT-2 and MRT-1 have OB2-folds that are predicted to interact with 3' ends of single-stranded telomeric DNA (Lei et al. 2003), and Argonaute proteins possess a PAZ domain that uses a variant OB fold to interact with the 3' termini of single-stranded RNA (Song et al. 2003).

ITS sites have been observed in number of different species but have not been comprehensively characterized within a fully assembled metazoan genome. We located over one thousand ITS sites in the *C. elegans* genome and found a conspicuous depletion of ITS sites at exons

and UTRs and enrichment at introns. This localization pattern is not a general feature of hexanucleotide repeats, and is suggestive of specific formation mechanisms and/or selection pressures involved in ITS formation. These data suggest that ITS sites may be toxic to mRNAs, possibly because RNA composed of telomere repeats may form high order structures such as G-quadruplexes (Fay et al. 2017), and may also contribute to intron evolution. Although the origin of ITSs is uncertain, our data suggest that these genome sequences may function to promote gene silencing when in introns and to promote gene expression during development when in enhancers. Alternatively, it is possible that new ITS sites are created in regions of the genome that are silenced or heterochromatic, and that their association with silencing marks is a consequence of their mechanism of biogenesis. Given the abundance of telomeric siRNAs with mismatches that map to ITS tracts and the depletion of ITSs from 3'UTRs, we favor the possibility that ITS tracts may actively promote transcriptional and posttranscriptional gene silencing, a topic that could be addressed by nuclease-mediated genome engineering. Although telomerase generally functions at telomeres, there is evidence to suggest that its reverse transcriptase activity can result in the insertion of telomeric sequence at double-strand breaks, although this generally results in the creation of a new telomere (Churikov and Geli 2017). Therefore, ITSs may represent a contribution to intron and enhancer evolution by the telomerase enzyme.

We found that ITS sites are a shared feature among close relatives of *C. elegans*, although their compositions show some striking differences. ITSs in both *C. briggsae* and *C. remanei* are rarer and generally more degenerate than in *C. elegans*, although both species contain examples of ITS sites with perfect telomere homology. It is intriguing that there should be such a discrepancy in ITS divergence between these species, and this observation suggests differing selective pressures. Although it may be a coincidence that perfect telomeric small RNAs are far more abundant in *C. briggsae* and *C. remanei*, it is also possible that these RNAs are functionally linked to ITS formation and/or evolution. Telomeric small RNAs may actively inhibit formation or expansion of ITS sites, meaning that their very low levels in *C. elegans* could have led to an increasing number of ITSs in the *C. elegans* genome, many of which have been formed relatively recently.

MATERIALS AND METHODS

Analysis of high-throughput sequencing data

Publicly available RNA-seq data sets were downloaded from the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) (see Supplemental Table S9 for library information). Adapter trimming was performed as required using the `bbduk.sh` script from

strains MK104 and AF16, which were grown at 20°C on NGM plates. Southern blotting was performed using genomic DNA prepared from a Gentra kit and a digoxigenin-labeled TTAGGC probe, as previously described (Meier et al. 2009).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Yun Li, Will Valdar, and Erin Osborne for advice concerning genome analysis. This work was supported by National Institutes of Health R01 GM066228 to S.A.

Author contributions: S.F. and S.A. designed the experiments. S.F. and E.H.L.-S. carried out the experiments, S.F. performed computational analysis and E.H.L.-S. performed Southern blotting, S.F. and S.A. wrote the manuscript.

Received March 20, 2019; accepted May 28, 2019.

REFERENCES

- Aksenova AY, Greenwell PW, Dominska M, Shishkin AA, Kim JC, Petes TD, Mirkin SM. 2013. Genome rearrangements caused by interstitial telomeric sequences in yeast. *Proc Natl Acad Sci* **110**: 19866–19871. doi:10.1073/pnas.1319313110
- Aksenova AY, Han G, Shishkin AA, Volkov KV, Mirkin SM. 2015. Expansion of interstitial telomeric sequences in yeast. *Cell Rep* **13**: 1545–1551. doi:10.1016/j.celrep.2015.10.023
- Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, Cording AC, Doebley AL, Goldstein LD, Lehrbach NJ, Le Pen J, et al. 2012. piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. *Cell* **150**: 88–99. doi:10.1016/j.cell.2012.06.018
- Ashley T, Ward DC. 1993. A “hot spot” of recombination coincides with an interstitial telomeric sequence in the Armenian hamster. *Cytogenet Cell Genet* **62**: 169–171. doi:10.1159/000133464
- Axtell MJ. 2014. Butter: high precision genomic alignment of small RNA-seq data. *bioRxiv* doi:10.1101/007427
- Azzalin CM, Lingner J. 2015. Telomere functions grounding on TERRA firma. *Trends Cell Biol* **25**: 29–36. doi:10.1016/j.tcb.2014.08.007
- Azzalin CM, Nergadze SG, Giulotto E. 2001. Human intrachromosomal telomeric-like repeats: sequence organization and mechanisms of origin. *Chromosoma* **110**: 75–82. doi:10.1007/s004120100135
- Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J. 2007. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**: 798–801. doi:10.1126/science.1147182
- Bagijn MP, Goldstein LD, Sapetschnig A, Weick EM, Bouasker S, Lehrbach NJ, Simard MJ, Miska EA. 2012. Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science* **337**: 574–578. doi:10.1126/science.1220952
- Bah A, Wischniewski H, Shchepachev V, Azzalin CM. 2012. The telomeric transcriptome of *Schizosaccharomyces pombe*. *Nucleic Acids Res* **40**: 2995–3005. doi:10.1093/nar/gkr1153
- Bao Z, Eddy SR. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res* **12**: 1269–1276. doi:10.1101/gr.88502
- Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, Kasschau KD, Chaves DA, Gu W, Vasale JJ, Duan S, et al. 2008. PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol Cell* **31**: 67–78. doi:10.1016/j.molcel.2008.06.002
- Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363–366. doi:10.1038/35053110
- Billi AC, Fischer SE, Kim JK. 2014. Endogenous RNAi pathways in *C. elegans*. *WormBook* 1–49. doi:10.1895/wormbook.1.170.1
- Bosco N, de Lange T. 2012. A TRF1-controlled common fragile site containing interstitial telomeric sequences. *Chromosoma* **121**: 465–474. doi:10.1007/s00412-012-0377-6
- Buckley BA, Burkhart KB, Gu SG, Spracklin G, Kershner A, Fritz H, Kimble J, Fire A, Kennedy S. 2012. A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature* **489**: 447–451. doi:10.1038/nature11352
- Cao F, Li X, Hiew S, Brady H, Liu Y, Dou Y. 2009. Dicer independent small RNAs associate with telomeric heterochromatin. *RNA* **15**: 1274–1281. doi:10.1261/rna.1423309
- Cheng C, Shtessel L, Brady MM, Ahmed S. 2012. *Caenorhabditis elegans* POT-2 telomere protein represses a mode of alternative lengthening of telomeres with normal telomere lengths. *Proc Natl Acad Sci* **109**: 7805–7810. doi:10.1073/pnas.1119191109
- Chu HP, Cifuentes-Rojas C, Kesner B, Aeby E, Lee HG, Wei C, Oh HJ, Boukhali M, Haas W, Lee JT. 2017. TERRA RNA antagonizes ATRX and protects telomeres. *Cell* **170**: 86–101.e16. doi:10.1016/j.cell.2017.06.017
- Churikov D, Géli V. 2017. De novo telomere addition at chromosome breaks: dangerous liaisons. *J Cell Biol* **216**: 2243–2245. doi:10.1083/jcb.201705156
- Claycomb JM, Batista PJ, Pang KM, Gu W, Vasale JJ, van Wolfswinkel JC, Chaves DA, Shirayama M, Mitani S, Ketting RF, et al. 2009. The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* **139**: 123–134. doi:10.1016/j.cell.2009.09.014
- Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi DD, Noble LM, Rockman MV, Alkema MJ, Braendle C, Kammenga JE, et al. 2016. The genetic basis of natural variation in *Caenorhabditis elegans* telomere length. *Genetics* **204**: 371–383. doi:10.1534/genet.ics.116.191148
- Couvillion MT, Lee SR, Hogstad B, Malone CD, Tonkin LA, Sachidanandam R, Hannon GJ, Collins K. 2009. Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of *Tetrahymena thermophila*. *Genes Dev* **23**: 2016–2032. doi:10.1101/gad.1821209
- Cusanelli E, Romero CA, Chartrand P. 2013. Telomeric noncoding RNA TERRA is induced by telomere shortening to nucleate telomerase molecules at short telomeres. *Mol Cell* **51**: 780–791. doi:10.1016/j.molcel.2013.08.029
- Cutter AD. 2008. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Mol Biol Evol* **25**: 778–786. doi:10.1093/molbev/msn024
- Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, Sapetschnig A, Buhecha HR, Gilchrist MJ, Howe KL, Stark R, et al. 2008. Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Mol Cell* **31**: 79–90. doi:10.1016/j.molcel.2008.06.003
- de Lange T. 2018. Shelterin-mediated telomere protection. *Annu Rev Genet* **52**: 223–247. doi:10.1146/annurev-genet-032918-021921
- Dyer MA, Qadeer ZA, Valle-Garcia D, Bernstein E. 2017. ATRX and DAXX: mechanisms and mutations. *Cold Spring Harb Perspect Med* **7**: a026567. doi:10.1101/cshperspect.a026567
- Fay MM, Lyons SM, Ivanov P. 2017. RNA G-quadruplexes in biology: principles and molecular mechanisms. *J Mol Biol* **429**: 2127–2147. doi:10.1016/j.jmb.2017.05.017

- Fierst JL, Willis JH, Thomas CG, Wang W, Reynolds RM, Ahearne TE, Cutter AD, Phillips PC. 2015. Reproductive mode and the evolution of genome size and structure in *Caenorhabditis* nematodes. *PLoS Genet* **11**: e1005323. doi:10.1371/journal.pgen.1005323
- Fischer SE, Montgomery TA, Zhang C, Fahlgren N, Breen PC, Hwang A, Sullivan CM, Carrington JC, Ruvkun G. 2011. The ERI-6/7 helicase acts at the first stage of an siRNA amplification pathway that targets recent gene duplications. *PLoS Genet* **7**: e1002369. doi:10.1371/journal.pgen.1002369
- Fresno C, Fernandez EA. 2013. RDAVIDWebService: a versatile R interface to DAVID. *Bioinformatics* **29**: 2810–2811. doi:10.1093/bioinformatics/btt487
- Garrigues JM, Sidoli S, Garcia BA, Strome S. 2015. Defining heterochromatin in *C. elegans* through genome-wide analysis of the heterochromatin protein 1 homolog HPL-2. *Genome Res* **25**: 76–88. doi:10.1101/gr.180489.114
- Gaydos LJ, Rechtsteiner A, Egelhofer TA, Carroll CR, Strome S. 2012. Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in *C. elegans* germ cells. *Cell Rep* **2**: 1169–1177. doi:10.1016/j.celrep.2012.09.019
- Gent JI, Lamm AT, Pavelec DM, Maniar JM, Parameswaran P, Tao L, Kennedy S, Fire AZ. 2010. Distinct phases of siRNA synthesis in an endogenous RNAi pathway in *C. elegans* soma. *Mol Cell* **37**: 679–689. doi:10.1016/j.molcel.2010.01.012
- Gomes NM, Ryder OA, Houck ML, Charter SJ, Walker W, Forsyth NR, Austad SN, Venditti C, Pagel M, Shay JW, et al. 2011. Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging Cell* **10**: 761–768. doi:10.1111/j.1474-9726.2011.00718.x
- Gu W, Shirayama M, Conte D, Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ, et al. 2009. Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol Cell* **36**: 231–244. doi:10.1016/j.molcel.2009.09.020
- Hahne F, Ivanek R. 2016. Visualizing genomic data using Gviz and bioconductor. *Methods Mol Biol* **1418**: 335–351. doi:10.1007/978-1-4939-3578-9_16
- Han T, Manoharan AP, Harkins TT, Bouffard P, Fitzpatrick C, Chu DS, Thierry-Mieg D, Thierry-Mieg J, Kim JK. 2009. 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in *Caenorhabditis elegans*. *Proc Natl Acad Sci* **106**: 18674–18679. doi:10.1073/pnas.0906378106
- Ho JW, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn KA, Minoda A, Tolstorukov MY, Appert A, et al. 2014. Comparative analysis of metazoan chromatin organization. *Nature* **512**: 449–452. doi:10.1038/nature13415
- Holoch D, Moazed D. 2015. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet* **16**: 71–84. doi:10.1038/nrg3863
- Ijdo JW, Baldini A, Ward DC, Reeders ST, Wells RA. 1991. Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc Natl Acad Sci* **88**: 9051–9055. doi:10.1073/pnas.88.20.9051
- Ikegami K, Egelhofer TA, Strome S, Lieb JD. 2010. *Caenorhabditis elegans* chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. *Genome Biol* **11**: R120. doi:10.1186/gb-2010-11-12-r120
- Jacob NK, Skopp R, Price CM. 2001. G-overhang dynamics at *Tetrahymena* telomeres. *EMBO J* **20**: 4299–4308. doi:10.1093/emboj/20.15.4299
- Jacob NK, Kirk KE, Price CM. 2003. Generation of telomeric G strand overhangs involves both G and C strand cleavage. *Mol Cell* **11**: 1021–1032. doi:10.1016/S1097-2765(03)00131-X
- Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, Plasterk RH. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev* **15**: 2654–2659. doi:10.1101/gad.927801
- Kilburn AE, Shea MJ, Sargent RG, Wilson JH. 2001. Insertion of a telomere repeat sequence into a mammalian gene causes chromosome instability. *Mol Cell Biol* **21**: 126–135. doi:10.1128/MCB.21.1.126-135.2001
- Knight SW, Bass BL. 2001. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* **293**: 2269–2271. doi:10.1126/science.1062039
- Koster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**: 2520–2522. doi:10.1093/bioinformatics/bts480
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25. doi:10.1186/gb-2009-10-3-r25
- Lei M, Podell ER, Baumann P, Cech TR. 2003. DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA. *Nature* **426**: 198–203. doi:10.1038/nature02092
- Lin KW, Yan J. 2008. Endings in the middle: current knowledge of interstitial telomeric sequences. *Mutat Res* **658**: 95–110. doi:10.1016/j.mrrev.2007.08.006
- Lowden MR, Flibotte S, Moerman DG, Ahmed S. 2011. DNA synthesis generates terminal duplications that seal end-to-end chromosome fusions. *Science* **332**: 468–471. doi:10.1126/science.1199022
- Manakov SA, Pezic D, Marinov GK, Pastor WA, Sachidanandam R, Aravin AA. 2015. MIWI2 and MILI have differential effects on piRNA biogenesis and DNA methylation. *Cell Rep* **12**: 1234–1243. doi:10.1016/j.celrep.2015.07.036
- Meier B, Barber LJ, Liu Y, Shtessel L, Boulton SJ, Gartner A, Ahmed S. 2009. The MRT-1 nuclease is required for DNA crosslink repair and telomerase activity in vivo in *Caenorhabditis elegans*. *EMBO J* **28**: 3549–3563. doi:10.1038/emboj.2009.278
- Meyne J, Baker RJ, Hobart HH, Hsu TC, Ryder OA, Ward OG, Wiley JE, Wurster-Hill DH, Yates TL, Moyzis RK. 1990. Distribution of non-telomeric sites of the (TTAGGG)_n telomeric sequence in vertebrate chromosomes. *Chromosoma* **99**: 3–10. doi:10.1007/BF01737283
- Moore A, Dominska M, Greenwell P, Aksenova AY, Mirkin S, Petes T. 2018. Genetic control of genomic alterations induced in yeast by interstitial telomeric sequences. *Genetics* **209**: 425–438. doi:10.1534/genetics.118.300950
- Palm W, de Lange T. 2008. How shelterin protects mammalian telomeres. *Annu Rev Genet* **42**: 301–334. doi:10.1146/annurev.genet.41.110306.130350
- Pandey RR, Tokuzawa Y, Yang Z, Hayashi E, Ichisaka T, Kajita S, Asano Y, Kunieda T, Sachidanandam R, Chuma S, et al. 2013. Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice. *Proc Natl Acad Sci* **110**: 16492–16497. doi:10.1073/pnas.1316316110
- Phillips CM, Montgomery BE, Breen PC, Roovers EF, Rim YS, Ohsumi TK, Newman MA, van Wolfswinkel JC, Ketting RF, Ruvkun G, et al. 2014. MUT-14 and SMUT-1 DEAD box RNA helicases have overlapping roles in germline RNAi and endogenous siRNA formation. *Curr Biol* **24**: 839–844. doi:10.1016/j.cub.2014.02.060
- Phillips CM, Brown KC, Montgomery BE, Ruvkun G, Montgomery TA. 2015. piRNAs and piRNA-dependent siRNAs protect conserved and essential *C. elegans* genes from misrouting into the RNAi pathway. *Dev Cell* **34**: 457–465. doi:10.1016/j.devcel.2015.07.009
- Pickett HA, Reddel RR. 2015. Molecular mechanisms of activity and derepression of alternative lengthening of telomeres. *Nat Struct Mol Biol* **22**: 875–880. doi:10.1038/nsmb.3106
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033

- Raices M, Maruyama H, Dillin A, Karlseder J. 2005. Uncoupling of longevity and telomere length in *C. elegans*. *PLoS Genet* **1**: e30. doi:10.1371/journal.pgen.0010030
- Raices M, Verdun RE, Compton SA, Haggblom CI, Griffith JD, Dillin A, Karlseder J. 2008. *C. elegans* telomeres contain G-strand and C-strand overhangs that are bound by distinct proteins. *Cell* **132**: 745–757. doi:10.1016/j.cell.2007.12.039
- Reinhart BJ, Bartel DP. 2002. Small RNAs correspond to centromere heterochromatic repeats. *Science* **297**: 1831. doi:10.1126/science.1077183
- Rossiello F, Aguado J, Sepe S, Iannelli F, Nguyen Q, Pitchiaya S, Caminci P, d'Adda di Fagnana F. 2017. DNA damage response inhibition at dysfunctional telomeres by modulation of telomeric DNA damage response RNAs. *Nat Commun* **8**: 13980. doi:10.1038/ncomms13980
- Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**: 1193–1207. doi:10.1016/j.cell.2006.10.040
- Sapetschnig A, Sarkies P, Lehrbach NJ, Miska EA. 2015. Tertiary siRNAs mediate paramutation in *C. elegans*. *PLoS Genet* **11**: e1005078. doi:10.1371/journal.pgen.1005078
- Schoeffner S, Blasco MA. 2008. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol* **10**: 228–236. doi:10.1038/ncb1685
- Schoeffner S, Blasco MA. 2010. Chromatin regulation and non-coding RNAs at mammalian telomeres. *Semin Cell Dev Biol* **21**: 186–193. doi:10.1016/j.semcdb.2009.09.015
- Seluanov A, Chen Z, Hine C, Sasahara TH, Ribeiro AA, Catania KC, Presgraves DC, Gorbunova V. 2007. Telomerase activity coevolves with body mass not lifespan. *Aging Cell* **6**: 45–52. doi:10.1111/j.1474-9726.2006.00262.x
- Seo B, Kim C, Hills M, Sung S, Kim H, Kim E, Lim DS, Oh HS, Choi RM, Chun J, et al. 2015. Telomere maintenance through recruitment of internal genomic regions. *Nat Commun* **6**: 8189. doi:10.1038/ncomms9189
- Seth M, Shirayama M, Gu W, Ishidate T, Conte D Jr, Mello CC. 2013. The *C. elegans* CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. *Dev Cell* **27**: 656–663. doi:10.1016/j.devcel.2013.11.014
- Sfeir AJ, Chai W, Shay JW, Wright WE. 2005. Telomere-end processing the terminal nucleotides of human chromosomes. *Mol Cell* **18**: 131–138. doi:10.1016/j.molcel.2005.02.035
- Shen EZ, Chen H, Ozturk AR, Tu S, Shirayama M, Tang W, Ding YH, Dai SY, Weng Z, Mello CC. 2018. Identification of piRNA binding sites reveals the argonaute regulatory landscape of the *C. elegans* germline. *Cell* **172**: 937–951 e918. doi:10.1016/j.cell.2018.02.002
- Shi Z, Montgomery TA, Qi Y, Ruvkun G. 2013. High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes. *Genome Res* **23**: 497–508. doi:10.1101/gr.149112.112
- Shtessel L, Lowden MR, Cheng C, Simon M, Wang K, Ahmed S. 2013. *Caenorhabditis elegans* POT-1 and POT-2 repress telomere maintenance pathways. *G3 (Bethesda)* **3**: 305–313. doi:10.1534/g3.112.004440
- Smogorzewska A, de Lange T. 2004. Regulation of telomerase by telomeric proteins. *Annu Rev Biochem* **73**: 177–208. doi:10.1146/annurev.biochem.73.071403.160049
- Song JJ, Liu J, Tolia NH, Schneiderman J, Smith SK, Martienssen RA, Hannon GJ, Joshua-Tor L. 2003. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol* **10**: 1026–1032. doi:10.1038/nsb1016
- Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, et al. 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol* **1**: E45. doi:10.1371/journal.pbio.0000045
- Vasale JJ, Gu W, Thivierge C, Batista PJ, Claycomb JM, Youngman EM, Duchaine TF, Mello CC, Conte D Jr. 2010. Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. *Proc Natl Acad Sci* **107**: 3582–3587. doi:10.1073/pnas.0911908107
- Vrbisky J, Akimcheva S, Watson JM, Turner TL, Daxinger L, Vyskot B, Aufsatz W, Riha K. 2010. siRNA-mediated methylation of *Arabidopsis* telomeres. *PLoS Genet* **6**: e1000986. doi:10.1371/journal.pgen.1000986
- Wang G, Reinke V. 2008. A *C. elegans* Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Curr Biol* **18**: 861–867. doi:10.1016/j.cub.2008.05.009
- Watanabe T, Cheng EC, Zhong M, Lin H. 2015. Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline. *Genome Res* **25**: 368–380. doi:10.1101/gr.180802.114
- Wenda JM, Homolka D, Yang Z, Spinelli P, Sachidanandam R, Pandey RR, Pillai RS. 2017. Distinct roles of RNA helicases MVH and TDRD9 in PIWI slicing-triggered mammalian piRNA biogenesis and function. *Dev Cell* **41**: 623–637 e629. doi:10.1016/j.devcel.2017.05.021
- Yang D, Xiong Y, Kim H, He Q, Li Y, Chen R, Songyang Z. 2011. Human telomeric proteins occupy selective interstitial sites. *Cell Res* **21**: 1013–1027. doi:10.1038/cr.2011.39
- Zhang LF, Ogawa Y, Ahn JY, Namekawa SH, Silva SS, Lee JT. 2009. Telomeric RNAs mark sex chromosomes in stem cells. *Genetics* **182**: 685–698. doi:10.1534/genetics.109.103093
- Zhang D, Tu S, Stubna M, Wu WS, Huang WC, Weng Z, Lee HC. 2018. The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. *Science* **359**: 587–592. doi:10.1126/science.aao2840