Comparative Analysis of Nematode small-RNA pathways using Gametocyte specific factor 1 (GTSF-1)

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Shamitha Shamitha

geboren am 18 Juli 1996 in Ahmedabad

Mainz, 2024

Dekan: Prof. Dr. Eckhard Thines

1.Berichterstatter: Prof. Dr. René F. Ketting

2.Berichterstatter: Prof. Dr. Miguel Andrade

Tag der mündlichen Prüfung:

Table of Contents

Summary

Small-RNAs (sRNAs) regulate gene expression by binding to an Argonaute (Ago) protein and interfering with mRNA transcription or translation. Evolutionary arms race with transposable elements has greatly diversified the small-RNA pathways in individual species. This resonates in the Zn finger protein Gametocyte specific factor -1 (GTSF-1), which are essential for sRNA pathways in many species but show evolutionary plasticity by acting at different steps of the pathway in different species. In mice and insects, GTSF-1 is a PIWI Argonaute-associated protein crucial for piRNA-directed silencing. It plays a pivotal role by enhancing the endoribonuclease activity of catalytic PIWI Agos (Arif et al., 2022). However, in the nematode *C. elegans*, GTSF-1 associates with the RNA-Dependent RNA Polymerase (RdRP) RRF-3, diverging from the conventional PIWI association (Almeida et al., 2018). RRF-3 is essential for the biogenesis of 26G-RNAs (siRNAs), which regulate gene expression in the germline. The observation gains significance as several nematode lineages independently lost PIWI and the entire piRNA pathway, yet retained RRF-3-like RdRPs (Sarkies et al., 2015).

To investigate the conservation of GTSF-1 function within nematodes, we conducted a comparative analysis of small-RNA pathways in three nematode species: *C. elegans*, *C. briggsae* and *P. pacificus,* with a focus on GTSF-1 and the 26G-RNA pathway. Employing various methodologies including CRISPR-Cas9 transgenics, Next Generation Sequencing, Immunoprecipitation coupled to Mass-spectrometry, and AlphaFold2 predictions, we elucidated the interactions and functions of GTSF-1 within these nematode species.

Our findings reveal an interaction between GTSF-1 and RRF-3 in *C. briggsae* and *P. pacificus,* crucial for fertility and the regulation of 26G-RNAs. We identified sequence and structural differences in nematode GTSF-1 homologs potentially explaining the loss of PIWI binding. Utilizing AlphaFold2, we identified key domains and residues for GTSF-1 binding to RRF-3, validated through *in-vitro* and *in-vivo* assays.

We also described conserved functions of the PETISCO complex in *C. briggsae* embryogenesis and 21U-RNA biogenesis. Notably, we uncover a curious link between the PETISCO and 26G-RNA pathway in *C. briggsae,* potentially mediated through GTSF-1. In summary, these results shed light on the conservation and functional significance of small RNA pathways in clade V nematodes. Moreover, our study suggests ancient conservation of this association within nematodes, highlighting the functional adaptation of nematode GTSF-1 proteins. Through this work, we provide a compelling explanation for the loss of the highly conserved piRNA pathway across nematode lineages.

Zusammenfassung

Small RNA Moleküle (sRNAs) regulieren die Genexpression, indem sie an ein Argonauten(Ago) Protein binden und die mRNA-Transkription oder -Translation stören. Ein evolutionärer Wettlauf mit Transposons führte zu einer Diversifizierung in unterschiedliche Arten von sRNAs Signalwegen. Dies spiegelt sich in dem Zinkfingerprotein Gametocyte specific factor -1 (GTSF-1) wider, welches in vielen Spezies essenziell für sRNA-Signalwege ist, durch seine evolutionäre Plastizität jedoch an verschiedenen Schritten in unterschiedlichen Spezies beteiligt ist. In Mäusen und Insekten ist GTSF-1 mit dem PIWI-Argonaute Protein assoziiert, welches essenziell für piRNA-gesteuerte Inhibierung ist. So spielt es entscheidende Rolle, indem es die Endoribonukleaseaktivität katalytischer PIWI-Agos steigert (Arif et al., 2022). In der Nematode *C. elegans* interagiert GTSF-1 jedoch mit der RNA-abhängigen RNA-Polymerase (RdRP) RRF-3, abweichend von der konventionellen Assoziation mit PIWI (Almeida et al., 2018). RRF-3 ist essenziell für die Biogenese von 26G-RNAs (siRNAs), welche die Genexpression in der Keimbahn regulieren. Diese Beobachtung wird vor dem Hintergrund relevant, dass mehrere Klassen von Nematoden unabhängig PIWI und den gesamten piRNA-Signalweg verloren, RRF-3-ähnliche RdRPs jedoch beibehalten haben (Sarkies et al., 2015).

Um die Konservierung der Funktion von GTSF 1 in Nematoden zu untersuchen, haben wir eine vergleichende Analyse der sRNA Signalwege mit Fokus auf GTSF 1 und 26G-RNA Signalwegen in den folgenden drei Spezies durchgeführt: *C. elegans*, *C. briggsae* und *P. pacificus*. Durch die Anwendung verschiedener Methoden, einschließlich CRISPR-Cas9-Transgenics, Next Generation Sequencing, Immunopräzipitation gekoppelt mit Massenspektrometrie sowie mit AlphaFold2 generierte Vorhersagen, konnten wir die Interaktionen und Funktionen von GTSF 1 in diesen Spezies aufklären.

Unsere Ergebnisse offenbaren die Interaktion zwischen GTSF-1 und RRF-3 in *C. briggsae* und *P. pacificus*, welche essenziell für die Fruchtbarkeit und die Regulation von 26G-RNAs ist. Wir konnten Sequenz- und Strukturunterschiede der GTSF 1 Homologe in den untersuchten Nematoden aufdecken, welche eine mögliche Erklärung für den Verlust der Bindung mit PIWI darstellen. Mithilfe von AlphaFold2 konnten wir Schlüsseldomänen- und -aminosäuren für die Interaktion von GTSF-1 mit RRf 3 identifizieren, welche durch *in vitro*- und *in vivo*-Assays validiert wurden.

Zudem beschreiben wir konservierte Funktionen des PETISCO-Komplexes in der Embryogenese von *C. briggsae* und der Biogenese von 21U-RNAs. Nennenswert ist hier die entdeckte Verbindung zwischen PETISCO und dem 26G-RNA Signalweg in *C. briggsae,* welche möglicherweise durch GTSF 1 ermöglicht wird. Zusammenfassend lässt sich festhalten, dass diese Ergebnisse die Konservierung und funktionelle Signifikanz von sRNA Signalwegen in Nematoden der Klade V erleuchten.

Zudem legen unsere Ergebnisse eine antike Konservierung dieser Assoziation in Nematoden nahe, was wiederum die funktionale Anpassbarkeit von GTSF 1 in Nematoden betont. Mithilfe dieser Arbeit liefern wir eine fundierte Erklärung für den Verlust des hoch konservierten piRNA-Signalweges in verschiedenen Klassen von Nematoden.

List of Abbreviations

AGO	Argonaute
Erh	Enhancer of rudimentary
Eri	Enhanced RNAi
GA	Gravid-adult
GTSF-1	Gametocyte-specific factor-1
Mrt	Mortal Germline
PTGS	Post-transcriptional gene silencing
RISC	RNA-Induced Silencing Complex
RRM	RNA recognition motif
RdRPs	RNA-dependent RNA polymerases
RNAi	RNA interference
TE	Transposable element
TGS	Transcriptional gene silencing
USTC	Upstream sequence transcription complex
WT	Wild-type
dsRNA	Double-stranded RNA
eTudor	Extended Tudor
endo-siRNA	Endogenous siRNAs

miRNA	MicroRNAs
ssRNA	Single-stranded RNA

List of Tables & Figures

INTRODUCTION

The control of gene expression is a fundamental biological process vital to the functioning of all organisms. Within eukaryotic cells, organelles play a crucial role by physically segregating the events involved in gene expression. This spatial separation allows for regulation to occur in distinct locations. While transcriptional regulation is predominant, post-transcriptional processes also significantly contribute. These post-transcriptional events include precursor mRNA processing, such as 5′-end capping, 3'-end cleavage and polyadenylation, as well as intron splicing, RNA base editing, and modifications. This diversity of mechanisms is essential for driving the necessary changes in gene expression during eukaryotic development.

In 1998, Fire and Mello published a groundbreaking paper in *C. elegans* showing that the introduction of double-stranded RNA (dsRNA) triggers the degradation of endogenous mRNA. Further research showed that this dsRNA is converted into shorter intermediates capable of binding to their target mRNAs, resulting in transcript cleavage. These studies established a novel conceptual framework for post-transcriptional gene regulation through small RNAs, a process termed RNA interference (RNAi) (Sen and Blau, 2006). In this section, we describe key mechanistic principles of RNAi and then delve into details of this pathway in the model organism *C. elegans.*

1. Mechanisms of RNA Interference

RNA interference (RNAi) is a mode of gene regulation that employ small RNAs as guide for sequence specific degradation of RNA. The canonical RNAi pathway involves the cleavage of long dsRNA into 20-30nt siRNAs, with one strand loaded onto an Argonaute protein possessing endonucleolytic activity. The resulting complex is termed RNA-Induced Silencing Complex (RISC). This Argonaute-siRNA RISC targets and cleaves complementary mRNA, allowing cells to selectively silence specific genes.

RISCs regulate mRNA expression at both transcriptional and post-transcriptional level. RISCs can trigger transcript cleavage, deadenylation or decapping as well as inhibit translation (Höck and Meister, 2008; Hutvagner and Simard, 2008). Some RISCs also carry out silencing in the nucleus by triggering deposition of repressive chromatin marks and interfering with RNA polymerase (Guang *et al.*, 2008, 2010; Bethany A Buckley *et al.*, 2012; Castel and Martienssen, 2013; Meister, 2013; Mao *et al.*, 2015)

RNA silencing pathways differ in the origin and biogenesis of small RNAs, mechanisms leading to target repression, and biological roles. There are three broad classes of small RNAs namely miRNAs, piRNAs, and endo-siRNAs. They vary in size, and initial nucleotide identity, but their classification predominantly relies on their Argonaute cofactor and their downstream functions. MicroRNAs and endo-siRNAs bind to the AGO clade of Argonautes while piRNAs exclusively associate with the animal-specific PIWI clade Argonaute.

MicroRNAs

MicroRNAs (miRNA) were the first class of small RNAs to be described and are the most extensively studied branch of RNAi to date (reviewed in Ha and Kim, 2014; Bartel, 2018). Typically, miRNAs regulate the expression of protein-coding genes involved in various biological processes, including development, differentiation, and apoptosis. Hundreds of miRNAs have been annotated in species from all kingdoms of life. Each miRNA has a defined sequence and can guide silencing of many different mRNAs through imperfect base pairing.

In animals, miRNAs are derived from long Pol II transcripts that contain regions that fold back to create hairpin loops. Two RNAse III endonucleases, nuclear Drosha and cytoplasmic Dicer, tandemly process these hairpins, converting the transcripts into short double-stranded RNA duplexes (reviewed in Ha and Kim, 2014; Bartel, 2018). One strand of the resulting duplex is loaded onto Argonaute (AGO) proteins and guided to target mRNAs. Typically, miRNAs bind the 3' UTR region of mRNAs with partial complementarity. This results in translational repression or degradation (reviewed in Svoboda, 2020).

Endogenous siRNAs

Unlike miRNAs that are derived from the genome, siRNAs may be endogenous or arise via viral infection or other exogenous sources. Endogenous siRNAs (endo-siRNA) is an umbrella term for diverse types of small-RNAs produced endogenously in the cell. Endo-siRNAs are produced by cleavage of long double-stranded RNA (dsRNA) into short siRNAs, typically by Dicer endonucleases. There are several sources of dsRNA in the cell. In Drosophila and mice, bidirectional transcription of a locus can give rise to complementary sense and anti-sense transcripts that hybridize to form a dsRNA. Such complimentary transcripts can also arise from two separate loci with similar genetic content. Examples of these loci would be pseudogenes or multiple copies of a transposon sequence. Additionally, some transcripts may also contain inverted repeat sequences that can fold back into a hairpin structure, akin to miRNA. Furthermore, nematodes and plants employ RNA-dependent RNA polymerases (RdRPs) for the direct transcription of complementary strands.

Endo-siRNAs are deeply conserved but their features vary across species. The endo-siRNAs of flies and mice are produced by Dcr2, have a clear signature of 21nt and are loaded onto Ago2 (reviewed in Claycomb, 2014). In contrast, nematodes and plants have a complex network of endo-siRNAs, involving multiple RdRPs and Argonautes. *C. elegans* produces endo-siRNAs of lengths 22nt and 26nt, with only the latter dependent on Dicer. Both classes regulate germline gene expression but via the action of distinct Ago cofactors (Ketting and Cochella, 2021). Similarly, Arabidopsis produces diverse endo-siRNAs, spanning 21 to 24nt in length, with the involvement of multiple RdRPs, Argonautes, and Dicer proteins (Baulcombe, 2023). Despite these differences, a general function of endo-siRNAs is to provide an additional layer of control on protein-coding gene expression or genome surveillance against nucleic acid parasites (reviewed in Claycomb, 2014).

Piwi-interacting RNAs

Piwi-interacting RNAs are an animal specific class of small-RNAs that guide the PIWI clade of Argonautes. They are typically 20-30nt long and often restricted to the germline. In most animals, their primary function is to defend the germline against genomic parasite such as transposons and viruses (reviewed in Ozata et al., 2019).

The biogenesis of piRNAs differs from miRNA and siRNA in that it does not depend on dsRNA and Dicer. Instead, PIWI Argonautes themselves cleave long single stranded RNA precursors to create piRNAs. In most animals, piRNA producing loci are clustered together in discrete genomic regions. These “hotspots” can be either uni-stranded or dual-stranded, based on their capacity to generate piRNAs from one or both genomic strands. These precursor transcripts are exported to specialized peri-nuclear compartments where numerous factors facilitate their processing (reviewed in Czech et al., 2018). Specifically, two PIWI proteins in tandem action drive a piRNA amplification process known as the *ping-pong* cycle.

The first PIWI, guided by maternally inherited "initiator" piRNAs, cleaves the sense precursor transcript. This cleavage results in the generation of a new precursor with a monophosphorylated 5′ end. A second PIWI then cleaves the 5' monophosphorylated precursor to generate responder piRNAs. The cycle is completed by the responder piRNA itself acting as an initiator piRNA and guiding cleavage of the antisense copy of the transcript into more piRNAs. Hence, the ping-pong cycle continues until precursor transcripts levels are depleted, contributing to the exhaustion of these transcripts in the cell (reviewed in Ozata et al., 2019).

As these precursors are also the mRNAs targeted for silencing, the ping-pong cycle additionally functions as a mechanism for post-transcriptional gene silencing (PTGS). Cleavage competent PIWI proteins are essential for PTGS, such as Aub and Ago3 in flies, Siwi and Ago3 in silkmoths, and MILI and MIWI in mice. Most species also encode a third PIWI that mediates transcriptional gene silencing (TGS) in the nucleus. TGS can take place even when the PIWI is catalytically inactive. Instead of influencing transcript levels through cleavage, TGS operates through chromatin silencing. PIWI binds to and brings histone-modifying enzymes and DNA methyltransferases to target loci. This third PIWI is referred to as Piwi in flies and MIWI2 in mice (reviewed in Czech et al., 2018).

Evolution of RNAi

Phylogenetic analyses reveal that the core components of RNAi: Argonaute, Piwi, Dicer, and RdRP, can be traced back to the last common ancestor of eukaryotes, suggesting ancient origin of this pathway. The ancestral function of RNAi is proposed to be genome defense against viruses through cytoplasmic, Ago pathways and against transposons through nuclear, Piwi pathways (Obbard et al., 2009; Cerutti & Casas-Mollano, 2006; Shabalina & Koonin, 2008). Furthermore, RNAi genes working against viral and TE display rapid evolution, suggestive of an ongoing evolutionary arms race in RNAi pathways and genome parasites (Palmer et al., 2018b). This is also reflected in the high complexity of RNAi pathways, as key RNAi genes and accessory proteins repeatedly undergo high duplication or loss across species. It has been proposed that this defense mechanism of siRNA and piRNA pathways might have indirectly influenced the evolution of a present characteristic of eukaryotic gene expression, namely miRNAs (Cerutti and Casas-Mollano, 2006; Shabalina and Koonin, 2008).

2. The *C. elegans* Small-RNA Landscape

The nematode *C. elegans* has played a pioneering role in unraveling the complexities of RNAi pathways. Despite having only one Dicer gene, RNA silencing in *C. elegans* is remarkably intricate owing to the presence of more than 20 Argonaute proteins and distinct classes of small RNAs (Ketting, 2011; Ketting and Cochella, 2021). For instance, the PIWI Argonaute PRG-1 associates with piRNAs while AGO clade ALG-3/4 and ERGO-1 bind a subset of 26nt long endogenous siRNAs. Another Argonaute, RDE-1, from the AGO clade, binds siRNAs generated from external sources like viral infections and introduced dsRNA. These primary Argonautes are responsible for recognition of target transcript but not silencing. Instead, they trigger an amplification process similar to the ping-pong cycle of piRNA pathway.

Upon target recognition, primary Argonautes trigger the production of secondary 22nt long siRNAs by RNA-Dependent RNA polymerases (RdRPs). Being RdRP products, these secondary siRNAs have a 5' triphosphate and a preference for Guanosine at 5'end hence termed 22G-RNAs. 22G-RNAs preferentially associate with a worm-specific clade of Argonaute termed WAGOs. WAGOs constitute the majority of Argonautes in *C. elegans* and exhibit high functional redundancy. WAGO proteins loaded with 22G-RNAs repress transcripts both post-

transcriptionally in the cytoplasm, and transcriptionally in the nucleus. Notable, gene silencing by WAGOs can be transgenerationally inherited (Ketting, 2011; Xu et al., 2018; Ozata et al., 2019).

The PRG-1 pathway

The piRNAs in *C. elegans* have a length of 21-nucleotide and a strong preference for Uridine at the 5' end. They are thus termed as 21U-RNAs. They bind to the PIWI clade Argonaute PRG-1 and are solely expressed in the germline. 21U RNAs are individually transcribed from the genome. A minority of 21U RNAs are generated through bidirectional transcription from the transcription start sites of protein-coding genes (Ruby *et al.*, 2006; Cecere *et al.*, 2012; H.-C. Lee *et al.*, 2012). However, most 21U RNAs are encoded as individual miniature genes. There are approximately 30,000 such 21U-RNA coding genes in the *C. elegans* genome (Batista *et al.*, 2008; H. C. Lee *et al.*, 2012). These genes are clustered on chromosome IV in two genomic regions of 2.5 Mb and 3.7 Mb, respectively (Ruby *et al.*, 2006; Batista *et al.*, 2008). Each 21U RNA producing locus contains a distinctive 5' upstream motif –CTGTTTCA – and depend on a dedicated machinery for its transcription. A multi-protein complex called upstream sequence transcription complex (USTC) recognizes the motif and recruits RNA-pol II for the transcription of 21U-RNAs (Weng *et al.*, 2019). Some components of USTC are also involved in biogenesis of small nuclear RNA suggesting that the machinery for 21U RNA transcription may have been evolved from other sRNA pathways (Beltran *et al.*, 2019).

The transcription of 21U RNA genes by RNA-pol II results in short and atypical transcripts. These precursors undergo sequential processing to yield mature 21U-RNAs with a 5' monophosphate and a 2'-O-methyl group at the 3' end of the piRNA. Subsequent to transcription, these precursors exit the nucleus and enter perinuclear granules. Here, the 5' end undergoes processing to eliminate the m7G cap and the first two nucleotides by the PUCH complex, as detailed in the subsequent section (Podvalnaya *et al.*, 2023). The resulting 5' end recessed precursors are then loaded onto PRG-1. After loading, the 3' – 5' exonuclease PARN-1 trims the 3' end until a length of 21 nucleotides is achieved (Tang *et al.*, 2016). In the final step, the RNA methyltransferase HENN-1 2'-O-methylates the 3' ends, a crucial modification for ensuring the stability of 21U-RNAs (Billi *et al.*, 2012; Kamminga *et al.*, 2012; Montgomery *et al.*, 2012).

The synergy of PUCH and PETISCO:

Through RNAi screens, the proteins TOFU-1, TOFU-2, and PID-1 were identified as critical for 5' end processing (de Albuquerque *et al.*, 2014; Goh *et al.*, 2014). Further studies revealed the presence of both TOFU-1/2 and PID-1 as part of larger protein complexes that work together to process the 5' end of 21U RNA precursor.

PID-1 operates as an effector protein within PETISCO, a cytoplasmic protein complex encompassing PID-3, ERH-2, TOFU-6, and IFE-3. With each subunit represented in duplicate, PETISCO forms an octameric assembly boasting a mass of approximately ~231 kDa (Perez-Borrajero *et al.*, 2021). Apart from PID-1, PETISCO houses a second effector protein, TOST-1. Intriguingly, the roles of PID-1 and TOST-1 distinctly define the function of PETISCO: PID-1-bound PETISCO functions as a stabilizer of 21U precursor transcripts and as a platform for 5' end processing (Rodrigues *et al.*, 2018; Zeng *et al.*, 2018; Podvalnaya *et al.*, 2023), while unpublished data from our laboratory indicates that TOST-1 bound PETISCO plays a crucial role in stabilizing histone mRNA and is essential for early embryogenesis.

PETISCO is an RNA binding complex and consequently, its subunits contain domains frequently found in RNA binding proteins. PID-3 contains a MID domain similar to those found in Argonautes and TOFU-6 contains an extended Tudor (eTudor) domain. In addition, both PID-3 and TOFU-6 have RNA recognition (RRM) domains. PID-3, TOFU-6 and ERH-2 are found only in the nematode

phylum. IFE-3 and ERH like proteins are deeply conserved in eukaryotes. IFE-3 is a homolog of the eukaryotic translation initiation factor Eif4e (Huggins *et al.*, 2020). ERH-2 is a paralog of conserved "enhancer of rudimentary" (Erh) factors that is involved in nuclear RNA degradation and processing, of for instance miRNA precursors, in several eukaryotes (Sugiyama *et al.*, 2016; Chul Kwon *et al.*, 2020; Fang and Bartel, 2020; Hutter *et al.*, 2020).

TOFU-1 and TOFU-2 exhibit domains similar to mammalian Schlafen ribonucleases. These proteins interact with the Schlafen domain-containing proteins SLFL-3 or SLFL-4, forming a trimeric nuclease known as PUCH. Genetic and biochemical studies show that PUCH is responsible for the cleavage of 5′-end 21U RNA precursors (Podvalnaya et al., 2023). The TOFU-1 subunit of PUCH interacts with the eTudor domain of TOFU-6 within the PETISCO complex. Animals devoid of *tofu-2* lack mature 21U RNAs, accumulating instead the unprocessed precursors. Interestingly, in the double mutant of *tofu-2* and *pid-1*, not only are the levels of mature 21U RNAs diminished, but the precursors also show a reduction. This implies that PETISCO stabilizes the precursors and facilitates their processing by PUCH (Podvalnaya *et al.*, 2023).

21U-RNA mediated silencing:

21U RNAs loaded onto PRG-1 can regulate expression of both transposable elements as well as endogenous genes (Batista *et al.*, 2008; Wang and Reinke, 2008). 21U RNAs can tolerate up to four mismatches. The imperfect base pairing of 21U RNAs with their targets can explain the low specificity of *C. elegans* piRNAs. Indeed PRG-1 RISCs can target a wide range of transcripts and are hence critical for germ-cell development. Animals lacking PRG-1 are not immediately sterile but rather show reduced fertility across generations, a phenotype referred to as Mortal Germline (Mrt) (Wang and Reinke, 2008; Simon *et al.*, 2014). Unlike PIWI proteins in other species, the endonuclease activity of PRG-1 does not seem to be required for target silencing (Bagijn et al., 2012). Instead, the recognition of a target by PRG-1 RISC initiates the production of 22G RNAs that direct WAGOs to silence transcripts in the germline. Intriguingly, without PRG-1, WAGO proteins start to acquire "erroneous" 22G-RNAs that also target genes that are required in the germline (Barucci *et al.*, 2020; Wahba, Hansen and Fire, 2021). An important function of PRG-1 thus appears to be to ensure WAGO proteins focused on the correct targets.

RNA-Dependent RNA Polymerases (RdRP) - Enhancer of RNAi

RdRPs are the primary catalysts driving endo-siRNA amplification in *C. elegans.* There are four RdRPs in *C. elegans* namely RRF-1, RRF-2, RRF-3 and EGO-1. Among these, RRF-3 is solely responsible for synthesizing primary 26nt siRNAs (Gent *et al.*, 2009, 2010; Han *et al.*, 2009; Pavelec *et al.*, 2009; Vasale *et al.*, 2010) whereas RRF-1 and EGO-1 produce the secondary 22G-RNAs (Smardon *et al.*, 2000; Aoki *et al.*, 2007; Sijen *et al.*, 2007; Maniar and Fire, 2011). The function of RRF-2 is unknown.

Biochemically, RdRPs are enzymes that utilize an RNA template to transcribe RNA by catalyzing the formation of phosphodiester bonds between ribonucleotides. Transcription begins at the 3'-end of the template and progresses in the 5' to 3' direction. The process requires presence of divalent metal ions and can occur either in a primer-dependent or primer-independent manner (Venkataraman, Prasad and Selvarajan, 2018). The core catalytic domain of eukaryotic RdRP is a double-psi β-barrel characterized by a metal-coordinating motif called "DXDGD," where 'X' can be any amino acid. This motif is also present in the universally conserved β' subunit of DNA-dependent RNA polymerases, suggesting that DdRP and RdRP diverged from a shared ancestor. However, the absence of other common domains between the two polymerases strongly indicates

that RdRPs evolved as an independent class of enzymes early in evolution (Iyer, Koonin and Aravind, 2003).

RdRPs are classified into two distinct groups: eukaryotic and viral. RNA viruses utilize RdRPs for processes such as genome replication and transcription (Wang and Metzlaff, 2005). In contrast, eukaryotic RdRPs constitute a separate family and are frequently associated with RNAi. RdRP genes are found in a wide range of eukaryotes, from diverse protists to multicellular organisms including fungi and plants (Zong *et al.*, 2009). Among animals, RdRPs are unevenly scattered. They are present in invertebrate animals, such as nematodes, lancelet and sea anemone but lost independently in most vertebrates, insects, mollusks and deuterostomes (Zong *et al.*, 2009). Based on sequence similarity, eukaryotic RdRPs are divided into alpha, beta and gamma clades, suggesting that the eukaryotic ancestor contained three copies of RdRP genes. Animal RdRPs, including nematodes, occupy clade alpha (Zong *et al.*, 2009).

Two models explain the need for eukaryotic RdRPs in accumulation of siRNAs. In the first model, RdRPs generate long complementary RNAs onto specific template RNAs that hybridize to form a dsRNA. These dsRNAs are cleaved by Dicer nucleases into shorter duplexes, which are subsequently processed into small-interfering RNAs, loaded onto Argonaute proteins and utilized for gene silencing. Experimental evidence for this model comes from studies in diverse organisms. RdRP purified from plants polymerize long dsRNA on ssRNA in-vitro, are functionally linked with Dicer proteins and are essential for accumulation for siRNA (Schiebel *et al.*, 1993; Tang *et al.*, 2003; Baulcombe, 2023). Further, unicellular RdRPs and Dicer have been implicated together in RNAi and related mechanism. For example, *T. thermophiles* RdRP purifies as a multi-subunit complex with Dicer and this interaction stimulates Dicer activity to generate 23 and 24nt sRNAs (Lee and Collins, 2007).

However, some RdRPs are capable of generating short RNAs along the length of an ssRNA template and thus can function in RNAi independent of Dicer. This has been observed in purified fungal RdRPs, such as from *Neurospora crassa*, *Thielavia terrestris* and *Myceliophthora thermophile.* These RdRPs, such as QDE-1 from *N. crassa*, tend to polymerize essentially short (9–21 nt) RNAs *in-vitro* (Fagard *et al.*, no date; Makeyev and Bamford, 2002; Qian *et al.*, 2016).

In *C. elegans*, RdRPs follow both models of small-RNA biogenesis. RRF-3 physically interacts with Dicer and is essential for production of 26nt 5' mono-phosphorylated RNAs (Duchaine *et al.*, 2006; Pavelec *et al.*, 2009; Vasale *et al.*, 2010). On the other hand, the activity of RdRP RRF-1 and EGO-1 is unaffected by absence of Dicer and the products of these RdRPs are 22nt long bearing 5' triphosphate termed 22G-RNAs (Smardon *et al.*, 2000; Aoki *et al.*, 2007; Sijen *et al.*, 2007; Maniar and Fire, 2011). Both types of small-RNAs will be described in subsequent sections.

The 26G-RNA pathway

In the initial stages of deep sequencing for small RNAs in *C. elegans*, a unique group of 26 nucleotide molecules was identified (Ruby et al., 2006; Ambros et al., 2003). These molecules were predominantly antisense to annotated genes and hence classified as endogenous-siRNAs. Additionally, these siRNAs exhibited a monophosphorylated 5' end as well as a preference for guanosine, resulting in their designation as 26G-RNAs (Ruby et al., 2006; Ambros et al., 2003).

26G-RNAs are abundant in the *C. elegans* germline. The majority of these are antisense to protein-coding genes and a small fraction map to pseudogenes. Two subclasses of 26G-RNAs sort into specific RISCs. (Han et al., in 2009). In the spermatogenic gonad, 26G-RNAs are associated with Argonautes ALG-3 and ALG-4. Conversely, in the oogenic gonad and embryos, 26G-RNAs are bound to the ERGO-1 Argonaute (Han et al., in 2009, Conine et al., in 2010, Vasale et al., in 2010). These subclasses regulate distinct, non-overlapping mRNA targets (see below).

Animals lacking 26G-RNAs show defective spermatogenesis and spermiogenesis. This results in a reduced brood size at 20 °C and the characteristic sterility at elevated temperature of 25°C (Conine et al., 2010). Mutant spermatocytes display nuclear abnormalities and chromatin bridges, which results in their arrest as multinucleate masses. The few spermatocytes that develop into mature sperm have reduced motility and eventually fail to activate. In addition, the mutant animals frequently show X-chromosome nondisjunction that results in a high-incidence of males (*him*). (Gent et al., 2009; Pavelec et al., 2009; Conine et al., 2010). Another notable phenotype of 26G-RNA mutants in the enhanced sensitivity to exogenous RNAi (Eri). This is proposed to result from the release of limiting DCR-1 proteins, which are shared among other RNAi pathways. (Sijen et al., 2001)

The ERI Complex mediates biogenesis of 26G-RNAs:

Central to production of 26G-RNAs is the generation of a dsRNA. *C. elegans* depends on RdRPs for the synthesis of dsRNA on single-stranded RNA templates. Out of the three RdRPs in *C. elegans*, RRF-3 plays a specialized role in the production of 26G-RNAs (Lee et al., 2006). 26G-RNAs are inherently antisense and span exon-exon junctions. In addition, the RdRP activity of RRF-3 has shown to be necessary for accumulation of 26G-RNAs. These results strongly suggest that RRF-3 directly synthesizes 26G-RNAs from a spliced mRNA template. Furthermore, RdRPs exhibit a preference for initiating transcription with Guanosine triphosphate (GTP) which is likely responsible for the 5' Guanosine bias of 26G-RNAs. (Aoki et al., 2007, Duchaine et al., 2006; Lee et al., 2006; Gent et al., 2009; Han et al., 2009; Pavelec et al., 2009; Conine et al., 2010; Gent et al., 2010; Vasale et al., 2010).

In addition to RRF-3, some of the earliest factors identified for 26G-RNA biogenesis were the RNAse III enzyme DCR-1, exonuclease ERI-1 and Tudor-domain protein ERI-5. Mutants lacking these genes failed to express 26G-RNAs and overexpressed the target mRNAs (Duchaine et al., 2006; Lee et al., 2006; Asikainen et al., 2007). Functional proteomics point towards the existence of these proteins in a large complex. For instance, MudPIT analyses on DCR-1, ERI-1 and ERI-5 detected ERI-1, ERI-5 as well as RRF-3. Furthermore, gel filtration chromatography to separate DCR-1 complexes in embryonic extracts resulted in the co-elution these ERI proteins together with DCR-1 and RRF-3 (Duchaine et al., 2006).

This collective assembly of proteins is denoted as the ERI Complex (ERIC). ERIC also includes the Dicer related helicase DRH-3, dsRNA binding protein RDE-4 and another protein, ERI-3, lacking identifiable domains (Tabara et al., 2002; Duchaine et. al. 2006). ERIC is estimated to have an approximate molecular mass of 850kDa (Duchaine et. al. 2006, Thivierge et al., 2012). In the following section, we will describe the functions of key ERIC factors **(Figure 1)**.

DCR-1 endonuclease processes RRF-3 generated dsRNA:

Dicer is a highly conserved endonuclease belonging to the RNAse III family of enzymes (Bernstein et al., 2001). In the presence of ATP, Dicer recognizes long dsRNA in a sequence independent manner and digests it into 22-23nt small dsRNAs. (Bernstein et al., 2001, Ketting et al., 2001). *C. elegans* encodes a single homolog of DCR-1, absence of which results in severe defects in germline development. DCR-1 initiates the various RNAi pathways in *C. elegans* by processing the input trigger dsRNA. Indeed, loss of DCR-1 depletes levels of microRNA, exo-siRNAs and 26G-RNAs. Consequently, animals lacking DCR-1 produce malformed unfertilized oocytes, are deficient for RNAi and display heterochronic phenotypes (Grishok et al., 2001; Ketting et al., 2001; Knight and Bass, 2001; Han et al., 2009.

In the context of the ERI complex, DCR-1 likely processes the dsRNA precursor generated by RRF-3 into shorter duplexes. Surprisingly, point mutations in the helicase domain of DCR-1 exclusively disrupt the 26G-RNA pathway. Missense mutations in this domain that affects ATP hydrolysis and helicase activity are sterile in 25 °C, exhibit *him* and *eri* phenotypes, but do not show heterochronic defects caused by disruption of miRNA pathway (Pavelec et al., 2009, Welker et al., 2010). The effect of this mutation on DCR-1 cleavage efficiency has been studied through *in-vitro* biochemical assays. The helicase domain is required for efficient cleavage of dsRNA with blunt or 5'-overhanging termini, but not of dsRNA with 3'-overhanging termini (Welker et al., 2011). This study also describes the production of 26nt species from the 5' of a blunt ended dsRNA. These findings strongly suggest that the dsRNA substrate generated by ERIC possesses a blunt end that is digested into 26nt long RNA through DCR-1.

3'- 5' exonuclease ERI-1b generates blunt ends for DCR-1:

The ERI complex includes ERI-1, a conserved protein with DEDDh-like exonuclease and SAP/SAF-box nucleic acid binding domains. ERI-1 was one of the earliest identified factor displaying enhanced sensitivity to RNAi. *eri-1* null mutants are viable, but they display a weak *Him* phenotype and sterility at 25 °C. *C. elegans* encodes two isoforms of ERI-1 but only the longer isoform ERI-1b is found to co-purify with DCR-1 (Duchaine et al., 2006). Consequently, expression of ERI-1b, rather than ERI-1a, rescues the temperature sensitive sterility and 26G-RNA levels in an *eri-1* deletion animal. Instead, ERI-1a is required for 3' end processing of ribosomal RNAs. (Kennedy et al., 2004; Duchaine et al., 2006; Gabel and Ruvkun, 2008; Han et al., 2009; Conine et al., 2010; Gent et al., 2010).

While there is a clear significance of ERI-1b in the 26G-RNA pathway, its precise function is not understood. The 3' to 5' exonuclease activity of ERI-1 is required for 26G-RNA production as mutations that abrogate this activity fail to re-establish expression of 26G-RNAs in *eri-1* deletion animals (Gabel and Ruvkun, 2008). *In-vitro* experiments on immuno-purified ERI-1 show its ability to partially degrade siRNAs that possess 2-nt 3' overhangs (Kennedy et al., in 2004). It is proposed that RRF-3 might synthesize dsRNA with heterogeneous termini that are subsequently polished by ERI-1 to create blunt termini that require DCR-1's helicase domain for processing.

Tudor domain ERI-5 recruits RRF-3 module to DCR-1:

The tandem Tudor domain protein ERI-5 was identified to be part of ERIC as it co-purifies with both DCR-1 and ERI-1. The Tudor domain motifs consist of approximately 60 amino acids and are recognized by their distinct bent anti-parallel β-sheet structure, formed by five β-strands that give it a barrel-like fold (Sprangers et al., 2003). Tudor domain proteins are highly conserved and play pivotal roles in RNA processing, stability, and translation (reviewed in Pek J et al., 2012). They act as molecular adaptors, facilitating the binding of substrates containing methylated arginine or lysine residues, thereby promoting physical interactions and the assembly of macromolecular complexes (reviewed in Pek J et al., 2012).

RRF-3, DRH-3 and DCR-1 were consistently the top ERI-5 interactors, suggesting that ERI-5 is most intimately associated with these components of ERIC. In addition, DCR-1, ERI-5, RRF-3, and DRH-3 co-immunoprecipitated with ERI-5. In the absence of ERI-5, RRF-3 immunoprecipitation did not retreive DCR-1. These results suggest that ERI-5, RRF-3 and DRH-3 form an RdRP module and ERI-5 is important for tethering this module to DCR-1 (Duchaine et al., 2006; Thivierge et al., 2012).

As an Eri gene, loss of ERI-5 leads to sperm-origin sterility and Eri phenotype. However, overall 26G-RNA levels only reduces by ~50% in *eri-5* mutant embryos relative to wild-type. This is due

to the activity of EKL-1, a paralog of ERI-5, which partially compensates for the loss of ERI-5. (Duchaine et al., 2006; Thivierge et al., 2012). A small fraction of EKL-1 co-purifies with RRF-3

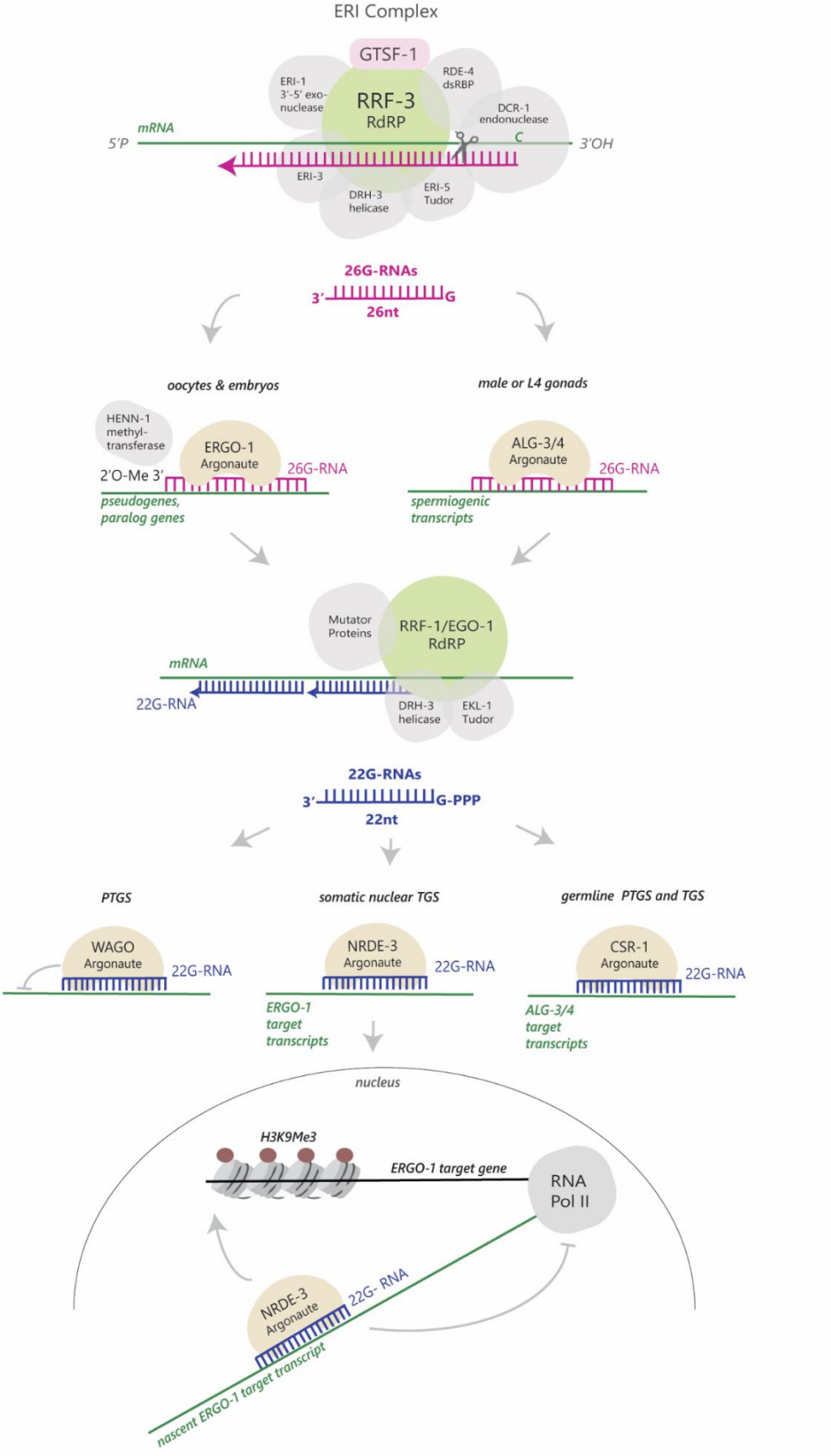


Figure 1 | Biogenesis and silencing mechanism of 26G-endo-siRNAs: The ERI complex composed of core RdRP module (RRF-3, DRH-3, ERI-5, and GTSF-1) along with accessory factors (DCR-1, ERI-1b, ERI-3, and RDE-4), generates two distinct subpopulations of 26G-RNAs. In the male germline and spermatogenic gonad, 26G-RNAs associate with ALG-3/4 to target spermatogenic transcripts. Conversely, in the oogenic gonad and embryos, 26G-RNAs associate with ERGO-1 and undergo 2'-O-methylation at their 3' end by HENN-1. Both classes of 26G siRNAs trigger secondary 22G siRNA generation by RdRPs RRF-1 and EGO-1, aided by DRH-3 and EKL-1. Mutator proteins form perinuclear foci required for 22G siRNA production. 22G-RNAs derived from the ALG-3/4 branch associate with the Argonaute CSR-1 to establish transgenerational memory of paternal gene expression. ERGO-1 derived 22G-RNAs further promote transcriptional gene silencing (TGS) through nuclear WAGO protein NRDE-3. This cytoplasmic secondary 22G siRNAs-NRDE-3 complex translocates to the nucleus, promoting histone H3K9 trimethylation and inhibiting transcription, while also inhibiting RNA polymerase II elongation. Nuclear RNAi signals are inherited across generations, directing H3K9 trimethylation in offspring. Finally, 22G-RNAs from both ALG-3/4 and ERGO-1 branch associate with other WAGO class Argonautes and negatively regulate target gene expression. PTGS" Post-transcriptional gene silencing; TGS: Transcriptional gene silencing; dsRBP: Double-stranded RNA-Binding protein.

equally well in both WT and *eri-5* backgrounds. A depletion of EKL-1 in the *eri-5* mutant background obliterates the remaining 26G-RNAs (Thivierge et al., 2012). EKL-1 itself exhibits robust interactions with both DRH-3 and another RdRP in *C.elegans*, known as RRF-1. RRF-1 functions downstream of RRF-3, its activity is required to transcribe 22G RNAs on the transcripts recognized by 26G-RNAs. Thus, tandem-tudor domain proteins in *C. elegans* specify RdRP modules for RNAi pathways (Thivierge et al., 2012).

ALG-3/4 Argonautes loaded 26G-RNAs are essential for sperm development and function:

In the spermatogenic germline 26G-RNAs are loaded onto the Argonaute proteins ALG-3 and ALG-4. ALG-3 and ALG-4 belong to the AGO clade (Seroussi et al., 2023) and are conserved in nematode clades I, III, IV and V (Sarkies et al., 2015). Both ALG-3/4 are expressed within the spermatogenic germline. Their expression begins in post-pachytene spermatocytes, where they are distributed across the cytoplasm and concentrated around peri-nuclear P granules. After spermatogenesis, ALG-3/4 are exclusively detected in the male germline, where it is confined to residual bodies (Han et al., 2009; Conine et al., 2010). Both Argonautes are absent from mature sperm (Conine et al., 2010).

The ALG-3/4 branch of the 26G-RNA pathway is essential for sperm development and function. Mutants show defects in sterility and the Him phenotype, but not the Eri phenotype. (Conine et al., 2010). ALG-3/4-dependent 26G-RNAs target genes are associated with gamete generation, particularly spermatogenesis. Many of these 26G-RNAs target transcripts encoding major sperm proteins, known to assemble into filamentous fibers in the pseudopod proteins (Asikainen et al., 2007; Han et al., 2009; Pavelec et al., 2009; Conine et al., 2010). This suggests a pivotal role for ALG-3 and ALG-4 in limiting excess accumulation of major sperm proteins, thereby promoting male fertility.

Interestingly, 26G-RNAs are preferentially generated from either the 5' or 3' ends of target transcripts. Gene Ontology analysis of the genes targeted at the 5' end showed associations with processes such as protein (de)phosphorylation, including phosphorus metabolic processes, protein modification, and dephosphorylation. On the other hand, an analysis of genes targeted at the 3' end revealed connections to defects in chromosome organization based on phenotype

enrichment. These include phenotypes such as progression of diakinesis during oogenesis and lack of nucleus in germ cells (Tsai et al., 2022). The 5′–targeted genes display a bias to be positive regulated by 26G-RNAs while the 3′–targeted genes have a tendency to be negatively regulated (Tsai et al., 2022).

The RNA tri-phosphatase PIR-1 has been implicated in the ALG-3/4 branch of 26G-RNAs. PIR-1 interacts with DCR-1 and exhibits 5' RNA tri-phosphatase activity, which results in the production of RNA di and mono phosphates from tri-phosphare RNA substrates. The absence of PIR-1 leads to a complete depletion of ALG-3/4 dependent 26G but has no discernible effect on ERGO-1 dependent 26G-RNAs (Chaves et al., 2021). Considering the enzymatic function of PIR-1, these findings indicate that PIR-1 plays a role in transforming the 5' triphosphate of nascent 26G-RNAs into 5' monophosphate, thereby allowing DCR-1 activity, promoting the stability of 26G-RNA and its subsequent loading onto ALG-3/4 (Chaves et al., 2021).

ERGO-1 Argonaute bound 26G-RNAs in oocytes and embryos:

The Argonaute ERGO-1 drives the oogenic branch of 26G-RNA pathway. ERGO-1 can be found in early larval stages but also displays prominent expression in the cytoplasm of oocytes and embryos. Targets of the ERGO-1 are largely somatic, most are protein-coding genes alongside a significant proportion of pseudogenes and poorly conserved duplicated genes (Fischer et al., 2011; Gent et al., 2010; Han et al., 2009; Vasale et al., 2010).

Notably, ERGO-1 bound 26G-RNAs are methylated on the 3’ ends by the methyltransferase HENN-1. This addition of a methyl group to the ribose ring is crucial for maintaining the stability of ERGO-1 26G-RNAs. (Billi et al., 2012; Kamminga et al., 2012; Montgomery et al., 2012). Despite these findings, the regulatory significance of ERGO-1 remains unclear The absence of ERGO-1 leads to the Eri phenotype (Vasale et al., 2010), which is not a direct defect in the oogenic germline. It is possible that the factors for RNAi such as helicases and dsRNA binding proteins are misregulated in the ERGO-1 mutants.

Further components required for 26G-RNA synthesis:

In addition to these factors, the zc3h12a-like ribonuclease proteins NYN-3 and RDE-8 have been implicated in cleaving target mRNAs upstream of ERIC activity. The initial cleavage by these nucleases primes recruitment of the ERI complex for siRNA synthesis. It also provides a 3’ end of mRNA for initiation of transcription by RdRP. Among these, NYN-3 deletion exclusively disrupts the ALG-3/4 class of 26G-RNAs and similarly RDE-8 deletion effects ERGO-1 class of 26G-RNAs (Tsai et al., 2022, Tsai et al., 2015. A largely unresolved question is how specific transcripts are selected by ERIC for the production of 26G-RNAs. An integral component of ERIC is the Gametocyte-specific factor-1 (GTSF-1) (Almeida et al., 2018). In the subsequent section, we introduce the GTSF-1 family of proteins and describe the molecular functions observed in various organisms.

ALG-3/4 or ERGO-1 loaded with 26G-RNAs mark the target transcript for secondary siRNA biogenesis. The final gene silencing is mediated by the secondary Argonautes called WAGOs through the 22G-RNA pathway **(Figure 1)**.

The 22G RNA and WAGO pathway

The predominant portion of small RNA population in *C. elegans* is composed of 22G endo-siRNAs. When a target transcript is identified by a primary Argonaute (RDE-1, PRG-1, ALG-3/4 or ERGO-

1), the transcript is not directly cleaved but rather designated as a template for the production of 22G-RNA by RdRPs **(Figure 1).** Thus, the 22G-RNA pathway is downstream of both 21U-RNA and 26-RNA pathway.

The initiation of 22G-RNA synthesis involves the Mutator protein complex, featuring the endoribonucleases NYN-1, NYN-2, and RDE-8. Subsequently, the ribonucleotidyltransferase RDE-3 catalyzes the non-templated addition of alternating uracil and guanine on the target RNA fragments. Once the final length of 17 nucleotides is achieved, The RdRPs RRF-1 or EGO-1 utilizes these fragments as templates to synthesize 22G-RNAs. These 22G-RNAs are synthesized from the entire gene body, forming a pool of randomly generated siRNAs that can be loaded onto WAGO Argonautes. The entire biogenesis process likely occurs within germline-specific peri-nuclear granules known as mutator foci (Zhang et al., 2011; Phillips et al., 2012, 2014; Tsai et al., 2015; Uebel et al., 2018; Shukla, Perales and Kennedy, 2021).

The resulting 22G-RNAs can enter various WAGO pathways. WAGO-1 localizes to cytoplasmic P granules and facilitates post-transcriptional silencing. WAGOs such as HRDE-1 and NRDE-3 shuttle into the nucleus, recruiting chromatin modifiers onto nascent transcripts. This process leads to transcriptional gene silencing by inducing the deposition of repressive chromatin marks and blocking RNA-pol II elongation (Burkhart *et al.*, 2011; Burton, Burkhart and Kennedy, 2011; Bethany A. Buckley *et al.*, 2012; Gu *et al.*, 2012; Mao *et al.*, 2015). Collectively, these 22G-RNAs target over 50% of the *C. elegans* genome. Not much is known, however, about the loading process of WAGO proteins.

A subset of 22G-RNAs is produced independently of the Mutator complex by EGO-1. These are bound by the germline-specific Argonaute CSR-1. The current model for CSR-1 regulation suggests that it licenses gene expression and counteracts the widespread targeting of WAGO 22G-RNAs derived from PRG-1 and 21U-RNAs (Seth et al., 2013; Claycomb et al., 2009; Maniar and Fire, 2011; Phillips et al., 2014; Smardon et al., 2000). Once the silencing of a specific transcript is established by a primary Argonaute, it can be inherited through the germ cells across multiple generations (Ashe et al., 2012; Bagijn et al., 2012; Lee et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012). This epigenetic silencing, mediated by WAGOs, can be considered as one of the most noteworthy features of *C. elegans* RNAi pathways.

3. Gametocyte Specific Factor-1 (GTSF-1): Small but Mighty Partners of RNAi

Gametocyte Specific Factor 1 (GTSF-1) proteins have garnered considerable attention due to their role in small-RNA function, fertility, and genome stability. These are small, ~20kDa proteins with tandem CHHC zinc-finger motifs located at its N-terminus and an unstructured C terminus. GTSF-1 proteins are widely conserved across metazoans, with some species encoding multiple paralogs (Ipsaro et al., 2022). Notably, GTSF-1 proteins are absent in fungi and plants, indicating their unique role in the animal kingdom (Tidow et al., 2009). As the name suggests, GTSF-1 proteins are expressed specifically in the germline and have a critical role in reproduction. The molecular functions of GTSF-1 has been characterized in *M.musculus (Mm)*, *B. mori (Bm)*, *D. melanogaster (Dm)* and *C. elegans (Ce)*. These animals possess multiple Gtsf paralogs; flies have four (Gtsf1/Asterix, CG14036, CG32625, and CG34283), mice have three (GTSF1, GTSF1L, and GTSF2), and silkworms have two (Gtsf1 and Gtsf1-like (Gtsf1L)) **(Figure 2).**

Molecular Function of GTSF1 in *M. musculus*

MmGTSF1 was discovered in *M. musculus* as a novel transcript with very high expression in adult mouse testes (Yoshimura et al., 2007). In male mice, MmGTSF1 expression begins around embryonic day 13.5 and persists into adulthood (Yoshimura et al., 2007). The gonads show a temporal expression of GTSF-1 in the early stages of germ cell differentiation. MmGTSF1 transcript is detected from the pre-leptotene stage of early meiosis I until the round spermatid stage (Yoshimura et al., 2007). Notably, MmGTSF1 is absent in mature sperm. Male mice homozygous for MmGTSF1 deletion are infertile and have a variety of gonadal abnormalities. The vas deferens and epididymis appear normal but the testes are highly reduced in size and the spermatocytes have enlarged nuclei with scattered chromatin (Yoshimura et al., 2009). Without MmGTSF1, the differentiation of spermatocytes arrests in meiotic prophase I and the cells undergo apoptosis.

MmGTSF1 is also found in female mice. Unfertilized eggs and ovaries show moderate expression of MmGTSF1. However, unlike male mice, females depleted of MmGTSF1 are fertile and show no obvious defects in the ovaries (Yoshimura et al., 2009). This lack of phenotype questions the significance of MmGTSF1 in the female germline. However, as several animals including humans have retained the expression of GTSF1 in the female gonads (Huntriss et al., 2017, Liperis et al., 2013), there may be a beneficial role of GTSF1 in oocytes or embryos that is yet to be identified.

The localization of MmGTSF1 is solely within cytoplasmic granules, together with various piRNA pathway components. MmGTSF1 associates with both MIWI and MILI complexes, demonstrated through enrichment of MILI, MIWI, MIWI2, as well as the Tudor proteins TDRD1 and TDRD9 in GTSF1 immunoprecipitates. Notably, MmGTSF1 specifically co-localizes with piP body components MIWI2 and TDRD9. The absence of MmGTSF1 disrupts the localization of piP body proteins, and in vitro experiments confirm direct binding between recombinantly expressed MmGTSF1 and TDRD9. Intriguingly, MmGTSF1's localization remains unaffected by mutations in MIWI2, while MILI mutants cause complete mislocalization of MmGTSF1 (Yoshimura et al., 2018).

One of the conserved functions of MmGTSF1 lies in its ability to enhance the endonuclease activity of PIWI Argonautes, which, by nature, are relatively inefficient enzymes **(Figure 2)** (Arif et al., 2022). In this context, MmGTSF1 is essential for the coordinated action of MILI and MIWI2 in secondary piRNA biogenesis. Without MmGTSF1, MIWI2 fails to load piRNAs and MILI loses its slicing capacity, both resulting in a depletion of secondary piRNAs. This causes the depression of LINE-1 and IAP transposons and apoptotic cell death of spermatocytes (Yoshimura et al., 2018).

M. musculus also express two paralogs of GTSF1: MmGTSF1L and MmGTSF2. These are also abundantly expressed in the testes but unlike MmGTSF1, MmGTSF1L and MmGTSF2 are expressed only at the stage of round spermatids and elongating spermatids, respectively. Both paralogs resemble MmGTSF1 at the N-terminus however, the central and C-terminus is quite divergent. They also interact with both MIWI and MILI, however single and double knockout mice show no defects in spermatogenesis and are fertile (Takemoto et al., 2016).

Remarkably, the paralogs of GTSF1 can even distinguish between the various PIWI proteins. MmGTSF1 enhances the cleavage of target RNAs by both MILI and MIWI. Conversely, the paralogs GTSF1L and GTSF2 effectively boost target cleavage by MIWI but not MILI. (Arif et al., 2022).

Molecular Function of Gtsf1 in *B. mori*

In the silkworm, *Bombyx mori*, Gtsf1 is found ubiquitously in all tissues, with its highest expression occurring in the gonads. Among the gonadal tissues, the ovaries display a higher Gtsf1 expression compared to the testes. Furthermore, Gtsf1 mRNA levels are notably elevated in early female

embryos as opposed to male embryos (Chen et al., 2020). Gtsf1 localizes to both nucleus and cytoplasm, specifically within the cytoplasmic granules (Chen et al., 2020).

BmGtsf1 is essential in the germline piRNA pathway. Mutants of Gtsf1 have highly atrophied testes and ovaries and both sexes are infertile. In both gonads, Gtsf1 localizes in cytoplasm and nucleus as well as interacts with BmSiwi. Consequently, Gtsf1 mutants show downregulation of piRNAs and de-repression of transposable elements (Chen et al., 2020).

Furthermore, BmGtsf1 also plays a significant role in the somatic piRNA pathway. In silkworms, the piRNA pathway is linked to the process of sex determination. A piRNA precursor named *Feminizer* (*Fem*) serves to suppress the expression of the *Masculinizer* gene (*Masc*) and thereby promotes female development. When essential components of the piRNA pathway, such as BmSIWI, are absent, females exhibit partial sex reversal. Notably, female mutants of BmGtsf1 experience reduced levels of Fem piRNAs and develop male-specific anatomical structures. In contrast, the morphology of male mutants remains unaffected (Chen et al., 2020).

Similar to MmGTSF1, BmGtsf1 homologs potentiates the cleavage activity of silkworm PIWI proteins **(Figure 2)** (Izumi et al., 2022). Gtsf1 and Gtsf1L co-localize with Siwi and Ago3 respectively as well as specifically immunoprecipitate their respective PIWI partner (Izumi et al., 2022; Bronkhorst 2023). The orthogonal activation of PIWI function by BmGtsf1 paralogs is particularly evident in the context of *Fem* mediated repression of *Masc*. Siwi is loaded with *Fem* piRNAs and is responsible for cleaving *Masc* mRNA thereby producing *Masc* piRNAs. These *Masc* piRNAs are subsequently loaded into Ago3, which conversely cleaves *Fem* mRNA and amplify *Fem* piRNA (Kiuchi et al., 2014). Loss of Gtsf1 and Gtsf1L specifically reduces *Masc* and *Fem* piRNAs, respectively (Izumi et al., 2022). This underscores the specificity of the pairing between Gtsf1 and PIWI in silkworms.

It is hypothesized that Gtsf1 stabilizes a target-bound confirmation of the PIWI complex which enhances the catalytic activity of PIWI (Arif et al., 2020). Why PIWI proteins need a co-factor for cleavage, while other Argonaute proteins can do without is not yet clear. However, the interaction between BmGtsf1L and BmAgo3 is mediated and enhanced through a Tudor domain protein called Vreteno, a molecular scaffold that also can attract unloaded Siwi protein. This suggests that the Gtsf1 dependence may allow the activation of PIWI proteins only in specific environments, such that their cleavage products can be used as substrates for new piRNAs, instead of being degraded by non-specific RNase activities (Bronkhorst et al., 2023).

Molecular Function of Gtsf1 in *D. melanogaster*

The *Drosophila melanogaster* equivalent of Gtsf1 was detected in three separate screenings aimed at identifying mutants with deficiencies in the piRNA pathway (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013). The absence of DmGtsf1 has a significant impact on the germline development and fertility of female flies. In contrast to mice and silkworms, loss of DmGtsf1 does not lead to a decrease in overall piRNA levels. Notably, PIWI proteins, when not loaded with piRNAs, become unstable and undergo degradation. In DmGtsf1 mutant PIWI protein levels and their distribution remain unaffected, providing further evidence of the presence of piRNAs in these cells (Donertas et al., 2013, Ohtani et al., 2013).

DmGtsf1 is predominantly nuclear where it co-localizes with DmPiwi. Ectopically expressed DmGtsf1 can immunoprecipitate DmPiwi, underlining their close association. The subcellular localization of DmGtsf1 is also dependent on DmPiwi, such that when mutant DmPiwi relocates to the nuage, DmGtsf1 follows suit. The primary function of DmPiwi is to suppress transposon transcription by assembling histone methyl-transferases onto target loci **(Figure 2)**. Strikingly, loss of DmGtsf1 results in a remarkable loss of Piwi-mediated transcriptional silencing of

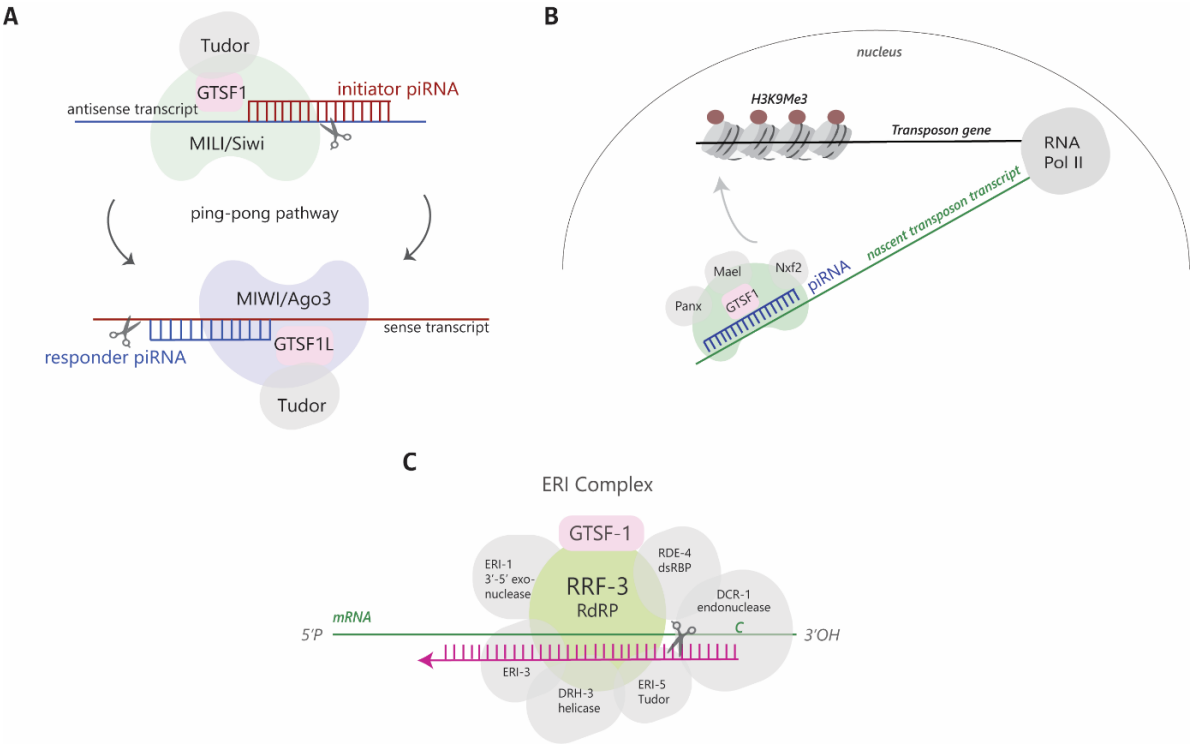


Figure 2 | Diverse functions of GTSF1: (A) GTSF1 proteins drive the ping-pong cycle of Piwi-piRNA mediated silencing. In *M. musculus* and *B. mori*, GTSF1 and its paralogs potentiate the slicing activity of piRNA-loaded Piwi silencing complexes (piRISCs). GTSF1 associates with piRISCs loaded with initiator piRNAs in the cytoplasm, facilitating slicing of antisense transcripts. The sliced targets generate responder piRNAs, which can then be loaded into another Piwi clade Argonaute molecule, such as MIWI or Ago3. Paralogs of GTSF1, such as GTSF1-like (GTSF1L) or GTSF2, enhance cleavage by the second piRISC, which is utilized for slicing of complementary sense transcripts. Together, GTSF1 and Piwi drive multiple rounds of slicing, amplifying the silencing effect until target transcripts are exhausted (Arif et al., 2022).

(B) GTSF1's Role in Nuclear piRISC-Mediated Heterochromatin Formation. In *D. melanogaster*, the nuclear piRNA pathway leads to the transcriptional silencing of transposons. GTSF1 directly interacts with target-engaged piRISC and facilitates recruitment of additional factors. These recruited factors such as Panoramix (Panx), Nxf2, and Maelstrom (Mael) coordinate with chromatin modifiers to silence transposons co-transcriptionally, promoting the formation of heterochromatin at targeted loci.

(C) GTSF-1 assembles an RdRP complex for endo-siRNA biogenesis. In *C. elegans*, GTSF-1 forms a pre-ERI complex together with RRF-3 and ERI-5. GTSF-1 and ERI-5 are both required to incorporate RRF-3 into the ERI complex for 26G-RNA biogenesis.

transposons in the nucleus. This is evident in the reduction of H3K9me2 marks and an increase in the occupancy of PolII at target transposons (Donertas et al., 2013, Ohtani et al., 2013).

Nuclear DmPiwi is non-catalytic and associates with a number of close factors such as Panoramix and Maelstrom to achieve silencing **(Figure 2)**. Importantly, in the absence of GTSF-1, these proteins lose their interaction with DmPiwi (Onishi et al., 2020). Recent unpublished work show that Gtsf1 binding allows a conformation switch of DmPiwi upon target recognition that facilitates the assembly of various cofactors onto DmPiwi, thereby enabling silencing of nascent transcripts.

GTSF-1 domain structure and interaction surfaces

The CHHC domain of GTSF-1 act as independent units that folds around a central zinc ion. These motifs are also found in two other protein families: the TRM13 tRNA modification enzymes and the minor spliceosomal protein U11-48K (Tidow et al., 2008). Both TRM13 and U11-48K proteins have RNA binding activity suggesting that CHHC motifs of GTSF-1 may potentially bind RNA (Tidow et al., 2008).

The solution structure of the first 115 residues of *M. musculus* GTSF1 has been determined by NMR spectroscopy. The two tandem Zn fingers have strand-strand helix architectures and are connected by a-helix-containing linker. In contrast, the N and C termini are intrinsically disordered (Ipsaro et al., 2021). According to the NMR structure, the first Zn finger contains a stretch of positively charged residues. In an in-vitro crosslinking coupled to pulldown experiments, these residues were implicated in GTSF-1 binding RNA (Ipsaro et al., 2021). Furthermore, in *M. musculus*, RNase A treatment considerably reduced the interaction of MmGTSF1 with MIWI2, suggesting that RNA in the MIWI2 complex is required for the binding of GTSF1 (Yoshimura et al., 2018).

On the other hand, the central region of GTSF-1 has been repeatedly implicated in binding to PIWI. While the central region of GTSF-1 is highly divergent, a stretch of negatively charged amino acids with two embedded aromatic residues is conserved in several homologs. In *D. melanogaster*, mutations of these aromatic residues (W89 and Y98) fail to recuse TE depression of endogenous protein in OSS cells. Furthermore, purified peptides of wild-type CTD pull down PIWI proteins from OSS nuclear extract, but this interaction is significantly reduced with CTD peptides mutant for the aromatic residues (Donertas et al., 2013). This was also observed in *M. musculus* where mutation of W107A most efficiently abrogated pulldown of PIWI from testis lysates (Yoshimura et al., 2018). In addition, mutations of W98A, W107A and W112A in the recombinant MmGTSF1 reduced the stimulatory effect of MmGTSF1 on MIWI in a concentration-dependent manner. (Arif et al., 2022).

These studies support the idea that central region of GTSF1 binds PIWI through conserved aromatic amino acids. However, one cannot eliminate the presence of additional factors that mediate this interaction. Strikingly, in *B. mori,* the conserved tryptophan residue (W99) within the C-terminal tail of BmGtsf1L is essential for interaction with BmVreteno and to enhance binding to BmAgo3 (Bronkhorst et al., 2023). In several species, Tudor domain proteins are observed to closely interact with GTSF-1 as well as PIWI. This strongly suggest that other species may also have a similar reliance on Tudor proteins in the association between GTSF-1 and PIWI.

Molecular Function of GTSF-1 in *C. elegans*

Surprisingly, the *C. elegans* homolog of GTSF-1 takes a different path than other species, as it does not form an association with a PIWI Argonaute but rather engages with the RdRP RRF-3. The loss of GTSF-1 in *C. elegans* has no discernible impact on piRNA levels and does not result in an increase in transposon expression. Instead, *C. elegans* GTSF-1 plays a vital role in the assembly of the ERI complex. It accomplishes this by binding to RRF-3 and ERI-5 and enables the formation of the ERI complex **(Figure 2)**. Unlike vertebrate homologs, *C. elegans* GTSF-1 does not contain the central aromatic residues and rather bind RRF-3 through the CHHC fingers. In absence of GTSF-1, RRF-3 and ERI-5 do not associate with the other ERIC proteins and this leads to a complete loss of 26G-RNAs from both sperm and oocyte (Almeida, Andrade-Navarro, and Ketting 2019b; Ozata et al., 2018).

In summary, this diversity in function underscores the versatility and adaptability of GTSF-1 in different genetic contexts.

4. Comparative Genomics and Evolutionary Insights through nematodes *C. briggsae* and *P. pacificus*

Nematodes, belonging to the phylum Nematoda, are a highly diverse and numerous group of organisms, capable of adapting to a wide range of ecological niches and lifestyles. They inhabit diverse environments, with lifestyles ranging from free-living to parasitic. According to a comprehensive nematode phylogeny, based on the Small Subunit Sequences of the ribosomal RNA gene (SSU), there are five distinct clades within the phylum Nematoda **(Figure 3)** (Blaxter et al., 1998). These clades are categorized into Enoplea and Chromadorea. Enoplea comprises two subclasses, Dorylaimia and Enoplia, corresponding to clade I and II, respectively. In contrast, Chromadorea constitutes a single subclass, Chromadoria, encompassing nematodes from clades III to V **(Figure 3)**.

The Clade V contains two notable genera, *Caenorhabditis* and *Pristionchus*. *Pristionchus* belongs to the *Diplogastridae* family that is paraphyletic to the *Rhabditida* family of *Caenorhabditis*.

The species *Caenorhabdits briggsae* diverged from *C. elegans* approximately 100 million years ago (Stein et al., 2003; Stevens et al., 2022). *C. elegans* and *C. briggsae* are found all across the globe in various niches, mostly consisting of rotting vegetation. However, unlike *C. elegans*, wild strains of *C. briggsae* shows greater molecular differentiation leading to the formation of distinct genetic groups separated by latitudes (Cutter et al., 2006). These phylogeographic groups demonstrate a range of local adaptations, one of which is the effect of temperature on animal development. Tropical strains of *C. briggsae* have been observed to display higher lifetime fecundity compared to temperate strains at elevated temperatures (30 °C). However, at lower temperatures (14 °C), they exhibit lower fecundity than temperate strains (Prasad et al., 2011). This is in stark contrast to *C. elegans*, which do show much genotypic variation with geography (reviewed in Cutter et al., 2015).

The distribution of *Pristionchus* strains is also extensive, with over 700 strains collected globally, mainly from the Japanese islands and La Réunion. *Pristionchus* nematodes are necromantic: living in close association with scarab beetles and the Colorado potato beetle, often in a species-specific manner. Specifically, the species *Pristionchus pacificus* is found within the oriental beetle *Exomala orientalis* (Herrmann et al., 2006).

P. pacificus diverged from *C. elegans* around 300 million years ago (Dietrich et al., 2008). It has a life cycle similar in speed to C. elegans, but it distinguishes itself by undergoing an embryonic molt from J1 to J2 before hatching. It is morphologically very different from *C. elegans* and *C. briggsae* (Sommer et al., 1996a). They can be cultured in the lab on OP50 bacteria, however, in the wild, *P. pacificus* have distinct feeding habits. *C. elegans* and other members of *Rhabditidae* specializes in bacterial consumption and are equipped with a posterior pharyngeal bulb grinder. *P. pacificus* lacks this grinder but possesses teeth-like structures at the pharynx tip, allowing it to feed on a wide range of sources, including bacteria, fungi, and other nematodes (Sommer et al., 1996a). Furthermore, *P. pacificus* exhibits greater resistance to pathogenic bacteria compared to *C. elegans* and *C. briggsae* (Wei et al., 2003; Rae et al., 2008).

Both *C. briggsae* and *P. pacificus*, like *C. elegans*, display an androdioecious mode of reproduction, a mating system comprised of males and self-fertile hermaphrodites. This is rare among clade V nematodes. In total, there are only three androdioecious species in *Caenorhabditis* and seven in *Pristionchus* genera (Kiontke et al., 2004; Rodelsperger et al., 2018). In fact, the sibling species of *C. elegans* and *C. briggsae* have male/female mating systems, implying that hermaphroditism likely arose independently in these species.

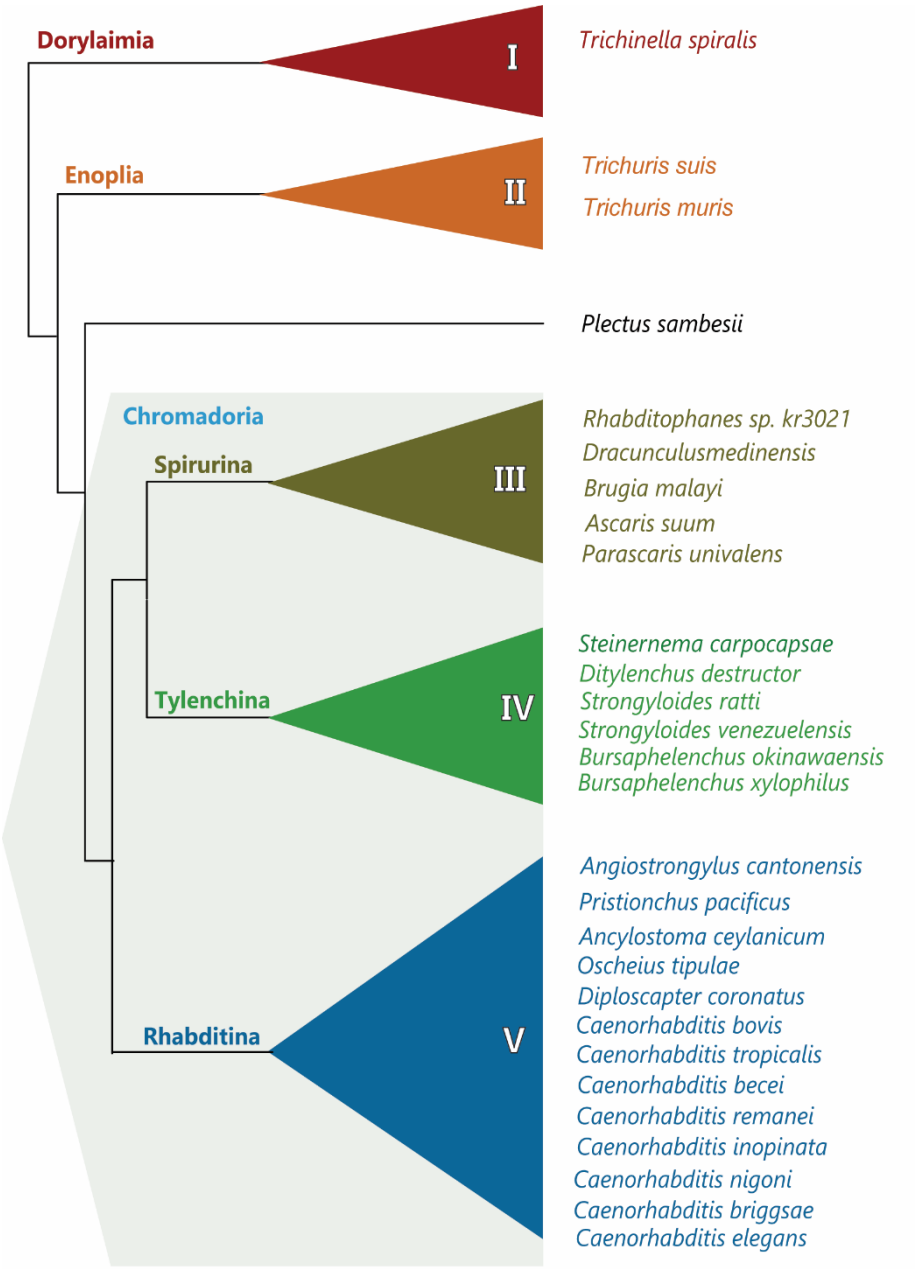


Figure 3 | Overview of Nematode Phylogeny: This schematic phylogenetic tree depicts the general organization of the phylum Nematoda, showcasing five major clades. Adapted from Blaxter and Koutsovoulos, 2015 the tree is reconstructed based on small subunit ribosomal RNA sequences. Notably, the *Pristionchus* and *Caenorhabditis* genera are classified within clade V of nematodes. Each clade encompasses a diversity of parasitic and free-living nematodes, as well as species exhibiting various mating systems, including male/female and hermaphroditic. The use of consistent color coding allows for easy identification of species belonging to different nematode clades.

C. briggsae and *P. pacificus* have high-quality genome sequences available. *C. elegans* was the first metazoan to have its complete genome sequenced in 1998 (The *C. elegans* Sequencing Consortium 1998). The sequencing of *C. briggsae* genome started along with *C. elegans* but was completed only in 2003 (Stein. et. al. 2003). 10 years later, *P. pacificus* followed suit with a genome based on shotgun sequencing (Dieterich et al., 2008). The genomes of *C. elegans* and *C. briggsae* are similar in size, around 100Mb and 106Mb, respectively, with *C. briggsae's* larger size attributed to repeat

sequences (Stein et al., 2003). They also have a similar number of protein-coding genes (approximately 20,000). In contrast, *P. pacificus* possesses a larger genome of 169Mb, more transcripts, and protein-coding genes (around 29,000), with a higher GC-content (43%) (Dieterich et al., 2008; Borchert et al., 2010). There is significant protein-level resemblances between *C. elegans* and both *P. pacificus* and *C. briggsae*. However, each of these also possess a number of species-specific genes. *C. briggsae* has approximately 2,000 species-specific genes, while *P. pacificus* has a much higher count of around 11,000 (Dieterich et al., 2008; Borchert et al., 2010).

In conclusion, the nematodes, *C. briggsae* and *P. pacificus*, make excellent models for comparative studies with *C. elegans* due to their rare hermaphroditic mode of reproduction, well-documented genome sequences, and unique genomic features, providing valuable insights into nematode evolution and adaptation.

Small-RNA pathways of Nematodes

The general framework of RNAi is shared among species from all clades of life. However, RNAi pathways are evolutionarily labile and show significant mechanistic differences between organisms. For example, while Ago proteins are conserved throughout species, their numbers range from one in *Schizosaccharomyces pombe*, five in *Drosophila*, eight in humans, ten in *Arabidopsis* to twenty-seven in *C. elegans* (Tolia and Joshua-Tor, 2007). In addition, RNAi pathway genes crucial for anti-viral and anti-TE defense show high rates of adaptive protein substitution (William H Palmer, Hadfield and Obbard, 2018). This is particularly evident in the phylum nematoda. Among the wild strains of *C. elegans*, genes responsible for mediating RNAi display substantial expression variation in comparison to other genes within the genome. Additionally, these genes exhibit pronounced allelic divergence and instances of strain-specific pseudogenization at the sequence level (Chou et al., 2023).

Notably, the rapid evolution of RNAi pathway genes seems to have created significant variations even among short evolutionary time scales. Studies from within clade V nematodes (*C. remanei*, *C. brenneri*, *C. briggsae* and *P. pacificus*) show that while these species express 21U- 26G- and 22G-RNAs similar to *C. elegans*, the individual targets are not conserved (De Wit *et al.*, 2009; Shi *et al.*, 2013a; Tu *et al.*, 2014; Holz and Streit, 2017). Moreover, *C. remanei* and *C. brenneri* express almost triple the number of 21-U RNAs and encode two times the number of Ago proteins than *C. elegans* (Shi *et al.*, 2013b).

The miRNA specific Argonaute, miRNA sequences and their developmental regulation exhibit a high degree of conservation across all nematode clades (de Wit et al., 2009; Sarkies et al., 2015). In contrast, PIWI protein and piRNAs are only deeply conserved in Clade V nematodes. Outside of Clade V, Piwi is only detected in plectids, which occupy a basal position relative to Clades III, IV, and V **(Figure 3)**. This indicates that the piRNA pathway appears to have been independently lost in Clades III and IV (Beltran et al., 2019; Sarkies et al., 2015). Furthermore, nematodes from Clades I and II also lack a homolog of PIWI or 21U RNAs. In most *Caenorhabditis* species, 21U RNAs tend to cluster in a manner similar to *C. elegans,* specifically on repressed chromatin regions located on chromosome IV. However, in other nematodes such as *P. pacificus*, 21U RNAs are distributed across various chromosomes, often within actively transcribing genes. (Beltran et al., 2019).

How do nematodes without PIWI Argonautes and piRNAs silence genome parasites such as transposons? Unlike PIWI, the RdRP RRF-3 is present in all nematode clades. Other RdRps RRF-1 and EGO-1 as well as 22G- RNAs are also well conserved at least in clades III-V (Wang *et al.*, 2011; Sarkies *et al.*, 2015; Holz and Streit, 2017). It has been suggested that transposon control mechanisms within "piRNA-less" nematodes is achieved through a diverse range of endo-siRNAs with varying lengths. For instance, the parasitic nematode from Clade III, *Ascaris suum*, possesses five WAGOs that interact with 5' triphosphate 22-24G RNAs, targeting both repetitive sequences

and fully mature mRNAs. Interestingly, outside of Clade V, only *A. suum* has 26G-RNAs that resemble those found in *C. elegans*. These 26G-RNAs are bound by the *A. suum* equivalent of ALG-4 and effectively repress specific transcripts in the testis, mirroring the mechanism observed in *C. elegans*. (Zagoskin et al., 2022). Thus, RdRPs play a central role in mediating RNAi mediated gene regulation in nematodes.

5. Aim of thesis

C. elegans GTSF-1 deviates from the conventional PIWI association by binding with the RdRP RRF-3. This observation gains significance as at least four independent lineages of nematodes have lost the entire PIWI pathway, yet retained RRF-3-like RdRPs. Our project seeks to understand the factors conferring specificity of *C. elegans* GTSF-1 to RdRP RRF-3 and investigate the conservation of this association across nematodes. To achieve this, we are characterizing GTSF-1 from the clade V nematodes *C. briggsae* and *P. pacificus* using transgenic techniques, transcriptomics and proteomics. We also aim to explore the 26G-RNA pathway of these nematodes in terms of functions, composition of ERIC and target genes. The combined studies will describe a general molecular function of nematode GTSF-1 proteins in small RNA biology. Further, it will reveal how small RNA pathways have evolved during nematode evolution in relation to genome characteristics and gene-regulatory needs of the individual species.

RESULTS

1. Identification of Nematode GTSF-1 orthologues

We identified orthologues for the nematode *gtsf-1* gene via reciprocal BLAST, with *C. elegans* GTSF-1 (CelGTSF-1) as the input sequence against the genomes of *C. briggsae* and *P. pacificus*. We identified one unambiguous hit in both species that we denoted as CbrGTSF-1 and PpaGTSF-1, respectively. Subsequent multiple sequence alignment (MSA) using Clustal Omega revealed that all three GTSF-1 sequences featured two CHHC Zinc finger motifs and an acidic C-terminal region **(Figure 4A)**. Consistent with observations in vertebrate GTSF-1 sequences (Ipsaro et al., 2021), the Zinc finger regions was similar across species, while the C-terminal region displayed high variation. Notably, PpaGTSF-1 contains an extended N-terminus with an additional 100 residues **(Figure 4A).**

We predicted GTSF-1 protein structure using Alpha-Fold2 and found a consistent pattern in all orthologues, with N-terminus CHHC Zinc fingers arranged within two alpha helices and the C-terminus comprising intrinsically disordered regions (IDRs) interspersed by alpha-helices **(Figure 4B).** In addition, the Zn-finger regions of CbrGTSF-1 and PpaGTSF-1 showed perfect alignment with CelGTSF-1 **(Figure 4C).** We noticed that PpaGTSF-1 displayed an extra four amino acids (EGER) between Cysteine and Histidine of the first Zn finger that created a disordered loop between the helices **(Figure 4C).** Nevertheless, considering the overall sequence conservation, all Zn fingers maintain the potential to be functional. Thus, *C. briggsae* and *P. pacificus* harbor a single GTSF-1 orthologue, each with the potential to be functionally significant.

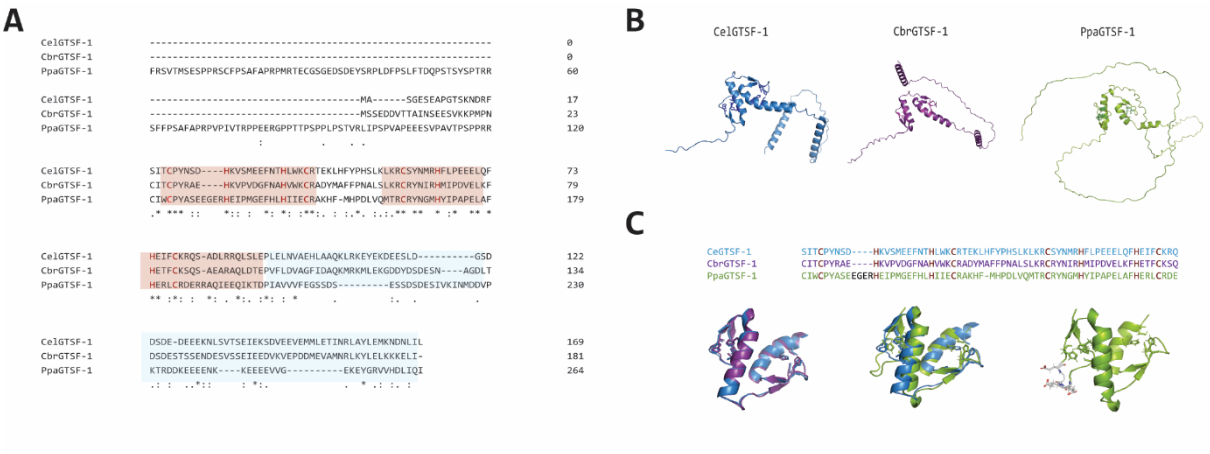


Figure 4 | Sequence and Structure of Nematode *gtsf-1* orthologues: (A) Multiple Sequence Alignment of GTSF-1 sequences was performed using Clustal Omega (EMBL-EBI), with CHHC Zinc fingers highlighted in red and C-terminal domain in blue. **(B)** Structure predictions for GTSF-1 were generated using the AlphaFold2 v2.3.2 monomer models. CHHC residues are displayed in stick format. **(C)** The alignment of AlphaFold2 (AF2) predicted structures of GTSF-1 is depicted. The top panel exhibits region of GTSF-1 used for alignment, color-coded based on species. In the bottom panels, CelGTSF-1 is aligned to CbrGTSF-1 or PpaGTSF-1, with CHHC residues presented in stick format. The third bottom panel exclusively showcases PpaGTSF-1, emphasizing the four distinct EGER residues in stick format along with their side chains.

2. Nematode GTSF-1 have distinct structural differences compared to *M. musculus* and *D. melanogaster* homologs

Co-precipitation experiments conducted on GTSF1 in *M. musculus* and *D. melanogaster* have pinpointed critical tryptophan residues in the C-terminal region that play a crucial role in interacting with PIWI. Additionally, a study from CSHL unveiled the solution structure of the first

100 amino acids of *M. musculus* GTSF-1. They identified a positively charged ridge on the first zinc finger, emphasizing that five basic residues along this ridge were essential for *in-vitro* RNA binding. These findings led to the proposal of a modular function for Mmus and Dmel GTSF-1, where the first CHHC zinc finger facilitates RNA interactions, and the C-terminus binds to PIWI (Dönertas, Sienski, and Brennecke 2013; Ohtani et al., 2013; Ipsaro et. al. 2022)

While these crucial residues show relatively high conservation in vertebrate and invertebrate GTSF-1 orthologues, a surprising lack of conservation in nematodes was observed. Notably, the highly conserved tryptophan is absent in nematode GTSF-1 **(Figure 5A).** Although the N-terminus does contain positively charged residues, these are distributed across both zinc fingers **(Figure 5B).** This difference is evident in the AF2 predicted structure of nematode GTSF-1. In particular, *C. elegans* and *C. briggsae* GTSF-1 exhibit a pocket of positive residues between Zinc-finger 1 and 2. This pocket is shorter and less pronounced in *P. pacificus* GTSF-1, which instead harbors highly negative residues on the opposite side of this pocket **(Figure 5B).**

In accordance with this difference in charge distribution, CelGTSF-1 purified from Sf9 cells did not co-purify RNA, as confirmed by Dr. Jonathan Ipsaro at CHSL, New York **(Figure 5C).** This suggests that CelGTSF-1 may not function as an RNA-binding protein. These results display clear sequence and structural distinctions in *C. elegans* GTSF-1 compared to *M. musculus* GTSF-1, potentially linking the shift in binding partner from PIWI to an RdRP. Further, these differences are well conserved in *C. briggsae* and relatively conserved in *P. pacificus*, suggesting that other nematode GTSF-1 proteins might also interact with RdRP.

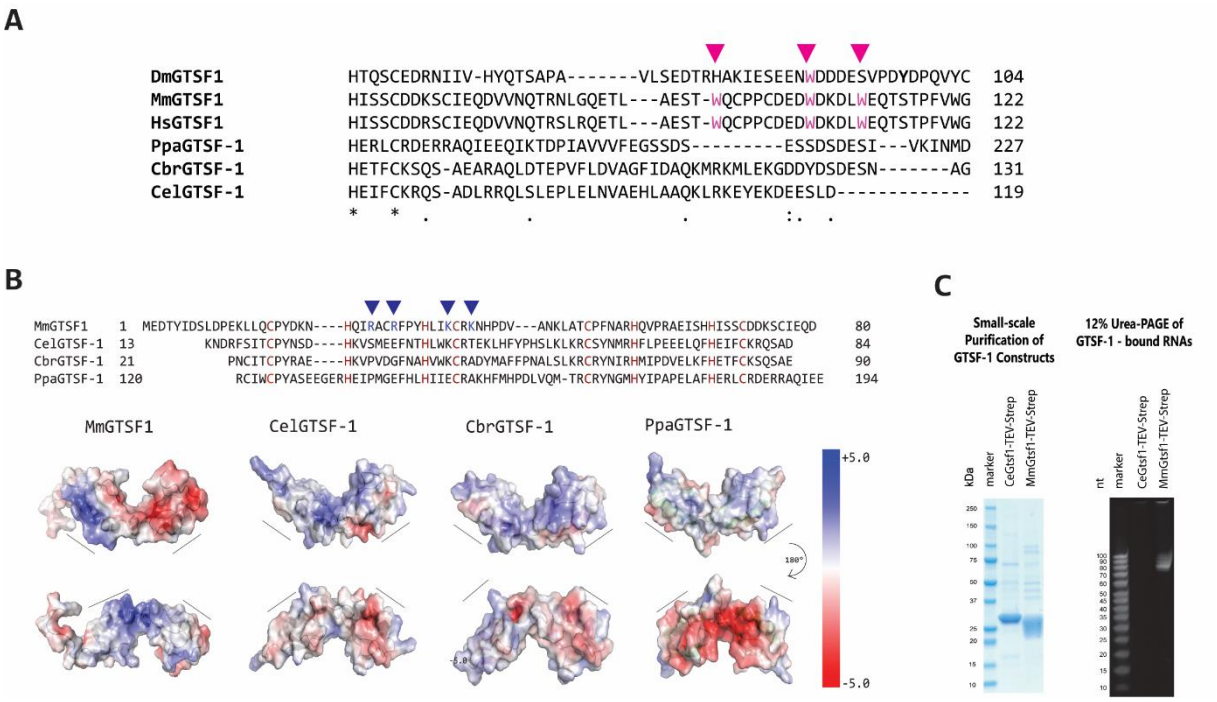


Figure 5 | Variations in Nematode *gtsf-1* homologs: (A) Multiple Sequence Alignment of GTSF-1 sequences. **(B)** Electrostatic charge distribution on GTSF-1: The electrostatic surface is scaled from -5kBT in red to +5kBT in blue. The top panel highlights the residues included in the structures with Zn finger CHHC in red. *M. musculus* GTSF1 (MmGTSF1) exhibits a positively charged ridge on ZnF1 as reported by Ipsaro et al. (2020). CelGTSF-1and CbrGTSF-1 display positive charge between Znf1 and Znf2, with a less pronounced effect in PpaGTSF-1 Instead, PpaGTSF-1features a highly negative charge on the opposite side of this region. **(C)** Left: SDS-PAGE of affinity-purified MmGtsf1 or CelGTSF-1 constructs. Each lane has 2ug protein. Right: Urea-PAGE analysis of RNAs that co-purify with GTSF-1 constructs. Recombinant MmGTSF-1 co-purifies nucleic acids 70-90 nucleotides in size. Each lane was loaded with 1ug protein.

3. Phenotypes of nematode *gtsf-1* mutants

Our next question was whether the observed structural similarities in nematode GTSF-1 were extended to function. We used the CRISPR/Cas9 system to create *gtsf-1* mutants in *C. elegans*, *C. briggsae* and *P. pacificus* **(Figure 6A).** In the previous study, Almeida et al., used mutant alleles of *Cel-gtsf-1* generated by inserting pre-mature stop codons. In our study, we generated deletion alleles for *C. elegans* and *C. briggsae* where we chose protospacer sequences at the first and third exons. In *C. elegans,* this resulted in a complete gene deletion. In *C. briggsae*, we obtained an allele with a significant deletion in the gene, preserving the first 30 amino acids in-frame before introducing point mutations leading to premature stop codons. In *P. pacificus*, we targeted the first exon with a single protospacer, producing two indel alleles that caused frameshifts and premature stop codons. The *xf208* allele in *P. pacificus* retained the first 100 amino acids in-frame, encoding the extended N terminus. In each species, a majority of the protein was deleted, especially the region containing the CHHC Zinc fingers **(Figure 6A).**

The phenotypes of *C. elegans* GTSF-1 mutants have been previously described (Almeida et al., 2018). *C. briggsae* and *P. pacificus* GTSF-1 mutants are viable and show no obvious morphological defects. They have a striking reduction in brood size at normal temperature of 20˚C. When cultured at an elevated growth temperature, *C. briggsae* and *P. pacificus* GTSF-1 mutants are completely sterile. **(Figure 6B).** At these temperatures, mutants mostly produce unfertilized oocytes. These phenotypes are identical to those observed in *C. elegans* (Almeida et al., 2018). Thus, there is a consistent role of GTSF-1 in nematode fertility, particularly at high temperatures.

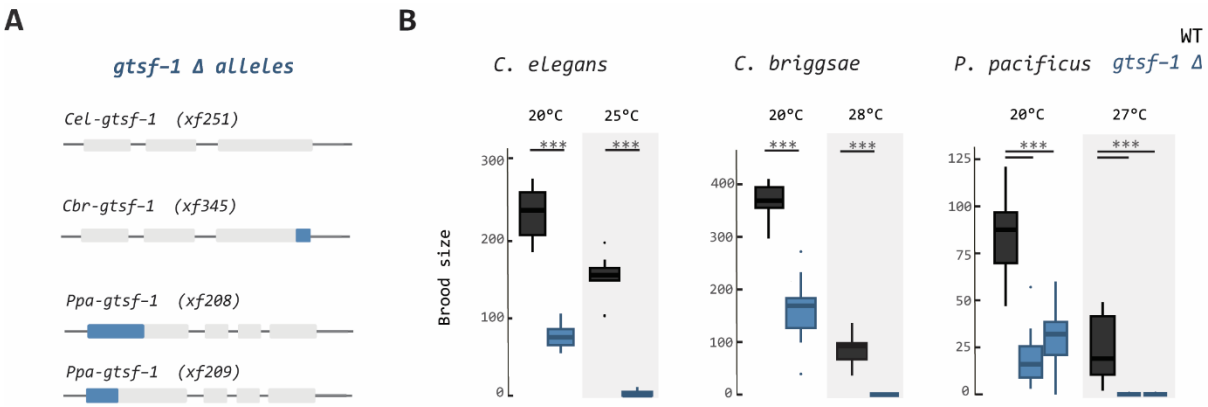


Figure 6 | Nematode *gtsf-1* is required for fertility: (A) Illustration representing mutant alleles of *gtsf-1* in nematodes. Exons are represented by boxes, and filled colors, blue or grey, indicate regions of the gene retained or deleted, respectively. *Cel-gtsf-1* (*xf251*) corresponds to a complete gene deletion**. (B)** Brood-size quantification of wild-type and *gtsf-1* mutant animals, cultured at permissive temperature of 20 ℃ or elevated temperature of 25 - 28 ℃ (*P. pacificus* data from Miguel Almeida). p-values calculated by Wilcoxon rank sum test.

4. Small-RNA sequencing of nematode *gtsf-1* mutants

In *C. elegans*, the temperature-sensitive sterility phenotype is associated with impaired 26-RNA pathway. To better understand this phenotype in other nematodes, we focused on elucidating the role of *gtsf-1* in the small-RNA landscape. To achieve this, small-RNA sequencing was conducted on wild-type (WT) and *gtsf-1* Δ animals from *C. elegans*, *C. briggsae*, and *P. pacificus*. To capture the diverse ALG-3/4 and ERGO-1 classes of 26G-RNAs, small RNAs were sequenced from both gravid-adult (GA) animals, which contains the oocytes and embryos, as well as the L4 larval stage

where the animals exclusively produce sperm. M. Almeida performed the experiment on *P. pacificus* exclusively on embryos in 2019, which is being repeated on GA and J4 animals.

Our results revealed minimal variations in miRNA and piRNA levels across the life stages **(Figure 7A).** These variations were more prominent in L4 samples compared to GA samples. This discrepancy could potentially stem from issues in synchronization of L4 samples upon collection, due to slower growth of *gtsf-1* Δ animals. In contrast, the impact of *gtsf-1* deletion on 26G-RNAs was striking, indicating a significant loss in both life stages across the two species **(Figure 7A).** The majority of affected 26G-RNAs were associated with protein-coding genes and pseudogenes, with a few instances observed in miRNA and transposon loci. The targets of 26G-RNAs are explored in detail in the subsequent section.

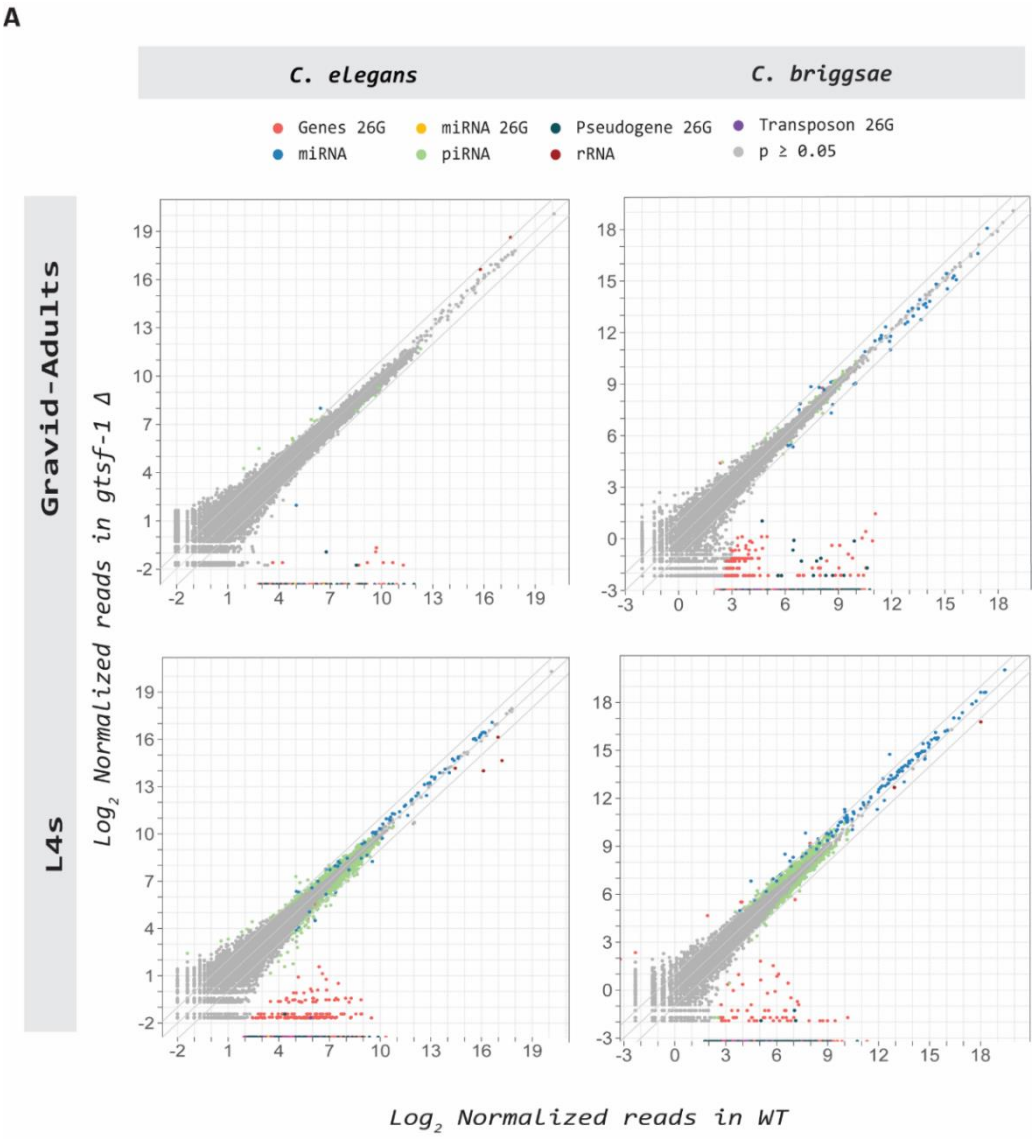
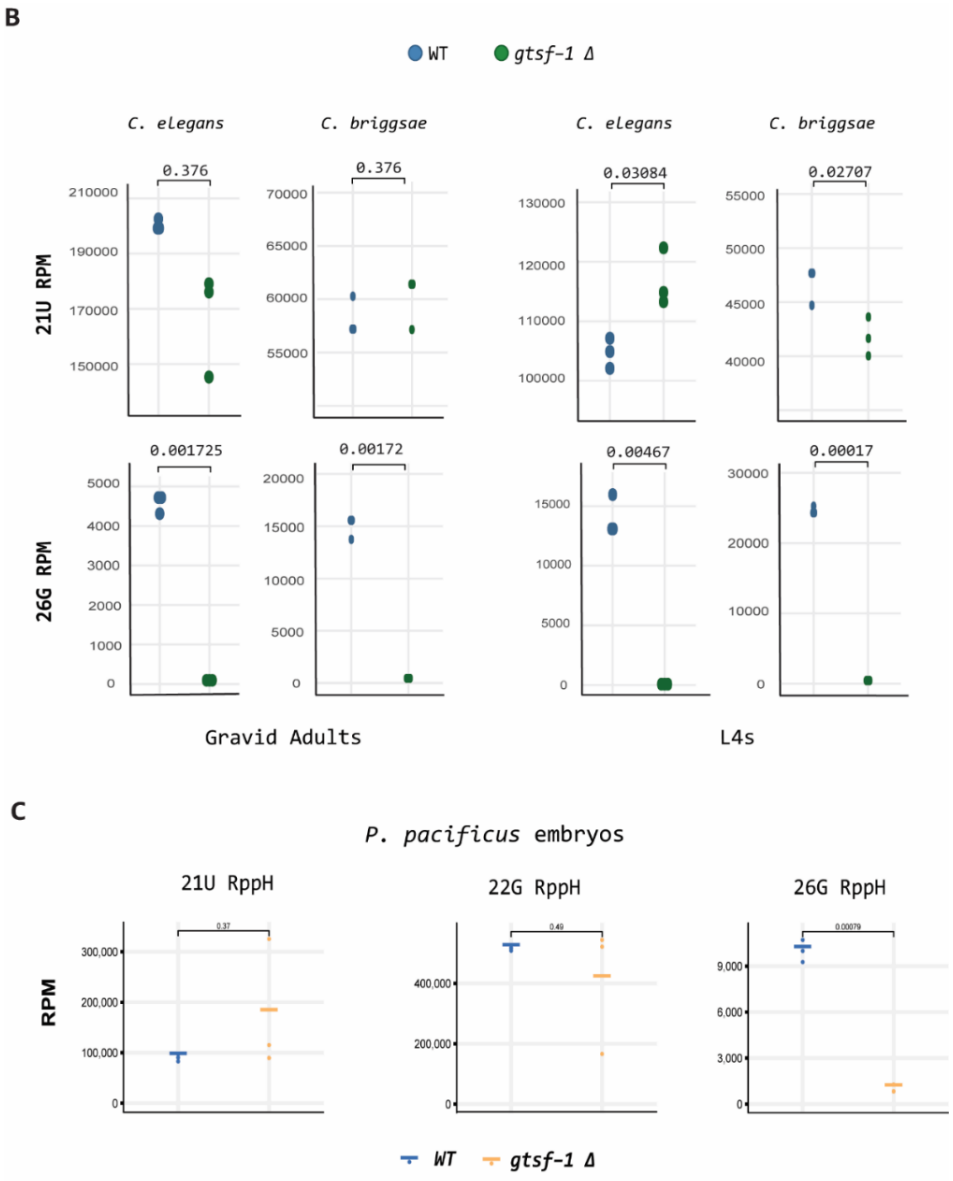


Figure 7 | Small-RNA sequencing of nematode *gtsf-1* Δ animals: (A) TinyRNA-produced scatterplot comparing all sample groups, with the x-axis and y-axis representing Log2 Mean normalized 26G-RNA counts per gene from WT and *gtsf-1* Δ samples, respectively. Strains used in the experiment: *C. elegans* WT (N2), *C. briggsae* WT (AF16), *Cel-gtsf-1* (*xf251*) and *Cbr-gtsf-1* (*xf345*). The normalized counts were generated using DEseq2. Classes of differentially expressed small RNAs are highlighted based on a P-value cutoff of 0.05. Genes labeled as 26G indicate 26G-RNAs mapping antisense to protein-coding genes. Similarly, miRNA 26G, pseudogene 26G, and transposon 26G indicate 26G-RNAs mapping antisense to loci annotated with the same. (Figure continued on next page)

Figure 7 | (B) 21U-RNAs and 26G-RNA levels from *C. elegans* or *C. briggsae* Gravid-Adult and L4 animals are normalized to total mapped reads and presented in reads per million (RPM). p-value calculated using Welch-t-test. **(C)** Reads-per-million (RPM) of 21U-RNAs, 22G-RNAs and 26G-RNA from RppH treated libraries of *P. pacificus* WT (PS312) and *gtsf-1* (*xf208*) embryos [collected and anlyzed by Miguel Almeida]. p-value calculated using Welch-t-test



Interestingly, species-specific differences were noted, as *C. briggsae* exhibited higher levels of 26Gs in both stages, contrasting with lower levels of 21U RNAs **(Figure 7B)**. The 21U-RNA RPM levels observed in *C. briggsae* align with those reported in a separate study by Pastore et al., (2022). Finally, in accordance to Almeida et al, *P. pacificus* embryos demonstrated a consistent pattern of complete loss of 26G-RNAs with minimal changes to 21U and 22G RNAs upon *gtsf-1* deletion **(Figure 7C)**. These results emphasize the specific requirement of *gtsf-1* in the 26-RNA pathway across nematodes.

5. GTSF-1 interacts with components of PETISCO and ERI complex in *C. briggsae* Gravid-Adults

In a previous study, the interactors of GTSF-1 in *C. elegans* were described using the immunoprecipitation (IP) coupled to label-free quantitative mass-spectrometry (MS) technique (Almeida et al.,2019). The most enriched interactor was found to be the RdRP RRF-3. In a slight overexpression of GTSF-1 using a transgene, GTSF-1 IP also enriched other known interactors of RRF-3, such as ERI-5 and DRH-3.

To compare the interactors of *C. elegans* GTSF-1 with those of *C. briggsae* GTSF-1, we conducted a similar immunoprecipitation experiment in *C. briggsae*. Affinity-purified antibodies against endogenous CbrGTSF-1 were developed, and whole-worm extracts of Gravid-Adult stage hermaphrodites were used to immunoprecipitate CbrGTSF-1. We performed the samples in quadruplicates and used the non- specific antibody against Immunoglobulin G (IgG) as a negative control **(Figure 8A)**. Under these conditions, we found 126 proteins to be co-precipitating with CbrGTSF-1. These protein cross the cutoff of at least two-fold enrichment in test IP compared to control as well as a p-value greater than 0.05. Our bait CbrGTSF-1 was highly enriched in the test IP but not in control, which validates a successful immunoprecipitation experiment.

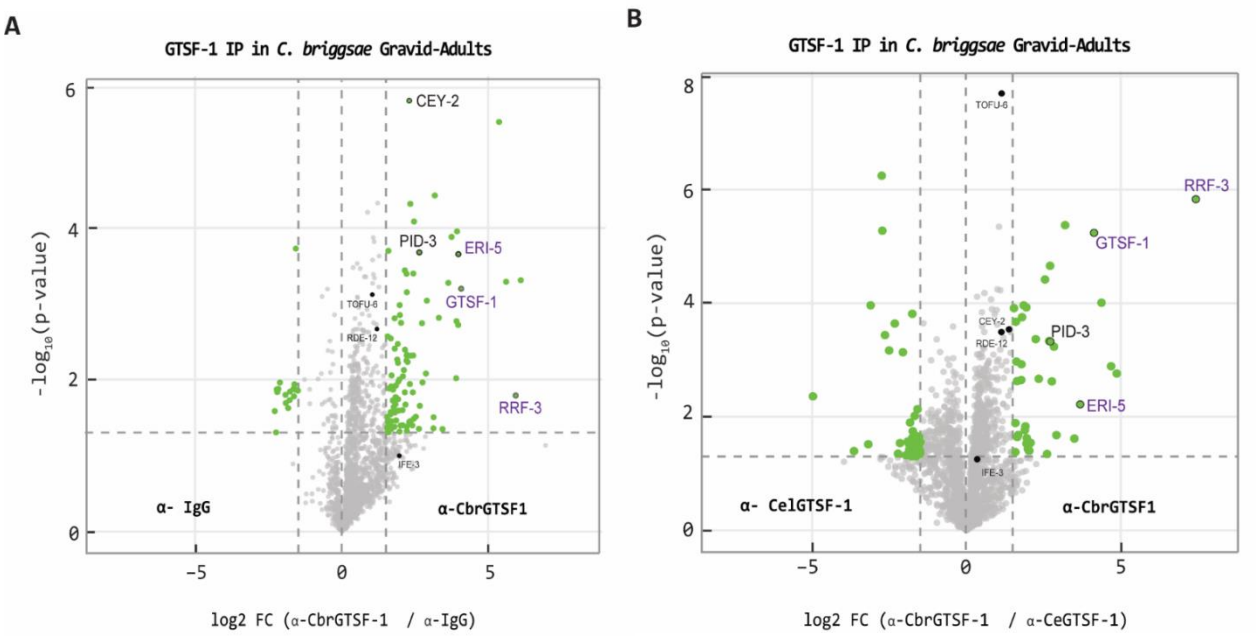


Figure 8 | IP-MS on GTSF-1 in *C. briggsae* Gravid-Adult animals: Volcano plots representing label-free proteomic quantification of quadruplicate Immunoprecipitates (IPs) from various *C. briggsae* WT Gravid adult extracts. 1.5mg of extract was incubated with 2ug of bead-bound antibody and subjected to three washes with a Tris buffer. After final wash, beads were boiled in SDS and analysed on a mass-spectrometer. **(A)** Two antibodies were used for this experiment: α-IgG is affinity purified goat anti-rabbit IgG with minimum cross-reactivity to other proteins to collect background interactions; α-CbrGTSF-1 is affinity purified antibody against the antigen of interest CbrGTSF-1. **(B)** The affinity purified α-CelGTSF-1 was used as a control as it does not show cross-reactivity to CbrGTSF-1.The x axis represents the median fold enrichment of individual proteins precipitated with one antibody over another. The y-axis indicates -Log10 (p-value) of observed enrichments. Dotted lines represents thresholds at p=0.05 and 1.5-fold enrichment. Data points highlighted in green represent statistically significant interactors.

As observed in *C. elegans*, RRF-3 was one of the most enriched interactor of GTSF-1. In addition, ERI-5 was also a statistically significant interactor. We did not detect other ERI proteins or DCR-1. However, RDE-12, an RNA helicase functioning downstream of the ERGO-1 pathway in *C. elegans*, was mildly enriched (Yang et al., 2014). Notably, no AGO protein was found in our experiment **(Figure 9A).**

Functional Profiling of CbrGTSF-1 Interactors

To better characterize the list of interactors, we conducted functional profiling using the g:GOst tool from the g:Profiler web server (Kolberg et al., 2023; Reimand et al., 2007). This gene set enrichment analysis utilizes an input gene list, mapping genes to known functional information sources, and identifying statistically significant enriched terms. We used the default settings for the analysis where g:GOSt utilizes the set of all protein-coding genes with at least one domain annotated as a background **(Figure 9B)**.

Significant portions of the genes in the input list had Gene Ontology terms related to transmembrane transporters **(Figure 9C)**. When we examined the ontology for cellular components, we observed an enrichment of hits in categories such as envelope, organelle membrane, and actin cytoskeleton. While the biological relevance of these findings cannot be dismissed, it is plausible that they may be incidental outcomes of the preparation of whole worm lysate samples, potentially leading to a nonspecific enrichment of membrane proteins.

To refine our focus, we specifically prioritized Gene Ontology terms with molecular functions and biological processes that are more pertinent to the investigation of GTSF-1 in small-RNA pathways **(Figure 9D)**.

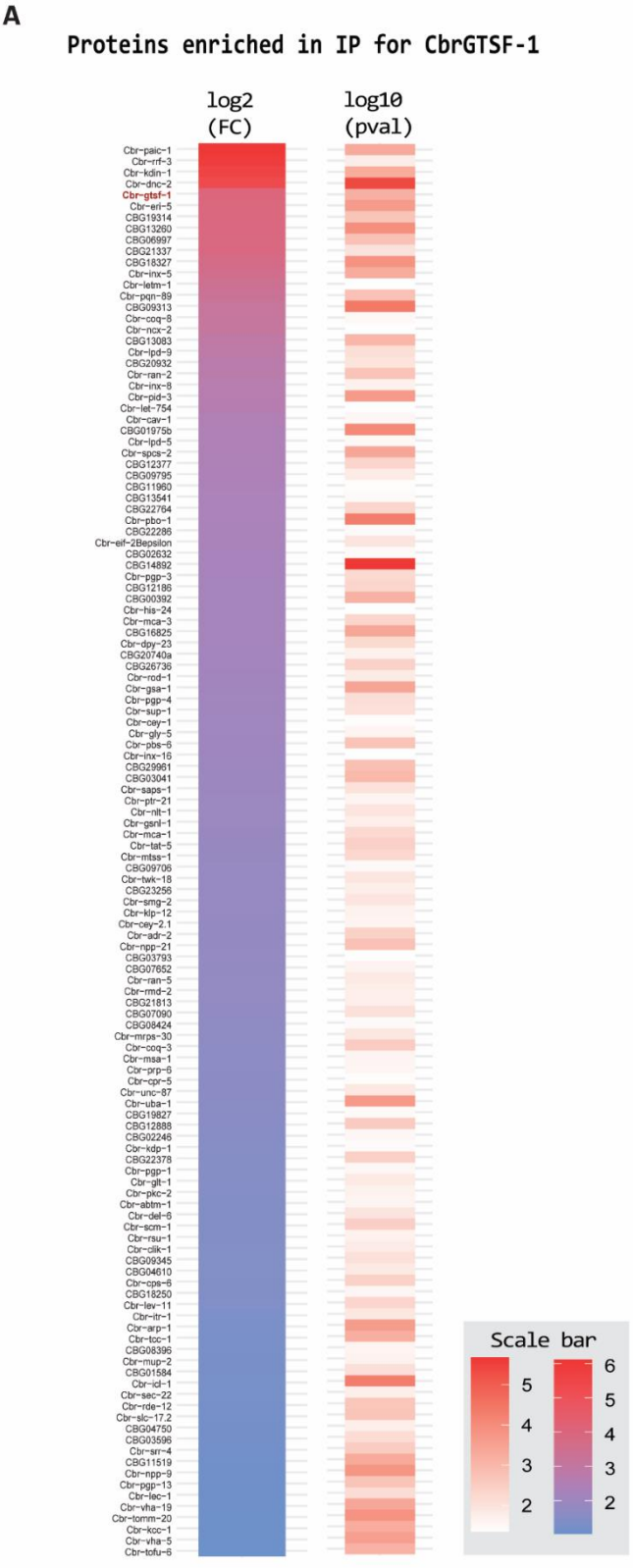
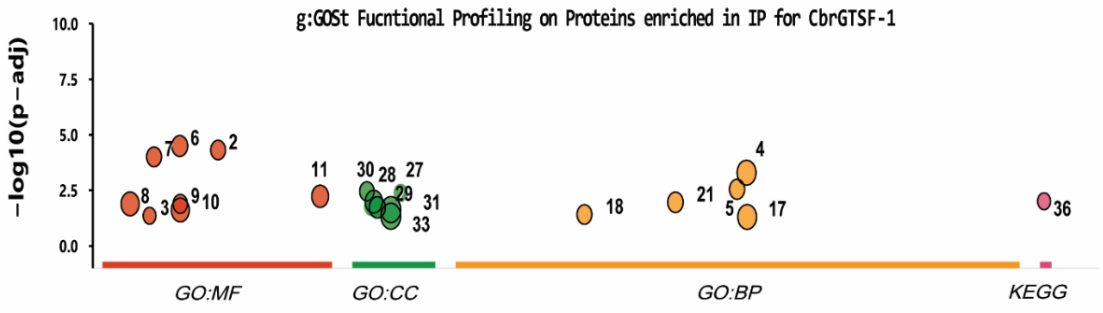


Figure 9 | Gene Ontology (GO) Term Enrichment Analysis of Proteins Co-immunoprecipitated with CbrGTSF-1: (A) (Page 45) Heat map depicting proteins significantly enriched in the immunoprecipitation (IP) for CbrGTSF-1 compared to IgG. Significance thresholds applied are Fold-Change (FC) greater than 2 and adjusted p-value greater than 0.05. **(B)** Manhattan plot illustrating g:GOst enrichment analysis results. Only significant enrichments with a p-value > 0.05 are shown. The x-axis represents functional terms, grouped and color-coded by data sources, while the y-axis displays adjusted enrichment p-values in a negative log10 scale. Circle sizes correspond to the respective term sizes. **(C)** Detailed results corresponding to the Manhattan plot, with IDs referencing individual terms displayed in (B). (Continued on next page)

B



C

id	source	term_id	term_name	term_size		p_value
2	GO:MF	GO:0042626	ATPase−coupled transmembrane transporter activity	94		4.8e−05
3	GO:MF	GO:0015085	calcium ion transmembrane transporter activity	31		4.4e−02
4	GO:BP	GO:0051179	localization	1751		5.0e−04
5	GO:BP	GO:0048878	chemical homeostasis	122		2.8e−03
6	GO:MF	GO:0022804	active transmembrane transporter activity	185		3.2e−05
7	GO:MF	GO:0015399	primary active transmembrane transporter activity	102		9.8e−05
8	GO:MF	GO:0005215	transporter activity	889		1.3e−02
9	GO:MF	GO:0022853	active monoatomic ion transmembrane transporter activity	72		1.3e−02
10	GO:MF	GO:0022857	transmembrane transporter activity	858		2.4e−02
11	GO:MF	GO:0140657	ATP−dependent activity	356		5.8e−03
17	GO:BP	GO:0051234	establishment of localization	1568		4.9e−02
18	GO:BP	GO:0019725	cellular homeostasis	97		3.8e−02
21	GO:BP	GO:0042592	homeostatic process	144		1.1e−02
27	GO:CC	GO:0031975	envelope	213		1.8e−02
28	GO:CC	GO:0031967	organelle envelope	213		1.8e−02
29	GO:CC	GO:0031090	organelle membrane	476		1.0e−02
30	GO:CC	GO:0015629	actin cytoskeleton	78		3.5e−03
31	GO:CC	GO:0043229	intracellular organelle	3490		2.3e−02
33	GO:CC	GO:0043226	organelle	3565		4.6e−02
36	KEGG	KEGG:02010	ABC transporters	33		9.7e−03

Figure 9 | (D) Table presenting relevant Gene Ontology terms associated with CbrGTSF-1 interactors. These terms are not significantly enriched in the input dataset. Source: GO subontology MF: Molecular Function, BP: Biological Processes, and CC: Cellular Compartment. Term_ID: Unique identifier associated with the specific Gene Ontology (GO) term. Adjusted_p_value: A custom algorithm g:SCS is used for computing multiple testing correction for p-values gained from GO and pathway enrichment analysis and reducing significance scores. The default threshold of a=0.05 is applied. Term_Size: Total genes associated with a GO term. Query_Size: Genes in the input dataset linked to the specific GO term. Intersection_Size: Number of overlapping genes between the input dataset and the GO term. Effective_Domain_Size: Statistically adjusted measure that considers the genomic background.)

D

source	term_name	term_id	adjusted_p_value	negative_log10_of_adjusted_p_value	term_size	query_size
GO:MF	catalytic activity, acting on RNA	GO:0140098	1	0	244	89
GO:MF	catalytic activity, acting on a protein	GO:0140096	1	0	1328	89
GO:MF	RNA polymerase activity	GO:0097747	1	0	32	89
GO:MF	double-stranded DNA endonuclease activity	GO:1990238	1	0	1	89
GO:MF	catalytic activity, acting on a nucleic acid	GO:0140640	1	0	368	89
GO:MF	5'-3' RNA polymerase activity	GO:0034062	1	0	32	89
GO:MF	ribonucleotide binding	GO:0032553	1	0	1169	89
GO:MF	chromatin DNA binding	GO:0031490	1	0	26	89
GO:BP	embryo development	GO:0009790	0.419219294	0.377558738	261	85
GO:BP	piRNA processing	GO:0034587	0.727294877	0.138289472	6	85
GO:BP	sperm ejaculation	GO:0042713	1	0	1	85
GO:BP	regulatory ncRNA-mediated post-transcriptional gene silencing	GO:0035194	1	0	89	85
GO:BP	ncRNA metabolic process	GO:0034660	1	0	277	85
GO:BP	embryonic cleavage	GO:0040016	1	0	1	85
GO:BP	positive regulation of cell cycle	GO:0045787	1	0	40	85
GO:BP	negative regulation of cell cycle	GO:0045786	1	0	53	85

intersection_size	effective_domain_size	intersections
3	10039	RRF-3,CPS-6,SMG-2
5	10039	CPR-5,UBA-1,CBG03805,CBG08424,PKC-2
1	10039	RRF-3
1	10039	CPS-6
3	10039	RRF-3,CPS-6,SMG-2
1	10039	RRF-3
17	10039	PGP-13,CBG01975b,MCA-3,UBA-1,PAIC-1,ABTM-1,KLP-12,SMG-2,RDE-12,PGP-1,PKC-2,TAT-5,PGP-4,PGP-3,MCA-1,LET-754,GSA-1
1	10039	HIS-24
9	9036	TOFU-6,SAPS-1,UBA-1,MUP-2,ITR-1,SMG-2,EMB-9,RAN-2,LET-2
2	9036	TOFU-6,PID-3
1	9036	ITR-1
1	9036	SMG-2
7	9036	RRF-3,TOFU-6,SMG-2,CBG13541,PID-3,CBG21813,CBG22378
1	9036	TOFU-6
2	9036	TOFU-6,PID-3
1	9036	ROD-1

Despite not showing statistical significance in g:GOst calculations, this list of proteins could be more relevant for subsequent analysis. Within this list, we were surprised to find two components associated with piRNA processing: TOFU-6 and PID-3 **(Figure 9D)**. These are core components of the PETISCO complex that is essential for processing piRNA precursors and histone-mRNA stability (Rodriguez et al., 2018, Zeng C et al., 2019, J. Pereirinha unpublished). In addition, our experiment also enriched for several CEY proteins (CBG12892, CEY-1, CEY-2.1). These proteins constitute a group of cold-shock domain proteins with an unknown function, recognized to interact with PETISCO components (Rodriguez et al., 2018, Zeng C et al., 2019). Another core PETISCO protein IFE-3 was detected in the immunoprecipitate however not enriched.

To validate the specificity of our findings, we conducted a parallel experiment using an anti-CelGTSF-1 antibody as a negative control **(Figure 8B)**. This antibody does not recognize CbrGTSF-1, and this is evident in the absence of CbrGTSF-1 peptides in the control IP. Under these conditions, we identified 64 enriched interactors of CbrGTSF-1. This lower number is likely due to the higher number of non-specific interactions exhibited by the control antibody anti-CelGTSF-1 (62 in contrast to 18 by anti-IgG). Nevertheless, RRF-3, ERI-5, PID-3, TOFU-6, and the CEY proteins were identified again as enriched for CbrGTSF-1.

6. GTSF-1 in *C. briggsae* continues to interact with PETISCO and ERI Complex in L4 stage

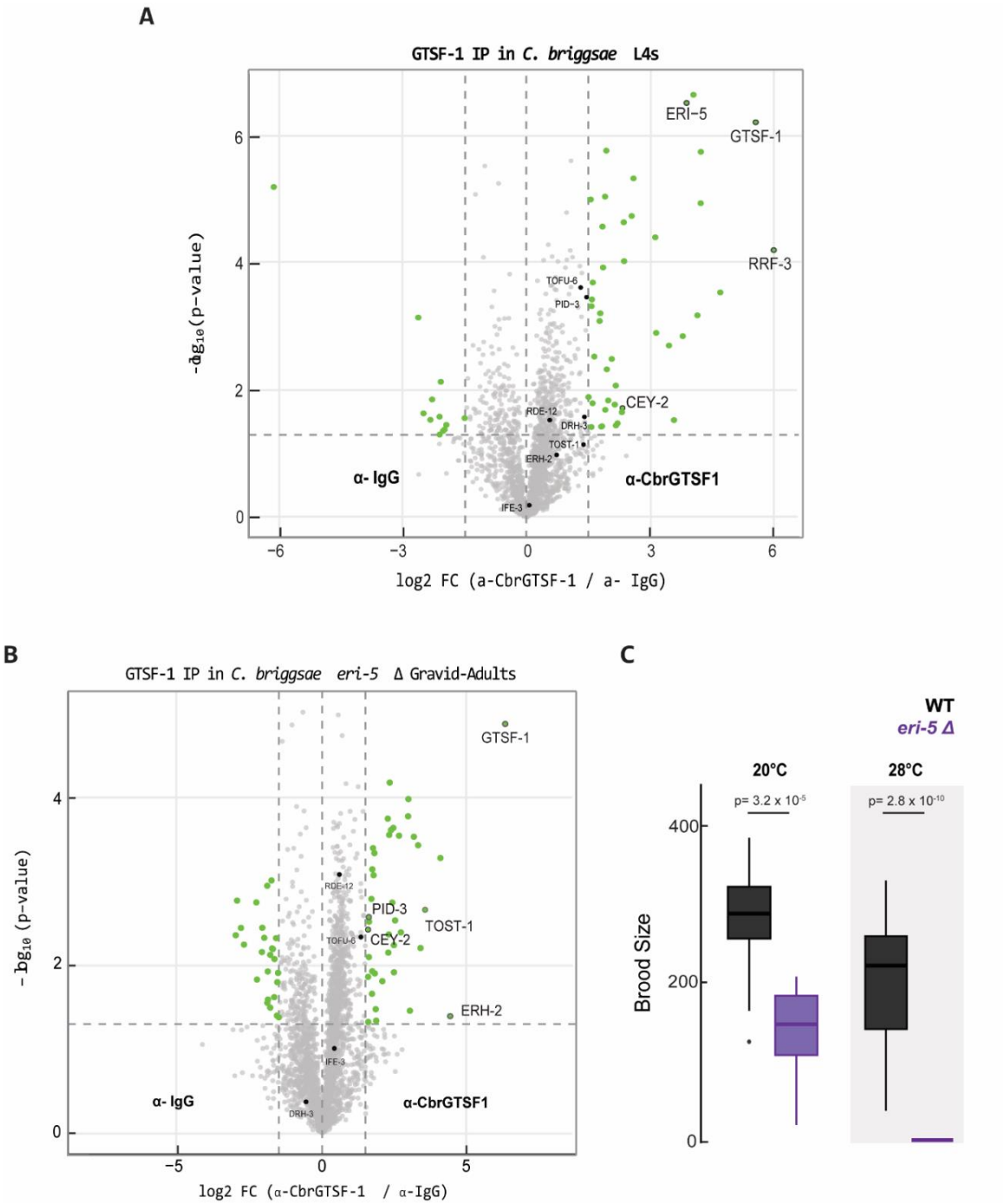
In *C. elegans* hermaphrodites, the L4 larval stage is characterized by germ cells initiating spermatogenesis, which shifts to oogenesis upon transitioning to adulthood. During this L4 stage, the 26G-RNA pathway is actively engaged, playing a vital role in regulating spermatogenesis and spermiogenesis (Conine et al., 2010, Conine et al., 2013). Our sRNA-seq data on *C. briggsae* L4 animals reveal active expression of 26G-RNAs at this stage, targeting genes similar to those in *C. elegans*. While the expression of TOFU-6 and PID-3 has been documented in the adult germline of *C. elegans* (Rodriguez et al., 2018; Zeng et al., 2019), their expression in L4 animals remains unexplored.

With confidence in the expression of GTSF-1 in *C. briggsae* L4 animals, we aimed to investigate its interaction with TOFU-6 and PID-3 at this developmental stage. Immunoprecipitation against GTSF-1 was conducted in L4 animals using the α-CbrGTSF-1 antibody, with α-IgG as a control **(Figure 10A)**. Under these conditions, IP-MS of GTSF-1 enriched for RRF-3 and ERI-5, mirroring observations in Gravid-Adults. RDE-12 was detected in the background but not enriched. Additionally, DRH-3, a Dicer-related helicase known to interact with RRF-3 in *C. elegans* and essential for 26G-RNAs biogenesis, was observed in the background (Vasale et al., 2010, Thivierge et al., 2012, Almeida et al., 2018). Notably, TOFU-6 and PID-3 were also enriched. In addition, we detected TOST-1, ERH-2, and IFE-3, comprising the entire PETISCO complex, in the background, indicating the presence of PETISCO in the *C. briggsae* L4 germline. Notably, PID-1, the second effector protein of PETISCO, was not found in our experiment **(Figure 10A)**.

In the Gravid-Adult animals, we observed TOST-1 to co-immunoprecipitate with CbrGTSF-1 when the amount of whole-worm lysate and antibody were increased (data not shown). In *C. elegans*, both TOST-1 and PID-1 are found in the Gravid-Adult germline but like other PETISCO proteins, their expression has not been tested in the L4 germline. Thus, we cannot exclude the possibility that PID-1 is not expressed in *C. briggsae* L4 animals. However, the absence of PID-1 from Gravid-Adult co-immunoprecipitates does suggest that the PETISCO co-immunoprecipitated with GTSF-1 is bound exclusively to TOST-1. In conclusion, during the L4 stage, CbrGTSF-1 interacts with both pre-ERIC and PETISCO, specifically TOST-1-bound PETISCO.

Previous studies have shown the existence of an RdRP module composed to RRF-3, DRH-3 and ERI-5. GTSF-1 is placed within the context of this module, as GTSF-1 pulldowns detect all three proteins with varying coverages. Other ERI complex proteins such as DCR-1, ERI-1 and ERI-3 are not detected in GTSF-1 IPs. In *C. briggsae*, GTSF-1 pulldowns detect RRF-3, DRH-3 as well as ERI-5. This suggest a similar RdRP module composition for RRF-3 in *C. briggsae*.

Figure 10 | IP-MS of GTSF-1 in *C. briggsae*: Volcano plots representing label-free proteomic quantification of quadruplicate Immunoprecipitates (IPs) from various *C. briggsae* extracts: **(A)** Wild-type L4 stage animals. **(B)** *eri-5* Δ Gravid-Adult animals. A total of 1.5mg worm lysate and 2ug of bead-bound antibody was used for the IP. The x-axis represents the median fold enrichment of individual proteins precipitated with one antibody over another. The y-axis indicates -Log10 (p-value) of observed enrichments. Dotted lines represent thresholds at p=0.05 and 1.5-fold enrichment. Data points highlighted in green represent statistically significant interactors. Relevant interactors falling below the applied cutoff are highlighted in black. **(C)** Brood size of *C. briggsae* WT and *eri-5* Δ animals at permissive temperature of 20˚C and elevated temperature of 28˚C. p-value calculated by Wilcoxon rank sum test.



7. The role of ERI-5 in GTSF-1 and RRF-3 interaction in *C. briggsae*

The tandem tudor domain protein ERI-5 is one of the core ERI complex proteins. ERI-5 is a direct binding partner of RRF-3 (Thivierge, Caroline et al., 2011) and is shown to mediate association between RRF-3 and the endoribonuclease DCR-1. To assess the role of ERI-5's significance in *C. briggsae*, we generated an *eri-5* deletion allele **(Figure 10C)**. We targeted the first and last exons of *eri-5* and isolated an allele with the whole gene deleted. Similar to GTSF-1, loss of ERI-5 resulted in animals with low brood size at 20˚C and complete sterility at 28˚C **(Figure 10C)**. Although we did not sequence small-RNAs from *eri-5* Δ animals, the ts sterility phenotype suggests a conserved role of ERI-5 in the *C. briggsae* 26G-RNA pathway.

In *C. elegans*, absence of ERI-5 partially destabilizes both GTSF-1 and RRF-3 in embryonic extracts. However, GTSF-1 continues to immunoprecipitate RRF-3 in an *eri-5* mutant background (Almeida et al., 2018). We immunoprecipitated GTSF-1 in *C. briggsae* Gravid-Adult *eri-5* Δ animals and analyzed the precipitates using label-free mass spectrometry **(Figure 10B).** In *C. briggsae*, GTSF-1 continues to be expressed upon deletion of ERI-5, as evident from the high GTSF-1 enrichment in our experiment. However, we did not detect any peptides of RRF-3. This could result either from complete destabilization of RRF-3 or loss of RRF-3 binding to GTSF-1.

Both scenarios are interesting as ERI-5 has a paralog named EKL-1. In *C. elegans*, a small fraction of EKL-1 co-immunoprecipitate with RRF-3, and in the absence of ERI-5, this interaction is sufficient to contribute to 26G-RNA biogenesis (Thivierge et al., 2011). It is likely that RRF-3 is not completely destabilized in *C. elegans* ERI-5 deletion due to the compensatory binding of RRF-3 by EKL-1. *C. briggsae* also encodes a copy of EKL-1. However, EKL-1 was not found in immunoprecipitation targeting GTSF-1 in ERI-5Δ background. This suggests that EKL-1 in *C. briggsae* is potentially unable to compensate for ERI-5 function.

The PETISCO proteins TOFU-6 and PID-3 continue to co-immunoprecipitate with GTSF-1 in the absence of ERI-5. We also notice a higher enrichment of TOST-1 and ERH-2 in this experiment, strengthening the hypothesis of TOST-1 bound PETISCO as the interactor of GTSF-1. These results also suggest that GTSF-1 interacts with PETISCO independent of the RRF-3 module.

8. PETISCO is a stable protein complex in *C. briggsae* adult animals and is essential for fertility

PETISCO is an oligomeric RNA-binding protein complex described in *C. elegans* **(Figure 11A)**. The consistent co-immunoprecipitation of PETISCO subunits with GTSF-1 prompted us to investigate whether GTSF-1 and PETISCO directly interact with each other. Additionally, identifying PETISCO encoding genes in *C. briggsae* sparked our interest in understanding if PETISCO proteins share similarities in structure and function across various organisms. We determined the conservation of PETISCO resorting to reciprocal BLASTp analysis of each subunit to the whole proteome of species representing the multiple clades of Nematoda **(Figure 11B)**. *C. briggsae* encodes one homolog of each PETISCO subunit. All subunits have a length and domain distribution similar to their homolog in *C. elegans* [data not shown].

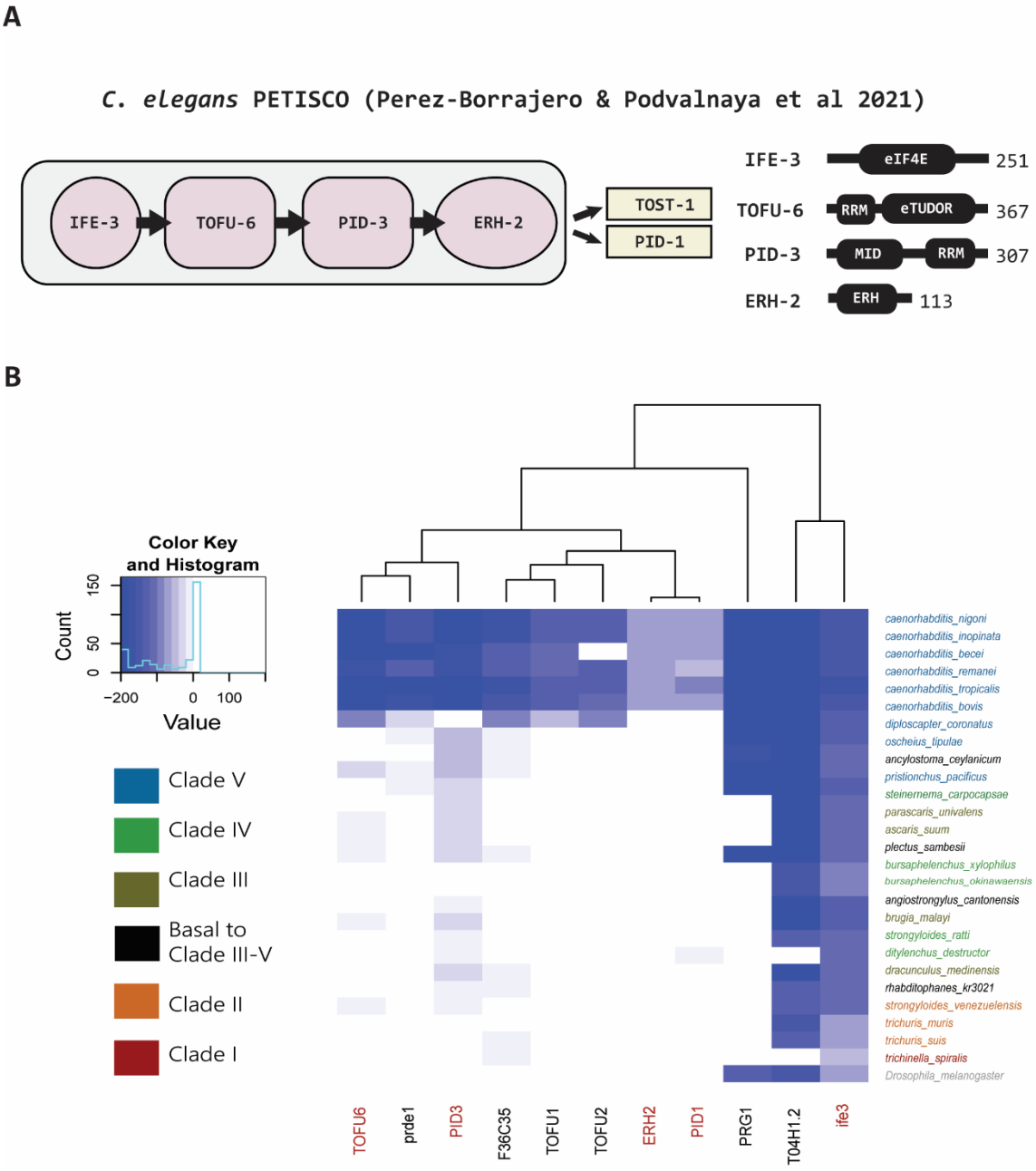


Figure 11 | PETISCO conservation in nematodes: (A) (Left) Illustration outlining the PETISCO topology, subunit interactions, and the binding of effector proteins TOST-1 and PID-1 (adapted from Perez-Borrajero & Podvalnaya et al., 2021). (Right) Configuration of domains in IFE-3, TOFU-6, PID-3, and ERH-2 from *C. elegans*. Predicted domains are represented by rounded rectangles. These effectors delineate distinct functions of the complex: TOST-1 oversees piRNA processing (Rodrigues et al., 2019), while PID-1 connects PETISCO to histone mRNA stabilization (J. Pereirinha, unpublished). **(B)** Conservation of selected *C. elegans* small RNA pathway proteins generate by Prof. Peter Sarkies (Department of Biochemistry, Oxford GB). Conservation is calculated as the logarithm of the score of the best blast hit in bits normalized to the length of the protein. Species are colored bases on their classification into clades I-V of phylum nematoda. *D. melanogaster* serves as an outgroup.

Yeast-Two-Hybrid describes the architecture of *C. briggsae* PETISCO

To test interactions of PETISCO subunits and ERIC in-vitro, we used the Yeast-two-Hybrid (Y2H) assay **(Figure 12A).** We designed a large interaction matrix, testing all subunits of PETISCO and GTSF-1 against each other **(Figure 12B).** We also included the CEY-2 homolog (CBG14892) and ERI-5, as both were highly enriched in the GTSF-1 IPs. We confirmed positive interactions through growth in selective media lacking either Adenine or Histidine.

In less stringent conditions (-TRP, -LEU, -HIS), we observed PID-3 interacting with ERH-2 and TOFU-6, and TOFU-6 interacting with IFE-3. Additionally, ERH-2 showed interactions with TOST-1 and PID-1. Under high stringency selection (-TRP, -LEU, -HIS, -ADE), we recapitulated the interaction of PID-3 with TOFU-6, while noting a mild interaction between ERH-2 and PID-1 **(Figure 12B).** These findings suggest an interaction map for PETISCO, following the sequence IFE-3 to TOFU-6 to PID-3 to ERH-2 to PID-1 or TOST-1, aligning with observations in *C. elegans* (Rodriguez et al., 2019). We could not find an interaction of CEY-2 with any subunit of PETISCO.

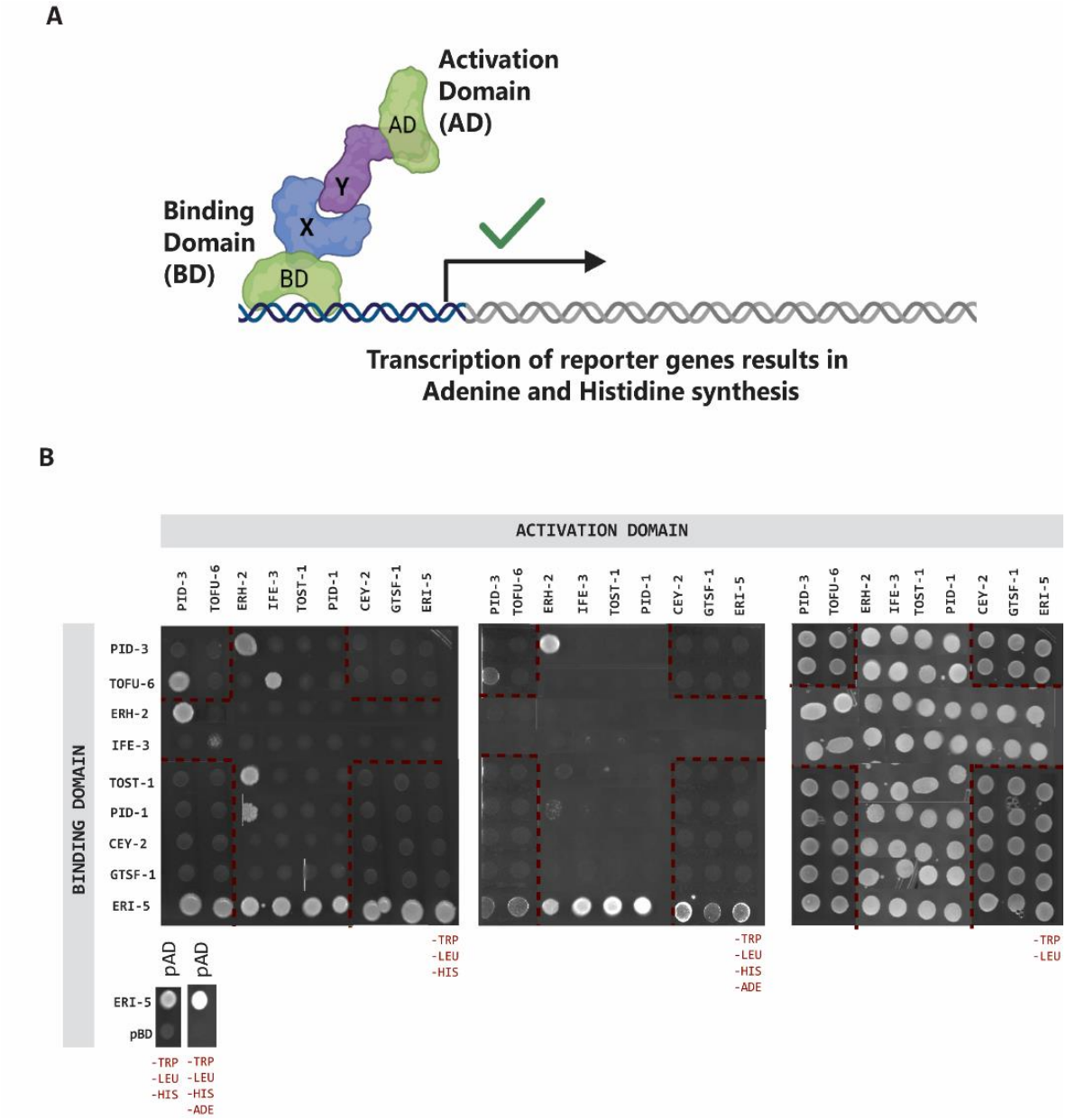


Figure 12 | PETISCO topology is conserved in *C. briggsae*: **(A)** Schematic of the Yeast-Two-Hybrid system: We transfect auxotrophic strains of *Saccharomyces cerevisiae* with two plasmids expressing a pair of proteins and score for protein-protein interaction based on the growth of *S. cervisiae.* The introduced proteins are genetically fused to either the Gal4 transcription factor's binding domain (BD) or activation domain (AD). Favorable interactions trigger expression of genes which compensate for auxotrophy of Histidine (HIS) and Adenine (ADE). The plasmids for BD and AD provide Leucine (LEU) and Tryptophan (TRP), respectively, and are used to select for successful double transformants. **(B)** Yeast-Two-Hybrid assay on *C. briggsae* proteins under control (-LEU, -TRP), low stringency (-LEU, -TRP, -HIS) or high stringency (-LEU, -TRP, -HIS -ADE) conditions. Images are stitched together from two seperate experiments seperated by dotted red lines. Lower panel tests Gal4 activation domain (pAD) against either Gal4 binding domain or ERI-5 fused to Gal4 binding domain (pBD). *C. briggsae* ERI-5 shows false positive interactions.

We compared our results to the Y2H assay conducted for *C. elegans* PETISCO subunits (Rodriguez et al., in 2019) as it was performed in our group under the same conditions with similar plasmids.

We note that interactions in *C. briggsae* are generally less robust compared to those in *C. elegans*. For instance, in low stringency conditions, the majority of protein pairs in *C. briggsae* interact only unidirectionally, in contrast with *C. elegans* (Rodriguez et al., 2019) where most interactions are bidirectional. Furthermore, under high stringency conditions, all combinations of *C. elegans* proteins interact at least in one direction. However, under the same conditions, most interactions are absent in *C. briggsae*. The only exception is binding of PID-3 to ERH- 2 in *C. briggsae*, which displays bidirectional interaction under low stringency as well as strong unidirectional interaction under high stringency. Finally, we also observe that *C. briggsae* ERH-2 does not show dimerization. Interestingly, we did not observe any interaction of GTSF-1 with either CEY-2 or PETISCO subunits. We note the same for ERI-5 fused to activation domain. ERI-5 fused to binding domain shows false-positive interactions.

IP-MS on CbrPID-3 detects all known PETISCO subunits

We proceeded to check if these proteins interact with each other *in-vivo*. Our group is extensively studying PETISCO in *C. elegans*, and thus we had access to robust affinity-purified antibody against various *C. elegans* PETISCO subunits. Through an IP-Western, we confirmed that α-CelPID-3 recognizes a protein around the expected size of CbrPID-3 of ~40kDa **(Figure 13A).** We used this antibody to immunoprecipitate PID-3 in Gravid-Adult *C. briggsae* animals and analyzed the precipitates using label-free mass spectrometry **(Figure 13B).** In accordance to the western blot, PID-3 was highly enriched in the immunoprecipitates. Further, all constituents of PETISCO, namely IFE-3, TOFU-6, ERH-2, PID-1 and TOST-1 were also highly enriched in the IP. These results strongly suggest that *C. briggsae* animals also form a PETISCO complex.

RNAi depletion of CbrPID-3 and CbrTOFU-6 recapitulates embryonic defects

In *C. elegans*, PETISCO is required for 21U RNA biogenesis and embryonic development (Rodrigues, et al.,2018, Zeng et al.,2019). Mutants of PETISCO show a maternal-effect-lethal (Mel) phenotype, where embryos are laid but they do not hatch. To test the function of PETISCO in *C. briggsae*, we attempted to generate a deletion allele of TOFU-6 using CRISPR-Cas9. We isolated

several potential founders; however, we could not maintain the strain, as eggs laid by homozygous animals never hatched **(data not shown).** This embryonic defect is consistent with the Mel phenotype of PETISCO mutants in *C. elegans*. We then opted to use RNAi and deplete *tofu-6* and *pid-3* RNA levels in *C. briggsae* **(Figure 13C).** Wild-type *C. briggsae* AF16 does not express SID-2, the intestinal luminal transmembrane protein required for environmental RNAi in *C. elegans* (Winston, et al.,2007). The JU1018 strain of *C. briggsae* contains a transgene expressing *Cel-sid-2* and is RNAi competent. Hence, we cultured JU1018 animals on agar plates seeded with *E. coli* expressing dsRNA encoding ~150bp of *Cbr-tofu-6* and *Cbr-pid-3*. Indeed, as we expected, RNAi against both genes led to a strong Mel phenotype, with treated animals not laying any viable progeny **(Figure 13C)**

Small-RNA sequencing on *pid-3* and *tofu-6* depleted animals show loss of 21U-RNAs

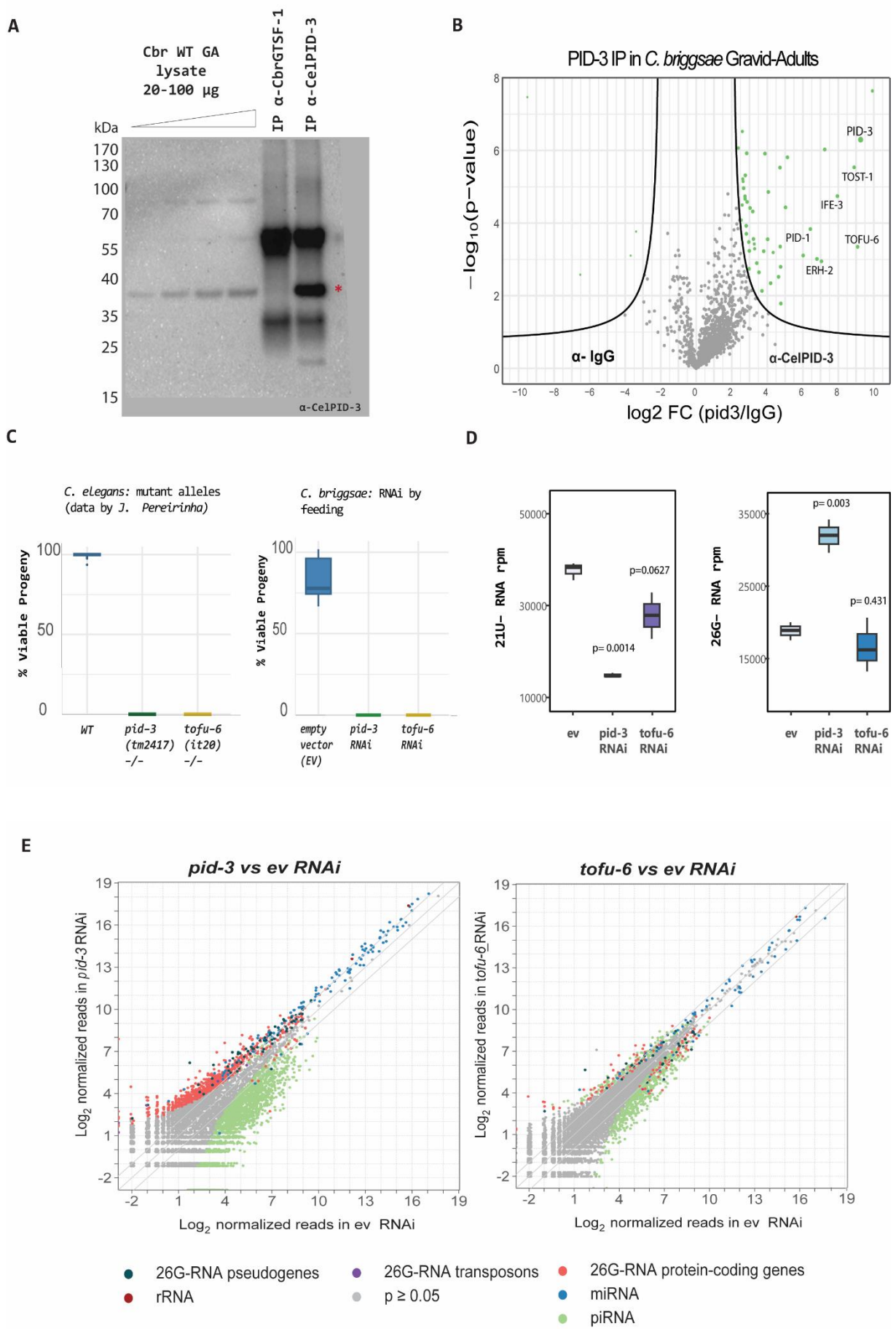
Next, we collected RNAi treated animals at gravid-adult stage for small-RNA sequencing. All experiments were conducted in triplicates. Samples were analyzed using the tinyRNA pipeline (Tate et al., 2022) using a custom annotations file for the *C. briggsae* genome. The reads for each of the small-RNA classes were normalized to library depth using DESeq2. Both RNAi treatments led to a reduction of global 21U-RNA levels, although the effect was much milder in *tofu-6* RNAi **(Figure 13D).** Strikingly, we also observed a moderate increase in total 26G-RNA levels in *pid-3* RNAi samples **(Figure 13D).** A scatterplot of normalized reads of individual genes in *pid-3* RNAi samples against empty-vector (ev) treated animals, show a significant increase in amount of 26G-RNAs per gene **(Figure 13E).** This increase is seen in general for all target genes, not specific towards a few set of targets. On the other hand, global miRNA levels were largely unaffected. Individual miRNAs did show variation. However, expression of miRNAs are highly sensitive to developmental stage of the animal. Thus, one explanation for variation of miRNA levels in our samples is the mild un-synchronization of all animals grown for RNAi treatment **(Figure 13E).**

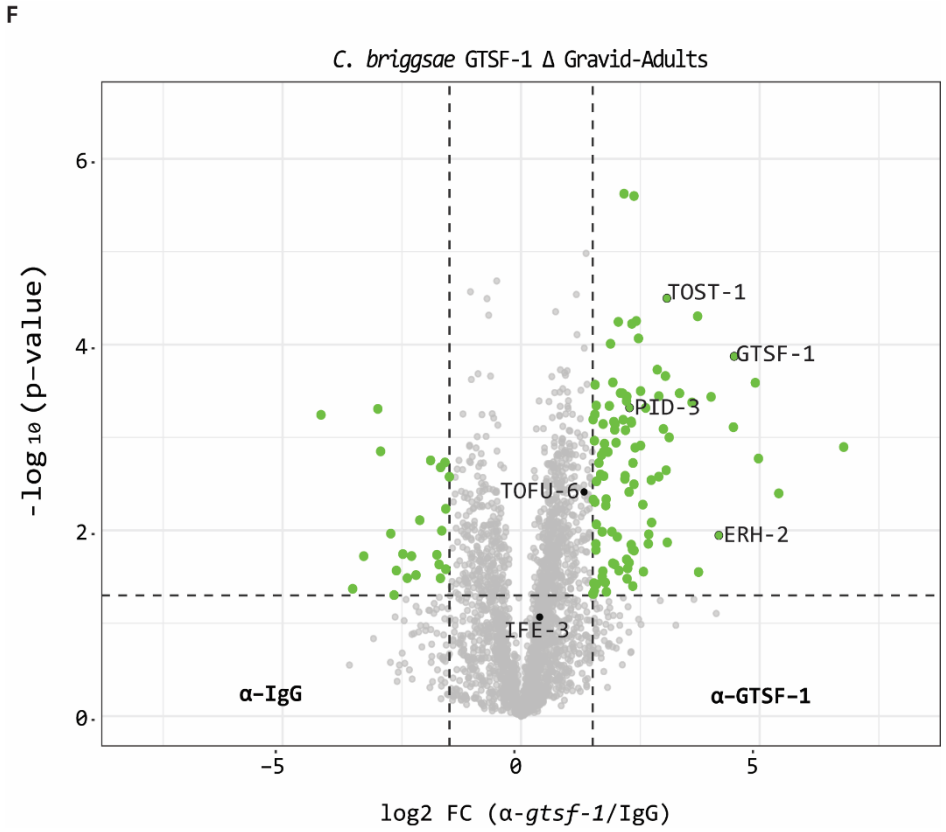
Figure 13 | PETISCO is a stable complex in *C. briggsae* with conserved functions:

(A) Western-Blot to detect PID-3 on *C. briggsae* wild-type lysates: Lanes 1-4 are loaded with increasing concentrations of synchronized GA extracts. Lanes 5 and 6 showcased lysate fractions immunoprecipitated with anti-CbrGTSF-1 or anti-CelPID-3 antibodies, using 1.5mg total protein and 2µg antibody. A 5% fraction of the IP was loaded. The red asterisk indicates the 35kDa CbrPID-3 band. GTSF-1 IP did not detect PID-3 on the western blot. **(B)** Volcano plots representing IPs from *C. briggsae* wild-type GA extracts using anti-CelPID-3 with anti-IgG as control. Green points depict interactors which cross the log2FC cutoff of 2 and p-value lower than 0.05.

(C) Percentage of Viable progeny scored in absence of *pid-3* and *tofu-6*: Left panel contains data from *C. elegans* mutant alleles by J. Perereirinha; Right panel depicts results of RNAi in a *C. briggsae* RNAi competent strain JU1018. **(D)** Global levels of 21U-RNAs and 26G-RNAs in empty-vector (ev) and *pid-3* / *tofu-6* RNAi worms, in RPM (reads per million). p values computed using Welch Two Sample t-test.

(E) TinyRNA-produced scatterplot comparing *C. briggsae* treated with either *pid-3* or *tofu-6* RNAi.x-axis and y-axis representing Log2 Mean normalized 26G-RNA counts per gene from empty-vector RNAi and *pid-3* or *tofu-6* RNAi samples, respectively. For more information on labelled gene-classes, go to Figure 4. **(F)** Volcano plots representing IPs from *C. briggsae gtsf-1 Δ* GA extracts using anti-CbrGTSF-1 with anti-IgG as control. Green points depict interactors which cross the log2FC cutoff of 2 and p-value lower than 0.05.


A
Cbr WT GA lysate 20-100 µg
IP α-CbrGTSF-1
IP α-CelPID-3
kDa
170
130
100
70
55
40
35
25
15
α-CelPID-3
B
PID-3 IP in C. briggsae Gravid-Adults
−log10(p−value)
log2 FC (pid3/IgG)
PID-3
TOST-1
IFE-3
PID-1
TOFU-6
ERH-2
α- IgG
α-CelPID-3
C
C. elegans: mutant alleles (data by J. Pereirinha)
C. briggsae: RNAi by feeding
% Viable Progeny
WT
pid-3 (tm2417) -/-
tofu-6 (it20) -/-
empty vector (EV)
pid-3 RNAi
tofu-6 RNAi
D
21U- RNA rpm
p=0.0627
p= 0.0014
26G- RNA rpm
p= 0.003
p= 0.431
ev
pid-3 RNAi
tofu-6 RNAi
E
pid-3 vs ev RNAi
tofu-6 vs ev RNAi
Log2 normalized reads in pid-3 RNAi
Log2 normalized reads in tofu-6 RNAi
Log2 normalized reads in ev RNAi
26G-RNA pseudogenes
rRNA
26G-RNA transposons
p ≥ 0.05
26G-RNA protein-coding genes
miRNA
piRNA



IP-MS on CbrGTSF-1 in GTSF-1 Δ background

The previous IP for GTSF-1 in *C. briggsae* animals were conducted with a non-specific antibody against IgG as control. This experiment ensures that the same proteome is being targeted by two different antibodies. However, it does not account for non-specific interactions of the α-CbrGTSF-1 antibody itself. To test this, we used this antibody for an IP-MS for GTSF-1 in the GTSF-1 Δ background (**Figure 13F**). Surprisingly, we detected GTSF-1 in the *gtsf-1 Δ* strain. We verified that the correct strain was used for this experiment by genotyping our cultures. In retrospect, the α-CbrGTSF-1 antibody was affinity purified from rabbit serum using CbrGTSF-1 as a ligand immobilized on a column. It is possible that α-CbrGTSF-1 is bound to low levels of CbrGTSF-1, which would result in peptides of GTSF-1 showing up in *gtsf-1 Δ* strain.

In this experiment, PETISCO subunits were enriched (**Figure 13F**). Thus, either the CbrGTSF-1 bound to α-CbrGTSF-1 is sufficient to interact with PETISCO, or the α-CbrGTSF-1 performs non-specific interactions to PETISCO. However, we did not detect any members of ERIC, so the latter explanation is highly likely. To get conclusive results, we would repeat this experiment with an epitope tagged GTSF-1.

Conclusions

Thus, the architecture and functions of PETISCO is conserved in *C. briggsae*. It is encouraging to see the uptick in total 26G-RNA levels in *pid-3* RNAi. However, this could also be a result of the depletion of piRNAs from the sequencing pool. To test this, we will be analyzing whether the 22G-RNAs downstream of 26G-RNAs show a similar increase in the *pid-3* and *tofu-6* RNAi samples. While these experiments fail to find a definitive link between PETISCO and the 26G-RNA pathway, further experiments with more controls will help define this link better.

9. CbrGTSF-1 can functionally replace GTSF-1 in *C. elegans*

Our results suggest a conserved function of *C. briggsae* GTSF-1 in 26G-RNA biogenesis with potential novel interactions within the PETISCO complex. To explore these findings further, we employed a cross-species complementation strategy. Specifically, we replaced the endogenous *gtsf-1* gene of *C. elegans* with that of *C. briggsae*. We generated a strain expressing *Cbr-gtsf-1::gfp::3x-flag* (*Cel-gtsf-1* (*xf264*)) in a *C. elegans gtsf-1* Δ background. The expression of CbrGTSF-1::GFP::3xFLAG was confirmed through Western blot analysis using anti-FLAG antibodies **(Figure 14A).**

The introduced transgene could partially rescue the temperature-sensitive sterility phenotype observed in *gtsf-1* Δ **(Figure 14B).** Notably, *Cel-gtsf-1* mutants exhibit heightened sensitivity to RNAi targeting somatic genes, such as *dpy-13* (Almeida et al., 2018). The transgene also rescued this phenotype, restoring RNAi sensitivity to practically wild-type levels **(Figure 14C).** Subsequently, we examined the subcellular localization of CbrGTSF-1::GFP::3xFLAG in gravid adult animals expressing the P-granule marker RFP::PGL-1 in the background. Similar to CelGTSF-1, we observed that CbrGTSF-1::GFP::3xFLAG diffused within the germline cytoplasm, without co-localization in the perinuclear P-granules **(Figure 14D).**

To verify that this rescue was indeed due to restoration in levels of 26G-RNAs, we performed small-RNA sequencing of *Cel-gtsf-1* (*xf264*) animals with wild-type (WT) and *gtsf-1* Δ samples as controls. As expected, *gtsf-1* Δ samples had a complete loss of 26G-RNAs. Expression of CbrGTSF-1::GFP::FLAG in *gtsf-1* (*xf264*) led to a small but significant restoration of total 26G-RNA levels **(Figure 15A).** This rescue was observed in both life-stages.

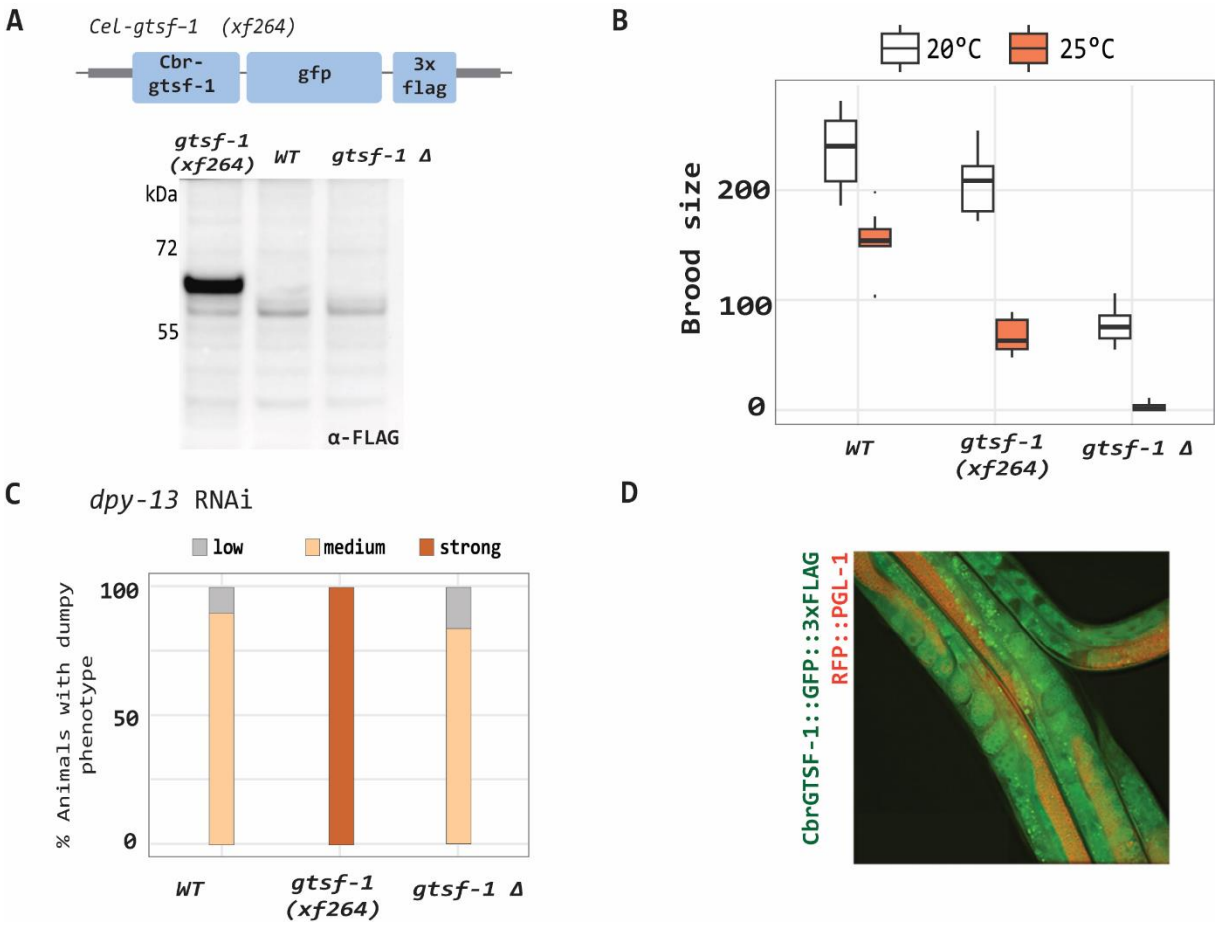
In GA samples, we identified 923 26G-RNA target genes. While most genes had partial restoration of 26G-RNA levels in *gtsf-1* (*xf264*) animals, 80 gene did not exhibit restoration **(Figure 15B).** Interestingly, a subset of genes (149) displayed more than double the 26G-RNA counts in *Cel-gtsf-1* (*xf264*) animals compared to the WT **(Figure 15B).** Within L4 samples, a total of 4026 26G-RNA target genes were identified **(Figure 15B).** Similar to GA samples, the majority of genes experienced partial restoration, with the exception of 867 genes that exhibited no rescue. In contrast to the GA samples, there was no instance where a gene was targeted more in *Cel-gtsf-1* (*xf264*) animals than in WT samples in the L4 stage **(Figure 15B).** This could suggest that 26G-RNA pathway in L4 stage is more stringent compared to GA stage. Notably, we did not identify any novel 26G-RNA target loci. Thus, the introduction of CbrGTSF-1::GFP::FLAG does not bias the *C. elegans* 26G-RNA pathway towards targeting loci that are originally targeted in *C. briggsae*.

Altogether, these results suggest that CbrGTSF-1::GFP::FLAG is capable of capturing the interacting partners of CelGTSF-1 and inducing the transcription of 26G-RNAs, albeit not to wild-type levels. Our western blot indicates high transgene expression, suggesting that the partial restoration of 26G-RNA levels may be attributed to the limited binding of CbrGTSF-1 to the *C. elegans* ERI complex. To investigate this, we conducted a pull-down on gravid-adult (GA) animals using antibodies against CbrGTSF-1 and analyzed the interacting partners through mass spectrometry **(Figure 16A).** The pull-down clearly enriched for the bait protein CbrGTSF-1::GFP::FLAG, but notably, no peptides against RRF-3 and ERI-5, the two prominent interactors of CelGTSF-1, were identified. We repeated this experiment under low-salt conditions to enhance potential weak interactions, yet once again, no proteins other than our bait were enriched **(Figure 16B).** We wondered if introducing the GFP::FLAG tag to C-terminus of GTSF-1 might have influenced the binding to RRF-3 and/or ERIC. Almeida et. al 2018 did similar experiments with CelGTSF-1::mCHERRY::3XFLAG expressing transgene, which was also an overexpression condition. IP of GTSF-1 in this transgenic strain could identify not just RRF-3 and ERI-5 but also other ERIC proteins (Almeida et al., 2018). In our experiment, the complete lack of any interactors of CbrGTSF-1::GFP::FLAG is thus, an interesting observation that highlights the biochemical differences in CelGTSF-1 and CbrGTSF-1.

Figure 14 | Expression of *C. briggsae* GTSF-1 in *C. elegans*: (A) Top: Schematic representation of *Cel-gtsf-1* (*xf264*) where endogenous GTSF-1 is replaced by a construct expressing GFP::3xFLAG fused CbrGTSF-1. Bottom: Western-Blot of Gravid-Adult animals probed with α-FLAG or α-CbrGTSF-1. **(B)** Brood-size quantification of wild-type and *gtsf-1* mutant animals, cultured at permissive temperature of 20℃ or elevated temperature of 25℃. p-values calculated by Wilcoxon rank sum test.

(C) Wild-type and *gtsf-1* mutant animals treated with RNAi against dpy-13. Animals were scored for strength and penetrance of Dumpy (Dpy) phenotype. The Dpy phenotype is characterized by animals harboring a thicker middle body and a shorter length. Severity levels were categorized as "low," "medium," or "strong," determined by visually observed ratios of animal length to body thickness. The most pronounced phenotype ("strong") occurred when the animal length resembled that of a wild-type L2 larva, but the body exhibited extreme thickness, rendering the animal immobile.

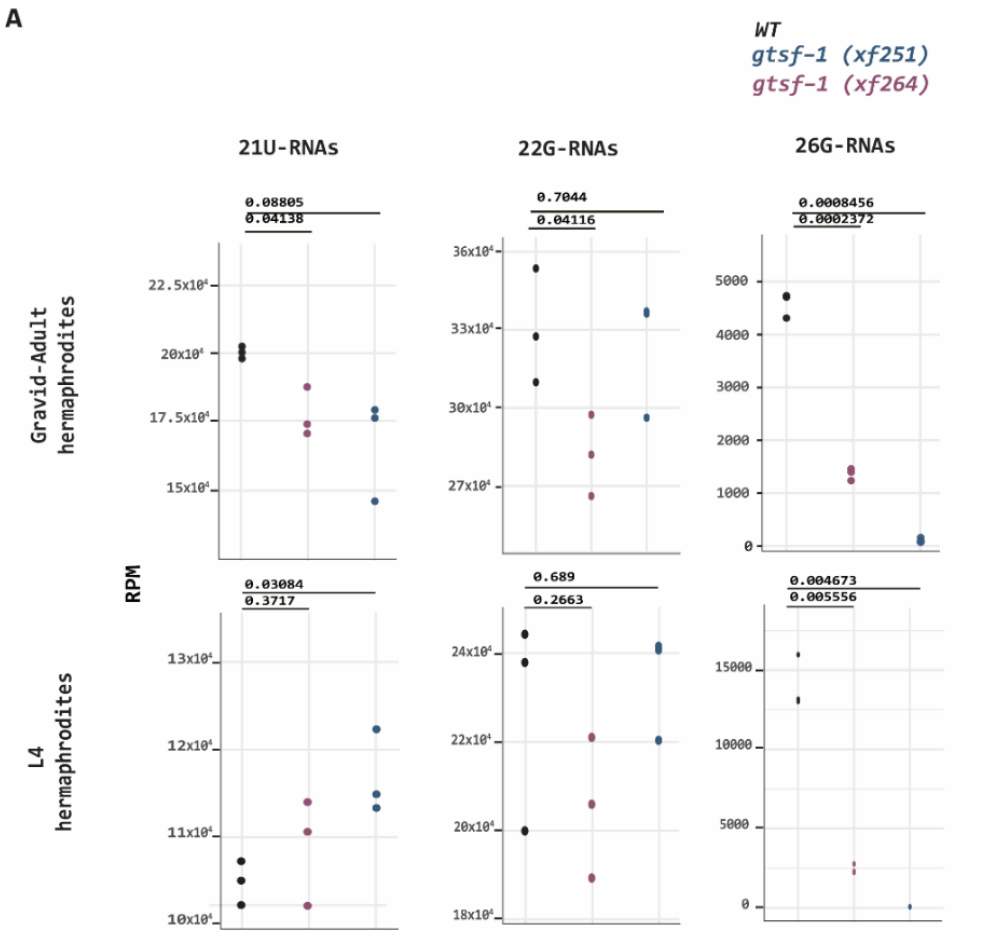
(D) Representative confocal fluorescence microscopy images of a Gravid-Adult animals showing the presence of GTSF-1 and PGL-1 tagged proteins in gonad.



In summary, we demonstrate that CbrGTSF-1::GFP::FLAG forms a functional ERI complex in *C. elegans* and partially restores 26G-RNA transcription. However, the association is notably weaker than with the endogenous CelGTSF-1, making it challenging to capture biochemically. The lack of novel loci for 26G-RNA targeting implies that CbrGTSF-1 does not specify the targets. Furthermore, we did not observe any associations of CbrGTSF-1 with *C. elegans* PETISCO nor note any differences to piRNA levels.

Figure 15 | Expression of CbrGTSF-1 in *C. elegans* can rescue 26G-RNA levels: (A) The global levels of 21U-RNAs or 26G RNAs were assessed in Gravid adult and L4 stage animals. The values were normalized to total mapped reads and are presented in reads per million (RPM). The p-value was calculated using the Welch t-test.

(B) A scatter plot was generated to display the normalized 26G-RNA count per gene, averaged across three replicates, on a log10 scale. The x-axis represents counts in the WT sample, while the y-axis illustrates the difference in counts between *gtsf-1* (*xf264*) [Rescue] samples and *gtsf-1* (*xf251*) [Deletion] samples. The black line signifies x=y. Only genes with WT counts >1 and WT/Deletion count >2 are shown. The color of each dot indicates the gene bio-type. Orange and yellow points representing genes that exhibit more than double counts in Rescue samples compared to WT. Black points are genes whose 26G-RNA levels are not restored in Rescue samples



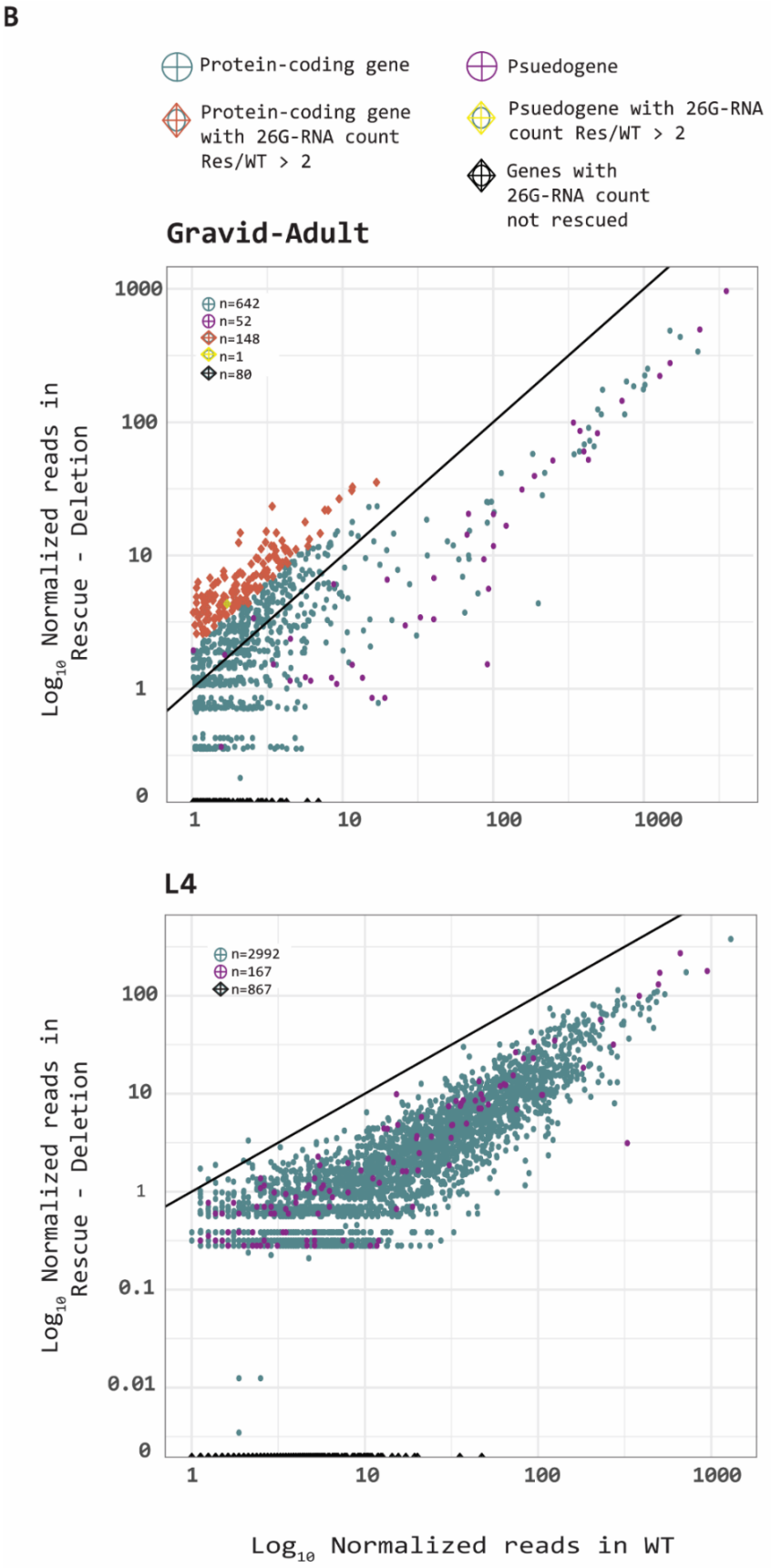
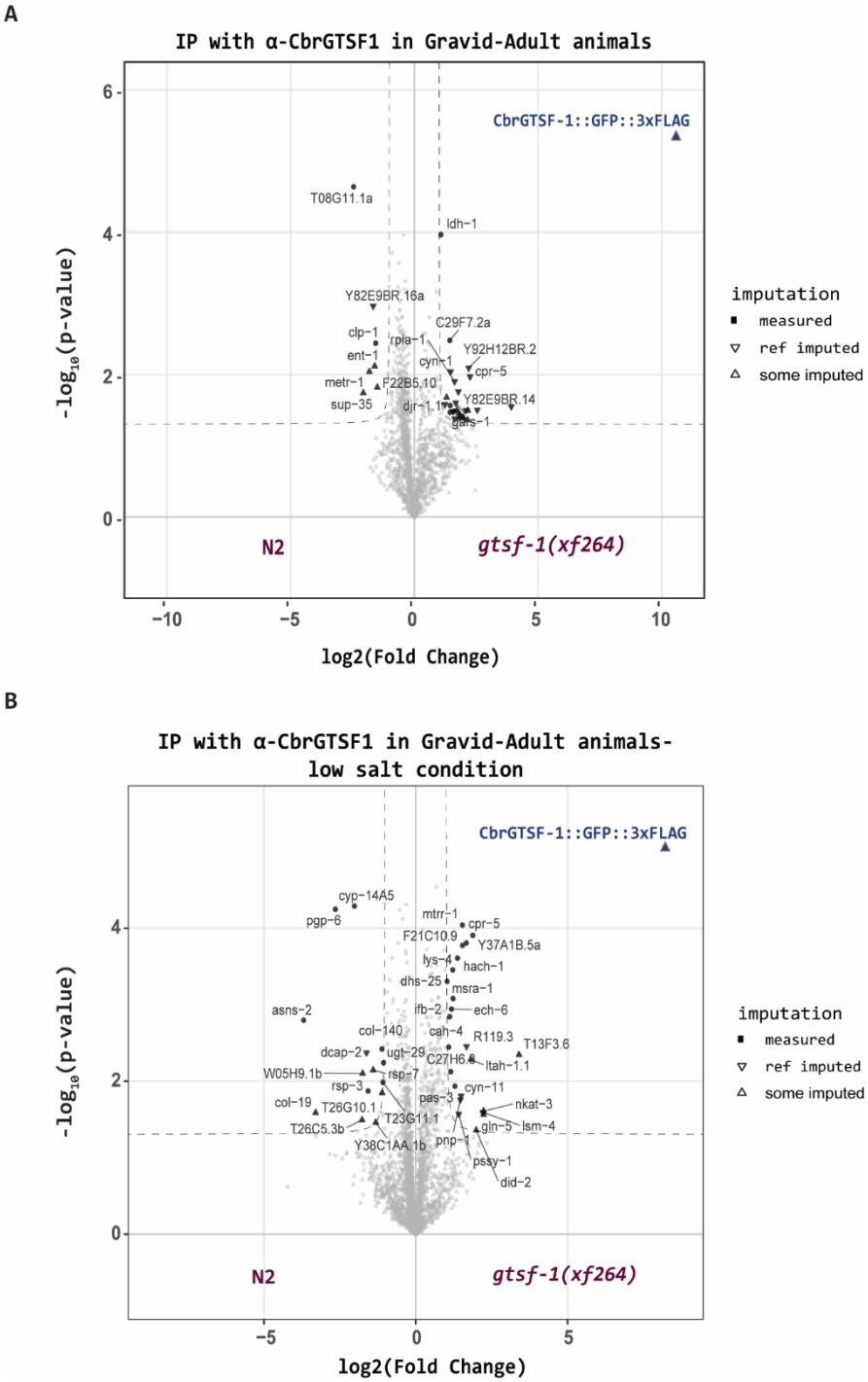

B
Protein-coding gene
Psuedogene
Protein-coding gene with 26G-RNA count Res/WT > 2
Psuedogene with 26G-RNA count Res/WT > 2
Genes with 26G-RNA count not rescued
Gravid-Adult
n=642
n=52
n=148
n=1
n=80
Log10 Normalized reads in Rescue - Deletion
1000
100
10
1
0
1
10
100
1000
L4
n=2992
n=167
n=867
Log10 Normalized reads in Rescue - Deletion
100
10
1
0.1
0.01
0
1
10
100
1000
Log10 Normalized reads in WT

Figure 16| Immunoprecipitation of *C. briggsae* GTSF-1 expressed in *C. elegans*: Volcano plots representing label-free proteomic quantification of quadruplicate Immunoprecipitates (IPs) from *C. elegans* Wild-type (WT) Gravid-Adult extracts using α-CbrGTSF-1 antibody. **(A)** 1.5mg of extract was incubated with 2ug of bead-bound antibody and subjected to three washes with a Tris buffer containing 150mM NaCl. **(B)** 3mg of extract was incubated with 4ug of bead-bound antibody and subjected to two washes with a Tris buffer containing 75mM NaCl. After final wash, beads were boiled in SDS and analysed on a mass-spectrometer. The x axis represents the log2 fold enrichment of individual proteins in one strain over another. The y-axis indicates -Log10 (p-value) of observed enrichments. Dotted lines represents thresholds at p=0.05 and 1.5-fold enrichment. The bait is highlighted in blue while all significant interactors are in black



10. A curious mutant of *C. briggsae gtsf-1* shows temperature-dependent sterility without loss of 26G-RNAs

In a previous effort aimed at generating mutants for *Cbr-gtsf-1*, we successfully identified a mutant allele characterized by a precise, in-frame deletion of 6 base pairs. This deletion resulted in the removal of two non-conserved amino acids situated between the highly conserved CPY and H residues within the first CHHC zinc finger of *gtsf-1*. Remarkably, these mutants exhibited normal viability and displayed no discernible morphological abnormalities. However, when subjected to an elevated temperature of 30˚C, the mutants exhibited sterility. Despite this temperature-induced sterility, small-RNA sequencing of the mutant embryos did not reveal any alterations in the profile of small RNAs (PhD Thesis, Miguel Almeida).

This intriguing outcome prompted further investigation. To mitigate potential off-target effects associated with CRISPR-Cas9, we recreated the same allele using an alternative guide RNA, designating it *Cbr-gtsf-1* (*xf346*) **(Figure 17A).** Following two rounds of outcrossing, we evaluated the brood size of *Cbr-gtsf-1* (*xf346*) animals cultivated at both 20˚C and 28˚C. The positive control consisted of *Cbr-gtsf-1* deletion animals, known to exhibit fertility defects at 20˚C and complete sterility at 28˚C. Notably, *Cbr-gtsf-1* (*xf346*) animals displayed negligible changes in brood size at 20˚C. However, at 28˚C, these animals mirrored the sterility observed in the complete gene deletion **(Figure 17A).** The deletion of just two amino acids surprisingly recapitulated the temperature-sensitive sterility observed in the absence of the entire gene.

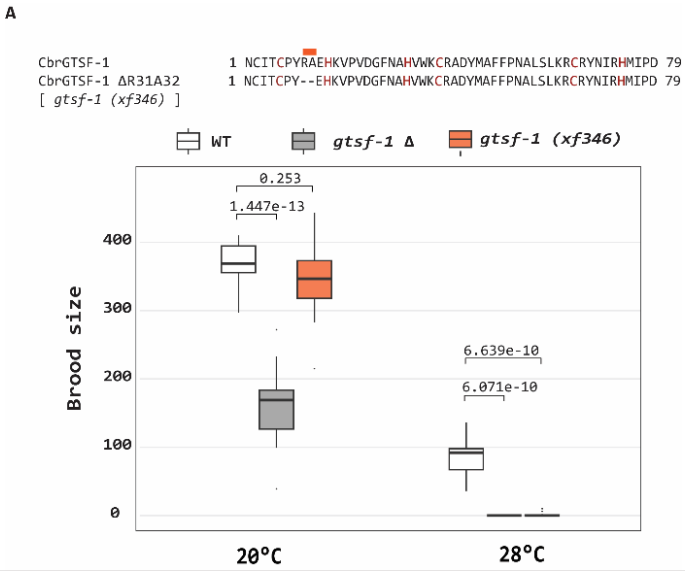
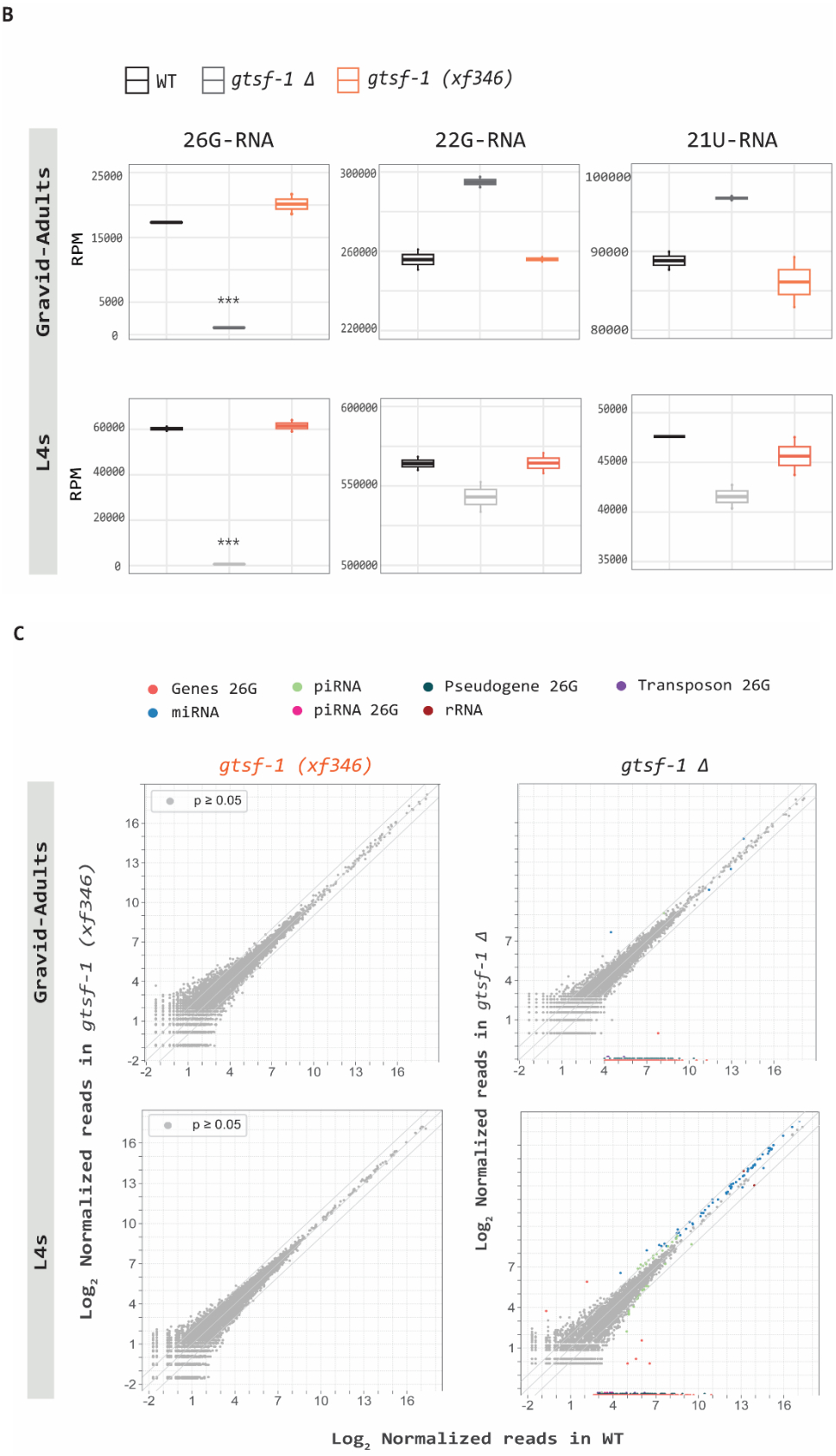


Figure 17 | Phenotypes and sRNA sequencing of a *C. briggsae* GTSF-1 mutant: (A) TOP: Alignment of GTSF-1 sequence indicating the in-frame deletion of R31 and A32 in *Cbr-gtsf-1 (xf346).* BOTTOM: Brood-size quantification of wild-type and *gtsf-1* mutant animals, cultured at permissive temperature of 20˚C or elevated temperature of 28˚C. Indicated *gtsf-1 Δ* is *Cbr-gtsf-1 (xf345).* p-values calculated by Welch two sample t-test. **(B)** Reads-per-million (RPM) of 21U-RNAs and 26G-RNA from *C. elegans* or *C. briggsae* Gravid-Adult and L4 animals. p-value calculated using Welch-t-test. Only p-values < 0.05 is depicted as ***. **(C)** TinyRNA generated scatterplot of log2 DEseq2 normalized reads in WT samples against either *gtsf-1 Δ* or *gtsf-1 (xf346).* Genes with deferentially expressed small-RNAs are identified using DESeq2. Genes with a P-value cutoff of 0.05 are highlighted in color according to their class. The class "Genes 26G: indicate 26G-RNAs mapping antisense to protein-coding genes. Similarly, the classes "miRNA 26G", "pseudogene 26G", and "transposon 26G" indicate 26G-RNAs mapping antisense to loci annotated with miRNA, pseudogene or transposon, respectively. (Continued on next page)



In the initial experiment, small-RNA sequencing was limited to embryos of *xf207* animals. To assess small-RNA populations comprehensively across various developmental stages, we sequenced Gravid-Adult and L4 animals from *Cbr-gtsf-1* (*xf346*). Unexpectedly, no discernible changes in small-RNA populations were observed in *xf346* animals in either sample **(Figure 17B).**

Additionally, no individual gene exhibited significant alterations in small-RNA levels **(Figure 17C).** Given that the temperature-sensitive sterility phenotype of *Cbr-gtsf-1* (*xf346*) manifests specifically at 28˚C, it is plausible that changes in small-RNA levels would only be evident in animals cultured at this elevated temperature. These findings imply that residues R31 and A32 of CbrGTSF-1 play a crucial role in fertility at elevated temperatures. However, conclusive data regarding the involvement of small RNAs in this phenotype is currently lacking.

11. Conservation of 26G-RNAs target genes between *C. elegans* and *C. briggsae*

As discussed in section 4, we sequenced small-RNAs of GTSF-1 deletion strains from *C. elegans* and *C. briggsae*, including both Gravid-Adult and L4 stage animals. We analyzed the sequencing files using the TinyRNA pipeline (Tate et al., 2023) and modified the genome annotation file of *C. briggsae* to include loci encoding piRNAs and miRNAs (Beltran et al., 2019). GTSF-1 is essential for 26G-RNA biogenesis in nematodes. Unlike 26G-RNAs, global 21U-RNA levels only marginally change upon GTSF-1 deletion in both *C. elegans* and *C. briggsae* **(Figure 18A).** The specific and significant depletion of 26G-RNAs in the absence of GTSF-1 suggests that we can utilize GTSF-1 to identify 26G-RNA targets precisely across nematodes. Our aim was to compare the targets for their conservation, gene function and strength of targeting.

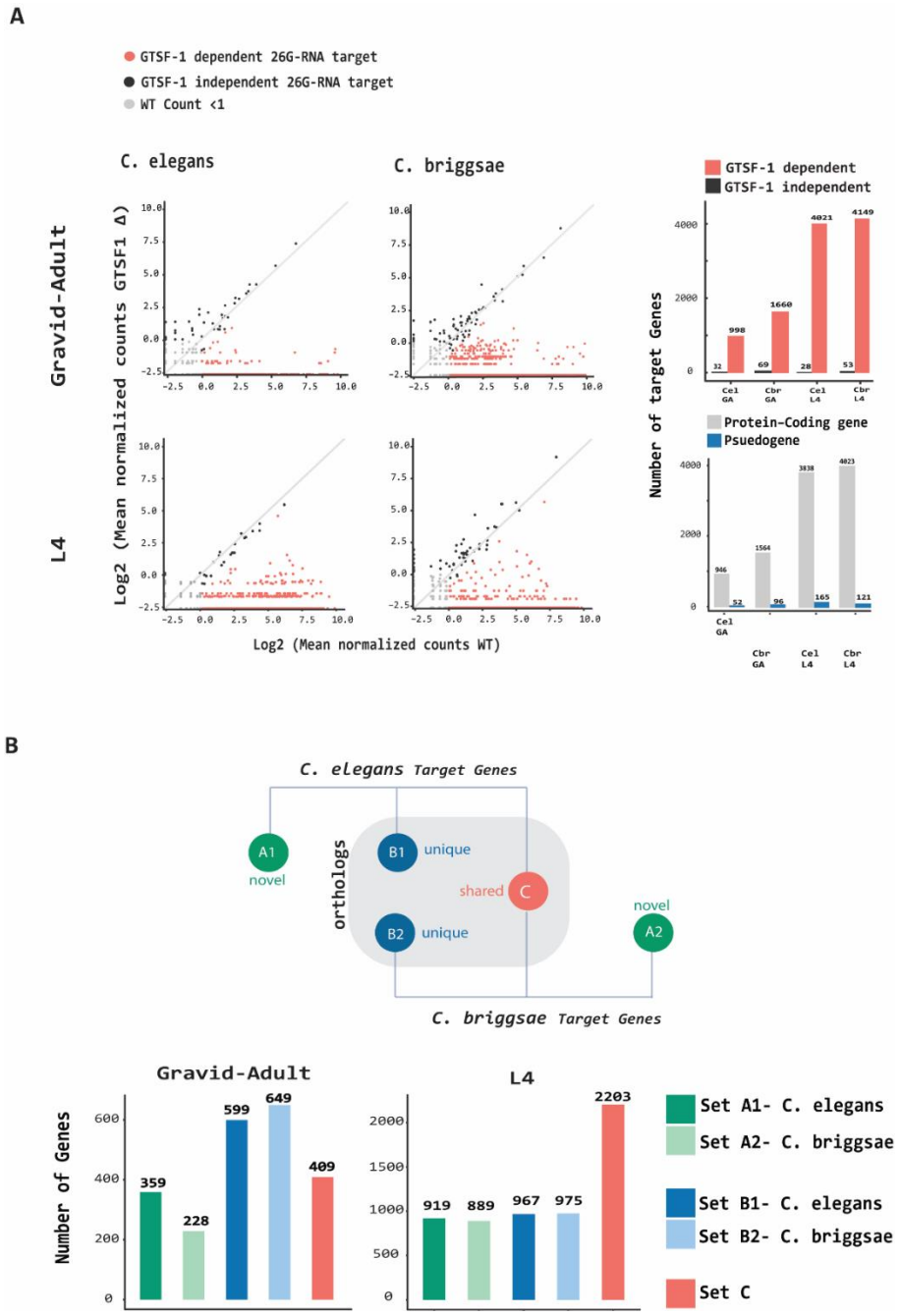
Identifying target genes

Our first task was to identify 26G-RNA target genes in the two species across both developmental stages. To achieve this, we utilized the DESeq2 software package within tinyRNA to generate a normalized count value for each original count. DESeq2 normalization takes into account both sequencing depth of each sample as well as RNA composition, adjusting for highly differentially expressed genes within samples. DESeq2 normalization is important for accurate comparison of expression between samples **(Figure 18A).**

We plotted the mean normalized 26G-RNA count for each gene in WT against *gtsf-1* deletion samples. Any gene with lower than one 26G-RNA count was considered biologically irrelevant and

Figure 18 | Comparison of 26G-RNA target genes from *C. elegans* and *C. briggsae*: (A) Scatterplot of Log2 Mean normalized 26G-RNA counts per gene from WT and *gtsf-1* Δ. GTSF-1 dependent targets are defined by a reduction in normalized average counts in the *gtsf-1* Δ strain to a level at least half of the WT levels, while the normalized average counts in the WT strain remain above 1. Diagonal line represents x=y. Inset | TOP: Barplot depicting number of GTSF-1 dependent and independent 26G-RNA target genes. BOTTOM: Barplot depicting biotype of GTSF-1 Dependent 26G-RNAs targets in both species.

(B) TOP: Schematic depicting classification of GTSF-1 dependent target genes into Set A B or C depending on 1:1 orthologs. We used reciprocal BLAST to identify 1:1 orthologs of all protein-coding genes between *C. elegans* and *C. briggsae*: SET-A includes target genes without a 1:1 ortholog. SET-B consists of genes that have a 1:1 ortholog but are only targeted in either one species. SET-C also includes genes that have a 1:1 ortholog, but the ortholog is a target in both species. BOTTOM: Barplot indicating number of genes in Set A, B or C highlighted in green, blue or red respectively. *C. elegans* gene sets are labelled A1 and B1 while *C. briggsae* samples are labelled A2 and B2.



thus discarded from further analysis. Among all the other genes, a significant portion lost at least half of their 26G-RNAs upon *gtsf-1* Δ **(Figure 18A).** We denoted them as "GTSF-1 dependent" target genes. However, both species demonstrate a small set of genes that retained more than half of their 26G-RNA levels. We denoted these as "GTSF-1 independent" target genes. While being an interesting observation, it is unclear whether these GTSF-1 independent genes are biologically relevant or artifacts. We conducted further analysis only with the "GTSF-1 dependent" target genes and for brevity, we refer to these as 26G-RNA target genes.

We observed that L4 stage animals have more 26G-RNAs as well more target genes, compared to GA animals. Within GA samples, *C. briggsae* animals exhibit both higher levels of 26G-RNAs as well as a higher number of target genes compared to *C. elegans* **(Figure 18B; Figure 15A).** This species specific difference was not observed for the L4 samples. This may suggest that specifically the oogenic branch of 26G-RNA pathway is more "active" in *C. briggsae* than *C. elegans*. This could also be a reflection of the higher copy number of ERGO-1 in *C. briggsae*. The best way to confirm this

would be to plot the 26G-RNA levels per target gene for both species. In both species, 26G-RNA target genes are predominantly protein coding genes, with a small proportion of pseudogenes **(Figure 18A).**

Categorizing target genes

Having identified the targets, our next questions were whether the target genes were conserved between the two species. Especially since a large majority are protein-coding, we wondered if the same genes or same pathways were targeted in both species.

To achieve this, we first defined the orthologs of all protein-coding genes between *C. elegans* and *C. briggsae*. We used reciprocal BLAST and filtered each gene pair with best e-value, to retain only 1:1 ortholog for each gene. We compared this list with the list of 26G-RNA target genes in both species and categorized all target genes into three sets **(Figure 18B):** Set A genes lack an ortholog in the other species. These constitute the set of genes regulated by 26G-RNAs that are found in one species over the other. Set B genes possess an ortholog but the ortholog itself is not a target. In other words, set B genes are found in both species but regulated in only one. Finally, Set C genes have an ortholog that is itself a target. These are the genes regulated in both species, and hence constitute a conserved target repertoire of the 26G-RNA pathway.

In *C. elegans* GA samples, we identified 359 genes in Set A and 228 genes in Set B. Meanwhile in *C. briggsae* GA samples, we found 599 genes in Set A and 649 genes in Set B. Finally, there were 409 genes in Set C representing the targets shared between both species. We also classified targets in the L4 samples in a similar manner. *C. elegans* L4 samples shows 919 genes in Set A and 889 genes in Set B. *C. briggsae* samples had similar numbers with 967 genes in Set A and 975 genes in Set B. The number of shared targets, i.e Set C was 2203 **(Figure 18B).**

This classification raised a number of questions. Firstly, are conserved genes are more prone to 26G-RNA targeting than non-conserved genes? How does this differ between Gravid-Adult and L4 stages? Both stages have a large number of shared targets encompassing Set C. Is this convergence merely coincidental, or does it indicate that targets in one nematode are probable targets in another nematode as well?

Statistical measures to compare target genes

The Fisher's exact test serves as a statistical tool to identify potential nonrandom associations between two categorical variables. This test calculates an odds ratio and a p-value, with the odds ratio representing the strength and direction of the association. An odds ratio of 1 implies no association, while values above or below 1 indicate positive or negative associations. Additionally, a low p-value (e.g., less than 0.05) suggests that the observed association is likely not due to random chance, indicating a genuine link between the variables.

We asked whether the likelihood of a gene being a target in both species is higher than what would be expected by chance. Using a Fisher's test for this analysis **(Figure 19A)**, we observed high odds ratios in both Gravid-Adult and L4 stages (25 and 35), and the results were statistically significant. This implies there is a strong association between a gene being targeted by 26G-RNA in both *C. elegans* and *C. briggsae*.

Next, we asked whether conserved genes are more prone to 26G-RNA targeting, and if so, how this differs between Gravid-Adult and L4 stages. We performed the Fisher's test on four conditions: *C. elegans* Gravid-Adult, *C. briggsae* Gravid-Adult, *C. elegans* L4, and *C. briggsae* L4

(Figure 19B). In L4 samples of both species, odds ratios exceeded 1 (1.57 and 1.78), indicating that conserved genes are more likely targets of 26G-RNAs. Conversely, in Gravid-Adult samples of both species, odds ratios were below 1 (0.75 and 0.85), suggesting that non-conserved genes are more likely targets **(Figure 19B).** All computed ratios were statistically significant. Given the high number shared targets in both developmental stages, we wondered if these genes also produce similar amount of 26G-RNAs in the two nematodes. To assess this, we plotted the normalized counts of each gene in one species over the other. We applied a linear regression model on the data, and calculated the Spearman's ranked correlation test **(Figure 20).**

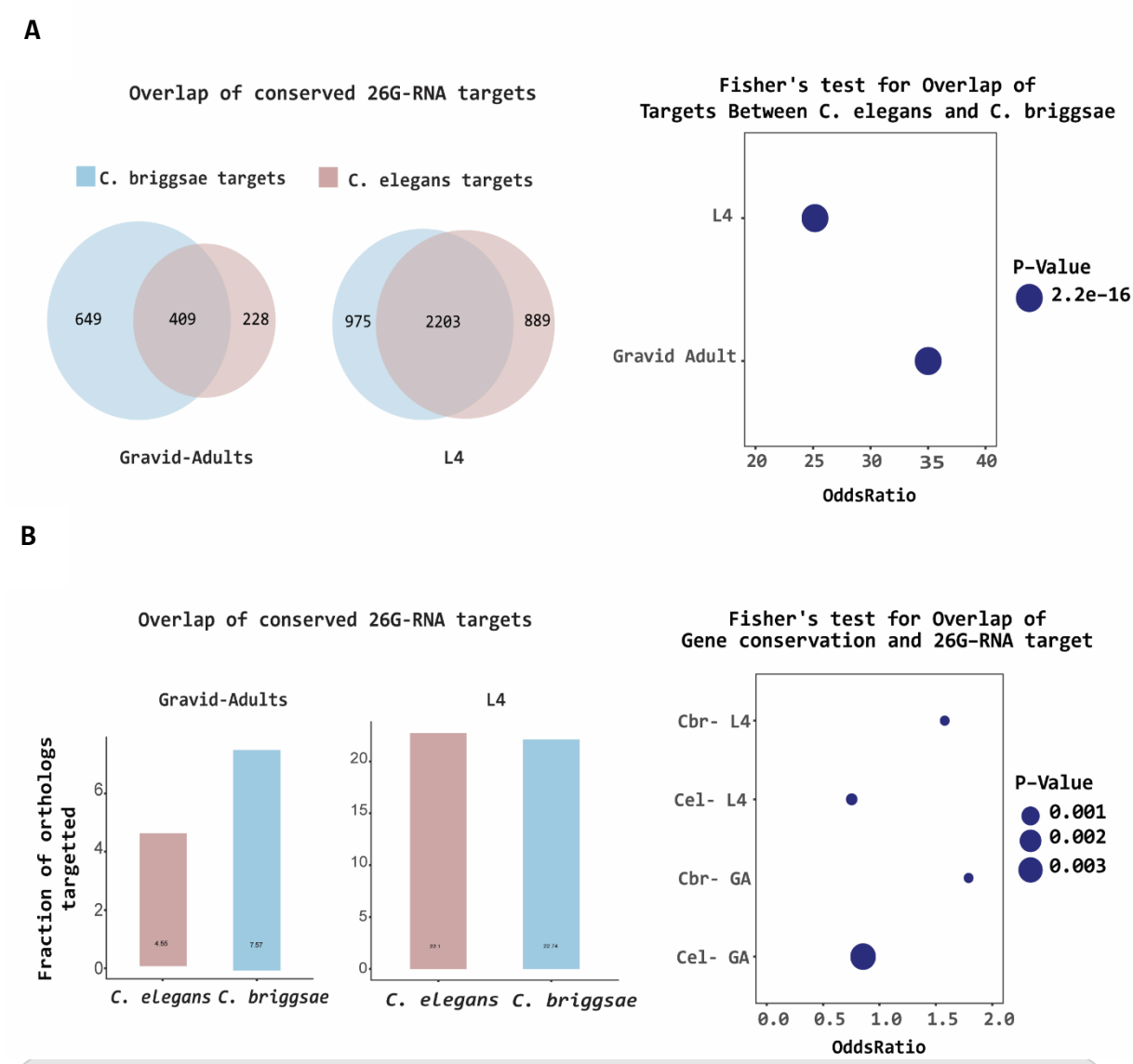


Figure 19 | Fisher's test on 26G-RNA target genes: (A) LEFT: Overlap of Set B1 and Set B2 to show Set C in Gravid-Adult and L4 samples. RIGHT: Bubble plot depicting results of Fishers test conducted on the variables: Number of 26G-RNA targets in *C. elegans* and Number of 26G-RNA targets in *C. briggsae*, considering only genes with a 1:1 ortholog in both species. x axis depicts odds ratio and size of circle indicates p-value. **(B)** LEFT: Barplot showing fraction of target genes conserved. y axis is the ratio of conserved targets (Set B+C) to total number of targets (Set A+B+C). RIGHT: Bubble plot depicting results of Fishers test conducted on the variables: Number of 26G-RNA targets in test sample set and Number of genes with 1:1 ortholog between *C. elegans* and *C. briggsae*. The test was conducted independently on all four samples sets. x axis depicts odds-ratio and size of circle indicates p-value.

Similar to Fisher's test, the Spearman's test assess strength of correlation between two set of variables, which in our case would be 26G-RNA counts in *C. elegans* and in *C. briggsae* for each gene in Set C. The resulting Spearman's rho (ρ) will be a value between -1 and 1. A positive value indicates a positive monotonic relationship, a negative value indicates a negative monotonic relationship, and a value of 0 indicates no monotonic relationship. The Spearman's ρ for Gravid-Adult and L4 datasets was 0.138 and 0.572 respectively **(Figure 20)**. Both tests were statistically significant. According to these results, there is a tendency to produce similar levels of 26G-RNAs from conserved genes between the two nematodes. However, this association much stronger in L4 stage compared to GA stage. Why would 26G-RNA levels per gene vary more in Gravid-Adult stage? One possibility is that GA Set C gene expression itself is more variable than L4.

Altogether, these results point to relatively high conservation of 26G-RNA targeting in L4 stages and poor conservation of targeting in GA stage of nematodes.

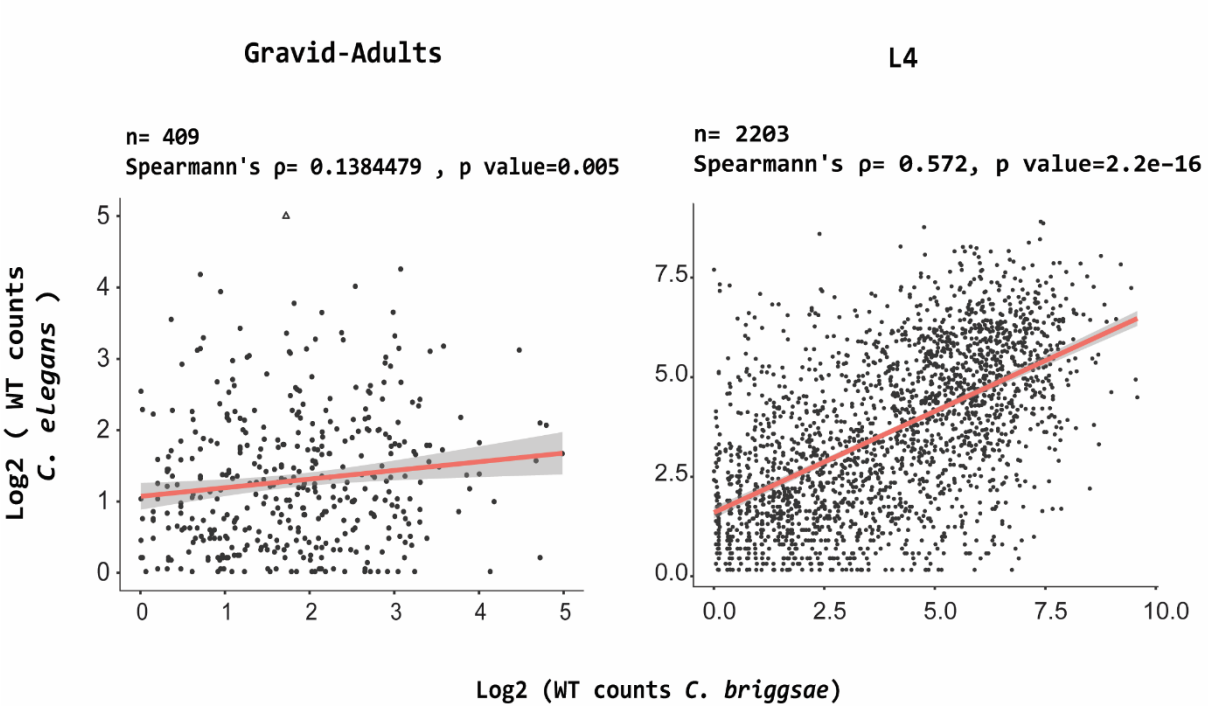


Figure 20 | Correlation of 26G-RNA counts per SetC gene*:* Scatterplot illustrating Log2 Mean Normalized 26G-RNA counts per gene in SetC. The x-axis represents counts in *C. briggsae* WT samples, while the y-axis represents counts in *C. elegans* WT samples. A linear regression model, implemented with the lm() function in R, was applied to the data using the formula Y ~ X, where Y is the outcome variable and X is the predictor variable. Spearmann's correlation coefficient and its associated p-value, depicted in the inset, were calculated in R using the cor.test() function. The displayed data include counts for n genes.

12. *P. pacificus* GTSF-1 interacts with the RRF-3 Module and ERI complex

So far, we observe the interaction of GTSF-1 with RRF-3 in two nematode species with a divergence of around 100 million years. In order to test this in *P. pacificus*. We developed affinity purified antibodies against PpaGTSF-1 and used this to immunoprecipitate GTSF-1 from *P.*

pacificus PS312 gravid-adult animals **(Figure 21A).** To reduce non-specific interactions, we used the *gtsf-1* Δ strain as a control. All experiments were performed in quadruplicates. Unfortunately, all *gtsf-1* Δ samples contained peptides of GTSF-1. This is similar to what we observed in *C. briggsae gtsf-1 Δ* **(Figure 13F).** Again, α-PpaGTSF-1 is also affinity-purified from rabbit serum, and might also be bound to low levels of PpaGTSF-1. As a result, the enrichment of GTSF-1 in wild-type samples was highly reduced. Likely due to this, we could detect only 19 proteins that were at least two-fold enriched with a p-value less than 0.05 **(Figure 21A).**

However despite this, we noticed a striking enrichment of RRF-3, DRH-3 and ERI-5 in wild-type samples **(Figure 21A).** We did not detect homologs of PIWI or any other AGO in the immunoprecipitate, even at an enrichment lower than the significance cutoffs. These results demonstrate that the preferential association of GTSF-1 with RRF-3 over PIWI persists into the *Pristionchus* genus. In addition, there is also evidence to suggest that *P. pacificus* homologs of GTSF-1, RRF-3, DRH-3 and ERI-5 assemble into an RdRP module similar to those found in *C. elegans*. As *gtsf-1* Δ causes a specific and complete loss of 26G-RNAs **(Figure 7C),** it is likely that *P. pacificus* RRF-3 module is necessary for 26G-RNA biogenesis.

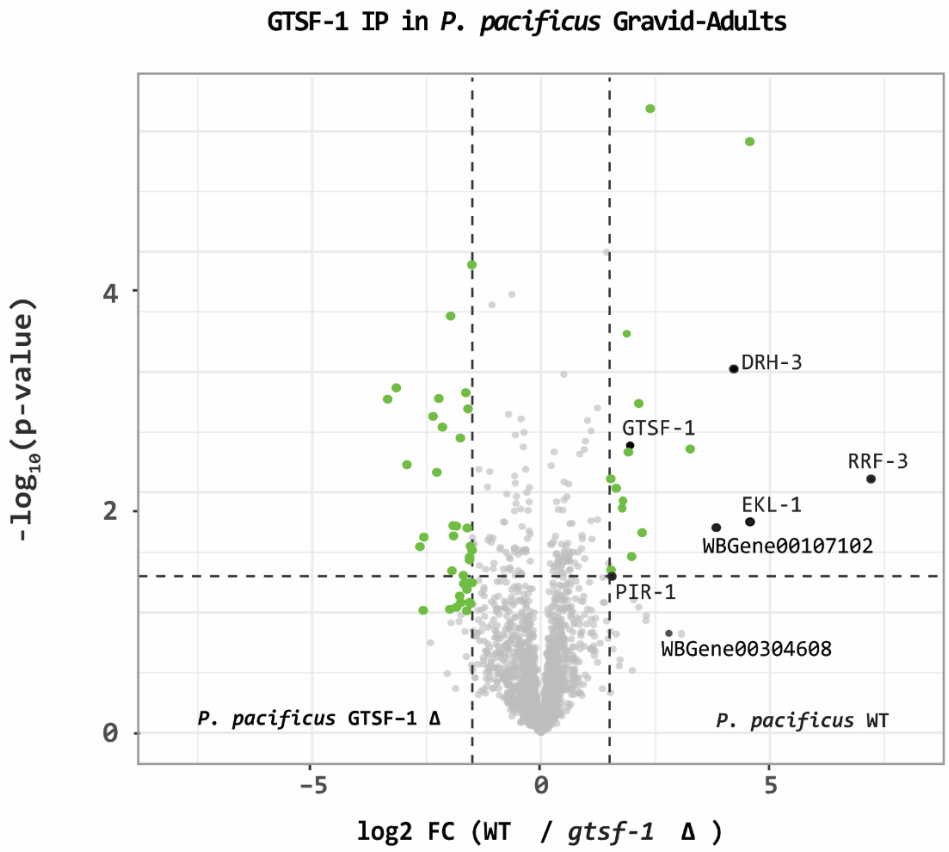
In *C. elegans*, DRH-3 is shared by both RRF-3 and RRF-1/EGO-1 RdRP modules (Gu et al., 2009, Aoki et al., 2007, Thiverge et al., 2013). Additionally, within the RRF-3 module, ERI-5 is present, while its paralog, EKL-1, is associated with the RRF-1/EGO-1 module. This particular module plays a crucial role in the biogenesis of 22G-RNA. Several species outside of the *Caenorhabditis* genus lack EKL-1 and this is also seen in *P. pacificus* **(Figure 29).** This intriguing observation raises the question if species without EKL-1 compensate for its absence by interacting with homologs of RRF-1/EGO-1 through ERI-5. Studying the phenotypes and interactors of RRF-3, RRF-1/EGO-1 and ERI-5 will shed light on their functions in specific sRNA pathways.

Further interactions with ERI proteins and Discovery of uncharacterized proteins

While peptides from GTSF-1 were found in all samples, peptides from RRF-3, DRH-3 and ERI-5 were only abundant in the wild-type samples, which gave us confidence that the experiment was successful despite the contamination of GTSF-1 in the control samples **(Figure 21B).** We clearly detected a number of peptides from a homolog of NYN-domain protein **WBGene00304608 (PPA46829)** and a conserved RNA-phosphatase PIR-1, **WBGene00117842 (PPA28288) (Figure 21B).** Similar to RRF-3, DRH-3 and ERI-5, peptides from both PPA46829 and PPA28288 were abundant in two wild-type samples and no *gtsf-1* Δ samples. This suggests that both PPA46829 and PPA28288 are associated with the RRF-3 module. PIR-1 is a conserved RNA-phosphatase that is required for accumulation of ALG-3/4 class of 26G-RNAs. PIR-1 interacts with the Dicer ERI complex proteins including RRF-3 and DRH-3 (Chaves et al., 2021). *P. pacificus* encodes a single homolog of PIR-1 - PA28288, which was found in our experiment.

The Zc3h12a-like NYN domain proteins are RNA nucleases (Habacher & Ciosk 2017) whose paralogs in *C. elegans*, such as RDE-8, NYN-1, NYN-2, NYN-3 and ERI-9 have been implicated in RNAi pathways (Habacher & Ciosk 2017, Tsai et al., 2022, Tsai et al., 2015, Pavlec et al., 2009). Specifically, RDE-8 and ERI-9 are required for accumulation of ERGO-1 class 26G-RNAs while NYN-3 is essential for ALG-3/4 class 26G-RNAs (Tsai et al., 2022, Tsai et al., 2015, Pavlec et al., 2009). RDE-8 and NYN-3 have a conserved catalytic domain and their function is to cleave targets mRNAs and recruit RdRP for siRNA production. ERI-9 on the other hand contains point-mutations in catalytic domain and possibly plays a structural role in the ERI complex (Habacher & Ciosk 2017). PPA46829 is phylogenetically closest to ERI-9 and likewise has lost the catalytic aspartic acid residues **(Figure 22A).**

A



B

Gene.names	Protein.names	Peptides_Δ_1	Peptides_Δ_2	Peptides_Δ_3	Peptides_Δ_4	Peptides_WT_1	Peptides_WT_2	Peptides_WT_3	Peptides_WT_4	log2FC	-log10(p.value)	significant
WBGene00098302	Uncharacterized protein	1	1	1	1	2	2	2	2	4.560828358	4.915148176	+
WBGene00304839;WBGene00089591	**RRF-3**	**0**	**0**	**1**	**1**	**71**	**73**	**62**	**80**	**7.209314124**	**2.110427644**	**+**
WBGene00102740	**DRH-3**	**4**	**3**	**6**	**8**	**13**	**18**	**11**	**24**	**4.212070433**	**3.024626343**	**+**
WBGene00097118	**EKL-1**	**1**	**0**	**0**	**1**	**9**	**17**	**4**	**19**	**4.566971492**	**1.753839823**	**+**
WBGene00103672;WBGene00205176	Insect cuticle protein;Uncharacterized protein	0	0	0	1	2	2	3	2	3.258580302	2.35990379	+
WBGene00113102	RNA binding protein	1	1	1	1	1	1	1	1	2.386011285	5.189266914	+
WBGene00107102	**Uncharacterized protein**	**1**	**2**	**2**	**1**	**11**	**16**	**6**	**24**	**3.828110937**	**1.705698036**	**+**
WBGene00108593	GYF domain-containing protein	0	0	0	1	1	1	1	1	2.13668091	2.737939752	+
WBGene00090521	GrpE protein homolog	6	7	7	7	9	9	10	9	1.873538686	3.318405028	+
WBGene00304918	**GTSF-1**	**9**	**12**	**12**	**12**	**13**	**12**	**14**	**13**	**1.946084094**	**2.389296511**	**+**
WBGene00093669	LNKN-1	1	0	0		1	1	1	1	2.20874336	1.664024637	
WBGene00203876	ZTF-18	1	1	1	1	3	3	2	3	1.907833768	2.33410927	
WBGene00098693	Ubiquitin	1	0	0	1	1	1	1	1	3.069649015	0.817515616	
WBGene00113096	Uncharacterized protein (Fragment)	5	5	4	4	6	6	6	5	1.788629287	1.922033997	
WBGene00276991	Ribonuclease kappa	0	0	1	1	1	1	1	1	1.97852696	1.466127417	
WBGene00280459	Uncharacterized protein	2	2	2	2	2	2	1	3	1.769117511	1.872350863	
WBGene00304608	**NYN domain protein**	**1**	**2**	**0**	**1**	**3**	**10**	**3**	**14**	**2.79393817**	**0.827304835**	
WBGene00117842	PIR-1	0	0	0	0	0	1	0	3	1.546708039	1.300671486	

Figure 21 | IP-MS of GTSF-1 in *P. pacificus*: (A) Volcano plots representing label-free proteomic quantification of quadruplicate Immunoprecipitates (IPs) from *P. pacificus* PS312 extracts from Wild-type or *gtsf-1* Δ [*Ppa-gtsf-1* (*xf208*)] gravid-adult animals. A total of 1.5mg worm lysate and 2ug of bead-bound antibody was used for the IP. The x axis represents the median fold enrichment of individual proteins precipitated with one antibody over another. The y-axis indicates -Log10 (p-value) of observed enrichments. Dotted lines represents thresholds at p=0.05 and 1.5-fold enrichment. Data points highlighted in green represent statistically significant interactors. Relevant interactors are highlighted in black. **(B)** Section of the MaxQuant output table for the IP-MS experiment in (A). Peptides coloumn indicates the number of unique peptides detected in the immunoprecipitate and assigned exclusively to the protein in each of the four replicates of GTSF-1 Δ and Wild-type PS312.

Interestingly, peptides from an uncharacterized protein WBGene00107102 (PPA17548) were also abundant in the wild-type samples. PPA17548 has no annotated domains and is mostly disordered **(Figure 22B).** We used BLAST on Wormbase Parasite to find homologs of PPA17548 in nematodes. Homologs of this protein were found in several *Pristionchus* species but not in any *Caenorhabditis* species **(Figure 22C).** Outside of *Pristionchus*, homologs were found in two clade V nematodes *Parapristionchus giblindavisi* and *Micoletzkya japonica* **(Figure 22C)**. The functions of this protein would be interesting to follow up on.

In summary, homologs of several known factors of *C. elegans* ERI complex and 26G-RNA biogenesis were found to immunoprecipitate with GTSF-1 in *P. pacificus*. This shows a strong conservation of not just GTSF-1 function but also a putative ERI complex. It would be interesting to study the dependence of *P. pacificus* RRF-3 module on DCR-1.

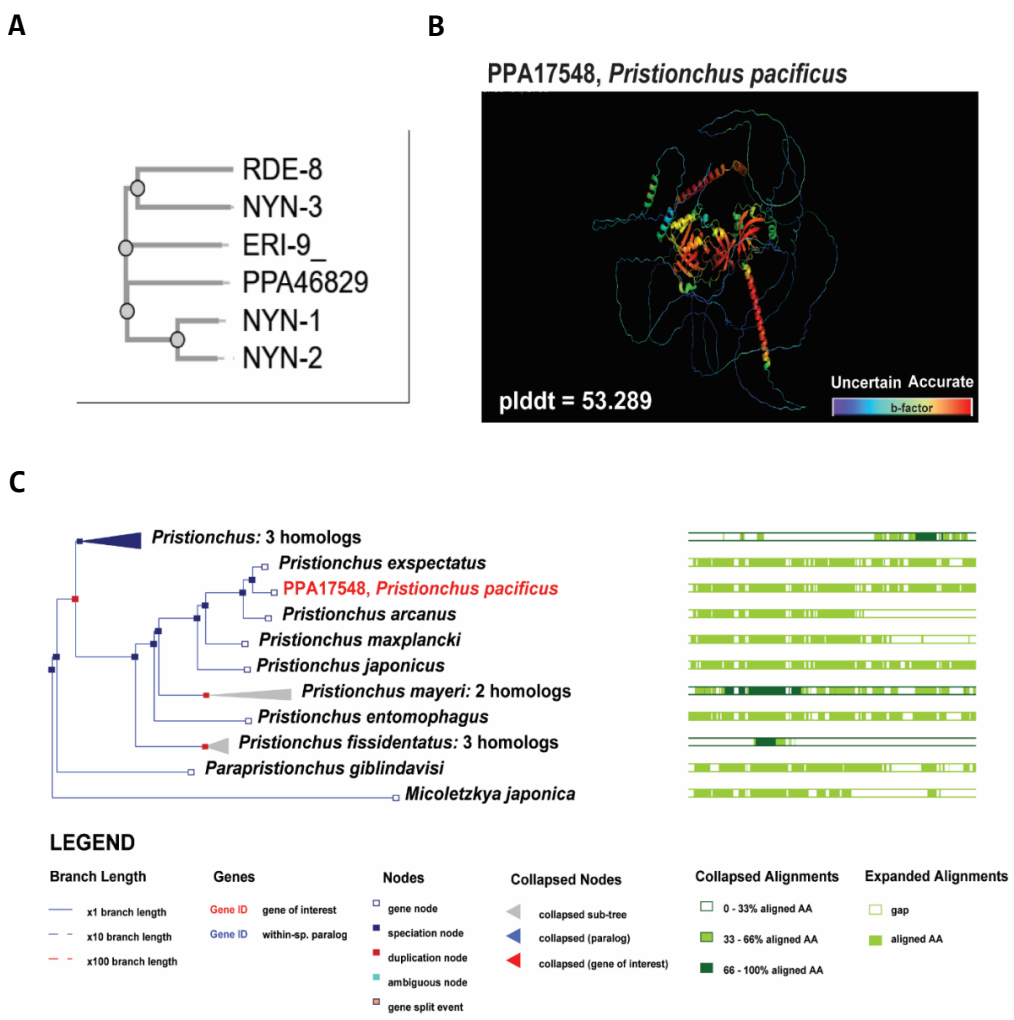


Figure 22 | Factors identified in PpaGTSF-1 IP-MS: (A) Phylogenetic relationship of NYN-domain proteins from *C. elegans* and PPA46829 from *P. pacificus*. The phylogenetic tree was computed using the multiple sequence alignment program ClustalOmega. Default settings without tree branch corrections were used. **(B)** Alpha-Fold predicted structure of PPA17548 using the monomer model. b-factor and pLDDT (average) values are residue specific and overall structure prediction accuracy scores respectively. **(C)** LEFT: Phylogenetic connections among homologs of PPA17548 across nematodes. RIGHT: Alignment of protein sequences from identified homologs. The WormBase Parasite webserver's BLASTP was employed to discover homologs in all documented nematode proteomes, with all recognized homologs represented here. Additional details are provided in the inset legend.

13. GTSF-1 Interaction with the N-Terminus Region of RRF-3: Insights from Biochemical and Structural Analyses

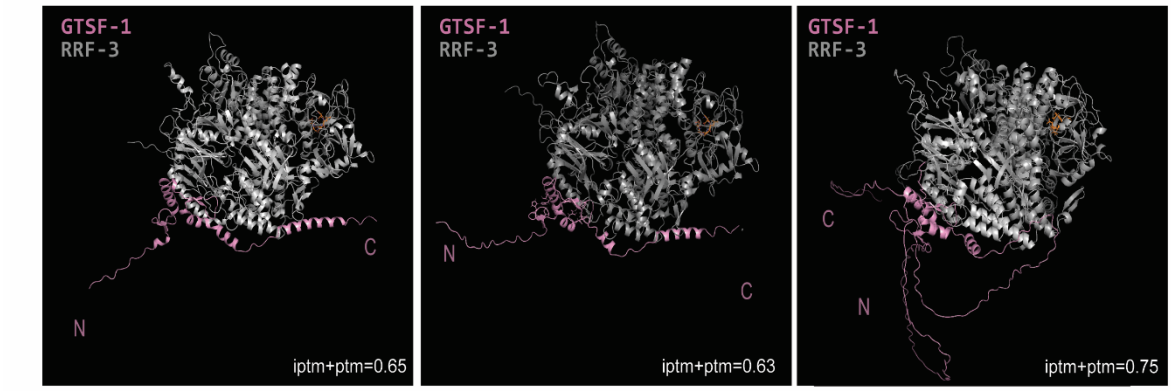
Our investigation into the association between GTSF-1 and RRF-3 in *C. elegans*, *C. briggsae* and *P. pacificus* has revealed consistent and highly conserved interactions. This specificity prompted further exploration into the biochemistry underlying this binding. Utilizing the Alpha-Fold multimer model, we predicted the structures and binding of GTSF-1 to RRF-3 for all three nematodes. The resulting models had robust scores ranging from 0.63 to 0.75 **(Figure 23A).** Notably, the N-terminus of GTSF-1 consistently associated with the RRF-3 protein in all three species. In *P. pacificus*, an interesting observation was made—the C-terminus of GTSF-1, characterized by long stretches of negatively charged amino acids **(Figure 4A)**, displayed increased interaction with RRF-3 compared to *C. elegans* and *C. briggsae*. This might contribute to the higher model confidence scores observed in *P. pacificus* predictions **(Figure 23A).**

Examining the region of RRF-3 bound by GTSF-1, we identified a consistent pattern among the three species. RRF-3 has a single annotated RdRP domain that houses the active site of triple Asp (D) motif. This motif is conserved across all three nematodes in our study **(Figure 23C).** According to our predictions, binding of GTSF-1 occurs distally to the RRF-3 active site **(Figure 23A),** distinguishing it from homologs in mice and flies, where the interaction occurs directly at the active site of the PIWI domain (Arif et al., 2022). The region of RRF-3 where GTSF-1 binds was pinpointed through a visual examination of structures in Pymol. We identified a unique segment spanning 400-600 amino acids that shows independent folding **(Figure 23B).** This specific area was labelled as the GTSF-1 Interacting Region (GID). The GID prediction displayed favorable b-factor scores **(Figure 23E).** We created a sequence alignment of RRF-3 GID across all three species and found a good degree of similarity **(Figure 23D).**

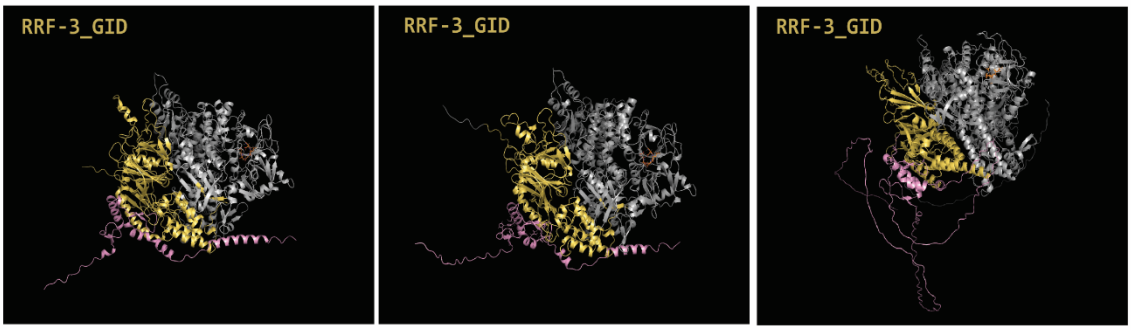
Through Pymol, we performed structural alignment of *P. pacificus* and *C. briggsae* RRF-3 GID to *C. elegans* and found these domains to align perfectly **(Figure 23E).** These results supporting the conserved nature of the GTSF-1-RRF-3 interaction and suggest an uncharacterized region within RRF-3 to serve as GTSF-1 binding interface.

Figure 23 | Prediction of GTSF-1 and RRF-3 interaction: (A) AlphaFold2 (v2.3.2) *in-silico* structural predictions depict the multimeric models of GTSF-1 and RRF-3. The RRF-3 catalytic site is highlighted in orange and displayed in stick format, while the N and C termini of GTSF-1 are indicated. **(B)** The region of RRF-3 interacting with GTSF-1 from the structures in (A) is emphasized in yellow. **(C)** The domain organization of RRF-3 is illustrated, with the RdRP domain annotated in InterPRO. GID is annotated based on AlphaFold2 predictions. Multiple Sequence Alignment (MSA) using ClustalOmega emphasizes the conservation of RRF-3 catalytic residues. Interpro annotation of RdRP overlaps with RRF-3_GID annotation. **(D)** MSA showcases various RRF-3_GID sequences. **(E) (Page 74)** Structural alignments of RRF-3_GID from *C. briggsae* and *P. pacificus* to *C. elegans*. The structures are colored by b-factor scores, depicting residue-level prediction accuracy.

A



B



C


```
CelRRF-3    ITKNPCHVPGDVRVFDAVWQPALAHLVDVVVFPQHGPRPHPDEMAGSDLDGDEYSIIWDQ 1146
CbrRRF-3    ITKNPCHVPGDVRVFDAVWQPALAHLVDVVVFPRHGPRPHPDEMAGSDLDGDEYSIIWDQ 1152
PpaRRF-3    VSKNPCHAPGDARVFEAVWQPALSHLSDVIVFPRYGERPHADEMAGSDLDGDEYSVIFDP 997
            ::******.***.***:*******:** **:***::* *** **************:*:*
```

D

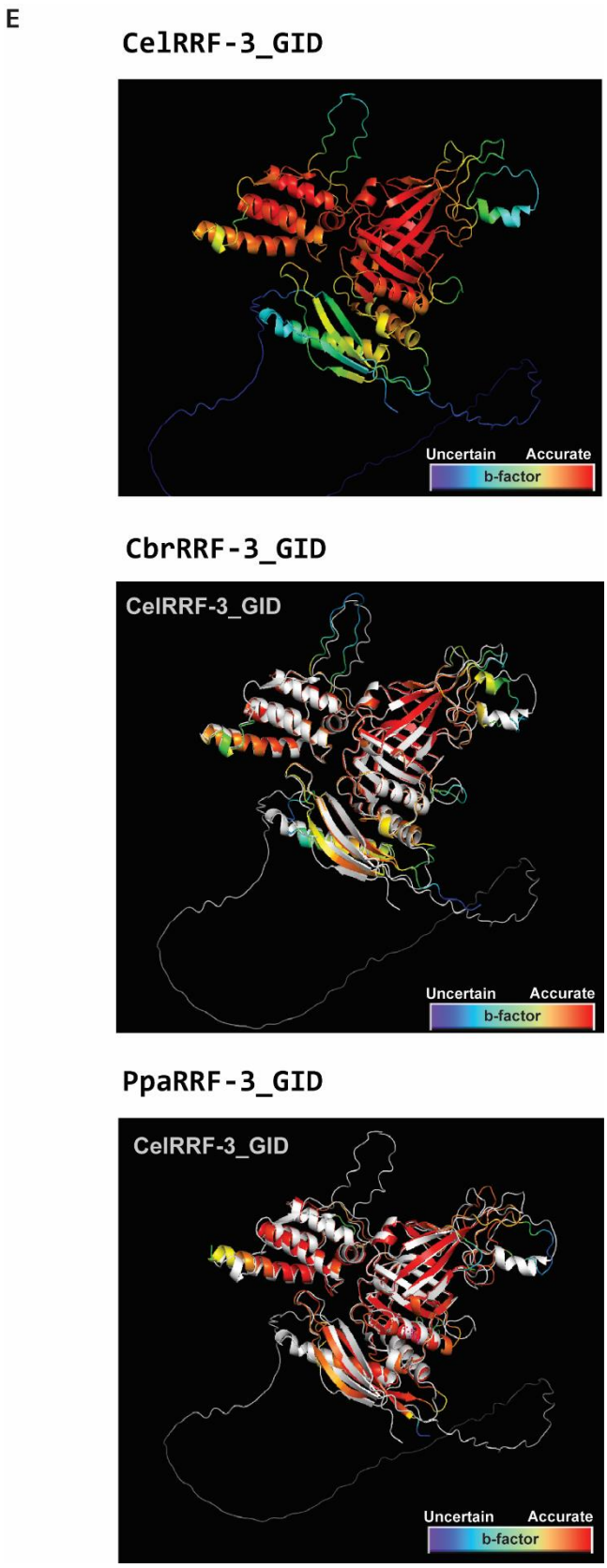
```
Ppa_GID     ------------------------------------------------------------  0
Cel_GID     SNKMLLTTFRMDRNSKSKSEVDVQEQPVHSSSSAFPGNHLNNFSYPVNRGYLRDYLLQSQ  60
Cbr_GID     ------------------------------------------------------------  0

Ppa_GID     ------------------------------------------------------------  0
Cel_GID     RPSTSKPVDCSVLKRHSLPSTHILYEKTKHRGGVNIEEQEKLVRMLWAAAEESETVAKTR  120
Cbr_GID     ----------------------------------------------------------AKP  3

Ppa_GID     ---------------IAGSTYSFVVRICLPLEMMPGIEAKRYIDAMIRKSGGRVESVVRC  45
Cel_GID     QFSKKQAIELNFDAKLIGSMNNDCFGYCRAH--ME--NIKDVLKTHLKLSKVDEVNWIKV  176
Cbr_GID     AVVKKAPIELSLDVKITGSMNNEMFGYCRTH--ME--TMKEFFSGKLKESNIDEVNWMKT  59

Ppa_GID     ARIPSADYEI---MMKEHIVNVQIPVSSESTVQIALAAFLSRFSTESRGGLQNQLMIEIP  102
Cel_GID     GMVPRAAYEDKSYVIDAHLVL--TPNGEVEDENELFSEFASSFTSRITGMLHDQVFLEVP  234
Cbr_GID     GMAPRAAYEDKAYVIEAHLIL--TPSVAAFDKNDLFAQCITKFTRKTKGMLQDQCFLEVP  117

Ppa_GID     KLSTVFNEIRTEDNQIGLGNFALGNCPNPGLFFVRGNYFTKVNSMPSNTKLSTGNLDK--  160
Cel_GID     KMHTLFTKITPQHMDINISAIAIGNCPNSGLFLVRGDFISQENTVCSVKLQSHHN-ADAS  293
Cbr_GID     KMHQFFPKITPQHVDINVSAIGLGNCINPGLFLIRGDYISNDNTVCSVKLQSQHNPSDAS  177

Ppa-GID     -NARIEKPFGTLNLLSYVHFEHDQKHISIKFAIPMSKPDEDGLAFRGFKMNLQYSAIQSI  219
Cel_GID     RENSSFKVAGSNKYLSYARFEHDKRLAVVYFGVRLAEFADDGLDHAGFRLNLYYNLFVRI  353
Cbr_GID     REFNSYKVAGSSKYLSFARFEHDKRVALLYFGVRLSEFADDGLDHAGFRLNLYYNTFVRI  237

Ppa_GID     TVDSGQKNLNRKSRSIDFSFRLNHPPSLYEVQPTTMADQKRIINLQKTIRTYNTWSRVLQ  279
Cel_GID     VVDMSHETTNS------IYIQMKNPPHLWEGIPKNTIFHPSK-SKVLNMETCTEWTRVLS  406
Cbr_GID     VVDLSAVNTNS------IYIQLKNPPQLWEGIPKSTMFHPSK-SKVLNLETCTEWIRVLS  290

Ppa_GID     WPGHD--RSHGCSKECLSDSSVLHLSIHLNNEKDRVR---MFSIAARIQSRVPTIKVYFG  334
Cel_GID     WPGDAEGRGVGCTSEAFSQSSWIRLTMRKDDDNDSVSSTQLMDIVTRLSA-RSKAKVMFG  465
Cbr_GID     WPGDAERRGIGCTPEAFYQSNWIRIMFRKETDDEKLCSANLIDVVTRLAS-RSNAKVSFG  349

Ppa_GID     SIVSMRRRVAKIRSLPPLGSYRADYALKAIISRGTVVTDELFDAAKENN-----------  383
Cel_GID     SIFSIRRKLAPSPAFHSLGSFRANYALQALITRGSVFTDQLFDATDENIPSSDNDNDEDD  525
Cbr_GID     HIFSIRRQLAPSPCLASLGSFRANYALQALITRGSVFMDQLFDCTDTNIPDPLEPEDEEG  409

Ppa_GID     --------------TVPFFIRRLLWCMKECQAACEEALELVLTAIDERRNITIERAFDRL  429
Cel_GID     DDDVDDTKKPMELVHEPLFLKLVRRGMKECSQATEETLEQLLNAFDERRQIDVVTAFTTM  585
Cbr_GID     --EGELLKQPMNLKDEPMFLKIVRRAMRECPQAAEESLEQLLNAFDERRSIDVIFAFSTM  467

Ppa_GID     YTTRVELYRRSASSH-------------------     444
Cel_GID     YQSRKIQYERLLKGESLQDVGLAKPLPKNCVSVA     619
Cbr_GID     YKTRKVQYERLLNGESLQDVGLAR----------     491
```

Experimental Validation through Yeast-Two-Hybrid Assay

To validate the predicted models of GTSF-1 and RRF-3 interactions further, we conducted Yeast-Two-Hybrid assays. The experiment involved the cloning and expression of full-length GTSF-1 and RRF-3, as well as specific domains, namely the N-terminal domain (NTD) and C-terminal domain (CTD) of GTSF-1, and the GID and RdRP domains of RRF-3 **(Figure 24A, 24B).**

In *C. elegans*, GTSF-1 exhibited false-positive interactions, particularly evident in the CTD **(Figure 24A).** The NTD of GTSF-1 interacted with full-length GTSF-1, suggesting a potential role in protein dimerization. The RRF-3 GID showed stable expression and robust interactions with full-length GTSF-1 **(Figure 24A).** GID also interacted with the RdRP domain of RRF-3, possibly forming the complete RRF-3 protein. Surprisingly, RRF-3 GID had stronger binding to GTSF-1 CTD compared to NTD **(Figure 24A).** This may be due to the tendency for false positive interactions of the CTD, however the biological relevance of this finding cannot be dismissed.

In *C. briggsae*, the results were more straightforward. None of the constructs showed false-positive interactions. A distinct interaction was observed between the NTD of GTSF-1 and the GID of RRF-3. This interaction persisted even under more stringent selection conditions, reinforcing the validity of the interaction. Intriguingly, *C. briggsae* GID did not bind with the RdRP domain.



These observed interactions align with the AF2-predicted models, confirming association between RRF-3 GID and GTSF-1. The *C. elegans* GTSF-1 CTD's false-positive interactions, not observed in *C. briggsae*, underscore biochemical distinctions between these homologs. Furthermore, in *C. briggsae*, the NTD of GTSF-1 emerged as a key mediator in the interaction with RRF-3 GID. Within the GTSF-1 NTD, we identified four residues that mediate majority of the interactions with RRF-3. These interactions are mostly polar, through the positively charged Arginines, and some pi-pi interactions through the aromatic Tryptophan **(Figure 25A).** We created GTSF-1 constructs with these residues mutated to alanine and tested binding to RRF-3 via Y2H. This experiment was

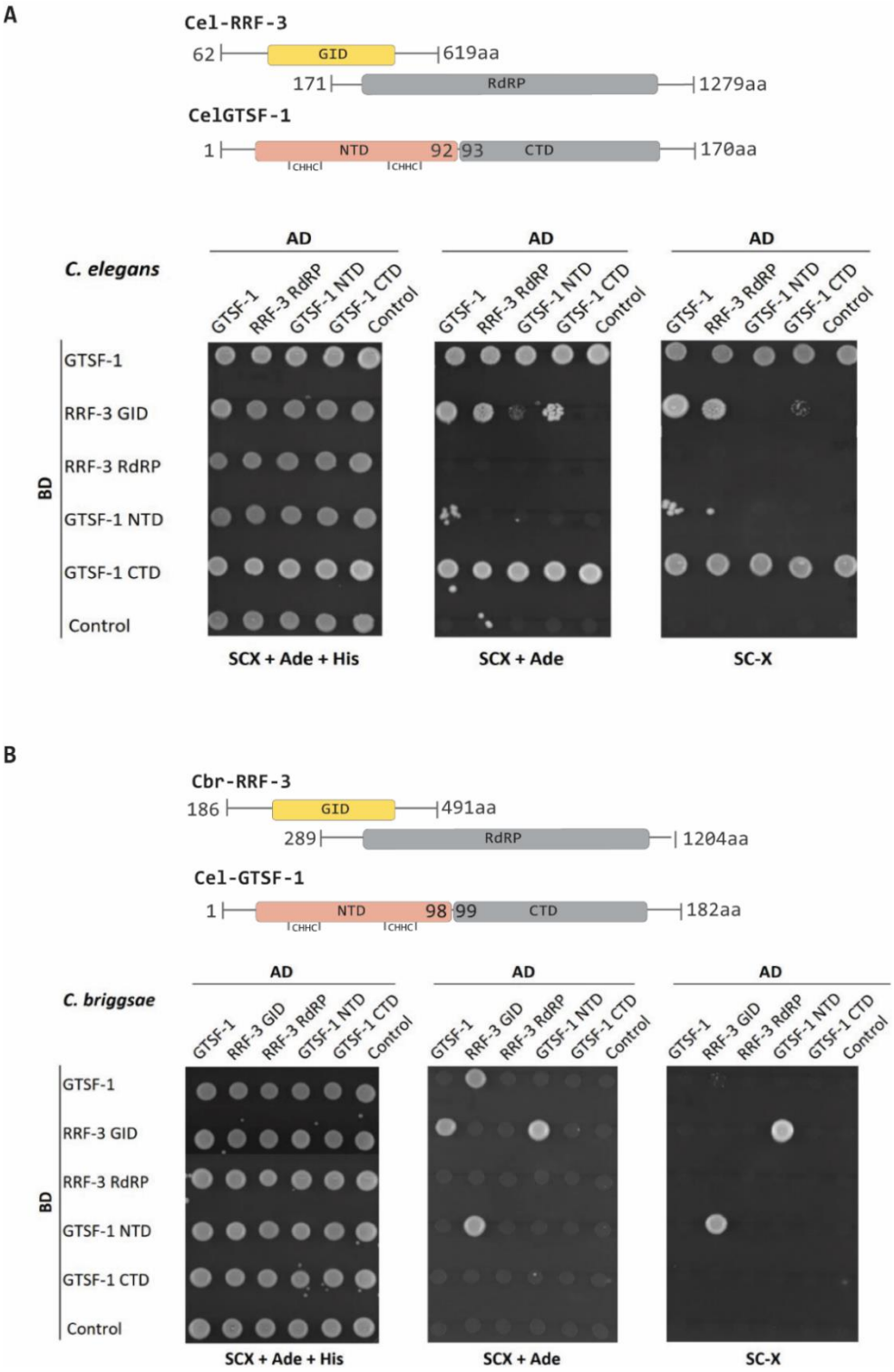


Figure 24 | Yeast-Two Hybrid to assess GTSF-1 and RRF-3 interaction: (A) *C. elegans* **(B)** *C. briggsae* TOP: Schematic illustration depicting domains and residues tested in the yeast-two-hybrid assay. BOTTOM: Yeast-Two-Hybrid assay under control (-LEU, -TRP), low stringency (-LEU, -TRP, -HIS) or high stringency (-LEU, -TRP, -HIS -ADE) conditions. GTSF-1 is full-length GTSF-1. Controls are either Activation Domain (AD) or Binding domain (BD) expressed without fusion to any protein. *C. elegans* GTSF-1 and GTSF-1 CTD shows false positive interaction.

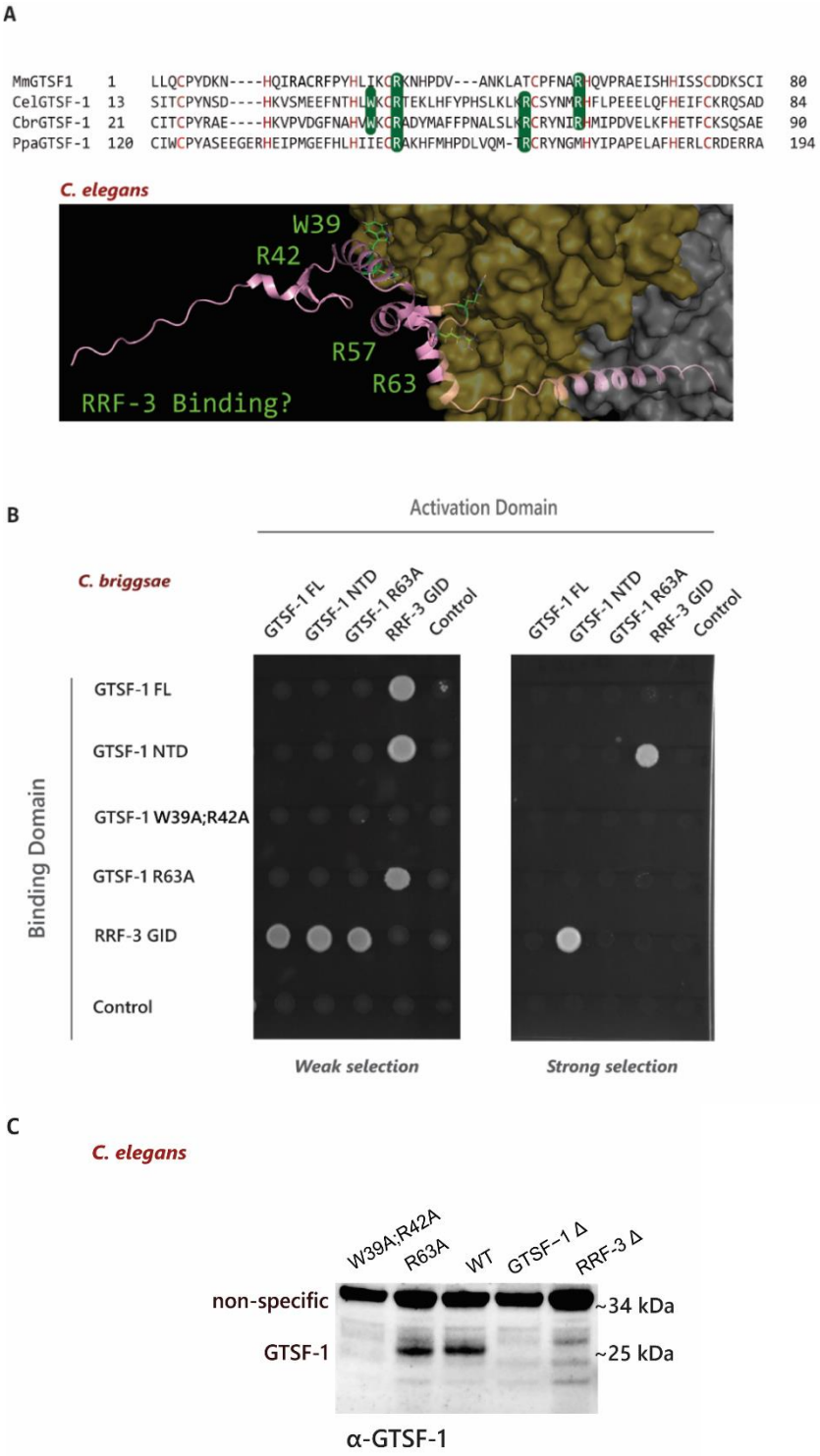


Figure 25 | Aromatic and Positively charged residues in NTD of GTSF-1 bind RRF-3: (A) TOP: Multiple sequence alignment displays a segment of GTSF-1 NTD, with CHHC Zn finger residues highlighted in red. Green bars indicate conservation of RRF-3 binding residues. BOTTOM: AlphaFold2 predictions highlight interactions, showcasing residues of *C. elegans* GTSF-1 against RRF-3 GID (depicted in yellow). **(B)** Yeast-Two-Hybrid assay with *C. briggsae* under weak selection (-LEU, -TRP, -HIS) or strong selection (-LEU, -TRP, -HIS -ADE) conditions. GTSF-1 FL is full-length GTSF-1. Controls are either Activation Domain (AD) or Binding domain (BD) expressed without fusion to any protein. **(C)** Western blot to detect GTSF-1 in various *C. elegans* mutant strains. First two lanes are point mutation in GTSF-1. Non-specific band serves as loading control.

conducted with *C. briggsae* constructs, as they lack false positive interactions. For the construct containing two point mutations (W39A, R42A), we observe a complete loss of binding to RRF-3 GID **(Figure 25B).** For the second construct, we tested for the point mutation R63A. This construct interacted with RRF-3 GID, however the colony growth is slightly reduced compared to WT, suggesting potential loss of interaction with RRF-3 **(Figure 25B).** Next, we generated these mutants in *C. elegans* using CRISPR/Cas9. To our surprise, the W39A,R42A mutation caused a complete loss of GTSF-1 *in-vivo*. But since deletion of RRF-3 also leads to complete loss of GTSF-1, it is likely that these mutations abolished binding to RRF-3, which caused instability and triggered degradation of GTSF-1. The mutation R63A did not alter protein levels **(Figure 25C).**

Overall, these results support the AF2 predicted binding models and provide valuable insights into the GTSF-1 and RRF-3 interaction through their N- terminal domains. The species-specific variations in GTSF-1 CTD is emphasized in the Y2H assay.

14. Predicting specificity of GTSF-1 binding to RRF-3 using AlphaFold2 predictive structure modeling

After identifying a domain in RRF-3 that interacts with GTSF-1, we questioned whether such a domain exists in the other RdRPs of *C. elegans*: RRF-1 and EGO-1. Employing AlphaFold2 (AF2), we predicted the individual structures of all three RdRPs. According to protein databases like InterPro, both RRF-1 and EGO-1 possess a single annotated RdRP domain, structurally akin to the RRF-3 RdRP domain. Our AF2 modeling revealed a long N-terminal IDR in RRF-3, absent in both RRF-1 and EGO-1 **(Figure 26A).** Aligning this N-terminal region of RRF-3, comprising both the IDR and the interacting domain (GID), with RRF-1 and EGO-1 **(Figure 26A)** showed good alignment among the structures, although not as precise as the RRF-3 GID across the three nematode species **(Figure 23E).** Subsequently, we conducted a Multiple Sequence Alignment of the N-terminal region of RdRPs, revealing that the residues in RRF-3 mediating interaction with GTSF-1 *in-silico* are entirely absent from RRF-1 and EGO-1 **(Figure 26B).** Consequently, AF2 predicts a much lower confidence in the binding of GTSF-1 to both RRF-1 and EGO-1 (ipTM+pTM $\sim$ 0.4) compared to RRF-3 or RRF-3 GID (ipTM+pTM $\sim$ 0.7) **(Figure 26C).** Thus, despite the structural similarity between the RdRPs, RRF-3 uniquely possesses the necessary residues for binding GTSF-1.

This explains why GTSF-1 selectively binds to RRF-3. Employing a similar approach, we asked why GTSF-1 fails to bind PRG-1. Despite the loss of PIWI from numerous nematode lineages, Clade V nematodes have retained this Argonaute deeply. PRG-1 in *C. elegans*, *C. briggsae*, and *P. pacificus* possess the catalytic residues (data not shown), yet mutations in these residues in *C. elegans* do not affect piRNA-mediated gene regulation (Lee et al., 2012). This leads us to question whether PRG-1 harbors differences from other PIWI orthologs, preventing binding by GTSF-1. To investigate this, we generated models of Mus musculus GTSF1 with *Mus musculus* PIWI (MIWI) and with *C. elegans* PIWI (PRG-1) **(Figure 27A, 27B).** Recent research has elucidated how MmGTSF1 associates with the PIWI domain of MIWI, with aromatic residues in the C-terminal domain mediating PIWI binding and positively charged residues in the N-terminal domain facilitating RNA binding (Arif et al., 2022). Surprisingly, MmGTSF-1 associates with PRG-1 exactly as it does with MIWI. Alignments of GTSF1-bound MIWI and PRG-1 reveal negligible differences **(Figure 27C).** These observations are clearly reflected in the confidence scores of the AF2 multimer model **(Figure 27D).** The interaction of MmGTSF1 with both MIWI and PRG-1 is predicted with very high confidence, while the interaction of CeGTSF-1 with both MIWI and PRG-1 is very low **(Figure 27D).**

These findings strongly suggest that differences in nematode GTSF-1 are responsible for the loss of interaction with PIWI.

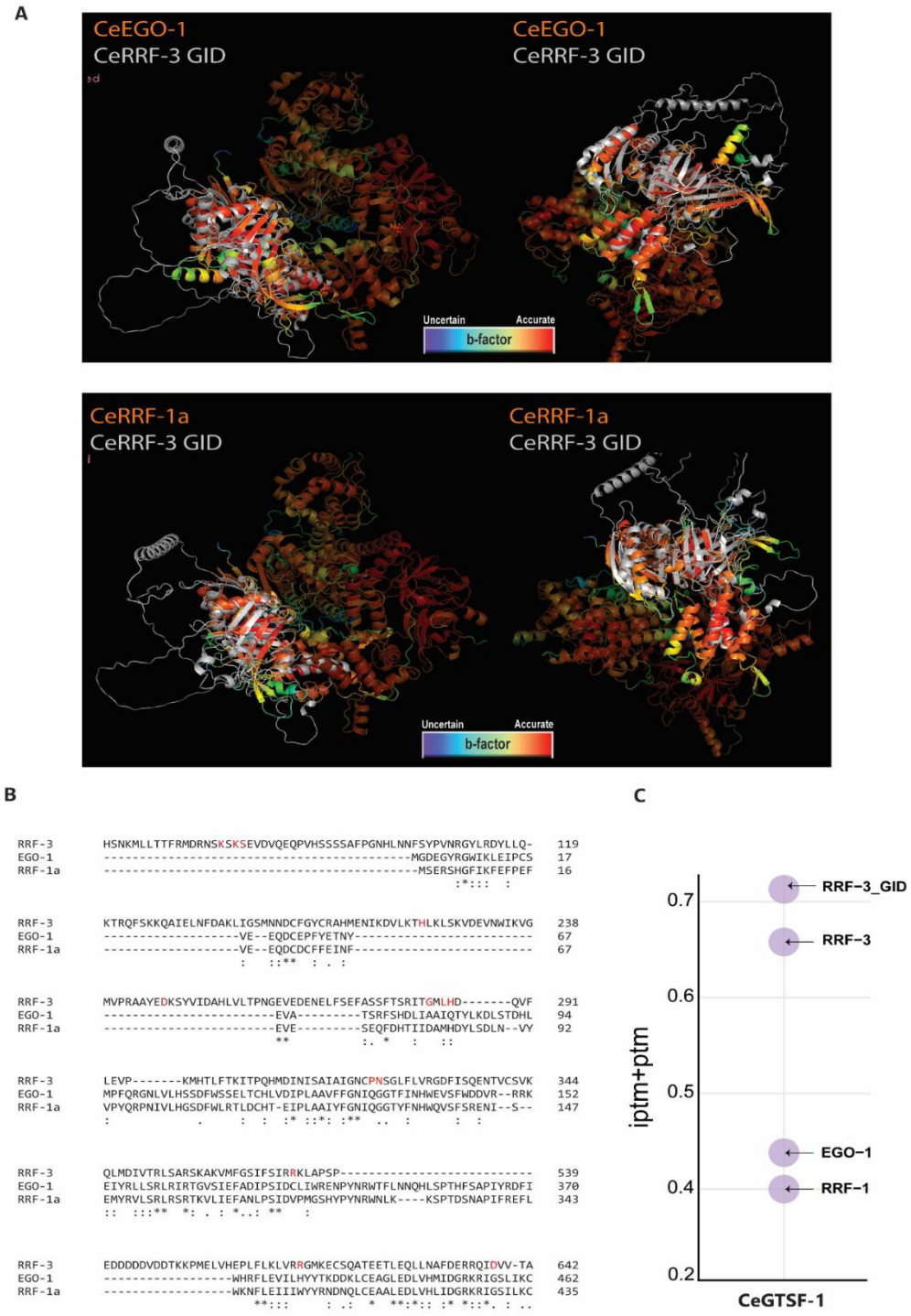


Figure 26 | Specificity of *C. elegans* GTSF-1 towards the RdRP RRF-3: (A) AlphaFold2 (AF2) *in-silico* structural predictions of RRF-1 or EGO-1 aligned with RRF-3 GID. The RRF-1 and EGO-1 models are colored based on AF2 confidence scores with highest confidence in red. Left and right panels are the same structures inverted to view opposite sides. RRF-3 has an IDR in its N terminus that overlaps minimally with GID. **(B)** Multiple Sequence Alignment (MSA) of RRF-3, EGO-1 and the long isoform of RRF-1, RRF-1a generated using ClustalOmega. The residues of RRF-3 that interact with GTSF-1 as per AF2 modelling are highlighted in red. **(C)** AF2 confidence scores (ipTM +pTM) of *C. elegans* GTSF-1 as a multimer with *C. elegans* RdRPs or RRF-3 GID. Multimers with scores below 0.6 are considered to have a low probability of interaction.

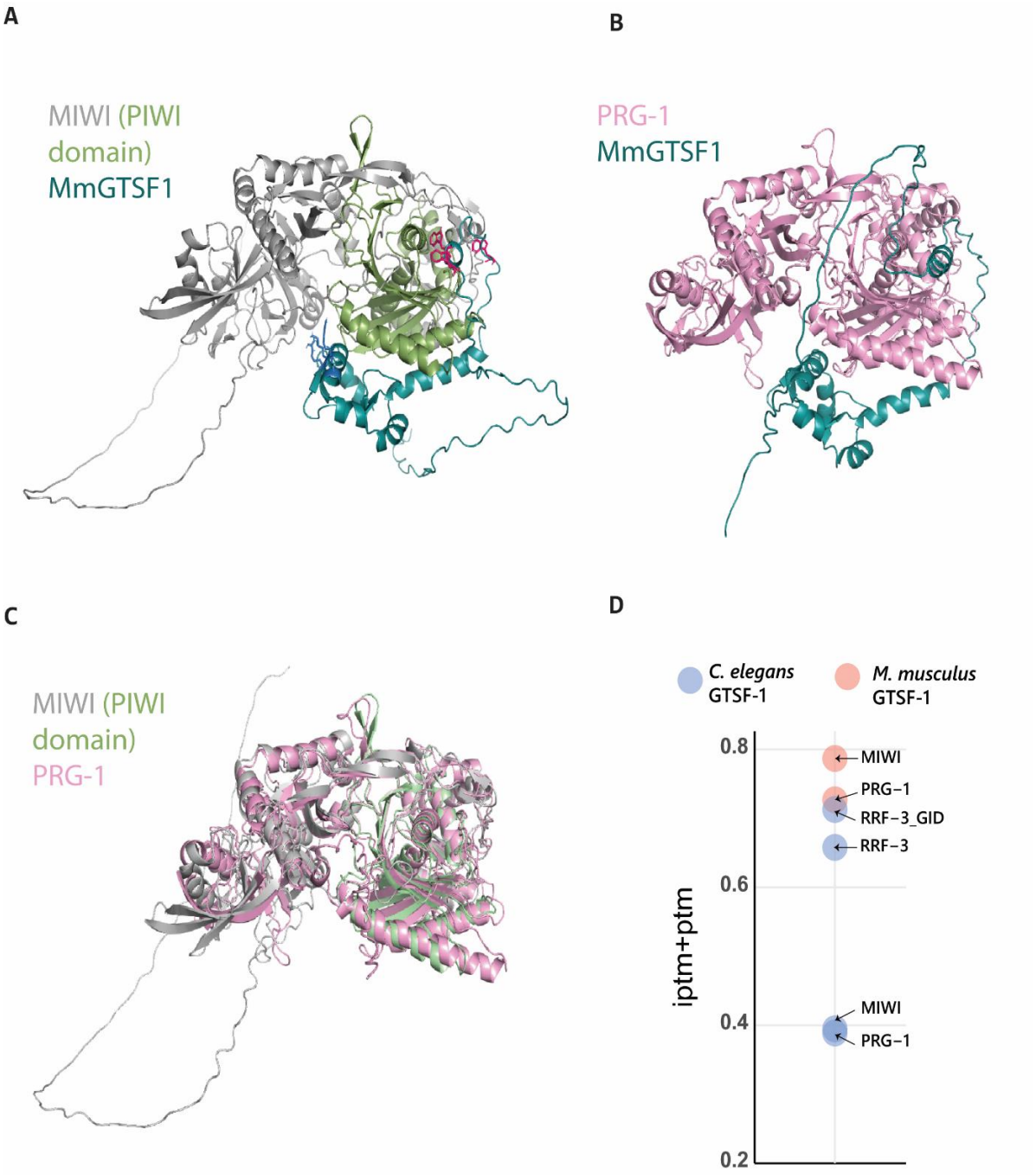


Figure 27 | AF2 predicts MmGTSF1 binding to PRG-1 but not CeGTSF-1: (A) AlphaFold2 (AF2) *in-silico* multimer model of MIWI (*M. musculus* PIWI) with *M. musculus* GTSF1. Residues of GTSF1 that are shown essential for MIWI binding (W98, W107, W112) are displayed as pink sticks while those necessary for RNA binding (R26, R29, K36, K39) are in blue. **(B)** AF2 multimer model of PRG-1 with *M. musculus* GTSF1. **(C)** Alignment of GTSF1 bound MIWI and PRG-1 isolated from structures A and B. **(D)** AF2 confidence scores (ipTM +pTM) of *C. elegans* or *M. musculus* GTSF-1 as a multimer with RdRPs or PIWI. Multimers with scores below 0.6 are considered to have a low probability of interaction.

DISCUSSION

Our research has discovered a consistent and robust interaction of GTSF-1 and the RdRP RRF-3 in *C. briggsae* and *P. pacificus*. Despite the presence of other RdRPs in these nematodes, such as RRF-1, RRF-2, and EGO-1, along with an additional subfamily known as RRF-4 in *C. briggsae* (Shi et al., 2013), the specificity of interaction with RRF-3 remains pronounced.

1. The NTD of RRF-3, labelled as GID associates with GTSF-1

Our Alpha-Fold2 predictions show a distinct region within RRF-3 that appears to fold independently and interacts with GTSF-1 **(Figure 23)**. Termed the GTSF-1 interacting domain (GID), this segment resides in the N-terminus of RRF-3, separate from its RdRP domain. What's crucial to note is that the GID is not only identifiable in RRF-3 from all three nematodes we studied but also aligns perfectly among them, indicating a deep conservation of this domain in clade V nematodes. Subsequently, through experimental validation using yeast two-hybrid (Y2H) assays, we confirmed the binding of GID to GTSF-1 and observed a stable interaction between them **(Figure 24)**.

All three *C. elegans* RdRPs all possess the RdRP core which is flanked by N- and C-terminal domains. Other RdRPs such as RDR2 from maize also feature an N-terminal extension (RRM) domain (Du et al., 2022; Fukudome et al., 2021), but the CTD appears to be specific to nematodes. Extensions of RdRP motifs at both the N and C terminals can interact with accessory factors and provide functional specificity. The three RdRPs have highly similar structures. However, our AF2 prediction of RRF-3 shows presence of a long N-terminal IDR that is absent in RRF-1 and EGO-1 **(Figure 26)**. This IDR may be used to mediate interactions with ERIC.

Our Y2H assay demonstrates that GTSF-1 can directly bind to RRF-3 **(Figure 24)**. However, *in-vivo*, this interaction may rely on additional factors such as ERI-5, as observed for the Tudor domain protein Vreteno in *B. mori* (Bronkhorst et al., 2023, Izumi et al., 2022). Is RNA involved in GTSF-1 and RRF-3 interaction? Almeida et al. (2019) found that RNase treatment does not disrupt the interaction between recombinant GTSF-1 and RRF-3 in embryo extracts. And an iCLIP analysis on GTSF-1 also showed no interaction with RNA. We support these findings by highlighting the absence of positively charged residues in nematode GTSF-1 **(Figure 5)**, which are known to facilitate RNA binding in fly and mouse orthologs (Ipsaro et al., 2021). Further, GTSF-1 did not co-precipitate RNA when purified from Sf9 cells. Thus, we argue that RNA binding by GTSF-1 isn't essential for its interaction with RRF-3. This goes in line with the fact that nematode GTSF-1 doesn't associate with the catalytic, RNA-binding active site of RRF-3 **(Figure 23)** in contrast to PIWI binding orthologs (Arif et al., 2022).

2. Residues around the NTD of GTSF-1 mediate binding to RRF-3

The high homology of various GTSF-1 proteins is attributed to the CHHC Zinc fingers domain, which are highly similar across the orthologs. In contrast, the central and CTD diverge substantially (Arif et al., 2022). Aromatic residues in the CTD are crucial for PIWI binding in mouse and fly orthologs. Through a multiple-sequence alignment we show that nematode GTSF-1 lacks these residues **(Figure 5).** Instead, previous research by Almeida et al. demonstrated the crucial role of the CHHC Zinc fingers of CelGTSF-1 in binding to RRF-3. Mutation of all cysteines to alanines in a GST-fused GTSF-1 construct abrogated the interaction with 3XFLAG::RRF-3. In addition, transgenes expressing the mutated CHHC GTSF-1 failed to immunoprecipitate RRF-3 (Almeida et al., 2019).

Our AF2 prediction models and Y2H assay support these findings. We find the N-terminal domain (NTD) of GTSF-1 to interact with RRF-3 **(Figure 24)**. However, according to AF2 structures, aromatic and positively charged residues surrounding the cysteine residues of CHHC mediate this interaction. It is possible that mutating the CHHC cysteines to alanines, as done in Almeida et al., 2019, alters the conformation of GTSF-1 in a way that prevents our identified residues from binding RRF-3. We have initial evidence to validate these models **(Figure 25)**. Moreover, the fact our observations are conserved in *C. briggsae* and *P. pacificus* strengthens the validity of our results. An important observation we made is that stability of GTSF-1 is directly dependent on interaction with RRF-3, as in RRF-3 deletion GTSF-1 is completely depleted **(Figure 25)**. This highlights the close interaction of the two proteins.

3. The impact of binding by GTSF-1 on RRF-3 activity

Almeida et al. 2019 demonstrated that in *C. elegans*, the absence of GTSF-1 resulted in RRF-3 no longer engaging with most ERI complex factors, except for ERI-5. This shift is evident in size-exclusion chromatography, where RRF-3's elution pattern changes in the absence of GTSF-1, from bimodal to entirely towards a smaller fraction (Almeida et al., 2019). These observations lead to the hypothesis that GTSF-1 is essential for integrating an RRF-3/ERI-5 pre-complex into ERIC, possibly via an RRF-3/ERI-5/GTSF-1 intermediate (Almeida et al., 2019).

But how might GTSF-1 achieve this feat? Investigating its effect on PIWI enzymatic activity in mice might provide insights. The addition of recombinant GTSF1 to recombinant MIWI programmed with an artificial, 30-nt piRNA (piRISC) enhances *pre-steady-state* rate target cleavage (Arif et al., 2022). This rate reflects the pace of piRISC assembly with target mRNA. However, GTSF1 doesn't influence the *steady-state rate* of target cleavage (Arif et al., 2022), which reflects piRISC product release and turnover post-reaction. The study proposes that GTSF1 recognizes the pre-catalytic piRISC, induces a conformational change and possibly stabilizes the catalytically active conformation. In other words, GTSF1 primes piRISC for target mRNA cleavage (Arif et al., 2022).

In *C. elegans* and other nematodes, GTSF-1 might similarly induce a conformational change upon binding to RRF-3 **(Figure 28A)**. This could boost the RdRP activity of RRF-3, enabling dsRNA synthesis from target mRNA. De-novo initiating RdRPs may not necessitate such a conformational change, thus potentially operating independently of GTSF-1. It is possible that, recognition of dsRNA by other ERI-factors, such as DRH-3, could prompt DCR-1 recruitment onto 26G-RNA production site. There are no studies yet testing whether RNA is necessary for ERIC/Dicer assembly on RRF-3. However, Almeida et al., detected the ERIC proteins associated with RRF-3 in the embryos but not in young-adults (Almeida et al., 2019). Assuming that the 26G-RNA pathway is active in embryos but not in young adults, it implies that the engagement of RRF-3 with its targets is necessary for the assembly of ERIC. In this model, loss of GTSF-1 would prevent dsRNA synthesis by RRF-3 and thus prevent ERI/Dicer to co-IP with RRF-3.

Alternatively, GTSF-1 might facilitate protein-protein interactions, bridging RRF-3 to ERIC. With the NTD directly binding RRF-3, the CTD could engage with another ERIC component like DRH-3 or ERI-5, effectively linking RRF-3 to ERIC **(Figure 28B)**.

We've also observed that in *C. briggsae* and *P. pacificus*, GTSF-1 co-immunoprecipitates with RRF-3 and ERI-5, and to a lesser extent with DRH-3. This suggests that the concept of pre-ERIC might be applicable in other nematodes as well. Additionally, the idea of an RdRP module comprising DRH-3, ERI-5, and RRF-3 may hold true in these nematodes, as discussed in more detail later. We haven't yet investigated whether RRF-3 co-immunoprecipitates with ERIC orthologs in these nematodes or whether, similar to *C. elegans*, GTSF-1 is necessary to link RRF-3 to ERIC. To elucidate these further, we are generating strains with epitope tagged RRF-3 and ERI-5.

A third possible role of GTSF-1 arises regarding the GID of RRF-3 potentially being an unidentified endonuclease, with GTSF-1 enhancing RNA cleavage through this domain. Despite identifying several factors vital for 26G-RNA synthesis, their precise functions remain elusive. Some of these factors are ribonucleases associated with ERIC, such as the NYN-domain endoribonucleases (ERI-9, NYN-3, RDE-8) and DCR-1. The current model of 26G-RNA biogenesis is as follows: NYN-domain proteins perform the first cleavage, downstream of 26G-RNA synthesis sites, by detecting target mRNAs, potentially with a stem-loop structure and an AG-rich motif (Tsai et al., 2022). RRF-3 then initiates 3' to 5' dsRNA production from a guanosine near the cleavage site (Tsai et al., 2022). *In-vitro* studies indicate that DCR-1 generates 26 nt endo-siRNAs through helicase-dependent cleavage of a blunt terminus (Welker et al., 2011). However, the identity of the ribonuclease generating the blunt terminus *in-vivo* remains unknown. ERI-1, a 3' to 5' exonuclease, is favored in the current model (Gabel and Ruvkun, 2008). Nonetheless, considering GTSF-1's deeply conserved binding to RRF-3 GID, speculation arises regarding whether RRF-3 GID might serve as an undiscovered endonuclease. This could either directly generate the blunt end post-dsRNA synthesis or perform the initial cleavage instead of/in addition to the NYN-domain proteins. If confirmed, nematode GTSF-1 would still associate with an endonuclease domain, albeit that of RRF-3 instead of PIWI Argonaute **(Figure 28C)**.

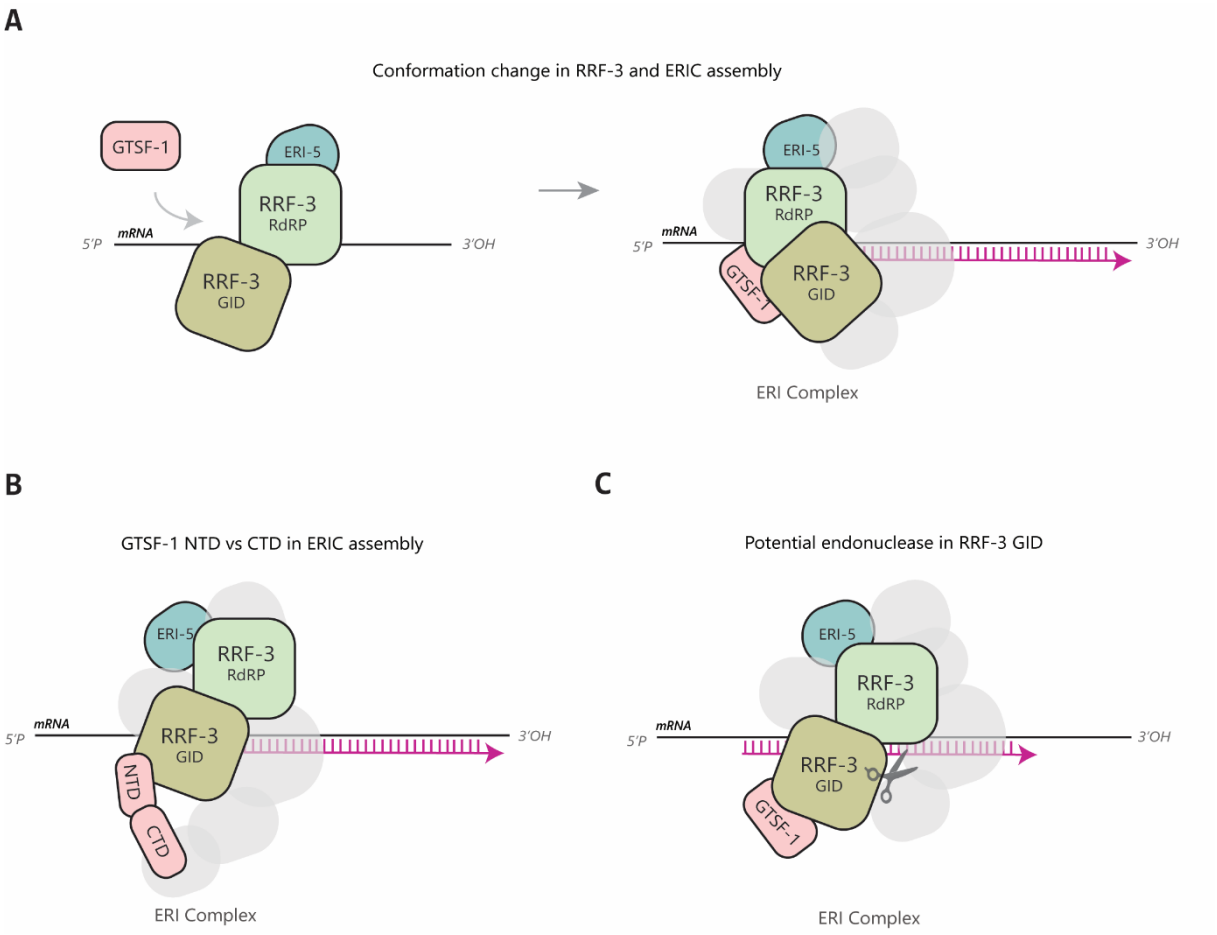


Figure 28 | Model for GTSF-1 function: (A) GTSF-1 triggers a conformational shift upon interaction with RRF-3, enabling the formation of novel surfaces crucial for ERIC assembly**. (B)** GTSF-1 acts as a mediator for protein-protein interactions with the NTD binding RRF-3 and CTD with additional ERIC components. **(C)** The GID of RRF-3 may serve as an unidentified endonuclease, with GTSF-1 potentially enhancing RNA cleavage via this domain.

We argue that, while the binding partner differs, the exact function of GTSF-1 in *C. elegans* might be quite similar to that observed in *D. melanogaster* and *M. musculus*. We propose that binding by this small protein induces a conformational change in its large enzyme cofactor, allowing either catalytic activity or generating new surfaces for assembly of large protein complexes on the enzyme.

4. Towards an ancient association of nematode GTSF-1 with RRF-3

It is known that PIWI proteins and piRNAs are absent in nematodes outside of clade V. These nematodes instead express 5' triphosphate RNAs against transposons (Sarkies et al., 2015). These are likely to be synthesized by RRF-3 as RRF-1, RRF-2, and EGO-1 are only present in clades III–V, while RRF-3-like genes are found in all clades (Sarkies et al., 2015). In addition, RRF-3 homologs from nematodes form a phylogenetic subgroup distinct from RRF-1/RRF-2/EGO-1. At the molecular level, RRF-1 type RdRPs have an insertion of residues forming a loop at the RdRP domain, which are typically absent from eukaryotic RdRPs (Sarkies et al., 2015). Thus, RRF-3 RdRP family likely participates in ancient small-RNA pathways of nematodes, while the RRF-1 family originated in the last common ancestor of clades III-V (Sarkies et al., 2015).

We tested for the presence of 1:1 orthologs of various small-RNA pathway proteins across nematoda **(Figure 29).** Orthologs of GTSF-1 are distributed among several nematodes from all clades. The evalue of BLASTp is low possibly due to high divergence of the CTD. According to this plot, majority of species from Rhabditida (which includes Clades III, IV, and V) mirror the conservation of GTSF-1 and RRF-3. This includes clade IV nematode *Ditylenchus destructor* and clade III nematode *Dracunculus medinensis.* Both species lose RRF-3, GTSF-1 as well as the Argonautes that bind 26G-RNA (ALG-3/4) suggesting loss of the entire 26G-RNA pathway. Some species from Rahbditida break this pattern of conservation. For example, two *Stronglyloides* from Clade IV and *Rhabditophanes kr3021* from Clade III do not encode GTSF-1 but have RRF-3. Outside of clade V, RRF-3 is conserved. On the other hand, an ortholog of GTSF-1 is present in *T. spiralis* from Clade I, but is absent in the two species of *Trichuris* from Clade II. The free living nematode *Plectus sambesii* is the only known species outside of clade V that retain PIWI. Similar to clade V species, *P. sambesii* also encodes GTSF-1 and RRF-3. In the list of nematodes we studied, Clade V *Angiostrongylus cantonensis* stands out as a curious case as it has GTSF-1 but no PIWI or RRF-3 **(Figure 29)**. This species would be an interesting case to target GTSF-1 through IP-MS.

Altogether, these observations suggest that loss of PRG-1 does not necessarily result in loss of GTSF-1. We would argue that this pattern of conservation hints towards a general conservation of GTSF-1 interacting with RRF-3 like RdRPs in nematodes, rather than PIWI. To test this, we will use AF2 multimer modelling to predict binding of GTSF-1 to either RRF-3 or PIWI from representative species in clades I-IV. If GTSF-1 from clade I species, *T. spiralis*, shows a good confidence score as a multimer with RRF-3, we can argue that the switch in binding to an RdRP occurred in the last common ancestor of all nematodes.

We also showed that the in nematodes, differences in GTSF-1 is likely responsible for loss of PIWI binding **(Figure 27)**. Interaction of *M. musculus* GTSF-1 with PRG-1 is predicted with high confidence, unlike *C. elegans* GTSF-1, which fails to show a good interaction score with both MIWI and PRG-1 **(Figure 27)**. Two possible scenarios can explain this observation. First, PIWI loss allowed GTSF-1 to find new binding partners such as RRF-3. However, this is not a parsimonious explanation as PIWI was lost independently at least four times in clades I-IV. It would be unlikely that GTSF-1 would repeatedly acquire binding to RRF-3 binding upon each loss event. The second possibility is that GTSF-1 preferred binding to RRF-3 like RdRPs, even in the presence of PIWI. Perhaps, the evolutionary pressure on Piwi proteins to cleave their targets somehow weakened in the lineage leading to nematodes, leading to GTSF-1 being free to take on another binding partner. Alternatively, it could be simple biochemistry where GTSF-1 has stronger binding

strength to RRF-3 than PRG-1. Unbound to GTSF-1, PRG-1 would likely have reduced endonuclease activity, as shown for several metazoan PIWI homologs (Arif et al., 2022, Izumi et al., 2022). Such a catalytically incompetent enzyme might not have a function and thus be lost in clades I-IV. However, the catalytic residues are present in *C. elegans* PRG-1 although its catalytic activity is dispensable for piRNA mediated silencing (Chi-Lee et al., 2012). Thus, it is possible that Clade V PRG-1 might have acquired additional functions, independent of its catalytic activity. We would like to check to what extent the catalytic residues of PRG-1 homologs conserved across clade V nematodes. Given the confidence in AF2 prediction of MmGTSF1 and PRG-1, an intriguing experiment would be to express MmGTSF-1 in *C. elegans* and test if PRG-1 nuclease can be "activated", leading to a gain of function of PRG-1 activity, possibly associated with phenotypes. Such an experiment would show PRG-1 to be an inefficient enzyme without binding by GTSF1 cofactors and shed light on the evolution of GTSF1 proteins.

Finally, the reason for absence of GTSF-1 in clade II nematodes despite the retention of RRF-3 remains unclear. Small-RNAs sequenced from clade II nematodes, *E. brevis* and *O. rectangular*, failed to reveal sRNAs with 5' triphosphate, in line with the loss of RRF-1/EGO-1. However, they did contain 5' monophosphate of varying sizes, including a bias towards 26nt. Such siRNAs are also found in clade I species *T. spiralis.* This suggests a potential role for RRF-3 from clade I - II nematodes in producing siRNAs with Dicer action. (Sarkies et al., 2015). The most parsimonious explanation would be an independent loss of GTSF-1 from clade II nematodes. Further investigation into the sequence and structure of RRF-3 from clade II nematodes compared to those of clades III-V might provide insight into this discrepancy.

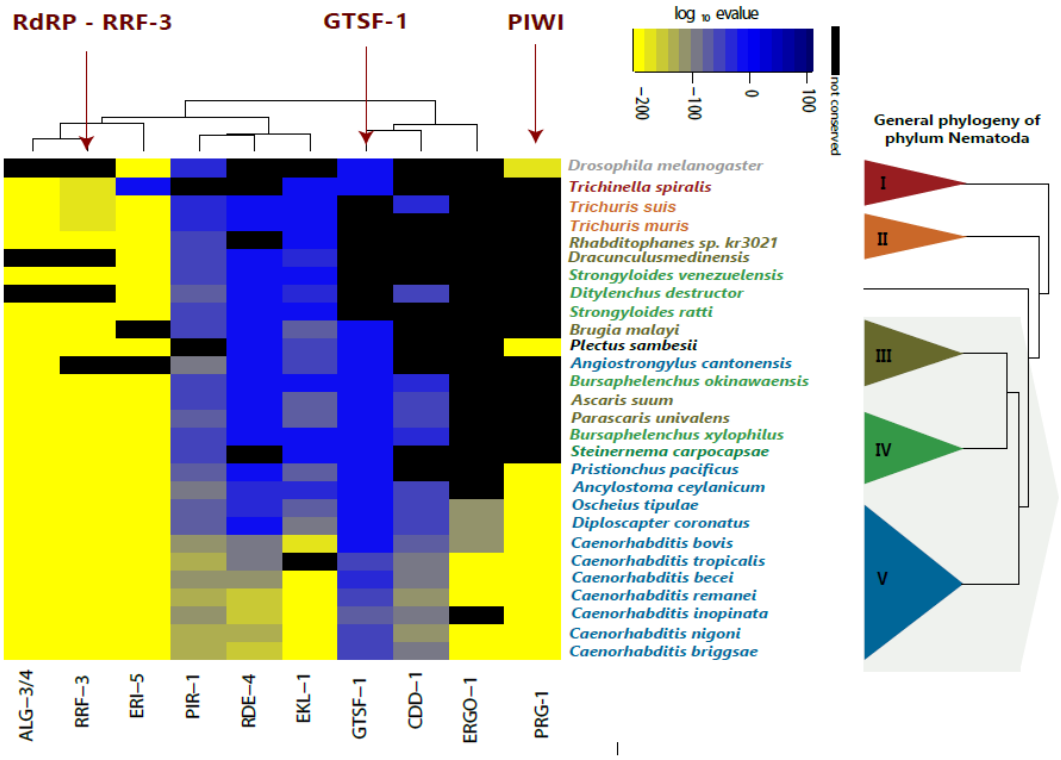


Figure 29 | Conservation of GTSF-1 and RRF-3 like RdRPs compared to PIWI in nematodes: Heat map displaying results of reciprocal BLASTp of selected small-RNA pathway proteins across phylum nematoda. Conservation is calculated as the logarithm of the score of the best blast hit in bits normalized to the length of the protein. The color legend atop indicates e-value, with black denoting no reciprocal BLASTp match. Nematode species are color-coded by phylogenetic position, with the legend on the left panel.

5. The role of DRH-3 and ERI-5 as cofactors of RRF-3

The presence of ERI-5 or DRH-3 as interactors of GTSF-1 in both *C. elegans* (Almeida et al., 2019), *C. briggsae* **(Figure 8, Figure 10A)** and *P. pacificus* **(Figure 21)** prompts us to question whether these interactions are direct or rather indirect, potentially mediated through their shared association with RRF-3. While the yeast-two-hybrid system revealed a direct interaction between GTSF-1 and RRF-3, false positive interactions of ERI-5 prevented us from testing for a direct interaction between ERI-5 and GTSF-1 **(Figure 12; Figure 24)**. Interactions of DRH-3 were not assessed in our experiment. However, DRH-3 and the ERI-5 paralog, EKL-1, participate with RRF-1 and EGO-1 in the 22G-RNA pathway, where GTSF-1 is not implicated (Aoki et al., 2007; Claycomb et al., 2009). Moreover, both EGO-1 and EKL-1 were highly enriched in DRH-3 immunoprecipitation experiments using multidimensional protein identification technology (MudPIT) (Gu et al., 2009). Consequently, it is reasonable to conclude that DRH-3 and ERI-5 act as cofactors of RRF-3.

DRH-3, which is orthologous to Dicer and shares the closest relation to the RIG-I-like receptors (RLRs), exhibits closer interactions with nematode RdRps than Dicer itself. This suggests the potential involvement of DRH-3 in RdRp enzymatic activity. Structurally, DRH-3 possesses a domain architecture like RLRs, with an N-terminal domain (NTD), a DExH/D helicase core, and a C-terminal domain (CTD). Its enzymatic activity has been characterized through various assays, including RNA binding and ATP hydrolysis experiments (Li et al., 2021). This extensive analysis provides compelling evidence supporting DRH-3's specialization as an enzyme tailored for endo-siRNA pathways.

The helicase domain of DRH-3 is crucial for both RNA binding and ATP hydrolysis (Gu et al., 2009; Fitzgerald et al., 2014). Interestingly, the length of the helicase domain is approximately 12-13bp, around half the length of endo-siRNAs. Indeed the N-terminal domain (NTD) of DRH-3 mediates ATP-dependent dimerization on double-stranded RNA (dsRNA), particularly 22–25bp in length (Fitzgerald et al., 2014). In the absence of the NTD, DRH-3 binds 12bp and 25bp dsRNAs equally, but full-length DRH-3 exhibits a stronger affinity for 25bp dsRNAs (Fitzgerald et al., 2014). DRH-3 displays peak ATPase activity when associated with RNAs ranging from 22 to 34bp, particularly with 22bp RNAs, and has a greater preference for 22 or 26nt dsRNAs over dsRNAs of 34, 12, or 18bp (Fitzgerald et al., 2014). These dsRNA have to be 5′-triphosphate to be recognized efficiently by DRH-3, and its CTD mediates this specificity (Li et al., 2021). The functional activation of DRH-3 relies significantly on its RNA binding and ATP hydrolysis status, as well as its interaction with protein partners, which may include members of ERIC.

How might DRH-3 support ERIC activity? While there are no formal studies on this matter, some speculations can be made based on biochemical understanding of DRH-3. Mutations affecting DRH-3 helicase domain result in a depletion of 22G-RNAs specifically at the upstream end of target mRNA (Gu et al, 2009), suggesting DRH-3 to facilitate the propagation of 22G-RNAs by the RdRP along the 3′ to 5′ direction of template mRNA. Indeed, DRH-3 may assemble in the wake of RRF-3 synthesis of dsRNA, dimerize and be activated. Along with a dual Tudor domain protein, DRH-3 may facilitate the translocation of RdRp along the template RNA and may enhance dsRNA synthesis by boosting RdRp processivity (Li et al., 2021). It has recently emerged that the role of RNA helicases may extend beyond merely modifying RNA structure. Instead, it appears they may also play a role in initiating protein-protein interactions and upholding the stability of the RNP complex over a prolonged period in a durable configuration (Owttrim, 2013). DRH-3 may also contribute to stability of ERIC to sustain long dsRNA synthesis.

The dual Tudor domain protein ERI-5 likely plays a structural role, potentially linking RdRP to the helicase function. ERI-5 is deeply conserved outside in nematodes, while its paralog EKL-1 is restricted to *Caenorhabditis* species (**Figure 29)**. In *C. elegans*, the two proteins preferentially bind to two different RdRPs (Thivierge et al., 2012). ERI-5 associates with RRF-3 and EKL-1 is associated with RRF-1 and EGO-1. It has also been shown that the two paralogs can compensate

for each others function (Thivierge et al., 2012). Thus, EKL-1 appears to be a gene duplication of ERI-5, which specialized to binding the RRF-1/EGO-1 class of RdRPs. As *P. pacificus* does not encode an ortholog of EKL-1 (the low evalues of BLASTp are likely reporting a match with ERI-5), it is possible that *P. pacificus* ERI-5 associates with all RdRPs, including RRF-1/EGO-1.

In summary, our results support the role of DRH-3 in RRF-3 and 26G-RNA pathway in *C. elegans* and *P. pacificus*. Its absence in *C. briggsae* GTSF-1 IPs might be due to weaker interactions with GTSF-1, and we would need RRF-3 IP-MS analysis to conclude this. We propose that in clade V nematodes, RRF-3 is part of a module, alongside a dual Tudor domain protein, a Dicer-related helicase, and a dual Zn finger protein GTSF-1.

6. Comparison of eukaryotic RdRP complexes

Eukaryotic RDRs are ubiquitous across protists, plants, fungi, and animals, where they play crucial roles in RNAi pathways. Numerous studies have been conducted to unveil proteins closely linked with these RdRPs.

In the ciliate *T. thermophiles*, the RdRp Rdr1 forms three distinct RNA-dependent RNA polymerase complexes (RDRCs). Rdr1 exhibits tight association with either ribonucleotidyl transferase Rdn1 or Rdn2, along with one of two additional subunits, Rdf1 or Rdf2. These additional subunits are small proteins lacking identifiable domains. When coupled with Dicer, these RDRCs produce approximately 24-nucleotide-long siRNAs from double-stranded RNA transcribed from non-coding RNA transcripts (Lee et al., 2009). A comparable RdRP complex, also known as RDRC, exists in the fission yeast *S. pombe*, wherein the RdRP Rdp1 directly associates with RNA helicase Hrr1 and polyA polymerase Cid12 (Motamedi et al., 2004). Cid12, a member of a protein family possessing a nucleotidyltransferase domain is similar to *T. thermophiles*' Rdn1 and Rdn2. Furthermore, Hrr1 is a DEAD box RNA helicase from the Smg2/Upf1 family, a domain also found in *C. elegans* ZNFX-1, SMG-2, and ERI-7.

In *S. pombe*, Hrr-1 acts as a bridge between RDRC and the RNA-induced transcriptional silencing (RITS) complex, consisting of Argonaute Ago1 and methyltransferase Clr4. RDRC, in conjunction with Dicer, generates siRNAs for loading onto Ago1 with RITS (Lee et al., 2009). Interestingly, a protein structurally similar to GTSF-1, called Stc1, bridges Ago1 to Clr4. Two tandem zinc finger domains in NTD of Stc1 binding Ago1, while its negatively charged and unstructured CTD interacts with Clr4 (He et al., 2013).

In contrast to these models, RdRP from the fungi *N. crassa* QDE-1 also contains a DNA-dependent RNA polymerase (DdRP) domain. Subsequently QDE-1 cofactors are Replication Protein A (RPA), a single-stranded DNA-binding complex and RecQ DNA helicase QDE-3 (Liu et al., 2010). In plants, *A.* thaliana's RdRP RDR6 interacts with at least two plant-specific proteins, SDE5 and SGS3. SDE5, classified as a putative nuclear export factor, lacks identifiable domains, whereas SGS3 facilitates liquid-liquid phase separation to promote RDR6 activity within condensates (Tan et al., 2023).

Nucleotidyl-transferases frequently associated with RdRP complexes may serve to initiate dsRNA synthesis. *In-vitro* analysis of *T. thermophiles* Rdr1 show that the uridyltransferases associated with RDRC tail the template mRNA at the 3' end with polyU. This stabilizes the looping of template 3' end to provide the hydroxyl group for RdR1 to synthesize dsRNA (Talsky et al., 2010). In *C. elegans,* RDE-3 or MUT-2 performs similar pUG tailing and marks mRNA as templates for RRF-1/EGO-1 activity (Shukla et al., 2020).

No nucleotidyl-transferases were identified in any of our immunoprecipitations, including those conducted for RRF-3 in *C. elegans* (Almeida et al., 2019). It is possible that the initiation of dsRNA synthesis by RRF-3 is probably not reliant on polyUG like tailing, emphasizing another distinction

from the RRF-1/EGO-1 family of RdRPs. We can speculate if this could be a reason specifically RRF-3 needs GTSF-1; that is to be an RdRP independent of nucleotidyl-transferase activity.

7. The 26G-RNA pathway in *C. briggsae* and *P. pacificus*

Our understanding of nematode small RNA pathways comes from detailed studies in *C. elegans*. However, the phylum nematoda has species from diverse lifestyles and ecosystems and not much is understood about the conservation and functional role of small RNA pathways in these divergent nematodes.

Our results show that GTSF-1 is crucial for fertility for both *C. briggsae* and *P. pacificus* (**Figure 6)**. In *C. elegans*, the temperature sensitive sterility is attributed to loss of 26G-RNAs which causes defects in the male germline. Mutations in ALG-3/4, the argonautes responsible for loading 26G-RNAs specifically in the male germline, result in decreased spermatid numbers and abnormalities in meiotic cell division at elevated temperatures (Conine et al., 2010). We have not yet assessed the precise impact of GTSF-1 on male gamete quantity and quality in *C. briggsae* and *P. pacificus*. Nevertheless, considering the reliance on RRF-3 and the significant similarity of 26G-RNA targets in the L4 stage **(Figure 18)**, it is probable that the downstream mechanisms are conserved at this developmental stage as well.

We show that both *C. briggsae* and *P. pacificus* express 21U RNAs, 22G RNAs and 26G-RNAs similar to those seen in *C. elegans* **(Figure 7)**. These small-RNAs are expressed both during spermatogenesis (L4 stage) and oogenesis (GA stage). The majority of small-RNAs are 5' triphosphorylated 22G RNAs. The read-per-million (RPM) values for individual sRNA populations match those observed in *C. briggsae* (Shi et al., 2013, Pastore et al., 2022, Fusca et al., 2022), while similar analyses for *P. pacificus* have not been conducted before. Previous attempts to identify 26G-RNA target genes in *C. briggsae* used a stringent cutoff of > 10 RPM of 26G-RNAs per gene (Shi et al., 2013). Using this threshold, the number of identified targets was low: the *C. elegans* embryo small RNA library identified 29 26G siRNA-producing genes, 121 such genes in *C. briggsae* embryos, 71 in C. remanei, and 272 in C. brenneri (Shi et al., 2013). In our study, we used the specificity of GTSF-1 to deplete 26G-RNAs and identify a comprehensive list of 26G-RNA targets in *C. briggsae*. We are currently extending this analysis for *P. pacificus*.

In *C. elegans*, it is established that embryonic 26G-RNAs target several duplicated genes, pseudogenes, or genes that typically possess fewer and less efficient introns (Fischer et al., 2011; Vasale et al., 2010). Besides, these genes are poorly conserved, even in other *Caenorhabditis* species (Shi et al., 2013). We observe a similar pattern in our comparison of 26G-RNA target genes in Gravid-adult *C. elegans* and *C. briggsae*, a developmental stage that consists of mostly oocytes and embryos **(Figure 18)**. In our analysis, only 20%–30% of target genes produce 26G-RNAs in other species. In addition, there is low correlation in 26G-RNA levels among orthologous target genes, indicating a poor conservation of 26G-RNA targeting strength **(Figure 18)**. We have not specifically examined the *C. briggsae* gene database for duplicated genes or the number of introns, so we cannot directly compare the gene features seen in *C. elegans*. Nevertheless, our study did observe a higher proportion of pseudogenes targeted in the GA stage compared to the L4 stage **(Figure 18)**.

The embryonic 26G-RNAs are loaded onto the Argonaute ERGO-1. The importance of this Argonaute is not clear. Argonautes encompass an endonuclease domain and in the case of *C. elegans* ERGO-1, the catalytic residues are conserved. In fact, the slicer activity of ERGO-1 has been shown to resolve siRNA duplexes downstream of DCR-1 processing (Fischer et al., 2011). However, absence of ERGO-1 does not directly affect the fertility of *C. elegans* (Vasale et al., 2010). The only phenotype displayed by these animals is enhanced response to RNAi (Eri) which is attributed to increased availability of Dicer for exogenous RNAi pathways (Simmer et al., 2002).

It would be interesting to check whether *C. briggsae* and *P. pacificus* GTSF-1 deletion animals also exhibit an Eri phenotype.

The conservation of ERGO-1 is noteworthy, with certain nematodes like *C. briggsae* and *C. remanei possessing* multiple copies, while the majority, including *P. pacificus*, have completely lost this gene (Shi et al., 2013) **(Figure 29).** In some nematodes, such as *A. summ*, CSR-1 bound 22G-RNAs take over regulation of gene expression in the embryos (Wang et al., 2011; Zagoskin et al., 2022). Hence, it is possible that nematodes without ERGO-1 might have other WAGO 22G-RNA pathway to compensate. In our analysis, we observe a larger amount of 26G-RNAs in *C. briggsae* compared to *C. elegans*, specifically in the Gravid-Adult samples **(Figure 18)**. This might be a consequence of multiple paralogs of ERGO-1 expressed in *C. briggsae*. Interestingly, despite the absence of ERGO-1 in *P. pacificus*, our sRNA sequencing attempt from *P. pacificus* embryos did identify a sizeable level of 26G-RNAs **(Figure 7)**. It would be interesting to check which Argonaute these are bound to and what the target genes are.

Given the substantial presence of pseudogenes, duplicated genes, and the limited conservation of targeting, ERGO-1 has been suggested as a regulator of duplicated genes. Gene duplications are notably frequent in *C. elegans* (Katju & Lynch 2003; Lipinski et al., 2011). Exploring the variations in the frequency and scope of gene duplications between nematodes with and without an orthologue of ERGO-1 could offer intriguing insights.

In contrast to the Gravid-Adult samples, our examination of 26G-RNA pathway in L4 stage reveals a remarkable conservation between *C. elegans* and *C. briggsae* **(Figure 20)**. Both species demonstrate similar levels of 26G-RNAs and possess a comparable number of target genes. This observation aligns with previous findings in sRNA sequencing studies conducted on male Caenorhabditis species (Shi et al., 2013). However, the most notable discovery from our analysis is the consistent targeting of orthologous genes to a similar extent **(Figure 20)**. This suggests a high degree of conservation in the 26G-RNA pathway during the L4 stage, particularly significant for spermatogenesis. Supporting this notion, ALG-3/4 orthologs, crucial components of the pathway, are deeply conserved across nematoda **(Figure 29)**

Further insights into the conservation and functionality of the 26G-RNA pathway are provided by a comprehensive study in *A. suum* (Wang et al., 2011; Zagoskin et al., 2022). ALG-4 RIP-seq analysis revealed its binding to 26G-RNAs, with AsALG-4 expression being specific to spermatogenesis. This expression pattern correlates with the expression of the RRF-3 ortholog RdRP-3. Additionally, there is an expansion of ALG AGO clade Argonautes closely related to ALG-4, all of which are expressed during spermatogenesis (Wang et al., 2011; Zagoskin et al., 2022). Overall, the identification of 26G-RNAs and their targets in *C. briggsae* and *P. pacificus* opens avenues for understanding the regulatory mechanisms underlying gene expression in these nematode species.

8. Explaining the curious link between PETISCO complex and the 26G-RNA pathway in *C. briggsae*

In our study, we describe a compelling association between GTSF-1 and the PETISCO complex in *C. briggsae*. Through IP-MS targeting *C. briggsae* GTSF-1, we consistently observed enrichment of several components of the PETISCO complex, in both Gravid-Adult **(Figure 8)** and L4 stages **(Figure 10)**. We also occasionally detected TOST-1 in our IPs. However, PID-1 was conspicuously absent. This suggests that TOST-1 bound PETISCO preferentially interacts with GTSF-1.

Our results from IP-MS on *C. briggsae* PID-3 and yeast- two-hybrid (Y2H) show that PETISCO is a stable complex with an architecture similar to *C. elegans* **(Figure 12; Figure 13)**. We also show that depletion of *pid-3* and *tofu-6* in *C. briggsae* through RNAi causes a Mel phenotype and loss of

21U-RNAs **(Figure 13)**. These results argue towards a deeply conserved function of the PETISCO complex in *Caenorhabditis* species.

To explain the link between GTSF-1 and PETISCO, we developed several hypotheses. First we questioned if GTSF-1, ERI-5 or RRF-3 interact with PETISCO directly. While PETISCO subunits are enriched in GTSF-1 IPs, no ERIC component is present in PID-3 IP. Additionally, Y2H failed to find any positive interactions between GTSF-1 and PETISCO subunits **(Figure 12; Figure 13)**. Thus, GTSF-1 is not a component of *C. briggsae* PETISCO. However, we cannot rule out any transient interactions of GTSF-1 with PETISCO *in vivo*. We could not test ERI-5 in Y2H assay, as it displays false positive interactions. However, the enrichment of PETISCO in GTSF-1 IPs persisted upon deletion of ERI-5 **(Figure 10)**. Thus, it is unlikely that ERI-5 is mediating this link.

Alternatively, does PETISCO act upstream, downstream or parallel to ERIC? Are these two complexes assembling on the same transcript? PETISCO is an RNA binding complex that appears to recognize unusual transcripts such as precursors of 21U RNAs and histone mRNAs (J. Pereirinha, unpublished). Both of these are clustered, multi-copy genes in the genome and lack introns. Most *C. elegans* transcripts are trans-spliced, where mRNAs contain a splice leader sequence with a 2,2,7-trimethylguanosine-containing cap. However, both 21U-RNA precursors and histone mRNAs lack trans-splicing (J. Pereirinha, unpublished). Lack of splicing has been implicated as a recognition factor for PETISCO binding (J. Pereirinha, unpublished).

What effect could PETISCO have on 26G-RNA pathway? Interestingly, 26G-RNA target genes have fewer introns and often these introns lack strong consensus splice sites, making them inefficient to splice (Newman et al., 2018). In this light, it would be a tempting hypothesis that *C. briggsae* PETISCO recognizes 26G-RNA targets transcripts. What would be the consequence of this recognition? Perhaps, PETISCO could recruit ERIC for 26G-RNA synthesis. Surprisingly, our results show a negative regulatory role of PETISCO on 26G-RNA levels. Animals with *pid-3* and *tofu-6* depletion by RNAi show moderate increase of 26G-RNAs compared to control **(Figure 13E)**. Thus, a more appropriate explanation is that recognition by PETISCO blocks 26G-RNA synthesis by ERIC.

We have not tested if RRF-3 is the link between ERIC and PETISCO. If RRF-3 is somehow sequestered by PETISCO, loss of PETISCO could free up RRF-3 and cause the moderate increase of 26G-RNAs that we observe. However, RRF-3 levels are likely perturbed upon ERI-5 deletion (as observed in *C. elegans* by Thivierge et al., 2012). As mentioned earlier, this strain continued to IP PETISCO subunits with GTSF-1 while no RRF-3 was observed in the IP. We would use this result to also rule out RRF-3 as the link.

To address concerns regarding technical artifacts, we performed an IP for GTSF-1 in the GTSF-1 deletion strain **(Figure 13F).** Surprisingly, we detected the bait protein. While no members of ERIC were detected, we did see an enrichment of PETISCO subunits. We attribute this to the affinity purified antibody against GTSF-1, which may have low levels of recombinant GTSF-1 bound to it. Due to this, we cannot exclude that the enrichment of PETISCO is due to non-specific interactions with the antibody. In addition, the RNAi treatment to deplete *pid-3* and *tofu-6* caused unsynchronized growth. Despite the intention to harvest gravid-adults, presence of L4 animals could have influenced the composition of harvested samples, possibly contributing to observed variations in 26G-RNA levels **(Figure 13E).**

In light of these findings, we would need an epitope tagged GTSF-1 in *C. briggsae*, which would allow us to repeat the IP-MS with a different antibody. In addition, we would need to repeat the small-RNA sequencing, by hand-picking gravid-adult animals from RNAi treated plates to avoid sequencing L4s. Such results would more conclusively probe the possible connection between PETISCO and the 26G-RNA pathway.

9. Conclusions

In summary, our study sheds light on the conservation and functional significance of small RNA (sRNA) pathways in clade V nematodes. Specifically, we conducted a comparative analysis of the 26G-RNA pathway in *C. briggsae* and *P. pacificus*. We show the deep conservation of GTSF-1 in RRF-3 activity and the 26G-RNA pathway in these species, as well as provide arguments for this association being present in ancient nematodes. We discuss conserved functions of the PETISCO complex in *C. briggsae* embryogenesis and 21U-RNA biogenesis. Our results linking PETISCO to ERIC might uncover diversified functions of more sRNA pathway proteins. We also show conservation of ERIC proteins in *P. pacificus* as well as identify novel, genus specific, uncharacterized factors. Our comprehensive transcriptomic and proteomic datasets will benefit further explorations of sRNA pathways in *C. briggsae* and *P. pacificus*.

Recent unpublished work has shown presence of GTSF-1 orthologs in the planaria *Schmidtea mediterranea* as well as the ciliates *Tetrahymena thermophilis* and *Paramecium tetraurelia.* In each species, GTSF-1 was discovered as interactor of PIWI, placing GTSF-1 proteins as core PIWI cofactors. These findings increase the relevance of our results, making the phylum nematoda the only known system where GTSF-1 is repurposed to bind an RdRP. In other words, our work highlights the functional adaptation of GTSF-1 proteins in nematodes and provides a compelling explanation for the loss of the highly conserved piRNA pathway from the majority of nematodes.

MATERIALS AND METHODS

Nematode growth and culture

The handling of *C. briggsae* and *P. pacificus* culture closely paralleled that of *C. elegans*, with general features described by Brenner, 1974 and Wood, 1988. The nematode stocks were maintained on 5 cm NG agar plates seeded with OPSO, a uracil-requiring mutant of *E. coli* (Brenner, 1974; Wood, 1988). Nematodes for daily growth were stored at 20°C, while long-term cultures were incubated at 15°C. To sustain cultures, several hermaphrodites were transferred to new OP50 plates regularly.

The wild-type strains employed in this research are the *C. elegans* N2 from Bristol, *C. briggsae* AF16 from Ahmedabad, India (Nigon and Dougherty, 1949), and *P. pacificus* PS312 from Pasadena, California (Sommer *et al.*, 1996). AF16 was obtained from Caenorhabditis Genetics Center (CGC), and PS312 was a generous gift from the Sommer Lab. For *P. pacificus* males, cultures were obtained by seeding plates with several hundred animals until bacterial depletion, followed by transfer to a new plate (Sommer et al., 1996). Details about the strains utilized and created in this study can be found in **Table 1.**

Genome Editing

The *C. elegans* strains *gtsf-1* (*xf251*) and *gtsf*-1 (*xf264*) were developed through the plasmid injection method. Protospacer sequences were identified using CRISPOR (http://crispor.tefor.net) (Haeussler *et al.*, 2016) and inserted into plasmids expressing Cas9 + sgRNA(F+E) or plasmids expressing sgRNA(F+E) with Cas9 deletion, utilizing site-directed, ligase-independent mutagenesis (SLIM) (Chiu *et al.*, 2004). For other strains in *C. elegans* as well as *C. briggsae*, the Cas9 protein injection approach was employed (Paix, Folkmann and Seydoux, 2017), and Cas9-NLS was produced by the IMB core facility. Alt-R CRISPR-Cas9 tracrRNA and Alt-R CRISPR-Cas9 Custom Guide RNAs were procured from Integrated DNA Technologies™ (IDT), with crRNA protospacer sequences determined using the Predesigned Alt-R™ CRISPR-Cas9 guide RNA tool **(Table 2).** In both methods, the *dpy-10* gene was co-targeted as a control.

For the integration of the CbrGTSF-1::GFP::3xFLAG sequence for *gtsf-1* (*xf264*), PCR products served as linear, double-stranded DNA donor templates. In other epitope tag insertions, co-conversions, and precise deletions, 4 nmole Ultramer® DNA oligodeoxynucleotides from Integrated DNA Technologies™ (IDT) were ordered as linear, single-stranded DNA (ssODN) donor templates.

The plasmid injection mix comprised 50 ng/uL each of plasmids encoding Cas9, sgRNA(F+E) against *gtsf-1* and *dpy-10*, 750 nM ssODN repair oligo and 400 ng/uL PCR product. For protein injection mixes, all RNAs were reconstituted in 5 mM Tris-HCl pH 7.5, achieving a final concentration of 200 µM for both crRNA and tracrRNA. The other components namely KCl, HEPES pH 7.4, tracrRNA, dpy-10 crRNA, targeted gene crRNA, and dpy-10 ssODN, was added sequentially to Cas9 protein, followed by ssODNA or PCR repair templates (Paix et al., 2017). This mix was incubated at 37 °C for 10–15 min before use (Paix et al., 2017).

The injection mixes were injected into both gonad arms of young adult N2 hermaphrodites maintained at 20°C. F1 progeny were screened for insertions or deletions by PCR, and successful editing events were confirmed by Sanger sequencing. All generated mutant strains underwent at

least two out-crosses before any subsequent cross or analysis. The Sommer Lab at MPI Tubingen generated all strains in *P. pacificus* according to (Witte et al., 2015).

Table 1 | List of CRISPR/Cas9-generated alleles:

RFK	Species	Notes	Genotype
RFK1317	*C. elegans*	*gtsf-1 deletion*	*gtsf-1(xf251[d10(replacement of A2 - L169 with d10)]) IV*
RFK1318	*C. elegans*	*Cbr-gtsf-1::gfp::flag expressed in C. elegans*	*gtsf-1(xf264[(replacement of endogenous Cel-gtsf-1 with Cbr-gtsf-1 (spliced)::gfp::3x flag)]) IV*
RFK1425	*C. elegans*	*Cbr-gtsf-1::gfp::flagx flag::rrf-3*	*gtsf-1 (xf264) IV; xfIs63[gld-1(prm)::flag::rrf-3::unc-54 (3'UTR)] IV*
RFK1493	*C. elegans*	*Cbr-gtsf-1::gfp::flag x pgl rfp*	*gtsf-1 (xf264) IV; pgl-1(xf233[pgl-1::mTagRFP-T]) IV*
RFK1730	*C. elegans*	*gtsf-1 W39A_R42A_HindIII*	*gtsf-1 (xf373[(W39A;R42A;synonymous mutation to add HindIII cleavage site)]) IV*
RFK1731	*C. elegans*	*gtsf-1 R63A_BstAPI*	*gtsf-1 (xf373[(R63A;synonymous mutation to add BstAPI cleavage site)]) IV*
RFK1732	*C. elegans*	*gtsf-1 CTD replaced by 3xflag_stop*	*gtsf-1 (xf373[(replacement of E101-L169 with 3XFLAG)]) IV*
RFK18	*C. elegans*	*rrf-3 Δ*	*gtsf-1 (pk1426)*
RFK837	*C.briggsae*	Wild-Type	Wild-Type AF16
RFK948	*C.briggsae*	*gtsf-1 ΔR30 & A31*	*gtsf-1(xf207[ΔR30;ΔA31]) IV*
RFK1124	*C.briggsae*	*express Cel-sid2 in C.briggsae for RNAi*	*mfIs42[Cel-sid-2 + Cel-myo-2::DsRed]*
RFK1609	*C.briggsae*	*gtsf-1 Δ*	*gtsf-1(xf345) IV*
RFK1610	*C.briggsae*	*gtsf-1 ΔR30 & A31_NdeI*	*gtsf-1(xf346[ΔR30;ΔA31;synonymous mutations to add NdeI cleavage site]) IV*
RFK1691	*C.briggsae*	*eri-5 Δ*	*eri-5 (xf357)*
RFK949	*P.pacificus*	*gtsf-1 deletion*	*gtsf-1(xf208)*
RFK838	*P.pacificus*	Wild-Type	Wild-Type PS312

Table 2 | List of sgRNA sequences used for CRISRPR-Cas9:

Gene	Species	sgRNA sequence	Allele Generated
cbr-dpy-1	*C. briggsae*	TGCTGATCATTGTGACTGA	co-crispr
cbr-dpy-1	*C. briggsae*	CTATCGCCACTGATGTTGG	co-crispr
cbr-gtsf-1	*C. briggsae*	GTGCTCCGCTCTATAAGGAC	*Cbr-gtsf-1 (xf345)*
cbr-gtsf-1	*C. briggsae*	AAACGATGAATCGGTGTCTT	*Cbr-gtsf-1 (xf345)*
Cbr-gtsf-1	*C. briggsae*	CATCACCTGTCCTTATAGAG	*Cbr-gtsf-1 (xf346)*
Cbr-eri-5	*C. briggsae*	TTCCACTTCTAGAAGCCTGT	*Cbr-eri-5 (xf357)*
Cbr-eri-5	*C. briggsae*	TTCGATATCCACTTGACAAG	*Cbr-eri-5 (xf357)*
Cel-gtsf-1	*C. elegans*	AGGAGTTCAACACGCACTTG	*Cel-gtsf-1 (xf373)*
Cel-gtsf-1	*C. elegans*	AAGGAAGTGACGCATGTTGT	*Cel-gtsf-1 (xf374)*
Cel-gtsf-1	*C. elegans*	GAACGTGGCGGAGCACCTGG	*Cel-gtsf-1 (xf375)*
Cel-gtsf-1	*C. elegans*	CGGATTCAACGCGCACGTC	*Cel-gtsf-1(xf264)*
Cel-gtsf-1	*C. elegans*	TCTTCTGAAAACGATGAAT	*Cel-gtsf-1(xf264)*

RNA interference

RNAi experiments were conducted as previously described (Fire *et al.*, 1998; Timmons, Court and Fire, 2001). For *C. elegans dpy-13* RNAi, animals were synchronized by bleaching. L1 worms were seeded into six-well plates and scored for dumpy phenotype on the second day of adulthood.

For *C. briggsae*, the RNAi competent strain JU1018 was used. A 100-200bp sequence from one exon of each target gene was amplified using polymerase chain reaction (PCR), with the sequence of T7 promoter at primer ends. The PCR products were cloned into L4440 plasmid using Gibson assembly (Gibson *et al.*, 2009) and transformed into ht115 *E. coli.* Exon sequences used for cloning can be found in **Table 3.** NGM plates were supplemented with IPTG to induce expression of dsRNA synthesis and seeded with the transformed *E. coli* (RNAi plates).

Nematodes were cultured on agar plates seeded by dsRNA expressing bacteria and either scored for fertility or harvested for small-RNA sequencing. *C. briggsae* JU1018 animals cultured on OP50 plates were bleached, hatched in M9 overnight and the L1 larvae were seeded onto RNAi plates. To assess maternal effect lethal (Mel) phenotype, a small number of synchronized L1 larvae were seeded onto 2cm RNAi plates and removed once a sufficient quantity of eggs had been laid. The number of eggs laid was counted, and 48 hours later, the count of both larvae and eggs was recorded. Embryonic lethality was calculated as the ratio of the number of hatched eggs to total eggs laid. For small-RNA sequencing, synchronized L1 larvae were seeded onto appropriate RNAi plates and grown until adulthood. Animals were then collected off the plates using standard M9 washes and snap-frozen in liquid Nitrogen before RNA extraction.

Table 3 | Exon sequences used for RNAi by feeding

Target gene	Sequence of dsRNA
Cbr-pid-3	GAAAAACGGGCTGATTTGAAAAAGTTTGCTGGATGGTTTATGGATACATTGAGAA TGGAAGGTTCTTTCATCGGACATTACTTCAATTACGATACCGCACCGGTGACAATC GTTGAAACTGAGCCAAACAATCTTGAATCATGCACCAATGCGTATCAACAAATTCA CAAGGATCATCCG
Cbr-tofu-6	TCGTCAGCAGCTAAGGATGTTTTCGCTCTAACGATTCTCCGTATCAACGATGCACA GTCAAATGAGAAGTACAATGCGATGCATGAAAAAATGAATTCGTATGCACAACTA GTGGCTTTCGATTCGGAGCTTCAGATCGGTTACGATGGTGTGTTCCGAGAGTCG

Brood size calculation

Nematodes strains were synchronized by bleaching. Synchronized L1 larvae were transferred to test temperature and singled when L4 stage was reached. At the onset of egg laying, animals were transferred to new plate every day until egg laying stopped. Larvae were counted two days after egg hatching.

Yeast-Two-Hybrid

The gene of interests were amplified from cDNA and cloned into the pGAD and pGBD plasmids as described in James, Halladay and Craig, 1996. All plasmids were transformed into haploid

Saccharomyces cerevisiae strain AH109 (James, Halladay and Craig, 1996) using the high efficiency LiAC method of Daniel Gietz and Schiestl, 1995). Invitrogen UltraPure™ Salmon Sperm DNA Solution (Catalog number: 15632011) was denatured at 95˚C for 5 minutes before use in transformation mix. Transformants were then cultured on agar plates supplemented with Adenine (ADE) and Histidine (HIS). For selection of protein interactions, transformants were further cultured at concentrations of 5×10^6 cells/ml and 5×10^5 cells/ml, on plates lacking either only ADE or both ADE and HIS.

For the preparation of all plates, Sigma Aldrich Yeast Synthetic Drop-out Medium Supplements without histidine, leucine, tryptophan, and adenine (Y2021) and Yeast Nitrogen Base Without Amino Acids (Y0626) were utilized at final concentrations of 1.399 g/L and 6.7 g/L, respectively. Sigma Aldrich Adenine (A2786) and L-Histidine monohydrochloride monohydrate (H5659) were added to selection plates at concentrations of 21 mg/L and 85.6 mg/L, respectively. Sigma Aldrich Bacto Agar (A5306) was then included at a concentration of 20 g/L. The final media were supplemented with 2% Glucose.

Expression and Purification of GTSF1 constructs in Sf9 cells (Performed by Dr. Jonathan Ipsaro in Joshua-Tor Group from CSHL, NY)

The experiment was performed according to the procedure described in Ipsaro et al 2021. Briefly, constructs of GTSF1 from *M. musculus,* and *C. elegans* were generated using SLIC cloning in DH5a cells. Each construct had a C-terminal TEV-Strep2 tag and was cloned into the vector pFL, then integrated into bacmids using DH10MultiBac cells. The bacmids were transfected into Sf9 cells for expression using the baculoviral system. After expression, cells were lysed, and the lysate was clarified by ultracentrifugation. The clarified lysate was applied to a Strep-Tactin column, washed, and the bound GTSF1 proteins were eluted. Protein purity was assessed by SDS-PAGE, and any co-purifying nucleic acids were isolated and assessed by Urea-PAGE.

Antibodies against CbrGTSF-1 and PpaGTSF-1 (Developed with assistance of Martin Möckel, IMB Protein-Production Core-Facility)

C. briggsae and *P. pacificus gtsf-1* sequences were amplified from cDNA and cloned into a plasmid with *gtsf-1* aligned in-frame with 6xHIs, maltose binding protein (MBP) from *E. coli*, and a recognition/cleavage site for the tobacco etch virus (TEV) protease. The engineered plasmid was then expressed in *E. coli* Rosetta™(DE3) Competent Cells, designed as BL21 derivatives to optimize the expression of eukaryotic proteins with infrequently used codons in *E. coli*, with overnight expression at 18˚C. Recombinant GTSF-1 were purified using Ni-NTA Agarose beads on a column. Following cleavage by the TEV protease, recombinant GTSF-1 was forwarded to Eurogentec for immunization in rabbits.

For antibody production, the serum from the immunized rabbits was collected and subjected to affinity purification using recombinant GTSF-1, tagged with 6xHis, SUMO (a cleavable ubiquitin-like protein tag), and HRV3C site (a recognition and cleavage site for human rhinovirus 3C and PreScission proteases) as bait. The resulting antibodies were then validated through Western blot analysis against purified recombinant GTSF-1 or worm lysates.

Western Blot

Synchronized animals were harvested at gravid-adult stage in M9 buffer, snap frozen in liquid nitrogen and stored as pellets in -20˚C. To prepare lysates, worm pellets were resuspended in lysis buffer (50 mM Tris HCl pH 7.5, 300 mM NaCl, 3 mM MgCl2, 2 mM DTT, 0.2% Triton™ X-100, and cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail from Roche), and then sonicated at 4°C using the Bioruptor® Plus device for 10 cycles at high energy, with intervals of 30 seconds ON and 30 seconds OFF. The protein concentration of the lysate was determined using the Pierce™ BCA™ Protein Assay and an Infinite® M200 Pro plate reader. Lysates were appropriately diluted in LDS buffer supplemented with 1M DTT, denatured at 95°C for 5 minutes, and stored at -20°C.

Lysates were loaded onto either a 4-12% or 12% Bis-Tris NuPAGE gel using 1x NuPAGE MES SDS Running Buffer, with 3 µl of Color Prestained Protein Standard, Broad Range from NEB (P7719S) serving as the standard. Gels were transferred overnight (15V, 8°C) to an Amersham Protran 0.45 µm nitrocellulose membrane (10600002) using 1x transfer buffer supplemented with 15% methanol.

For blocking, membranes were incubated with 1x PBS-Tween-20 (0.1%) supplemented with 5% skim milk for 30 minutes. Blocked membranes were then exposed to primary antibodies diluted in 1x PBS-Tween-20 (0.1%) supplemented with 5% skim milk for 1 hour at room temperature [1:500 dilution of rabbit anti-GTSF-1 (Q5963) and rabbit anti-PID-3 (Q5885), mouse anti-FLAG (Sigma Aldrich, F7425) diluted 1:1000]. After three washes of 15 minutes each in wash buffer (PBS-Tween-20 (1%)), the membranes were incubated with secondary antibodies diluted 1:10000 in PBS-Tween-20 (0.1%) for 1 hour [anti-rabbit HRP-linked antibody (CST, 7074S), anti-mouse HRP-linked antibody (CST, 7076S)]. The membranes were washed again three times in wash buffer. Detection was carried out using the SuperSignal™ West Atto Ultimate Sensitivity Substrate kit (ThermoFisher, A38554) according to the manufacturer's instructions.

Immunoprecipitation

Synchronized nematodes were cultivated at 20°C, harvested at either the Gravid-Adult (GA) or late-L4 stages, and subsequently frozen in 200 µl aliquots on dry ice in sterile water. Upon thawing, aliquots were combined with an equal volume of 2x lysis buffer (comprising 50 mM Tris HCl pH 7.5, 300 mM NaCl, 3 mM MgCl2, 2 mM DTT, 0.2% Triton™ X-100, and cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail from Roche). Sonication was carried out using a Bioruptor® Plus device at 4°C for 10 cycles, with intervals of 30 seconds ON and 30 seconds OFF. Samples were centrifuged at 4°C and 21,000 xg for 10 minutes. The insoluble pellets were reconstituted in Novex™ NuPAGE™ LDS sample buffer (Invitrogen™) supplemented with 100 mM DTT. Protein concentrations of the soluble worm extract were quantified using the Pierce™ BCA™ Protein-Assay and an Infinite® M200 Pro plate reader. Extracts were diluted with 1x lysis buffer. Around 1 to 3mg of total protein was used for immunoprecipitation experiments.

For each IP, 30 µl Novex™ DYNAL™ Dynabeads™ Protein G were conjugated to 1.5 to 2µg of the relevant antibody according to the manufacturer's instructions. Concurrently, another 30 µl of Dynabeads™, were washed three times with 1x wash buffer (consisting of 25 mM Tris HCl pH 7.5, 150 mM NaCl, 1.5 mM MgCl2, 1 mM DTT, and cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail) and incubated with worm extract for 1 to 2 hours at 4°C. Extracts were separated from non-conjugated Dynabeads™, combined with antibody-conjugated Dynabeads™, and incubated with rotation for 2 hours at 4°C. These antibody-conjugated Dynabeads™ were washed thrice with 1x wash buffer and resuspended in 25 µl 1.2x Novex™ NuPAGE™ LDS sample buffer supplemented with 120 mM DTT. The antibodies used for Immunoprecipitation experiments in this study are Polyclonal Rabbit-ANTI-GTSF-1 produced in-house against various nematode GTSF-1, Rabbit-

ANTI-IgG from Cell-Signaling Technologies (Article No. 2729) and Monoclonal ANTI-FLAG® M2. The immunoprecipitation samples were used for mass-spectrometry or western blot analysis.

Quantitative mass-spectrometry (Performed by Emily Nischwitz, Butter Group, IMB & Proteomics core facility, IMB)

Enzymatic protein digestion

For samples RK_0460 to RK_0467, the eluted proteins were separated briefly in a 10% NuPAGE Bis-Tris gel, stained with Coomassie blue and cut into small gel cubes, followed by destaining in 50% ethanol/25 mM ammonium bicarbonate. The proteins were then reduced in 10 mM DTT at 56°C and alkylated by 50 mM iodoacetamide in the dark at room temperature. Afterwards, proteins were digested by trypsin in 50 mM ammonium bicarbonate buffer overnight at 37°C. Following peptide extraction through sequential incubation of gel cubes in 30% and 100% acetonitrile, the sample volume was reduced in a centrifugal evaporator to remove residual acetonitrile. The resultant peptide solution was purified by solid phase extraction in C18 StageTips (PMID: 12585499).

For samples RK_0468 to RK_0495, the eluted proteins were processed using the SP3 approach (PMID: 30464214). The proteins were then digested using trypsin overnight at 37°C. The resultant peptide solution was purified in C18 StageTips.

Liquid chromatography tandem mass spectrometry

Peptides were separated in an in-house packed 30-cm analytical column (inner diameter: 75 µm; ReproSil-Pur 120 C18-AQ 1.9-µm beads, Dr. Maisch GmbH; heated at 50°C) by online reverse phase chromatography through a 105-min non-linear gradient of 1.6-32% acetonitrile with 0.1% formic acid at a nanoflow rate of 225 nl/min. The eluted peptides were sprayed directly by electrospray ionization into a Q Exactive Plus Orbitrap mass spectrometer (Thermo Scientific). Mass spectrometry measurement was conducted in data-dependent acquisition mode using a top10 method with one full scan (mass range: 300 to 1,650 m/z; resolution: 70,000, target value: 3 × 106, maximum injection time: 20 ms) followed by 10 fragmentation scans via higher energy collision dissociation (HCD; normalised collision energy: 25%, resolution: 17,500, target value: 1 × 105, maximum injection time: 120 ms, isolation window: 1.8 m/z). Precursor ions of unassigned or +1 charge state were rejected. Additionally, precursor ions already isolated for fragmentation were dynamically excluded for 20 s.

Mass spectrometry data processing and statistical analysis

Mass spectrometry raw data were processed by MaxQuant software package (version 2.1.3.0) (PMID: 19029910) using its built-in Andromeda search engine (PMID: 21254760). Spectral data were searched against a target-decoy database consisting of the forward and reverse sequences of the UniProt *C. briggsae* (release 2023_02; 21,756 entries) or *P. pacificus* (release 2023_03; 26,122 entries) reference proteome, the UniProt *E. coli* (release 2023_01; 5,064 entries) reference proteome and a list of common contaminants. Trypsin/P specificity was assigned. Carbamidomethylation of cysteine was set as fixed modification. Methionine oxidation and protein N-terminal acetylation were chosen as variable modifications. A maximum of 2 missed cleavages were tolerated. The “second peptides” options were switched on. The “match between

runs" function was activated. The minimum peptide length was set to be 7 amino acids. False discovery rate (FDR) was set to 1% at both peptide and protein levels.

The MaxLFQ algorithm (PMID: 24942700) was employed for label-free protein quantification without using its default normalization option. Minimum LFQ ratio count was set to be one. Both the unique and razor peptides were used for quantification. Detected *E. coli* proteins, reverse hits, potential contaminants and "only identified by site" protein groups were filtered out. Data normalization was performed on the log-transformed data using a median-centering approach. Proteins were further filtered to retain only those detected in at least two out of the four replicates in either group of each comparison. Following imputation of the missing LFQ intensity values, the statistical significance of the difference between the two groups was assessed using a modified t-statistic (t(SAM, statistical analysis of microarrays) (PMID: 11309499) and visualized in a volcano plot. The combined significance threshold (hyperbolic curve) was defined as t0 = 1.0~1.4 and s0 = 1.5.

Functional Enrichment Profiling

Protein IDs enriched in the IP-MS experiment were used as input for g:GOst in the g:Profiler web server (https://biit.cs.ut.ee/gprofiler/gost). The genome *C. briggsae* PRJNA10731 [WS285] was used as reference. The anlaysis was conducted using default settings with a 0.05 g:SCS significance threshold. All annotated protein-coding genes was used as the statistical domain scope (Kolberg et al., 2023; Reimand et al., 2007).

RNA extraction

Synchronized animals were harvested in M9 buffer at Gravid-Adult or Late-L4 stages and snap-frozen into 50 µl pellets on dry ice. For RNA extraction, a liquid nitrogen freeze-thawing-based cell-lysis procedure was employed. In this method, the 50 µl worm pellets was resuspended in 500 µl of TRIzol™ LS Reagent (Invitrogen). Subsequently, the sample underwent a rapid freeze in liquid nitrogen for 30 seconds followed by thawing in a 37°C water bath for 1 minute or until the entire sample thawed. This freeze-thaw cycle was repeated at least eight times to ensure efficient cell lysis. After each cycle, the sample was vortexed well. Microscopic examination confirmed the absence of intact worms, with only visible worm debris. In the case of starting with gravid adults, intact eggs remained unaffected by the lysis method. The samples were then cleared by centrifuging at 21,000xg for 5 minutes. RNA isolation from the supernatant was carried out using the Direct-zol™ RNA Miniprep kit following the manufacturer's instructions.

NGS Library Preparation (Performed by Genomics Core Facility, IMB)

RppH Treatment: 200 to 1000 ng RNA was subjected to RNA 5' Pyrophosphohydrolase (RppH) treatment (NEB, M0356S) at 37°C for 1 hour to eliminate pyrophosphate from the 5´ end of triphosphorylated RNA, leaving a 5´ monophosphate RNA. Following purification, the samples were quantified using the Qubit RNA High Sensitivity Assay Kit.

NGS Library Preparation and Sequencing: Next-generation sequencing (NGS) library preparation employed the NEXTflex Small RNA-Seq Kit V3, following Bioo Scientific's standard protocol (V19.01) from Step A to Step G. The process utilized the NEXTFlex 3' SR Adaptor and 5' SR Adaptor (5'rApp/NNNNTGGAATTCTCGGGTGCCAAGG/3ddC/ and

5'GUUCAGAGUUCUACAGUCCGACGAUCNNNN, respectively). Step A (NEXTflex 3` 4N Adenylated Adapter Ligation) was conducted overnight at 20°C.

Libraries were prepared with an initial amount of 500 to 900 ng and amplified through 14 to 15 PCR cycles (specific amounts for individual experiments detailed in the supplementary table). Amplified libraries underwent purification via an 8% TBE gel, and fragments in the 15–40 nt range were size-selected. Profiling occurred using a High Sensitivity DNA Chip on a 2100 Bioanalyzer (Agilent Technologies), and quantification was performed with the Qubit dsDNA HS Assay Kit on a Qubit 2.0 Fluorometer (Life Technologies). Samples were equimolarly pooled and sequenced on a Highoutput NextSeq 500/550 Flowcell, single-read for 1x 84 cycles, plus 7 cycles for the index read.

For Projects 8 to 10, the NextSeq 2000 sequencer was employed, necessitating the use of NextFlex small RNAseq PreArrayed UDI Barcode primers to prevent index hopping (James P et al., 1996). Amplified libraries were purified using a PippinHT (Sage Science) on a 3% agarose gel cassette, size-selected for library fragments in the 150–196 bp range (adapter size + insert size), and pooled equimolarly. Sequencing took place on a NextSeq 2000 P2 (100 cycles) Flowcell, single-read for 1x 84 cycles, plus 2x8 cycles for the dual index read. Details of projects in **Table 4.**

Table 4 | NGS projects and library details:

LIMS ID	Species	Strains	Life stages	Treatments	Sequencing Platform	Flowcell type and reagents	Read Length	Starting Material and Number of PCR cycles
1683	*C. elegans*	WT, RFK1317 ,RFK1318	GA; L4	None	NextSeq500 (illumina)	High Output - 75 cycle kit (NextSeq)	1x75bp (Single End) - NextSeq	500ng and amplified in 15 PCR cycles
1685	*C. elegans*	WT, RFK1317, RFK1318	GA; L4	RppH	NextSeq500 (illumina)	High Output - 75 cycle kit (NextSeq)	1x75bp (Single End) - NextSeq	650 ng and amplified in 14 PCR cycles
1946	*C. briggsae*	WT; RFK1609	GA;L4	RppH	NextSeq500 (illumina)	High Output - 75 cycle kit (NextSeq)	1x75bp (Single End) - NextSeq	900 ng and amplified in 14 PCR cycles
1947	*C. briggsae*	WT; RFK1609	GA;L4	None	NextSeq500 (illumina)	High Output - 75 cycle kit (NextSeq)	1x75bp (Single End) - NextSeq	900 ng and amplified in 14 PCR cycles
2075	*C. briggsae*	WT, RFK1609 ,RFK1610	GA;L4	None	NextSeq 2000 (illumina)	P2 - 100 cycle kit (NextSeq 2000)	1x50 P3 (Single End) - NextSeq 2000	273ng and amplified in 18 PCR cycles.
2077	*C. briggsae*	WT, RFK1609, RFK1610	GA;L4	RppH	NextSeq 2000 (illumina)	P2 - 100 cycle kit (NextSeq 2000)	1x50 P3 (Single End) - NextSeq 2000	273ng and amplified in 18 PCR cycles.
2078	*C.briggsae*	RFK RFK1124	GA	None	NextSeq 2000 (illumina)	P2 - 100 cycle kit (NextSeq 2000)	1x50 P3 (Single End) - NextSeq 2000	200ng and amplified in 18 PCR cycles.
2079	*C.briggsae*	RFK RFK1124	GA	RppH	NextSeq 2000 (illumina)	P2 - 100 cycle kit (NextSeq 2000)	1x50 P3 (Single End) - NextSeq 2000	200ng and amplified in 18 PCR cycles.

NGS Data analysis with TinyRNA:

Analysis of small-RNA sequencing datasets was performed using TinyRNA (Tate et al 2023). Within this pipeline, the preprocessing of FASTQ files, with adapter trimming and quality filtering, was executed through the utilization of fastp (Chen *et al.*, 2018) -a TGGAATTCTCGGGTGCCAAGG, -q 20 -e 25 –u 10). 5' and 3' random UMIs (NNNN-RNA sequence-NNNN) were then removed and reads shorter than 15 nucleotides were filtered out. Subsequently, unique sequences were counted and collapsed using tiny-collapse. Collapsed reads were aligned to *C. elegans* (WS285) and *C.briggsae* (WS279) reference genome using bowtie (Langmead *et al.*, 2009). Features were

assigned using tiny-count that uses the Genomic Array of Sets and GFF reader from HTSeq (Anders, Pyl and Huber, 2014). Feature assignment is based on features annotated in custom GFF input files. Feature selection rules are specified in a separate features table. All read counts are normalized to numbers of genomic hits and feature matches before assignment to features. Differential expression analysis is performed by tiny-deseq, which uses DESeq2 (Love, Huber and Anders, 2014). Finally, the outputs generated by tiny-count and tiny-deseq are plotted using tinyRNA's plotter utility, tiny-plot.

For *C. briggsae*, we modified the genome annotation (c_briggsae.PRJNA10731. WS285.annotations) by incorporating miRNA loci from miRbase genome-build-accession (NCBI_Assembly:GCA_000004555.3) and piRNA loci from Beltran et al. 2019. PiRNAs that were not assembled into chromosomes were excluded from the annotation file. The categorization of features involved assigning 21U RNAs based on size selection, specifically targeting 18-21nt long reads that map sense to annotated piRNA loci. Similarly, 26G-RNAs were selected based on a size criterion of 26nt, with a 5' nucleotide as G, and mapped antisense to annotated protein-coding genes, pseudogene, or transposon loci.

Conservation Heat Map (Performed by Prof. Peter Sarkies, Department of Biochemistry, Oxford).

Reciprocal BLASTP was conducted with the *C. elegans* WS235 proteome against the most updated versions (as of 2023) of the proteomes of other nematodes listed in the heatmap. Hits were filtered to retain only 1:1 orthologs with the lowest e-value. The e-values were then plotted into a heatmap in R using heatmap.2 function.

Alpha-Fold

C. elegans and *C. briggsae* GTSF-1 structure predictions were downloaded from AlphaFold Protein Structure Database (EMBL-EBI) with identifiers A3QMD6 and A8XIH1 respectively. The GTSF-1 structure for *P. pacificus* was generated using AlphaFold2 (v2.3.2) using monomer parameters. All other protein complex predictions were performed using AlphaFold2 (v2.3.2) using the AlphaFold-Multimer model as described in (Jumper et al., 2021; Evans et al., 2021).

REFERENCES

de Albuquerque, B.F.M. *et al.* (2014) 'PID-1 is a novel factor that operates during 21U-RNA biogenesis in Caenorhabditis elegans', *Genes & development*, 28(7), pp. 683–688. Available at: https://doi.org/10.1101/GAD.238220.114.

Almeida, M.V., Andrade-Navarro, M.A. and Ketting, R.F. (2019) 'Function and Evolution of Nematode RNAi Pathways', *Non-Coding RNA*, 5(1), p. 8. Available at: https://doi.org/10.3390/ncrna5010008.

Anders, S., Pyl, P.T. and Huber, W. (2014) 'HTSeq—a Python framework to work with high-throughput sequencing data', *Bioinformatics*, p. btu638. Available at: https://doi.org/10.1093/bioinformatics/btu638.

Aoki, K. *et al.* (2007) 'In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans', *The EMBO Journal*, 26(24), p. 5007. Available at: https://doi.org/10.1038/SJ.EMBOJ.7601910.

Bartel, D.P. (2018) 'Metazoan MicroRNAs', *Cell*, 173(1), pp. 20–51. Available at: https://doi.org/10.1016/J.CELL.2018.03.006.

Barucci, G. *et al.* (2020) 'Small RNA-mediated transgenerational silencing of histone genes impairs fertility in piRNA mutants', *Nature cell biology*, 22(2), p. 235. Available at: https://doi.org/10.1038/S41556-020-0462-7.

Batista, P.J. *et al.* (2008) 'PRG-1 and 21U-RNAs Interact to Form the piRNA Complex Required for Fertility in C. elegans', *Molecular Cell*, 31(1), pp. 67–78. Available at: https://doi.org/10.1016/j.molcel.2008.06.002.

Baulcombe, D.C. (2023) 'An RNA World', *Annual Review of Plant Biology*, 74(Volume 74, 2023), pp. 1–20. Available at: https://doi.org/10.1146/ANNUREV-ARPLANT-070622-021021/CITE/REFWORKS.

Beltran, T. *et al.* (2019) 'Comparative Epigenomics Reveals that RNA Polymerase II Pausing and Chromatin Domain Organization Control Nematode piRNA Biogenesis', *Developmental cell*, 48(6), pp. 793-810.e6. Available at: https://doi.org/10.1016/J.DEVCEL.2018.12.026.

Billi, A.C. *et al.* (2012) 'The Caenorhabditis elegans HEN1 Ortholog, HENN-1, Methylates and Stabilizes Select Subclasses of Germline Small RNAs', *PLoS Genet*, 8(4), p. e1002617. Available at: https://doi.org/10.1371/journal.pgen.1002617.

Blaxter, M. and Koutsovoulos, G. (2015) 'The evolution of parasitism in Nematoda', *Parasitology*, 142(Suppl 1), p. S26. Available at: https://doi.org/10.1017/S0031182014000791.

Brenner, S. (1974) 'The Genetics of CAENORHABDITIS ELEGANS', *Genetics*, 77(1), p. 71. Available at: https://doi.org/10.1093/GENETICS/77.1.71.

Buckley, Bethany A *et al.* (2012) 'A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality', *Nature*, 489(7416), pp. 447–451. Available at: https://doi.org/10.1038/nature11352.

Buckley, Bethany A. *et al.* (2012) 'A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality', *Nature*, 489(7416), pp. 447–451. Available at: https://doi.org/10.1038/NATURE11352.

Burkhart, K.B. *et al.* (2011) 'A Pre-mRNA–Associating Factor Links Endogenous siRNAs to Chromatin Regulation', *PLOS Genetics*, 7(8), p. e1002249. Available at: https://doi.org/10.1371/journal.pgen.1002249.

Burton, N.O., Burkhart, K.B. and Kennedy, S. (2011) 'Nuclear RNAi maintains heritable gene silencing in Caenorhabditis elegans', *Proceedings of the National Academy of Sciences*, 108(49), pp. 19683–19688. Available at: https://doi.org/10.1073/pnas.1113310108.

Castel, S.E. and Martienssen, R.A. (2013) 'RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond', *Nature Reviews Genetics 2013 14:2*, 14(2), pp. 100–112. Available at: https://doi.org/10.1038/nrg3355.

Cecere, G. *et al.* (2012) 'Promoters recognized by forkhead proteins exist for individual 21U-RNAs', *Molecular cell*, 47(5), pp. 734–745. Available at: https://doi.org/10.1016/J.MOLCEL.2012.06.021.

Cerutti, H. and Casas-Mollano, J.A. (2006) 'On the origin and functions of RNA-mediated silencing: from protists to man', *Current genetics*, 50(2), pp. 81–99. Available at: https://doi.org/10.1007/S00294-006-0078-X.

Chen, S. *et al.* (2018) 'fastp: an ultra-fast all-in-one FASTQ preprocessor', *Bioinformatics*, 34(17), pp. i884–i890. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTY560.

Chiu, J. *et al.* (2004) 'Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h', *Nucleic Acids Research*, 32(21), p. e174. Available at: https://doi.org/10.1093/NAR/GNH172.

Chul Kwon, S. *et al.* (2020) 'ERH facilitates microRNA maturation through the interaction with the N-terminus of DGCR8', *Nucleic Acids Research*, 48(19), pp. 11097–11112. Available at: https://doi.org/10.1093/NAR/GKAA827.

Claycomb, J.M. (2014) 'Ancient endo-siRNA pathways reveal new tricks', *Current biology : CB*, 24(15). Available at: https://doi.org/10.1016/J.CUB.2014.06.009.

Czech, B. *et al.* (2018) 'piRNA-Guided Genome Defense: From Biogenesis to Silencing'. Available at: https://doi.org/10.1146/annurev-genet-120417.

Daniel Gietz, R. and Schiestl, R.H. (no date) 'High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method'. Available at: https://doi.org/10.1038/nprot.2007.13.

Duchaine, T.F. *et al.* (2006) 'Functional proteomics reveals the biochemical niche of C. elegans DCR-1 in multiple small-RNA-mediated pathways', *Cell*, 124(2), pp. 343–354. Available at: https://doi.org/10.1016/j.cell.2005.11.036.

Evans, R. *et al.* (2021) 'Protein complex prediction with AlphaFold-Multimer', *bioRxiv*, p. 2021.10.04.463034. Available at: https://doi.org/10.1101/2021.10.04.463034.

Fagard, M. *et al.* (no date) *AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals*. Available at: www.pnas.orgcgidoi10.1073pnas.200217597.

Fang, W. and Bartel, D.P. (2020) 'MicroRNA Clustering Assists Processing of Suboptimal MicroRNA Hairpins through the Action of the ERH Protein', *Molecular cell*, 78(2), pp. 289-302.e6. Available at: https://doi.org/10.1016/J.MOLCEL.2020.01.026.

Fire, A. *et al.* (1998) 'Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans', *Nature 1998 391:6669*, 391(6669), pp. 806–811. Available at: https://doi.org/10.1038/35888.

Gent, J.I. *et al.* (2009) 'A Caenorhabditis elegans RNA-directed RNA polymerase in sperm development and endogenous RNA interference', *Genetics*, 183(4), pp. 1297–1314. Available at: https://doi.org/10.1534/genetics.109.109686.

Gent, J.I. *et al.* (2010) 'Distinct Phases of siRNA Synthesis in an Endogenous RNAi Pathway in C. elegans Soma', *Molecular Cell*, 37(5), pp. 679–689. Available at: https://doi.org/10.1016/j.molcel.2010.01.012.

Gibson, D.G. *et al.* (2009) 'Enzymatic assembly of DNA molecules up to several hundred kilobases', *Nature Methods 2009 6:5*, 6(5), pp. 343–345. Available at: https://doi.org/10.1038/nmeth.1318.

Goh, W.S.S. *et al.* (2014) 'A genome-wide RNAi screen identifies factors required for distinct stages of C. elegans piRNA biogenesis', *Genes & development*, 28(7), pp. 797–807. Available at: https://doi.org/10.1101/GAD.235622.113.

Guang, S. *et al.* (2008) 'An Argonaute transports siRNAs from the cytoplasm to the nucleus', *Science (New York, N.Y.)*, 321(5888), pp. 537–541. Available at: https://doi.org/10.1126/SCIENCE.1157647.

Guang, S. *et al.* (2010) 'Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription', *Nature*, 465(7301), pp. 1097–1101. Available at: https://doi.org/10.1038/nature09095.

Gu, S.G. *et al.* (2012) 'Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint', *Nature Genetics*, 44(2), pp. 157–164. Available at: https://doi.org/10.1038/ng.1039.

Haeussler, M. *et al.* (2016) 'Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR', *Genome biology*, 17(1). Available at: https://doi.org/10.1186/S13059-016-1012-2.

Ha, M. and Kim, V.N. (2014) 'Regulation of microRNA biogenesis', *Nature reviews. Molecular cell biology*, 15(8), pp. 509–524. Available at: https://doi.org/10.1038/NRM3838.

Han, T. *et al.* (2009) '26G endo-siRNAs regulate spermatogenic and zygotic gene expression in Caenorhabditis elegans', *Proceedings of the National Academy of Sciences of the United States of America*, 106(44), pp. 18674–18679. Available at: https://doi.org/10.1073/pnas.0906378106.

Holz, A. and Streit, A. (2017) 'Gain and loss of small RNA classes-characterization of small rnas in the parasitic nematode family strongyloididae', *Genome Biology and Evolution*, 9(10), pp. 2826–2843. Available at: https://doi.org/10.1093/gbe/evx197.

Huggins, H.P. *et al.* (2020) 'Distinct roles of two eIF4E isoforms in the germline of Caenorhabditis elegans', *Journal of Cell Science*, 133(6). Available at: https://doi.org/10.1242/JCS.237990.

Hutter, K. *et al.* (2020) 'SAFB2 Enables the Processing of Suboptimal Stem-Loop Structures in Clustered Primary miRNA Transcripts', *Molecular cell*, 78(5), pp. 876-889.e6. Available at: https://doi.org/10.1016/J.MOLCEL.2020.05.011.

Iyer, L.M., Koonin, E. V and Aravind, L. (2003) *Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases*. Available at: http://www.biomedcentral.com/1472-6807/3/1.

James, P., Halladay, J. and Craig, E.A. (1996) 'Genomic Libraries and a Host Strain Designed for Highly Efficient Two-Hybrid Selection in Yeast'.

Kamminga, L.M. *et al.* (2012) 'Differential impact of the HEN1 homolog HENN-1 on 21U and 26G RNAs in the germline of Caenorhabditis elegans', *PLoS Genetics*, 8(7). Available at: https://doi.org/10.1371/journal.pgen.1002702.

Ketting, R.F. (2011) 'The Many Faces of RNAi', *Developmental Cell*, 20(2), pp. 148–161. Available at: https://doi.org/10.1016/j.devcel.2011.01.012.

Ketting, R.F. and Cochella, L. (2021) 'Concepts and functions of small RNA pathways in C. elegans', *Current topics in developmental biology*, 144, pp. 45–89. Available at: https://doi.org/10.1016/BS.CTDB.2020.08.002.

Langmead, B. *et al.* (2009) 'Ultrafast and memory-efficient alignment of short DNA sequences to the human genome', *Genome Biology*, 10(3), pp. 1–10. Available at: https://doi.org/10.1186/GB-2009-10-3-R25/TABLES/5.

Lee, H.-C. *et al.* (2012) 'C. elegans piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts', *Cell*, 150(1), pp. 78–87. Available at: https://doi.org/10.1016/j.cell.2012.06.016.

Lee, H.C. *et al.* (2012) 'C. elegans piRNAs mediate the genome-wide surveillance of germline transcripts', *Cell*, 150(1), p. 78. Available at: https://doi.org/10.1016/J.CELL.2012.06.016.

Lee, S.R. and Collins, K. (2007) 'Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs', *Nature Structural & Molecular Biology 2007 14:7*, 14(7), pp. 604–610. Available at: https://doi.org/10.1038/nsmb1262.

Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12). Available at: https://doi.org/10.1186/s13059-014-0550-8.

Makeyev, E. V and Bamford, D.H. (2002) 'Cellular RNA-Dependent RNA Polymerase Involved in Posttranscriptional Gene Silencing Has Two Distinct Activity Modes', *Molecular Cell*, 10(6), pp. 1417–1427. Available at: https://doi.org/10.1016/S1097-2765(02)00780-3.

Maniar, J.M. and Fire, A.Z. (2011) 'EGO-1, a C. elegans RdRP, modulates gene expression via production of mRNA-templated short antisense RNAs', *Current Biology*, 21(6), pp. 449–459. Available at: https://doi.org/10.1016/j.cub.2011.02.019.

Mao, H. *et al.* (2015) 'The Nrde Pathway Mediates Small-RNA-Directed Histone H3 Lysine 27 Trimethylation in Caenorhabditis elegans', *Current Biology*, 25, pp. 2398–2403. Available at: https://doi.org/10.1016/j.cub.2015.07.051.

Meister, G. (2013) 'Argonaute proteins: functional insights and emerging roles', *Nature reviews. Genetics*, 14(7), pp. 447–459. Available at: https://doi.org/10.1038/NRG3462.

Montgomery, T.A. *et al.* (2012) 'PIWI associated siRNAs and piRNAs specifically require the Caenorhabditis elegans HEN1 ortholog henn-1', *PLoS Genetics*, 8(4). Available at: https://doi.org/10.1371/journal.pgen.1002616.

Nigon, V. and Dougherty, E.C. (1949) 'Reproductive patterns and attempts at reciprocal crossing of Rhabditis elegans Maupas, 1900, and Rhabditis briggsae Dougherty and Nigon, 1949 (Nematoda: Rhabditidae)', *The Journal of experimental zoology*, 112(3), pp. 485–503. Available at: https://doi.org/10.1002/JEZ.1401120307.

Owttrim, G.W. (2013) 'RNA helicases', *RNA Biology*, 10(1), pp. 96–110. Available at: https://doi.org/10.4161/rna.22638.

Ozata, D.M. *et al.* (2018) 'PIWI-interacting RNAs: small RNAs with big functions', *Nature Reviews Genetics*, p. 1. Available at: https://doi.org/10.1038/s41576-018-0073-3.

Ozata, D.M. *et al.* (2019) 'PIWI-interacting RNAs: small RNAs with big functions', *Nature Reviews Genetics*. Nature Publishing Group, pp. 89–108. Available at: https://doi.org/10.1038/s41576-018-0073-3.

Paix, A., Folkmann, A. and Seydoux, G. (2017) 'Precision genome editing using CRISPR-Cas9 and linear repair templates in C. elegans', *Methods*, 121–122, pp. 86–93. Available at: https://doi.org/10.1016/j.ymeth.2017.03.023.

Palmer, William H, Hadfield, J.D. and Obbard, D.J. (2018) 'RNA-Interference Pathways Display High Rates of', 208(April), pp. 1585–1599. Available at: https://doi.org/10.1534/genetics.117.300567/-/DC1.1.

Palmer, William H., Hadfield, J.D. and Obbard, D.J. (2018) 'RNA-interference pathways display high rates of adaptive protein evolution in multiple invertebrates', *Genetics*, 208(4), pp. 1585–1599. Available at: https://doi.org/10.1534/GENETICS.117.300567/-/DC1/FILES2.DOCX.

Pavelec, D.M. *et al.* (2009) 'Requirement for the ERI/DICER complex in endogenous RNA interference and sperm development in Caenorhabditis elegans', *Genetics*, 183(4), pp. 1283–1295. Available at: https://doi.org/10.1534/genetics.109.108134.

Perez-Borrajero, C. *et al.* (2021) 'Structural basis of PETISCO complex assembly during piRNA biogenesis in C. elegans', *Genes and Development*, 35(17–18), pp. 1304–1323. Available at: https://doi.org/10.1101/GAD.348648.121/-/DC1.

Phillips, C.M. *et al.* (2012) 'MUT-16 promotes formation of perinuclear mutator foci required for RNA silencing in the C. elegans germline', *Genes & development*, 26(13), pp. 1433–1444. Available at: https://doi.org/10.1101/GAD.193904.112.

Phillips, C.M. *et al.* (2014) 'MUT-14 and SMUT-1 DEAD box RNA helicases have overlapping roles in germline RNAi and endogenous siRNA formation', *Current biology : CB*, 24(8), pp. 839–844. Available at: https://doi.org/10.1016/J.CUB.2014.02.060.

Podvalnaya, N. *et al.* (2023) 'piRNA processing by a trimeric Schlafen-domain nuclease', *Nature*, 622(7982), pp. 402–409. Available at: https://doi.org/10.1038/S41586-023-06588-2.

Qian, X. *et al.* (2016) 'Functional Evolution in Orthologous Cell-encoded RNA-dependent RNA Polymerases', *Journal of Biological Chemistry*, 291(17), pp. 9295–9309. Available at: https://doi.org/10.1074/JBC.M115.685933.

Rodrigues, R.J.C. *et al.* (2018) 'PETISCO is a novel protein complex required for 21U RNA biogenesis and embryonic viability', *bioRxiv*, p. 463711. Available at: https://doi.org/10.1101/463711.

Ruby, J.G. *et al.* (2006) 'Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in C. elegans', *Cell*, 127(6), pp. 1193–1207. Available at: https://doi.org/10.1016/j.cell.2006.10.040.

Sarkies, P. *et al.* (2015) 'Ancient and Novel Small RNA Pathways Compensate for the Loss of piRNAs in Multiple Independent Nematode Lineages', *PLOS Biol*, 13(2), p. e1002061. Available at: https://doi.org/10.1371/journal.pbio.1002061.

Schiebel, W. *et al.* (1993) 'RNA-directed RNA polymerase from tomato leaves. I. Purification and physical properties', *Journal of Biological Chemistry*, 268(16), pp. 11851–11857. Available at: https://doi.org/10.1016/s0021-9258(19)50278-2.

Sen, G.L. and Blau, H.M. (2006) 'A brief history of RNAi: the silence of the genes', *The FASEB Journal*, 20(9), pp. 1293–1299. Available at: https://doi.org/10.1096/FJ.06-6014REV.

Shabalina, S.A. and Koonin, E. V (2008) 'Origins and evolution of eukaryotic RNA interference', *Trends in Ecology & Evolution*, 23(10), pp. 578–587. Available at: https://doi.org/10.1016/j.tree.2008.06.005.

Shi, Z. *et al.* (2013a) 'High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes', *Genome Research*, 23(3), pp. 497–508. Available at: https://doi.org/10.1101/gr.149112.112.

Shi, Z. *et al.* (2013b) 'High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes', *Genome Research*, 23(3), pp. 497–508. Available at: https://doi.org/10.1101/gr.149112.112.

Shukla, A., Perales, R. and Kennedy, S. (2021) 'piRNAs Coordinate poly(UG) Tailing to Prevent Aberrant and Permanent Gene Silencing', *bioRxiv*, p. 2021.01.30.428010. Available at: http://biorxiv.org/content/early/2021/02/01/2021.01.30.428010.abstract.

Sijen, T. *et al.* (2007) 'Secondary siRNAs Result from Unprimed RNA Synthesis and Form a Distinct Class', *Science*, 315(5809), pp. 244–247. Available at: https://doi.org/10.1126/science.1136699.

Simon, M. *et al.* (2014) 'Reduced Insulin/IGF-1 Signaling Restores Germ Cell Immortality to Caenorhabditis elegans Piwi Mutants', *Cell Reports*, 7(3), pp. 762–773. Available at: https://doi.org/10.1016/j.celrep.2014.03.056.

Smardon, A. *et al.* (2000) *EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans*.

Sommer, R.J. *et al.* (1996) 'Morphological, genetic and molecular description of Pristionchus pacificus sp. n. (Nematoda : N eodiplogastridae)', *Fundam. appl. NemalOl*, 19(6), pp. 511–521.

Sugiyama, T. *et al.* (2016) 'Enhancer of Rudimentary Cooperates with Conserved RNA-Processing Factors to Promote Meiotic mRNA Decay and Facultative Heterochromatin Assembly', *Molecular cell*, 61(5), pp. 747–759. Available at: https://doi.org/10.1016/J.MOLCEL.2016.01.029.

Svoboda, P. (no date) 'Key Mechanistic Principles and Considerations Concerning RNA Interference'. Available at: https://doi.org/10.3389/fpls.2020.01237.

Tang, G. *et al.* (2003) 'A biochemical framework for RNA silencing in plants', *Genes & Development*, 17(1), pp. 49–63. Available at: https://doi.org/10.1101/GAD.1048103.

Tang, W. *et al.* (2016) 'The RNase PARN-1 Trims piRNA 3' Ends to Promote Transcriptome Surveillance in C. elegans', *Cell*, 164(5), pp. 974–984. Available at: https://doi.org/10.1016/J.CELL.2016.02.008.

Timmons, L., Court, D.L. and Fire, A. (2001) 'Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans', *Gene*, 263(1–2), pp. 103–112. Available at: https://doi.org/10.1016/S0378-1119(00)00579-5.

Tolia, N.H. and Joshua-Tor, L. (2007) 'Slicer and the Argonautes', *Nature Chemical Biology*. Nature Publishing Group, pp. 36–43. Available at: https://doi.org/10.1038/nchembio848.

Tsai, H.Y. *et al.* (2015) 'A ribonuclease coordinates siRNA amplification and mRNA cleavage during RNAi', *Cell*, 160(3), pp. 407–419. Available at: https://doi.org/10.1016/J.CELL.2015.01.010.

Tu, S. *et al.* (2014) 'Comparative functional characterization of the CSR-1 22G-RNA pathway in Caenorhabditis nematodes', *Nucleic Acids Research*, p. gku1308. Available at: https://doi.org/10.1093/nar/gku1308.

Uebel, C.J. *et al.* (2018) 'Distinct regions of the intrinsically disordered protein MUT-16 mediate assembly of a small RNA amplification complex and promote phase separation of Mutator foci', *PLOS Genetics*, 14(7), p. e1007542. Available at: https://doi.org/10.1371/JOURNAL.PGEN.1007542.

Vasale, J.J. *et al.* (2010) 'Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/ Argonaute pathway', *Proceedings of the National Academy of Sciences of the United States of America*, 107(8), pp. 3582–3587. Available at: https://doi.org/10.1073/pnas.0911908107.

Venkataraman, S., Prasad, B.V.L.S. and Selvarajan, R. (2018) 'RNA dependent RNA polymerases: Insights from structure, function and evolution', *Viruses*. MDPI AG. Available at: https://doi.org/10.3390/v10020076.

Wahba, L., Hansen, L. and Fire, A.Z. (2021) 'An essential role for the piRNA pathway in regulating the ribosomal RNA pool in C. elegans', *Developmental cell*, 56(16), pp. 2295-2312.e6. Available at: https://doi.org/10.1016/J.DEVCEL.2021.07.014.

Wang, G. and Reinke, V. (2008) 'A C. elegans Piwi, PRG-1, Regulates 21U-RNAs during Spermatogenesis', *Current Biology*, 18(12), pp. 861–867. Available at: https://doi.org/10.1016/j.cub.2008.05.009.

Wang, J. *et al.* (2011) 'Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles', *Genome Research* [Preprint]. Available at: https://doi.org/10.1101/gr.121426.111.

Wang, M.B. and Metzlaff, M. (2005) 'RNA silencing and antiviral defense in plants', *Current opinion in plant biology*, 8(2), pp. 216–222. Available at: https://doi.org/10.1016/J.PBI.2005.01.006.

Weng, C. *et al.* (2019) 'The USTC co-opts an ancient machinery to drive piRNA transcription in C. Elegans', *Genes and Development*, 33(1–2), pp. 90–102. Available at: https://doi.org/10.1101/GAD.319293.118/-/DC1.

De Wit, E. *et al.* (2009) 'Repertoire and evolution of miRNA genes in four divergent nematode species', *Genome Research* [Preprint]. Available at: https://doi.org/10.1101/gr.093781.109.

Witte, H. *et al.* (2015) 'Gene inactivation using the CRISPR/Cas9 system in the nematode Pristionchus pacificus', *Development genes and evolution*, 225(1), pp. 55–62. Available at: https://doi.org/10.1007/S00427-014-0486-8.

Xu, F. *et al.* (2018) 'A Cytoplasmic Argonaute Protein Promotes the Inheritance of RNAi', *Cell Reports*, 23(8), pp. 2482–2494. Available at: https://doi.org/10.1016/j.celrep.2018.04.072.

Zeng, C. *et al.* (2018) 'Differential phase partition of a PICS complex is required for piRNA processing and chromosome segregation in C. elegans', *bioRxiv*, p. 463919. Available at: https://doi.org/10.1101/463919.

Zhang, C. *et al.* (2011) 'mut-16 and other mutator class genes modulate 22G and 26G siRNA pathways in Caenorhabditis elegans', *Proceedings of the National Academy of Sciences of the United States of America*, 108(4), pp. 1201–1208. Available at: https://doi.org/10.1073/PNAS.1018695108.

Zong, J. *et al.* (2009) 'Evolution of the RNA-dependent RNA polymerase (RdRP) genes: Duplications and possible losses before and after the divergence of major eukaryotic groups', *Gene*, 447(1), pp. 29–39. Available at: https://doi.org/10.1016/J.GENE.2009.07.004.