



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN
SCIENZE DELLA TERRA, DELLA VITA E DELL'AMBIENTE

Ciclo 38

Settore Concorsuale: 05/B1 - ZOOLOGIA E ANTROPOLOGIA

Settore Scientifico Disciplinare: BIO/05 - ZOOLOGIA

A NOVEL MITOCHONDRIALLY DRIVEN RNA INTERFERENCE MECHANISM:
DISCOVERY, EVOLUTION AND REGULATORY EFFECTS OF SMITHRNAS

Presentata da: Diego Carli

Coordinatore Dottorato

Barbara Cavalazzi

Supervisore

Marco Passamonti

Co-supervisore

Federico Plazzi

Esame finale anno 2026

Abstract

Intergenomic communication is imperative for eukaryotic life, particularly the dialogue between the nucleus and mitochondria. These coordinative activities rely on diverse molecular components. A prime example is RNA interference (RNAi), where protein complexes are guided by nucleic acid complementarity to achieve various effects on gene networks. Thus, small RNAs (sRNAs) are emerging as critical mediators in this wider process. This thesis investigates a novel class of molecules, named small mitochondrial highly-transcribed RNAs (smithRNAs). It proposes them as an essential and extensive mechanism of cellular regulation, unveils their origins, and explores a functional parallel in a bacterium-host endosymbiotic system.

A primary challenge hindering the probing of smithRNAs was the lack of standardized methods. I addressed this issue by developing and validating a robust and reproducible bioinformatic workflow designed to identify candidate smithRNAs and their nuclear targets. The pipeline's efficacy was confirmed by successfully reproducing and standardizing the discovery of smithRNAs in bivalve species where they were first identified, which in turn supported its application in the following analyses.

Subsequently, I conducted the first large-scale comparative genomic analysis of smithRNAs across a diverse panel of 14 metazoan species. This study revealed that smithRNAs are not a bivalve anomaly but a common and widespread feature of Metazoa. My findings show that in most animals, smithRNAs originate preferentially from tRNA and rRNA genes, in contrast to the non-coding regions seen in bivalves. This observation aligns with the increasing interest in tRNA fragments (tRFs), which are known to be involved in development and signalling. The functional predictions for their nuclear targets are highly varied and lineage-specific, pointing to smithRNAs as a dynamic and rapidly evolving regulatory system.

Building on the concept of sRNA-mediated intergenomic control, I inspected its evolutionary and functional analogues in a cross-kingdom context: the *Wolbachia pipientis* infection of *Zyginidia pullula* is a peculiar system where the bacterial parasite drives host feminization. My studies revealed that *Wolbachia* profoundly alters its host transcriptome, shifting the male molecular profile towards a female phenotype. I identified bacterial sRNAs, originating primarily from a highly conserved non-coding region, consistent with previous clues of independent transcription of these sequences. Secondly, they arose from tRFs, showing clear similarities to the proposed genesis of the novel smithRNAs. This bacterial-derived RNAi is predicted to target host genes involved in several biological routes such as spermatogenesis, immunity, and behavior, thereby suggesting a direct molecular mechanism for *Wolbachia*'s known manipulation of host biology.

Collectively, these results provide comprehensive evidence that sRNA-mediated signaling is a conserved and influential layer of biological regulation. From mitochondrial-nuclear communication to cross-kingdom interactions, these molecules act as powerful intermediaries capable of shaping symbiotic evolution and influencing host phenotypes, from the modulation of cellular pathways to the alteration of complex behaviours and systems.

Index

1. Introduction	3
1.1 Aims of the thesis	7
2. A workflow for the identification of novel smithRNAs	9
2.1 Introduction	9
2.2 Implementation	13
2.3 Results	18
2.4 Discussion	33
3. SmithRNAs as a common feature among Metazoa	36
3.1 Introduction	36
3.2 Results	39
3.3 Discussion	48
3.4 Materials and Methods	53
3.5 Supplementary Materials	57
4. Wolbachia pipientis RNA interference in the leafhopper Zyginidia pullula	64
4.1 Introduction	64
4.2 Results	66
4.3 Discussion	74
4.4 Materials and Methods	78
4.5 Supplementary Materials	81
5. Conclusion	88
5.1 Future perspective	91
6. References	93

1. Introduction

Intergenomic communication within the Eukaryotic Cell

The eukaryotic cell is widely understood to have arisen from a pivotal endosymbiotic event, resulting in a chimaeric organism from the integration of archaeal and bacterial lineages. Although the archaeal ancestor provided fundamental cellular machinery and information-processing systems (Spang et al. 2015), the subsequent evolution of eukaryotic complexity was critically dependent on the profound bioenergetic advantage conferred by the endosymbiont (Lane 2018). As for other symbiotic bacteria (McCutcheon and Moran 2012), the proto-mitochondrion progressively underwent extensive genome reduction. This energetically convenient process resulted in the retention of only a small fraction of the ancestral genetic repertoire still observed in their free-living alphaproteobacterial relatives (Timmis et al. 2004; Kelly 2021). The persistence of a residual organellar genome, rather than its complete elimination, is predominantly attributed to the Co-location for Redox Regulation (CoRR) hypothesis, which posits that core mitochondrial encoded proteins are crucial for rapid and precise energetic modulation (Allen 2015). Together with protein hydrophobicity and GC content, it will explain the common shared set of coding genes found in mitochondria (Johnston and Williams 2016). Notwithstanding the relatively high conservation of these genes, mitochondrial genomes display a great variability in terms of structure, gene content and genome size (Smith and Keeling 2015). This phenomenon is most pronounced within Viridiplantae, while it is less prominent in other kingdoms, like Fungi. Metazoa, despite common beliefs, do not exhibit constant values, but rather portray notable exceptions, especially in terms of genome size (Formaggioni et al. 2021). In fact, large deviations from the norm have been found, usually for the proliferation of non-coding regions. While in plants it can constitute up to 99% of the mitochondrial genome, its effect in driving the size variation can be observed for animals too: in bivalve molluscs, for example, the mitogenome ranges from approximately 14 to 67 kilobases (Ghiselli, Iannello, et al. 2021).

As a consequence of its limited coding capacity, the mitochondrion is critically dependent on the nuclear genome. The vast majority of its proteome is encoded by nuclear genes and imported post-translationally to perform the essential functions of the organelle (Schmidt et al. 2010). In conjunction with transcription factors and other components of the anterograde signaling, they constitute the primary mechanism for monitoring mitochondrial health, copy number, and morphology.

To maintain cellular homeostasis, the mitochondrion must signal its metabolic and physiological state back to the nucleus. This process, which constitutes the mitochondrial retrograde response (MRR), allows the cell to adapt to mitochondrial dysfunction or changing energetic demands. Ca^{2+} , O_2 , metabolic products and reactive oxygen species have been all identified as retrograde signals, affecting transcription of nuclear gene expression (Kleine and Leister 2016). Moreover, there is evidence for small peptides and mitochondrial protein not involved in the ATP production which have been linked to retrograde signalling. The first class is recognized to have protective functions against aging and various disease (Lee et al. 2013; Cobb et al. 2016), while the second one have no clear homologies or functions and seems to be restricted to some animal lineages (Mitchell et al. 2016; Niais et al. 2021; Selifanova et al. 2023).

Mitochondrial non-coding retrograde response

Protein-based signaling is but one facet of biological regulation (Mattick and Makunin 2006). The discovery of a complex transcriptional landscape of mitochondria, yielding a variety of non-coding RNAs, suggests that these transcripts fulfil multiple functional roles (Dietrich et al. 2015). At a basic level, they can act intra-mitochondrially, regulating their complementary transcripts through base-pairing. Beyond this local function, evidence indicates that these RNAs can be exported to the nucleus, and their expression linked to various pathological conditions (Burzio et al. 2009; Kumarswamy et al. 2014). Of particular interest are the recently discovered mitochondrially-encoded small non-coding RNAs (sRNAs), although their precise biological functions remain to be elucidated (Meseguer 2021). Most investigations into these molecules have concentrated on mammalian systems due to their clinical relevance, leaving a broader, comparative analysis across diverse taxa wanting. Within these model species, which are characterized by compact and streamlined mitochondrial genomes, the identified sRNAs have been reported to originate from the D-loop region or tRNA structures (Meseguer 2021). In particular, a specific class of these molecules, named small mitochondrial highly-transcribed RNAs (smithRNAs), was first identified as originating from non-coding regions in the bivalve *Ruditapes philippinarum* and was demonstrated to regulate nuclear transcripts (Pozzi et al. 2017). The initial characterization of smithRNAs linked them to fundamental biological processes related to sex determination in other Metazoa (Pozzi et al. 2017). Indeed, subsequent *in vivo* validation revealed their ability to influence the nuclear epigenome by regulating histone methylation and acetylation (Passamonti et al. 2020). Their potential significance is amplified within

the context of Doubly Uniparental Inheritance (DUI), a unique mitochondrial vertical transmission system restricted, but frequently reported, in bivalves (Zouros 2013; Passamonti and Plazzi 2020). This atypical mode maintains two distinct mitochondrial lineages that are inherited separately through the male and female germlines, leading to independent evolutionary trajectories and substantial genetic divergence (Doucet-Beaupré et al. 2010). The convergence of these factors, coupled with the presence of extensive non-coding regions, constitutes a fertile evolutionary landscape for smithRNAs to acquire a role in modulating reproductive pathways, a hypothesis strongly supported by their recent identification in *Potamilus streckersoni* (Smith et al. 2023). Moreover, smithRNAs have a marked propensity to evolve new functions (Plazzi et al. 2024), which would promote them as a dynamic way to control cellular activity.

Endosymbiotic RNA Interference towards hosts

Mitochondrial non-coding based communication would therefore embody a discrete mode of RNA interference (RNAi). In recent years, RNAi has garnered increasing interest as a sophisticated layer of post-transcriptional regulation (Ghildiyal and Zamore 2009). RNAi is a radical biological process in which sRNA molecules, such as microRNAs (miRNAs) and small interfering RNAs (siRNAs), guide protein complexes to target specific nucleic acid sequences for silencing (Ghildiyal and Zamore 2009; Shabalina and Koonin 2008). The widespread conservation of RNAi machinery points to its ancient evolutionary origins, where it likely first served as a defense mechanism against invasive genetic elements like viruses and transposons (Koonin 2017). The efficacy of RNA interference is established upon the principle of Watson-Crick base pairing, the elementary rule governing nucleic acid interactions. This universality dictates that an sRNA's regulatory potential is not confined to its genome of origin. Although the protein effectors of RNAi pathways diverge between prokaryotes and eukaryotes (Shabalina and Koonin 2008), the underlying mechanism, wherein an sRNA guides a regulatory complex to its target, is functionally conserved. Therefore, sRNA can, in principle, target any complementary sequence, irrespective of its genomic source. This property transforms RNAi from a mechanism of mere endogenous gene regulation into a potential medium for communication and interaction between separate genetic entities. It provides a molecular basis for a "genetic cross-talk" that can occur between different organisms in symbiotic or parasitic relationships (Felden and Augagneur 2021). Thus, it is not unexpected to find them between organelles and the nucleus (Cognat et al. 2017). While bacterial sRNAs commonly function endogenously by associating with Hfq

and ProQ chaperones to regulate internal gene transcription (Melamed et al. 2020), they can also be exported to function exogenously within a host cell, taking advantage of the host's RNAi machinery (Sahr et al. 2022). Recent studies confirm that parasitic bacterial sRNAs are able to target the host's immune system, allowing bacteria to evade host defence response (Moriano-Gutierrez et al. 2020; H. Zhang et al. 2020; Sahr et al. 2022). Compellingly, similar RNA-mediated mechanisms are employed by *Wolbachia* (Mayoral et al. 2014), an endosymbiont belonging to the same bacterial lineage that gave rise to mitochondria (Wang and Wu 2014). *Wolbachia* is a master manipulator of arthropod biology (Sullivan and Porter 2023), inducing profound phenotypic alterations to facilitate its maternal transmission. Among the repertoire of reproductive modification caused, *Wolbachia*-induced feminization is the only one which does not occur during the syncytial and cellularized embryonic stages (Sullivan and Porter 2023). Feminization represents a precise intervention where the bacterium uncouples the host's phenotypic development from its genotypic sex. Such molecular specificity might be the reason why these effects are so taxonomically limited, having them been observed solely in a few species of the Lepidoptera, Isopoda and Hemiptera, orders (Bouchon et al. 1998; Negri et al. 2006; Kern et al. 2015). Adding to that, *Wolbachia* influences other host physiological traits and abilities, suggesting a broad capacity to regulate diverse molecular pathways (Peng et al. 2008; Maistrenko et al. 2016; Vandana et al. 2024), not strictly related to reproduction.

Thus, numerous parallels can be drawn between mitochondria and endosymbiotic prokaryotes, especially those who shared a phylogenetic history and ecological demands with the APT-producing organelle. The most striking of these is the communal ability to swirl sex of its host, likely with the support of sRNAs. Even if they represent a small portion of the many connections and tools engaged in the intergenomic communication, unveiling the basis of these mechanisms would shed light on an underestimated propriety with clear biological and medical relevance.

1.1 Aims of the thesis

The overarching goal of my doctoral research would have been to investigate the role of small non-coding RNAs as a universal medium of cross-talk in eukaryotic cells. This thesis tried to address essential questions regarding the prevalence, function, and evolutionary significance of sRNA-mediated interchange by examining both intracellular mitochondrial-nuclear systems and their bacterium-host equivalent.

A significant initial bottleneck in this field was the lack of standardized methods to reliably identify mitochondrially-encoded small RNAs from complex sequencing datasets. In spite of many others being able to detect sRNAs signatures (Friedländer et al. 2012; Rueda et al. 2015), a comprehensive methodology was required to specifically discriminate true biological signals from background transcriptional noise. Based on the framework which first led to the disclosure of smithRNAs (Pozzi et al. 2017), the earliest aim of this thesis was to develop a dedicated and solid bioinformatic workflow, for the systematic identification of candidate smithRNAs and the prediction of their nuclear targets. This reproducible pipeline provides the essential methodological foundation upon which I built further inquiries.

The research then proceeded to perform a wide, comparative analysis across Metazoa, with the scope of determining the occurrence, genetic origin, and putative roles of smithRNAs. By applying the SmithHunter workflow to a diverse panel of 14 species, this large-scale exploration sought to test the hypothesis that smithRNAs are a conserved feature of animal biology, providing the first panoramic view of these molecules as a potentially fundamental mechanism of the MRR. I scouted up mitochondrial encoded small RNAs for all the inspected organisms, which originated principally from tRNA structures. This observation is consistent with the low frequency of mitochondrial non-coding regions of the considered species. Additionally, I could not delineate a single, unifying function for these elements. Rather, their high evolutionary dynamism implies they fulfil diverse and lineage-specific roles, reinforcing their proposed proneness for rapid functional evolution (Plazzi et al. 2024).

Finally, to trace back the origin of such a potentially important mechanism, I explored the versatility of sRNA-mediated communication in a different biological context of endosymbiosis. The *Wolbachia pipientis* manipulation of the leafhopper *Zyginidia pullula* provides an exemplary system for this purpose. This detailed case study of cross-kingdom RNAi, which investigates how a bacterial

symbiont can hijack host cellular pathways to induce feminization, gives a powerful functional analogue to the mitochondrial-nuclear communication system. This study confirms *Wolbachia* utilizes sRNAs to adjust its host gene expression according to its needs. My findings not only confirm a highly conserved sRNA originating from non-coding region (Woolfit et al. 2015), but identify others which would target key players in reproduction, immunity and behavioural processes.

Collectively, my approach integrates bioinformatic software development, large-scale comparative genomics, and an accurate probe to illuminate a previously underappreciated mode of biological regulation. The underlined RNAi would be capable of carrying genetic information across the boundaries that separate genomes, thereby shaping the cooperative and conflicting interactions that define life.

2. A workflow for the identification of novel smithRNAs

This chapter has been published in the journal *BMC Bioinformatics*. Supplementary materials are available via the link to the original article (<https://doi.org/10.1186/s12859-024-05909-0>).

2.1 Introduction

Proteins must interact with each other to assist in the dynamic processes of living cells in a functional organism. However, the intricate dance of protein interactions often deviates from the proportions dictated by their genome occurrences. Consequently, the modulation of gene transcription becomes a critical factor in maintaining the balance necessary for proper cellular function. One possible mechanism for achieving this equilibrium is through post-transcriptional modifications, a process frequently involving small noncoding RNAs—short fragments capable of modulating gene expression by silencing genes (Ghildiyal and Zamore 2009). Among these entities, microRNAs (miRNAs) have emerged as some of the most pervasive gene-regulatory molecules in the animal kingdom (Moran et al. 2017; Bartel 2018; Szczepanek et al. 2018; Biswas et al. 2021; Formaggioni et al. 2024).

Elements such as miRNAs guide the RNA-induced silencing complex (RISC) to regulate the translation of specific mRNAs through sequence complementarity (Shabalina and Koonin 2008; Ghildiyal and Zamore 2009; Bofill-De Ros et al. 2020) and their post-transcriptional silencing activity extends to various developmental processes and diseases (Jiao et al. 2014; Larriba and Del Mazo 2016; Riggs et al. 2018; Wang et al. 2018; Kim and Lee 2019).

Notably, miRNAs have been predominantly studied in the context of nuclear-nuclear interactions (i.e., miRNAs encoded in the nuclear genome that modulate the expression of nuclear transcripts), although some mitochondrial targets have also been described for miRNAs encoded in the nuclear genome (Li et al. 2012). At the other extreme, mitochondrial-mitochondrial interactions have been described, where mitochondrially-encoded microRNAs can actually regulate gene expression in the mitochondrion (Ro et al. 2013; Fan et al. 2019; Paramasivam and Vijayashree Priyadharsini 2020). Only limited consideration, in turn, has been given to the mitochondrial genome as a potential source of RNA interference acting on the nuclear genome (Pozzi et al. 2017).

In animal cells, mitochondrial DNA (mtDNA) is a small (~16 Kb) molecule, that is usually characterized by the absence of introns, a circular structure, and a conserved repertoire of 13 protein-coding genes, two ribosomal genes and 22 tRNA-coding genes (Boore 1999), although many

anomalies are known (Formaggioni et al. 2021). Molluscan mtDNA is unique in many respects (Ghiselli, Gomes-dos-Santos, et al. 2021) including, among others, the Double Uniparental Inheritance (DUI), a peculiar process of vertical transmission to offspring observed in bivalves (Zouros and Rodakis 2019; Passamonti and Plazzi 2020). Nevertheless, it generally conforms, at least in its structure, to the model previously described for animals (Zouros and Rodakis 2019; Passamonti and Plazzi 2020; Ghiselli et al. 2021). Given its unique transcription mechanism, which involves the production of long transcripts that are further cleaved to produce single gene transcripts and liberate functional RNA molecules, like rRNAs and tRNAs (D'Souza and Minczuk 2018), it is reasonable to hypothesize that the mitochondrial genome may serve as an efficient source of miRNA-like molecules (Plazzi et al. 2024). This possibility has been explored in the Manila clam *Ruditapes philippinarum*, which was selected as a model species for the study of mitochondrially encoded microRNAs for several reasons. First of all, it harbors two genetically distinct mitochondrial genomes, one per sex, inherited according to the DUI model (Passamonti and Scali 2001; Passamonti and Plazzi 2020), thus allowing a solid establishment of the mitochondrial source of miRNAs by comparing the small transcriptome against a male or female background (Pozzi et al. 2017). Moreover there is evidence that the mitochondrial genome (namely, the presence and activity of either the male or the female genome in the developing embryos) is involved in sex determination (Milani et al. 2016). Finally, it is worth noting that, at variance with the typical metazoan mitochondrial DNA, bivalve mitochondrial genomes are characterized by large intergenic spacers and unassigned regions (Breton et al. 2014; Plazzi et al. 2016), that may be deployed to develop novel roles in the cell, including, following maturation, regulatory RNAs (Plazzi et al. 2024).

The small transcriptome of the Manila clam has been characterized in detail, and multiple highly transcribed small RNAs of mitochondrial origin have been identified (Pozzi et al. 2017) *in silico*. Two of these were further validated *in vivo* through RNA injection experiments that demonstrated their biological activity (Passamonti et al. 2020). Results of these studies have highlighted an intriguing interplay between mitochondrial and nuclear transcripts, possibly leading to gonad formation in *R. philippinarum* (Pozzi et al. 2017; Passamonti et al. 2020).

This, in turn, led to the proposal of smithRNAs (Small MITochondrial Highly-transcribed RNAs) as a novel class of small RNAs defined as (a) of mitochondrial origin, (b) highly transcribed, and (c) regulating a nuclear transcript (Pozzi et al. 2017). While (b) is somewhat arbitrary and (c) requires that

their function is confirmed experimentally by RNA injection, this provides a clear definition of the class. Worth of note, some overlaps are envisionable with other classes of small RNAs, such as tRNA fragments (tRFs) (Meseguer 2021; Shaukat et al. 2021), rRNA fragments (rRFs) (Chen et al. 2017) and degradation fragments (Xu et al. 2019; Jin et al. 2020). Nevertheless, the unique combination of expression, origin and function provides a clear and workable definition of this novel small RNA class. Concurrently, other studies have proposed the involvement of smithRNAs in sex determination in the bivalve *Potamilus streckersoni* (Smith et al. 2023). One smithRNA, encoded in the male mitochondrial genome, was identified and predicted to target a nuclear transcript that is both differentially regulated in males, compared to females, and presumably involved in female development. Henceforth, while not biologically validated in the strict sense as for *R. philippinarum*'s smithRNAs (Passamonti et al. 2020), this smithRNA receives substantial support and, while it is never called so (Smith et al. 2023), its features nicely conform to the definition of smithRNAs introduced above.

While still awaiting *in vivo* validation, the presence of smithRNAs have been further suggested, based on bioinformatic analyses, in *Danio rerio*, *Drosophila melanogaster* and *Mus musculus* (Passamonti et al. 2020), where they appear to be characterized by a high degree of sequence conservation in line with other functional mitochondrial loci. Adding to their significance, it has been proposed that new smithRNAs can readily evolve from mitochondrial RNAs through an exaptation process (Plazzi et al. 2024). This evolutionary trajectory, combined with a remarkably high probability of finding nuclear targets (Plazzi et al. 2024), underlines the central role of sncRNAs in mediating the elaborate interaction between mitochondrial and nuclear genomes during metazoan evolution. Nevertheless, the question remains open whether smithRNAs are a species-specific mechanism related to sex determination in the Manila clam, or rather a mechanism of more general interest, possibly shared by the entire Metazoa.

To evaluate the potential of this phenomenon as a novel signaling pathway in the broader context of mito-nuclear cross-talking, a comprehensive investigation across a panel of representative metazoan species was therefore imperative. This systematic approach contributed to the understanding of the functional implications of retrograde mitochondrial RNAi across animal lineages, with implications going as far as to the origin of the eukaryotic cell (Plazzi et al. 2024).

To this end, I developed SmithHunter, a new workflow designed for the identification and characterization of candidate smithRNAs. The pipeline reproduces, in a unified workflow, the original

procedure used to identify smithRNAs in the Manila clam (Pozzi et al. 2017) and in other metazoans (Passamonti et al. 2020). Improvements in this implementation rely on the possibility of using replicate samples, remapping on the nuclear genome to select reads of unequivocal mitochondrial origin, improved clustering and cluster filtering methods, and computation of free energy of pre-smithRNAs secondary structures. Moreover, multiple analytical parameters of SmithHunter can be customized, allowing users to adapt the analysis to the studied system.

2.2 Implementation

Overview

The pipeline is composed of two main components (FIGURE. 1). The first module is essentially data-driven and focuses on the identification and filtering of presumptive smithRNA sequences, defined as centroids of clusters with significant transcription levels and narrow transcription boundaries. One or more small RNA libraries (replicates), the sequence of the mitochondrial genome and, optionally, the sequence of the nuclear genome of the species of interest, are used as inputs. The main output is a list of presumptive smithRNA sequences, as well as graphics depicting: (a) raw coverage over the mitochondrial genome, (b) cluster position/abundance on the mitochondrial genome, and (c) 5' and 3' transcription end conservation.

The second module is more predictive in nature and deals with the identification of possible nuclear targets and pre-miRNA-like precursor structures for presumptive smithRNAs. The list of smithRNAs identified by the first module and the transcriptome of the species of interest, with annotated 5' and 3' UTR regions, are used as inputs. The main output is a list of nuclear transcripts putatively targeted by individual smithRNAs, information regarding the Gibbs free energy (dG) of smithRNA/target pairs, and putative precursor structures. After passing filters in the first module, sequences that find a putative nuclear target in the second module are regarded as candidate smithRNAs. Presumptive and candidate smithRNAs are given a unique identifier in the form of a string reporting, in order, the cluster number, its depth, start and end position with respect to the genome and strand.

The main pipeline is written in bash and R (R Core Team 2025), with calls to external software. The pipeline setup entails cloning of the GitHub repository and running an installation script that will install SmithHunter as well as, if required, conda (Anaconda Inc 2020) and PITA (Kertesz et al. 2007). Additionally, the installation script will create a conda environment within which SmithHunter is executed. The two modules are invoked independently via command line and all inputs (file paths, options) are specified as command line arguments. The pipeline was created and tested on the Ubuntu 22.04.3 LTS platform, but is portable to any Linux OS provided that the conda environment can be successfully created and PITA can be installed. Detailed installation and usage information, with examples, can be found in the project's GitHub repository.

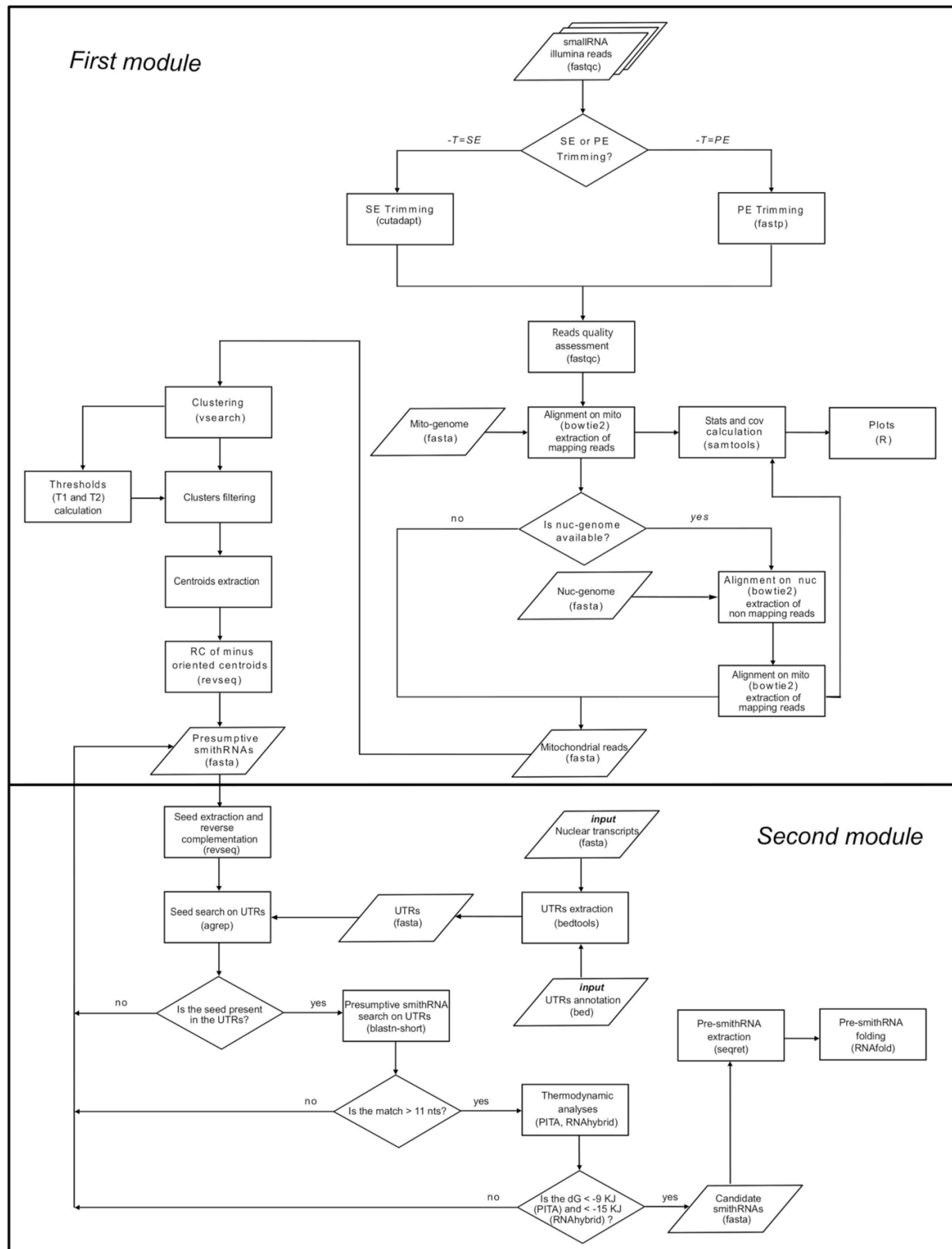


FIGURE 1: SmithHunter workflow: first module (data to presumptive smithRNAs) and second module (presumptive to candidate smithRNAs).

First module

Sequence reads (multiple replicates are accepted) are preprocessed using CUTADAPT v. 4.3 (Martin 2011) if single-end (SE) or fastp v. 0.23.2 (Chen et al. 2018) if paired-end (PE). This behavior is selected with option `-T`, which takes as values either SE, for single end, or PE, for paired end. Value NO can be used for pre-trimmed data. In PE mode, reverse reads are exploited to correct overlapping regions in forward reads and subsequently discarded. Forward and reverse adapters are specified via options `-a` and `-A`, respectively.

The trimmed reads are initially aligned to the mitochondrial genome with Bowtie2 v. 2.5.1 (Langmead and Salzberg 2012) using the end-to-end option and allowing a single nucleotide mismatch. Mitochondrion-mapping reads are extracted with SAMtools v. 1.13 (Li et al. 2009) and aligned, as above, to the nuclear genome, if available. This step is justified by the need to exclude reads that map to both the mitochondrial and the nuclear genome, as these may originate from nuclear mitochondrial pseudogenes (NUMTs) (Bensasson et al. 2001) and therefore their mitochondrial origin cannot be assured. In the absence of a nuclear genome sequence for the species under study, this step is not performed. Nuclear nonmapping reads, representing bona fide mitochondrial-unique reads, are eventually remapped on the mitochondrial genome for counting and coverage calculation. Alignment statistics are calculated at each step using the flagstats and depth modules of SAMtools and mitochondrial-unique alignments are produced from each remapping file using the bamtobed module of BEDtools v. 2.30.0 (Quinlan and Hall 2010).

Mitochondrial-unique reads are sorted by abundance and clustered using VSEARCH v. 2.22.1 (Rognes et al. 2016). The percent identity used for clustering can be specified via option `-I`. Clusters are then subjected to several filters. To discard clusters characterized by low copy number and/or variable expression, two abundance thresholds are calculated based on the empirical distribution of cluster depths and the stringency parameter defined by option `-S`. One (T1, global) is calculated on combined reads and the second (T2, possibly different across replicates) is calculated on reads from individual replicates. Both thresholds are defined as the percentile of order S of the relevant distribution of unique cluster sizes. This definition of abundance thresholds is inherently different from the manually defined, hard coded, threshold of > 200 reads of the original paper (Pozzi et al. 2017). This option, nevertheless, has the advantage that different thresholds are applied to different datasets,

thus accounting for differences in sequencing depth across species/replicates while retaining comparability across different datasets by the use of a unique stringency value.

Clusters not reaching T1 overall are discarded. Clusters are then filtered based on a minimum number of replicates where each cluster reaches T2. The number of replicates can be specified by the user via option `-M`. This step is different from the previous method (Pozzi et al. 2017), where data from different replicates were combined (equivalent to `-M 1`). I discussed later in the appropriate section the utility of setting `-M` depending on the nature of the samples and the expectations of the user. All the clusters not passing both filters are discarded, while the others are retained as multifasta files. Representative sequences for each cluster, selected as the most abundant within the cluster, are extracted and those mapping in minus orientation are reverse/complemented using the `revseq` module of `EMBOSS` (ver. 6.6.0.0; Rice et al. 2000) to obtain smithRNAs in 5' to 3' orientation. Custom R scripts are used to generate coverage plots of the mitochondrial alignments and visualize cluster size and distribution, as well as the distribution of the 5' and 3' ends of the reads within each cluster (i.e., transcription end conservation; equivalent to Figure 3 in (Pozzi et al. 2017)).

The endpoint of this first module is a multifasta of presumptive smithRNAs in 5' to 3' orientation. This list can be manually inspected and edited. Among other possible ad hoc analyses, I envision the possibility that the user may visualize the plot of transcription end conservation and manually select only those smithRNAs that appear to have the tightest 5' and 3' transcription boundaries. A script that automatically identifies smithRNAs with narrow 5' and 3' ends is provided to optionally help the user in this step (see GitHub repository for its documentation).

Second module

5' and 3' UTRs of nuclear transcripts are extracted from the transcriptome sequence as a multifasta file based on UTR annotations provided as a BED file. Seed regions (nucleotides 4–10, as in (Pozzi et al. 2017), see below) of each smithRNA are identified, reverse/complemented and searched against 5' and 3' UTRs of nuclear transcripts using the approximate `grep` algorithm (Wu and Manber 1992). For each matching target, UTR regions are converted in a BLAST database and the full length sequence of the smithRNA is used as query in a BLAST search executed with the `blastn-short` option (S. F. Altschul et al. 1990). A minimum of 11 nucleotides aligned on the minus strand (alignment length) are necessary for the presumptive smithRNA and target pair to pass to the next step. Following the

observation that the default e-value threshold can bias the results in favor of smithRNAs with a limited number of targets (i.e. due to differences in database size) the e-value threshold was removed and filtering is therefore based on seed matching, alignment length and RNA/RNA stability.

While the previous definition of the seed as nucleotides 4–10 was retained, I acknowledge that different options exist. Furthermore, while I tentatively assimilate smithRNAs with miRNAs, the exact fine scale molecular processes involved in smithRNA-target interaction are largely unknown. Through sequence complementarity, miRNAs and their targets interact into a seed region represented by nucleotides 2–8 of the 5' region of the miRNA and the 3' untranslated region (UTR) of the target mRNA (Lewis et al. 2003). Although a perfect seed match between the miRNA and its target is generally needed, noncanonical matches, as well as cases where the seed is shifted, have been reported and appear to be common (Shin et al. 2010; Tan et al. 2014; Bartel 2018). Moreover, central seed pairing has been reported to be more predictive in mammals (Shin et al. 2010). To overcome a fixed definition of the seed region, and take into account noncanonical seed matches and seed shifts, I implemented the possibility to define the seed region (options -X and -Y for the first and last nucleotides, included) and to allow 0 (by default) to 2 mismatches in the seed region (option -m).

Presumptive smithRNAs with evidence of similarity with UTRs in the seed region are subjected to additional thermodynamic analyses using PITA v. 6 (Kertesz et al. 2007) and RNAHybrid v. 2.1.2 (Krüger and Rehmsmeier 2006)). Those smithRNA-target pairs with dG levels < -15 kJ and 3–10 helix constraints from RNAhybrid, as well as DG < -9 kJ from PITA, are eventually retained. The former threshold (< -15) differs from the threshold (< -20) applied in previous studies (Pozzi et al. 2017; Passamonti et al. 2020). This finds a justification in the observation that the dG values calculated here are marginally, but consistently lower than those reported in Pozzi et al. 2017. This observation, coupled with the fact that the current workflow is aimed at finding candidate smithRNAs that need to be further validated biologically, and therefore a false positive is better tolerated than a false negative, suggested that a less stringent position was to be taken in this respect.

Presumptive smithRNAs that find at least one target in the nuclear transcriptome are further referred to as candidate smithRNAs. Putative pre-smithRNA sequences of candidate smithRNAs are extracted using the seqret module of EMBOSS (Rice et al. 2000). Under the assumption that, in line with miRNAs, mature RNAs can be located on either strand of the pre-miRNA hairpin, I tentatively identified pre-smithRNA sequences as either region -15 to +50, or region -50 to +15, relative to the

smithRNA. The secondary structure of both possible pre-smithRNAs is computed using RNAfold v. 2.3.3 (Lorenz et al. 2011) at the default folding temperature of 25 °C and both structures are reported. Users can adjust the folding temperature through the $-R$ parameter and pre-smith coordinates with the -1 and -2 parameters. The identification of putative pre-smithRNA regions differs from that of Pozzi et al. 2017 and Passamonti et al. 2020, who, in turn, conducted this calculation on a region that was manually selected based on gene annotation of the mitochondrial genome. Briefly, the entire noncoding region encompassing the smithRNA was selected in the case of smithRNA that were found within a noncoding region, while the entire tRNA, plus neighboring small noncoding regions, was selected in the case of smithRNAs that were found within a tRNA. This is a sensible option in the Manila clam, whose mitochondrial genome is characterized by multiple large noncoding regions, and has been convincingly justified by the hypothesis that noncoding regions, released during the splicing of the initial mitochondrial transcript, may actually act as pre-smithRNAs (Plazzi et al. 2024). However, this approach does not seem feasible in general, as metazoan mitochondrial genomes are almost invariably devoid of intergenic noncoding regions of sufficient size. As such, I revert to a more standard working definition of putative pre-smithRNAs that does not rely on the presence of large intergenic spacers.

2.3 Results

Fourteen candidate smithRNAs were identified in the Manila clam (Pozzi et al. 2017). Laterally, the activity of two of these genes was validated *in vivo* by Passamonti et al. 2020, thus establishing this species as the reference organism for smithRNA studies. Given the absence of alternative software for smithRNA detection, evaluating the reproducibility of the results obtained by Pozzi et al. for the Manila clam, with special attention given to the two validated smithRNAs, appears to be the most effective way to assess and present the functionalities of the SmithHunter pipeline.

I reproduced the analysis performed by Pozzi et al. using the same data, with minimal modifications due to specific SmithHunter functionalities and parameters. Specifically, raw reads of the small transcriptome of *R. philippinarum* were downloaded from the NCBI Short Read Archive (SRA; accession numbers SRR3662624-SRR3662629), while male and female mitochondrial genomes, as well as the nuclear genome of the species, were recovered from GenBank (accession numbers

AB065375.1, AB065374.1, and GCA_026571515.2, respectively). The original, unannotated transcriptome (Ghiselli et al. 2012) was downloaded from GenBank (accession numbers JO101212-JO124029). In the absence of the original UTR annotations, updated annotations, produced by the same authors, were retrieved from (Ghiselli and Iannello 2015). The SmithHunter analysis was performed, independently and in parallel, on the male and female small transcriptome and mitochondrial genomes. Note that the species name *Venerupis philippinarum*, used in some records, is a synonym of *Ruditapes philippinarum* (World Register of Marine Species database).

Additional testing was performed on data from *P. streckersoni*, focusing on a male-related smithRNA that, while not biologically validated in strict terms, received substantial support as a masculinizing agent and regulator of the nuclear transcript for GCNT1 (Smith et al. 2023). Raw reads of the small transcriptome were obtained from SRA (accession numbers SRR23195578-SRR23195582, SRR23195559), while the male and female mitochondrial genomes, as well as the nuclear genome, were recovered from GenBank (accession numbers ON881148, MW413895, and JAEAOA01, respectively). Transcriptome UTR sequences were in turn retrieved from Supplementary Materials (GitHub: <https://github.com/raqmejtru/mitonuclear-sd>, Smith et al. 2023). The SmithHunter analysis was performed, as above, independently and in parallel on the male and female small transcriptome and mitochondrial genomes, under parameters designed to replicate Smith et al. as well under a more stringent parameter set stemming from parameter optimization.

Trimming and remapping

A total of approximately 69 and 72 million reads belonging to male and female individuals of *R. philippinarum*, respectively, (21 to 27 million in individual replicates) were analyzed using the procedure implemented in the first module of SmithHunter. Reads were trimmed in SE-mode (option -T SE) and aligned to the mitochondrial genomes of both sexes, as well as to the nuclear genome of the species. There were 207,149 and 276,748 reads mapped on the mitochondrial genome for the male and female, respectively, representing 0.31% and 0.39% of the trimmed reads. Among these, 187,735 and 259,719, representing 0.28% and 0.36%, respectively, of the trimmed reads, were identified as mitochondrial-unique reads: i.e., they did not align to the nuclear genome (TABLE S1, FIGURE 2).

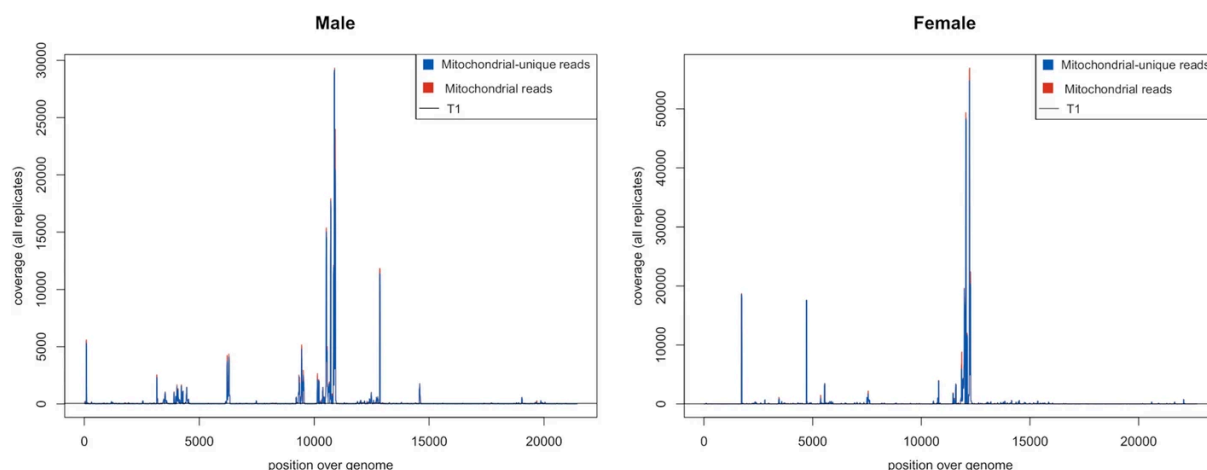


FIGURE 2: Remapping of the small transcriptome over the male and female mitochondrial genomes. Coverage along the mitochondrial genome (all replicates, combined) is shown based on mitochondrial reads (red) and mitochondrial-unique reads (blue). The black horizontal line represents the T1 coverage threshold ($-S\ 0.50$).

The addition of a remapping step to the nuclear genome, which was not available to Pozzi et al., confirmed what was hypothesized herein, i.e., that the source of the vast majority of mitochondrion-remapping reads is the mitochondrial genome and not the nuclear genome. In fact, only 9.37% of the mitochondrial reads mapping to the male mitochondrial genome, and 6.15% of reads mapping to the female genome, map to the nuclear genome as well. Despite suggestive evidence (Pozzi and Dowling 2019), these reads are cautiously considered to be of uncertain origin in the SmithHunter pipeline, possibly originating from nuclear mitochondrial pseudogenes (NUMTS, Bensasson et al. 2001), and are discarded.

For each of the six sequencing libraries and for the two sexes, the number of raw reads, the number/percentage of reads after trimming, the number/percentage of trimmed reads mapping to the mitochondrial genome and the number/percentage of trimmed reads uniquely mapping to the mitochondrial genome (i.e., not mapping to the nuclear genome), is reported in Table S1.

Clustering

Mitochondrial-unique reads were clustered, and each cluster was filtered for size using the procedure implemented in the first module of SmithHunter. In line with Pozzi et al., clustering was performed at 99% identity (option $-I\ 0.99$). The stringency was set as the 50th percentile of unique cluster depth (option $-S\ 0.50$), which corresponds to 125 and 116 reads for males and females, respectively. The minimum number of replicates was not enforced (option $-M\ 1$). A total of 89 and 97 clusters were observed that passed all the thresholds for the male and female genomes, respectively. All the

clusters appear to be on the plus strand (FIGURE 3), the same strand from which all the mitochondrial genes are transcribed in the species (Ghiselli et al. 2013).

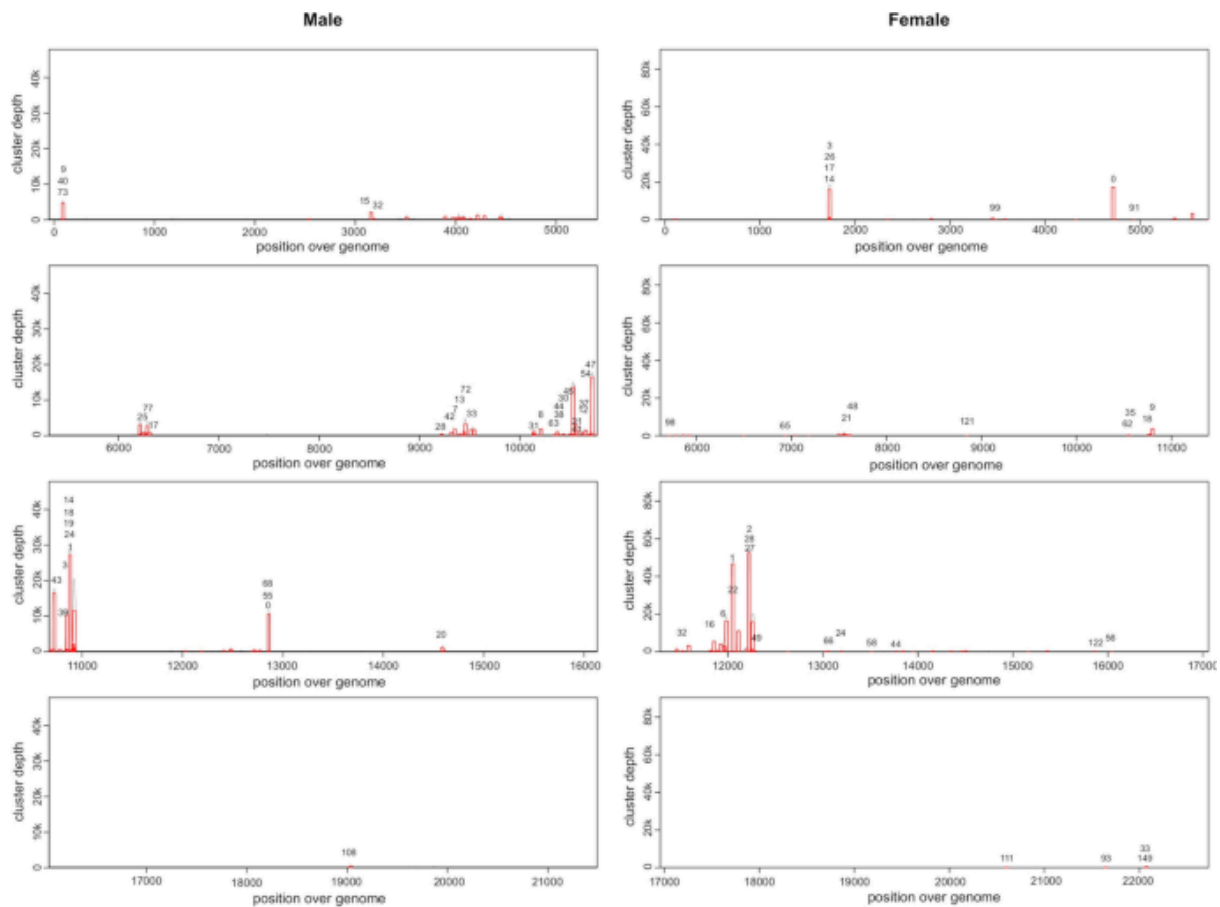


FIGURE 3: Distribution and depth of clusters along the male and female mitochondrial genomes. Grey peaks in the background represent total coverage. The horizontal black line represents the T1 coverage threshold ($-S$ 0.50). All clusters are observed on the plus strand and are shown in red (clusters on the minus strand would be shown in blue). To improve readability, only clusters giving rise to candidate smithRNAs are numbered.

Among these clusters, 42 and 32 exhibited conserved 5' transcription ends in the male and female genomes, respectively, and their centroid sequences were considered presumptive smithRNAs. All 14 smithRNAs selected by Pozzi et al. (hereafter referred to as reference smithRNAs) were found to be presumptive smithRNAs in my analyses, and their coverage at the 5' and 3' transcription ends was almost identical to that previously reported (FIGURE 4).

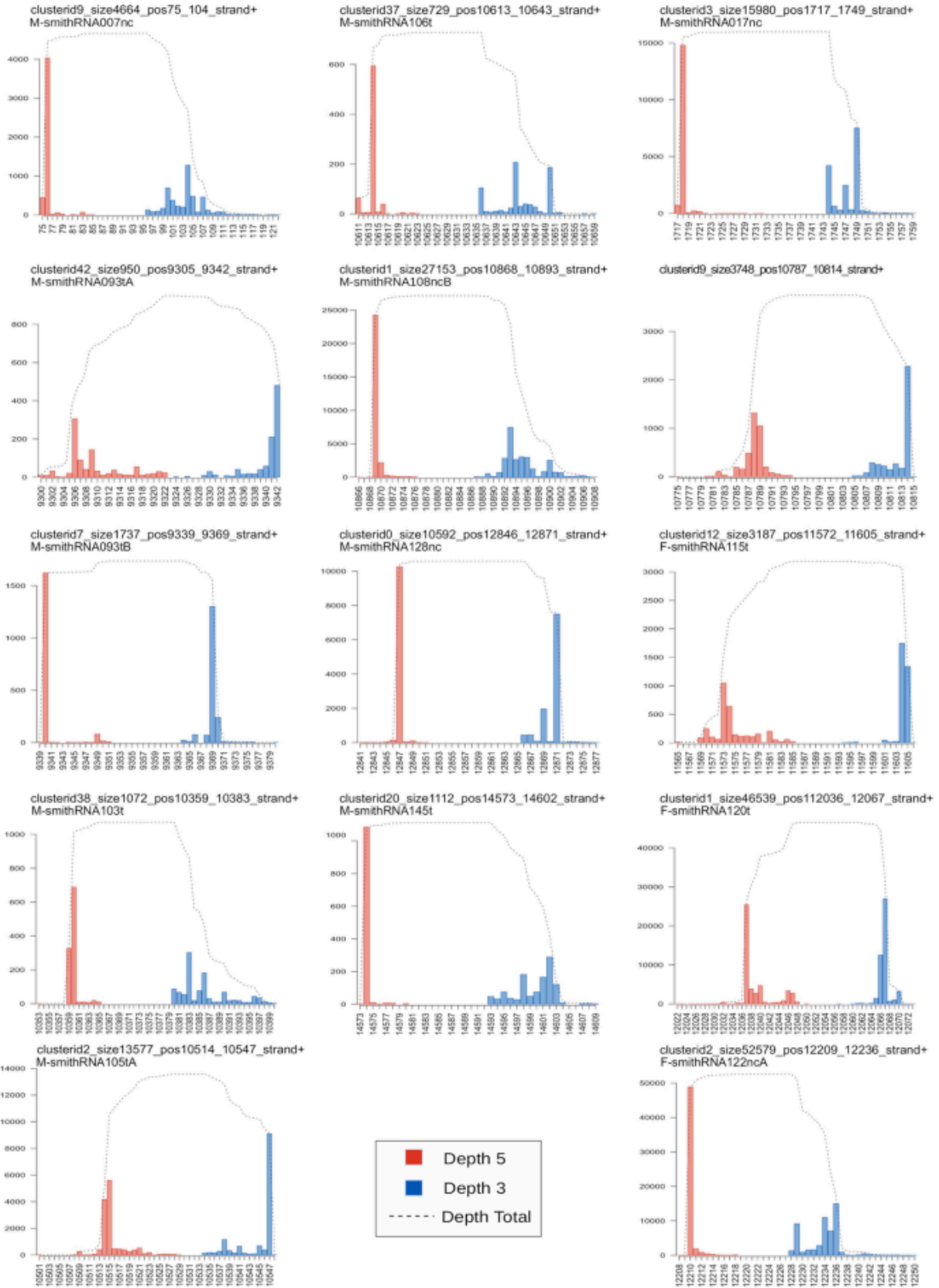


FIGURE 4: Transcription end conservation of the 14 reference smithRNAs. Names of smithRNAs are shown following the nomenclature of SmithHunter]. Red and blue bars represent the distribution of unique start/ends of reads mapping on the mitochondrial genomes, respectively. The black dotted line represents overall, per base, coverage. Genome positions on the horizontal axis refer to sequences AB065375.1 (male mitochondrial genome) and AB065374.1 (female).

Target prediction

Presumptive smithRNAs were searched against UTR regions of the *R. philippinarum* transcriptome using the procedure implemented in the second module of SmithHunter. A total of 39 and 30 presumptive smithRNAs from the male and female genomes, respectively, were hypothesized to target at least one nuclear transcript and are here referred to as candidate smithRNAs. Of the total of 69 candidates, 12 were reference smithRNAs, and 10 of these were associated with the same 13 nuclear targets hypothesized herein (TABLE 1). Most importantly, the two reference smithRNAs that were validated *in vivo* (M_smithRNA106t and M_smithRNA145t, Passamonti et al. 2020) passed through all the filters and were associated with the previously documented targets. In particular, M-smithRNA106t was hypothesized to target the Manila clam homolog of the human Histone-lysine N-methyltransferase and M-smith145t was hypothesized to target the Manila clam homolog of the human polymerase epsilon (TABLE 1). Incidentally, 11 out of 24 smithRNA-target pairs reported by Pozzi et al. were not identified in the current analyses. These sequences were considered individually, and the smithRNA sequence was searched over the UTR regions of the *R. philippinarum* transcriptome using BLAST. A total of 10 out of the 11 smithRNAs actually matched the previously reported target in the plus-plus orientation, suggesting a problem with a directionality filter in the original analysis. These were subsequently excluded from the analysis. In its basic implementation, SmithHunter was therefore shown to largely duplicate previous results. In the following section I aimed at additional testing and the deployment of additional features implemented in SmithHunter.

SmithRNA	Target	PITA	RNAhybrid	Protein name
M-smithRNA108ncB	Locus_4366	- 26.5	- 29.4	Dynein 1 Heavy Chain 1
M-smithRNA145t	Locus_2953	- 22	- 24.5	DNA polymerase epsilon
M-smithRNA106t	Locus_14	- 15.2	- 17	Histone-lysine N-methyltransferase SETD2
M-smithRNA103t	Locus_1096	- 18.4	- 22.7	Microsomal triglyceride transfer protein large subunit
M-smithRNA103t	Locus_6539	- 18.4	- 19.4	Centrosomal protein of 131 kDA
M-smithRNA103t	Locus_62148	- 19.4	- 19.8	Kinein-like protein KIF21A
M-smithRNA093tA	Locus_15177	- 16.4	- 15.8	Elongator complex protein 5

M-smithRNA093tB	Locus_2534	- 15.9	- 17.6	Trifunctional enzyme subunit alpha, mitochondrial-like
M-smithRNA007nc	Locus_5815	- 17.2	- 18.2	U3 small nucleolar RNA-associated protein 6
M-smithRNA007nc	Locus_3650	- 19.2	- 18	eukaryotic translation initiation factor 3 subunit I
F-smithRNA120t	Locus_472	- 20.2	- 19.6	Serine/arginine repetitive matrix protein 2-like
F-smithRNA122ncA	Locus_15925	- 17.6	- 27.9	Nuclear Receptor Subfamily 0
F-smithRNA107t	Locus_31245	- 18.5	- 23.5	Cullin-5-like

TABLE 1: The first column reports smithRNA names as in the original paper. Additional columns indicate the nuclear target, Gibbs Free Energy (dG, in kJ) of RNA-RNA hybrids calculated using PITA and RNAHybrid and the human homolog of the target according to UniProt.

Parameter optimization

The effects of user choice on selectivity related parameters, a new implementation of SmithHunter that was not available to before (Pozzi et al. 2017; Passamonti et al. 2020), have been evaluated by repeating the analysis under 27 different parameter combinations as well as allowing 0 to 2 mismatches in the seed region (see below).

All combinations of cluster identity (option -I; tested: 0.99, 0.95, 0.90), stringency (option -S; tested: 0.2, 0.5, 0.8) and minimum number of replicates (option -M; tested: 1, 2, 3) were studied on both the male and female data. Predictably, the number of presumptive smithRNAs identified by the first module was directly correlated with the identity (I) parameter and inversely correlated with the replicates (M) and stringency (S) parameters. By adopting the less selective parameter combination (-I 0.99 -S 0.2 -M 1) a total of 71 and 65 presumptive smithRNAs were identified for males and females, respectively. Out of this set, 67 candidate smithRNAs, targeting 320 nuclear genes, were identified for males and 59 candidate smithRNAs, targeting 323 nuclear targets for females (TABLE S2, FIGURE 5). In contrast, by using the most selective parameter combination (-I 0.90 -S 0.8 -M 3), eight and six presumptive smithRNAs were found in male and female individuals, respectively. Six and five of these were identified as candidate smithRNAs and were associated with 25 and 16 nuclear targets, respectively, in males and females (TABLE S2, FIGURE 5).

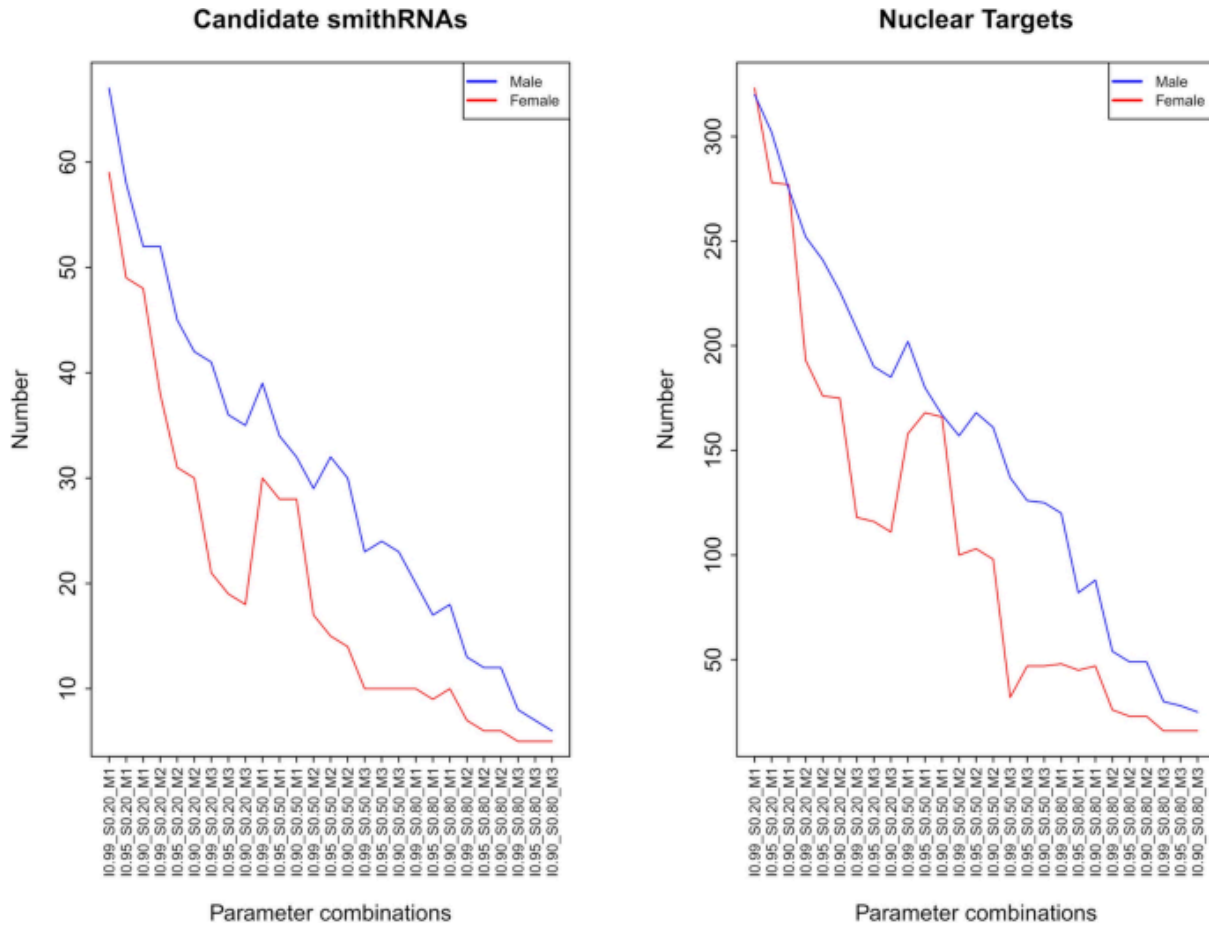


FIGURE 5: Number of candidate smithRNAs and of their associated nuclear targets recovered under different parameter combinations. Parameter combinations are indicated on the horizontal axis as follows: I[identity parameter] _S[stringency parameter] _M[replicates parameter].

Taking the number of presumptive smithRNAs produced at the end of the first module as a proxy for selectivity, it was possible to visualize the effects of different parameters. Stringency (S) had the most substantial effect, leading to the filtering, in the range examined, of 74.5% of clusters. The number of replicates (M) had a more limited effect, with the filtering of 48.9% of clusters. The cluster identity parameter (I), on the other hand, displayed a minimal effect, at least on the Manila clam data, leading to the filtering of 11.5% of clusters (FIGURE S1). Notably, reference smithRNAs were often identified even under the most selective criteria. Eight out of twelve reference smithRNAs were identified under all combinations tested, while the remaining four were identified under a minimum of 21 parameter combinations (FIGURE 6). Moreover, candidate smithRNAs identified under the most selective criteria were also identified under all the tested combinations. This finding suggested that the most selective parameter combinations were optimal starting points and that the resulting candidate smithRNAs, in turn, constitute a reasonably reliable group and starting point for subsequent validation analyses.

Sex	SmithRNA	Target	PITA	RNAhybrid	Protein name
M	clusterid0	Locus_690	- 17.1	- 15.0	Microtubule-associated protein RP/EB family member 1-like isoform X2
M	clusterid0	Locus_1177	- 12.39	- 15.2	Hydrocephalus-inducing protein homolog isoform X19
M	clusterid0	Locus_1711	- 13.7	- 22.2	Uncharacterized protein LOC123531196 isoform X12
M	clusterid0	Locus_4783	- 16.9	- 21.4	Integrator complex subunit 2-like
M	clusterid0	Locus_9761	- 14.3	- 16.2	Matrix-remodeling-associated protein 7-like
M	clusterid0	Locus_3140	- 14.23	- 15.0	Dynein axonemal heavy chain 2-like isoform X3
M	clusterid0	Locus_4916	- 21.7	- 20.7	Dynein heavy chain domain-containing protein 1-like
M	clusterid0	Locus_16697	- 13.29	- 16.2	WD repeat-containing protein 1-like
M	clusterid15	Locus_7476	- 23.4	- 27.2	Uncharacterized protein LOC123561027
M	clusterid15	Locus_7478	- 15	- 16.5	Hypoxia-inducible factor 1-alpha
M	clusterid15	Locus_3651	- 12.9	- 18.7	Protein polybromo-1-like isoform X13
M	clusterid1	Locus_4366	- 26.5	- 29.4	Cytoplasmic dynein 1 heavy chain 1-like isoform X6
M	clusterid3	Locus_1455	- 22.6	- 20.9	Uncharacterized protein LOC123556528
M	clusterid3	Locus_3369	- 15.35	- 17.7	Pericentriolar material 1 protein-like isoform X5
M	clusterid3	Locus_34514	- 22.2	- 23.6	Uncharacterized protein LOC123532869
M	clusterid3	Locus_3229	- 23.7	- 19.4	MAM and LDL-receptor class A domain-containing protein 2-like
M	clusterid7	Locus_2534	- 15.9	- 17.6	Trifunctional enzyme subunit alpha, mitochondrial-like isoform X1
M	clusterid7	Locus_2153	- 12.25	- 16.3	ATP-dependent translocase ABCB1-like isoform X2
M	clusterid7	Locus_17431	- 14	- 16.1	Putative inhibitor of apoptosis
M	clusterid7	Locus_1312	- 11.9	- 15.4	Testis-expressed protein 45-like
M	clusterid9	Locus_5815	- 17.2	- 18.2	U3 small nucleolar RNA-associated protein 6 homolog

M	clusterid9	Locus_3650	- 19.2	- 18.0	Eukaryotic translation initiation factor 3 subunit I-like
M	clusterid9	Locus_9171	- 16.1	- 17.3	Peripheral-type .. receptor-associated protein 1-like isoform X3
M	clusterid9	Locus_5848	- 14.2	- 15.0	E3 ubiquitin-protein ligase rnf213-alpha-like isoform X2
M	clusterid9	Locus_3816	- 17.9	- 17.3	DNA-directed RNA polymerases I, II, and III subunit RPABC3-like
F	clusterid0	Locus_1177	- 15.1	- 15.2	Hydrocephalus-inducing protein homolog isoform X19
F	clusterid0	Locus_7826	- 12.3	- 21.3	Zinc finger and BTB containing protein 17-like isoform X1
F	clusterid0	Locus_1982	- 12.51	- 15.1	Uncharacterized protein LOC123551470
F	clusterid0	Locus_4867	- 15.7	- 19.2	Zinc finger and BTB containing protein 17-like isoform X1
F	clusterid0	Locus_1117	- 9.9	- 18.3	Dynein axonemal heavy chain 3-like
F	clusterid1	Locus_472	- 20.2	- 19.6	msx2-interacting protein-like isoform X2
F	clusterid1	Locus_25748	- 25.91	- 16.9	SH3 and multiple ankyrin repeat domains protein 2-like isoform X2
F	clusterid1	Locus_2909	- 9.5	- 18.3	Syntenin-1-like
F	clusterid1	Locus_1117	- 18.6	- 30.1	Dynein axonemal heavy chain 3-like
F	clusterid2	Locus_15925	- 17.16	- 27.9	Uncharacterized protein LOC123527919
F	clusterid3	Locus_13268	- 15.3	- 16.4	Transcription initiation factor TFIIID subunit 1-like isoform X1
F	clusterid3	Locus_1747	- 14.1	- 17.7	Dynein axonemal heavy chain 6-like isoform X2
F	clusterid3	Locus_6770	- 17.9	- 17.5	Acetyl-CoA acetyltransferase, mitochondrial-like
F	clusterid9	Locus_3552	- 26.8	- 20.2	Cilia- and flagella-associated protein 47-like isoform X8
F	clusterid9	Locus_31245	- 18.5	- 23.5	Cullin-5-like isoform X1
F	clusterid9	Locus_1096	- 25.1	- 16.1	Apolipoporphins-like

TABLE 2: Name of candidate smithRNA (simplified to cluster id) is given in the second column. Other columns report the targeted transcript, Gibbs Free Energy (dG, in kJ) of RNA–RNA hybrids calculated using PITA and RNAHybrid, as well as the targeted gene name.

Their length distribution, compared with those of the total reads mapped to mitochondrial genomes, is shown in Figure S2. The length distribution compares well with the results of Pozzi et al. (compare to

FIGURE 1 therein) over the 22–35 bp range. This finding further suggested that, unlike miRNAs, smithRNAs may exhibit broad variation in length, with substantial peaks in the 20–34 bp range. Noteworthy, this may have a relevance in the context of smithRNA maturation and AGO2 binding, as recently proposed (Pozzi and Dowling 2022).

At least six candidate smithRNAs obtained under the most selective parameter combination were predicted to form bona fide pre-miRNA-like harpins (FIGURE 7). Their free energy is generally lower than that presented in the original study and their shape, in most cases, better conforms to the expectation of a long hairpin. This, in turn, supports the tentative identification of pre-smithRNAs based on position rather than on the span of tRNA and unassigned region annotations in the mitochondrial genome.

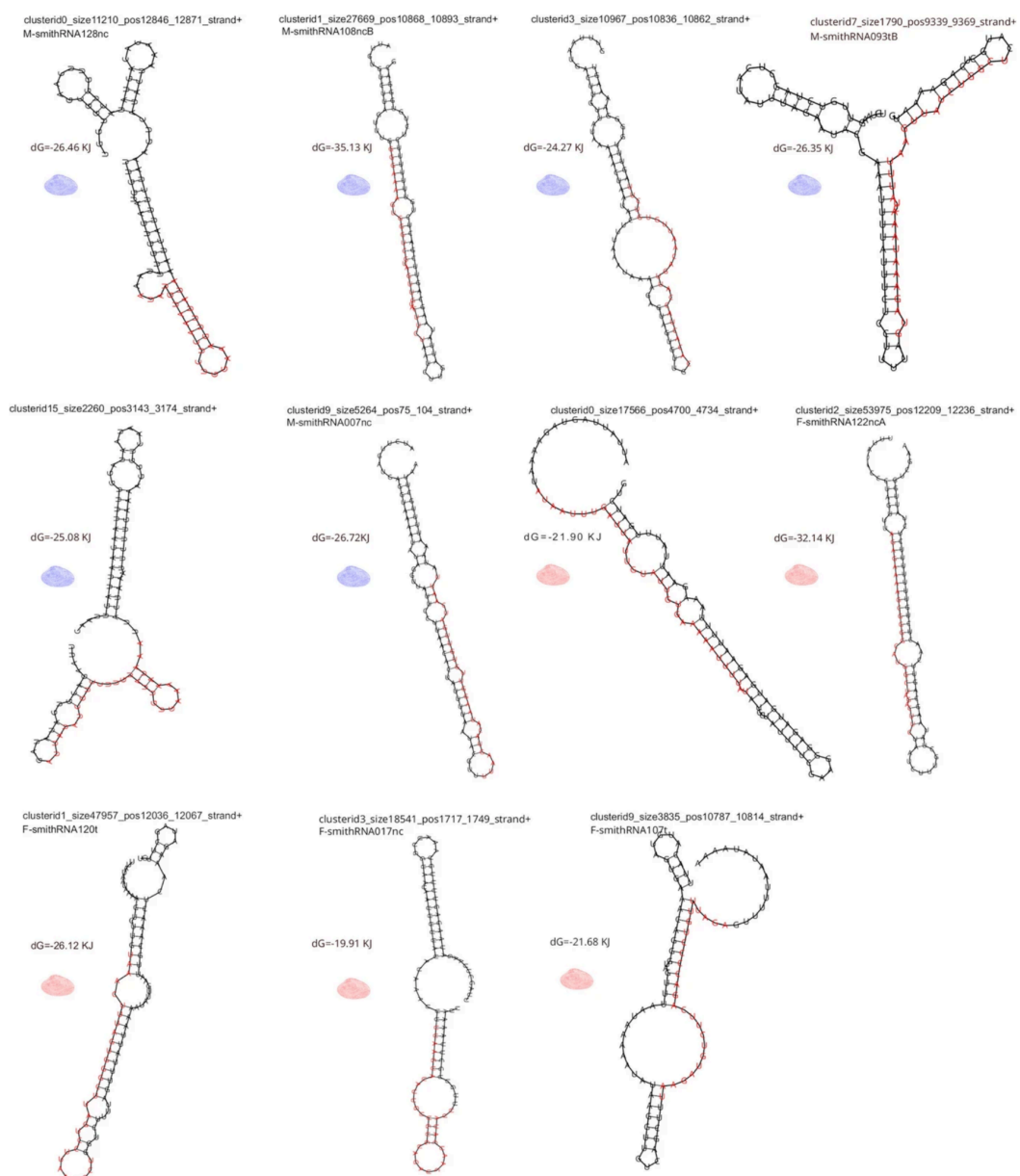


FIGURE 7: Putative secondary structure of pre-smithRNAs for smithRNAs obtained under the most selective parameter combination. Structure labels report the name of the gene, dG of the folded structure, sex and, if available, the corresponding name in the original paper. SmithRNA sequences are highlighted in red. Blue mussel marks indicate male smithRNAs, while pink marks indicate female smithRNAs.

In order to evaluate the effects of allowing non-perfect alignments in the seed region in the target identification step, target identification was repeated under all parameter combinations allowing no mismatch in the seed region (as above) as well as allowing 1 or 2 mismatches (option `-m`).

The number of candidate smithRNAs did not increase significantly when mismatches were allowed in the seed region. With one mismatch allowed, up to two additional candidate smithRNAs were identified under the less selective parameter combinations, and no additional candidate was identified by allowing two mismatches. Conversely, the number of targets identified for each candidate smithRNA increased significantly if nonperfect alignments in the seed region were considered (FIGURE 8). Compared to the case in which no mismatch was allowed, allowing for one mismatch led to the identification of almost twice as many targets (average $1.76 \times$ in male data and $2.03 \times$ in female data across different parameter combinations). Allowing two mismatches, in turn, did not lead to a further increase in the number of identified targets (average $1.01 \times$ and $1.05 \times$, respectively; FIGURE 8). According to this evaluation, allowing mismatches in the seed region appears to have a marginal effect on candidate smithRNA identification. On the other hand, in the context of target identification, allowing mismatches in the seed region leads to a possibly unwarranted increase in targets that are less, or marginally, supported. As such, my advice is not to use this option in standard applications of SmithHunter, where the purpose is to identify a restricted number of high confidence candidate smithRNAs and their targets. This option, in turn, may be used in specific contexts, such as the evaluation of imperfect alignments in identifying one specific target that has previously been validated based on external evidence, or a study set of known targets for parameter optimization. Moreover, given that no difference was observed between one or two mismatches allowed, I suggest using one mismatch, if any, to reduce the run time of the second module.

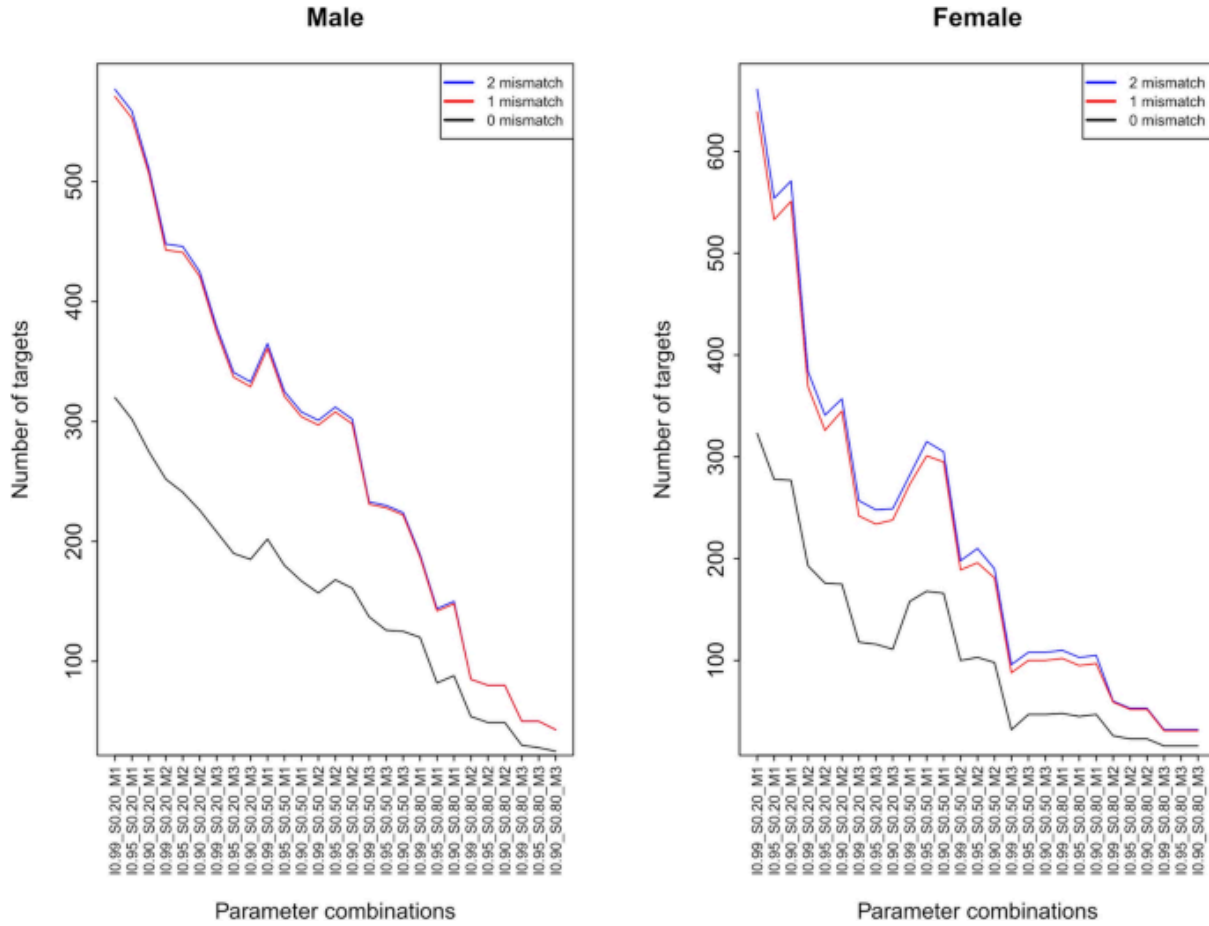


FIGURE 8: Number of targets identified as a function of the number of mismatches allowed in the seed region. Parameter combinations, on the horizontal axis, are labelled as follows: I[identity parameter] _S[stringency parameter] _M[replicates parameter].

Additional testing on *Potamilus streckersoni*

Additional testing in *P. streckersoni* entailed the use of two different sets of analytical parameters and was conducted on both male and female data, focusing on the retrieval of smithRNA M-9, which received substantial support in Smith et al. 2023. The first parameter combination was designed to be grossly similar, in terms of stringency, and acknowledged the differences among the SmithHunter procedure and the procedure applied in the original *P. streckersoni* analysis (Smith et al. 2023). Stringency parameters were applied as: -I 0.99 -S 0.5 -M 1, the nuclear filter was not applied, and end conservation was evaluated following the guidelines of Smith et al, which appear to be more relaxed than in my protocol. Compared to my testing setup, this corresponds to a medium/low stringency. A total of 28 out of the 33 originally identified smithRNAs, namely 8/9 in males and 20/24 in females, were recovered by SmithHunter as candidate smithRNAs.

The second parameter combination applied (-I 0.90 -S 0.8 M 2) corresponds to the most stringent parameter set in my testing regime. Here, the nuclear filter was also applied to exclude small RNAs of

uncertain (nuclear or mitochondrial) origin, and end conservation was evaluated following the guidelines of Pozzi et al., 2017. This time, a total of 10 out of the 33 smithRNAs identified by Smith et al., 5/9 in males and 5/24 in females, were recovered as candidate smithRNAs by SmithHunter. Most importantly, the focal smithRNA M-9, that was singled out as the best candidate based on differential, sex related, expression, as well as its putative role in the sex determining pathway (Smith et al. 2023), was recovered by SmithHunter even under this most stringent parameter combination in association with its proposed target (GCNT1).

Based on a comparison among the two runs, the reduction of the number of candidate smithRNA from the medium/low stringency to high stringency was due to the nuclear filter (5 candidates), coverage thresholds and replicate filter (4) and end conservation (4) or a combination thereof (5).

2.4 Discussion

As outlined above, the starting point of this work was the analysis of Manila clam data (Pozzi et al. 2017) which, in turn, led to the first proposal of the existence of smithRNAs and to the identification of the only two smithRNAs that have been biologically validated (Passamonti et al. 2020). While the analytical procedure is described in sufficient detail in the original paper, the actual bioinformatic implementation is no longer usable because some tools have been updated/surpassed, and more generally the script has been further modified and improved through time (Passamonti et al. 2020). The purpose of this work was therefore to retain the original implementation whenever possible, modify outdated tools with current implementations whenever needed, and increase the level of customizability as appropriate for its intended use across different animal species. Incidentally, I share the view of the original paper that at this early stage of smithRNA research it is preferable to avoid machine learning algorithms specifically trained on different types of small non coding RNAs and rely on more transparent parameters such as binding energy and sequence matching.

Testing, ex post, of the potential of the new analytical pipeline to duplicate the results of Pozzi et al. in *R. philippinarum* gave positive insights. Most smithRNAs and smithRNA/target pairs hypothesized in the original study were recovered. Some differences were nevertheless observed. Among the 14 candidate smithRNAs identified in the original study, 12 were confirmed in the present study. The two unconfirmed reference smithRNAs were found to be associated with nuclear targets previously reported by Pozzi et al., but in plus/plus direction. Additionally, our analyses identified 13 of 24 reference smithRNA-target pairs. Once more, the associations not detected in this study mainly exhibited a plus/plus orientation, suggesting a possible failure in a directionality filter in the original implementation. Incidentally, both smithRNA/target pairs whose biological function has been confirmed experimentally (Passamonti et al. 2020), were recovered in the correct orientation.

Additional testing on *P. streckersoni* confirmed this view. SmithHunter was in fact capable to identify almost all smithRNAs (Smith et al. 2023) at medium/low stringency and, most significantly, to recover the focal M-9 smithRNA, in conjunction with its target, even under the most stringent parameter combination.

Concerning the possibility, in SmithHunter, to customize different analytical steps and thresholds, I proposed some a priori considerations of their possible applicability, and, at the same time, I tested the behavior of the script across different parameter combinations. The minimum identity for

clustering, stringency and minimum number of replicates were under scrutiny, as well as the possibility to allow mismatches in the seed region.

Clustering identity reflects the minimum identity of reads that are combined in a cluster. The purpose of this parameter is to reflect the expected level of genetic variability across replicates. In an ideal situation where the genetic background is actually identical (i.e., all libraries come from one single individual and/or a single mitochondrial background), the parameter may be safely set to 0.99, whereas if libraries come from genetically different individuals some flexibility (0.95 or 0.90, grossly corresponding to 1 or 2 mutations out of approximately 20–25 bases) is to be allowed to avoid excess splitting of clusters across replicates.

The stringency parameter relates to the coverage (i.e., cluster size) threshold. In fact, higher stringency results in a higher threshold for cluster size. This definition of the cluster size threshold was preferred to a hard-coded indication of a minimum ad hoc cluster size (as in the original implementation) to foster comparability across libraries with unequal sequencing output and different species. I envision that, if the user is interested in a limited number of highly expressed clusters, this parameter may be set to 0.8 or higher. If, in turn, the operator is willing to expand its exploration to smithRNAs characterized by medium–low expression, thus maximizing the discovery rate, this parameter may be decreased. A medium level of filtering should nevertheless be retained at this step to avoid the background noise that is generally observed in sequencing libraries, possibly associated with the widespread presence of degradation fragments (Xu et al. 2019), and becomes evident in the coverage plot of mitochondrial reads. The intended use of the replicate parameters, i.e., the minimum number of replicates where a cluster reaches the minimum coverage threshold, in turn relates to the nature of the replicates in the study system and the expectations of the operator concerning the nature of the observed variability. If replicates come from different tissues/conditions/developmental stages and the operator is interested even in smithRNAs that may be expressed in one condition only, thus maximizing the discovery rate, this parameter may be set to 1 (in fact mimicking the original study). If, in turn, the operator is interested only in smithRNAs that are consistently expressed across conditions, thus reducing the number of false positives, the parameter may be set to the number of replicates or to the number of replicates minus one.

The effect of parameter customization was assessed analytically by exploring multiple parameter combinations. The most influential parameter was stringency (S), whereas replicates (M) had a more

limited effect and identity (I) a minimal effect (FIGURE S1). Notably, while I expected that the stringency parameter will be influential regardless of the study system, the more limited effect of the replicate and cluster identity parameters observed here may be related to specific features of the Manila clam data. More specifically, tissue uniformity, as only gonads were used, may have led to a reduction of the relevance of the replicate parameter, and the substantial genetic uniformity, as all the samples came from one and the same location/sampling date, may have led to a reduction of the relevance of the clustering parameter. These two latter parameters may, in turn, become more influential in different study systems characterized by higher genetic and tissue diversity.

Allowing mismatches in the seed region led to minimal differences in the number of candidate smithRNAs identified but, in turn, to a large, possibly unwarranted, expansion of the number of targets identified. Based on this observation, allowing of mismatches is recommended only in specific contexts.

Relevant for target identification within the second module is the availability of a high quality transcriptome sequence, inclusive of reliable 5' and 3' annotations. While this is not always the case, especially in non-model organisms, it can be noted that the most common inaccuracies relate to missing and/or incomplete transcripts, as well as a poor identification of splicing variants, while the occurrence of supernumerary transcripts is more of an unlikely occurrence. As such, using a low-quality transcriptome sequence is liable to obscure the presence of some targets, but may not generally lead to the inclusion of nonexistent targets.

In my view, the proposed analytical pipeline is of substantial interest in two different contexts. On the one hand, it allowed me to produce initial data about the presence and characteristics of smithRNAs in different Metazoan species. While such are necessarily limited to the identification of putative smithRNA/target interactions, this approach is liable to produce data that are comparable across different species. On the other hand, by focusing on one or a few individual species that may be targeted in functional studies, the proposed pipeline will allow the identification of candidate smithRNAs for biological validation. Finally, I currently see the first SmithHunter module as solid and efficient and I do not prefigure modifications in the short term. Instead, I consider the second module to be more experimental. Particularly, the foreseeable availability in the medium term of datasets containing true positives/negatives from *in vivo* experiments will make it easier to better gauge run parameters and to analytically evaluate the performance of different options and thresholds.

3. SmithRNAs as a common feature among Metazoa

This chapter has been published in the journal *Genome Biology and Evolution*. Supplementary materials are available via the link to the original article (<https://doi.org/10.1093/gbe/evaf208>).

3.1 Introduction

While the centrality of smithRNAs in the mito-nuclear communication remains to be explored, many other types of RNAi, more thoroughly understood in the nuclear-nuclear signalling, have been described in animals. Based on maturation, biogenesis, and regulatory mechanisms, most of them are classified in three primary classes: microRNAs (miRNAs), small interfering RNAs (siRNAs), and PIWI-interacting RNAs (piRNAs). Both miRNAs and siRNAs are cleaved from a double-stranded precursor by RNase III-type enzymes, while piRNAs are cleaved from a single-stranded precursor. Moreover, miRNAs and siRNAs depend on proteins of the Argonaute subfamily, while piRNAs count on proteins of the PIWI subfamily (Czech et al. 2018; Chen and Kim 2024). The main role of siRNAs appears to be the suppression of transposons and viruses in different animal species (Chen and Kim 2024). piRNAs similarly suppress transposons in germ cells (Czech et al. 2018; Onishi et al. 2021; Iki et al. 2023), but additional functions have been identified involving mRNAs as targets (Iki et al. 2023; Blumenstiel 2025; Shivam et al. 2025). miRNAs, in turn, are essentially ubiquitous, have a typical length of 22 nt, and play diverse regulatory roles (Ha and Kim 2014; Bartel 2018; Bofill-De Ros and Vang Ørom 2024; Chen and Kim 2024). They originate from RNA hairpin structures and, following processing by the endoribonuclease DICER, are loaded onto an Argonaute protein to produce the silencing complex (Swarts et al. 2014; Iwakawa and Tomari 2022).

Several other pathways, known as ‘non-canonical’ pathways, are liable to generate additional, and currently less understood, types of sncRNAs (Chen and Kim 2024). For instance, mirtrons, present in multiple animal and plant lineages (Westholm and Lai 2011; Chen and Kim 2024), can arise from intronic hairpins (Okamura et al. 2007) and do not require DICER or Microprocessor (Daugaard and Hansen 2017) for their maturation. On the other hand, the so-called tRNA-related fragments are produced from tRNAs (Meseguer 2021; Shaukat et al. 2021). Their maturation pathway is unclear, but similarly may not involve either DICER or Microprocessor (Kumar et al. 2014; Kuscu et al. 2018). Small ribosomal RNA-derived fragments, originating from ribosomal RNA genes, may in turn be recruited as miRNAs or piRNAs (Asha and Soniya 2017; Chen et al. 2017; Lambert et al. 2019; 2021).

From an evolutionary perspective, smallRNAs, including miRNAs, piRNAs and endo-siRNAs, arose early in the history of the eukaryotic cell, before the divergence of extant eukaryotic lineages, perhaps in response to the invasion of parasitic retroviruses and transposons (Calcino et al. 2018; Dexheimer and Cochella 2020). They are present in animals, where they appear to have been lost and acquired independently multiple times (Dexheimer and Cochella 2020; Li and Hui 2023; Formaggioni et al. 2024), as well as plants, algae and fungi (Dexheimer and Cochella 2020; Lax et al. 2020; Wang et al. 2021; Alhassan et al. 2025).

In contrast, our understanding of smithRNAs is still in its infancy. Initial findings on their origin and maturation suggested that smithRNAs can bind to Argonaute 2 (Pozzi and Dowling, 2022) and they might be implicated in sex determination (Pozzi et al. 2017, Smith et al. 2023), but their prevalence within Metazoa is unknown.

In this chapter I assessed whether smithRNAs are to be regarded as a unique feature, restricted to one or a limited number of lineages where they have been initially described, or, else, may represent a previously unexplored mechanism of MRR of more general interest, that may be widespread across animals, or even among all eukaryotes at large. This has been done by analyzing genomic data (small RNAseq, transcriptome, nuclear and mitochondrial genomes) of metazoan species representative of all major lineages (FIGURE 1) using the SmithHunter pipeline (Marturano et al. 2024) to identify putative smithRNA and their nuclear targets. SmithRNA/target pairs identified were further analyzed to identify arising patterns, potentially offering a more detailed understanding of the distribution, origin, and evolutionary pathways of this novel class of molecules.

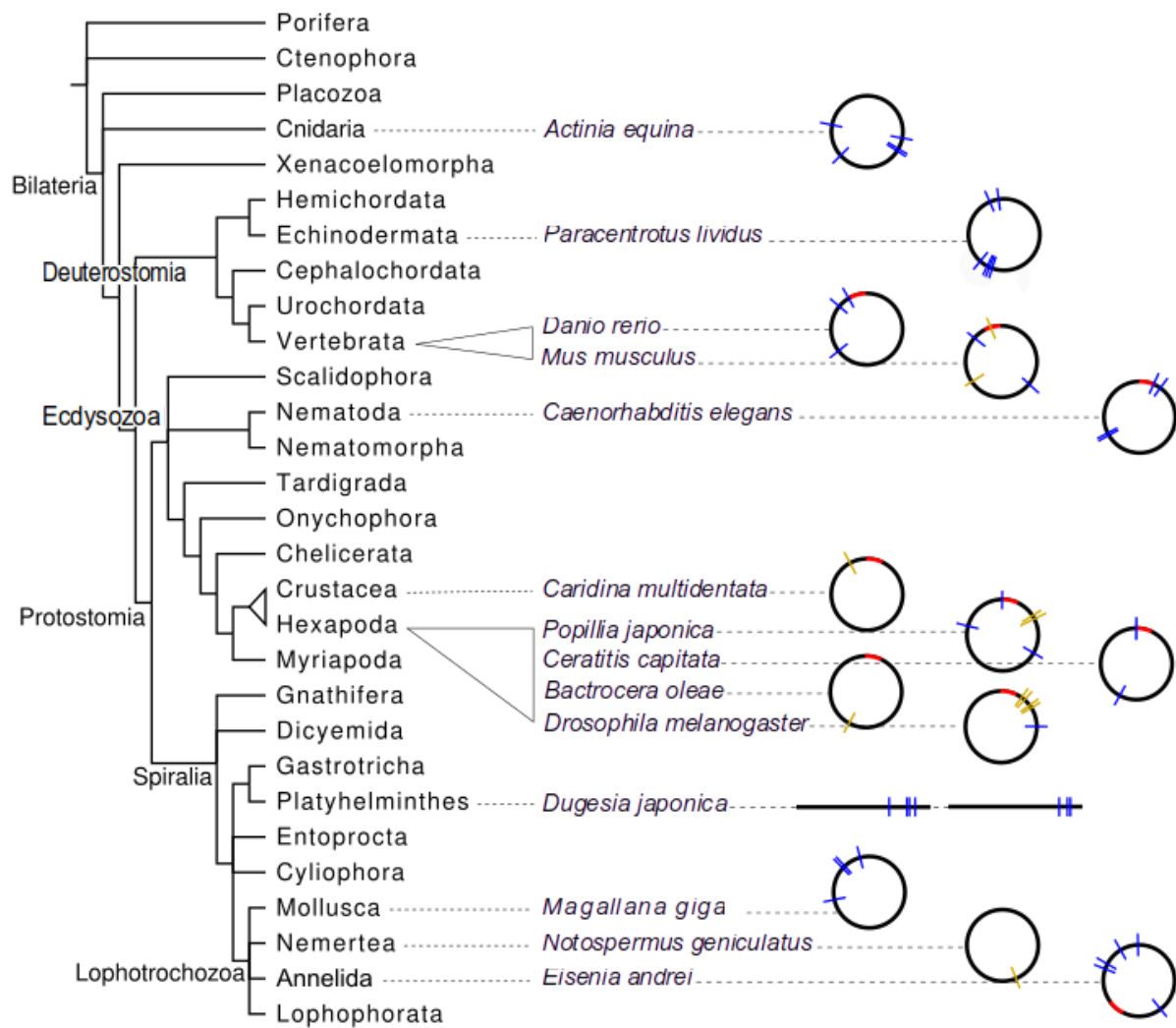


FIGURE 1: List of the species studied here over a consensus phylogeny of Metazoa from Brusca et al. 2023. Circles and lines represent mitochondrial genomes, and coloured segments represent the smRNAs identified (blue: forward strand; yellow: reverse strand; red: major non-coding region, if present).

3.2 Results

Analyses were carried out on a panel of 14 metazoan species (see discussion), including the sea anemone *Actinia equina*, the olive fly *Bactrocera oleae*, the nematode *Caenorhabditis elegans*, the freshwater shrimp *Caridina multidentata*, the Medfly *Ceratitis capitata*, the zebrafish *Danio rerio*, the fruitfly *Drosophila melanogaster*, the planarian *Dugesia japonica*, the earthworm *Eisenia andrei*, the Pacific Oyster *Magallana gigas*, the house mouse *Mus musculus*, the ribbon worm *Notospermus geniculatus*, the sea urchin *Paracentrotus lividus* and the Japanese beetle *Popillia japonica* (TABLE 1). Data for the nuclear and mitochondrial genome, transcriptome, as well as small RNAs, were newly produced or retrieved from previous studies and reanalyzed, as described in TABLE 1.

Species	Abbr. ^a	Classification	Transcriptome	Mitochondrial genome	Small RNAs	Putative smithRNAs	Candidate smithRNAs	n° targets
<i>Actinia equina</i>	AEQ	Cnidaria	<i>de novo</i> ^b	<i>de novo</i>	<i>de novo</i>	5	5	235
<i>Bactrocera oleae</i>	BOL	Hexapoda: Diptera	reassembled ^c	database ^d	<i>de novo</i>	1	1	28
<i>Caenorhabditis elegans</i>	CEL	Nematoda	reassembled	database	data base	4	4	197
<i>Caridina multidentata</i> ^e	CMU	Crustacea	<i>de novo</i>	<i>de novo</i>	<i>de novo</i>	1	1	200
<i>Ceratitis capitata</i>	CCA	Hexapoda: Diptera	reassembled	<i>de novo</i>	<i>de novo</i>	2	2	21
<i>Danio rerio</i>	DRE	Vertebrata: Actinopterygii	reassembled	database	data base	3	3	31
<i>Drosophila melanogaster</i>	DME	Hexapoda: Diptera	reassembled	database	data base	5	5	87
<i>Dugesia japonica</i>	DJA	Platyhelminthes	<i>de novo</i>	<i>de novo</i>	<i>de novo</i>	9	6	126
<i>Eisenia andrei</i>	EAN	Anellida	<i>de novo</i>	<i>de novo</i>	<i>de novo</i>	5	5	60
<i>Magallana gigas</i> ^f	MGI	Mollusca	reassembled	database	data base	4	4	314
<i>Mus musculus</i>	MMU	Vertebrata: Mammalia	reassembled	database	data base	4	4	402
<i>Notospermus geniculatus</i>	NGE	Nemertea	reassembled	database	data base	1	1	61
<i>Paracentrotus lividus</i>	PLI	Echinodermata	<i>de novo</i>	<i>de novo</i>	<i>de novo</i>	6	6	166
<i>Popillia japonica</i>	PJA	Hexapoda: Coleoptera	reassembled	database	<i>de novo</i>	5	5	194
TOTAL						55	52	2122

TABLE 1: Species under study, their classification, abbreviation, overview of dataset construction, number of putative and candidate smithRNAs, as well as the overall number of targets identified by the SmithHunter pipeline.^a three letter code used to designate the species in the following figures; ^b sequences produced and analyzed in this study; ^c raw reads downloaded from database and reanalyzed; ^d downloaded from NCBI, see Materials and Methods section for complete source information; ^e formerly *C. japonica*; ^f formerly *Crassostrea gigas*.

Transcriptome assembly and annotation

Transcriptome sequencing produced 175 to 374 million read pairs per each of the five newly sequenced species, namely *A. equina*, *C. multidentata*, *D. japonica*, *E. andrei* and *P. lividus* (TABLE 1, TABLE S1). Raw data retrieval produced 40 to 366 million read pairs per each of the remaining nine species. After trimming and filtering, 17 to 350 million read pairs were assembled for each species (TABLE S1). The substantially higher number of reads in *P. japonica* and *P. lividus* is justified by the inclusion of stage- or sex-specific libraries to promote transcriptome completeness (see Cucini et al. 2024 and the Materials and Methods section). The assembly process produced between 12,861 and 74,674 genes, resulting in 21,776 to 153,757 isoforms in the 14 species. A global remapping rate of 87% to 98.5%, as well as a proportion of complete BUSCOs ranging from 63.8 to 98.9% and almost invariantly over 95% (TABLE S2), indicate that, even in the presence of primary sequence data of different source/quality, and applying a standardized assembly/annotation procedure to all taxa to foster comparability, transcriptomes of good quality could nevertheless be obtained. Following structural annotation, 28,316 to 226,316 UTRs were identified and used for SmithHunter analysis (TABLE S2).

Mitochondrial and nuclear genomes

Mitochondrial genomes were produced de novo in six species: *A. equina*, *C. multidentata*, *C. capitata*, *D. japonica*, *E. andrei*, *P. lividus* (TABLE 1). Illumina whole-genome sequencing produced 74 to 85 million of read pairs per library. Complete circular mitochondrial genomes could be readily assembled for five out of six species: the expected genes were annotated and they were consistent with genomes of the same species already available in GenBank, both in structure and in primary sequence . Assembly of *D. japonica*, in turn, produced two different contigs encompassing the same area (~85%) of the genome. Based on the known sharing of mutations (Bessho et al. 1992; 1997), I considered the two sequences as the outcome of heteroplasmy, as previously described for this species and associated with asexuality. Both sequences (named A and B variants) were included in the analysis. The mitochondrial genome sequences of the remaining eight species were retrieved from NCBI as complete genomes.

Due to the availability of high-quality nuclear genome assemblies in public databases, all the nuclear genomes of the species analyzed were downloaded from NCBI and CNCB and not sequenced de novo.

SmithRNAs sequencing and identification

Small RNA sequencing produced 137 to 424 million read pairs per each of the eight species for which small RNA libraries were generated de novo (TABLE 1). Raw data retrieval produced 70 to 179 million read pairs per each of the remaining six species. After trimming, 31 to 411 million reads were processed per species. The number of replicates for the newly sequenced species was fixed to 5. The number of replicates for species retrieved from the database varied between 3 and 6. Between 0.03% and 1.58% of reads, corresponding almost invariably to well over 100 thousand reads, mapped to the mitochondrial genome, suggesting the availability of a sizable amount of data as a basis for the analysis. The fraction of reads uniquely mapping to the mitochondrial genome (i.e. mapping to the mitochondrial genome and not to the nuclear genome), provides a gross estimate of the amount of smithRNA reads over the global smallRNA transcriptome. This ranged between 0.01% and 0.88%, with the lowest figures observed in the two dipterans *C. capitata* and *B. oleae* and the highest in the deuterostomes *M. musculus* and *P. lividus*. The ratio between the percentage of uniquely mitochondrial reads and the total percentage of reads mapping to the mitochondrial genome, accounting for the possibility that small RNAs characterized by a mitochondrial sequence could be transcribed from nuclear mitochondrial pseudogenes (NUMTs) ([Bensasson et al. 2001](#)) in the nuclear genome, ranged from 16.05% to 94.11%. The lowest figures came from the two dipterans *C. capitata* and *B. oleae*, suggesting the existence of extensive nuclear mitochondrial pseudogenes, whereas for other species the percentages are almost invariably over 50%. The distribution of read length after trimming, representing the global population of small RNAs in the species analyzed, as well as of mito-unique reads, of mitochondrial origin, and of reads that following analysis were hypothesized to come from bona fide smithRNAs (see below) is shown in FIGURE 2. Trimmed reads display a bimodal distribution, with two peaks at ~21-22 bp and ~30 bp. Mito-unique reads display a broader distribution, with visible peaks at ~20-22 bp, ~27-28 bp and over 40 bp in different species. Reads associated with bona fide smithRNAs are mostly in the 20-32bp range, but with evidence of substantial peaks corresponding to longer molecules in *M. gigas*, *M. musculus* and *A. equina* (see below).

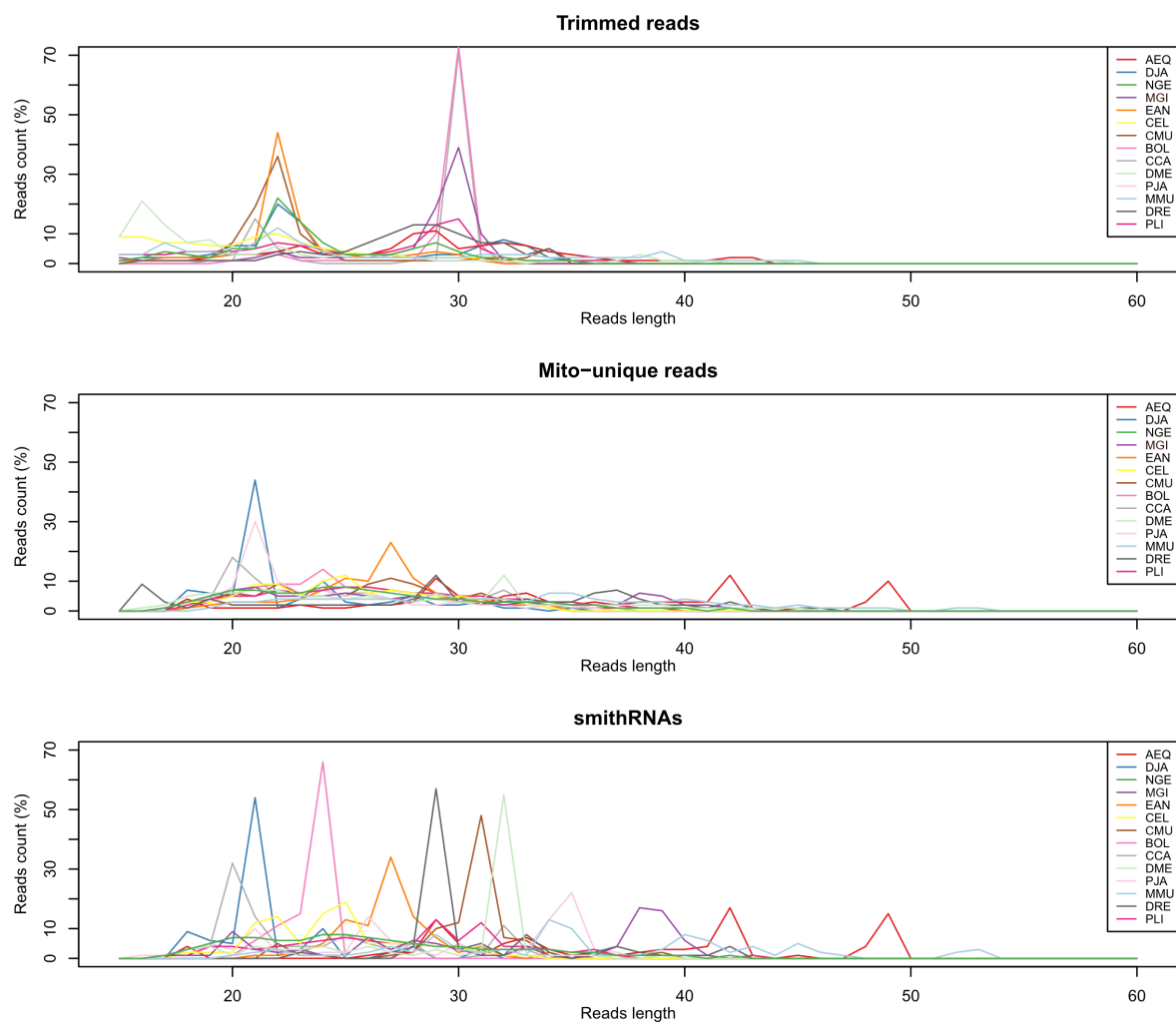


FIGURE 2: Length distribution of all reads (trimmed), reads uniquely mapping to the mitochondrial genome and reads from bona fide smithRNAs. See TABLE S6 for details and TABLE 1 for species abbreviation.

Initial read clustering and filtering of mito-unique reads in the first module of SmithHunter module produced a total of 179 surviving clusters in the 14 species, ranging from one in *B. oleae* to 34 in *P. japonica*. Additional filtering for cluster 5'-end conservation led to the identification of 55 putative smithRNAs globally (1 to 9 in different species; TABLE 1). Final filtering in the second module of SmithHunter, based on target matching, slightly decreased this figure to 52 candidate smithRNAs (1 to 6 in different species, henceforth referred to as candidate smithRNAs) overall (TABLE 1).

Therefore, smithRNAs were identified in all the 14 species tested, including both heteroplasmic variants in *D. japonica*. The number of smithRNAs identified per genome ranged from one in *B. oleae*, *C. multidentata* and *N. geniculatus* to six in *D. japonica* and *P. lividus* (TABLE 1). This suggested that the number of smithRNAs, while generally low, is liable to vary across species, and the variation in number does not have an evident phylogenetic trend. Given the distribution of read length (FIGURE 2), the length of smithRNAs ranged between 18 and 34 bases in most cases. Nevertheless, rare

examples of longer smithRNAs were observed in *M. gigas* (38bp), *M. musculus* (41 bp) and, most noticeably, *A. equina* (42 and 49bp; FIGURE 2, TABLE S3). SmithRNAs were observed to be transcribed from both the forward (n=41) and the reverse (n=11) strand (FIGURE 1, TABLE S3). Most interestingly, in all (n=47) unambiguous cases, i.e. where the smithRNA overlaps with a gene whose strandedness can be ascertained unequivocally, smithRNAs are transcribed from the same strand as the encompassing gene. In the remaining cases (n=5), i.e. where the smithRNAs is located in a non-coding region, the same correlation could be confirmed as follows: a) three smithRNAs encoded in intergenic regions are transcribed from the same strand as the preceding as well as the following gene; b) among the two smithRNAs encoded in the control region, the one (CEL_smith4) encoded at the leftmost end of the control region is transcribed from the forward strand, whereas the one (MMU_smith2) encoded at the rightmost of the control region is transcribed from the reverse strand, in agreement with the general model of mitochondrial transcription (D'Souza and Minczuk 2018; TABLE S3).

Concerning the genomic location from which smithRNAs are transcribed, it appears that they can originate from all locations: protein coding genes (PCGs), rRNAs, tRNAs, as well as intergenic regions and the control region (FIGURE 1, TABLE S3). In line with the general model of metazoan mitochondrial genomes, where intergenic spacers are extremely reduced (Boore 1999), most (90.4%) smithRNAs are encoded within other genes or largely overlap with other genes. Specifically, 20 smithRNAs appear to be encoded within rRNAs, 17 within tRNAs, 10 within PCGs and 5 from noncoding regions (3 from intergenic regions, 2 from control regions; FIGURE 1, TABLE S3, FIGURE S1). The distribution observed across regions (19.2% in PCGs, 32.7% in tRNAs, 38.5% in rRNAs and 9.6% in non-coding regions; FIGURE S1), compared with the sheer extension of the said regions in mitochondrial genomes (69.4%, 8.3%, 13.8%, 8.5% on average, respectively), results in a nonuniform distribution (Chi-square = 2336.8, df = 3, p-value < 2.2e-16) and a marked preference for tRNAs and rRNAs over PCGs. Overall, the diverse mitochondrial origin of smithRNAs does not seem to correlate with taxonomy, as smithRNAs originating from the same mitochondrial component were frequently identified in unrelated species. At the same time, related species generally have smithRNAs originating from different mitochondrial components (TABLE S3).

The twenty smithRNAs derived from rRNAs were identified in 8 species including Deuterostomia (2 species), Ecdysozoa (2), Spiralia (2) and Cnidaria (1). Among these, the large ribosomal subunit

accounts for 12 smithRNAs, while on the small subunits 8 smithRNAs were detected. Similarly, the 17 smithRNAs derived from tRNAs were found in 11 species, including Deuterostomia (2), Ecdysozoa (5), Spiralia (3) and Cnidaria (1). Noteworthy, the two tRNAs for serine accounted for 5 over 17. SmithRNAs encoded within PCGs were found in three species, all protostomes: *P. japonica*, *E. andrei*, and *C. elegans*. Specifically, 5 smithRNAs were observed within genes of the NAD dehydrogenase complex, four within COX genes and one within the ATP6 gene. Overall, no pattern seems to emerge concerning a possible differential distribution of smithRNAs originating from different genomic compartments across taxa.

SmithRNAs do not appear to be conserved at the sequence level (TABLE S3), not even in the rare cases where they are observed within the same gene in related species, whereas a clear pattern emerges concerning their position with respect to the gene from which they originated. In fact, a marked propensity is observed for smithRNAs to be encoded at the extremes, generally at the 5'-end, of the encompassing gene, especially in the case of rRNAs and tRNAs (FIGURE S1). This is especially evident for tRNA-derived smithRNAs, which always originate from one end of the encompassing tRNA, and originate from the 5'-end in as many as 81% cases (FIGURE 3).

The two heteroplasmic mitochondrial genomes observed in *D. japonica* produced a similar remapping profile and clusters, in line with their marked similarity (91% identity in primary sequence), with two shared smithRNAs, two smithRNAs exclusive to the A variant and two exclusive to the B variant, the latter characterized by minimal expression. These differences were considered of limited interest and are not discussed further.

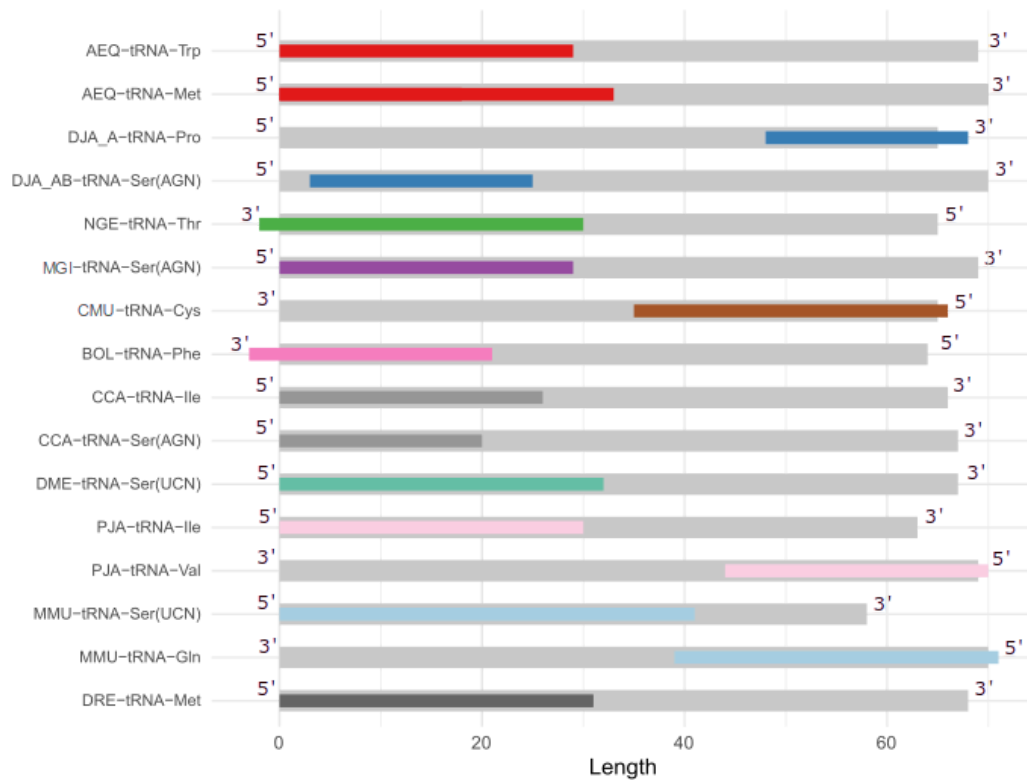


FIGURE 3: Position of smithRNAs identified within tRNA genes. SmithRNAs are represented in colour, tRNAs are represented in grey and their strandedness is indicated. See TABLE 1 for species abbreviations.

smithRNAs nuclear targets

SmithRNA/target matching in the second module of SmithHunter against the assembled transcriptomes led to the identification of possible targets for the smithRNAs described above. Overall, 2,122 nuclear targets were identified for the 52 smithRNAs, with 21 to 402 identified in different species (TABLE 1). The number of targets identified per smithRNA is generally in the 1-25 range, with notable exceptions where one single smithRNA was associated with as many as 200 (CMU_smith2) or 214 (MGI_smith0) nuclear targets (FIGURE 4). As expected, the number of nuclear targets identified is inversely correlated with stringency (dG threshold) applied. Nevertheless, even filtering for RNAHybrid dG as low as -20, all smithRNAs identified retain one or more targets (FIGURE S2).

The possibility that smithRNAs could be preferentially implicated in the regulation of specific biological functions, at the level of individual species or globally, was investigated in a GO enrichment analysis. In most instances, although smithRNAs from different species do not generally target the exact same function (enriched term), most targets were variously associated with mechanisms of differential expression, providing a first potential pattern for smithRNAs function.

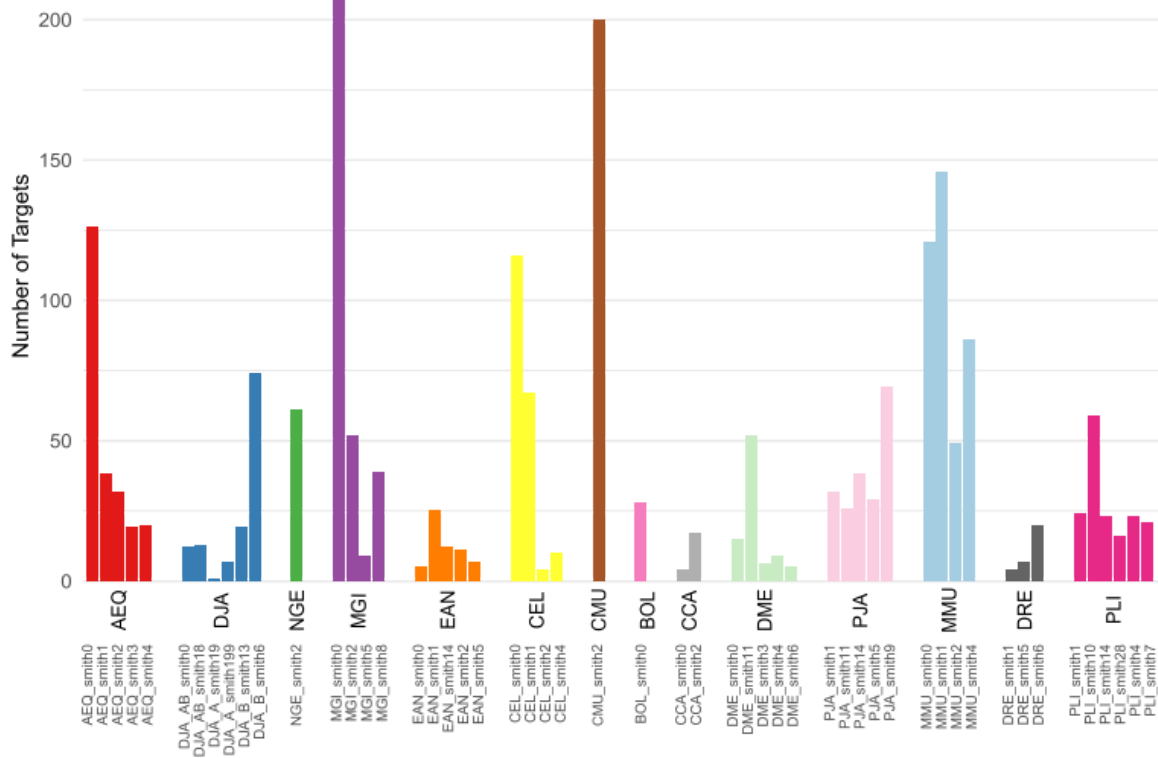


FIGURE 4: Number of nuclear targets identified per smithRNA. See TABLE 1 for species abbreviations.

Within the category Biological Processes (FIGURE 5 and S3), terms related to regulatory processes were found to be enriched in most species, accounting for 42 out of 121 enriched GO terms with P-value < 0.005. Similarly, terms related to the biological response to different stresses and stimuli were enriched in the sea anemone *A. equina*, the ribbon worm *N. geniculatus*, the zebrafish *D. rerio* and the earthworm *E. andrei* (19/121), and terms related to demethylation processes were enriched in *E. andrei* (2/121). Functions specifically associated with mitochondrial functions and/or biology were observed to be enriched in *A. equina* (regulation of protein targeting mitochondrion; negative regulation of release of cytochrome c from mitochondria) and in the nematode *C. elegans* (mitochondrial respiratory chain complex III assembly). Sex-related functions were further identified in *E. andrei* (reproductive senescence, male germ-line stem cell population maintenance and binding of sperm to zona pellucida), in *A. equina* (germ cell attraction) and in the house mouse *M. musculus* (androgen receptor signaling pathway; FIGURE 5 and S3).

Limited evidence was, in turn, obtained for conservation among the targeted genes, with seemingly homologous targets in only two pairs of species: *ncoa7*, involved in nuclear receptors binding, in *C.*

multidentata and the sea urchin *P. Lividus*, and actn1 (actin) in *C. multidentata* and the Japanese beetle *P. japonica*. At variance, smithRNAs encoded from the same gene in different species were sometimes observed to target functionally related genes even in the absence of an evident sequence conservation. Examples are smithRNAs in tRNA-Ile that target serine proteases, smithRNAs in tRNA-Ser(AGN) that target guanine nucleotide-releasing/exchanging factors and the mediator of RNA polymerase II transcription subunit, and smithRNAs in tRNA-Ser(UCN) that target the protein disulfide isomerase. Evidence along this line is nevertheless fairly sparse at present and was not investigated further.



FIGURE 5: GO enrichment analysis of biological processes performed on the supposed nuclear targets of smithRNAs, by species. Only GO supported by more than one gene are shown. See FIGURE S3 for complete information. The circle area represents the number of genes, while the p-value is colour coded. Go terms associated with mitochondrial and sex related functions are highlighted in green and red, respectively. See TABLE 1 for species abbreviations.

3.3 Discussion

Taxonomic distribution

The primary aim of this study was to assess, using a reasonably large and taxonomically representative sample of metazoan species, whether smithRNAs represent a unique feature of the Manila clam and related species, or, else, constitute a more broadly conserved mechanism, potentially widespread across animals or even among eukaryotes.

The data and analyses presented here led to the identification of a total of 52 smithRNAs in the 14 species studied, with each species harboring between one and six different smithRNAs. Additionally, 2,122 supposed nuclear targets were detected, ranging from 21 to 402 in different species, and from 1 to 214 for different smithRNAs (FIGURES 1, 4 and TABLES 1, S3). The sheer dimension of these numbers supports the hypothesis that smithRNAs, though originally described in the Manila clam (Pozzi et al. 2017), are not a peculiar feature of this species, nor of bivalves at large, but are much more widespread across Metazoa. The large taxonomic sampling presented here (FIGURE 1) spans over Deuterostomia (3 species), Ecdysozoa (6 species), Spiralia (4 species), and Cnidaria (1 species). Present results can be combined with previous fragmentary reports of miRNA-like sequences across various taxa, including fruitfly (Passamonti et al. 2020), mouse (Blumental-Perry et al. 2020), humans (Meseguer et al. 2019; Blumental-Perry et al. 2020), fish (Riggs et al. 2019; Passamonti et al. 2020), and bivalves (Pozzi et al. 2017; Smith et al. 2023). Together, they suggest that smithRNAs may be a shared feature that was present in the mitochondrial genome since the outset of the diversification of Metazoa.

Some differences were observed among taxonomic groups: a higher propensity for smithRNAs generation in the reverse strand in Ecdysozoa, a substantial reduction of smithRNA reads in Diptera and their preference to target different pathways, according to the GO term analysis. This may be the secondary outcome of specificities associated to a given group more than real differences in the biology of smithRNAs across taxa. For instance, once it is acknowledged that the strandedness of the smithRNA corresponds almost always to the strandedness of the encompassing gene, the distribution of smithRNA strandedness depends on the gene order of the species, a feature that is obviously related to its taxonomic position (Boore 1999). Similarly, the reduction in smithRNA reads in the two dipterans *B. oleae* and *C. capitata* may be the outcome of the presence of NUMTs, that similarly have

a taxonomic trend, or be one of the many additional specificities of Diptera concerning epigenetic regulation at large (Bewick et al. 2017).

Functions

An additional goal of the study was to identify arising patterns, liable to offer a more detailed understanding of the distribution, origin, and evolutionary pathways of this still relatively unknown class of molecules.

In line with what observed in the bivalves *R. philippinarum* (Pozzi et al. 2017) and *Potamilus streckersoni* (Smith et al. 2023), some smithRNAs were found to be associated with sex-related functions in the earthworm *E. andrei*, the sea anemone *A. equina* and the house mouse and *M. musculus* (FIGURES 5, S3). Furthermore, smithRNAs involved in the regulation of spermatogenesis and sperm motility were identified in most species, leading credibility to the idea that this phenomenon may be broadly relevant across the groups.

However, the regulatory role of smithRNAs does not seem to be confined to sex-specific functions, as they appear to be implicated in a broader spectrum of biological functions. Many different regulatory processes, response to stress, as well as binding activity, were indeed observed to be enriched in the GO analysis across the panel of species analyzed, including some compelling functions such as histone deacetylase in the ribbon worm *N. geniculatus* (FIGURES 5, S3). Notably, these functions find some parallel in the proposed role of response to stress in the annual killifish *A. limnaeus* (Riggs et al. 2019) and mice (Blumental-Perry et al. 2020), as well as histone acetylation in the Manila clam (Passamonti et al. 2020).

Overall, smithRNAs seem to be involved in regulatory networks which differ across species. This suggests smithRNAs as a changing regulatory mechanism that can be recruited to control different pathways in different taxa, as needed. This dynamics is reflected in both the number of nuclear targets identified and the range of different processes and functions in which smithRNAs are involved. Furthermore, the limited sequence conservation of seemingly related smithRNAs across different species, at least at the taxonomic level studied here, suggests that they may be subject to a rapid mode of evolution. This is in agreement with a previous study that hypothesized that smithRNAs can commonly arise de novo during evolution through exaptation from pre-existing mitochondrial RNAs and easily recruit nuclear targets (Plazzi et al. 2024). In fact, a simulation study in bivalves suggested

that newly arisen smithRNAs can efficiently find suitable targets in the nuclear transcriptome with a non negligible probability (Plazzi et al. 2024).

From a methodological standpoint, I acknowledge that the bioinformatic analysis performed here cannot replace expression profiling via Real-Time PCR and/or functional tests *in vivo*. As such, evidence for each one smithRNA/target pair is treated as provisional, whereas I think that evidence for the modulation of a given pathway or function at large (e.g. sex determination, regulation, stress), integrating over multiple smithRNA/target pairs, is substantially more reliable.

Genomic localization

Concerning the genomic localization of smithRNAs, some patterns similarly emerge across the species analyzed here. At variance with *R. philippinarum*, where smithRNAs are mostly encoded within the long intergenic spacers that characterize the bivalve mitochondrial genome (Pozzi et al. 2017), in most other species smithRNAs do not appear to be associated to noncoding regions. As a matter of fact, smithRNAs were identified from all components of the mitochondrial genome (PCGs, rRNAs, tRNAs), with only 9.9% coming from non-coding regions and as little as 5.8% from intergenic spacers (TABLE S3). In general terms, this may be a consequence of the peculiar structure of the mitochondrial genome of bivalves, where long unassigned regions (i.e., intergenic spacers) are observed, at variance with canonical metazoan mitochondrial genomes, that are very compact in structure with virtually no intergenic spacer of substantial size apart from one, variously named as Control Region, A+T-rich region or D-loop in different taxa (Boore 1999). While smithRNAs were observed to originate from all genetic components of the mitochondrial genome, the majority was identified within rRNA and tRNA genes, with the latter, considering their comparatively smaller size, being the mitochondrial component which is more prone to generate smithRNAs. This observation is not unprecedented (Villegas et al. 2007; Burzio et al. 2009; Mercer et al. 2011), and has been frequently associated to the presence of sequences capable to give rise to extensive secondary structures within rRNA and tRNA genes, that can serve as a basis for hairpin formation in smallRNAs. Non-canonical mechanisms of miRNAs formation involving tRNAs hairpins have been repeatedly observed (Kawaji et al. 2008; Thompson et al. 2008; Zhang et al. 2009). Moreover, mitochondrial tRNA-derived sncRNAs have been identified in the human 143B cell line, where they were found to be the most abundant compared to sncRNAs derived from other mitochondrial components (Mercer et al. 2011). Similarly, rRNA-derived small RNAs, originating through a process of fragmentation, were

identified in several organisms (Asha and Soniya 2017; Chen et al. 2017; Cherlin et al. 2020; Kusch et al. 2023). In addition, both the 12S and 16S mitochondrial rRNAs, as well as nuclear rRNAs, have been identified as a source of smallRNAs in humans from the 1000 Genomes Project (Cherlin et al. 2020). Additionally, putative smithRNAs located within tRNA and rRNA genes were previously identified in *D. melanogaster*, *D. rerio*, and mice (Passamonti et al. 2020). Nevertheless, the pre-existence of sequences capable of forming secondary structures, facilitating as it can be, does not appear to be a prerequisite, as a non-negligible fraction of smithRNAs (~20%) are actually encoded within PCGs.

A possible mechanism of transcription

The orientation of candidate smithRNAs consistently matches the orientation of the tRNAs, rRNAs, or PCGs from which they originate (TABLE S3), and, with some caveats, this correspondence appears to extend also to smithRNAs that are encoded within intergenic spacers and the control region. While not a rule, a strong tendency was also observed for smithRNAs encoded within tRNAs and rRNAs, but not PCGs, to be encoded at one of the extremes of the gene, generally the 5'-end in tRNA-derived smithRNAs (FIGURE 3 and S1).

In turn, this is strongly suggestive of a possible transcription mechanism for smithRNAs. On one hand, the correspondence in strandedness between the smithRNA and the enclosing gene suggests that smithRNA transcription is concomitant to the transcription of the enclosing gene, and therefore it has to be evaluated in the context of the general transcription model of the mitochondrial genome (D'Souza and Minczuk 2018). It can be envisioned that, after the formation of the two long multi-gene primary transcripts, and during transcript maturation, pre-smithRNA sequences are cleaved off to be released and then follow their own maturation pathway. On the other hand, the non random distribution of smithRNAs with respect to gene boundaries, and the fact that they are almost invariantly located at the 5'-end of tRNAs, suggests that first the genes, and then smithRNAs are cleaved off (generally from the 5'-end) from secondary, gene-specific, transcripts. Noteworthy, the proposed mechanism is not overly different from the one proposed in Pozzi et al. 2017, concerning the possible mechanism of transcription of smithRNAs in the Manila clam, but rather a variation on the same theme. There, smithRNAs, encoded not within genes but within the large intergenic spacers that characterize the species' mitochondrial genome, may be similarly transcribed as part of the two primary multi-gene transcripts and then cleaved during the process of transcript maturation and

released concomitantly to other RNA molecules. Incidentally, this is different from the typical transcription process of nuclear miRNAs, that generally derive from specific transcripts, with few overlapping other genes (Shang et al. 2023).

The variability observed in the length of smithRNAs (FIGURE 2) currently precludes a definitive interpretation. Although the length distribution of trimmed reads, characterized by two peaks at approximately 21–22 nt and ~30 nt, aligns with miRNA-like length expectations (Ha and Kim 2014), the detection of smithRNAs spanning the entire 20–32 nt range, including some longer variants, remains unexplained. It is important to note that the expectation of a narrow length distribution centered around 22 nt derives from the well-characterized biogenesis and maturation pathways of miRNAs (Ha and Kim 2014). These pathways, however, may not apply to smithRNAs, for which distinct transcriptional or maturation mechanisms have been proposed. Consequently, smithRNAs appear to constitute a subset of the broader family of small regulatory non-coding RNAs in mitochondria (Khatri and Blumental-Perry 2025) but do not correspond directly to the currently better-characterized miRNA, piRNA, or siRNA classes. This observation suggests that the biogenesis and functional role of smithRNAs are still poorly understood and calls for a more comprehensive investigation. Moreover, this study demonstrates that smithRNAs share overlapping properties with mitochondrial-derived tRNA fragments, which are known to localize to the cytoplasm (Meseguer et al. 2019). Critically, for both classes, the mode of mitochondrial export remains elusive. While several translocation systems have been proposed (Meseguer 2021; Muneretto et al. 2024), definitive experimental evidence of their operation is currently lacking, and future experimental efforts must elucidate these routes to validate how these molecules reach their functional targets.

3.4 Materials and Methods

Overview

The data herein analyzed includes the nuclear genome, transcriptome, mitochondrial genome and small RNAseq of 14 species representing all major metazoan lineages (TABLE 1). Primary criteria in assembling the dataset were the availability of all four components for a given species, necessary for a full SmithHunter analysis, and taxonomic representativeness across Metazoa. Actions taken to further foster comparability where: (a) transcriptomes were assembled/annotated de novo for all species applying the same methodology; (b) libraries from whole body samples were preferred; (c) mitochondrial genomes were sequenced, whenever possible, from a genetically related individual to the one used for small RNA sequencing; (d) small RNAs were sequenced ex novo in most cases, applying the same methodology.

Transcriptome sequencing and annotation

The transcriptome sequence was obtained de novo in 5 species and reassembled based on raw reads from NCBI in the remaining 9 species (TABLE 1). Total RNA was extracted from single specimens, or a pool for *D. japonica*, using the RNeasy Mini Kit (Qiagen) and assessed for quality using a NanoDrop ONE and Qubit 5.0 (Thermo Fisher Scientific). The whole body was used in all species with the exception of *P. lividus*, for which gonads were used. This finds a justification in the body size and internal texture that made processing of the whole animal for RNA extraction problematic. mRNA was sequenced at Macrogen (Amsterdam, The Netherlands) using TruSeq stranded mRNA libraries and a 2×150bp strategy in a NovaSeq6000 machine, targeting 9 Gigabases per sample. Five to ten replicates were sequenced for each species, and raw sequence collections from NCBI were selected to have between 3 and 6 replicates per species. Whenever possible, different stages and sexes were included to foster completeness and representativeness of the transcriptome.

Raw sequences were trimmed in fastp v. 0.23.2 (Chen et al. 2018). Contaminant reads of non-metazoan origin were filtered using kraken2 v. 2.1.3 (Wood et al. 2019) based on a custom database including RefSeq complete bacterial, archaeal, plant, fungal and human genomes, as well as a collection of known vectors from the UniVec_Core database. Reads of ribosomal origin were filtered using SortMeRNA v. 4.3.6 (Kopylova et al. 2012) based on the smr_v4.3_default database. All 14 transcriptomes were assembled using Trinity v. 2.8.5 (Grabherr et al. 2011) with default parameters. The quality of individual assemblies was evaluated based on: (a) fragmentation, using

Assembly-stats (<https://github.com/sanger-pathogens/assembly-stats>); (b) completeness, using BUSCO v. 4.1.4 (Manni et al. 2021) against the metazoan_odb10 database; (c) the fraction of properly remapping reads, using Bowtie2 v. 2.3.5.1 (Langmead and Salzberg 2012) and SAMtools v. 1.6 (Danecek et al. 2021). Any remaining mitochondrial transcript, identified in BLAST v. 2.12.0 (S. F. Altschul et al. 1990) as contigs sharing over 90% similarity across over 80% of their length with the corresponding mitochondrial genome, was discarded. Transcripts were structurally and functionally annotated in TransDecoder v. 5.5.0 and Trinotate v. 3.11. In summary, open reading frames longer than 100 amino acids were searched against the SwissProt, PFAM, and eggNOG databases, and transmembrane helices and signal peptides were predicted using tmHMM and signalP. Coding regions were then predicted incorporating results from homology searches. In rare cases where more than one coding region was predicted from the same isoform, the longest protein (i.e. the one with the highest score) was retained. The transcriptome in FASTA format, as well as a BED file with positional information on 5'- and 3'-UTRs, was used as input for SmithHunter.

Mitochondrial genome sequencing and annotation

The mitochondrial genome sequence was obtained de novo in 6 species (TABLE 1). Total DNA was extracted from single specimens, or a pool for *C. capitata* and *D. japonica*, using the Wizard Genomic DNA Purification kit (Promega) and assessed for quality using a NanoDrop ONE and Qubit (Thermo Fisher Scientific). The DNA was sequenced at MacroGen (Amsterdam, The Netherlands) using TruSeq PCR free libraries and a 2×150bp strategy in a NovaSeq6000 machine targeting 20 Gigabases per sample. The library TruSeq Nano (Illumina) was used for *D. japonica* due to the limited size of the sample. Raw sequences were trimmed in fastp v. 0.23.2 and assembled in NOVOplasty v. 4.3.5 (Dierckxsens et al. 2017) using known COX1 sequences from BOLD (Ratnasingham and Hebert 2007) as seed. The assembly of *D. japonica* was further verified using MEGAHIT (Li et al. 2015). All the mitochondrial assemblies were annotated using MitoS2 (Bernt et al. 2013). The mitochondrial genome sequence of the remaining 8 species was retrieved from NCBI.

Nuclear genome retrieval

Nuclear genome assemblies were downloaded from NCBI or CNCB for all species (TABLE 1). Following a BLAST search, nuclear contigs sharing over 90% similarity across over 50% of their length with the corresponding mitochondrial genome were considered as of likely mitochondrial origin

and removed. This was necessary given the filtering procedure (see: mito-unique reads) implemented in SmithHunter.

Small RNA sequencing

The sequence of small RNAs was obtained de novo in 8 species (TABLE 1). Total RNA was extracted using the miRNeasy Micro kit (Qiagen) and assessed for quality using a NanoDrop ONE and Qubit 5.0 (Thermo Fisher Scientific). Small RNAs were sequenced at Macrogen (Amsterdam, The Netherlands) using NEBNext Small RNA libraries (New England Biolabs) and a 2×150bp strategy in a NovaSeq6000 machine targeting 7.5 Gigabases per sample. TruSeq Small RNA libraries (Illumina) were used for *C. capitata* in an initial phase of the study. Five replicates were sequenced for each species to foster completeness and representativeness in the small RNAseq. Due to size and texture of the sample, gonads (male) only were used for sequencing of *P. lividus*. While I acknowledge a limitation in terms of comparability, I consider its impact on the analyses to be negligible as long as the discussion is based on positive observations only and the number/features of smithRNAs are not compared numerically across taxa. Raw sequences of small RNAs for the remaining 6 species were retrieved from NCBI, selecting library collections that had between 3 and 6 replicates per species.

Candidate smithRNAs identification

smithRNAs were identified independently in each of the 14 species using SmithHunter (Marturano et al. 2024). RNA sequences of *D. japonica* were analyzed against both variants, supposedly heteroplasmic, of the mitochondrial genome of the species. In summary, presumptive (sensu Marturano et al. 2024) smithRNAs were initially identified by running the SmithHunter A module as follows. After filtering against sequences of possible nuclear origin, mitochondrial-unique reads were clustered at 90% of identity (-I parameter), clusters were filtered to be above the 80th percentile of unique cluster depth (-S parameter) and a minimum of N-1 replicates (-M parameter; N=number of available replicates) passing filters was required for a cluster to be retained. The remaining clusters were further evaluated for end-conservation using the sharp_smith.R script adopting a T-five threshold of 0.5. Presumptive smithRNAs were subsequently evaluated using SmithHunter B module (default settings) against 5'- and 3'-UTR regions of transcripts to identify possible nuclear targets. Plots were produced using SmithHunter. While, given the nature of the starting information, it cannot be excluded that some of the identified clusters may represent degradational fragments with no associated

function, these settings correspond to the most stringent among the parameter combinations tested in Marturano et al. (2024) and the one that was identified as most appropriate to identify a reliable set of candidate smithRNAs for subsequent analyses, contributing to the overall credibility to the analysis.

Inferred nuclear targets were further scrutinized in a Gene Ontology enrichment analysis to evaluate the possibility that specific pathways/functions may be preferentially targeted by smithRNAs. GO enrichment was carried out using TopGO at P-value < 0.005. Plots were produced using ggplot in R (R Core Team 2025). Nuclear transcripts targeted by candidate smithRNAs identified in all the species analyzed were further evaluated for orthology relationships with OrthoFinder under default parameters (Emms and Kelly 2019).

3.5 Supplementary Materials

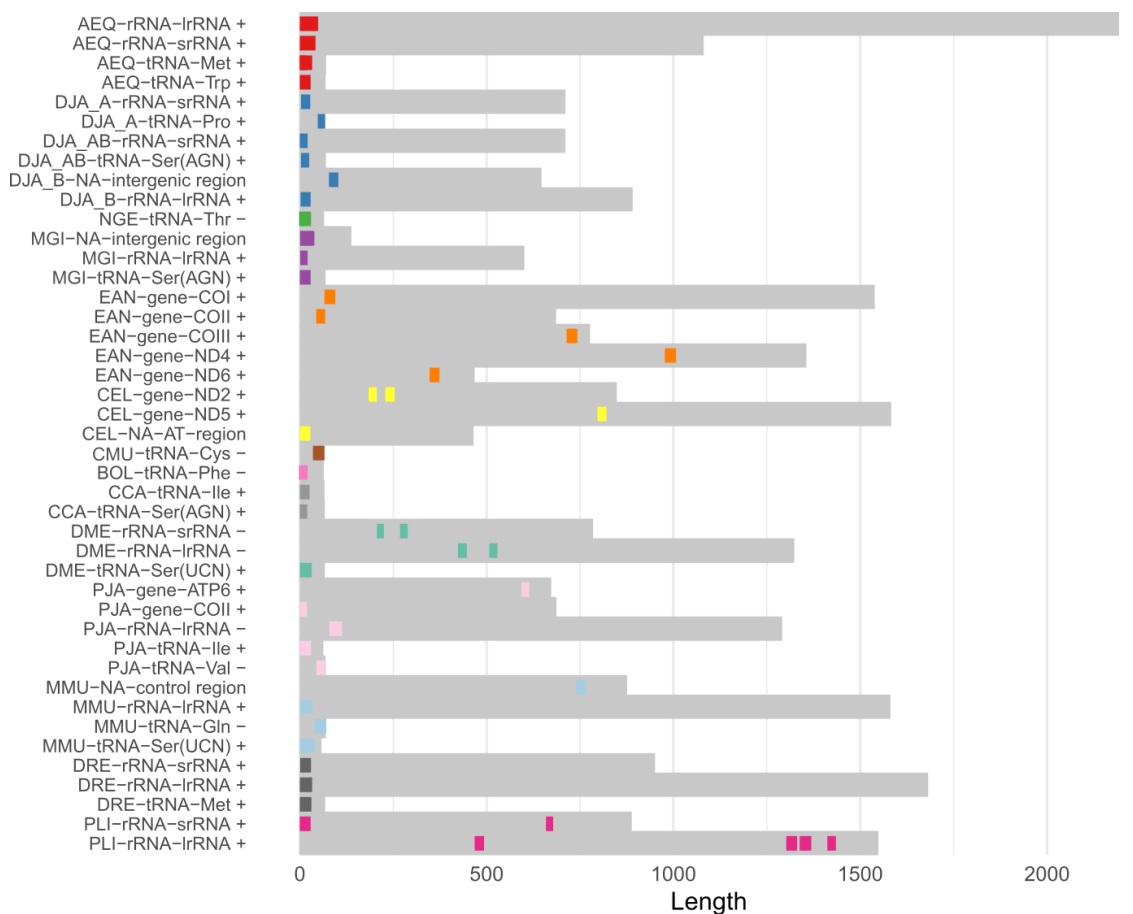


FIGURE S1: Position of smithRNAs relative to the genomic region where they are encoded. Grey and coloured boxes represent the genomic region and smithRNAs, respectively. Species are colour coded. Gene strandedness is indicated as + (sequenced strand) or – (complement strand).

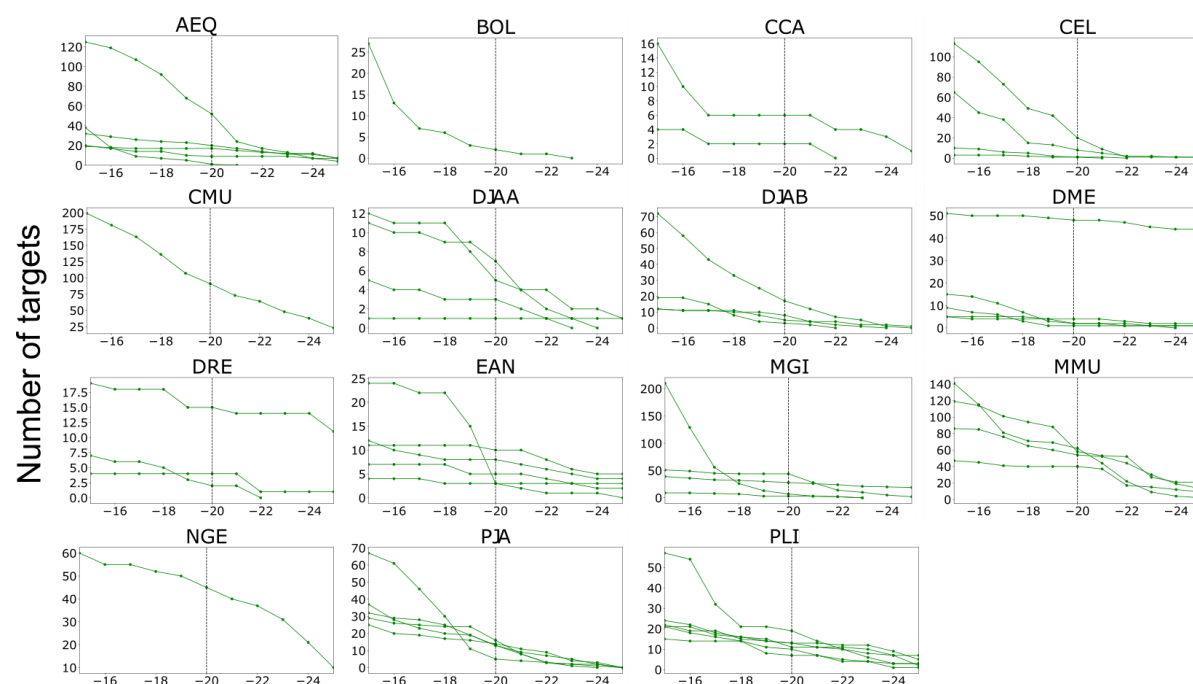


FIGURE S2: Number of identified nuclear targets as a function of smithRNA-target binding energy (dG). Each green line identifies a different smithRNA.

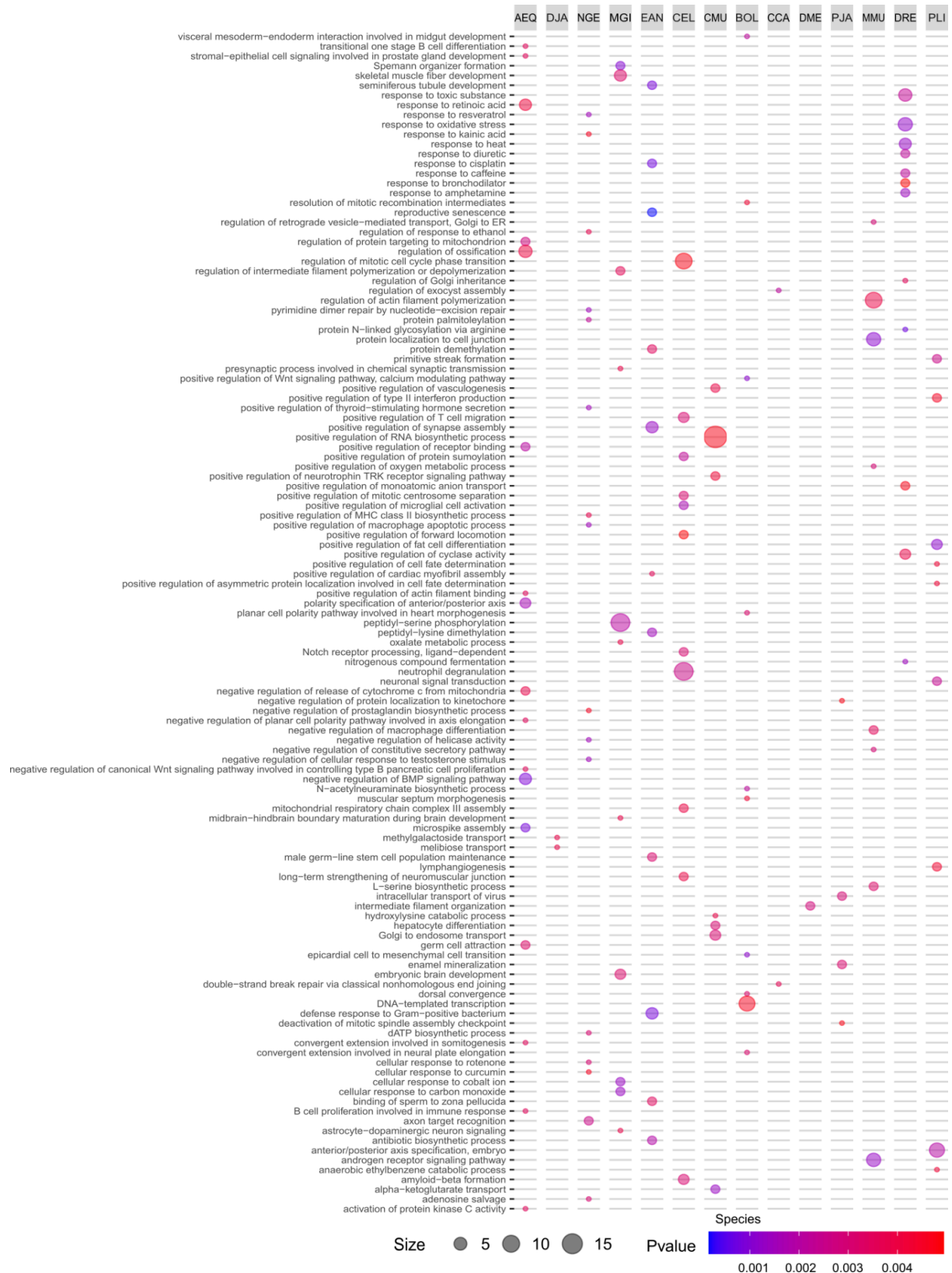


FIGURE S3: Gene Ontology enrichment analysis of biological processes performed on the supposed nuclear targets of smithRNAs, grouped by species. Circle size represents the number of genes belonging to the specified GO class. The p-value is colour coded.

Species	SRA accession	Tissue	Raw (read pairs)	Filtered (reads pairs)	Total filtered (read pairs)
<i>Actinia equina</i>	SRR28330942	whole body	35.295.370	5.501.922	136.339.866
	SRR28330941	whole body	35.453.947	33.134.786	
	SRR28330940	whole body	33.906.265	31.493.426	
	SRR28330939	whole body	35.343.874	33.240.197	
	SRR28330938	whole body	35.064.132	32.969.535	
<i>Bactrocera oleae</i>	SRR8800824	whole body	34.849.995	24.583.040	112.607.041
	SRR8800826	whole body	40.630.107	31.532.742	
	SRR8800835	whole body	38.678.071	27.468.455	
	SRR8800833	whole body	37.651.087	29.022.804	
<i>Caenorhabditis elegans</i>	SRR7443965	hypodermis	37.717.253	2.117.657	19.769.408
	SRR7443967	neurons	31.348.071	7.252.421	
	SRR7443968	neurons	33.110.795	7.435.022	
	SRR7443975	neurons	32.346.851	2.964.308	
<i>Caridina multidentata</i>	SRR28330958	whole body	39.609.960	35.007.681	164.491.374
	SRR28330957	whole body	39.034.511	35.216.020	
	SRR28330956	whole body	36.448.034	31.006.198	
	SRR28330955	whole body	35.244.991	31.488.607	
	SRR28330954	whole body	35.930.691	31.772.868	
<i>Ceratitis capitata</i>	SRR836188	whole body	27.777.737	5.219.824	16.858.275
	SRR836189	whole body	25.891.361	5.421.987	
	SRR836190	whole body	31.016.416	6.216.464	
<i>Crassostrea gigas</i>	SRR13528977	gonads	55.680.106	47.427.817	252.137.516
	SRR13528978	gonads	47.608.066	39.912.255	
	SRR13528979	gonads	47.772.090	41.250.563	
	SRR13528980	gonads	63.158.655	43.569.069	
	SRR13528983	gonads	43.471.193	37.079.882	
	SRR13528984	gonads	50.260.430	42.897.930	
<i>Danio rerio</i>	SRR5181405	gonads	40.090.189	29.551.196	104.938.438
	SRR5181406	gonads	37.897.175	27.536.049	
	SRR5181411	gonads	38.346.079	25.503.458	
	SRR5181412	gonads	33.550.183	22.347.735	
<i>Drosophila melanogaster</i>	SRR5181405	gonads	12.157.192	7.575.689	24.777.739
	SRR5181406	gonads	10.144.040	6.312.856	
	SRR5181411	gonads	6.868.154	3.973.422	
	SRR5181412	gonads	10.341.764	6.915.772	
<i>Dugesia japonica</i>	SRR28330935	whole body	37.439.968	34.657.203	175.739.234
	SRR28330934	whole body	35.985.210	33.170.338	
	SRR28330933	whole body	38.664.038	35.703.103	
	SRR28330932	whole body	35.703.235	33.097.739	
	SRR28330931	whole body	42.317.149	39.110.851	
<i>Eisenia andrei</i>	SRR28330960	whole body	42.035.147	39.028.360	
	SRR28330959	whole body	40.333.571	37.990.828	

	SRR28330948	whole body	42.274.042	39.949.408	
	SRR28330937	whole body	35.142.458	33.078.754	
	SRR28330936	whole body	32.989.091	31.169.563	181.216.913
<i>Mus musculus</i>	SRR16832607	gonads	37.861.037	30.449.676	
	SRR16832608	gonads	39.905.457	32.216.464	
	SRR16832609	gonads	38.364.179	31.394.377	
	SRR16832610	gonads	41.127.811	30.266.048	
	SRR16832611	gonads	40.789.383	33.227.390	
	SRR16832612	gonads	36.632.747	27.423.844	184.977.799
<i>Notospermus geniculatus</i>	SRR5811964	gonads	17.233.380	14.564.966	
	SRR5811966	gonads	17.250.875	16.154.404	
	SRR5811967	gonads	16.130.732	15.187.895	
	SRR5811970	gonads	16.473.717	15.330.559	61.237.824
<i>Paracentrotus lividus</i>	SRR28330943	gonads	38.270.799	36.128.014	
	SRR28330947	gonads	33.187.852	31.379.661	
	SRR28330946	gonads	34.967.849	31.958.121	
	SRR28330945	gonads	40.025.085	37.765.963	
	SRR28330944	gonads	33.289.808	31.151.480	
	SRR28330943	gonads	39.115.050	36.719.822	
	SRR28330952	gonads	39.022.595	36.782.161	
	SRR28330951	gonads	39.366.661	36.720.744	
	SRR28330950	gonads	37.032.699	33.640.372	
	SRR28330949	gonads	39.787.721	37.373.548	349.619.886
<i>Popilia japonica</i>	SRR20647943	whole body	46.697.613	34.010.274	
	SRR20647941	whole body	51.997.688	44.228.788	
	SRR20647945	whole body	43.120.800	38.558.938	
	SRR20647948	whole body	43.680.037	38.997.322	
	SRR20647950	whole body	41.568.118	35.096.751	
	SRR20647932	whole body	51.649.684	42.629.648	
	SRR20647936	whole body	41.014.055	34.752.395	
	SRR20647934	whole body	45.901.875	39.950.256	308.224.372

TABLE S1: Information about transcriptome sequencing. Accession number (SRA), biological origin and the number of raw reads and filtered reads are indicated for each library. The total number of filtered reads, used to assemble transcriptomes, is indicated per species.

Species	Transcript size	Number of contigs	Average contig size	N50	BUSCO	% Properly mapping reads	N° genes	N° isoforms	N° UTRs
<i>Actinia equina</i>	251862622	243629	1033	2168	C:96.8%[S:41.5%,D:55.3%], F:1.2%,M:2.0%	91,55	26018	71398	102793
<i>Bactrocera oleae</i>	115706881	90631	1276	2998	C:95.3%[S:61.3%,D:34.0%], F:0.3%,M:4.4%	96,99	12861	28775	48135
<i>Caenorhabditis elegans</i>	47979731	45481	1055	1875	C:63.8%[S:35.6%,D:28.2%], F:9.7%,M:26.5%	86,96	13341	21776	28316
<i>Caridina japonica</i>	347221602	267470	1298	3241	C:98.1%[S:37.5%,D:60.6%], F:0.6%,M:1.3%	95,06	25080	71878	112211
<i>Ceratitis capitata</i>	167719830	203846	822	1538	C:93.8%[S:64.0%,D:29.8%], F:1.2%,M:5.0%	92,27	15405	29486	43700
<i>Crassostrea gigas</i>	478101424	554331	862	1695	C:98.5%[S:37.1%,D:61.4%], F:1.4%,M:0.1%	95,77	38587	104641	145566
<i>Danio rerio</i>	127959264	115414	1108	2253	C:98.9%[S:61.2%,D:37.7%], F:0.6%,M:0.5%	90,57	23123	42883	63774
<i>Drosophila melanogaster</i>	76100984	72514	1049	2059	C:89.2%[S:60.3%,D:28.9%], F:7.9%,M:2.9%	91,57	14166	27296	37278
<i>Dugesia japonica</i>	59356547	64267	923	1627	C:96.8%[S:41.5%,D:55.3%], F:1.2%,M:2.0%	91,85	16939	34510	41507
<i>Eisenia andrei</i>	223980358	275628	812	1478	C:92.0%[S:47.8%,D:44.2%], F:4.7%,M:3.3%	90,38	39508	83710	92503
<i>Mus musculus</i>	1009106044	1568414	643	884	C:98.7%[S:20%,D:78.7%], F:0.8%,M:0.5%	91,8	74674	151227	226316
<i>Notospermus geniculatus</i>	305559287	334915	912	1825	C:98.9%[S:25.4%,D:73.5%], F:0.7%,M:0.4%	88,86	31131	80721	112193
<i>Paracentrotus lividus</i>	724935223	1099158	659	936	C:96.7%[S:12.7%,D:84.0%], F:2.2%,M:1.1%	91,16	46806	153757	164988
<i>Popilia japonica</i>	279438688	278215	1004	2495	C:98.3%[S:30.8%,D:67.5%], F:0.1%,M:1.6%	98,5	34896	79707	112879

TABLE S2: Transcriptome assembly statistics and annotation. BUSCO score shows values for Completeness (C), Single Copy (S), Duplicated (D), Fragmented (F) and Missing (M) genes. Transcript size, average contigs size and N50 refers to base pairs.

ID	Species	Genomic origin	Strand origin	Strand smithRNA
AEQ_smith1	<i>Actinia equina</i>	lrRNA	+	+
AEQ_smith0	<i>Actinia equina</i>	srRNA	+	+
AEQ_smith2	<i>Actinia equina</i>	tRNA-Met	+	+
AEQ_smith3	<i>Actinia equina</i>	tRNA-Met	+	+
AEQ_smith4	<i>Actinia equina</i>	tRNA-Trp	+	+
BOL_smith0	<i>Bactrocera oleae</i>	tRNA-Phe	-	-
CEL_smith0	<i>Caenorhabditis elegans</i>	NAD2	+	+
CEL_smith1	<i>Caenorhabditis elegans</i>	NAD2	+	+
CEL_smith2	<i>Caenorhabditis elegans</i>	NAD5	+	+
CEL_smith4	<i>Caenorhabditis elegans</i>	control region	na	+
CMU_smith2	<i>Caridina multidentata</i>	tRNA-Cys	-	-
CCA_smith2	<i>Ceratitis capitata</i>	tRNA-Ile	+	+
CCA_smith0	<i>Ceratitis capitata</i>	tRNA-Ser(AGN)	+	+
CGI_smith0	<i>Magallana gigas</i>	intergenic region	+	+
CGI_smith5	<i>Magallana gigas</i>	intergenic region	+	+
CGI_smith2	<i>Magallana gigas</i>	lrRNA	+	+
CGI_smith8	<i>Magallana gigas</i>	tRNA-Ser(AGN)	+	+
DRE_smith1	<i>Danio rerio</i>	srRNA	+	+
DRE_smith5	<i>Danio rerio</i>	lrRNA	+	+
DRE_smith6	<i>Danio rerio</i>	tRNA-Met	+	+
DME_smith3	<i>Drosophila melanogaster</i>	srRNA	-	-
DME_smith6	<i>Drosophila melanogaster</i>	srRNA	-	-
DME_smith11	<i>Drosophila melanogaster</i>	lrRNA	-	-
DME_smith4	<i>Drosophila melanogaster</i>	lrRNA	-	-
DME_smith0	<i>Drosophila melanogaster</i>	tRNA-Ser(UCN)	+	+
DJA_A_smith199	<i>Dugesia japonica</i> var. A	srRNA	+	+
DJA_A_smith19	<i>Dugesia japonica</i> var. A	tRNA-Pro	+	+
DJA_AB_smith0	<i>Dugesia japonica</i> var. A and B	srRNA	+	+
DJA_B_smith13	<i>Dugesia japonica</i> var. B	intergenic region	+	+
DJA_B_smith6	<i>Dugesia japonica</i> var. B	lrRNA	+	+
DJA_AB_smith18	<i>Dugesia japonica</i> var. B	tRNA-Ser(AGN)	+	+
EAN_smith1	<i>Eisenia andrei</i>	COX1	+	+
EAN_smith2	<i>Eisenia andrei</i>	COX2	+	+
EAN_smith5	<i>Eisenia andrei</i>	COX3	+	+
EAN_smith14	<i>Eisenia andrei</i>	NAD4	+	+
EAN_smith0	<i>Eisenia andrei</i>	NAD6	+	+
MMU_smith2	<i>Mus musculus</i>	control region	na	-

MMU_smith0	<i>Mus musculus</i>	lrRNA	+	+
MMU_smith4	<i>Mus musculus</i>	tRNA-Gln	-	-
MMU_smith1	<i>Mus musculus</i>	tRNA-Ser(UCN)	+	+
NGE_smith2	<i>Notospermus geniculatus</i>	tRNA-Thr	-	-
PLI_smith4	<i>Paracentrotus lividus</i> M	srRNA	+	+
PLI_smith7	<i>Paracentrotus lividus</i> M and F	srRNA	+	+
PLI_smith10	<i>Paracentrotus lividus</i> M	lrRNA	+	+
PLI_smith14	<i>Paracentrotus lividus</i> M	lrRNA	+	+
PLI_smith1	<i>Paracentrotus lividus</i> M	lrRNA	+	+
PLI_smith28	<i>Paracentrotus lividus</i> M	lrRNA	+	+
PLI_smith0	<i>Paracentrotus lividus</i> F	tRNA-Leu(UUR)	+	+
PLI_smith5	<i>Paracentrotus lividus</i> F	lrRNA	+	+
PLI_smith2	<i>Paracentrotus lividus</i> F	tRNA-Ser(AGN)	+	+
PJA_smith9	<i>Popilia japonica</i>	ATP6	+	+
PJA_smith11	<i>Popilia japonica</i>	COX2	+	+
PJA_smith5	<i>Popilia japonica</i>	lrRNA	-	-
PJA_smith14	<i>Popilia japonica</i>	tRNA-Ile	+	+
PJA_smith1	<i>Popilia japonica</i>	tRNA-Val	-	-

TABLE S3: Candidate smithRNAs identified. Each smithRNA is named based on the species of the corresponding cluster number from the SmithHunter pipeline. Origin from heteroplasmic variant A or B in *D. japonica* is indicated.

4. *Wolbachia pipientis* RNA interference in the leafhopper *Zyginidia pullula*

4.1 Introduction

The profound impact of RNA interference (RNAi) on the evolutionary trajectory of life is becoming increasingly evident. The ubiquity of small RNAi across all the three domains (i.e. Archaea, Bacteria, and Eukarya), strongly suggests its ancient origins and fundamental importance (Koonin 2017). This widespread presence implies that RNAi is not a recent innovation, but rather an ancestral mechanism that has been conserved and adapted over billions of years of evolution. Given its consistent presence across such diverse life forms, it is plausible to hypothesize that RNAi played a critical role in major evolutionary transitions. One such pivotal moment may have been the emergence of eukaryotes. This event, characterized by the fusion and coevolution of Archaeal and Bacterial lineages, likely presented a complex genomic landscape that required sophisticated regulatory mechanisms (Torri et al. 2022). RNAi could have been instrumental in managing this genomic integration, acting initially as a regulator of the genomic regions from which it originated. Over time, however, this function may have expanded or been exapted, repurposed for novel roles, such as mediating intercommunication between distinct genomic entities. Indeed, emerging evidence supports the idea that small RNA (sRNAs) fragments serve as signaling molecules facilitating communication not only within the nucleus but also between the nuclear genome and other organelles, particularly mitochondria and chloroplasts (Cognat et al. 2017). This bidirectional regulation, where sRNAs from organelles influence nuclear gene expression and vice versa, highlights the intricate and dynamic nature of gene regulation in eukaryotic cells. While numerous studies have validated the existence and functionality of sRNAs in mitochondria (Meseguer 2021), and in bacteria too (Lalaouna et al. 2013), the exploration of this mechanism in their closest prokaryotic relatives, the alphaproteobacteria, remains poorly studied. Therefore, investigating whether this sRNA-mediated communication is a conserved feature in alphaproteobacteria could provide crucial insights into the evolutionary origins and conservation of this regulatory system.

Specifically, I focus on *Wolbachia pipientis*, an endosymbiont prokaryote that parasitizes half of known insects (Weinert et al. 2015). It is able to induce several drastic changes in its host, typically driving sex reproduction distortion in order to amplify its spread (Werren et al. 2008). Cytoplasmic Incompatibility (CI), Male-Killing (MK), parthenogenesis and feminization are all *Wolbachia*-induced

effects that can alter the sex ratio of a population, favouring female, pathogen-carrying individuals. While certain *Wolbachia* genes are supposed to induce sex-altering effects (Christensen and Serbus 2015; Ote et al. 2016; Sheehan et al. 2016), our understanding of how these bacteria influence host gene expression and behaviour remains incomplete. Partly, because many proteins are suspected to be species or genus specific (Rice et al. 2017), as with other cellular parasitism (Burstein et al. 2016). Recent findings suggest a connection between RNA interference and *Wolbachia* infection, revealing alterations not only in the host's microRNA expression, but also in sRNAs originating from the bacteria themselves (Mayoral et al. 2014; Woolfit et al. 2015).

The aim of this study is to shed light on how *Wolbachia pipientis* utilizes RNA interference to induce phenotypic and molecular chancing in its host. This would gain some insights on how mitochondria would have exploited an ancient machinery, shared by its free living relatives, to instaure connections and take control in their new organellar life-style.

4.2 Results

Identify Wolbachia strains and host genome

The *Wolbachia pipientis* genome has been assembled separately for the two infected samples, one per sex, resulting in a genome of 1.029 million bases. Both of them are relatively highly fragmented, compared to other bacterial genomes, with a number of 264 and 267 (female and male, respectively) contigs. Such fragmentation was expected, since only short reads were used for this study. Colinearity analysis revealed that the two genomes shared about 99.5% of the genome, with equal coding density (FIGURE S3). This good score of alignment between the two allowed us to take one of the two as a reference when tracking down the sRNAs expression along *Wolbachia*'s genome. The equal number and type of tRNAs (in terms of number, the majority of sRNAs coding regions) detected among the two genomes is another reason why I felt confident in using a unique bacterial genome for the following analyses.

The nuclear genome of *Zyginidia pullula* shows a totally different magnitude of fragmentation. As before, I independently assembled two genomes from two different individuals (8.455M female contigs and 7.936M male contigs), both not infected with *Wolbachia* strains. Total contigs are quite high for a genome with an expected size of 1.3 Gb. Thus, N50 is very low, of 570 and 500 kb, respectively. Completeness scores assessed on eukaryote BUSCO genes highlighted the suboptimal state of the assembly, with only 21.5% of complete genes and similar percentage of missing genes. The vast majority (59%) of genes belong to the fragmented BUSCO genes. The poor quality of the genome is likely resulting from the very high estimated heterozygosity (3.71%) of this population, and the lack of long reads data that can be used as scaffolds.

Wolbachia Infected male individuals resemble female molecular signature

The assembled transcriptome results in a very high number of transcripts. The BUSCO score, assessed against insects' most shared orthologues, shows a high score of completeness (C:98.4%[S:15.1%,D:83.3%],F:1.0%,M:0.6%,n:1367). The 810,430 transcripts have been shrunk down to 411,386 while counting transcripts, grouping together isoforms of the same gene. Then, considering only genes with a count per million (cpm) greater than or equal to 3, I end up with 260,793 genes. As suggested by FIGURE 1A, the coding sequence transcripts were the most suited to explaining the diversity among different conditions.

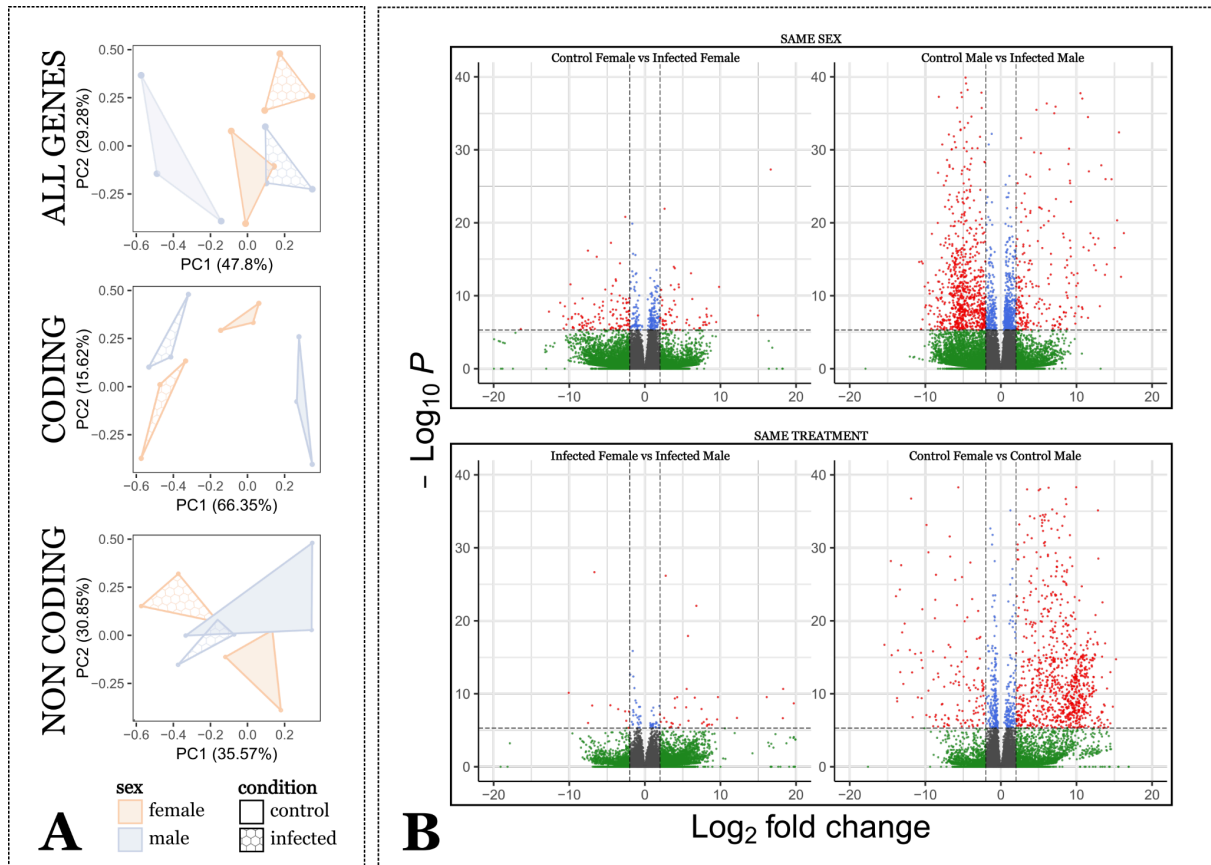


FIGURE 1: A) Distribution on a multidimensional space of the *Zyginidia-Wolbachia* system transcriptome. **B)** Volcano plots of paired comparisons between conditions. Dots are coloured when meeting a threshold on $\log_2$ FC (green when out of ± 2), p-value (blue when $< 5 \times 10^{-6}$), or both (red).

The distribution is not even across different conditions comparisons. Female samples display a relatively low variation between the infected condition and the control (FIGURE 1B). The same could be said for *Wolbachia* infected males and females. This low variation at the molecular level is mirrored in the phenotypic similarity observed at a macroscale. The major differences appear with the biased gene expression of males, where genes are produced at a totally different level in healthy males, compared to infected ones. GO terms associated with these genes show a high enrichment in biological processes leading to the cilium assembly and, generally, to microtubule related activities. Additionally, infected males increase the expression of genes involved in carbohydrate metabolism and lipid deposition. Overall, infection triggers defense response in both sexes (FIGURE S1).

sRNA libraries analyses detect 83 microRNAs, produced by *Z. pullula* genome. Sex seems to be critical in determining the expression of these elements, with 20 microRNAs differentially expressed between controls. But *Wolbachia* infection affects microRNAs levels as well: 16 of them are differentially expressed in infected males, compared to controls. On the other hand, only 2 microRNAs are overexpressed in infected females. Interestingly, one of these two (Bge-Nov-1-P2_pre_11)

displays a curious pattern: while being much more expressed in females, infection promotes its transcription in both sexes (TABLE S1).

Non-coding regions and tRNAs as the main sources of bacterial sRNA

The first module of the SmithHunter pipeline was run on sRNAs libraries, separated by sex. It reported a total of 81 and 72 clusters in females and males, respectively. They originate from all types of genomic features: from coding sequences (44%), tRNAs (36%), non coding regions (16%), and rRNAs (4%). If considering the significance of each cluster, and so its size, bacterial sRNAs arise mainly from tRNAs (47%), then non coding regions (33%), coding sequences (18%) and rRNAs (2%). When considering the distribution of bacterial sRNA production, the total reads arise overwhelmingly from tRNAs (47%) and non coding regions (33%), followed by coding sequences (18%) and rRNAs (2%). To assess whether this production is proportional to the genomic size of the features, a chi-squared test was performed comparing the sRNA read counts to the total base pair length of each genomic element (CDS: 889.9 kb; tRNA: 2.6 kb; NC: 111.6 kb; rRNA: 22.1 kb). The analysis resulted in a highly significant value ($\chi^2 \approx 98.6$ Million, $p\text{-value} = 2.2 \times 10^{-16}$), confirming that sRNA production is not proportional to the genomic length of the features. This difference is predominantly driven by tRNAs, which represent only 0.3% of the total genomic feature length yet contribute nearly half of all sRNA reads, providing clear evidence that minimal genomic regions are highly efficient sources for sRNA biogenesis. Additionally, while 26 out of 34 tRNAs produce a detectable signature of sRNAs, in non coding regions sRNAs originate from only 7 hotspots. Moreover, the clusters are not evenly distributed in structures of the same kind. Almost all RNA fragments in non coding regions derive from a specific area I called ncA. This region is located within *ncrwme102*, a transcribed non-coding region which shows its maximum expression in *D. melanogaster* and *D. simulans* infected male testis (Woolfit et al. 2015). Its expression is stable above the 30k reads in infected female hosts, with one strand (ncA-a) more expressed than the other in all the three samples (FIGURE 2A). In males its expression follows a similar pattern across samples, with one strand preferred, but its expression is erratic. On the other hand, *Wolbachia* sRNAs mapping on tRNAs arise only from one strand of the genome, the coding one (i.e the strand from which the corresponding tRNA is transcribed). The expression is overall not comparable to the ncA region, though in two specific cases it is. tRNA-Glu and tRNA-Ile manifest a male-biased expression, although following the same wobbly scheme

observed for ncA. Additionally, tRNA-Glu seems to be expressed at a steady and significant level in females as well (FIGURE 2A).

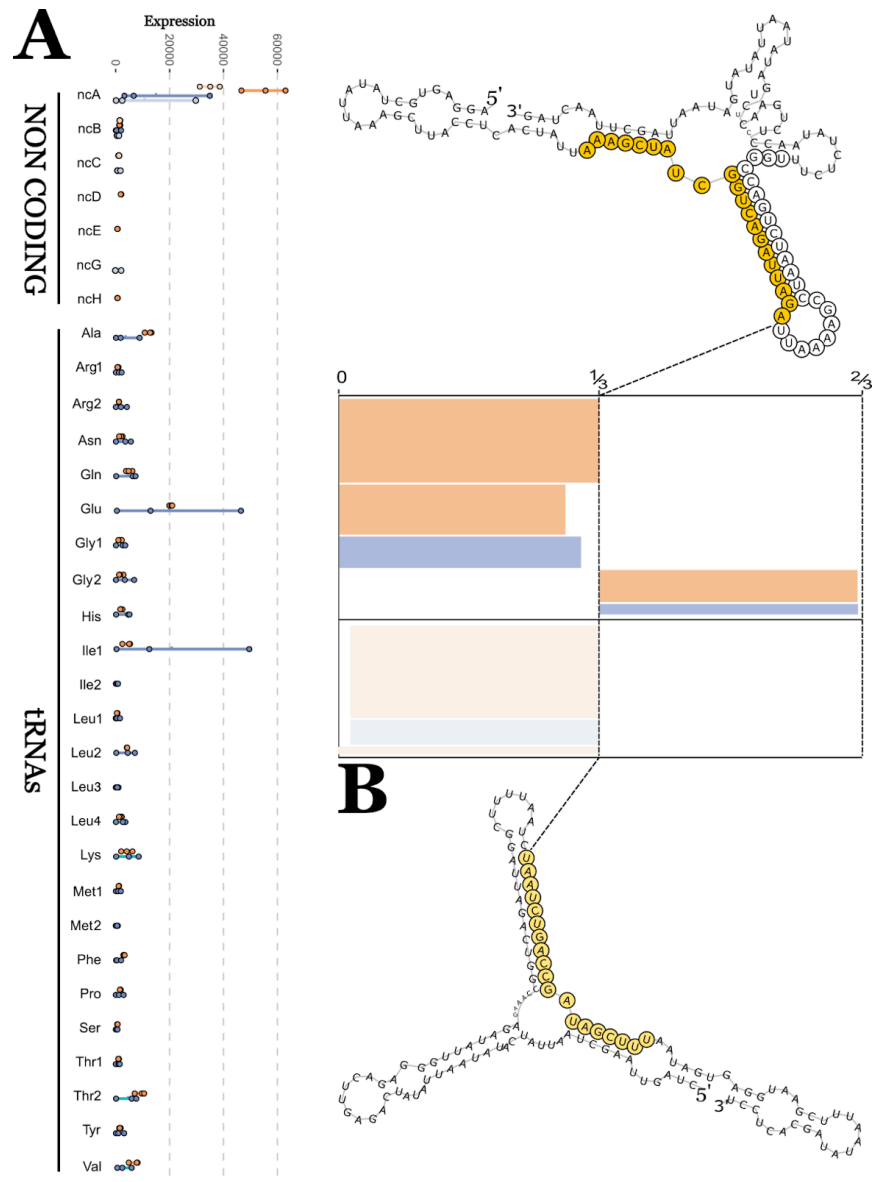


FIGURE 2: A) Expression of *Wolbachia pipientis* smallRNAs across samples **B)** ncA region produces smallRNAs from different parts of its precursor hairpin structures. The less expressed strand (lower part) is filled with faded colours.

Regarding the cleavage signatures of these bacterial RNA elements, ncA alone are able to generate three different elements from the same structure. The most expressed strand gives birth to two, while the third comes from the complementary sequence of the one with the highest expression (FIGURE 2B). Two are produced in the ncB region, while other non-coding regions and tRNAs usually give birth to only one, with very few exceptions (FIGURE S2).

Wolbachia RNA interference affects reproduction and behaviour in its host

The second module of the pipeline was run to find out which would be the potential targets of the RNA interference network I detected, with a focus on the most expressed sRNAs. Among the many potential targets with a high thermodynamic affinity and an extensive complementarity, there are some key genes who could be accounted for alterations usually associated with *Wolbachia* infection in the studied system, including some related to sex-determination and host behaviour.

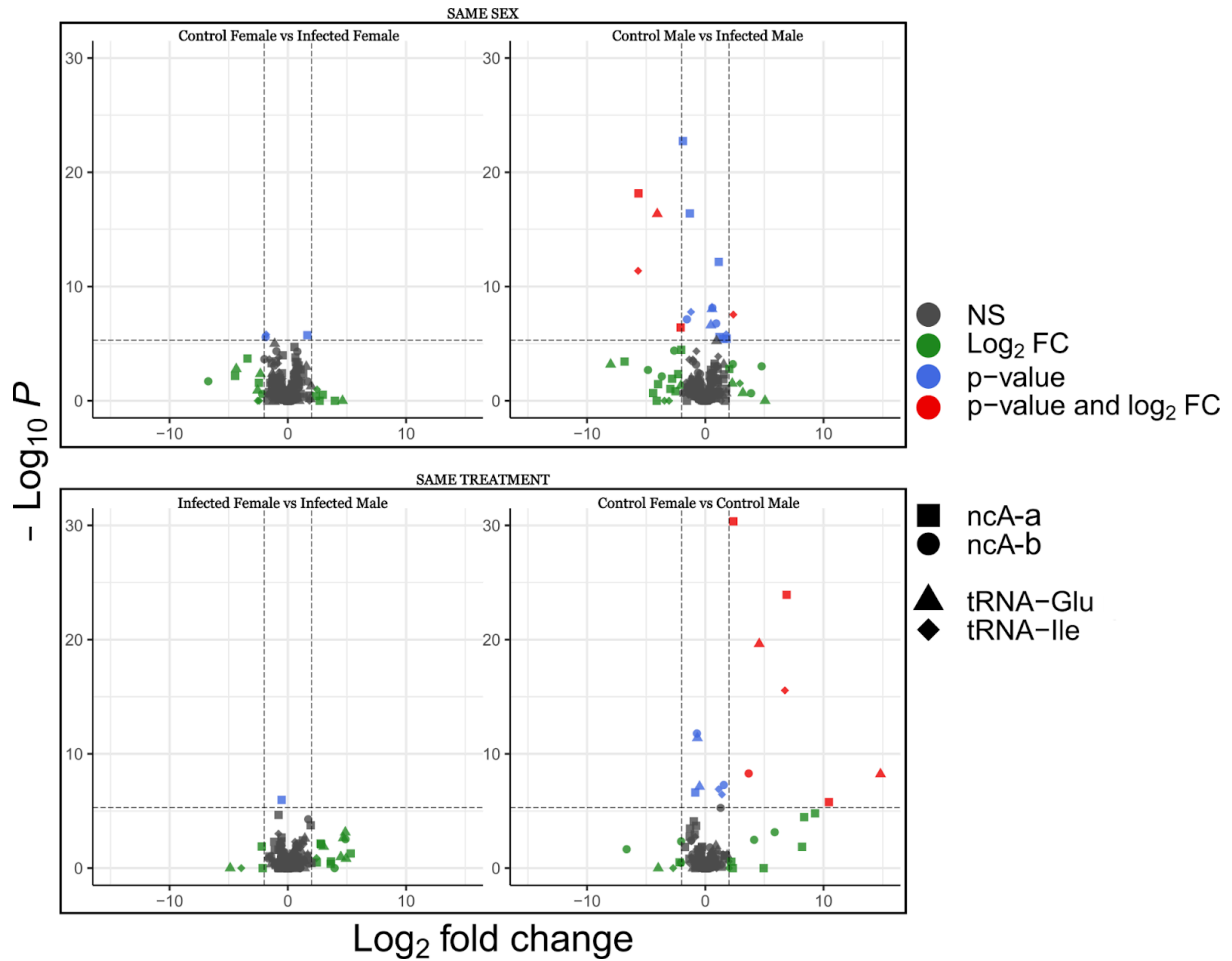


FIGURE 3: Differential expressed genes targeted by *Wolbachia pipientis* most expressed sRNAs. ncA-a is the highest expressed strand of ncA, while ncA-b is the other one. Dots are coloured when meeting a threshold on $\log_2$ FC (green when out of ± 2), p-value (blue when $< 5 \times 10^{-6}$), or both (red).

Interestingly, there are only 6 *Wolbachia* transcripts among sRNAs' targets, thus pointing out bacterial RNA interference is addressed almost entirely towards its host. As expected, I found *Wolbachia pipientis* transcripts to be present and upregulated only in infected samples. When comparing male infected and healthy samples, I can hypothesize a mainly silencing effect of bacterial sRNAs on their targets (FIGURE 3), and the same can be said while observing their distribution in uninfected samples.

With a first focus on ncA-originating RNA interference, I followed expressions of targeted genes across different comparisons. ncA region produces up to 5 sRNAs, but only 3 at a significant level (FIGURE S2). To enhance clarity, I grouped the two sRNAs from the dominant strand together and named them ncA-a (FIGURE 3). The sRNAs from the other strand were named ncA-b. Due to their similar expression profiles, I treated the targets of ncA-a and ncA-b as a single set of ncA targets.

Among them, I found many proteins involved in sex-related regulation: the most significant is HELLS_MOUSE, which has been localized in testis and seems to be essential for spermatogenesis (Zeng et al. 2011). Its expression is significantly lower in infected feminized males compared to control males. Furthermore, ncA's sRNAs can lower the presence of OVCH2_RHIAE and NEP2_DROME, which have crucial roles in sperm maturation and retention (Sitnik et al. 2014; Kent et al. 2025).

Moreover, ncA-mapping sRNAs target proteins that are not strictly related to sex differentiation or reproduction, but allow the bacterial endosymbiont to thrive in the host's cellular environment. As an example, they target a protein similar to GRAF_MOUSE, involved in immunoresponse towards malfunctioning cells and pathogens (Lord et al. 2003). It is usually a lot more expressed in males, but the infection with *Wolbachia* is twisting its expression to the same level observed for females. Additionally, there are many more proteins targeted by these sRNAs whose expression drops. Among them, I found CLB9_ANOGA, that is involved in the recognition of pathogens (Sousa et al. 2020), and SNCP4_MOUSE, likely affecting sRNAs transcriptions (Su et al. 2024). On the contrary, there are less, but fundamental proteins over-expressed in infected samples. KLC_DROME, a motor protein, is fundamental in organellar transport (Hirokawa et al. 2009). An homolog to ARID2_HUMAN is clearly upregulated in females, both conditions, and in infected males. ARID2_HUMAN is involved in chromatin remodeling (Li et al. 2011), thus altering the production of many other genes. It has been observed to be frequently targeted by other microRNAs, promoting it as a key player in the control of cellular processes.

As a second matter of investigation, I conducted the same analyses on *Wolbachia*'s tRNAs-originating sRNAs: the concentration comparable with ncA fragments suggested that similar potential effects were underlying.

The tRNA-Glu bacterial structure is able to generate a sRNA element targeting, again, several biological processes of the host. The most significant, in terms of p-value, is MOEH_CULQU, a protein intrinsically associated, in similar way to ncA target KLC_DROME, with cytoskeletal actin (Tsukita et

al. 1997) and upregulated in infected males and both condition females, compared to control males. The second target partially matches RLR78_PLAVT, observed to suppress cell death triggering process (Tomczynska et al. 2018), and colocated with other organellar entities such as mitochondria and chloroplasts. Another related activity it can have is, as seen before with ncA targets, chromatin state alteration (Giannakouros et al. 2011), a sometimes chronic condition that could lead to CI, as observed in other *Wolbachia*-host systems (Horard et al. 2022; Terretaz et al. 2023). The same sRNA upregulates DOM_DROME too, with function both in chromatin remodeling and oogenesis. It targets TIM_HUMAN as well, and thus its multiple activities (Cai and Chiu 2022). Finally, SCN60_DROME is an ion channel responsible for olfactory response (Kulkarni et al. 2002), and its significant downexpression in both sex of infected individuals agrees with other previous observation of bacterial-induced olfactory alterations (Baião et al. 2019; Warecki et al. 2025).

The other overexpressed bacterial sRNA originates on tRNA-Ile and regulates numerous aspects of host molecular response as well. The most interesting scheme of expression comes from SPE_DROME, with a definitely higher degree in control males, compared to all the three other conditions. This gene has the ability to affect both immunity and development (Jang et al. 2006), together with PFAF2_GALME and other downregulated proteins. tRNA-Ile's sRNA targets an olfactory protein as SNA_DROME too. Indeed, similar to the other olfactory related target, it is downregulated in infected individuals. Of particular interest for this sRNA target is CPE1A_XENLA, a gene often involved in RNAi (John et al. 2004). Its upregulation in infected males would play as a transcription regulator for the oocyte maturation (Sha et al. 2017).

High identity of ncA region in Wolbachia genome

Finally, I wondered if the observed RNAi, or its blueprint, could be detected in other *Wolbachia pipientis* strains. Regarding tRNAs-derived sRNAs, it is likely that different strains would have similar, if not the same, tRNAs. Since they have to accomplish the common function of tRNAs (i.e. loading amino acids on a template), I expect them to be under constraints. The same could not be said for the ncA region, especially given its being theoretically “non coding”. As underlined by previous studies (Woolfit et al. 2015), I found a shared identity of this region in 40 of the 43 genomes investigated. These high conservation degrees would suggest that the ncA region is very important among *Wolbachia* strains, likely because of the same RNA interfering function it has in the system I studied here. The distribution of non coding region conservation scores deviated from normality

(Anderson-Darling, p -value $< 2.2e-16$). The ncA region was significantly more conserved than the background distribution of other 70 bp long non-coding fragments (FIGURE 4). Additionally, I observed that this high conservation is remarkable among more than one thousand *Wolbachia* genomes (Scholz et al. 2020), regardless of their supergroup affiliations. The conservation is not limited to the ncA sequence itself, but it concerns the whole region around it, suggesting the presence of promoters and transcription regulators (Lamoureux et al. 2024). At the other end, it could be explained by considering the ncA region as an accessory sequence of the closest gene, and so, transcribed along with it (Ponath et al. 2022). Indeed, promoters prevision predict a suitable sequence for the precursor of ncA-b, starting at -35 bases from the Transcription Start Site (TSS), at the beginning of the hairpin structure. No promoters were found in the proximity of the predicted ncA-a precursor, suggesting a different biogenesis for this element.

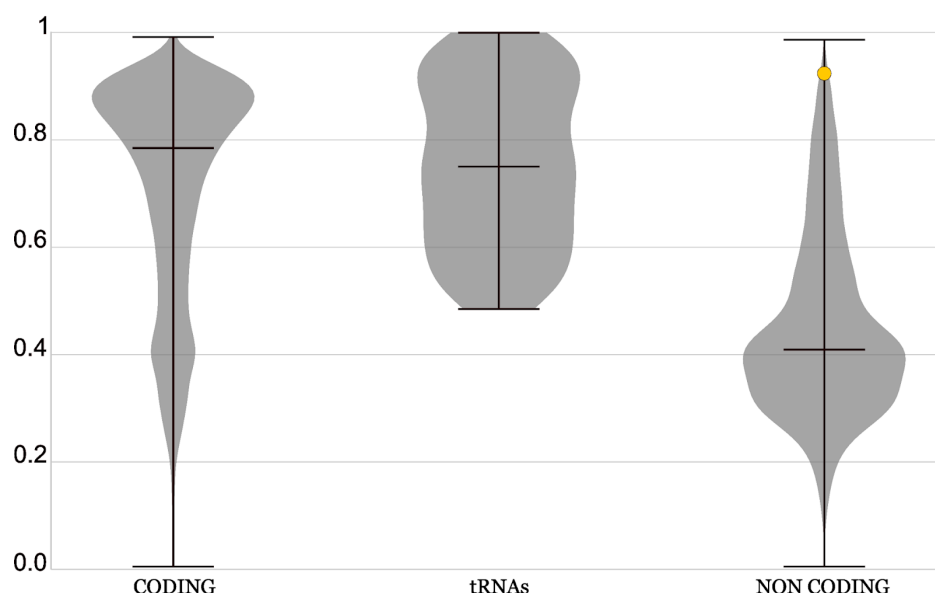


FIGURE 4: Conservation scores, grouped by type, of the 43 available reference genomes of *Wolbachia pipientis*. Yellow point highlights ncA fragment's value.

4.3 Discussion

Large alteration of gene expression related to Wolbachia infection

Wolbachia pipientis parasites a wide range of hosts invading various host tissues. This success would not be possible if not by interacting with various host systems and pathways to ensure successful intracellular maintenance (Werren et al. 2008; Sullivan and Porter 2023). In insects, *Wolbachia* can alter host fitness through changes in pathogen susceptibility, immunity response, lifespan and behaviour (Maistrenko et al. 2016; Vandana et al. 2024; Pimentel et al. 2025; Warecki et al. 2025). Many different effectors are exploited by the intracellular symbionts (Rice et al. 2017), including sRNAs (Mayoral et al. 2014; Woolfit et al. 2015).

My results confirm that *Wolbachia* heavily impacts the host transcriptome (FIGURE 1). Infected samples have similar gene expression profiles (Asgharian et al. 2014), while control samples stand aside, with sex-specific molecular signatures. DE analyses and GO terms enrichment confirm that the upregulated sperm-related development processes in healthy males is inhibited in infected individuals (Yuan et al. 2015), in line with other feminization effects observed in this species (Negri et al. 2006). Furthermore, *Wolbachia* enhances metabolic pathways and promotes lipids production, otherwise rapidly consumed by both the bacteria population and its host cells (Baião et al. 2019; Dou et al. 2021). However, such effects are likely occurring at a detectable degree when parasite titer is high in the host cell, suggesting this was the case. Indeed, immune response overexpression manifests in both sexes, highlighting *Wolbachia*'s ability in triggering host's defense mechanisms (Zhang et al. 2020). Although *Wolbachia* itself is able to circumvent the alert state of the host immune system (Wilmes et al. 2017), this sustained defense mechanism yields a protective benefit for the leafhopper, shielding it against opportunistic pathogens (Vandana et al. 2024).

Wolbachia induced alterations concern the host's microRNAs too, in line with multiple other studies (Osei-Amo et al. 2012; Zhang et al. 2013; Bi et al. 2019; Bishop et al. 2022). As observed for canonical genes, the infection shifts the healthy male's microRNA expression towards the female's profile. In *Drosophila melanogaster*, alteration of nov-miR-12 alone leads to a decreased hatching rate (Zheng et al. 2019). The 16 differential expressed microRNAs between control and infected males would be critical for sex-related features development, resulting in a gene network alteration that supports feminized phenotypes.

Wolbachia produces a massive amount of sRNAs

The pervasive role of proteins capable of generating and processing functional sRNAs across life, spanning eukaryotes (Iwakawa and Tomari 2022) and bacteria (Dutta and Srivastava 2018), establishes sRNA-mediated modulation as a fundamental backbone of cellular regulatory mechanisms. The majority of identified sRNAs derive from the non-coding regions and as tRNA-derived fragments (tRFs). tRFs exhibit a distinct biogenesis, originating preferentially from one end of the tRNA structure, a polarity previously noted for bacterial, mitochondrial and nuclear tRNAs (Kumar et al. 2014; Shaukat et al. 2021; Diallo et al. 2022). The major source, the ncA region, also displays a favored sequence for sRNA production. However, fragments from other structural arms and the complementary strand are also significantly expressed. The antisense fragments suggest a complex, potentially dual-strand biogenesis (Lasa et al. 2011). While a canonical promoter was found for one strand, the origin of the other could involve undetected, highly divergent promoter motifs (Rangel-Chávez et al. 2021), a plausible scenario given the reductive genome evolution of endosymbionts, or co-transcription with neighboring genes (Yang et al. 2013; Ponath et al. 2022). Regardless of their initial transcription, the stable hairpin precursors are ideally suited for processing by bacterial endonucleases. Furthermore, the confirmed high conservation of the ncA sequence across various *Wolbachia* strains strongly suggests its functional significance. Other prior investigations (Woolfit et al. 2015) reported its over transcription in testis of infected *Drosophila* individuals. The lower (and fluttering) expression observed in *Z. pullula* males could be explained by the pooling of individuals and whole-body sequencing, which might have smoothed out existing differences in expression between males and females. On the other hand, a genuinely variable expression of ncA's sRNAs would partially account for the diverse degrees of feminine feature development occurring in feminized males (Negri et al. 2006). Future studies should concurrently monitor ncA sRNA expression and upper pygofer appendage development to investigate a potential link between the two. My findings indicate that ncA-related sRNAs can downregulate numerous sex-related genes via direct binding. Their presence in *Wolbachia* strains that induce cytoplasmic incompatibility (Woolfit et al. 2015), however, complicates their role. Either these sRNAs are not the specific effectors of feminization, or their function is context-dependent, producing different reproductive manipulations based on the host's distinct sex-determination network. Consequently, additional studies on other biological systems where *Wolbachia* induces feminization are required to

validate the generality of these specific sRNA regulatory mechanisms. As for ncA, other *Wolbachia* sRNAs show stable expression levels across female replicates, suggesting a consistent *Wolbachia* concentration. In stark contrast, male samples displayed broader sRNA expression variation, particularly for tRNA-Glu and tRNA-Ile fragments. This high variability in males, as previously mentioned, would likely stem from a broader spectrum of bacterial titers across samples. Regardless of this concentration variability, all functional sRNAs must ultimately traverse the bacterial membrane to exert their regulatory effects on the host, either by being exported to the cytoplasm within vesicles (H. Zhang et al. 2020; Sahr et al. 2022), or facilitated by Hfq proteins (Malabirade et al. 2017). Upon arrival, these fragments necessitate loading onto host Argonaute proteins to form the RNA-induced silencing complex (RISC) and execute their regulatory function imitating host's microRNAs, as reported for other endoparasitic bacteria (Sahr et al. 2022). Further investigation is therefore required to confirm interactions between *Wolbachia* sRNAs and host RNAi machinery.

Wolbachia RNA interference as cross-kingdom communication

Numerous studies emphasized that pathogenic bacteria utilize RNAi pathways not only for internal gene regulation but also as a mechanism to manipulate external cellular environments (Mayoral et al. 2014; Woolfit et al. 2015; Singh et al. 2015; Koeppen et al. 2016; H. Zhang et al. 2020; Sahr et al. 2022). My analysis revealed that *Wolbachia* sRNAs target both bacterial and host transcripts. The few *Wolbachia* targets identified are primarily controlled by cis-acting elements, with one key exception: an sRNA originating from the ncG region modulates the expression of the P44/Msp2 family outer membrane protein. As this protein is essential for host cell invasion (Ren et al. 2025), its regulation is likely critical to the establishment and persistence of the endosymbiosis. Still, the predominant and most impactful role of *Wolbachia* sRNAs lies in their ability to directly target and manipulate host transcripts. The predicted cross-kingdom RNA interference would be a critical mechanism for the endosymbiont to exert broad control over host physiology, including sex determination, immune responses, and behavior. Through reproductive manipulation, the silencing effects of ncA fragments on host transcripts suggest a direct molecular pathway for compromising sperm development, consistent with *Wolbachia*'s handling of host fertility (Werren et al. 2008). For immune evasion, despite a general defence system activation in infected samples, specific genes related to pathogen recognition are repressed (Jang et al. 2006). Targeting these components would represent a practical mechanism for *Wolbachia* to bypass host immune surveillance and avoid clearance. RNAi would also

provide cellular and genetic control by exerting heavy influence over key genes in the chromatin remodeling network, a CI inducing mechanism that could globally change the transcription of many other host cellular processes (Riparbelli et al. 2012). Furthermore, *Wolbachia* sRNAs mediated regulation would be critical to ensure robust transmission. Internally, they promote cytoskeletal development and mobility, an action that facilitates the bacteria's crucial germline localization (Ferree et al. 2005; Newton et al. 2015). Externally, the bacterial sRNAs target the host's olfactory system, an ability shared with other host's microRNAs (Bi et al. 2019), providing a molecular basis for observed alterations in host behavior and fitness, both of which contribute to stable vertical inheritance.

4.4 Materials and Methods

Female and male individuals of *Zyginidia pullula* were collected on grass in different localities, and then reared individually in the laboratory inside climatic chambers ($23\pm1^{\circ}\text{C}$, $60\pm10\%$ relative humidity) on potted maize plantlets. *Wolbachia pipientis* infections have been assessed through observation at the microscopy. Total DNA for each condition (female and male controls, female and male infected by *Wolbachia*) have been extracted from single specimens through the Wizard Genomic DNA Purification kit (Promega) and assessed for quality using a NanoDrop ONE and Qubit (Thermo Fisher Scientific). Libraries have been prepared and sequenced at Macrogen (Amsterdam, The Netherlands) using TruSeq PCR free libraries and a $2\times150\text{bp}$ strategy in a NovaSeq6000 machine targeting 20 Gigabases per sample. Total RNA has been extracted from a pool of individuals for each condition, with three replicates each, using the RNeasy Mini Kit (Qiagen) and assessed in the same way of DNA. mRNA and smallRNA were sequenced at Macrogen using TruSeq stranded mRNA libraries, $2\times150\text{bp}$ strategy in a NovaSeq6000 machine, targeting 9 Gigabases per sample, and NEBNext Small RNA libraries (New England Biolabs), $2\times150\text{bp}$ strategy in a NovaSeq6000 machine targeting 7.5 Gigabases per sample, respectively.

I used Whole Genome Sequence libraries to assemble both the *W. pipientis* and *Z. pullula* genome. For the bacterial genome, I mapped trimmed reads (provided by the sequencing facility) against *Wolbachia* pangenome, made on the same 43 reference genomes available from Scholz et al., 2020 (TABLE S2). Mapping reads have been processed independently for each sample with the Read Assembly and Annotation Pipeline Tool (RAPT, v.0.5.5), which assembly using SKESA v. 2.5.0 (Souvorov et al. 2018) and predict coding and non coding genes with PGAP v.2023-05-17.build6771. Checkm v. 1.2.3 (Parks et al. 2015) tested for the correct lineage assignment of the resulting genome. Then, I used mummer v. 4.0.1 to check for the identity degree of the two bacterial genomes, and scanned them for promoter recognition using Promotech v. 1.0 (Chevez-Guardado and Peña-Castillo 2021).

Heterozygosity level of the remaining reads (i.e. not mapping on *Wolbachia* pangenome) has been assessed through k-mer frequency inference with genomescope 2.0 (Ranallo-Benavidez et al. 2020). I assembled the genome of *Z. pullula* with SPAdes v. 4.0.0 and checked it with BUSCO v. 5.8.2 (Manni et al. 2021).

TruSeq libraries have been used to assemble the transcriptome and assess gene expression. I trimmed the reads with fastp v. 0.23.2 (Chen et al. 2018) and filtered out reads of ribosomal origin with SortMeRNA v. 4.3.6 (Kopylova et al. 2012). Transcriptome has been assembled with Trinity v. 2.8.5 (Grabherr et al. 2011), with default settings. I evaluated transcriptome's completeness with BUSCO, against insect orthologues dataset (insecta_odb10). I predicted the transcripts coding sequences with Transdecoder v. 5.5.0 and Trinotate v. 3.11, and retrieved associated Gene Ontology terms with Interproscan v. 5.74.105.0 (Jones et al. 2014). I produced expression counts using align_and_estimate_abundance.pl and abundance_estimate_to_matrix.pl from the Trinity packages, with Kallisto v. 0.28.0 (Bray et al. 2016) as the estimation method. I ran the differential expression (DE) analyses in R v. 4.5.0 (R Core Team 2025), using edgeR v. 4.6.2 (Chen et al. 2025) for the count per million function (CPM), keeping only transcripts with a minimal expression of 3 CPM, and DESeq2 v. 1.28.1 (Love et al. 2014) for the weighted comparisons between groups. I plotted VolcanoPlots for the DE results using the EnhancedVolcano library v. 1.26.0 (Blighe et al. 2025). GO terms enrichment, associated with each comparison, has been displayed with REVIGO (Supek et al. 2011).

miRNA families were annotated on the *Z. pullula* genome using MirMachine (Umu et al. 2023), and validated by sRNA libraries using miRDeep2 (Friedländer et al. 2012), providing high-confidence *Z. pullula* pre-miRNAs predicted by MirMachine and the miRNA sequences of five insects (namely, *Blattella germanica*, *Tribolium castaneum*, *Heliconius melpomene*, *Drosophila melanogaster*) present on MirGeneDB 3.0. (Clarke et al. 2025). I locally aligned *Z. pullula* miRNAs against the MirGeneDB 3.0 database using blastn to assign miRNAs to the respective miRNA family. To confirm the miRNA family annotation, I manually checked that the seed of the *Z. pullula* miRNA matched the seed of the assigned miRNA family. I performed the DE analysis on *Z. pullula* miRNAs between the different conditions using the miRNA-specific sRNA read counts per condition generated by miRDeep2 and following the normalization steps used for gene expression.

I investigated *Wolbachia pipientis* smallRNAs production and targeting with SmithHunter. Although the tool is fine-tuned to work with circular mitochondrial genomes, which size around 15-20 thousand bases, it is able to process larger genomes as well. I ran the first module with default parameters (identity at 90%, stringency at 80% and 5' threshold of 0.5) but not accounting for a limit in the number of replicates passing filters (i.e. a cluster could be present in only one replicate to be considered as a

presumptive smallRNA). The analyses filtered out all the reads that originate from the host (both the nuclear and mitochondrial genome) and kept only smallRNAs transcribed from the *Wolbachia* genome. With the second module of SmithHunter, I selected putative targets of presumptive smallRNAs, allowing 1 mismatch in the seed region of the smallRNA. Given that small RNAs can bind to coding sequence as well (Hausser et al. 2013), a finding complemented by evidence of *Wolbachia* sRNAs targeting host CDS (Mayoral et al. 2014), I decided to run the prediction tool against the entire transcriptome of *Z. pullula*.

The conservation of the non-coding A (ncA) region was determined against 43 *Wolbachia pipientis* reference genomes (Scholz et al. 2020), analyzing coding, non-coding, and tRNA regions independently. Concatenated sequences were fragmented into 70 bp segments (with a 10 bp sliding window). With blastn v. 2.12.0 (S F. Altschul et al. 1990), I retained only the best hit per genome for each fragment, and the final conservation score was defined as the mean identity across all best hits. RNA secondary structures (FIGURE 2B) of ncA has been predicted by the RNAfold WebServer (Gruber et al. 2008).

4.5 Supplementary Materials

Comparison	microRNA	Exp1°	Exp2°	log2FC	Padj
Control, male vs female	Dme-Mir-965_pre_51	6,46E+01	4,53E+00	3,85	4,60E-10
	Dme-Mir-965_pre_52	6,46E+01	4,53E+00	3,85	4,60E-10
	Bge-Novel-1-P2_pre_11	2,34E-01	3,71E+01	-7,16	2,70E-07
	Dme-Mir-965_pre_50	4,31E+01	4,23E+00	3,39	6,38E-07
	Bge-Mir-252-P1_pre_37	0,00E+00	1,42E+01	-6,74	3,83E-06
	Bge-Mir-10-P3_pre_48	5,90E+01	1,59E+01	1,89	2,24E-05
	Bge-Novel-1-P2_pre_25	1,60E+00	4,60E+01	-4,81	2,24E-05
	Bge-Novel-1-P2_pre_26	1,60E+00	4,60E+01	-4,81	2,24E-05
	Bge-Novel-1-P2_pre_27	1,60E+00	4,60E+01	-4,81	2,24E-05
	Bge-Mir-10-P3_pre_47	1,06E+02	3,07E+01	1,78	6,67E-04
	Tca-Mir-3825_pre_61	1,69E+01	1,14E+00	3,92	8,55E-04
	Tca-Mir-2-P5c_pre_73	1,90E+01	2,90E+00	2,72	2,19E-03
	Bge-Mir-2765_pre_13	6,89E+01	1,38E+02	-1,01	2,19E-03
	Bge-Mir-2765_pre_14	6,89E+01	1,38E+02	-1,01	2,19E-03
	Bge-Mir-71_pre_7	7,06E+03	1,13E+04	-0,68	2,19E-03
	Bge-Mir-2765_pre_23	5,06E+01	1,13E+02	-1,15	2,56E-03
	Bge-Mir-2765_pre_24	5,06E+01	1,13E+02	-1,15	2,56E-03
	Bge-Mir-10-P3_pre_31	3,65E+02	6,31E+02	-0,79	1,72E-02
	Bge-Mir-96-P3_pre_59	1,13E+01	7,77E-01	3,98	2,01E-02
	Tca-Mir-2-P5c_pre_0	1,95E+02	3,30E+02	-0,76	2,01E-02
Infected, male vs female	Bge-Novel-1-P2_pre_11	3,65E+01	1,15E+02	-1,72	0,036576
	Bge-Mir-252-P1_pre_37	3,82E+00	2,51E+01	-2,96	0,036576
Female, infected vs control	Bge-Mir-10-P3_pre_48	5,86E+01	1,59E+01	1,89	0,000273
	Bge-Novel-1-P2_pre_11	1,15E+02	3,71E+01	1,63	0,014383

Male, infected vs control	Bge-Novel-1-P2_pre_11	3,65E+01	2,34E-01	7,07	2,44E-06
	Bge-Mir-9-P10_pre_2	6,84E+02	3,72E+02	0,89	1,89E-03
	Tca-Mir-2-P5c_pre_0	4,44E+02	1,95E+02	1,14	1,89E-03
	Bge-Novel-1-P2_pre_25	2,33E+01	1,60E+00	3,71	7,01E-03
	Bge-Novel-1-P2_pre_26	2,33E+01	1,60E+00	3,71	7,01E-03
	Bge-Novel-1-P2_pre_27	2,33E+01	1,60E+00	3,71	7,01E-03
	Dme-Mir-965_pre_51	2,12E+01	6,46E+01	-1,88	8,67E-03
	Dme-Mir-965_pre_52	2,12E+01	6,46E+01	-1,88	8,67E-03
	Bge-Mir-252-P1_pre_37	3,82E+00	0,00E+00	4,59	1,55E-02
	Tca-Mir-3825_pre_60	0,00E+00	1,16E+01	-4,68	2,00E-02
	Dme-Mir-965_pre_50	1,05E+01	4,31E+01	-1,86	2,59E-02
	Bge-Mir-10-P3_pre_47	3,85E+01	1,06E+02	-1,41	2,75E-02
	Bge-Mir-9-P10_pre_6	1,90E+01	4,93E+00	2,03	3,57E-02
	Tca-Mir-3863_pre_49	0,00E+00	7,16E+00	-4,00	4,63E-02
	Tca-Mir-3825_pre_61	3,19E+00	1,69E+01	-2,94	5,00E-02
	Bge-Mir-12_pre_35	9,14E+01	1,52E+02	-0,76	5,00E-02

TABLE S1: Differential expression of *Zyginidia pullula* microRNAs between conditions. Exp1° is the expression of the first term in the comparison. Exp2° is the expression of the second term in the comparison. Log2FC is the logarithmic fold change between the two conditions. Only microRNAs with a Padj (adjusted P-value) < 0.05 are shown.



Sperm-related

Cytoskeletal-related activity

Chromosome and chromatin related



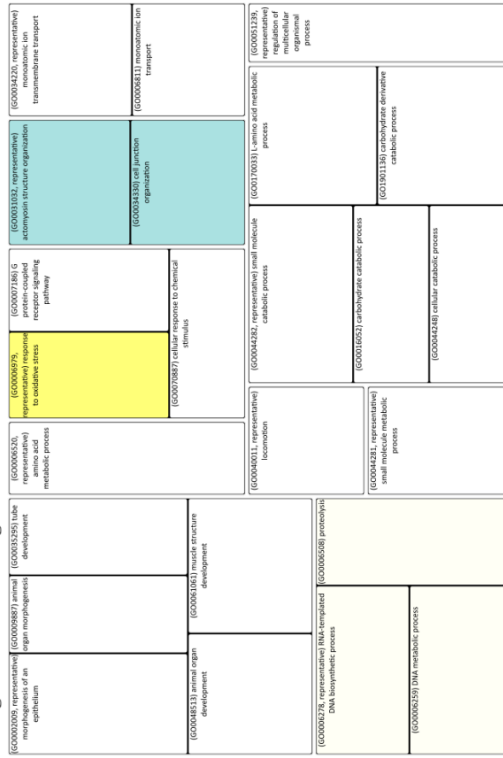
Microtubule activities

Lipid transport and deposition

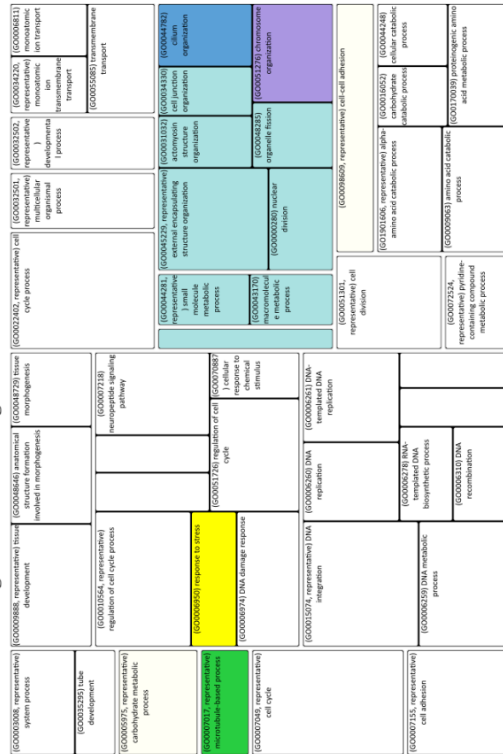
Stress and immune response

FEMALES, CONTROL VS INFECTED

Biological Process UP regulated in CONTROL FEMALES

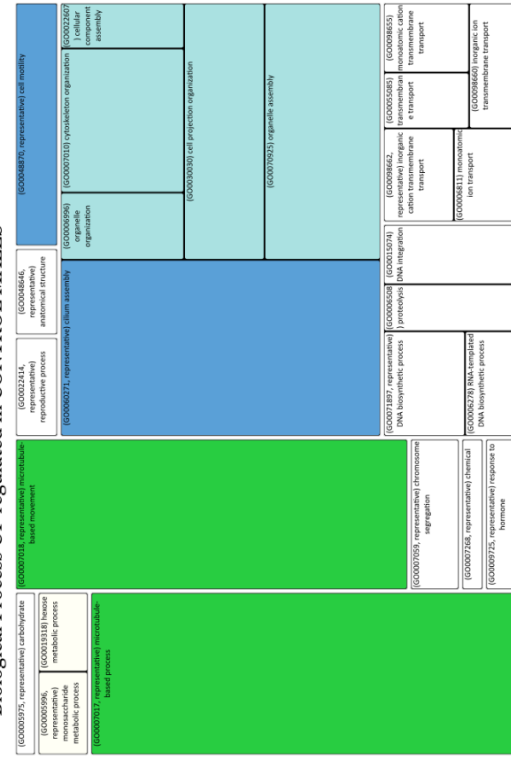


Biological Process UP regulated in INFECTED FEMALES

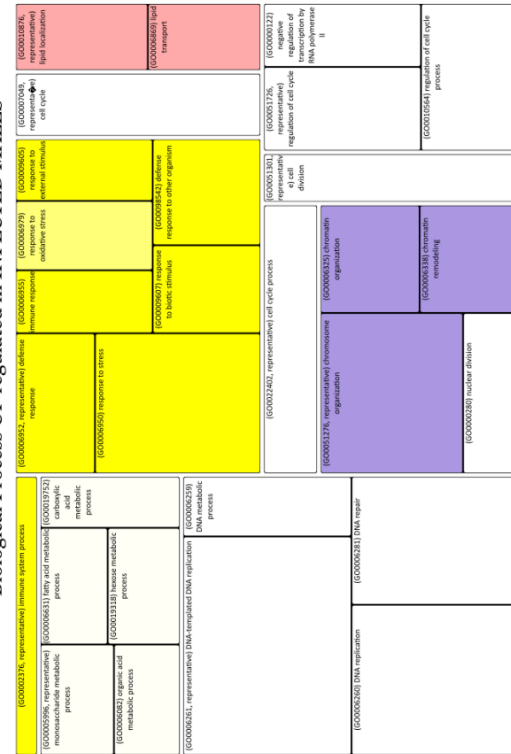


MALE, CONTROL VS INFECTED

Biological Process UP regulated in CONTROL MALES



Biological Process UP regulated in INFECTED MALES



[illegible][illegible]

84

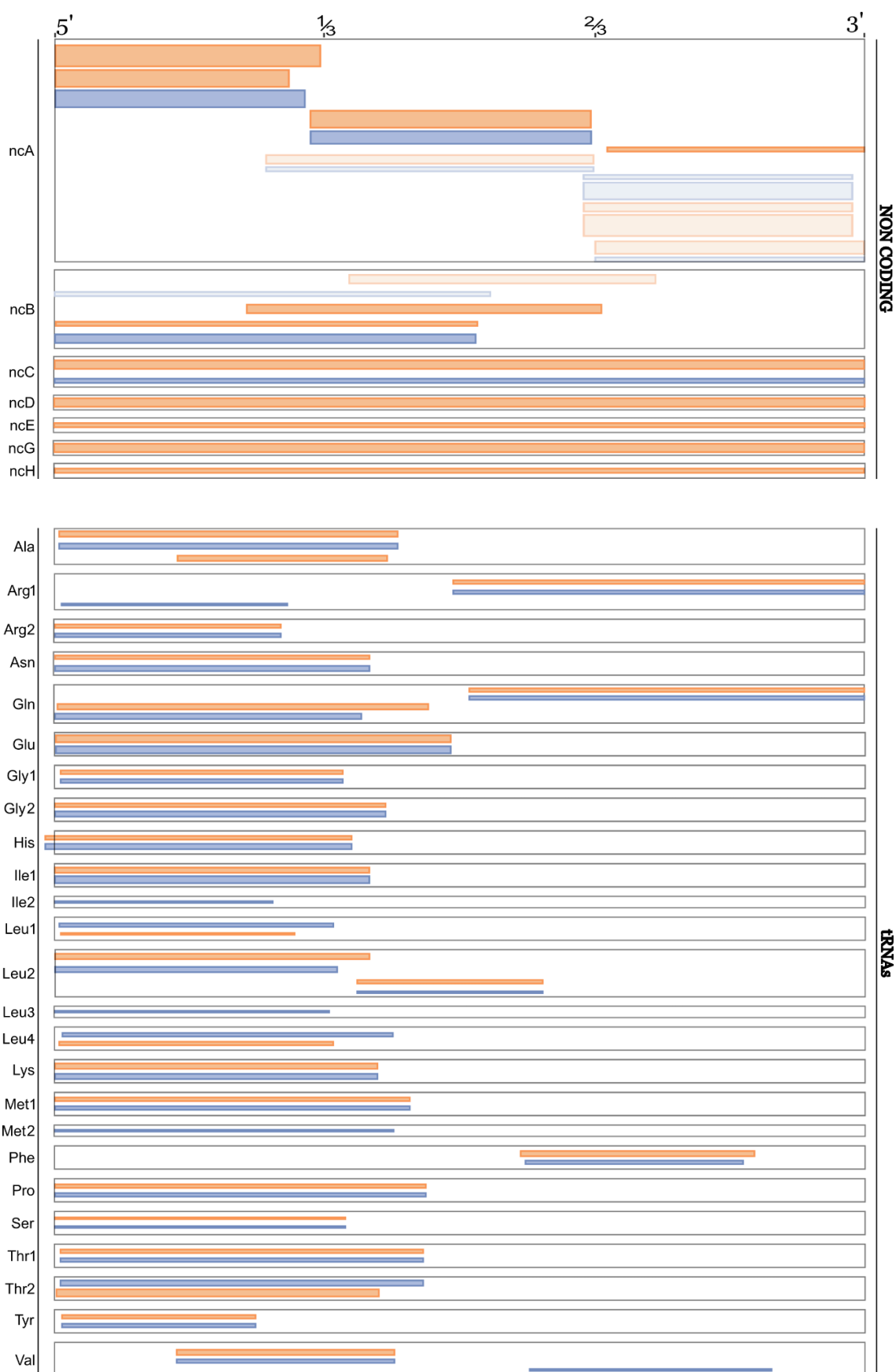


FIGURE S2: Genomic location of *Wolbachia* sRNAs. Only sRNAs originating on non-coding regions and tRNAs are displayed. Faint colours indicate sRNAs on the alternative strand (i.e. strand with a lower expression of sRNAs). To improve readability of sRNAs of small cluster size, the bar widths are not drawn to scale.

Strain	NCBI accession number	Contigs	Length	Group	Host
wOv_Cameroon	Cameroon_GCF_000530755	1	960618	C	<i>Onchocerca volvulus</i>
wPip_JHB	JHB_GCF_000156735	16	1543661	B	<i>Culex quinquefasciatus</i>
wOb	Ob_Wba_GCA_001266585	120	1121046	B	<i>Operophtera brumata</i>
wAlbB	wAlbB_GCF_000242415	156	1162431	B	<i>Aedes albopictus</i>
wAna	wAna_GCA_000167475	464	1440750	A	<i>Drosophila ananassae</i>
wAu	wAu_GCA_000953315	1	1268461	A	<i>Drosophila simulans</i>
wBn	wBn_GCA_000338395	763	1052327	D	<i>Wuchereria bancrofti</i>
wBol1-b	wBol1-b_GCA_000333775	144	1377933	B	<i>Hypolimnas bolina</i>
wCle	wCle_GCF_000829315	1	1250060	F	<i>Cimex lectularius</i>
wCoc	wCoc1_GCA_000832695	1064	1124974	B	<i>Dactylopius coccus</i>
wDim	wDim_nematodes_org	2	921012	C	<i>Dirofilaria immitis</i>
wDi	wDi_WACP3_GCF_000331595	104	1240904	B	<i>Diaphorina citri</i>
wGmm	wGmm_GCF_000689175	201	1019510	A	<i>Glossina morsitans morsitans</i>
wHa	wHa_GCF_000376605	1	1295804	A	<i>Drosophila simulans</i>
wLs	wLs_nematodes_org	10	1048936	D	<i>Litomosoides sigmodontis</i>
wMel	wMel_GCF_000008025	1	1267782	A	<i>Drosophila melanogaster</i>
wMelPop	wMelPop_GCF_000475015	25	1239155	A	<i>Drosophila melanogaster</i>
wBm_TRS	wMel_wBm_TRS_GCF_000008385	1	1080084	D	<i>Brugia malayi</i>
wNo	wNo_GCF_000376585	1	1301823	B	<i>Drosophila simulans</i>
wOo	wOo_GCF_000306885	1	957990	C	<i>Onchocerca ochengi</i>
wOv	wOv_GCA_000338375	341	444344	C	<i>Onchocerca volvulus</i>
wPip	wPipMol2010_GCF_000208785	888	1479531	B	<i>Culex pipiens molestus</i>
wPipMol	wPipMol_GCF_000723225	1	1340443	B	<i>Culex molestus</i>
wPipPel	wPipPel_GCF_000073005	1	1482455	B	<i>Culex quinquefasciatus</i>
wRec	wRec_GCF_000742435	43	1126656	A	<i>Drosophila recens</i>
wRi	wRi_GCF_000022285	1	1445873	A	<i>Drosophila simulans</i>
wSim	wSim_GCF_000167495	629	1063100	A	<i>Drosophila simulans</i>
wSuzi	wSuz_GCF_000333795	110	1415350	A	<i>Drosophila suzukii</i>
wTpre	wTPRE_GCA_001439985	1	1133809	B	<i>Trichogramma pretiosum</i>
wUni	wUni_GCF_000174095	256	867873	A	<i>Muscidifurax uniraptor</i>
wVitB	wVitB_GCF_000204545	426	1107643	B	<i>Nasonia vitripennis</i>
wVul	wVulC_GCF_001027565	10	1663741	B	<i>Armadillidium vulgare</i>
wWil	wWil_GCA_000153585	53	1145915	A	<i>Drosophila willistoni</i>
wStri	wStri_GCF_001637495	99	1232346	B	<i>Laodelphax striatella</i>
wDacA	wDacA_GCF_001648025	157	1170639	A	<i>Dactylopius coccus</i>
wDacB	wDacB_GCF_001648015	198	1497749	B	<i>Dactylopius coccus</i>
wNfe	wNfe_GCF_001675785	231	1337641	A	<i>Nomada ferruginata</i>

wNfla	wNfla_GCF_001675695	167	1332780	A	<i>Nomada flava</i>
wNleu	wNleu_GCF_001675715	182	1366719	A	<i>Nomada leucophthalma</i>
wNpa	wNpa_GCF_001675775	191	1344355	A	<i>Nomada panzeri</i>
wPpe	wPpe_GCF_001752665	12	975127	L	<i>Pratylenchus penetrans</i>
wInc_Cu	wInc_Cu_GCF_001758565	1	1267840	A	<i>Drosophila incompta</i>
wInc_SM	wInc_SM_GCA_001758585	1	1267664	A	<i>Drosophila incompta</i>

TABLE S2: Metadata of the 43 *Wolbachia* reference genome used for 1) building *Wolbachia* metagenome, to retrieve *Wolbachia* sequence in the genome assembly process and 2) evaluate conservation scores of CDS, tRNAs and non-coding regions. From Scholz et al., 2020.

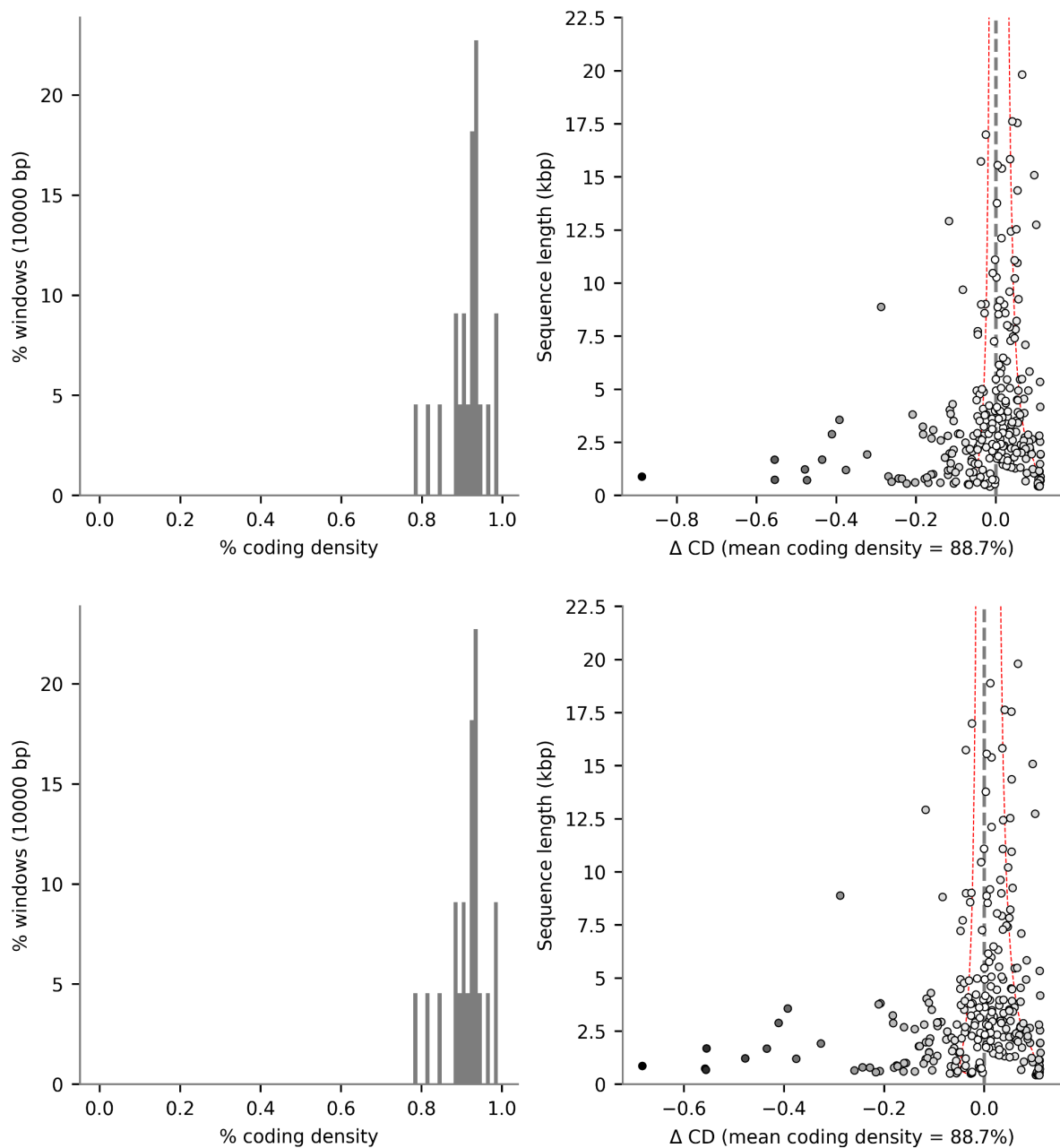


FIGURE S3: Coding density of the *Wolbachia* genome assembled from infected females (upper plot), and infected males (lower plot). The equal coding density reflects the 99.5% of collinearity measured between the two genomes.

5. Conclusion

Cellular life necessitates intricate regulation to ensure the cooperation of its multiple genomic entities. The emerging complexity requires a precise mechanism to fine tune each compartment with other components in the cytosolic environment. This is particularly evident in mitochondrial and nuclear intercommunication, which relies on a diverse array of signaling molecules. Among these, RNA-mediated regulation has emerged as a sophisticated layer of control able to profoundly alter gene expression.

In this context, the recent discovery of a new class of sRNAs, defined as small mitochondrial highly expressed RNAs (smithRNAs), nominate them as key mediators in this intergenomic dialogue (Pozzi et al. 2017). The initial identification of these elements came from *Ruditapes philippinarum*, a species with Doubly Uniparental Inheritance in which mitochondria might be implicated in sex determination. Furthermore, smithRNAs have been later found in another DUI bivalve (Smith et al. 2023), as well as model organisms (Passamonti et al. 2020). In the first chapter of my thesis I presented the development of SmithHunter (Marturano et al. 2024), a robust and customizable tool for the detection of smithRNAs from sequencing data. This pipeline would set a standard in the recognition of mitochondrial sRNAs in further studies, and was later applied in the analyses of the following research. Instead of the need for the users to assess each putative smithRNAs independently, SmithHunter employs a combination of parameters to evaluate all potential read clusters. When validated, the tool successfully retrieved nearly all smithRNAs proposed in the original work (Pozzi et al. 2017), as well as the male restricted smithRNA of *Potamilus streckersoni* (Smith et al. 2023), whose discovery were based on a different methodology. Therefore, SmithHunter proved to be a well suited instrument for detecting genuine mitochondrial encoded sRNA signals, capable of selecting any region with significant expression and clear signs of *ad hoc* transcription or processing. Indeed, when the above conditions are met, the pipeline also reports other mitochondrial sRNAs, such as tRNA-derived fragments (tRFs). Mitochondrial tRFs have been previously described as relevant in biological processes like development (Hirose et al. 2015) and stress response (Meseguer et al. 2019). So, it is not unexpected that mitochondria would leverage stable tRNA hairpin structures to produce sequences that could then be loaded into the RISC complex.

This hypothesis is especially important for Metazoa, whose densely packed mitochondria typically offer little space for non-coding regions (Formaggioni et al. 2021). My studies on the distribution of

smithRNAs in this group confirmed this concept. All 14 species considered produce smithRNAs. However, in contrast to the predominant non-coding origin observed in *R. philippinarum*, they exhibit a preferential origin from the processing of other transcribed features, namely tRNAs and rRNAs. This finding, coupled with the multiple and species-specific length distribution, would hint at different biogenesis and, as a consequence, dissimilar maturation pathways. While DICER is well-suited for canonical miRNAs or miRNA-like structure (Zapletal et al. 2022), other proteins are likely responsible for generating elements from transcripts with alternative conformations (Kumar et al. 2014; Fontanesi et al. 2020). Additionally, the analyses of potential smithRNAs target did not point to a single, specific function. Rather, the high variability in both the smithRNAs sequences (particularly the seed region) and the target-binding sites would likely determine unconserved, lineage specific roles. This supports the supposition of a regulatory system whose propensity to affect distinct gene networks results from a high probability to enter a new pathway, given the relatively high chance to be, or become, competent for a target (Plazzi et al. 2024).

The final chapter sought to trace the evolutionary origin of such a powerful tool within an endosymbiotic system which shared many traits with the mitonuclear duality. Both mitochondria and endosymbiotic bacteria are usually maternally transmitted, and their success is intrinsically linked to their ability to manipulate the host's sex ratio (Perlman et al. 2015). I purposefully choose a species, *Zyginidia pullula*, in which the infection of the intracellular bacterium *Wolbachia pipientis* promotes feminization (Negri et al. 2006), and thus, mirroring the hypothetical sex-determining role that smithRNAs seem to have in bivalves (Pozzi et al. 2017; Smith et al. 2023), the group in which they were first discovered. Generally, I found that *Wolbachia* infection induces a large-scale alteration of the host transcriptome, including the microRNA network, effectively shifting the male's sex-related gene expression toward the female phenotype and causing effects consistent with feminization. I confirmed that *W. pipientis* exploits RNA interference to induce modifications in its host's gene network, as previously observed for other bacteria (Sahr et al. 2022; Koeppen et al. 2016; Zhang et al. 2020), and other organisms who enter a thigh symbiosis with their hosts (Weiberg et al. 2013; Wang et al. 2017). The most expressed bacterial sRNAs originate from non-coding regions for which promoters of transcription were identified. The independent transcription of this region has been already assessed in different *Wolbachia* strains (Woolfit et al. 2015), where tissue specific expression analysis showed a significant upregulation in testis, compared to ovaries. The presence of this sRNA

in strains known to cause cytoplasmic incompatibility would suggest its effect is context-dependent, varying with the host's specific gene regulatory network. Alternatively, sRNAs originating from tRNAs were also predicted to target sex-related pathways, confirming the aptitude of tRFs to be transferred from symbionts (Cognat et al. 2017; Diallo et al. 2022) to the host cytoplasm. In agreement with this scenario, I found bacterial sRNAs to target genes not strictly related to sex related function, providing a mechanism for the *Wolbachia* induced effects on the host's immune program (Eleftherianos et al. 2013) and behaviour (Peng et al. 2008). All these effects, taken together, would contribute to stable vertical inheritance of the parasite.

In conclusion, the picture emerging from my thesis depicts mitochondrial small RNAs as an influential actor in the retrograde signalling. Furthermore, in systems where the regulatory landscape permits, they would have acquired functions capable of overriding the host's underlying genotypic programming. This substantial alteration is not a mitochondrial novelty, but rather a conserved feature of symbiotic interactions, dating back at the early origins of life. Thus, the layer of complexity that cross-kingdom RNAi represents has not only been a vital force in the evolution of the eukaryotic cell, but has continued to shape inter-organismal communication.

5.1 Future perspective

The work presented in this dissertation has resolved certain issues while simultaneously opening several new avenues of inquiry. This section describes the present area of exploration and outlines the most promising perspective and follow-up experiments.

A primary objective is to prove the functionality of smithRNAs: precise experimental validation will be indispensable for eliminating any doubts about their functionality inside cells. Previous studies tried to address this problem (Passamonti et al. 2020), but clear proof on target gene alteration is lacking. In an ongoing project I am examining the presence of smithRNAs in *Mytilus edulis*. This commercially valuable species would give the third proof of smithRNAs significance in the Bivalvia class and the first within the Mytilida order. Analyses on small RNA-seq libraries have already highlighted a high production of smithRNAs, especially in male type mitochondria. Interestingly, the most highly expressed smithRNA derives from the maturation of tRNA-Asn, confirming the importance of tRFs in mitonuclear interaction. Further steps would include the transfection of the male germline, growth on plate as cell cultures (Khorami et al. 2024), with synthetic RNAs designed to mimic or repress smithRNAs. A subsequent RNA sequencing would then reveal a precise map of genes affected by the treatment, supplying evidence that smithRNAs work efficiently as RNAi mediators. Moreover, distinctive methods have been used to identify which are the parameters that best predict the affinity of tRFs sequences for their cognate targets (Parikh et al. 2022). Applying this tool on the *M. edulis* dataset would determine its efficacy for smithRNAs target prediction and indicate whether the underlying machine learning model requires additional training. Enhancing precise target recognition would accelerate and increase interest in this RNAi mechanism, whose relevance has just been demonstrated, at least in Metazoa.

The findings of this thesis also suggest that a similar RNAi may be active in *Wolbachia*, a widespread endosymbiotic bacterium of insects (Weinert et al. 2015). Here, bacterial sRNAs appear able to alter several biological networks of its host, *Zyginidia pullula*. Simultaneous, tissue-specific sequencing of both RNA and small RNAs would clarify the expression patterns of the identified sRNAs within the animal and test if a correlation exists with the expression of the target they predict to interfere with. Furthermore, if the expression trend observed for ncA-originating sRNAs in *Drosophila melanogaster* infection (Woolfit et al. 2015) is maintained in this system, it would imply that the associated feminizing effects result from a specific architecture of the sex determination gene pathway. Additional

studies on other species where *Wolbachia* induce host feminization would elucidate whether the same RNAi mechanism is responsible for these phenotypic alterations. In addition, I am currently investigating the influence of *Wolbachia*-derived RNAi on host immunity. To do so, this work compares the bacterial-resistant red flour beetle, *Tribolium castaneum*, with its susceptible relative *Tribolium confusum* (Goodacre et al. 2015). *T. castaneum* is a premier model for functional genetics (Denell 2008), possessing a high quality genomic assembly and a comprehensive map of regulatory regions (Herndon et al. 2020), along with a well characterized immune system (Zou et al. 2007). Although the two species diverged approximately 66 Mya (Ramesh et al. 2021), regulatory elements would probably be at least partially shared in the *Tribolium* genus, given examples of similar target genes and microRNA expression between *D. melanogaster* and *T. castaneum* (Ninova et al. 2016; Herndon et al. 2020), whose orders split more than 360 Mya (Rainford et al. 2014). In particular, a comparison of potential *Wolbachia* sRNA target sites within genes involved in pathogen recognition (Zou et al. 2007) could help explain the differential vulnerability observed in *T. confusum* in respect to its resistant counterpart.

Pursuing these research directions could illuminate the mode of action, effects and implications of a mechanism that may support intracellular symbiosis in their multiple forms.

6. References

- Alhassan, A.R., Dzinyela, R., Dargiri, S.A., Suglo, P., Yang, L., Movahedi, A., 2025. Regulatory roles of small RNAs in plant growth, breeding, and stress adaptation. *Plant Growth Regul.* 105, 295–313.
- Allen, J.F., 2015. Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *Proc. Natl. Acad. Sci.* 112, 10231–10238.
- Altschul, Stephen F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Anaconda Inc, 2020. *Anaconda Software Distribution*, Available at: <https://www.anaconda.com>.
- Asgharian, H., Chang, P.L., Mazzoglio, P.J., Negri, I., 2014. Wolbachia is not all about sex: male-feminizing Wolbachia alters the leafhopper *Zyginidia pullula* transcriptome in a mainly sex-independent manner. *Front. Microbiol.* 5.
- Asha, S., Soniya, E.V., 2017. The sRNAome mining revealed existence of unique signature small RNAs derived from 5.8SrRNA from *Piper nigrum* and other plant lineages. *Sci. Rep.* 7, 41052.
- Baião, G.C., Schneider, D.I., Miller, W.J., Klasson, L., 2019. The effect of Wolbachia on gene expression in *Drosophila paulistorum* and its implications for symbiont-induced host speciation. *BMC Genomics* 20, 465.
- Bartel, D.P., 2018. Metazoan MicroRNAs. *Cell* 173, 20–51.
- Bensasson, D., Zhang, D.-X., Hartl, D.L., Hewitt, G.M., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol.* 16, 314–321.
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol., Mitogenomics and Metazoan Evolution* 69, 313–319.
- Bessho, Y., Ohama, T., Osawa, S., 1992. Planarian mitochondria I. Heterogeneity of cytochrome c oxidase subunit I gene sequences in the freshwater planarian, *Dugesia japonica*. *J. Mol. Evol.* 34, 324–330.

- Bessho, Y., Tamura, S., Hori, H., Tanaka, H., Ohama, T., Osawa, S., 1997. Planarian mitochondria sequence heterogeneity: relationships between the type of cytochrome c oxidase subunit I gene sequence, karyotype and genital organ. *Mol. Ecol.* 6, 129–136.
- Bewick, A.J., Vogel, K.J., Moore, A.J., Schmitz, R.J., 2017. Evolution of DNA Methylation across Insects. *Mol. Biol. Evol.* 34, 654–665.
- Bi, J., Zheng, Y., Wang, R.-F., Ai, H., Haynes, P.R., Brownlie, J.C., Yu, X.-Q., Wang, Y.-F., 2019. Wolbachia infection may improve learning and memory capacity of *Drosophila* by altering host gene expression through microRNA. *Insect Biochem. Mol. Biol.* 106, 47–54.
- Bishop, C., Hussain, M., Hugo, L.E., Asgari, S., 2022. Analysis of *Aedes aegypti* microRNAs in response to Wolbachia wAlbB infection and their potential role in mosquito longevity. *Sci. Rep.* 12, 15245.
- Biswas, K., Jolly, M.K., Ghosh, A., 2021. First passage time properties of miRNA-mediated protein translation. *J. Theor. Biol.* 529, 110863.
- Blighe, K, S Rana, M Lewis. 2018. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling.
- Blumenstiel, J.P., 2025. From the cauldron of conflict: Endogenous gene regulation by piRNA and other modes of adaptation enabled by selfish transposable elements. *Semin. Cell Dev. Biol.* 164, 1–12.
- Blumental-Perry, A., Jobava, R., Bederman, I., Degar, A.J., Kenche, H., Guan, B.J., Pandit, K., Perry, N.A., Molyneaux, N.D., Wu, J., Prendergas, E., 2020. Retrograde signaling by a mtDNA-encoded non-coding RNA preserves mitochondrial bioenergetics. *Communications Biology*, 3(1), p.626.
- Bofill-De Ros, X., Vang Ørom, U.A., 2024. Recent progress in miRNA biogenesis and decay. *RNA Biol.* 21, 36–43.

- Bofill-De Ros, X., Yang, A., Gu, S., 2020. IsomiRs: expanding the miRNA repression toolbox beyond the seed. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1863(4), p.194373.
- Boore, J.L., 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767–1780.
- Bouchon, D., Rigaud, T., Juchault, P., 1998. Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1401), 1081-1090.
- Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527.
- Breton, S., Milani, L., Ghiselli, F., Guerra, D., Stewart, D.T., Passamonti, M., 2014. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. *Trends Genet.* 30, 555–564.
- Brusca, R., Giribet, G., Moore, W., 2023. *Invertebrates*, 4th ed. Oxford University Press.
- Burstein, D., Amaro, F., Zusman, T., Lifshitz, Z., Cohen, O., Gilbert, J.A., Pupko, T., Shuman, H.A., Segal, G., 2016. Genomic analysis of 38 *Legionella* species identifies large and diverse effector repertoires. *Nat. Genet.* 48, 167–175.
- Burzio, V.A., Villota, C., Villegas, J., Landerer, E., Boccardo, E., Villa, L.L., Martínez, R., Lopez, C., Gaete, F., Toro, V., Rodriguez, X., Burzio, L.O., 2009. Expression of a family of noncoding mitochondrial RNAs distinguishes normal from cancer cells. *Proc. Natl. Acad. Sci.* 106, 9430–9434.
- Cai, Y.D., Chiu, J.C., 2022. Timeless in animal circadian clocks and beyond. *FEBS J.* 289, 6559–6575.
- Calcino, A.D., Fernandez-Valverde, S.L., Taft, R.J., Degnan, B.M., 2018. Diverse RNA interference strategies in early-branching metazoans. *BMC Evol. Biol.* 18, 160.
- Chen, L.-L., Kim, V.N., 2024. Small and long non-coding RNAs: Past, present, and future. *Cell* 187, 6451–6485.

- Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890.
- Chen, Y., Chen, L., Lun, A.T.L., Baldoni, P.L., Smyth, G.K., 2025. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res.* 53, gkaf018.
- Chen, Z., Sun, Y., Yang, X., Wu, Z., Guo, K., Niu, X., Wang, Q., Ruan, J., Bu, W., Gao, S., 2017. Two featured series of rRNA-derived RNA fragments (rRFs) constitute a novel class of small RNAs. *PLOS ONE* 12, e0176458.
- Cherlin, T., Magee, R., Jing, Y., Pliatsika, V., Loher, P., Rigoutsos, I., 2020. Ribosomal RNA fragmentation into short RNAs (rRFs) is modulated in a sex- and population of origin-specific manner. *BMC Biol.* 18, 38.
- Chevez-Guardado, R., Peña-Castillo, L., 2021. Promotech: a general tool for bacterial promoter recognition. *Genome Biol.* 22, 318.
- Christensen, S., Serbus, L.R., 2015. Comparative analysis of *Wolbachia* genomes reveals streamlining and divergence of minimalist two-component systems. *G3: Genes, Genomes, Genetics*, 5(5), 983-996.
- Clarke, A.W., Høye, E., Hembrom, A.A., Paynter, V.M., Vinther, J., Wyrożemski, Ł., Biryukova, I., Formaggioni, A., Ovchinnikov, V., Herlyn, H., Pierce, A., Wu, C., Aslanzadeh, M., Cheneby, J., Martinez, P., Friedländer, M.R., Hovig, E., Hackenberg, M., Umu, S.U., Johansen, M., Peterson, K.J., Fromm, B., 2025. MirGeneDB 3.0: improved taxonomic sampling, uniform nomenclature of novel conserved microRNA families and updated covariance models. *Nucleic Acids Res.* 53, D116–D128.
- Cobb, L.J., Lee, C., Xiao, J., Yen, K., Wong, R.G., Nakamura, H.K., Mehta, H.H., Gao, Q., Ashur, C., Huffman, D.M., Wan, J., Muzumdar, R., Barzilai, N., Cohen, P., 2016. Naturally occurring mitochondrial-derived peptides are age-dependent regulators of apoptosis, insulin sensitivity, and inflammatory markers. *Aging* 8, 796–808.

- Cognat, V., Morelle, G., Megel, C., Lalande, S., Molinier, J., Vincent, T., Small, I., Duchêne, A.-M., Maréchal-Drouard, L., 2017. The nuclear and organellar tRNA-derived RNA fragment population in *Arabidopsis thaliana* is highly dynamic. *Nucleic Acids Res.* 45, 3460–3472.
- Cucini, C., Boschi, S., Funari, R., Cardaioli, E., Iannotti, N., Marturano, G., Paoli, F., Bruttini, M., Carapelli, A., Frati, F., Nardi, F., 2024. De novo assembly and annotation of *Popillia japonica*'s genome with initial clues to its potential as an invasive pest. *BMC Genomics* 25, 275.
- Czech, B., Munafò, M., Ciabrelli, F., Eastwood, E.L., Fabry, M.H., Kneuss, E., Hannon, G.J., 2018. piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu. Rev. Genet.* 52, 131–157.
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008.
- Daugaard, I., Hansen, T.B., 2017. Biogenesis and Function of Ago-Associated RNAs. *Trends Genet.* 33, 208–219.
- Denell, R., 2008. Establishment of *Tribolium* as a Genetic Model System and Its Early Contributions to Evo-Devo. *Genetics* 180, 1779–1786 (2008).
- Dexheimer, P.J., Cochella, L., 2020. MicroRNAs: From Mechanism to Organism. *Front. Cell Dev. Biol.* 8.
- Diallo, I., Ho, J., Lalaouna, D., Masse, E., Provost, P., 2022. RNA sequencing unveils very small RNAs with potential regulatory functions in bacteria. *Front Mol Biosci* 9: 914991
- Dierckxsens, N., Mardulyn, P., Smits, G., 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45, e18.
- Dietrich, A., Wallet, C., Iqbal, R.K., Gualberto, J.M., Lotfi, F., 2015. Organellar non-coding RNAs: Emerging regulation mechanisms. *Biochimie, Special Issue: Regulatory RNAs* 117, 48–62.
- Dou, W., Miao, Y., Xiao, J., Huang, D., 2021. Association of *Wolbachia* with Gene Expression in *Drosophila* Testes. *Microb. Ecol.* 82, 805–817.

- Doucet-Beaupré, H., Breton, S., Chapman, E.G., Blier, P.U., Bogan, A.E., Stewart, D.T., Hoeh, W.R., 2010. Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. *BMC Evol. Biol.* 10, 50.
- D'Souza, A.R., Minczuk, M., 2018. Mitochondrial transcription and translation: overview. *Essays Biochem.* 62, 309–320.
- Dutta, T., Srivastava, S., 2018. Small RNA-mediated regulation in bacteria: A growing palette of diverse mechanisms. *Gene* 656, 60–72.
- Eleftherianos, I., Atri, J., Accetta, J., Castillo, J., 2013. Endosymbiotic bacteria in insects: guardians of the immune system? *Front. Physiol.* 4, (2013).
- Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238.
- Fan, S., Tian, T., Chen, W., Lv, X., Lei, X., Zhang, H., Sun, S., Cai, L., Pan, G., He, L., Ou, Z., Lin, X., Wang, X., Perez, M.F., Tu, Z., Ferrone, S., Tannous, B.A., Li, J., 2019. Mitochondrial miRNA Determines Chemoresistance by Reprogramming Metabolism and Regulating Mitochondrial Transcription. *Cancer Res.* 79, 1069–1084.
- Felden, B., Augagneur, Y., 2021. Diversity and Versatility in Small RNA-Mediated Regulation in Bacterial Pathogens. *Front. Microbiol.* 12.
- Ferree, P.M., Frydman, H.M., Li, J.M., Cao, J., Wieschaus, E., Sullivan, W., 2005. Wolbachia Utilizes Host Microtubules and Dynein for Anterior Localization in the *Drosophila* Oocyte. *PLOS Pathog.* 1, e14.
- Fontanesi, F., Tigano, M., Fu, Y., Sfeir, A., Barrientos, A., 2020. Human mitochondrial transcription and translation. in *The Human Mitochondrial Genome* (eds Gasparre, G. & Porcelli, A. M.) 35–70 (Academic Press, 2020).
- Formaggioni, A., Cavalli, G., Hamada, M., Sakamoto, T., Plazzi, F., Passamonti, M., 2024. The Evolution and Characterization of the RNA Interference Pathways in Lophotrochozoa. *Genome Biol. Evol.* 16, evae098.

- Formaggioni, A., Luchetti, A., Plazzi, F., 2021. Mitochondrial Genomic Landscape: A Portrait of the Mitochondrial Genome 40 Years after the First Complete Sequence. *Life* 11, 663.
- Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., Rajewsky, N., 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52.
- Ghildiyal, M., Zamore, P.D., 2009. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* 10, 94–108.
- Ghiselli, F., Gomes-dos-Santos, A., Adema, C.M., Lopes-Lima, M., Sharbrough, J., Boore, J.L., 2021a. Molluscan mitochondrial genomes break the rules. *Philos. Trans. R. Soc. B Biol. Sci.* 376, 20200159.
- Ghiselli, F., Iannello, M., 2015. A Transcriptome Annotation Pipeline for Non-model Organisms.
- Ghiselli, F., Iannello, M., Piccinini, G., Milani, L., 2021b. Bivalve Molluscs as Model Systems for Studying Mitochondrial Biology. *Integr. Comp. Biol.* 61, 1699–1714.
- Ghiselli, F., Milani, L., Chang, P.L., Hedgecock, D., Davis, J.P., Nuzhdin, S.V., Passamonti, M., 2012. De Novo Assembly of the Manila Clam *Ruditapes philippinarum* Transcriptome Provides New Insights into Expression Bias, Mitochondrial Doubly Uniparental Inheritance and Sex Determination. *Mol. Biol. Evol.* 29, 771–786.
- Ghiselli, F., Milani, L., Guerra, D., Chang, P.L., Breton, S., Nuzhdin, S.V., Passamonti, M., 2013. Structure, Transcription, and Variability of Metazoan Mitochondrial Genome: Perspectives from an Unusual Mitochondrial Inheritance System. *Genome Biol. Evol.* 5, 1535–1554.
- Giannakouros, T., Nikolakaki, E., Mylonis, I., Georgatsou, E., 2011. Serine-arginine protein kinases: a small protein kinase family with a large cellular presence. *FEBS J.* 278, 570–586.
- Goodacre, S. L., Fricke, C. & Martin, O. Y. , 2015. A screen for bacterial endosymbionts in the model organisms *Tribolium castaneum*, *T. confusum*, *Callosobruchus maculatus*, and related species. *Insect Science* 22, 165–177 (2015).

- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652.
- Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., Hofacker, I.L., 2008. The Vienna RNA Websuite. *Nucleic Acids Res.* 36, W70–W74.
- Ha, M., Kim, V.N., 2014. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* 15, 509–524.
- Hausser, J., Syed, A.P., Bilen, B., Zavolan, M., 2013. Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Res.* 23, 604.
- Herndon, N., Shelton, J., Gerischer, L., Ioannidis, P., Ninova, M., Dönitz, J., Waterhouse, R.M., Liang, C., Damm, C., Siemanowski, J., Kitzmann, P., Ulrich, J., Dippel, S., Oberhofer, G., Hu, Y., Schwirz, J., Schacht, M., Lehmann, S., Montino, A., Posnien, N., Gurska, D., Horn, T., Seibert, J., Vargas Jentzsch, I.M., Panfilio, K.A., Li, J., Wimmer, E.A., Stappert, D., Roth, S., Schröder, R., Park, Y., Schoppmeier, M., Chung, H.-R., Klingler, M., Kittelmann, S., Friedrich, M., Chen, R., Altincicek, B., Vilcinskas, A., Zdobnov, E., Griffiths-Jones, S., Ronshaugen, M., Stanke, M., Brown, S.J., Bucher, G., 2020. Enhanced genome assembly and a new official gene set for *Tribolium castaneum*. *BMC Genomics* 21, 47.
- Hirokawa, N., Noda, Y., Tanaka, Y., Niwa, S., 2009. Kinesin superfamily motor proteins and intracellular transport. *Nat. Rev. Mol. Cell Biol.* 10, 682–696.
- Hirose, Y., Ikeda, K.T., Noro, E., Hiraoka, K., Tomita, M., Kanai, A., 2015. Precise mapping and dynamics of tRNA-derived fragments (tRFs) in the development of *Triops cancriformis* (tadpole shrimp). *BMC Genet* 16, 83.
- Horard, B., Terretaz, K., Gosselin-Grenet, A.-S., Sobry, H., Sicard, M., Landmann, F., Loppin, B., 2022. Paternal transmission of the Wolbachia CidB toxin underlies cytoplasmic incompatibility. *Curr. Biol.* 32, 1319-1331.e5.

- Iki, T., Kawaguchi, S., Kai, T., 2023. miRNA/siRNA-directed pathway to produce noncoding piRNAs from endogenous protein-coding regions ensures *Drosophila* spermatogenesis. *Sci. Adv.* 9, eadh0397.
- Iwakawa, H., Tomari, Y., 2022. Life of RISC: Formation, action, and degradation of RNA-induced silencing complex. *Mol. Cell* 82, 30–43.
- Jang, I.-H., Chosa, N., Kim, S.-H., Nam, H.-J., Lemaitre, B., Ochiai, M., Kambris, Z., Brun, S., Hashimoto, C., Ashida, M., Brey, P.T., Lee, W.J., 2006. A Spätzle-processing enzyme required for toll signaling activation in *Drosophila* innate immunity. *Dev. Cell* 10, 45–55.
- Jiao, Y., Zheng, Z., Du, X., Wang, Q., Huang, R., Deng, Y., Shi, S., Zhao, X., 2014. Identification and Characterization of MicroRNAs in Pearl Oyster *Pinctada martensii* by Solexa Deep Sequencing. *Mar. Biotechnol.* 16, 54–62.
- Jin, X., Cheng, Z., Wang, B., Yau, T. on, Chen, Z., Barker, S.C., Chen, D., Bu, W., Sun, D., Gao, S., 2020. Precise annotation of human, chimpanzee, rhesus macaque and mouse mitochondrial genomes leads to insight into mitochondrial transcription in mammals. *RNA Biol.* 17, 395–402.
- John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., Marks, D.S., 2004. Human MicroRNA Targets. *PLOS Biol.* 2, e363.
- Johnston, I.G., Williams, B.P., 2016. Evolutionary Inference across Eukaryotes Identifies Specific Pressures Favoring Mitochondrial Gene Retention. *Cell Syst.* 2, 101–111.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240.
- Kawaji, H., Nakamura, M., Takahashi, Y., Sandelin, A., Katayama, S., Fukuda, S., Daub, C.O., Kai, C., Kawai, J., Yasuda, J., Carninci, P., Hayashizaki, Y., 2008. Hidden layers of human small RNAs. *BMC Genomics* 9, 157.

- Kelly, S., 2021. The economics of organellar gene loss and endosymbiotic gene transfer. *Genome Biol.* 22, 345.
- Kent, K., Nozawa, K., Jain, A., Malovannaya, A., Garcia, T.X., Matzuk, M.M., 2025. Ovocymase 2 is a key regulatory factor modulating proteolytic pathways and sperm maturation in the mammalian epididymis†. *Biol. Reprod.* 113, 127–140.
- Kern, P., Cook, J.M., Kageyama, D., Riegler, M., 2015. Double trouble: combined action of meiotic drive and Wolbachia feminization in *Eurema* butterflies. *Biol. Lett.* 11, 20150095.
- Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., Segal, E., 2007. The role of site accessibility in microRNA target recognition. *Nat. Genet.* 39, 1278–1284.
- Khorami, H.H., Breton, S., Angers, A., 2024. In vitro proliferation of *Mytilus edulis* male germ cell progenitors. *PLOS ONE* 19, e0292205.
- Kim, S.S., Lee, S.-J.V., 2019. Non-Coding RNAs in *Caenorhabditis elegans* Aging. *Mol. Cells* 42, 379–385.
- Kleine, T., Leister, D., 2016. Retrograde signaling: Organelles go networking. *Biochimica Et Biophysica Acta (BBA)-Bioenergetics*, 1857(8), 1313-1325.
- Koeppen, K., Hampton, T.H., Jarek, M., Scharfe, M., Gerber, S.A., Mielcarz, D.W., Demers, E.G., Dolben, E.L., Hammond, J.H., Hogan, D.A., Stanton, B.A., 2016. A Novel Mechanism of Host-Pathogen Interaction through sRNA in Bacterial Outer Membrane Vesicles. *PLOS Pathog.* 12, e1005672.
- Koonin, E.V., 2017. Evolution of RNA- and DNA-guided antiviral defense systems in prokaryotes and eukaryotes: common ancestry vs convergence. *Biol. Direct* 12, 5.
- Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28, 3211–3217.
- Krüger, J., Rehmsmeier, M., 2006. RNAhybrid: MicroRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* 34, W451–4.

- Kulkarni, N.H., Yamamoto, A.H., Robinson, K.O., Mackay, T.F.C., Anholt, R.R.H., 2002. The DSC1 Channel, Encoded by the smi60E Locus, Contributes to Odor-Guided Behavior in *Drosophila melanogaster*. *Genetics* 161, 1507–1516.
- Kumar, P., Anaya, J., Mudunuri, S.B., Dutta, A., 2014. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biol.* 12, 78.
- Kumarswamy, R., Bauters, C., Volkmann, I., Maury, F., Fetisch, J., Holzmann, A., Lemesle, G., de Groote, P., Pinet, F., Thum, T., 2014. Circulating Long Noncoding RNA, LIPCAR, Predicts Survival in Patients With Heart Failure. *Circ. Res.* 114, 1569–1575.
- Kusch, S., Singh, M., Thieron, H., Spanu, P.D., Panstruga, R., 2023. Site-specific analysis reveals candidate cross-kingdom small RNAs, tRNA and rRNA fragments, and signs of fungal RNA phasing in the barley–powdery mildew interaction. *Mol. Plant Pathol.* 24, 570–587.
- Kuscu, C., Kumar, P., Kiran, M., Su, Z., Malik, A., Dutta, A., 2018. tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA* 24, 1093–1105.
- Lalaouna, D., Simoneau-Roy, M., Lafontaine, D., Massé, E., 2013. Regulatory RNAs and target mRNA decay in prokaryotes. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1829(6-7), 742-747.
- Lambert, M., Benmoussa, A., Diallo, I., Ouellet-Boutin, K., Dorval, V., Majeau, N., Joly-Beauparlant, C., Droit, A., Bergeron, A., Têtu, B., Fradet, Y., Pouliot, F., Provost, P., 2021. Identification of Abundant and Functional dodecaRNAs (doRNAs) Derived from Ribosomal RNA. *Int. J. Mol. Sci.* 22, 9757.
- Lambert, M., Benmoussa, A., Provost, P., 2019. Small Non-Coding RNAs Derived from Eukaryotic Ribosomal RNA. *Non-Coding RNA* 5, 16.
- Lamoureux, C.R., Phaneuf, P.V., Palsson, B.O., Zielinski, D.C., 2024. *Escherichia coli* non-coding regulatory regions are highly conserved. *NAR Genomics and Bioinformatics*, 6(2), p.lqae041.

- Lane, N., 2018. *Power, Sex, Suicide: Mitochondria and the meaning of life*. Oxford University Press.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Larriba, E., Del Mazo, J., 2016. Role of Non-Coding RNAs in the Transgenerational Epigenetic Transmission of the Effects of Reprotoxicants. *Int. J. Mol. Sci.* 17, 452.
- Lasa, I., Toledo-Arana, A., Dobin, A., Villanueva, M., de los Mozos, I.R., Vergara-Irigaray, M., Segura, V., Fagegaltier, D., Penadés, J.R., Valle, J., Solano, C., Gingeras, T.R., 2011. Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc. Natl. Acad. Sci.* 108, 20172–20177.
- Lax, C., Tahiri, G., Patiño-Medina, J.A., Cánovas-Márquez, J.T., Pérez-Ruiz, J.A., Osorio-Concepción, M., Navarro, E., Calo, S., 2020. The Evolutionary Significance of RNAi in the Fungal Kingdom. *Int. J. Mol. Sci.* 21, 9348.
- Lee, C., Yen, K., Cohen, P., 2013. Humanin: a harbinger of mitochondrial-derived peptides? *Trends Endocrinol. Metab.* 24, 222–228.
- Lewis, B.P., Shih, I. -hung, Jones-Rhoades, M.W., Bartel, D.P., Burge, C.B., 2003. Prediction of Mammalian MicroRNA Targets. *Cell* 115, 787–798.
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, M., Zhao, H., Zhang, X., Wood, L.D., Anders, R.A., Choti, M.A., Pawlik, T.M., Daniel, H.D., Kannangai, R., Offerhaus, G.J.A., Velculescu, V.E., Wang, L., Zhou, S., Vogelstein, B., Hruban, R.H., Papadopoulos, N., Cai, J., Torbenson, M.S., Kinzler, K.W., 2011. Inactivating

- mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. *Nat. Genet.* 43, 828–829.
- Li, P., Jiao, J., Gao, G., Prabhakar, B.S., 2012. Control of mitochondrial activity by miRNAs. *J. Cell. Biochem.* 113, 1104–1110.
- Li, Y., Hui, J.H.L., 2023. Small RNAs in Cnidaria: A review. *Evol. Appl.* 16, 354–364.
- Lord, S.J., Rajotte, R.V., Korbitt, G.S., Bleackley, R.C., 2003. Granzyme B: a natural born killer. *Immunol. Rev.* 193, 31–38.
- Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L., 2011. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6, 26.
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Maistrenko, O.M., Serga, S.V., Vaiserman, A.M., Kozeretska, I.A., 2016. Longevity-modulating effects of symbiosis: insights from *Drosophila*–*Wolbachia* interaction. *Biogerontology* 17, 785–803.
- Malabirade, A., Morgado-Brajones, J., Trépout, S., Wien, F., Marquez, I., Seguin, J., Marco, S., Velez, M., Arluison, V., 2017. Membrane association of the bacterial riboregulator Hfq and functional perspectives. *Sci. Rep.* 7, 10724.
- Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M., 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 38, 4647–4654.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12.
- Marturano, G., Carli, D., Cucini, C., Carapelli, A., Plazzi, F., Frati, F., Passamonti, M., Nardi, F., 2024. SmithHunter: a workflow for the identification of candidate smithRNAs and their targets. *BMC Bioinformatics* 25, 286.
- Mattick, J.S., Makunin, I.V., 2006. Non-coding RNA. *Hum. Mol. Genet.* 15, R17–R29.

- Mayoral, J.G., Hussain, M., Joubert, D.A., Iturbe-Ormaetxe, I., O'Neill, S.L., Asgari, S., 2014. Wolbachia small noncoding RNAs and their role in cross-kingdom communications. *Proc. Natl. Acad. Sci.* 111, 18721–18726.
- McCutcheon, J.P., Moran, N.A., 2012. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26.
- Melamed, S., Adams, P.P., Zhang, A., Zhang, H., Storz, G., 2020. RNA-RNA Interactomes of ProQ and Hfq Reveal Overlapping and Competing Roles. *Mol. Cell* 77, 411-425.e7.
- Mercer, T.R., Neph, S., Dinger, M.E., Crawford, J., Smith, M.A., Shearwood, A.-M.J., Haugen, E., Bracken, C.P., Rackham, O., Stamatoyannopoulos, J.A., Filipovska, A., Mattick, J.S., 2011. The Human Mitochondrial Transcriptome. *Cell* 146, 645–658.
- Meseguer, S., 2021. MicroRNAs and tRNA-Derived Small Fragments: Key Messengers in Nuclear–Mitochondrial Communication. *Front. Mol. Biosci.* 8.
- Meseguer, S., Navarro-González, C., Panadero, J., Villarroja, M., Boutoual, R., Sánchez-Alcázar, J.A., Armengod, M.-E., 2019. The MELAS mutation m.3243A>G alters the expression of mitochondrial tRNA fragments. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* 1866, 1433–1449.
- Milani, L., Ghiselli, F., Passamonti, M., 2016. Mitochondrial selfish elements and the evolution of biological novelties. *Curr. Zool.* 62, 687–697.
- Mitchell, A., Guerra, D., Stewart, D., Breton, S., 2016. In silico analyses of mitochondrial ORFans in freshwater mussels (Bivalvia: Unionoida) provide a framework for future studies of their origin and function. *BMC Genomics* 17, 597.
- Moran, Y., Agron, M., Praher, D., Technau, U., 2017. The evolutionary origin of plant and animal microRNAs. *Nat. Ecol. Evol.* 1, 0027.
- Moriano-Gutierrez, S., Bongrand, C., Essock-Burns, T., Wu, L., McFall-Ngai, M.J., Ruby, E.G., 2020. The noncoding small RNA SsrA is released by *Vibrio fischeri* and modulates critical host responses. *PLOS Biol.* 18, e3000934.

- Muneretto, G., Plazzi F., e Passamonti M., 2024. Mitochondrion-to-nucleus communication mediated by RNA export: a survey of potential mechanisms and players across eukaryotes. *Biology Letters* 20 (7): 20240147. <https://doi.org/10.1098/rsbl.2024.0147>.
- Negri, I., Pellecchia, M., Mazzoglio, P. j, Patetta, A., Alma, A., 2006. Feminizing Wolbachia in Zyginidia pullula (Insecta, Hemiptera), a leafhopper with an XX/X0 sex-determination system. *Proc. R. Soc. B Biol. Sci.* 273, 2409–2416.
- Newton, I.L.G., Savytskyy, O., Sheehan, K.B., 2015. Wolbachia Utilize Host Actin for Efficient Maternal Transmission in *Drosophila melanogaster*. *PLOS Pathog.* 11, e1004798.
- Niaison, T., Guerra, D., Breton, S., 2021. The complete mitogenome of the inarticulate brachiopod *Glottidia pyramidata* reveals insights into gene order variation, deviant ATP8 and mtORFans in the Brachiopoda. *Mitochondrial DNA Part B* 6, 2701–2703.
- Ninova, M., Ronshaugen, M., Griffiths-Jones, S., 2016. MicroRNA evolution, expression, and function during short germband development in *Tribolium castaneum*. *Genome Res.* 26, 85–96.
- Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., Lai, E.C., 2007. The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in *Drosophila*. *Cell* 130, 89–100.
- Onishi, R., Yamanaka, S., Siomi, M.C., 2021. piRNA- and siRNA-mediated transcriptional repression in *Drosophila*, mice, and yeast: new insights and biodiversity. *EMBO Rep.* 22, e53062.
- Osei-Amo, S., Hussain, M., O'Neill, S.L., Asgari, S., 2012. Wolbachia-Induced aae-miR-12 miRNA Negatively Regulates the Expression of MCT1 and MCM6 Genes in Wolbachia-Infected Mosquito Cell Line. *PLOS ONE* 7, e50049.
- Ote, M., Ueyama, M., Yamamoto, D., 2016. Wolbachia Protein TomO Targets nanos mRNA and Restores Germ Stem Cells in *Drosophila* Sex-lethal Mutants. *Curr. Biol.* 26, 2223–2232.
- Paramasivam, A., Vijayashree Priyadharsini, J., 2020. MitomiRs: new emerging microRNAs in mitochondrial dysfunction and cardiovascular disease. *Hypertens. Res.* 43, 851–853.

- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055.
- Parikh, R., Wilson, B., Marrah, L., Su, Z., Saha, S., Kumar, P., Huang, F., Dutta, A., 2022. tRForest: a novel random forest-based algorithm for tRNA-derived fragment target prediction. *NAR Genom Bioinform* 4, lqac037.
- Passamonti, M., Calderone, M., Delpero, M., Plazzi, F., 2020. Clues of in vivo nuclear gene regulation by mitochondrial short non-coding RNAs. *Sci. Rep.* 10, 8219.
- Passamonti, M., Plazzi, F., 2020. Doubly Uniparental Inheritance and beyond: The contribution of the Manila clam *Ruditapes philippinarum*. *J. Zool. Syst. Evol. Res.* 58, 529–540.
- Passamonti, M., Scali, V., 2001. Gender-associated mitochondrial DNA heteroplasmy in the venerid clam *Tapes philippinarum* (Mollusca Bivalvia). *Curr. Genet.* 39, 117–124.
- Peng, Y., Nielsen, J.E., Cunningham, J.P., McGraw, E.A., 2008. Wolbachia Infection Alters Olfactory-Cued Locomotion in *Drosophila* spp. *Appl. Environ. Microbiol.* 74, 3943–3948.
- Perlman, S.J., Hodson, C.N., Hamilton, P.T., Opit, G.P., Gowen, B.E., 2015. Maternal transmission, sex ratio distortion, and mitochondria. *Proceedings of the National Academy of Sciences* 112, 10162–10168.
- Pimentel, A.C., Cesar, C.S., Martins, A.H.B., Martins, M., Cogni, R., 2025. Wolbachia Offers Protection Against Two Common Natural Viruses of *Drosophila*. *Microb. Ecol.* 88, 24.
- Plazzi, F., Le Cras, Y., Formaggioni, A., Passamonti, M., 2024. Mitochondrially mediated RNA interference, a retrograde signaling system affecting nuclear gene expression. *Heredity* 132, 156–161.
- Plazzi, F., Puccio, G., Passamonti, M., 2016. Comparative Large-Scale Mitogenomics Evidences Clade-Specific Evolutionary Trends in Mitochondrial DNAs of Bivalvia. *Genome Biol. Evol.* 8, 2544–2564.

- Ponath, F., Hör, J., Vogel, J., 2022. An overview of gene regulation in bacteria by small RNAs derived from mRNA 3' ends. *FEMS Microbiol. Rev.* 46, fuac017.
- Pozzi, A., Dowling, D.K., 2022. New Insights into Mitochondrial–Nuclear Interactions Revealed through Analysis of Small RNAs. *Genome Biol. Evol.* 14, evac023.
- Pozzi, A., Dowling, D.K., 2019. The Genomic Origins of Small Mitochondrial RNAs: Are They Transcribed by the Mitochondrial DNA or by Mitochondrial Pseudogenes within the Nucleus (NUMTs)? *Genome Biol. Evol.* 11, 1883–1896.
- Pozzi, A., Plazzi, F., Milani, L., Ghiselli, F., Passamonti, M., 2017. SmithRNAs: Could Mitochondria “Bend” Nuclear Regulation? *Mol. Biol. Evol.* 34, 1960–1973.
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- R Core Team, 2025. R: A Language and Environment for Statistical Computing.
- Rainford, J.L., Hofreiter, M., Nicholson, D.B., Mayhew, P.J., 2014. Phylogenetic Distribution of Extant Richness Suggests Metamorphosis Is a Key Innovation Driving Diversification in Insects. *PLOS ONE* 9, e109085.
- Ramesh, B., Firreno, T.J., Demuth, J.P., 2021. Divergence time estimation of genus *Tribolium* by extensive sampling of highly conserved orthologs. *Molecular Phylogenetics and Evolution* 159, 107084.
- Ranallo-Benavidez, T.R., Jaron, K.S., Schatz, M.C., 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* 11, 1432.
- Rangel-Chávez, C.P., Galán-Vásquez, E., Pescador-Tapia, A., Delaye, L., Martínez-Antonio, A., 2021. RNA polymerases in strict endosymbiont bacteria with extreme genome reduction show distinct erosions that might result in limited and differential promoter recognition. *PLOS ONE* 16, e0239350.
- Ratnasingham, S., Hebert, P., 2007. bold: The Barcode of Life Data System.

- Ren, L., Men, Y.-J., Bing, X.-L., Hoffmann, A., Hong, X.-Y., 2025. Transovarial transmission of *Wolbachia* bacteria via P44/Msp2-IMP2 mediated endocytosis. *Insect Science*.
- Rice, D.W., Sheehan, K.B., Newton, I.L.G., 2017. Large-Scale Identification of *Wolbachia pipientis* Effectors. *Genome Biol. Evol.* 9, 1925–1937.
- Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* 16, 276–277.
- Riggs, C.L., Summers, A., Warren, D.E., Nilsson, G.E., Lefevre, S., Dowd, W.W., Milton, S., Podrabsky, J.E., 2018. Small Non-coding RNA Expression and Vertebrate Anoxia Tolerance. *Front. Genet.* 9.
- Riggs, C.L., Woll, S.C., Podrabsky, J.E., 2019. MitosRNAs and extreme anoxia tolerance in embryos of the annual killifish *Austrofundulus limnaeus*. *Sci. Rep.* 9, 19812.
- Riparbelli, M.G., Giordano, R., Ueyama, M., Callaini, G., 2012. *Wolbachia*-Mediated Male Killing Is Associated with Defective Chromatin Remodeling. *PLOS ONE* 7, e30045.
- Ro, S., Ma, H.-Y., Park, C., Ortogero, N., Song, R., Hennig, G.W., Zheng, H., Lin, Y.-M., Moro, L., Hsieh, J.-T., Yan, W., 2013. The mitochondrial genome encodes abundant small noncoding RNAs. *Cell Res.* 23, 759–774.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584.
- Rueda, A., Barturen, G., Lebrón, R., Gómez-Martín, C., Alganza, Á., Oliver, J.L., Hackenberg, M., 2015. sRNAtoolbox: an integrated collection of small RNA research tools. *Nucleic Acids Res.* 43, W467–W473.
- Sahr, T., Escoll, P., Rusniok, C., Bui, S., Pehau-Arnaudet, G., Lavieu, G., Buchrieser, C., 2022. Translocated *Legionella pneumophila* small RNAs mimic eukaryotic microRNAs targeting the host immune response. *Nat. Commun.* 13, 762.
- Schmidt, O., Pfanner, N., Meisinger, C., 2010. Mitochondrial protein import: from proteomics to functional mechanisms. *Nat. Rev. Mol. Cell Biol.* 11, 655–667.

- Scholz, M., Albanese, D., Tuohy, K., Donati, C., Segata, N., Rota-Stabelli, O., 2020. Large scale genome reconstructions illuminate Wolbachia evolution. *Nat. Commun.* 11, 5235.
- Selifanova, M., Demianchenko, O., Noskova, E., Pitikov, E., Skvortsov, D., Drozd, J., Vatulkina, N., Apel, P., Kolodyazhnaya, E., Ezhova, M.A., Tzetlin, A.B., Neretina, T.V., Knorre, D.A., 2023. ORFans in Mitochondrial Genomes of Marine Polychaete *Polydora*. *Genome Biol. Evol.* 15, evad219.
- Sha, Q.-Q., Dai, X.-X., Dang, Y., Tang, F., Liu, J., Zhang, Y.-L., Fan, H.-Y., 2017. A MAPK cascade couples maternal mRNA translation and degradation to meiotic cell cycle progression in mouse oocytes. *Development* 144, 452–463.
- Shabalina, S.A., Koonin, E.V., 2008. Origins and evolution of eukaryotic RNA interference. *Trends Ecol. Evol.* 23, 578–587.
- Shang, R., Lee, S., Senavirathne, G., Lai, E.C., 2023. microRNAs in action: biogenesis, function and regulation. *Nat. Rev. Genet.* 24, 816–833.
- Shaukat, A.-N., Kaliatsi, E.G., Stamatopoulou, V., Stathopoulos, C., 2021a. Mitochondrial tRNA-Derived Fragments and Their Contribution to Gene Expression Regulation. *Front. Physiol.* 12.
- Sheehan, K.B., Martin, M., Lesser, C.F., Isberg, R.R., Newton, I.L.G., 2016. Identification and Characterization of a Candidate *Wolbachia pipientis* Type IV Effector That Interacts with the Actin Cytoskeleton. *mBio* 7, 10.1128/mbio.00622-16.
- Shin, C., Nam, J.-W., Farh, K.K.-H., Chiang, H.R., Shkumatava, A., Bartel, D.P., 2010. Expanding the MicroRNA Targeting Code: Functional Sites with Centered Pairing. *Mol. Cell* 38, 789–802.
- Shivam, P., Ball, D., Cooley, A., Osi, I., Rayford, K.J., Gonzalez, S.B., Edwards, A.D., McIntosh, A.R., Devaughn, J., Pugh-Brown, J.P. and Misra, S., 2025. Regulatory roles of PIWI-interacting RNAs in cardiovascular disease. *American Journal of Physiology-Heart and Circulatory Physiology*, 328(4), H991-H1004.

- Singh, P.P., Li, L., Schorey, J.S., 2015. Exosomal RNA from *Mycobacterium tuberculosis*-Infected Cells Is Functional in Recipient Macrophages. *Traffic* 16, 555–571.
- Sitnik, J.L., Francis, C., Hens, K., Huybrechts, R., Wolfner, M.F., Callaerts, P., 2014. Neprilysins: An Evolutionarily Conserved Family of Metalloproteases That Play Important Roles in Reproduction in *Drosophila*. *Genetics* 196, 781–797.
- Smith, C.H., Mejia-Trujillo, R., Breton, S., Pinto, B.J., Kirkpatrick, M., Havird, J.C., 2023. Mitonuclear Sex Determination? Empirical Evidence from Bivalves. *Mol. Biol. Evol.* 40, msad240.
- Smith, D.R., Keeling, P.J., 2015. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *PNAS*, 112(33), 0177-10184.
- Sousa, G.L., Bishnoi, R., Baxter, R.H.G., Povelones, M., 2020. The CLIP-domain serine protease CLIPC9 regulates melanization downstream of SPCLIP1, CLIPA8, and CLIPA28 in the malaria vector *Anopheles gambiae*. *PLOS Pathog.* 16, e1008985.
- Souvorov, A., Agarwala, R., Lipman, D.J., 2018. SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biol.* 19, 153.
- Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., Ettema, T.J.G., 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521, 173–179.
- Su, Y., Wu, J., Chen, W., Shan, J., Chen, D., Zhu, G., Ge, S., Liu, Y., 2024. Spliceosomal snRNAs, the Essential Players in pre-mRNA Processing in Eukaryotic Nucleus: From Biogenesis to Functions and Spatiotemporal Characteristics. *Adv. Biol.* 8, 2400006.
- Sullivan, W., Porter, J., 2023. The cellular lives of *Wolbachia*. *Nat. Rev. Microbiol.* 21.
- Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* 6, e21800.
- Swarts, D.C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R.F., Koonin, E.V., Patel, D.J., van der Oost, J., 2014. The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* 21, 743–753.

- Szczepanek, J., Pareek, C.S., Tretyn, A., 2018. The role of microRNAs in animal physiology and pathology. *Transl. Res. Vet. Sci.* 1, 13–33.
- Tan, G.C., Chan, E., Molnar, A., Sarkar, R., Alexieva, D., Isa, I.M., Robinson, S., Zhang, S., Ellis, P., Langford, C.F., Guillot, P.V., Chandrashekan, A., Fisk, N.M., Castellano, L., Meister, G., Winston, R.M., Cui, W., Baulcombe, D., Dibb, N.J., 2014. 5' isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res.* 42, 9424–9435.
- Terretaz, K., Horard, B., Weill, M., Loppin, B., Landmann, F., 2023. Functional analysis of *Wolbachia* Cid effectors unravels cooperative interactions to target host chromatin during replication. *PLOS Pathog.* 19, e1011211.
- Thompson, D.M., Lu, C., Green, P.J., Parker, R., 2008. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* 14, 2095–2103.
- Timmis, J.N., Ayliffe, M.A., Huang, C.Y., Martin, W., 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* 5, 123–135.
- Tomczynska, I., Stumpe, M., Mauch, F., 2018. A conserved RxLR effector interacts with host RABA-type GTPases to inhibit vesicle-mediated secretion of antimicrobial proteins. *Plant J.* 95, 187–203.
- Torri, A., Jaeger, J., Pradeu, T., Saleh, M.-C., 2022. The origin of RNA interference: Adaptive or neutral evolution? *PLOS Biol.* 20, e3001715.
- Tsukita, Sachiko, Yonemura, S., Tsukita, Shoichiro, 1997. ERM (ezrin/radixin/moesin) family: from cytoskeleton to signal transduction. *Curr. Opin. Cell Biol.* 9, 70–75.
- Umu, S.U., Paynter, V.M., Trondsen, H., Buschmann, T., Rounge, T.B., Peterson, K.J., Fromm, B., 2023. Accurate microRNA annotation of animal genomes using trained covariance models of curated microRNA complements in MirMachine. *Cell Genomics* 3.
- Vandana, V., Dong, S., Sheth, T., Sun, Q., Wen, H., Maldonado, A., Xi, Z., Dimopoulos, G., 2024. *Wolbachia* infection-responsive immune genes suppress *Plasmodium falciparum* infection in *Anopheles stephensi*. *PLOS Pathog.* 20, e1012145.

- Villegas, J., Burzio, V., Villota, C., Landerer, E., Martinez, Ronny, Santander, M., Martinez, Rodrigo, Pinto, R., Vera, M.I., Boccardo, E., Villa, L.L., Burzio, L.O., 2007. Expression of a novel non-coding mitochondrial RNA in human proliferating cells. *Nucleic Acids Res.* 35, 7336–7347.
- Wang, B., Sun, Y., Song, N., Zhao, M., Liu, R., Feng, H., Wang, X., Kang, Z., 2017. *Puccinia striiformis* f. sp. *tritici* microRNA-like RNA 1 (Pst-milR1), an important pathogenicity factor of Pst, impairs wheat resistance to Pst by suppressing the wheat pathogenesis-related 2 gene. *New Phytologist* 215, 338–350.
- Wang, M., Jiang, S., Wu, W., Yu, F., Chang, W., Li, P., Wang, K., 2018. Non-coding RNAs Function as Immune Regulators in Teleost Fish. *Front. Immunol.* 9.
- Wang, S., Liang, H., Xu, Y., Li, L., Wang, H., Sahu, D.N., Petersen, M., Melkonian, M., Sahu, S.K., Liu, H., 2021. Genome-wide analyses across Viridiplantae reveal the origin and diversification of small RNA pathway-related genes. *Commun. Biol.* 4, 412.
- Wang, Z., Wu, M., 2014. Phylogenomic Reconstruction Indicates Mitochondrial Ancestor Was an Energy Parasite. *PLOS ONE* 9, e110685.
- Warecki, B., Vega, G., Fowler, S., Hartzog, G., Karr, T.L., Sullivan, W., 2025. Wolbachia-mediated reduction in the glutamate receptor mGluR promotes female promiscuity and bacterial spread. *Cell Rep.* 44.
- Weiberg, A., Wang, M., Lin, F.-M., Zhao, H., Zhang, Z., Kaloshian, I., Huang, H.-D., Jin, H., 2013. Fungal Small RNAs Suppress Plant Immunity by Hijacking Host RNA Interference Pathways. *Science* 342, 118–123.
- Weinert, L.A., Araujo-Jnr, E.V., Ahmed, M.Z., Welch, J.J., 2015. The incidence of bacterial endosymbionts in terrestrial arthropods. *Proc. R. Soc. B Biol. Sci.* 282, 20150249.
- Werren, J.H., Baldo, L., Clark, M.E., 2008. Wolbachia: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* 6, 741–751.

- Westholm, J.O., Lai, E.C., 2011. Mirtrons: microRNA biogenesis via splicing. *Biochimie*, 93(11), 1897-1904.
- Wilmes, M., Meier, K., Schiefer, A., Josten, M., Otten, C.F., Klöckner, A., Henrichfreise, B., Vollmer, W., Hoerauf, A., Pfarr, K., 2017. AmiD Is a Novel Peptidoglycan Amidase in Wolbachia Endosymbionts of *Drosophila melanogaster*. *Front. Cell. Infect. Microbiol.* 7.
- Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 257.
- Woolfit, M., Algama, M., Keith, J.M., McGraw, E.A., Popovici, J., 2015. Discovery of Putative Small Non-Coding RNAs from the Obligate Intracellular Bacterium *Wolbachia pipientis*. *PLOS ONE* 10, e0118595.
- WoRMS Editorial Board (2025). World Register of Marine Species. Available from <https://www.marinespecies.org> at VLIZ.
- Wu, S., Manber, U., 1992. AGREP - A FAST APPROXIMATE PATTERN-MATCHING TOOL.
- Xu, X., Ji, H., Jin, X., Cheng, Z., Yao, X., Liu, Y., Zhao, Q., Zhang, T., Ruan, J., Bu, W., Chen, Z., Gao, S., 2019. Using Pan RNA-Seq Analysis to Reveal the Ubiquitous Existence of 5' and 3' End Small RNAs. *Front. Genet.* 10.
- Yang, Q., Li, L., Xue, Z., Ye, Q., Zhang, L., Li, S., Liu, Y., 2013. Transcription of the Major *Neurospora crassa* microRNA-Like Small RNAs Relies on RNA Polymerase III. *PLOS Genet.* 9, e1003227.
- Yuan, L.-L., Chen, X., Zong, Q., Zhao, T., Wang, J.-L., Zheng, Y., Zhang, M., Wang, Z., Brownlie, J.C., Yang, F., Wang, Y.-F., 2015. Quantitative Proteomic Analyses of Molecular Mechanisms Associated with Cytoplasmic Incompatibility in *Drosophila melanogaster* Induced by *Wolbachia*. *J. Proteome Res.* 14, 3835–3847.
- Zapletal, D., Taborska, E., Pasulka, J., Malik, R., Kubicek, K., Zanova, M., Much, C., Sebesta, M., Buccheri, V., Horvat, F., Jenickova, I., Prochazkova, M., Prochazka, J., Pinkas, M., Novacek, J., Joseph, D.F., Sedlacek, R., Bernecky, C., O'Carroll, D., Stefl, R., Svoboda, P., 2022.

- Structural and functional basis of mammalian microRNA biogenesis by Dicer. *Molecular Cell* 82, 4064–4079.e13.
- Zeng, W., Baumann, C., Schmidtman, A., Honaramooz, A., Tang, L., Bondareva, A., Dores, C., Fan, T., Xi, S., Geiman, T., Rathi, R., de Rooij, D., De La Fuente, R., Muegge, K., Dobrinski, I., 2011. Lymphoid-Specific Helicase (HELLS) Is Essential for Meiotic Progression in Mouse Spermatocytes¹. *Biol. Reprod.* 84, 1235–1241.
- Zhang, D., Wang, Y., He, K., Yang, Q., Gong, M., Ji, M., Chen, L., 2020. Wolbachia limits pathogen infections through induction of host innate immune responses. *PLOS ONE* 15, e0226736.
- Zhang, G., Hussain, M., O'Neill, S.L., Asgari, S., 2013. Wolbachia uses a host microRNA to regulate transcripts of a methyltransferase, contributing to dengue virus inhibition in *Aedes aegypti*. *Proc. Natl. Acad. Sci.* 110, 10276–10281.
- Zhang, H., Zhang, Y., Song, Z., Li, R., Ruan, H., Liu, Q., Huang, X., 2020. sncRNAs packaged by *Helicobacter pylori* outer membrane vesicles attenuate IL-8 secretion in human cells. *Int. J. Med. Microbiol.* 310, 151356.
- Zhang, S., Sun, L., Kragler, F., 2009. The Phloem-Delivered RNA Pool Contains Small Noncoding RNAs and Interferes with Translation. *Plant Physiol.* 150, 378–387.
- Zheng, Y., Shen, W., Bi, J., Chen, M.-Y., Wang, R.-F., Ai, H., Wang, Y.-F., 2019. Small RNA analysis provides new insights into cytoplasmic incompatibility in *Drosophila melanogaster* induced by Wolbachia. *J. Insect Physiol.* 118, 103938.
- Zou, Z., Evans, J.D., Lu, Z., Zhao, P., Williams, M., Sumathipala, N., Hetru, C., Hultmark, D., Jiang, H., 2007. Comparative genomic analysis of the *Tribolium* immune system. *Genome Biol* 8, R177.
- Zouros, E., 2013. Biparental Inheritance Through Uniparental Transmission: The Doubly Uniparental Inheritance (DUI) of Mitochondrial DNA. *Evol. Biol.* 40, 1–31.
- Zouros, E., Rodakis, G.C., 2019. Doubly uniparental inheritance of mtDNA: An unappreciated defiance of a general rule. In *Cellular and Molecular Basis of Mitochondrial Inheritance: Mitochondrial Disease and Fitness* (25-49). Cham: Springer International Publishing.