GBE

SmithRNAs: A Common Feature among Metazoa

Giovanni Marturano [1,†], Diego Carli [2,†], Claudio Cucini [1,3], Elena Cardaioli [1], Antonio Carapelli [1,3], Federico Plazzi [2], Francesco Frati [1,3], Marco Passamonti [2,*], Francesco Nardi [1,3]

[1]Department of Life Sciences, University of Siena, Siena 53100, Italy

[2]Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna 40126, Italy

[3]National Biodiversity Future Center (NBFC), Palermo 90133, Italy

[†]Equal contribution.

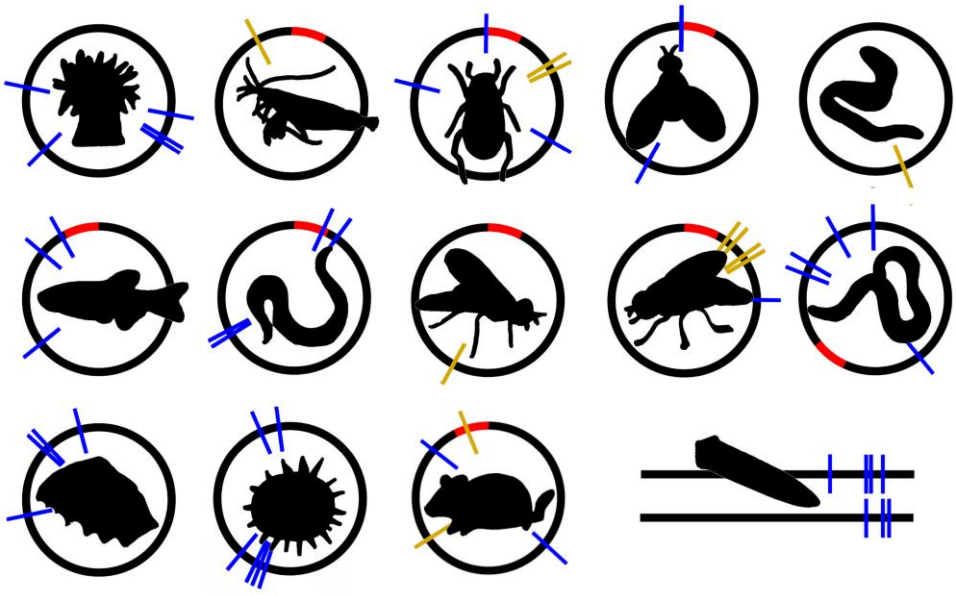
*Corresponding author: Marco Passamonti, University of Bologna, Bologna, Italy. E-mail: marco.passamonti@unibo.it.

Accepted: October 30, 2025

Abstract

SmithRNAs are a novel class of small noncoding RNAs that are encoded in the mitochondrial genome and regulate the expression of nuclear transcripts. They have been recently described in the Manila clam *Ruditapes philippinarum* and their biological function has been confirmed in vivo. It is currently unclear whether smithRNAs are a unique feature of this species, possibly related to the peculiar mechanism of sex determination observed in bivalves, or a common feature of Metazoa. Aiming at a broader survey on the presence and biological features of smithRNAs across Metazoa, 14 species were selected to represent major lineages, and for each species small RNAseq data, as well as the transcriptome, mitochondrial and nuclear genomes, were collected from the literature or sequenced/assembled de novo. Data were analyzed using the SmithHunter pipeline, a recently published tool specifically designed to identify smithRNAs and their targets. Candidate smithRNAs were identified in all species studied, supporting the notion that smithRNAs are a common feature across Metazoa. SmithRNAs are generally encoded within other genes, on the same strand, and with a preference for mitochondrial rRNAs and tRNAs. Based on their strandedness and preferential position at the 5′-end of the encompassing gene, a transcription mechanism is proposed where smithRNAs are cleaved off from gene-specific transcripts after the maturation of the two primary mitochondrial transcripts. A substantial variability was identified concerning the possible nuclear targets of smithRNAs, with a preference for regulation/response terms, mitochondrial functions, and sex/germline associated terms.

Graphical abstract



Key words: microRNAs, small RNAs, sncRNAs, mitochondrial genome, regulation, mito-nuclear interactions, mitochondrial retrograde response.

Significance

Multiple classes of small RNAs, including miRNAs, piRNAs, and siRNAs, are known to be implicated in the regulation of gene expression. A novel class, dubbed smithRNAs, was recently described in the Manila clam *Ruditapes philippinarum*. Unique to this class is that they are encoded in the mitochondrial genome but regulate the expression of nuclear transcripts. It is currently unknown whether these small RNAs are to be considered a unique feature of some odd taxa or a common feature of Metazoa. Here we present a broad survey for the presence of smithRNAs across Metazoa, based on the analysis of small RNAseq data, transcriptome, mitochondrial, and nuclear genome of 14 animal species, representative of all major metazoan taxa, using the recently published SmithHunter pipeline. Candidate smithRNAs were identified in all species analyzed, suggesting that this novel class of small RNAs may be a shared feature across all Metazoa and not a unique feature of the Manila clam or bivalves at large. Features like strandedness, genomic position, length, and possible targets are discussed and compared across taxa.

Introduction

The mitochondrial DNA (mtDNA) in animals is characterized by a highly compact structure, with minimal intergenic regions except for the so-called control region (A + T-rich region in insects), and this has led to the assumption that mtDNA function is restricted to its core set of 36 to 37 genes (Boore 1999). However, recent studies have revealed a substantial structural and functional variability in mitochondrial genomes (Zardoya 2020; Formaggioni et al. 2021), as well as the presence of supernumerary genes and complex signal sequences (Ghiselli et al. 2013; Breton et al. 2014; Weber-Lotfi et al. 2015). This, in turn, suggests that mtDNA transcription may support additional functions, including, but not limited to, genes within intergenic regions that are processed during primary transcript maturation (Ro et al. 2013).

Mitochondria display extensive interactions with the nucleus and other organelles through different signaling pathways, vesicle transport, and contact sites. Signals from the nucleus to mitochondria are well-characterized (Weber-Lotfi et al. 2015), and the possibility has been suggested that mitochondria can similarly influence the expression of nuclear transcripts via the Mitochondrial Retrograde Response

(MRR; Butow and Avadhani 2004), a "process whereby mitochondria send signals to the nucleus to modulate its gene expression in response to changes in their metabolic status. Overall, this evidence strongly supports the idea that a sizable interplay takes place between the mitochondrion and the nucleus." Having the potential to carry specific information and enable targeted signaling, mitochondrial noncoding RNA molecules may, in turn, play a pivotal role in this process (Cagin et al 2015; Ovciarikova et al. 2022; Jiang and Ye 2025; Khatri and Blumental-Perry 2025; Zhao et al 2025).

Focusing on nuclear–nuclear signaling, that is currently much more thoroughly understood, multiple types of transacting small noncoding RNAs (sncRNAs) 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 depend 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). Similarly, piRNAs 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; Bofill-De Ros and Vang Ørom 2024; Chen and Kim 2024). They originate from RNA hairpin structures (Bartel 2018; Bofill-De Ros and Vang Ørom 2024; Chen and Kim 2024) 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 "noncanonical" 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).

Focusing, in turn, on the possibility of a parallel mitochondrial-nuclear signaling, and RNA-mediated MRR specifically, a newly identified class of mitochondrial RNAs, termed small mitochondrial highly transcribed RNAs (smithRNAs; Pozzi et al. 2017), has been implicated in regulating nuclear genes, with at least two smithRNAs proven to have functional effects in vivo (Passamonti et al. 2020). SmithRNAs were first discovered in the Manila clam *Ruditapes philippinarum*, where they may play a role in gonad formation and Doubly Uniparental Inheritance of mitochondria (see Pozzi et al. 2017; Passamonti and Plazzi 2020 for details). Although a comprehensive understanding of their origin and role remains elusive, initial information on their presence in other species, and hypotheses on their possible maturation pathways, are emerging in the literature. In more detail, smithRNA-like molecules were identified in the unionid bivalve *Potamilus streckersoni*, where they have been proposed to play a role in sex determination (Smith et al. 2023). A preliminary bioinformatic analysis has provided evidence for the presence of smithRNAs in *Drosophila melanogaster*, *Mus musculus*, and *Danio rerio* (Passamonti et al. 2020), and several studies have also reported the presence of small mtDNA-encoded RNAs in vertebrates (Pozzi and Dowling 2022; Khatri and Blumental-Perry 2025). Initial findings on their origin and maturation suggested that smithRNAs are primarily transcribed from the mitochondrial genome rather than nuclear pseudogenes (Pozzi and Dowling 2019) and can bind to Argonaute 2 (Pozzi and Dowling 2022). Concerning their potential function, small mitochondrial RNAs were variously implicated in sex determination (Pozzi et al. 2017; Smith et al. 2023) and histone H3 acetylation (Passamonti et al. 2020) in bivalves, response to anoxia in the killifish *Austrofundulus limnaeus* (Riggs et al. 2019), and may also be implicated in malignant cell proliferation (Khatri and Blumental-Perry 2025) as well as mitochondrial diseases (Breton 2021; Pozzi and Dowling 2021) in humans. However, the biology of smithRNAs, including their production, target genes, mechanism of gene silencing, evolutionary conservation, and distribution, remain largely unknown. Furthermore, and despite considerable support for a functional role of these small RNAs, the extensive heterogeneity in experimental approaches and

detection methods warrants caution, as it cannot be excluded that some may represent mere degradational byproducts with no specific functional role (Xu et al. 2019).

In this study we wish to assess 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 will be done by analyzing genomic data (small RNAseq, transcriptome, nuclear and mitochondrial genomes) of metazoan species representative of all major lineages (Fig. 1) using the recently published SmithHunter pipeline (Marturano et al. 2024) to identify putative smithRNA and their nuclear targets. SmithRNA/target pairs identified will be 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.

Results

Analyses were carried out on a panel of 14 metazoan species, 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 *D. rerio*, the fruitfly *D. melanogaster*, the planarian *Dugesia japonica*, the earthworm *Eisenia andrei*, the Pacific oyster *Magallana gigas*, the house mouse *M. musculus*, the ribbon worm *Notospermus geniculatus*, the sea urchin *Paracentrotus lividus*, and the Japanese beetle *Popillia japonica* (see Table 1 for complete taxonomic information). 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.

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 foster 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*, and *P. lividus* (Table 1). Illumina whole-genome sequencing produced 74 to 85 million of read pairs per library (Table S3). 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 (Table S3). Assembly of *D. japonica*, in turn, produced two different contigs encompassing the same area (~85%) of the genome. Based on the sharing of mutations with sequences in Bessho et al. (1992, 1997), we considered the two sequences as the outcome of heteroplasmy, as previously described for this species and associated to 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 (Table S4) 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 and Table S5). 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. 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 (Table S5). The fraction of reads uniquely mapping to the mitochondrial genome (ie 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% (Table S5), 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

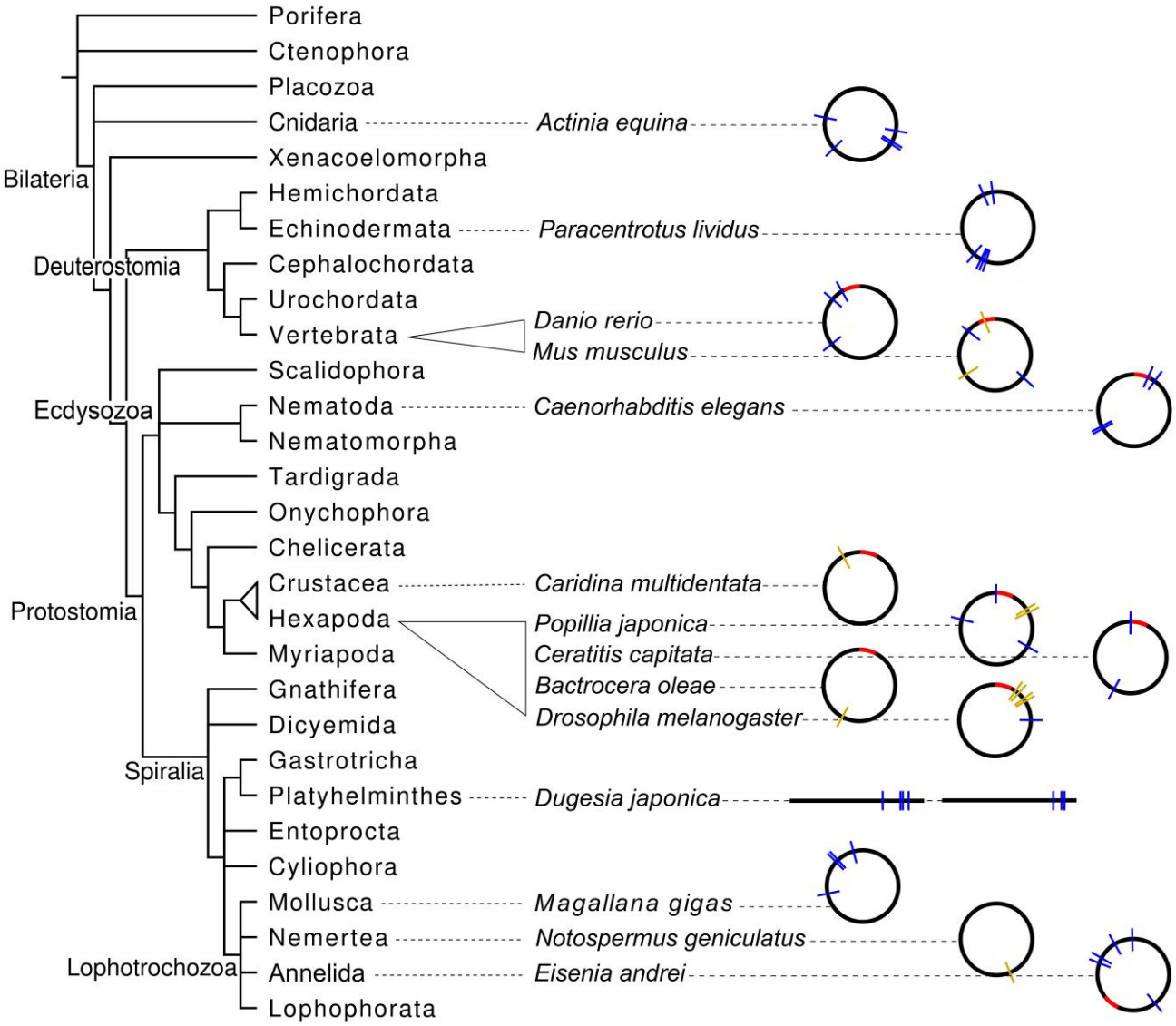


Fig. 1. List of the species studied here over a consensus phylogeny of Metazoa from Brusca et al. (2023). Circles and lines represent mitochondrial genomes, and colored segments represent the smithRNAs identified (blue, forward strand; yellow, reverse strand; red, major noncoding region, if present).

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% (Table S5). 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% (Table S5). 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 Fig. 2. Trimmed reads display a bimodal distribution, with two peaks at ~21 to 22 bp and ~30 bp. Mito-unique reads display a broader distribution, with visible peaks at ~20 to 22 bp, ~27 to 28 bp, and over 40 bp in different species. Reads associated to bona fide smithRNAs are mostly in the 20 to 32 bp range, but with evidence of substantial peaks corresponding to longer molecules in *M. gigas*, *M. musculus*, and *A. equina* (see below).

Initial read clustering and filtering of mito-unique reads in SmithHunter module A 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 (sensu

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

Species	Abbr.[a]	Classification	Transcriptome	Mitochondrial genome	Small RNAs	Putative smithRNAs	Candidate smithRNAs	No. of targets
Actinia equina	AEQ	Cnidaria	De novo[b]	De novo	De novo	5	5	235
Bactrocera oleae	BOL	Hexapoda: Diptera	Reassembled[c]	Database[d]	De novo	1	1	28
Caenorhabditis elegans	CEL	Nematoda	Reassembled	Database	Database	4	4	197
Caridina multidentata[e]	CMU	Crustacea	De novo	De novo	De novo	1	1	200
Ceratitis capitata	CCA	Hexapoda: Diptera	Reassembled	De novo	De novo	2	2	21
Danio rerio	DRE	Vertebrata: Actinopterygii	Reassembled	Database	Database	3	3	31
Drosophila melanogaster	DME	Hexapoda: Diptera	Reassembled	Database	Database	5	5	87
Dugesia japonica	DJA	Platyhelminthes	De novo	De novo	De novo	9	6	126
Eisenia andrei	EAN	Annelida	De novo	De novo	De novo	5	5	60
Magallana gigas[f]	MGI	Mollusca	Reassembled	Database	Database	4	4	314
Mus musculus	MMU	Vertebrata: Mammalia	Reassembled	Database	Database	4	4	402
Notospermus geniculatus	NGE	Nemertea	Reassembled	Database	Database	1	1	61
Paracentrotus lividus	PLI	Echinodermata	De novo	De novo	De novo	6	6	166
Popillia japonica	PJA	Hexapoda: Coleoptera	Reassembled	Database	De novo	5	5	194
…	…	…	…	…	TOTAL	55	52	2122

[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*.

Marturano et al. 2024) smithRNAs globally (1 to 9 in different species; Table 1). Final filtering in SmithHunter module B based on target matching slightly decreased this figure to 52 candidate smithRNAs (1 to 6 in different species; sensu Marturano et al. (2024); henceforth referred to as "smithRNAs" *proper*) 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 (Fig. 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* (38 bp), *M. musculus* (41 bp), and, most noticeably, *A. equina* (42 and 49 bp; Fig. 2, Table S6). SmithRNAs were observed to be transcribed from both the forward ($n = 41$) and the reverse ($n = 11$) strand (Fig. 1, Table S6). Most interestingly, in all ($n = 47$) unambiguous cases, ie 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 ($n = 5$), ie where the smithRNAs is located in a noncoding region, the same correlation could be confirmed as follows: (i) three smithRNAs encoded in intergenic regions are transcribed from the same strand as the preceding as well as the following gene; (ii) 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 S6).

Concerning the genomic location from which smithRNAs are transcribed, it appears that they can originate from all locations: protein coding gene (PCGs), rRNAs, tRNAs, as well as intergenic regions and the control region (Fig. 1, Table S6). 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; Fig. 1, Table S6, Fig. S1). The distribution observed across regions (19.2% in PCGs, 32.7% in tRNAs,

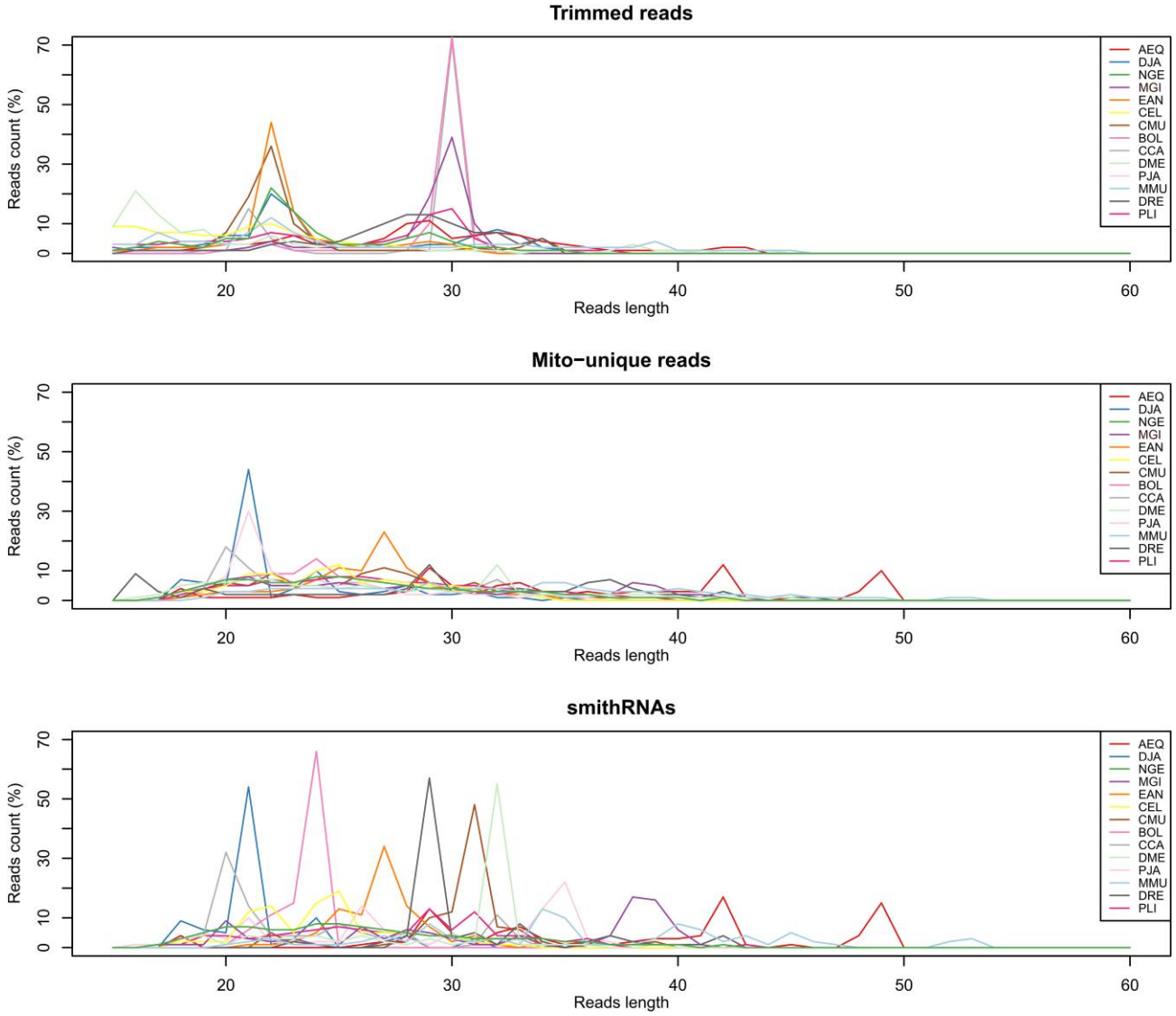


Fig. 2. Length distribution of all reads (trimmed), reads uniquely mapping to the mitochondrial genome and reads from bona fide smithRNAs. See Table S6 for details and Table 1 for species abbreviation.

38.5% in rRNAs and 9.6% in noncoding regions; Fig. 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 < 2.2×10^{-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 S6).

The 20 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, whereas the small subunit for 8. 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, five 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 S6), 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 are 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 (Fig. 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 (Fig. 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.

smithRNAs Nuclear Targets

SmithRNA/target matching in SmithHunter module B against the assembled transcriptomes led to the identification of possible targets for all 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 to 25 range, with notable exceptions where one single smithRNA was associated to as many as 200 (CMU_smith2) or 214 (MGI_smith0) nuclear targets (Fig. 4, Table S7). 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 (Table S7, Fig. 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 smithRNA function.

Within the category Biological Processes (Fig. 5 and Fig. 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; Fig. 5 and Fig. S3). Within the category Molecular Functions (Fig. S4a), 10 out of 31 enriched terms relate to binding activities and were identified in *A. equina*, *M. musculus*, *C. elegans*, the freshwater shrimp *C. multidentata*, the Pacific oyster *M. gigas*, *N. geniculatus*, and *E. andrei*. Moreover, *NAD*-dependent histone H3K9 deacetylase activity was enriched in *N. geniculatus*. Among Cellular Components (Fig. S4b), terms related to germline biology (spectrosome and spectrine) were enriched in *C. elegans*.

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. Evidences along this line are nevertheless fairly sparse at present and were not investigated further.

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 1 and 6 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 (Figs. 1 and 4; Table 1, Tables S6 and S7). The sheer dimension of these numbers supports the hypothesis that smithRNAs, though

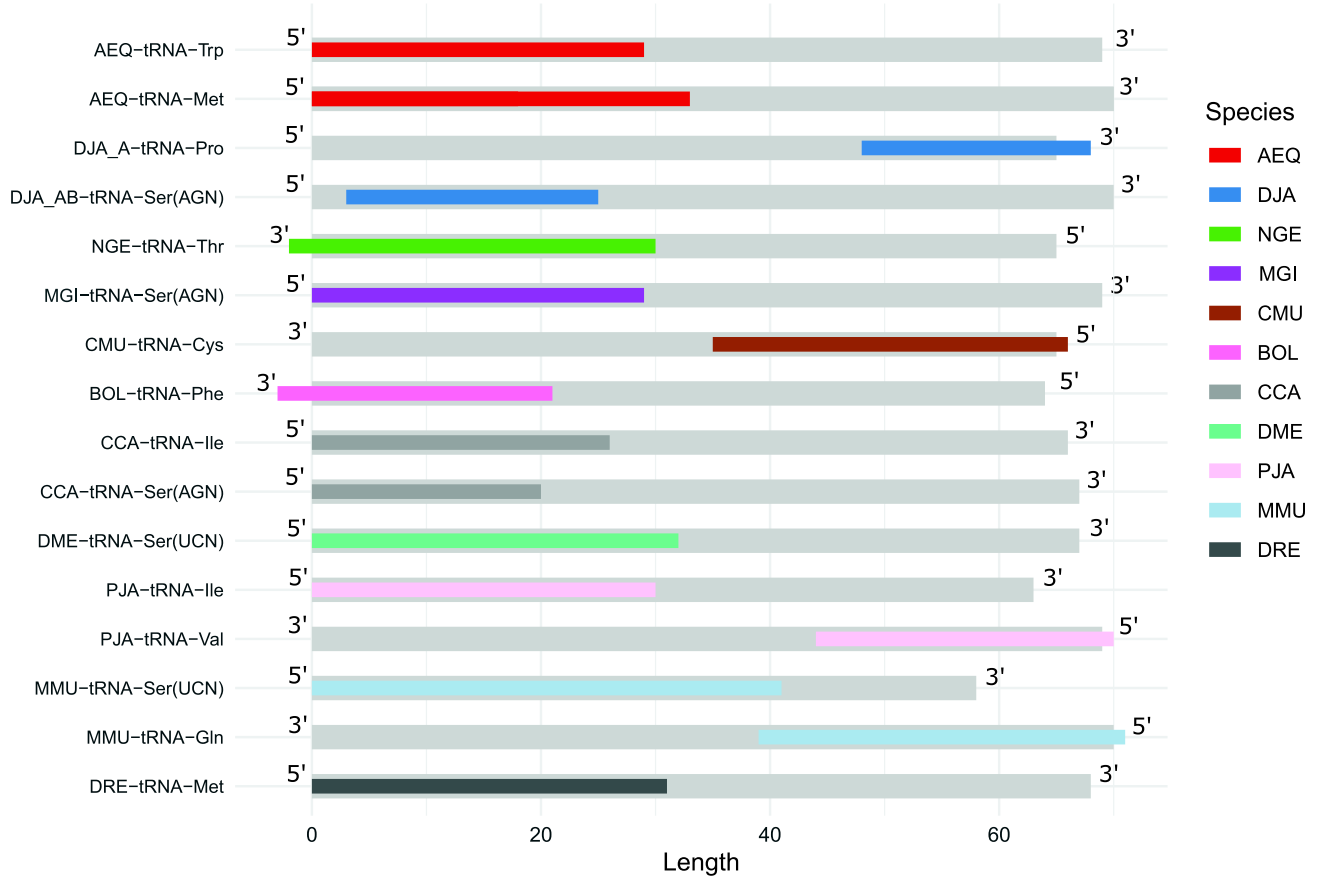


Fig. 3. Position of smithRNAs identified within tRNA genes. SmithRNAs are represented in color, tRNAs are represented in background and their strandedness is indicated. See Table 1 for species abbreviations.

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 (Fig. 1), spanning over Deuterostomia (3 species), Ecdysozoa (6 species), Spiralia (4 species), and Cnidaria (1 species), combined with previous fragmentary reports of miRNA-like sequences across various taxa, including *D. melanogaster* (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), further suggests that smithRNAs may be a shared feature that was present in the mitochondrial genome since the outset of the diversification of Metazoa.

While some differences were observed among taxonomic groups (eg a higher propensity for smithRNAs in the reverse strand in Ecdysozoa compared to other taxa; a substantial reduction of smithRNA reads in Diptera compared to other taxa, and especially Deuterostomia; smithRNAs preferentially targeting different pathways, according to the GO term analysis), we deem 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. 2016).

Function

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 *P. streckersoni* (Smith et al.

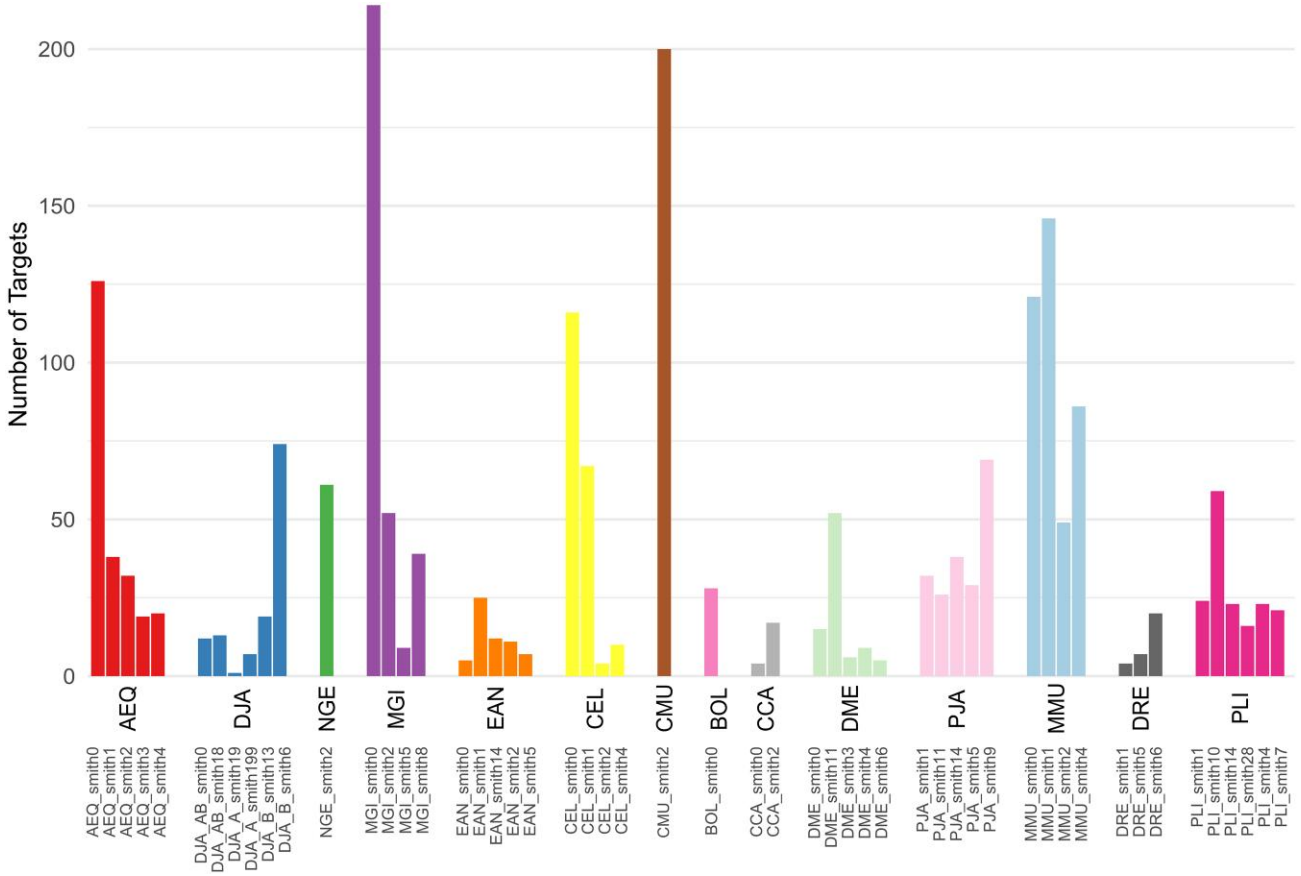


Fig. 4. Number of nuclear targets identified per smithRNA. See Table 1 for species abbreviations.

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* (Fig. 5, Figs. S3 and S4, Table S7). Furthermore, smithRNAs involved in the regulation of spermatogenesis and sperm motility were identified in most species (Table S7), 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* (Fig. 5, Figs. S3 and S4, Table S7). Notably, these functions find some parallel in the role suggested by Riggs et al. (2019) and Blumental-Perry et al. (2020) concerning response to stress in the annual killifish *A. limnaeus* and mice, respectively, as well as Passamonti et al. (2020) concerning histone acetylation in the Manila clam.

Overall, smithRNAs seem to be involved in regulatory networks which differ across species. This suggests smithRNAs as a dynamic regulatory mechanism that can be recruited to control different pathways in different taxa as needed. This dynamism 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, we 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 we think that evidence for the modulation of a given pathway or function at large (eg 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 noncoding regions and as little as 5.8% from intergenic spacers (Table S6). In general terms, this may be a consequence of the peculiar structure of the mitochondrial genome of bivalves, where long unassigned regions (ie intergenic spacers) are observed. By contrast, canonical metazoan mitochondrial genomes 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, structures 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), as well as in the killifish (Riggs et al. 2019), and have been discussed in Khatri and Blumental-Perry (2025). 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), and both the 12S and 16S mitochondrial rRNAs, as well as nuclear rRNAs, have been identified as 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 in a preliminary analysis conducted by 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 tRNA, rRNA, or PCGs from which they originate (Table S6), 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 (Fig. 3 and Fig. 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 mechanisms 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 multigene 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 (Fig. 2) currently precludes a definitive interpretation. The length distribution of trimmed reads, characterized by two peaks at approximately 21 to 22 nt and ~30 nt, aligns with expectations, particularly given that miRNAs, the most

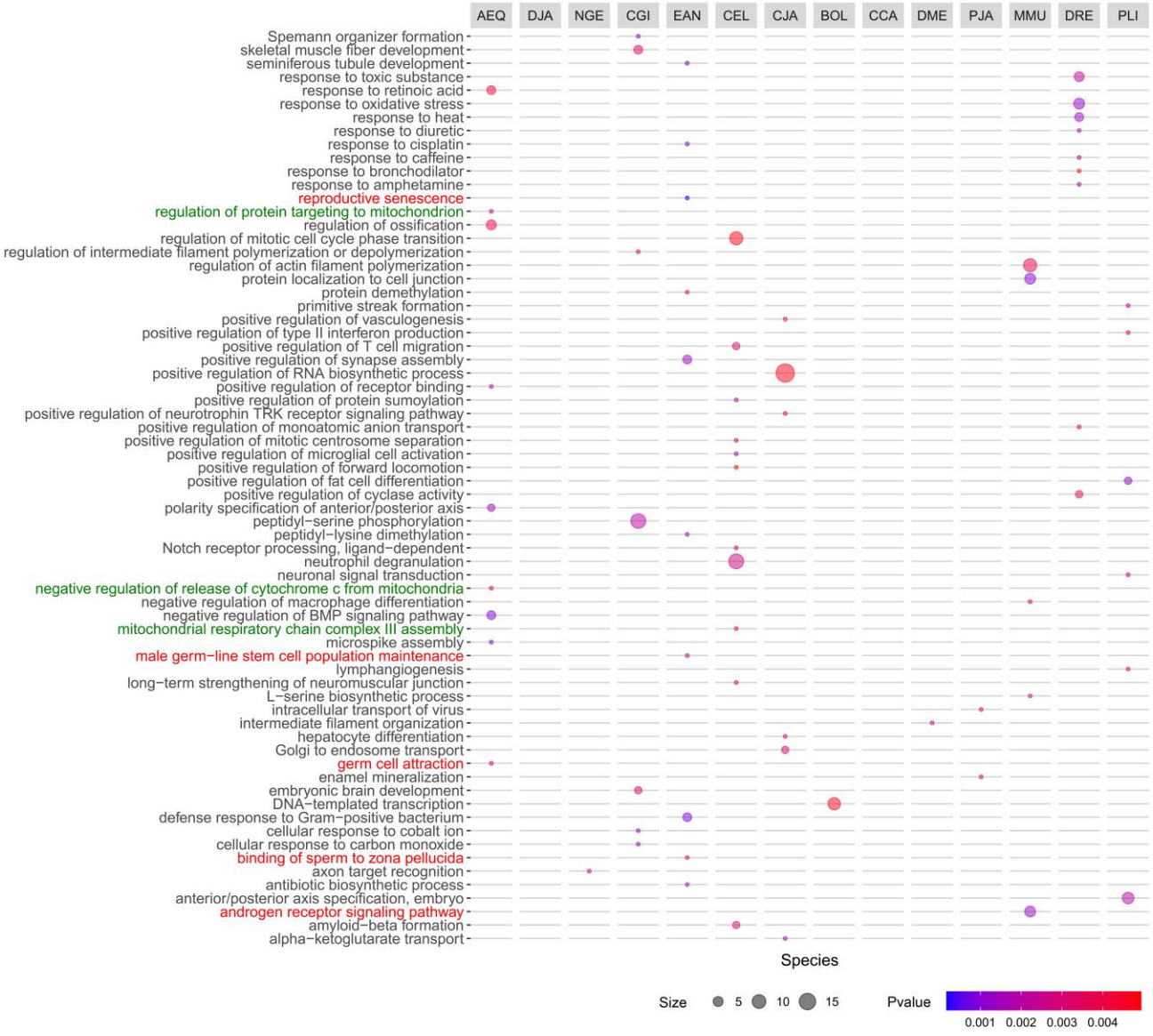


Fig. 5. GO enrichment analysis of biological processes performed on the supposed nuclear targets of smithRNAs, by species. Only GO supported by more than one gene are shown. See Fig. S3 for complete information and Fig. S4 for molecular functions and cellular components. Circle area represents the number of genes, the *P*-value is color coded. Go terms associated to mitochondrial and sex related functions are highlighted in green and red, respectively. See Table 1 for species abbreviations.

abundant class in adult cells, typically measure around 22 nt (Ha and Kim 2014). On the other hand, the detection of smithRNAs spanning the entire 20 to 32 nt range, including some longer variants, remains unexplained. Nevertheless, 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 pathway of miRNAs (Ha and Kim 2014). This pathway, however, may not apply to smithRNAs, for which distinct transcriptional or maturation mechanisms have been proposed. In summary, smithRNAs appear to constitute a subset of the broader family of small regulatory noncoding RNAs (Khatri and Blumental-Perry 2025 for examples in mitochondria) but do not correspond directly to the currently better-characterized miRNA, piRNA, or siRNA classes. This observation highlights that the biogenesis and functional role of smithRNAs are still poorly understood and calls for a more comprehensive investigation.

Conclusions

Following the initial description and characterization of smithRNAs in the Manila clam, the question was posed whether smithRNAs are to be considered "an odd feature

of an odd system" or rather a mechanism of more general interest, possibly shared by the entire Metazoa (Pozzi et al. 2017; Marturano et al. 2024). Based on the results presented here, we think we can safely conclude that smithRNAs are a widespread, possibly ubiquitous, feature of Metazoa. We further advocate the use of a single category to include all small, mitochondrially encoded, RNAs that regulate nuclear gene expression, minimally smithRNAs (Pozzi et al. 2017) and mito-ncRNAs (Khatri and Blumental-Perry 2025). We tend to prefer the former denomination as it has already been adopted in the literature.

Noteworthy, the possibility that sncRNAs originating form mitochondria could influence the expression of nuclear genes introduces a novel dimension to mitochondrial biology. An improved understanding of mitochondrial functions could, in fact, provide additional insights into the role of mitochondria in eukaryotic cells, as well as in the origins and evolution of Eukarya at large. Furthermore, the evolutionary roots of mitochondria within α-proteobacteria—some of which are known to interact with hosts and influence their gene expression—raise questions about which ancestral mechanisms may have been inherited from free-living α-proteobacterial ancestors and how these mechanisms contributed to the evolution of eukaryotic cells. Our attention is currently directed toward these themes.

In our view, the current priority in smithRNA research is the possibility to confirm the functionality of multiple smithRNAs in vivo using microinjection experiments, as in Passamonti et al. (2020). This, in turn, will (i) provide a final in vivo proof of smithRNA functionality across Metazoa; (ii) identify a number of experimental and reporting systems that can be further deployed to study smithRNAs in vivo; (iii) allow a revision of SmithHunter filtering procedures to enhance its discriminatory power and, ultimately, improve its predictions.

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 as follows: (i) transcriptomes were assembled/annotated de novo for all species applying the same methodology; (ii) libraries from whole body samples were preferred; (iii) mitochondrial genomes were sequenced, whenever possible, from a genetically related individual to the one used for small RNA sequencing; (iv) 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 five species and reassembled based on raw reads from NCBI in the remaining nine 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 2×150 bp 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. See Table S1 for full information. 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 (ver. 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 (ver. 4.3.6; Kopylova et al. 2012) based on the smr_v4.3_default database. See Table S1 for filtering information. All 14 transcriptomes were assembled using Trinity (ver. 2.8.5; Grabherr et al. 2011) with default parameters. The quality of individual assemblies was evaluated based on the following: (i) fragmentation, using Assembly-stats (https://github.com/sanger-pathogens/assembly-stats); (ii) completeness, using BUSCO (ver 4.1.4; Manni et al. 2021) against the metazoan_odb10 database; and (iii) the fraction of properly remapping reads, using Bowtie2 (ver. 2.3.5.1; Langmead and Salzberg 2012) and SAMtools (ver. 1.6; Danecek et al. 2021; Table S2). Any remaining mitochondrial transcript, identified in BLAST (ncbi-blast ver. 2.12.0; 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 (ver. 5.5.0; written by BJ Haas) and Trinotate (ver. 3.11; Bryant et al. 2017 and references herein). 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 (ie 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 six 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 × 150 bp 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; Chen et al. 2018) and assembled in NOVOplasty (v. 4.3.5; Dierckxsens et al. 2016) 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 eight species was retrieved from NCBI (see Table S3 for full information).

Nuclear Genome Retrieval

Nuclear genome assemblies were downloaded from NCBI or CNCB for all species (Table 1 and Table S4). Following a BLAST search (ncbi-blast ver. 2.12.0), 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 eight species (Table 1). Total RNA was extracted using the miRNeasy Micro kit (Quiagen) 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 × 150 bp 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 we acknowledge a limitation in terms of comparability, we 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 sequence of small RNAs for the remaining six species was retrieved from NCBI, selecting library collections that had between three and six replicates per species. See Table S5 for full information.

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, the analytical settings applied 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 (Alexa and Rahnenfuhrer 2024). Plots were produced using ggplot in R (Wickham 2016; R Core Team 2022). 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).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

The authors wish to thank Prof. Alessandra Salvetti from the University of Pisa for providing samples of *D. japonica* as well as all colleagues from the research groups at the University of Siena and at the University of Bologna for their daily contributions and useful discussions.

Funding

The study was supported by the Italian Ministry of University and Research under the program PRIN2020 (Progetti di Ricerca di Rilevante Interesse Nazionale) funded to M.P. and F.N. (project MitoMicro; 2020BE2BC3) as well as NBFC to University of Siena/Department of Life Sciences, funded by the Italian Ministry of University and Research, PNRR, Missione 4 Componente 2, "Dalla ricerca all'impresa", Investimento 1.4, Project CN00000033.

Conflict of Interest

The authors declare no conflict of interest

Data Availability

All new data produced for species *E. andrei*, *C. capitata*, *D. japonica*, *C. multidentata*, *P. lividus*, *A. equina*, *B. oleae*, and *P. japonica* have been deposited in the NCBI database under BioProject PRJNA1037026. Mitochondrial genome sequencing data have been deposited as follows. Biosamples: SAMN40215639 (*E. andrei*, *C. capitata*, pool), SAMN40215640 (*D. japonica*, pool), SAMN40215641 (*C. multidentata*), SAMN40215642 (*P. lividus*), SAMN40215643 (*A. equina*). Raw sequence data: SRR28211706 (*E. andrei*, *C. capitata*), SRR28211705 (*D. japonica*), SRR28211704 (*C. multidentata*), SRR28211703 (*P. lividus*), SRR28211702 (*A. equina*). Annotated mitochondrial genomes: PP438738 (*E. andrei*), PP438740 (*C. capitata*), PP484941 (*D. japonica* genome A), PP484942 (*D. japonica* genome B), PP438739 (*C. multidentata*), PP438737 (*P. lividus*), PP438735 (*A. equina*). Transcriptome sequencing data have been deposited as follows. Biosamples: SAMN40418221-SAMN40418225 (*E. andrei*, whole body), SAMN40418226-SAMN40418230 (*D. japonica*, whole body, pool), SAMN40418231-SAMN40418235 (*C. multidentata*, whole body), SAMN40418236-SAMN40418245 (*P. lividus*, gonads), SAMN40418246-SAMN40418250 (*A. equina*, whole body). Raw sequence data: SRR28330936, SRR28330937, SRR28330948, SRR28330959, SRR28330960 (*E. andrei*), SRR28330931-SRR28330935 (*D. japonica*), SRR28330954-SRR28330958 (*C. multidentata*), SRR28330943-SRR28330947, SRR28330949-SRR28330953 (*P. lividus*), SRR28330938-SRR28330942 (*A. equina*). Small RNA sequencing data have been deposited as follows. Biosasmples: SAMN40545089-SAMN40545093 (*C. capitata*, whole body, pool), SAMN40545094-SAMN40545098 (*E. andrei*, whole body), SAMN40545099-SAMN40545103 (*D. japonica*, whole body, pool), SAMN40545104-SAMN40545108 (*B. oleae*, whole body, pool), SAMN40545109-SAMN40545113 (*C. multidentata*, whole body), SAMN40545114-SAMN40545118 (*P. lividus*, male gonads), SAMN40545124-SAMN40545128 (*P. japonica*, whole body), SAMN40418246-SAMN40418250 (*A. equina*, whole body). Raw sequence data: SRR28406999, SRR28407010, SRR28407021, SRR28407032, SRR28407033 (*C. capitata*), SRR28406989-SRR28406993 (*E. andrei*), SRR28407027-SRR28407031 (*D. japonica*), SRR28407022-SRR28407026 (*B. oleae*), SRR28407016-SRR28407020 (*C. multidentata*), SRR28407011-SRR28407015 (*P. lividus*), SRR28407000-SRR28407004 (*P. japonica*), SRR28406994-SRR28406998 (*A. equina*).

Literature Cited

Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. R package version 2.58.0. 2024.

Alhassan AR, et al. Regulatory roles of small RNAs in plant growth, breeding, and stress adaptation. Plant Growth Regul. 2025:105:295–313. https://doi.org/10.1007/s10725-024-01271-6.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990:215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Asha S, Soniya EV. The sRNAome mining revealed existence of unique signature small RNAs derived from 5.8SrRNA from *Piper nigrum* and other plant lineages. Sci Rep. 2017:7:40152. https://doi.org/10.1038/srep41052.

Bartel DP. Metazoan MicroRNAs. Cell. 2018:173:20–51. https://doi.org/10.1016/j.cell.2018.03.006.

Bensasson D, Zhang D, Hartl DL, Hewitt GM. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol. 2001:16:314–321. https://doi.org/10.1016/S0169-5347(01)02151-6.

Bernt M, et al. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 2013:69:313–319. https://doi.org/10.1016/j.ympev.2012.08.023.

Bessho Y, et al. Planarian mitochondria sequence heterogeneity: relationships between the type of cytochrome c oxidase subunit I gene sequence, karyotype and genital organ. Mol Ecol. 1997:6:129–136. https://doi.org/10.1046/j.1365-294X.1997.00162.x.

Bessho Y, Ohama T, Osawa S. Planarian mitochondria. I. Heterogeneity of cytochrome c oxidase subunit I gene sequences in the freshwater planarian, *Dugesia japonica*. J Mol Evol. 1992:34:324–330. https://doi.org/10.1007/BF00160239.

Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. Evolution of DNA methylation across insects. Mol Biol Evol. 2016:34:654–665. https://doi.org/10.1093/molbev/msw264.

Blumenstiel JP. From the cauldron of conflict: endogenous gene regulation by piRNA and other modes of adaptation enabled by selfish transposable elements. Semin Cell Dev Biol. 2025:164:1–12. https://doi.org/10.1016/j.semcdb.2024.05.001.

Blumental-Perry A, et al. Retrograde signaling by a mtDNA-encoded non-coding RNA preserves mitochondrial bioenergetics. Commun Biol. 2020:3:626. https://doi.org/10.1038/s42003-020-01322-4.

Bofill-De Ros X, Vang Ørom UA. Recent progress in miRNA biogenesis and decay. RNA Biol. 2024:21:36–43. https://doi.org/10.1080/15476286.2023.2288741.

Boore JL. Animal mitochondrial genomes. Nucleic Acids Res. 1999:27:1767–1780. https://doi.org/10.1093/nar/27.8.1767.
Breton S. Mitochondrial Russian doll genes may explain some discrepancies in links between mtDNA mutations and mitochondrial diseases. Bioessays. 2021:43:e2100104. https://doi.org/10.1002/bies.202100104.
Breton S, et al. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. Trends Genet. 2014:30:555–564. https://doi.org/10.1016/j.tig.2014.09.002.
Brusca RC, Giribet G, Moore W. Invertebrates. 4th edition Oxford University Press; 2023.
Bryant DM, et al. A tissue-mapped axolotl *de novo* transcriptome enables identification of limb regeneration factors. Cell Rep. 2017:18:762–776. https://doi.org/10.1016/j.celrep.2016.12.063.
Burzio VA, et al. Expression of a family of noncoding mitochondrial RNAs distinguishes normal from cancer cells. Proc Natl Acad Sci U S A. 2009:106:9430–9434. https://doi.org/10.1073/pnas.0903086106.
Butow RA, Avadhani NG. Mitochondrial signaling: the retrograde response. Mol Cell. 2004:14:1–15. https://doi.org/10.1016/S1097-2765(04)00179-0.
Cagin U, et al. Mitochondrial retrograde signaling regulates neuronal function. Proc Natl Acad Sci U S A. 2015:112:E6000–E6009. https://doi.org/10.1073/pnas.1505036112.
Calcino AD, Fernandez-Valverde SL, Taft RJ, Degnan BM. Diverse RNA interference strategies in early-branching metazoans. BMC Evol Biol. 2018:18:160. https://doi.org/10.1186/s12862-018-1274-2.
Chen L-L, Kim VN. Small and long non-coding RNAs: past, present, and future. Cell. 2024:187:6451–6485. https://doi.org/10.1016/j.cell.2024.10.024.
Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018:34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
Chen Z, et al. Two featured series of rRNA-derived RNA fragments (rRFs) constitute a novel class of small RNAs. PLoS One. 2017:12:e0176458. https://doi.org/10.1371/journal.pone.0176458.
Cherlin T, et al. Ribosomal RNA fragmentation into short RNAs (rRFs) is modulated in a sex- and population of origin-specific manner. BMC Biol. 2020:18:38. https://doi.org/10.1186/s12915-020-0763-0.
Cucini C, et al. *De novo* assembly and annotation of *Popillia japonica*'s genome with initial clues to its potential as an invasive pest. BMC Genomics. 2024:25:275. https://doi.org/10.1186/s12864-024-10180-x.
Czech B, et al. piRNA-guided genome defense: from biogenesis to silencing. Annu Rev Genet. 2018:52:131–157. https://doi.org/10.1146/annurev-genet-120417-031441.
Danecek P, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021:10:giab008. https://doi.org/10.1093/gigascience/giab008.
Daugaard I, Hansen TB. Biogenesis and function of ago-associated RNAs. Trends Genet. 2017:33:208–219. https://doi.org/10.1016/j.tig.2017.01.003.
Dexheimer PJ, Cochella L. MicroRNAs: from mechanism to organism. Front Cell Dev Biol. 2020:8:409. https://doi.org/10.3389/fcell.2020.00409.
Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2016:45:e18. https://doi.org/10.1093/nar/gkw955.
D'Souza AR, Minczuk M. Mitochondrial transcription and translation: overview. Essays Biochem. 2018:62:309–320. https://doi.org/10.1042/EBC20170102.
Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019:20:238. https://doi.org/10.1186/s13059-019-1832-y.
Formaggioni A, et al. The evolution and characterization of the RNA interference pathways in lophotrochozoa. Gen Biol Evol. 2024:16:evae098. https://doi.org/10.1093/gbe/evae098.
Formaggioni A, Luchetti A, Plazzi F. Mitochondrial genomic landscape: a portrait of the mitochondrial genome 40 years after the first complete sequence. Life. 2021:11:663. https://doi.org/10.3390/life11070663.
Ghiselli F, et al. Structure, transcription, and variability of metazoan mitochondrial genome: perspectives from an unusual mitochondrial inheritance system. Gen Biol Evol. 2013:5:1535–1554. https://doi.org/10.1093/gbe/evt112.
Grabherr MG, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011:29:644–652. https://doi.org/10.1038/nbt.1883.
Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014:15:509–524. https://doi.org/10.1038/nrm3838.
Iki T, Kawaguhi S, Kai T. miRNA/siRNA-directed pathway to produce noncoding piRNAs from endogenous protein-coding regions ensures *Drosophila* spermatogenesis. Sci Adv. 2023:9:eadh0397. https://doi.org/10.1126/sciadv.adh0397.
Iwakawa H-o, Tomari Y. Life of RISC: formation, action, and degradation of RNA-induced silencing complex. Mol Cell. 2022:82:30–43. https://doi.org/10.1016/j.molcel.2021.11.026.
Jiang H, Ye J. The Warburg effect: the hacked mitochondrial-nuclear communication in cancer. Semin Cancer Biol. 2025:112:93–111. https://doi.org/10.1016/j.semcancer.2025.03.006.
Kawaji H, et al. Hidden layers of human small RNAs. BMC Genomics. 2008:9:157. https://doi.org/10.1186/1471-2164-9-157.
Khatri S, Blumental-Perry A. Mitochondrial ncRNAs beyond the mitochondrion: coordinators of organelle crosstalk. Curr Opin Physiol. 2025:45:100831. https://doi.org/10.1016/j.cophys.2025.100831.
Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012:28:3211–3217. https://doi.org/10.1093/bioinformatics/bts611.
Kumar P, Anaya J, Mudunuri SB, Dutta A. 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. 2014:12:78. https://doi.org/10.1186/s12915-014-0078-0.
Kusch S, Singh M, Thieron H, Spanu PD, Panstruga R. 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. 2023:24:570–587. https://doi.org/10.1111/mpp.13324.
Kuscu C, et al. tRNA fragments (tRFs) guide ago to regulate gene expression post-transcriptionally in a dicer-independent manner. RNA. 2018:24:1093–1105. https://doi.org/10.1261/rna.066126.118.
Lambert M, et al. Identification of abundant and functional dodecaRNAs (doRNAs) derived from ribosomal RNA. Int J Mol Sci. 2021:22:9757. https://doi.org/10.3390/ijms22189757.
Lambert M, Benmoussa A, Provost P. Small non-coding RNAs derived from eukaryotic ribosomal RNA. Noncoding RNA. 2019:5:16. https://doi.org/10.3390/ncrna5010016.
Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012:9:357–359. https://doi.org/10.1038/nmeth.1923.
Lax C, et al. The evolutionary significance of RNAi in the fungal kingdom. Int J Mol Sci. 2020:21:9348. https://doi.org/10.3390/ijms21249348.
Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. Bioinformatics. 2015:31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.
Li Y, Hui JHL. Small RNAs in Cnidaria: a review. Evol Appl. 2023:16:354–364. https://doi.org/10.1111/eva.13445.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021:38:4647–4654. https://doi.org/10.1093/molbev/msab199.

Marturano G, et al. SmithHunter: a workflow for the identification of candidate smithRNAs and their targets. BMC Bioinformatics. 2024:25:286. https://doi.org/10.1186/s12859-024-05909-0.

Mercer TR, et al. The human mitochondrial transcriptome. Cell. 2011:146:645–658. https://doi.org/10.1016/j.cell.2011.06.051.

Meseguer S. MicroRNAs and tRNA-derived small fragments: key messengers in nuclear–mitochondrial communication. Front Mol Biosci. 2021:8:643575. https://doi.org/10.3389/fmolb.2021.643575.

Meseguer S, et al. The MELAS mutation m. 3243A>G alters the expression of mitochondrial tRNA fragments. BBA-Mol Cell Res. 2019:1866:1433–1449. https://doi.org/10.1016/j.bbamcr.2019.06.004.

Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. Cell. 2007:130:89–100. https://doi.org/10.1016/j.cell.2007.06.028.

Onishi R, Yamanaka S, Siomi MC. piRNA- and siRNA-mediated transcriptional repression in *Drosophila*, mice, and yeast: new insights and biodiversity. EMBO Rep. 2021:22:e53062. https://doi.org/10.15252/embr.202153062.

Ovciarikova J, Shikha S, Sheiner L. Nuclear interactions: a spotlight on nuclear mitochondrial membrane contact sites. Contact. 2022:5:1–7. https://doi.org/10.1177/25152564221096217.

Passamonti M, Calderone M, Delpero M, Plazzi F. Clues of in vivo nuclear gene regulation by mitochondrial short non-coding RNAs. Sci Rep. 2020:10:8219. https://doi.org/10.1038/s41598-020-65084-z.

Passamonti M, Plazzi F. Doubly uniparental inheritance and beyond: the contribution of the Manila clam *Ruditapes philippinarum*. J Zool Syst Evol Res. 2020:58:529–540. https://doi.org/10.1111/jzs.12371.

Plazzi F, Le Cras Y, Formaggioni A, Passamonti M. Mitochondrially mediated RNA interference, a retrograde signaling system affecting nuclear gene expression. Heredity (Edinb). 2024:132:156–161. https://doi.org/10.1038/s41437-023-00650-5.

Pozzi A, Dowling DK. The genomics origins of small mitochondrial RNAs: are they transcribed by the mitochondrial DNA or by mitochondrial pseudogenes within the nucleus (NUMTs)? Genome Biol Evol. 2019:11:1883–1896. https://doi.org/10.1093/gbe/evz132.

Pozzi A, Dowling DK. Small mitochondrial RNAs as mediators of nuclear gene regulation, and potential implications for human health. Bioessays. 2021:43:e2000265. https://doi.org/10.1002/bies.202000265.

Pozzi A, Dowling DK. New insights into mitochondrial–nuclear interactions revealed through analysis of small RNAs. Gen Biol Evol. 2022:14:evac023. https://doi.org/10.1093/gbe/evac023.

Pozzi A, Plazzi F, Milani L, Ghiselli F, Passamonti M. SmithRNAs: could mitochondria "bend" nuclear regulation? Mol Biol Evol. 2017:34:1960–1973. https://doi.org/10.1093/molbev/msx140.

R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2022.

Ratnasingham S, Hebert PD. Bold: the barcode of life data system (http://www.barcodinglife.org). Mol Ecol Notes. 2007:7:355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x.

Riggs CL, Woll SC, Podrabsky JE. MitosRNAs and extreme anoxia tolerance in embryos of the annual killifish *Austrofundulus limnaeus*. Sci Rep. 2019:9:19812. https://doi.org/10.1038/s41598-019-56231-2.

Ro S, et al. The mitochondrial genome encodes abundant small noncoding RNAs. Cell Res. 2013:23:759–774. https://doi.org/10.1038/cr.2013.37.

Shang R, Lee S, Senavirathne G, Lai EC. microRNAs in action: biogenesis, function and regulation. Nat Rev Genet. 2023:24:816–833. https://doi.org/10.1038/s41576-023-00611-y.

Shaukat A-N, Kaliatsi EG, Stamatopoulou V, Stathopoulos C. Mitochondrial tRNA-derived fragments and their contribution to gene expression regulation. Front Physiol. 2021:12:729452. https://doi.org/10.3389/fphys.2021.729452.

Shivam P, et al. Regulatory roles of PIWI-interacting RNAs in cardiovascular disease. Am J Physiol Heart Circ. 2025:328:H991–H1004. https://doi.org/10.1152/ajpheart.00833.2024.

Smith CH, et al. Mitonuclear sex determination? Empirical evidence from bivalves. Mol Biol Evol. 2023:40:msad240. https://doi.org/10.1093/molbev/msad240.

Swarts DC, et al. The evolutionary journey of argonaute proteins. Nat Struct Mol Biol. 2014:21:743–753. https://doi.org/10.1038/nsmb.2879.

Thompson DM, Lu C, Green PJ, Parker R. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. RNA. 2008:14:2095–2103. https://doi.org/10.1261/rna.1232808.

Villegas J, et al. Expression of a novel non-coding mitochondrial RNA in human proliferating cells. Nucleic Acids Res. 2007:35:7336–7347. https://doi.org/10.1093/nar/gkm863.

Wang S, et al. Genome-wide analyses across viridiplantae reveal the origin and diversification of small RNA pathway-related genes. Commun Biol. 2021:4:412. https://doi.org/10.1038/s42003-021-01933-5.

Weber-Lotfi F, et al. Nucleic acid import into mitochondria: new insights into the translocation pathways. Biochim Biophys Acta Mol Cell Res. 2015:1853:3165–3181. https://doi.org/10.1016/j.bbamcr.2015.09.011.

Westholm JO, Lai EC. Mirtrons: microRNA biogenesis via splicing. Biochimie. 2011:93:1897–1904. https://doi.org/10.1016/j.biochi.2011.06.017.

Wickham H. Ggplot2: elegant graphics for data analysis. Springer; 2016.

Wood DE, Lu J, Langmead B. Improved metagenomic analysis with kraken 2. Genome Biol. 2019:20:257. https://doi.org/10.1186/s13059-019-1891-0.

Xu X, et al. Using pan RNA-Seq analysis to reveal the ubiquitous existence of 5′ and 3′ End small RNAs. Front Genet. 2019:10:1–11. https://doi.org/10.3389/fgene.2019.00105.

Zardoya R. Recent advances in understanding mitochondrial genome diversity. F1000Res. 2020:9:270. https://doi.org/10.12688/f1000research.21490.1.

Zhang S, Sun L, Kragler F. The phloem-delivered RNA pool contains small noncoding RNAs and interferes with translation. Plant Physiol. 2009:150:378–387. https://doi.org/10.1104/pp.108.134767.

Zhao L, et al. hnRNPH1 maintains mitochondrial homeostasis by establishing NRF1/DRP1 retrograde signaling under mitochondrial stress. Cell Death Differ. 2025:32:118–133. https://doi.org/10.1038/s41418-024-01331-4.

Associate editor: Soojin Yi