



OPEN

# Clues of *in vivo* nuclear gene regulation by mitochondrial short non-coding RNAs

Marco Passamonti<sup>1,3</sup>, Marco Calderone<sup>1</sup>, Manuel Delpero<sup>1,2</sup> & Federico Plazzi<sup>1,3</sup>✉

Gene expression involves multiple processes, from transcription to translation to the mature, functional peptide, and it is regulated at multiple levels. Small RNA molecules are known to bind RNA messengers affecting their fate in the cytoplasm (a process generically termed 'RNA interference'). Such small regulatory RNAs are well-known to be originated from the nuclear genome, while the role of mitochondrial genome in RNA interference was largely overlooked. However, evidence is growing that mitochondrial DNA does provide the cell a source of interfering RNAs. Small mitochondrial highly transcribed RNAs (smithRNAs) have been proposed to be transcribed from the mitochondrion and predicted to regulate nuclear genes. Here, for the first time, we show *in vivo* clues of the activity of two smithRNAs in the Manila clam, *Ruditapes philippinarum*. Moreover, we show that smithRNAs are present and can be annotated in representatives of the three main bilaterian lineages; in some cases, they were already described and assigned to a small RNA category (e.g., piRNAs) given their biogenesis, while in other cases their biogenesis remains unclear. If mitochondria may affect nuclear gene expression through RNA interference, this opens a plethora of new possibilities for them to interact with the nucleus and makes metazoan mitochondrial DNA a much more complex genome than previously thought.

Expression of nuclear genes may be enhanced or reduced by several cellular factors<sup>1,2</sup>, including a wide array of short non-coding RNA transcripts (sncRNAs)<sup>3–6</sup>. MicroRNAs (miRNAs) are possibly the most studied ones<sup>4,7</sup>, as they are the commonest regulators of gene expression at the posttranscriptional level<sup>8–11</sup>. However, other types of sncRNAs have been characterized<sup>3–6</sup>. For instance, piRNAs are sncRNAs that bind to PIWI Argonaute proteins, and are often associated to the maintenance of germline genome stability, transposition suppression, and chromatin remodeling<sup>3,6,12–14</sup>.

Many ways are known for sncRNAs to originate from transcribing DNA and to mature towards the working molecule which will take part in the regulating complex<sup>3,5,6,10,11,15</sup>. The biogenesis of these small molecules normally involves a maturation from longer transcripts, as is the case for mature miRNAs, resulting from pre-miRNAs<sup>6,9,11,16–25</sup>, which, in turn, are obtained from pri-miRNAs<sup>6,10,11,15,16,26–31</sup>.

The interaction between the mature miRNA and the target mRNA takes place in the RNA induced silencing complex (RISC)<sup>8–11,15</sup> and it typically involves the 'seed' region, i.e. bases 2–8 in the 5' end of a miRNA<sup>6,15,16,32</sup>, even if other modes of interaction have been published<sup>15</sup>. The 'seed' (i.e., miRNA nucleotides 2–7) normally basepairs with a region within the 3' UTR of the target mRNA, and many different types of seed-target interaction are known<sup>16,32</sup>.

The nuclear DNA is not the only DNA in the eukaryotic cell. More specifically, in animal cells, mitochondria do contain their own genome, a small, normally circular molecule encoding for a dozen of protein-coding genes, two ribosomal subunits and many tRNAs<sup>33–35</sup>. Mitochondrial DNA (mtDNA) is known to transcribe for sncRNAs as well. In the human mtDNA, a total of 31 sncRNAs mapping to 17 loci (prevalently tRNAs) were originally annotated<sup>36</sup>. These loci prevalently code for tRNAs, and it is known that tRNAs may be processed by Dicer and other RNAses in the cytoplasm into a wide array of small RNA molecules that enter RNA-silencing pathways<sup>37,38</sup>. Mitochondrial genome-encoded small RNAs (mitosRNAs<sup>39</sup>) are now known from different species<sup>39–41</sup>; however, although involved in many physiological processes, like anoxic response<sup>41</sup> or gametogenesis<sup>40</sup>, mitosRNAs

<sup>1</sup>Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy. <sup>2</sup>Present address: Department of Crop and Animal Sciences, Humboldt-Universität zu Berlin, Berlin, Germany. <sup>3</sup>These authors contributed equally: Marco Passamonti and Federico Plazzi. ✉e-mail: [federico.plazzi@unibo.it](mailto:federico.plazzi@unibo.it)

were never directly associated to nuclear target mRNAs. Conversely, the possible involvement of mitochondrial non-coding RNAs in nuclear regulation is an expanding research field<sup>42,43</sup>.

A new class of small mitochondrial highly-transcribed RNAs (smithRNAs) was recently described in the Manila clam *Ruditapes philippinarum* (Bivalvia: Veneridae) and associated to specific nuclear targets<sup>44</sup>. *R. philippinarum* shows a peculiar way of mitochondrial inheritance, called Doubly Uniparental Inheritance (DUI)<sup>33,45–47</sup>, which leads to the presence of two separate, sex-linked mitochondrial lineages, along with their highly divergent genomes (M-type for male, and F-type for female mitochondria). Putative smithRNAs were found in small RNA libraries of *R. philippinarum* gonads, mapping to both M- and F-type mitochondrial genomes, and they were analyzed *in silico*<sup>44</sup>. Although we used a very conservative pipeline for their validation, smithRNA functionality was never demonstrated *in vivo*.

Finally, many observations point to the fact that smithRNAs cannot be transcribed from mitochondrial pseudogenes (NUMTs): in the Manila clam, for instance, male-encoded smithRNAs are not expressed in females, which in fact lack male mitochondria, a pattern of expression that is expected for mitochondrially-encoded genes of DUI species<sup>44</sup>. Moreover, additional clues on the issue were recently provided<sup>48</sup>, further supporting the origin of smithRNAs from mitochondrial DNA.

In the present paper, we present several clues about the actual *in vivo* activity of smithRNAs in regulating nuclear gene expression in the Manila clam: loci associated to smithRNAs are indeed more conserved than unassigned regions of the mitochondrial genome, and *in vivo* interference effects were observed. Finally, we present preliminary evidence about the existence of smithRNA in distantly related eukaryotic lineages, irrespective of the biogenetic pathway, which appears to be variable or unclear.

## Results

**Conservation of smithRNA-encoding loci.** When smithRNAs were firstly described in the Manila clam *Ruditapes philippinarum*<sup>44</sup>, nothing was known about their biogenesis and way of regulating target genes. As a preliminary hypothesis, a miRNA-like biogenesis was assumed, so that it is possible that functional sncRNAs arising from different pathways were disregarded. Conversely, stringent thresholds were used for all computations, aiming to a robust description of at least a subset of existing smithRNAs. Nonetheless, at that point, bioinformatics was the only source of data for characterization and actual functionality of these smithRNAs.

In the present paper, we focused on gathering clues of functionality of these sncRNAs in regulating nuclear genes. Specifically, we devised three different lines of experimental evidence to support their functionality: (i) by sequence conservation, (ii) by *in vivo* observation of interference effects, and (iii) by distribution among metazoans.

A first rationale was as follows: many (putative) smithRNAs map to mitochondrial unassigned regions (URs)<sup>44</sup>, that are typically abundant in bivalve mtDNAs<sup>49,50</sup>. URs should experience relaxed selective pressure with respect to surrounding genomic regions, unless they are actually functional in some way: for instance, unless they do encode for a sncRNA.

To test the hypothesis of functionality of smithRNAs, we sequenced smithRNA-encoding loci and flanking regions in 15 female and 27 male specimens of *R. philippinarum* and partitioned sequenced regions into different functional regions: protein-coding genes (PCGs), rRNA-encoding genes (rRNAs), tRNA-encoding genes (tRNAs), regions predicted to transcribe for a pre-smithRNA, regions predicted to transcribe for a mature smithRNA, and (re-annotated) URs. Nucleotide diversity ( $\pi$ ) was computed along a sliding window on the resulting alignments (Fig. 1).

Levels of nucleotide diversity of the five types of coding regions were not significantly different; conversely, they were all significantly different from the level of nucleotide diversity measured within URs (Kruskal-Wallis rank sum test with 5 degrees of freedom,  $\chi^2 = 67.8566$ ,  $P = 0$ ; two-tailed pairwise Dunn's test with Bonferroni correction; Fig. 1). The only exception to this was the comparison between PCG and tRNA regions: however, the relatively high level of diversity detected within PCGs is largely due to synonymous mutations, which experience relaxed selective pressure (Supplementary Table S1).

In sum, the conservation of putative smithRNA loci is comparable to that of other well-known functional mitochondrial loci, and, even if some of them were originally annotated within URs, nucleotide diversity significantly drops for smithRNA-coding regions with respect to the remainder of URs (Fig. 1).

***In vivo* assay of M\_smithRNA106t and 145t.** The original *in silico* predictions allowed us to identify one or more nuclear targets for all the 14 putative smithRNAs<sup>44</sup>. The only putative target of M\_smithRNA106t, which is annotated between *cox3* and *atp6* on the male mtDNA (Fig. 1), is the clam homolog of a human Histone-lysine N-methyltransferase (SETD2; Uniprot entry Q9BYW2). Therefore, when M\_smithRNA106t is present, cells are expected to show reduced levels of methylation on histone H3<sup>51</sup>.

100  $\mu$ g of M\_smithRNA106t were injected into 61 clams at the concentration of 100 ng/ $\mu$ L; the same volume (1 mL) of ddH<sub>2</sub>O was injected into 60 control specimens. Levels of methylation were assayed 2 h and 24 h after injection, along with levels of methylation in 34 untreated specimens.

The injection itself entailed a significant reduction of methylation levels after 24 h (one-tailed Mann-Whitney test,  $W = 358.5$ ,  $P = 0.0383$ ); however, methylation levels are significantly lower in individuals injected with the smithRNA targeting a methyltransferase, when compared to control individuals injected with distilled water (one-tailed Mann-Whitney test,  $W = 362$ ,  $P = 0.033$ ; Fig. 2). The smithRNA-triggered effect seems especially evident in female specimens (Supplementary Fig. S1). Recall that it is not possible to *a priori* predict the sex of specimen, groups, including the untreated group, show some differences in size; however, the use of nonparametric tests accounts for the different dimension of resulting groups.

M\_smithRNA145t was associated with three different targets<sup>44</sup>; among them, a homolog of another human Histone-lysine N-methyltransferase (SETD8; Uniprot entry Q9NQR1), and the homolog of the human DNA

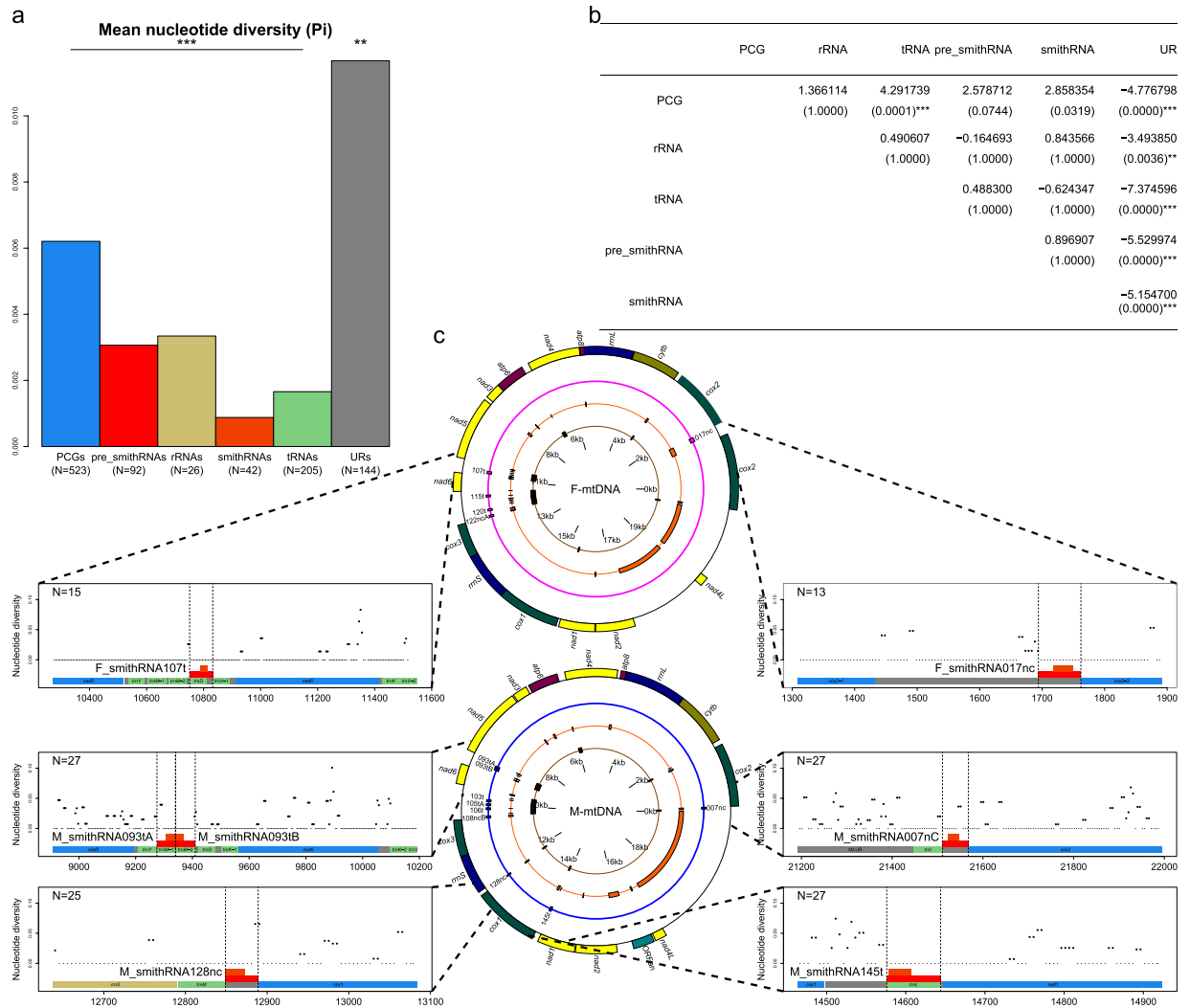


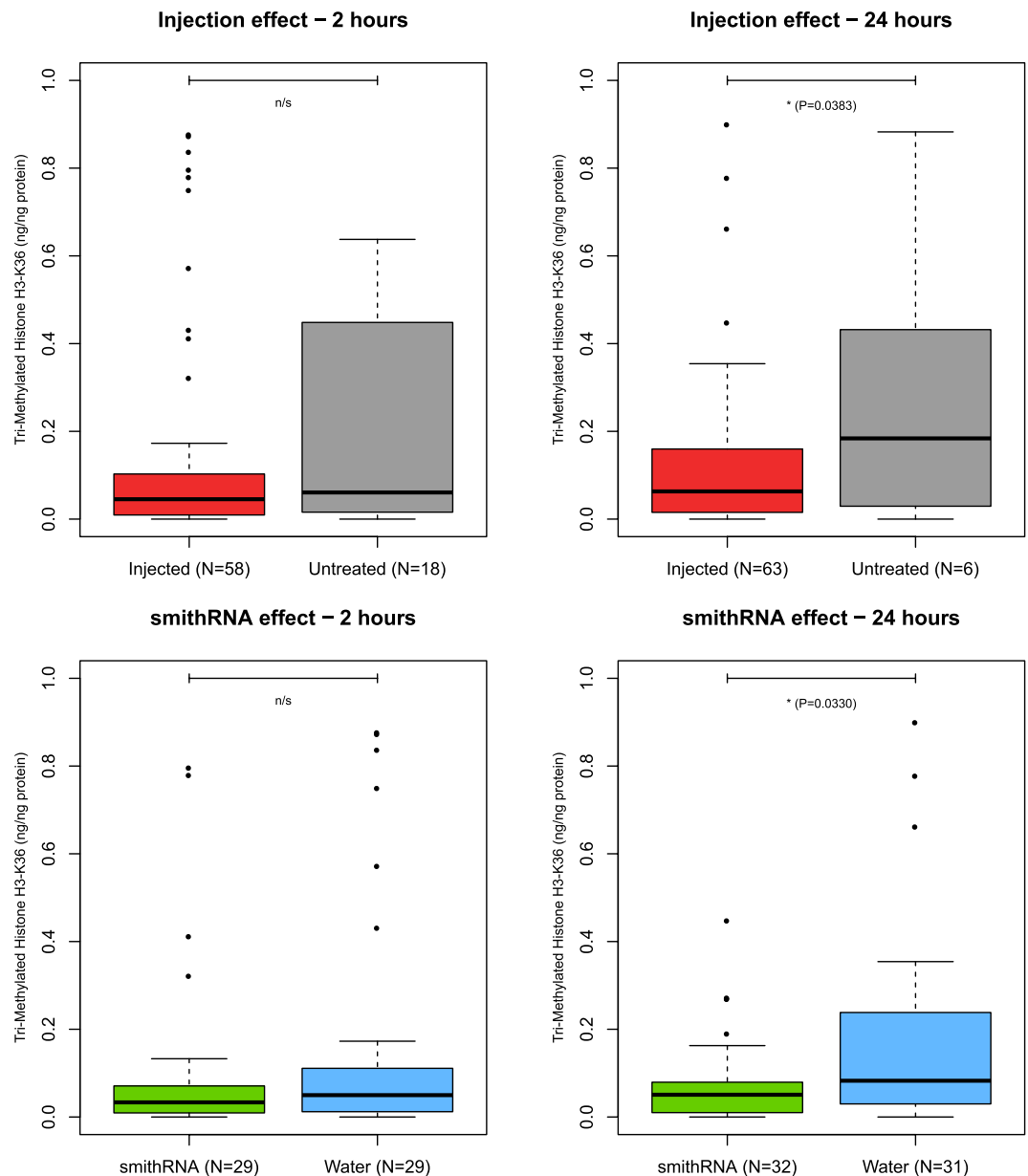
Figure 1

**Figure 1.** Conservation of Manila clam smithRNA loci. (a), mean nucleotide diversity ( $\pi$ ) for five type of coding regions, including putative pre\_smithRNAs and smithRNAs, and URs. (b), results of Dunn's test for the six types of regions; sample sizes are shown in (a). (c), nucleotide diversity over a sliding window in six genomic regions: colours as in a. Windows with nucleotide diversity equal to 0 are shown with small circles, while those with nucleotide diversity greater than 0 are shown with larger circles. Annotation of mtDNA is reported: from inner to outer, circles bear tRNAs, URs, smithRNAs, and PCGs.

polymerase epsilon subunit 3 (POLE3/CHRAC17; Uniprot entry Q9NRF9), which has been involved into the process of chromatin remodeling, and particularly to histone acetylation<sup>52</sup>. Again, when M\_smithRNA145t is present, cells are thus expected to show altered levels of histone acetylation.

As above, 100  $\mu$ g of M\_smithRNA145t were injected into 50 clams, while 41 control clams were injected with F\_smithRNA107t (which is not expected to affect chromatin remodelling<sup>44</sup>) and 53 clams were not treated. We decided to use a different ssRNA as control in this case to rule out the injection effect which was observed in the previous experiment: both groups were injected with a ssRNA, therefore differences in acetylation levels ought to be associated to the specific ssRNA itself.

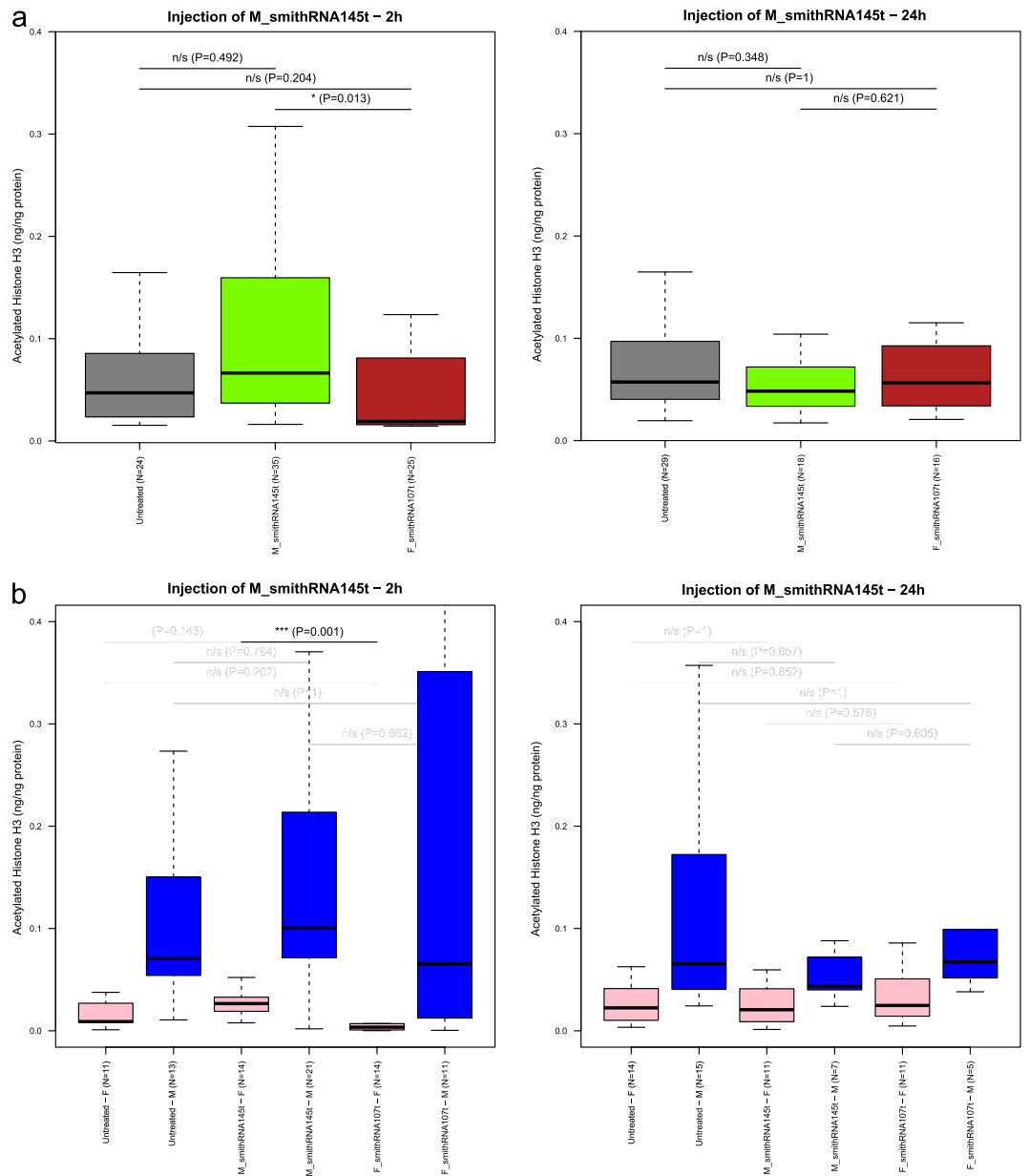
In this case, we observed a significant increase of histone H3 acetylation 2 hours after injection in specimens injected with M\_smithRNA145t with respect to control and untreated specimens (Kruskal-Wallis rank sum test with 2 degrees of freedom,  $\chi^2 = 6.8685$ ,  $P = 0.03$ ; two-tailed pairwise Dunn's test with Bonferroni correction); however, this effect became negligible after 24 h (Fig. 3). Again, when considering females and males separately, the significance holds for female specimens only (Kruskal-Wallis rank sum test with 2 degrees of freedom,  $\chi^2 = 11.3299$ ,  $P = 0$ ; two-tailed pairwise Dunn's test with Bonferroni correction; Supplementary Table S2). The high complexity of the involvement of POLE3/CHRAC17 in chromatin remodeling may well lead to the emergence of the observed upregulation effect by means of M\_smithRNA145t, instead of the expected silencing.



**Figure 2.** Effect of the injection of *M\_smithRNA106t* on Manila clams. Injected, individuals injected with either *M\_smithRNA106t* or ddH<sub>2</sub>O; untreated, uninjected individuals. The black line is the median; the two hinges of the box approximate the first and the third quartile; whiskers extend to a roughly 95% confidence interval.

**Distribution of smithRNAs among animals.** Wide-scale characterization of smithRNAs in different metazoan clades are beyond the aim of the present paper, which mainly aims to demonstrate the *in vivo* functionality of these sncRNAs. However, smithRNAs are not ‘weird RNA species’ of (at most) some clam species. We applied the same pipeline of the original study<sup>44</sup> to other metazoan species in order to conservatively recover good smithRNA candidates from other systems. As smithRNAs were originally described from gonad tissues (given the peculiarities of the DUI phenomenon), putative smithRNAs were also identified from gonad samples.

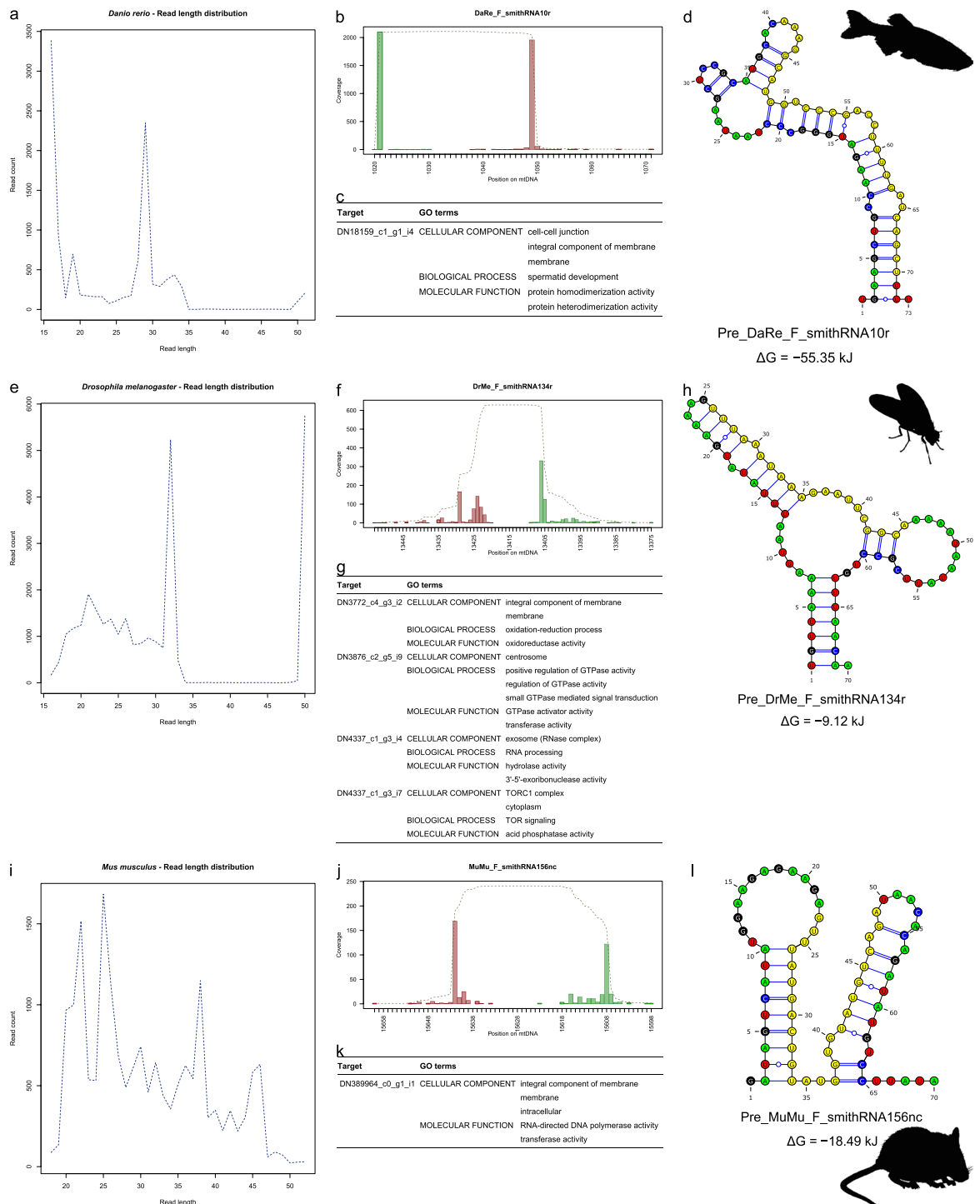
The fruitfly, *Drosophila melanogaster*, belongs to Ecdysozoa, while *R. philippinarum* belongs to Lophotrochozoa. We detected 2 putative smithRNAs from *Dr. melanogaster* ovaries and 8 putative smithRNAs from *Dr. melanogaster* testes (Fig. 4a–d). The zebrafish, *Danio rerio*, and the mouse, *Mus musculus*, are widely used model species as well; contrastingly with the two aforementioned protostome species, they belong to Deuterostomia. We detected 3 putative smithRNAs from *Da. rerio* ovaries and 3 putative smithRNAs from *Da. rerio* testes (Fig. 4e–h). Contrastingly, only one putative smithRNA was detected in *M. musculus* ovaries; however, when relaxing the read cluster size threshold from 200 to 50, we retrieved four more putative smithRNA in mouse ovaries and one putative smithRNA in mouse testes (Fig. 4i–l).



**Figure 3.** Effect of the injection of *F\_smithRNA107t* and *M\_smithRNA145t* on Manila clams. Untreated, uninjected individuals. (a), two sexes together; (b), injection effect shown by sex. Only significant pairwise Dunn's test comparisons are shown for the sake of clarity; for complete results, see Supplementary Table S2. The black line is the median; the two hinges of the box approximate the first and the third quartile; whiskers extend to a roughly 95% confidence interval.

Complete details on the annotation of putative smRNAs in *Dr. melanogaster*, *Da. rerio*, and *M. musculus* are given in Supplementary Dataset S1, Supplementary Dataset S2, and Supplementary Table S3; here we provide an example from each model species. *F\_smithRNA134r* from *Dr. melanogaster* is predicted to target four transcripts, involved in signal transduction and RNA processing. *F\_smithRNA10r* from *Da. rerio* has a single predicted target, connected to cell junction. *F\_smithRNA156nc* from *M. musculus* has a single predicted target as well, possibly an integral component of membrane.

Putative smRNAs from *R. philippinarum* and these three model species were queried against three different sncRNA databases: miRbase<sup>53–58</sup>, piRbase<sup>59–62</sup>, and fRNAdb<sup>63,64</sup>. Interestingly, all the smRNAs identified in the mouse have been described as piRNAs elsewhere, and four out of six do map on the mouse mitochondrial genome (Supplementary Table S4). Conversely, only five out of ten fly smRNAs were found in piRbase, again mapping on the mitochondrial genome. All *Drosophila* smRNAs have similarities with fly small RNAs in fRNAdb, though sometimes with high e-values. Finally, no significant similarities were detected for zebrafish and Manila clam smRNAs in these databases, with the exception of low-similarity hits for three out of six zebrafish



**Figure 4.** Three putative smithRNAs from *Danio rerio* (a–d), *Drosophila melanogaster* (e–h), and *Mus musculus* (i–l). (a,e,i): read length distribution for clusters with size greater than 200, whose centroid were retained for downstream analyses; (b,f,j): total and 5'/3' coverage on the mitochondrial genome. 5' coverage is shown in green and 3' coverage is shown in red; note that two putative smithRNAs would encode on the minus strand, hence the inversion of the horizontal axis (f,j). (c,g,k): top scoring GO terms associated with predicted targets; for complete ARGOT results, see Supplementary Table S3. (d,h,l): secondary structure of the genomic contexts. Mature putative smithRNAs are shown in yellow.

smithRNAs that were associated to ape/plant miRNAs and a mouse mitochondrial piRNA, in all cases with large e-values (Supplementary Table S4).

We carried out the same analysis of nucleotide diversity that was carried out for *R. philippinarum* on available complete mitochondrial genomes of *Dr. melanogaster* and *M. musculus* (Supplementary Table S5). Again, the

amount of nucleotide diversity in unassigned regions is significantly higher than in any other coding region of the genome, including regions where putative smithRNAs and pre-smithRNAs were mapped; moreover, with minor exceptions, the amount of nucleotide diversity is comparable among all coding regions (Kruskal-Wallis rank sum test with 5 degrees of freedom,  $\chi^2 = 303.0833$  and  $192.0627$  for fruit fly and mouse, respectively,  $P = 0$ ; two-tailed pairwise Dunn's test with Bonferroni correction; Supplementary Fig. S2).

## Discussion

Our analysis evidenced that smithRNA loci on both mitochondrial genomes of *Ruditapes philippinarum* show variability levels similar to coding loci. The conservation to an extent comparable with other coding regions is an evidence of selective constraints acting on these regions, which in turn constitutes a strong clue of functionality for these loci.

Furthermore, both injection experiments provided the first *in vivo* evidence of regulation of nuclear genes through small RNAs encoded by the mitochondrion; moreover, since the targets are broad-range epigenetic effectors, these smithRNAs may have a major impact on nuclear genome regulation. Finally, we showed that smithRNAs are present and can be annotated in representatives of the three main bilaterian lineages.

The characterization of molecular details of sncRNA biogenesis and silencing mechanism, as well as their scope on different organisms, is a growing research field. There is virtually no physiological process that has not been somehow linked to these molecules, including tumorigenesis, immune and stress response, reproduction, and development<sup>3,4,6,65–69</sup>.

Besides clams<sup>44</sup>, small RNAs coded by the mitochondrion (mitosRNAs) were recently described in plants<sup>70</sup> and many metazoans, including humans<sup>36,39</sup>, mice<sup>39,40</sup>, chickens<sup>71</sup>, sharks, teleost fishes, turtles, and frogs<sup>41</sup>. Different roles have been proposed for these mitochondrially-encoded RNAs, ranging from anoxia tolerance<sup>41</sup> to gametogenesis and fertilization<sup>40</sup>, however, to our knowledge, the only targets that were explicitly linked to mitochondrial sncRNAs are mitochondrial PCGs themselves<sup>36,39,71</sup>. Conversely, the ability of mitochondria to shape nuclear gene expression has already been noted<sup>72,73</sup>. Thus, smithRNAs may be the tile to complete this puzzle about retrograde (i.e., mitochondrion-to-nucleus) signalling<sup>74–76</sup>.

Mitochondrially-encoded smithRNAs must exit the mitochondrion to regulate nuclear transcripts in the cytoplasm, if not enter the nucleus itself. In the present paper we do not provide direct evidence of the presence of smithRNA molecules outside the mitochondrion; however, mitochondrial RNA outside the source organelle has been already observed. Several mitochondrially-encoded tRNAs were found in the cytoplasm of human cells<sup>77</sup>, in one case associated to Ago2, the protein involved in the final cleavage of many sncRNA biogenetic pathways. A release of mitochondrial material has been observed in *R. philippinarum*<sup>78</sup>, and this would provide a very obvious mechanism for smithRNAs to enter cytoplasm. Long non-coding RNAs transcribed by the mitochondrial genome were reported in the nucleus<sup>79–81</sup>, and therefore the presence of short non-coding ones would not be surprising.

Here we provide the first experimental data on smithRNA functionality on specific nuclear targets, demonstrating that in eukaryotic cells small mitochondrial transcripts do exit the mitochondrion to exert regulatory roles. Moreover, targets that are investigated here have a potential genome-scale effect, being epigenetic regulators: thus, not only can mitochondria affect nuclear expression, but also their scope is possibly massive.

Some putative smithRNAs were already described as piRNAs (all the putative smithRNAs we detected in *Mus musculus* and half of the putative smithRNAs we detected in *Drosophila melanogaster*). Previous reports showed that most mitosRNAs are indeed piRNAs in *M. musculus*<sup>40</sup>, and piRNAs are a class of sncRNAs which is typically associated to germline cells<sup>13,82,83</sup>, a fact which was recently confirmed in mollusks<sup>84</sup>. Moreover, the PIWI biogenetic pathway, involved in piRNA biogenesis<sup>85,86</sup> is associated to mitochondria<sup>87</sup> and piRNAs were already connected to mitochondrial communications to the nucleus<sup>79</sup>. Consistently with our mapping, piRNAs were previously found to originate from mitochondrial tRNAs and rRNAs<sup>83</sup>.

However, many other putative smithRNAs have unclear affinities with already described sncRNAs (half of the putative smithRNAs we detected in *Dr. melanogaster*) or no affinity at all (*Danio rerio*; *R. philippinarum*; Supplementary Table S4). It is tempting to conclude that smithRNAs transcribed by the mitochondrial genome were exapted from many different biogenetic pathways, most notably from that typical of piRNAs. Similarly, they may act with different mechanisms, either silencing or enhancing nuclear genes, as was observed in our *in vivo* assays.

These recently discovered tools may thus be heterogeneous in terms of biogenesis and activity, yet they share a functional role: a retrograde signaling directed towards nuclear genes. Moreover, given the available, stringent bioinformatic pipeline, they normally appear to be highly expressed. In fact, mitochondrial genomes already have some features (including tRNAs and secondary structures in intergenic regions) that are likely to exapt towards the evolution of regulatory RNAs. Put in other words, we regard to smithRNAs as a class of mitosRNAs which by definition affect nuclear gene expression to some extent.

These small genetic elements provide an effective way for mitochondria to largely influence nuclear functioning: while more smithRNAs should be experimentally validated in the future, it is now important to achieve a clearer picture of the distribution and abundance of these RNAs among living beings. Our present findings are necessary conservative, in that miRNA-like biogenesis and action mechanism were assumed, and it is likely that many smithRNAs were overlooked. Yet, our results highlight the presence of smithRNAs in the three major lineages of bilaterians (Ecdysozoa, Lophotrochozoa, and Deuterostomia), suggesting that the emergence of these regulatory elements is (at least) as ancient as bilaterians and that they represent a widespread mechanism for mitochondria to influence nuclear expression.

## Methods

**Conservation of smithRNA-encoding loci.** DNA from gametes of 15 female and 27 male *Ruditapes philippinarum* clams was collected and extracted as in<sup>50</sup>. PCR amplifications of mitochondrial regions flanking putative smithRNAs were carried out on a 2720 Thermal Cycler (Applied Biosystem) with GoTaq Flexi DNA Polymerase (Promega), as follows: 10  $\mu$ L 5  $\times$  Green GoTaq Flexi Buffer, MgCl<sub>2</sub> (3 mM), nucleotides (800  $\mu$ M each), primers (500 nM each), 1 U GoTaq DNA Polymerase, 40 ng template DNA, ddH<sub>2</sub>O up to 50  $\mu$ L. Primers were designed using Primer3<sup>88–90</sup> and are listed in Supplementary Table S6. PCR cycles were set following manufacturer's instructions, as follows: initial denaturation at 95 °C for 2'; 35 cycles of denaturation at 95 °C for 1', annealing at 48–56 °C for 1', and extension at 72 °C for 1'; final extension at 72 °C for 5'. PCR results were visualized using a 1% electrophoresis agarose gel stained with ethidium bromide. Amplicons were purified through a standard isopropanol protocol and sequenced at the MacroGen Europe facilities. Electropherograms were edited using MEGA7<sup>91</sup>. Mean nucleotide diversity is defined as

$$\hat{\pi} = \frac{N}{N-1} \sum_{i=1}^N \sum_{j>i}^N p_i p_j \pi_{ij}$$

where  $N$  is the number of sequences,  $p_i$  is the frequency of the  $i$ -th haplotype,  $p_j$  is the frequency of the  $j$ -th haplotype and  $\pi_{ij}$  is the uncorrected (p) distance between haplotypes  $i$  and  $j$ . Mean nucleotide diversity was computed over a 10-bp sliding window with 5-bp steps using the software Variscan<sup>92</sup>; general data organization, the Kruskal–Wallis rank sum test, and the two-tailed pairwise Dunn's test were carried out using the software R<sup>93</sup> and the package dunn.test<sup>94</sup>. Mitochondrial DNAs were drawn using the software GenomeVx<sup>95</sup>.

**In vivo assay of M\_smithRNA106t and 145t.** *R. philippinarum* specimens were sampled in Italy (Goro) in June/July 2017 and 2018 during two sampling campaigns. Individuals were collected during the reproductive season. After sampling, clams were placed in different beakers containing reverse osmosis water with Red Sea Coral Pro aquariology sea salt (Red Sea Europe).

1,500 nmol of custom ssRNAs corresponding to M\_smithRNA106t, M\_smithRNA145t, and F\_smithRNA107 were synthesized at the Integrated DNA Technologies, Inc. facilities. ssRNAs were resuspended in ddH<sub>2</sub>O to a final concentration of 100 ng/ $\mu$ L. 1 mL of ssRNA or ddH<sub>2</sub>O was injected into clams by slightly widening valves' opening and inserting the sterile 2.5-mL syringe needle into the mature gonad, immediately above the mollusk's foot. Two or twenty-four hours after the injection, specimens were collected and sexed by microscopic dissection of gonadal tissue. Gonadal tissues were then sampled and stored at –80 °C for histone extraction.

Total histone extraction and tri-methylation quantification were carried out using the EpiQuik Total Histone Extraction Kit and the EpiQuik Global Tri-Methyl Histone H3-K36 Quantification Kit (Colorimetric) (EpiGentek); acetylation quantification was carried out using the EpiQuik Global Histone H3 Acetylation Assay Kit (EpiGentek), following manufacturer's instruction. Protein concentration was quantified using the Quick Start Bradford Protein Assay (BioRad) on a NanoGenius Photometer Onda. Colorimetric assays were quantified on a Benchmark Microplate Reader (Bio-Rad). General data organization, the Mann–Whitney test, the Kruskal–Wallis rank sum test, and the two-tailed pairwise Dunn's test were carried out using the software R and the package dunn.test.

**Distribution of smithRNAs among animals.** The pipeline of Pozzi and colleagues<sup>44</sup> was strictly followed in order to obtain comparable results from *Danio rerio*, *Drosophila melanogaster*, and *Mus musculus*. An overview of the bioinformatic pipeline is provided as Supplementary Fig. S3.

Raw reads were downloaded from the GenBank repository: SRA Accession Numbers are listed in Supplementary Table S7. Reads were filtered with Trimmomatic 0.38<sup>96</sup> and potential contaminations were discarded using kraken2 2.0.8<sup>97</sup>. Settings of the filtering stage were ILLUMINACLIP:2:30:10 AVGQUAL:20 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 for smallRNA-Seq and ILLUMINACLIP:2:30:10 AVGQUAL:20 LEADING:3 TRAILING:3 SLIDINGWINDOW:25:33 MINLEN:75 (if suitable) for RNA-Seq; the database used for the software kraken2 was the standard database built on the 20<sup>th</sup> of June, 2019. Transcriptomes were assembled using the software Trinity 2.6.6<sup>98</sup> with the `-no_normalize_reads` option. The completeness of transcriptomes was assessed using the online tool gVolante<sup>99</sup> against the relevant database. ORFs and relative 3' UTRs were identified using the software ExUTR 0.1.0<sup>100</sup> against the Swissprot database with options `-x 2500 -m 1`.

Small transcriptomes' reads were mapped against the *Da. rerio*, *Dr. melanogaster*, and *M. musculus* mitochondrial genomes (GenBank Accession Numbers NC\_002333.2, KJ947872.2, and AY172335.1, respectively) using the software Bowtie2 2.3.0<sup>101</sup> in the end-to-end mode. The maximum number of tolerated mismatches was set to 1 using the options `-N 1 -i C,1 -L 18`. Reads mapping to the mitochondrial genome were mapped against the latest release of the nuclear genome (GenBank Accession Numbers GCA\_000002035.4, GCA\_000001215.4, and GCA\_000001635.8, respectively), again using end-to-end Bowtie2 alignments (options: `-i C,1 -L 22`). Reads not mapping on the nuclear genome were mapped again on the mitochondrial genome (as before) and analyzed further. Clusters of reads were created using USEARCH 11.0.667<sup>102</sup> setting the identity score to 0.99 (which, for short smallRNA-Seq reads, means total identity).

According to our original publication<sup>44</sup>, multiple requirements were mandatory for a centroid of a cluster to be considered a putative smithRNA: (1) a cluster size greater than 200; (2) sharp 5' and 3' coverage measured with bedtools 2.26.0<sup>103</sup>, meaning that the sncRNA has strongly preferred starting and ending sites for transcription; (3) perfect matching of nucleotides 4–10 of the centroid with a 3' UTR from the transcriptome, accounting for the phenomenon of seed shift (see<sup>44</sup> for details); (4) at least 11 matches between the centroid and a 3' UTR, taking advantage of the EMBOSS 6.6.0 suite<sup>104</sup> and following alignment scores computed by BLAST + 2.6.0<sup>105</sup> using options `-task blastn-short -strand minus`; (5) a  $\Delta\Delta G$  score lower than –9 kJ for the centroid-target UTR



interaction, as computed by the software PITA<sup>6106</sup> considering 3 and 15 nucleotides as 5' and 3' flanking sites, respectively; (6) a Gibbs free energy score lower than  $-20$  kJ for the centroid-mRNA duplex, as computed by RNAhybrid 2.1.2<sup>107</sup> (options: -f 3,10 -e -20 -p 0.05 -s 3utr\_fly -t).

Centroids meeting all these requirements were considered *bona fide* putative smithRNAs and the secondary structure of the genomic context was computed using RNAfold from the ViennaRNA 2.4.13 package<sup>108</sup>, setting the folding temperature to 25 °C. Most putative smithRNAs fell into larger ribosomal genes, and the pre-smithRNA regions were arbitrarily set to the 70-bp long region centered on the putative smithRNA sequence, which gives an approximation of secondary structures in that genomic context. The only exception to this is DaRe\_F\_smithRNA10r: mapping at the very beginning of a ribosomal gene, we included the 3' region of the upstream tRNA gene, in order to have the putative smithRNA approximately in the middle of the considered sequence. For those putative smithRNAs mapping to a tRNA gene, the whole tRNA region, along with flanking unassigned nucleotides, was used as pre-smithRNA. Secondary structures were drawn using the software VARNA<sup>109</sup>.

Putative smithRNAs were named according to our previous publication<sup>44</sup>: after the species abbreviation ('DaRe', 'DrMe', or 'MuMu'), the first letter (either 'F' or 'M') denotes the sex of the individual (and, thus, the gonad), while numbers refer to the 100-based position of the smithRNA on the mtDNA and 'r', 't', or 'nc' stands for 'mapping to a ribosomal gene', 'mapping to a tRNA gene', or 'mapping to a non-coding region', respectively. GO terms were associated to predicted targets of putative smithRNAs using ARGOT2<sup>110-112</sup>.

To investigate the variability of loci coding for putative smithRNAs in model species, 43 complete mitochondrial genomes of *D. melanogaster* and 162 complete mitochondrial genomes of *M. musculus* were downloaded from GenBank using CLC Main Workbench (QIAGEN), aligned by region using Muscle<sup>113</sup>, and concatenated. *Da. rerio* was excluded because of the paucity of available annotated complete mitochondrial genomes. As above, we used Variscan to compute nucleotide diversity along a sliding window and the R environment for general data organization, the Kruskal-Wallis rank sum test, and the two-tailed pairwise Dunn's test.

## Data availability

The *Ruditapes philippinarum* sequences generated and analysed during the current study are available in the DDBJ/ENA/GenBank repository under the Accession Numbers: MN814873-MN815006. Putative *Drosophila melanogaster*, *Danio rerio*, and *Mus musculus* smithRNA sequences are available in the Third Party Annotation Section of the DDBJ/ENA/GenBank databases under the accession numbers TPA: BK010906-15, BK011029-34, and BK011035-40, respectively. All custom-tailored scripts used for data analysis are available from the corresponding author on reasonable request.

Received: 4 December 2019; Accepted: 22 April 2020;

Published online: 19 May 2020

## References

- Brouwer, I. & Lenstra, T. L. Visualizing transcription: key to understanding gene expression dynamics. *Curr. Opin. Chem. Biol.* **51**, 122–129 (2019).
- Buchberger, E., Reis, M., Lu, T.-H. & Posnien, N. Cloudy with a chance of insights: context dependent gene regulation and implications for evolutionary studies. *Genes* **10**, 492 (2019).
- Kim, S. S. & Lee, S.-J. V. Non-coding RNAs in *Caenorhabditis elegans* aging. *Mol. Cells* **42**, 379–385 (2019).
- Larriba, E. & del Mazo, J. Role of non-coding RNAs in the transgenerational epigenetic transmission of the effects of reprotoxicants. *Int. J. Mol. Sci.* **17**, 452 (2016).
- Cech, T. R. & Steitz, J. A. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
- Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**, 94–108 (2009).
- Fromm, B. *et al.* A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. *Annu. Rev. Genet.* **49**, 213–242 (2015).
- Wessels, H. H. *et al.* Global identification of functional microRNA-mRNA interactions in *Drosophila*. *Nat Commun.* **10**, 1626 (2019).
- O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Front. Endocrinol.* **9**, 402 (2018).
- Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014).
- García-López, J., Briño-Enríquez, M. A. & del Mazo, J. MicroRNA biogenesis and variability. *BioMol Concepts* **4**, 367–380 (2013).
- Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D. & Zamore, P. D. PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* **20**, 89–108 (2019).
- Weick, E. M. & Miska, E. A. piRNAs: from biogenesis to function. *Development* **141**, 3458–3471 (2014).
- Bamezai, S., Rawat, V. P. & Buske, C. Concise review: The Piwi-piRNA axis: pivotal beyond transposon silencing. *Stem Cells* **30**, 2603–2611 (2012).
- Cloonan, N. Re-thinking miRNA-mRNA interactions: intertwining issues confound target discovery. *Bioessays* **37**, 379–388 (2015).
- Bartel, D. P. Metazoan microRNAs. *Cell* **173**, 20–50 (2018).
- Lee, Y. *et al.* The role of PACT in the RNA silencing pathway. *EMBO J.* **25**, 522–532 (2006).
- Chendrimada, T. P. *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740–744 (2005).
- Förstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3**, e236 (2005).
- Jiang, F. *et al.* Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev.* **19**, 1674–1679 (2005).
- Saito, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLoS Biol.* **3**, e235 (2005).
- Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
- Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).

24. Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293**, 834–838 (2001).
25. Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
26. Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. & Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231–235 (2004).
27. Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240 (2004).
28. Han, J. *et al.* The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016–3027 (2004).
29. Landthaler, M., Yalcin, A. & Tuschl, T. The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr Biol.* **14**, 2162–2167 (2004).
30. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
31. Lee, Y., Jeon, K., Lee, J. T., Kim, S. & Kim, V. N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* **21**, 4663–4670 (2002).
32. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
33. Breton, S. *et al.* A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. *Trends Genet.* **30**, 555–564 (2014).
34. Gissi, C. *et al.* Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* **101**, 301–320 (2008).
35. Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**, 1767–1780 (1999).
36. Mercer, T. R. *et al.* The human mitochondrial transcriptome. *Cell* **146**, 645–658 (2011).
37. Hausacker, D. *et al.* Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* **16**, 673–695 (2010).
38. Lee, Y. S., Shibata, Y., Malhotra, A. & Dutta, A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* **23**, 2639–2649 (2009).
39. Ro, S. *et al.* The mitochondrial genome encodes abundant small noncoding RNAs. *Cell Res.* **23**, 759–774 (2013).
40. Larriba, E., Rial, E. & del Mazo, J. The landscape of mitochondrial small non-coding RNAs in the PGCs of male mice, spermatogonia, gametes and in zygotes. *BMC Genomics* **19**, 634 (2018).
41. Riggs, C. L. *et al.* Small non-coding RNA expression and vertebrate anoxia tolerance. *Front. Genet.* **9**, 230 (2018).
42. Hill, G. E. *Mitochondrial Ecology* (Oxford University Press, 2019).
43. Dong, Y., Yoshitomi, T., Hu, J.-F. & Cui, J. Long noncoding RNAs coordinate functions between mitochondria and the nucleus. *Epigenet. Chromatin* **10**, 41 (2017).
44. Pozzi, A., Plazzi, F., Milani, L., Ghiselli, F. & Passamonti, M. SmithRNAs: could mitochondria “bend” nuclear regulation? *Mol. Biol. Evol.* **34**, 1960–1973 (2017).
45. Passamonti, M. & Plazzi, F. Doubly Uniparental Inheritance and beyond: The contribution of the Manila clam *Ruditapes philippinarum*. *J. Zool. Syst. Evol. Res.*, <https://doi.org/10.1111/jzs.12371> (2020).
46. Gusman, A., Lecomte, S., Stewart, D. T., Passamonti, M. & Breton, S. Pursuing the quest for better understanding the taxonomic distribution of the system of doubly uniparental inheritance of mtDNA. *PeerJ* **4**, e2760 (2016).
47. Zouros, E. Biparental Inheritance Through Uniparental Transmission: The Doubly Uniparental Inheritance (DUI) of Mitochondrial DNA. *Evol. Biol.* **40**, 1–31 (2013).
48. Pozzi, A. & Dowling, D. K. 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 (2019).
49. Plazzi, F., Puccio, G. & Passamonti, M. Comparative large-scale mitogenomics evidences clade-specific evolutionary trends in mitochondrial DNAs of Bivalvia. *Genome Biol. Evol.* **8**, 2544–2564 (2016).
50. Ghiselli, F. *et al.* Structure, transcription, and variability of metazoan mitochondrial genome: perspectives from an unusual mitochondrial inheritance system. *Genome Biol. Evol.* **5**, 1535–1554 (2013).
51. Yang, S. *et al.* Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase. *Genes Dev.* **30**, 1611–1616 (2016).
52. Wang, Y. L., Faiola, F., Xu, M., Pan, S. & Martinez, E. Human ATAC is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. *J. Biol. Chem.* **283**, 33808–33815 (2008).
53. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162 (2019).
54. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68–D73 (2014).
55. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–D157 (2011).
56. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158 (2008).
57. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34**, D140–D144 (2006).
58. Griffiths-Jones, S. The microRNA Registry. *Nucleic Acids Res.* **32**, D109–D111 (2004).
59. Wang, J. *et al.* piRBase: a comprehensive database of piRNA sequences. *Nucleic Acids Res.* **47**, D175–D180 (2019).
60. Yuan, J. *et al.* Computational identification of piRNA targets on mouse mRNAs. *Bioinformatics* **32**, 1170–1177 (2016).
61. Zhang, P. *et al.* MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. *Cell Res.* **25**, 193–207 (2015).
62. Zhang, P. *et al.* piRBase: a web resource assisting piRNA functional study. *Database (Oxford)* **2014**, bau110 (2014).
63. Mityama, T. *et al.* The Functional RNA Database 3.0: databases to support mining and annotation of functional RNAs. *Nucleic Acids Res.* **37**(Database issue), D89–D92 (2009).
64. Kin, T. *et al.* fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. *Nucleic Acids Res.* **35**(Database issue), D145–D148 (2007).
65. Fernandes, J. C. R., Acuna, S. M., Aoki, J. I., Floeter-Winter, L. M. & Muxel, S. M. Long non-coding RNAs in the regulation of gene expression: physiology and disease. *Noncoding RNA* **5**, 17 (2019).
66. Oberbauer, V. & Schaefer, M. R. tRNA-derived small RNAs: biogenesis, modification, function and potential impact on human disease development. *Genes* **9**, 607 (2018).
67. Shin, H., Kim, Y., Kim, M. & Lee, Y. BC200 RNA: an emerging therapeutic target and diagnostic marker for human cancer. *Mol. Cells* **41**, 993–999 (2018).
68. Szczepanek, J., Pareek, C. S. & Tretyn, A. The role of microRNAs in animal physiology and pathology. *Transl. Res. Vet. Sci.* **1**, 13–33 (2018).
69. Wang, M. *et al.* Non-coding RNAs function as immune regulators in teleost fish. *Front. Immunol.* **9**, 2801 (2018).
70. Wu, Z., Stone, J. D., Storchová, H. & Sloan, D. B. high transcript abundance, RNA editing, and small RNAs in intergenic regions within the massive mitochondrial genome of the angiosperm *Silene noctiflora*. *BMC Genomics* **16**, 938 (2015).
71. Bottje, W. G. *et al.* Identification and differential abundance of mitochondrial genome encoding small RNAs (mitosRNA) in breast muscles of modern broilers and unselected chicken breed. *Front. Physiol.* **8**, 816 (2017).
72. Baris, T. Z. *et al.* Evolved genetic and phenotypic differences due to mitochondrial-nuclear interactions. *PLoS Genetics* **13**, e1006517 (2017).

73. Innocenti, P., Morrow, E. H. & Dowling, D. K. Experimental evidence supports a sex-specific selective sieve in mitochondrial genome evolution. *Science* **332**, 845–848 (2011).
74. Arnould, T., Michel, S. & Renard, P. Mitochondria retrograde signaling and the UPR<sup>mt</sup>: where are we in mammals? *Int. J. Mol. Sci.* **16**, 18224–18251 (2015).
75. Cagina, U. & Enriqueza, J. A. The complex crosstalk between mitochondria and the nucleus: What goes in between? *Int. J. Biochem. Cell Biol.* **63**, 10–15 (2015).
76. Monaghan, R. M. & Whitmarsh, A. J. Mitochondrial proteins moonlighting in the nucleus. *Trends Biochem. Sci.* **40**, 728–735 (2015).
77. Maniatakis, E. & Mourelatos, Z. Human mitochondrial tRNAMet is exported to the cytoplasm and associates with the Argonaute 2 protein. *RNA* **11**, 849–852 (2005).
78. Milani, L., Ghiselli, F., Maurizii, M. G. & Passamonti, M. Doubly uniparental inheritance of mitochondria as a model system for studying germ line formation. *PLoS One* **6**, e28194 (2011).
79. Vendramin, R., Marine, J. C. & Leucci, E. Non-coding RNAs: the dark side of nuclear-mitochondrial communication. *EMBO J.* **36**, 1123–1133 (2017).
80. Landerer, E. *et al.* Nuclear localization of the mitochondrial ncRNAs in normal and cancer cells. *Cell. Oncol. (Dordr.)* **34**, 297–305 (2011).
81. Rackham, O. *et al.* Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. *RNA* **17**, 2085–2093 (2011).
82. Ku, H.-Y. & Lin, H. PIWI proteins and their interactors in piRNA biogenesis, germline development and gene expression. *Natl. Sci. Rev.* **1**, 205–218 (2014).
83. Kwon, C. *et al.* Detection of PIWI and piRNAs in the mitochondria of mammalian cancer cells. *Biochem. Biophys. Res. Commun.* **446**, 218–223 (2014).
84. Jehn, J. *et al.* PIWI genes and piRNAs are ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation. *Comm. Biol.* **1**, 137 (2018).
85. Le Thomas, A., Toth, K. F. & Aravin, A. A. To be or not to be a piRNA: genomic origin and processing of piRNAs. *Genome Biol.* **15**, 204 (2014).
86. Siomi, M. C. *et al.* PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12**, 246–258 (2011).
87. Aravin, A. A. & Chan, D. C. piRNAs meet mitochondria. *Dev. Cell.* **20**, 287–288 (2011).
88. Köressaar, T. *et al.* Primer3\_masker: integrating masking of template sequence with primer design software. *Bioinformatics* **34**, 1937–1938 (2018).
89. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
90. Köressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291 (2007).
91. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
92. Vilella, A. J., Blanco-Garcia, A., Hutter, S. & Rozas, J. VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics* **21**, 2791–2793 (2005).
93. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org> (2019).
94. Dinno, A. *dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums*. R package version 1.3.5. Available at <https://CRAN.R-project.org/package=dunn.test> (2017).
95. Conant, G. C. & Wolfe, K. H. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* **24**, 861–862 (2008).
96. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* **30**, 2114–2120 (2014).
97. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
98. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
99. Nishimura, O., Hara, Y. & Kuraku, S. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics* **33**, 3635–3637 (2017).
100. Huang, Z. & Teeling, E. C. ExUTR: a novel pipeline for large-scale prediction of 3'-UTR sequences from NGS data. *BMC Genomics* **18**, 847 (2017).
101. Langmead, B. & Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
102. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
103. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
104. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
105. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2008).
106. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nature Genet.* **39**, 1278–1284 (2007).
107. Krüger, J. & Rehmsmeier, M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* **34**, W451–W454 (2006).
108. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
109. Darty, K., Denise, A. & Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).
110. Radivojac, P. *et al.* A large-scale evaluation of computational protein function prediction. *Nat. Methods* **10**, 221–227 (2013).
111. Falda, M. *et al.* Argot2: a large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. *BMC Bioinformatics* **13**, S14 (2012).
112. Fontana, P., Cestaro, A., Velasco, R., Formentin, E. & Toppo, S. Rapid annotation of anonymous sequences from genome projects using semantic similarities and a weighting scheme in gene ontology. *PLoS One* **4**, e4619 (2009).
113. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

## Acknowledgements

We wish to acknowledge Edoardo Turolla (Istituto Delta, Ferrara, Italy) for providing the specimens for this analysis. Authors are grateful to Beatrice Baldi, Giada Bolognesi, Alberto Di Giovanni and Andrea Piccinini for wetlab assistance. This work was supported by Ministero dell'Istruzione dell'Università e della Ricerca (MIUR) RFO grant and the Canziani Bequest, both funded to M. P. Thanks are also due to three anonymous reviewers, whose comments and criticism greatly improved a previous version of the manuscript.

### Author contributions

M.P. conceived the study and supervised all stages of the research. M.C., M.D. and F.P. performed wetlab experiments. F.P. analysed data about sequence conservation and *in vivo* functionality and carried out the annotation of putative smithRNAs in *Danio rerio*, *Drosophila melanogaster*, and *Mus musculus*. M.P. and F.P. drafted the manuscript. All authors read and approved the final manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-65084-z>.

**Correspondence** and requests for materials should be addressed to F.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020