bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Functional lability of RNA-dependent RNA polymerases in animals

Natalia Pinzón[1,2], Stéphanie Bertrand[3], Lucie Subirana[3], Isabelle Busseau[1], Hector Escrivá[3] and Hervé Seitz[1,4]

[1] Institut de Génétique Humaine, UMR 9002 CNRS and université de Montpellier, 141, rue de la Cardonille, 34396 Montpellier CEDEX 5, France
[2] Current address: INSERM U981, Gustave Roussy Cancer Campus, Paris-Saclay University, Villejuif, France
[3] Sorbonne Université, CNRS, Biologie Intégrative des Organismes Marins, BIOM, F-66650 Banyuls-sur-Mer, France
[4] Corresponding author; telephone: (+33)434359936; fax: (+33)434359901; email: herve.seitz@igh.cnrs.fr

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

RNA interference (RNAi) requires RNA-dependent RNA polymerases (RdRPs) in many eukaryotes, and RNAi amplification constitutes the only known function for eukaryotic RdRPs. Yet in animals, classical model organisms can elicit RNAi without possessing RdRPs, and only nematode RNAi was shown to require RdRPs. Here we show that RdRP genes are much more common in animals than previously thought, even in insects, where they had been assumed not to exist. RdRP genes were present in the ancestors of numerous clades, and they were subsequently lost at a high frequency. In order to probe the function of RdRPs in a deuterostome (the cephalochordate *Branchiostoma lanceolatum*), we performed high-throughput analyses of small RNAs from various *Branchiostoma* developmental stages. Our results show that *Branchiostoma* RdRPs are active, generating antisense RNAs from spliced RNA templates, yet they do not appear to participate in RNAi: we did not detect any obvious siRNA population. Our results show that RdRPs have been independently lost in dozens of animal clades, and even in a clade where they have been conserved (cephalochordates) their function in RNAi amplification has been lost. Such a dramatic functional variability reveals an unexpected plasticity in RNA silencing pathways.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 Introduction

Small interfering RNAs (siRNAs) play a central role in the RNA interference (RNAi) response. Usually loaded on a protein of the Ago subfamily of the Argonaute family, they recognize specific target RNAs by sequence complementarity and typically trigger their degradation by the Ago protein (Ghildiyal and Zamore, 2009). In many eukaryotic species, normal siRNA accumulation requires an RNA-dependent RNA polymerase (RdRP). For example in plants, RdRPs are recruited to specific template RNAs and they generate long complementary RNAs (Schiebel et al., 1993 ; Tang et al., 2003 ; Curaba and Chen, 2008). The template RNA and the RdRP product are believed to hybridize, forming a long double-stranded RNA which is subsequently cleaved by Dicer nucleases into double-stranded siRNAs (reviewed in Voinnet, 2008). In fungi, RdRPs have also been implicated in RNAi and in RNA-directed heterochromatinization (Cogoni and Macino, 1999 ; Volpe et al., 2002 ; Hall et al., 2002 ; Sigova et al., 2004), but the exact nature of their products remains elusive: fungal RdRPs are frequently proposed to polymerize long RNAs which can form Dicer substrates after annealing to the RdRP template (Allshire, 2002 ; Motamedi et al., 2004 ; Martienssen and Moazed, 2015). But the purified *Neurospora crassa*, *Thielavia terrestris* and *Myceliophthora thermophila* QDE-1 RdRPs tend to polymerize essentially short (9–21 nt) RNAs *in vitro*, suggesting that they may generate Dicer-independent small RNAs (Makeyev and Bamford, 2002 ; Qian et al., 2016). In various unicellular eukaryotes, RdRPs have also been implicated in RNAi and related mechanisms (*e.g.*, see Kuhlmann et al., 2005 ; Marker et al., 2010). It is usually believed that their products are long RNAs that anneal with the template to generate a Dicer substrate, and that model has gained experimental support in one organism, *Tetrahymena* (Lee and Collins, 2007).

Among eukaryotes, animals are thought to constitute an exception: most classical animal model organisms (*Drosophila* and mammals) can elicit RNAi without the involvement of an RdRP (Ghildiyal and Zamore, 2009). Only one animal model organism was shown to require RdRPs for RNAi: the nematode *Cænorhabditis elegans* (Smardon et al., 2000 ; Sijen et al., 2001). In nematodes, siRNAs made by Dicer only constitute a minor fraction of the total siRNA pool: such "primary" siRNAs recruit an RdRP on target RNAs, triggering the production of short antisense RNAs named "secondary siRNAs" (Pak and Fire, 2007 ; Sijen et al., 2007 ; Gu et al., 2009). Secondary siRNAs outnumber primary siRNAs by ≈ 100-fold (Pak and Fire, 2007) and the major class of secondary siRNAs (the so-called "22G RNAs") is loaded on proteins of the Wago subfamily of the Argonaute family (Yigit et al., 2006 ; Gu et al., 2009). Wago proteins appear to be unable to cleave RNA targets (Yigit et al., 2006). Yet Wago/secondary siRNA/cofactor complexes appear to be much more efficient at repressing mRNA targets than Ago/primary siRNA/cofactor complexes (Aoki et al., 2007), possibly by recruiting another, unknown, nuclease. In contrast to Dicer products (which bear a 5´ monophosphate), direct RdRP products bear a 5´ triphosphate. 22G RNAs are thus triphosphorylated on their 5´ ends (Pak and Fire, 2007). Another class of nematode RdRP products, the "26G RNAs", appears to bear a 5´ monophosphate, and it is not clear whether they are matured from triphosphorylated precursors, or whether they are directly produced as monophosphorylated RNAs (Ruby et al., 2006 ; Han et al., 2009 ; Vasale et al., 2010).

The enzymatic activity of RNA-dependent RNA polymerization can be mediated by several unrelated protein families (Wassenegger and Krczal, 2006). Most of these families are specific to viruses (*e.g.*, PFAM ID #PF00680, PF04196 and PF00978). RdRPs involved in RNAi in plants, fungi and nematodes belong to a family named "eukaryotic RdRPs" (PFAM ID #PF05183). While viral RdRPs are conceivably frequently acquired by virus-mediated horizontal transfer, members of the eukaryotic RdRP family are thought to be inherited vertically only (Burroughs et al., 2014).

Besides eukaryotic RdRPs, other types of RdRP enzymes have been proposed to exist in various animals. It has been suggested that human cells express an atypical RdRP, composed of the catalytic subunit of telomerase and a non-coding RNA (Maida et al., 2009). While that complex exhibits RdRP activity *in vitro*, functional relevance of that activity is unclear, and other mammalian cells were shown to perform RNAi without RdRP activity (Stein et al., 2003). More recently, bat species of the *Eptesicus* clade were shown to possess an RdRP of viral origin, probably acquired upon endogenization of a viral gene at least 11.8 million years ago (Horie et al., 2016).

Here we took advantage of the availability of hundreds of metazoan genomes to draw a detailed map of predicted RdRP genes in animals. We found RdRP genes in a large diversity of animal clades, even in insects, where they had escaped detection so far. Even though RdRP genes are found in diverse animal clades, they are lacking in many species, indicating that they were frequently and independently lost in many lineages. Furthermore, the presence of RdRP genes in non-nematode genomes raises the possibility that additional metazoan lineages

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

possess an RdRP-based siRNA amplification mechanism. We sequenced small RNAs from various developmental stages in one such species with 6 candidate RdRP genes, the cephalochordate *Branchiostoma lanceolatum*, using experimental procedures that were designed to detect both 5´ mono- and tri-phosphorylated RNAs. Our analyses did not reveal any evidence of the existence of secondary siRNAs in that organism. While RNAi is the only known function for eukaryotic RdRPs, we thus propose that *Branchiostoma* RdRPs do not participate in RNAi.

2 Materials and Methods

Bioinformatic analyses of protein sequences

Predicted animal proteome sequences were downloaded from the following databases: NCBI (`ftp://ftp.ncbi.nlm.nih.gov/genomes/`), VectorBase (`https://www.vectorbase.org/download/`), FlyBase (`ftp://ftp.flybase.net/releases/FB2015_03/`), JGI (`ftp://ftp.jgi-psf.org/pub/JGI_data/`), Ensembl (`ftp://ftp.ensembl.org/pub/release-81/fasta/`), WormBase (`ftp://ftp.wormbase.org/pub/wormbase/species/`) and Uniprot (`http://www.uniprot.org/`). The predicted *Branchiostoma lanceolatum* proteome was obtained from the *B. lanceolatum* genome consortium. RdRP HMMer profiles were downloaded from PFAM v. 31.0 (`http://pfam.xfam.org/`): 19 viral RdRP family profiles (PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03035, PF03431, PF04196, PF04197, PF05788, PF05919, PF07925, PF08467, PF12426, PF17501) and 1 eukaryotic RdRP family profile (PF05183). Candidate RdRPs were selected by `hmmsearch` with an E-value cutoff of 10^{-2}. Only those candidates with a complete RdRP domain according to NCBI's Conserved domain search tool (`https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi`) were considered (tolerating up to 20% truncation on either end of the domain). One identified candidate, in the bat *Rhinolophus sinicus*, appears to be a plant contaminant (it is most similar to plant RdRPs, and its genomic scaffold [ACC# LVEH01002863.1] only contains that gene): it was not included in Figure 1 and in Supplementary Figure 1.

The *Branchiostoma* Hen1 candidate was identified using `HMMer` on the predicted *B. lanceolatum* proteome, with an HMMer profile built on an alignment of *Drosophila melanogaster*, *Mus musculus*, *Danio rerio*, *Nematostella vectensis* and *Arabidopsis thaliana* Hen1 sequences.

Phylogenetic tree reconstruction

Amino acid sequences of the eukaryotic RdRP domain (Pfam #PF05183) were retrieved from PFAM (Sonnhammer et al., 1998), and supplemented with the RdRP domains of the proteins identified in the 538 animal proteomes (*cf* above). Sequences were aligned using `hmmalign` (Eddy, 2009) using the HMM profile of the PF05183 RdRP domain. Sequences for which the domain was incomplete were deteled from the alignment. Sites used to reconstruct the phylogenetic tree were selected using `trimAl` (Capella-Gutiérrez et al., 2009) on the Phylemon 2.0 webserver (Sánchez et al., 2011). Bayesian inference (BI) tree was inferred using `MrBayes 3.2.6` (Ronquist et al., 2012), with the model recommended by `ProtTest 1.4` (Darriba et al., 2011) under the Akaike information criterion (LG+Γ), at the CIPRES Science Gateway portal (Miller et al., 2015). Two independent runs were performed, each with 4 chains and one million generations. A burn-in of 25% was used and a fifty majority-rule consensus tree was calculated for the remaining trees. The obtained tree was customized using `FigTree v.1.4.0`.

Sample collection

Mediterranean amphioxus (*Branchiostoma lanceolatum*) males and females were collected at le Racou (Argelès-sur-mer, France) and were induced to spawn as previously described (Fuentes et al., 2007). Embryos were obtained after fertilization in Petri dishes filled with filtered sea water and cultivated at 19°C. Total RNA was extracted from 8, 15, 36 and 60 hours post fertilization (hpf) embryos (three independent batches for each stage, pooled before small RNA gel purification) as well as from males (6 pooled individuals) and females (4 pooled individuals) using the RNeasy mini kit (for embryonic samples) and the RNeasy midi kit (for adult samples) (Qiagen).

Sequencing analyses

The BL09945 locus was PCR-amplified from adult female DNA, cloned in the pGEM-T easy vector (Promega cat. #A1360) and sequenced by MWG Eurofins Genomics.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

For Small RNA-Seq, 18–30 nt RNAs were gel-purified from total RNA (using between 92 and 228 μg total RNA per sample). One quarter of the small RNA preparation was kept untreated before library preparation (for "Libraries #1"). One quarter was incubated for 10 min at room temperature in 100 μL of freshly-prepared 60 mM sodium borate (pH=8.6), 25 mM sodium periodate, then the reaction was quenched with 10 μL glycerol (for "Libraries #2"). One quarter was treated with 1.25 U Terminator exonuclease (Epicentre) in 25 μL 1X Terminator reaction buffer A for 1h at 30°C, then the reaction was quenched with 1.25 μL 500 mM EDTA (pH=8.0) and ethanol-precipitated. RNA was then treated with 5 U Antarctic phosphatase (New England Biolabs) in 20 μL 1X Antarctic phosphatase buffer for 30 min at 37°C, the enzyme was heat-inactivated, then RNA was precipitated, then phosphorylated by 15 U T4 PNK (New England Biolabs) with 50 nmol ATP in 50 μL 1X T4 PNK buffer for 30 min at 37°C, then the enzyme was heat-inactivated (for "Libraries #3"). One quarter was treated successively with Terminator exonuclease, Antarctic phosphatase, T4 PNK then boric acid and sodium periodate, with the same protocols (for "Libraries #4"). Small RNA-Seq libraries were then generated using Illumina's TruSeq Small RNA library preparation kit, following the manufacturer's instructions.

Libraries were sequenced by the MGX sequencing facility (CNRS, Montpellier, France). Read sequences were aligned on the *B. lanceolatum* genome assembly (Marlétaz *et al.*, submitted) using `bowtie2`. A database of abundant non-coding RNAs was assembled by a search for orthologs for human and murine rRNAs, tRNAs, snRNAs, snoRNAs and scaRNAs; deep-sequencing libraries were also mapped on that database using `bowtie2`, and matching reads were flagged as "abundant ncRNA fragments". For pre-miRNA annotation, every *B. lanceolatum* locus with a Blast E-value $\leqslant 10^{-6}$ to any of the annotated *B. floridae* or *B. belcheri* pre-miRNA hairpins in miRBase v.22 was selected. Reads matching these loci were identified using `bowtie2`. For the measurement of miRNA abundance during development, hairpins were further screened for their `RNAfold`-predicted secondary structure and their read coverage: Supplementary Table 1 only lists unbranched hairpins with at least 25 bp in their stem, with a predicted $\Delta G_{folding} \leqslant -15\ kcal.mol^{-1}$, generating mostly 21- to 23-mer RNAs, and with at least 20 ppm read coverage on any nucleotide of the hairpin.

RNA-Seq data was taken in Marlétaz *et al.* (submitted) for embryonic and juvenile samples. Adult sample libraries were prepared and sequenced by "Grand plateau technique régional de génotypage" (SupAgro-INRA, Montpellier). mRNA abundance data was extracted using `vast-tools` (Tapial et al., 2017).

3 Results

A sporadic phylogenetic distribution of RdRP genes

Previous analyses showed that a few animal genomes contain candidate RdRP genes (Wassenegger and Krczal, 2006 ; Zong et al., 2009 ; Horie et al., 2016 ; Lewis et al., 2018). Rapid development of sequencing methods recently made many animal genomes available, allowing a more complete coverage of the phylogenetic tree. A systematic search for RdRP candidates (including every known viral or eukaryotic RdRP family) in 538 predicted metazoan proteomes confirms that animal species possessing RdRPs are unevenly scattered in the phylogenetic tree, but they are much more abundant than previously thought: we identified 98 metazoan species with convincing eukaryotic RdRP genes (see Figure 1**A**). Most RdRPs identified in animal predicted proteomes belong to the eukaryotic RdRP family, but 3 species (the Enoplea *Trichinella murrelli*, the Crustacea *Daphnia magna* and the Mesozoa *Intoshia linei*) possess RdRP genes belonging to various viral RdRP families (in green, dark blue and light blue on Figure 1**A**), which were probably acquired by horizontal transfer from viruses. Most sequenced nematode species appear to possess RdRP genes. But in addition, many other animal species are equipped with eukaryotic RdRP genes, even among insects (the Diptera *Clunio marinus* and *Rhagoletis zephyria*), where RdRPs were believed to be absent (Li et al., 2018 ; Lewis et al., 2018).

Our observation of eukaryotic family RdRPs in numerous animal clades therefore prompted us to revisit the evolutionary history of animal RdRPs: eukaryotic RdRPs were probably present in the last ancestors for many animal clades (including insects, mollusks, deuterostomes) and they were subsequently lost independently in most insects, mollusks and deuterostomes. It has been recently shown that the last ancestor of arthropods possessed an RdRP, which was subsequently lost in some lineages (Lewis et al., 2018): that result appears to be generalizable to a large diversity of animal clades. The apparent absence of RdRPs in some species may be due to genome incompleteness, or to defective proteome prediction. Excluding species with low numbers of long predicted proteins ($\geqslant$ 500 or 1,000 amino acids) indeed eliminates a few dubious proteomes, but the resulting distribution of RdRPs in the phylogenetic tree is only marginally affected, and still suggests multiple recent RdRP losses in diverse

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

lineages (see Supplementary Figure 1).

Alternatively to multiple gene losses, such a sporadic phylogenetic distribution could be due to frequent horizontal transfer of RdRP genes in animals. In order to assess these two possibilities, it is important to better understand the evolution of metazoan RdRPs in the context of the whole eukaryotic RdRP family. We therefore used sequences found in all eukaryotic groups for phylogenetic tree reconstruction. The supports for deep branching are low and do not allow us to propose a complete evolutionary history scenario of the whole eukaryotic RdRP family (see Figure 2**A**). However, metazoan sequences are forming three different groups, which were named RdRP α, β and γ according to the pre-existing nomenclature (Zong et al., 2009), and their position in relation to non-metazoan eukaryotic sequences does not support an origin through horizontal gene transfer. The only data that would support horizontal gene transfer pertains to the metazoan sequences of the RdRP β group (see Figure 2**C**). Indeed, sequences of stramenopiles and a fungus belonging to parasitic species are embedded in this clade. For the RdRP α and γ groups, the phylogeny strongly suggests that they derive from at least two genes already present in the common ancestor of cnidarians and bilaterians and that the scarcity of RdRP presence in metazoans would be the result of many secondary gene losses. Even the *Strigamia maritima* RdRP was probably not acquired by a recent horizontal transfer from a fungus, as has been proposed (Lewis et al., 2018): when assessed against a large number of eukaryotic RdRPs, the *S. maritima* sequence clearly clusters within metazoan γ RdRP sequences. In summary, we conclude that RdRPs were present in the last ancestors of many animal clades, and they were recently lost independently in diverse lineages.

Experimental search for RdRP products in *Branchiostoma*

In an attempt to probe the functional conservation of RdRP-mediated RNAi amplification among metazoans, we decided to search for secondary siRNAs in an organism where RdRP candidates could be found, while being distantly related to *C. elegans*. We reasoned that endogenous RNAi may act as a gene regulator during development or as an anti-pathogen response. Thus siRNAs are more likely to be detected if several developmental stages are probed, and if the analyzed specimens are gathered in a natural ecosystem, where they are naturally challenged by pathogens. From these considerations it appears that the most appropriate organism is a cephalochordate species, *Branchiostoma lanceolatum* (Bertrand and Escrivá, 2011). In good agreement with the known scarcity of gene loss in that lineage (Louis et al., 2012), cephalochordates also constitute the only bilaterian clade for which both RdRP α and γ sequences can be found, thus increasing the chances of observing RNAi amplification despite the diversification of eukaryotic RdRPs into three groups. According to our HMMer-based search, the *B. lanceolatum* genome encodes 6 candidate RdRPs, three of which containing an intact active site DbDGD (with b representing a bulky amino acid; Iyer et al., 2003) (see Figure 1**B**). The current *B. lanceolatum* genome assembly contains a direct 1,657 bp repeat in one of the 6 RdRP genes, named BL09945. This long duplication appears to be an assembly artifact: we cloned and re-sequenced that locus and identified two alleles (with a synonymous mutation on the 505th codon; deposited at GenBank under accession numbers MH261373 and MH261374), and none of them contained the repeat. In subsequent analyses, we thus used a corrected version of that locus, where the 1,657 bp duplication is removed.

In most metazoan species, siRNAs (as well as miRNAs) bear a 5´ monophosphate and a 3´ hydroxyl (Elbashir et al., 2001 ; Hutvágner et al., 2001). The only known exceptions are "22G" secondary siRNAs in nematodes (they bear a 5´ triphosphate; Pak and Fire, 2007), which may be primary polymerization products by an RdRP; and Ago2-loaded siRNAs and miRNA in *Drosophila*, which are 3´-methylated on their 2´ oxygen after loading on Ago2 and unwinding (Pélisson et al., 2007 ; Horwich et al., 2007). Note that one class of RdRP products (nematode "26G" RNAs) bears a 5´ monophosphate (Ruby et al., 2006 ; Han et al., 2009 ; Vasale et al., 2010).

In order to detect small RNAs with any number of 5´ phosphates, bearing either an unmodified or a methylated 3´ end, we prepared multiple Small RNA-Seq libraries (see Figure 3**A**). Total RNA was extracted from various embryonic stages: gastrula (8 hours post-fertilization, hpf), early neurula (15 hpf), premouth neurula (36 hpf) and larvae (60 hpf), as well as from adult male and female specimens collected from their natural ecosystem. Small (18 to 30 nt long) RNAs were gel-purified, then Small RNA-Seq libraries were prepared using either the standard Small RNA-Seq protocol (which detects 5´ monophosphorylated small RNAs, whether they bear a 3´ methylation or not; "Library #1"); or by oxidizing small RNAs with $NaIO_4$ in the presence of H_3BO_3 prior to library preparation (such treatment renders unmodified 3´ RNAs non-ligatable, hence undetectable by deep-sequencing; Ghildiyal et al., 2008; "Library #2"); or by treating small RNAs with the Terminator exonuclease (which degrades 5´ monophosphorylated RNAs) then with phosphatase then T4 PNK (to convert 5´ polyphosphorylated RNAs and

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

5´ hydroxyl RNAs into monophosphorylated RNAs, suitable for Small RNA-Seq library preparation; "Library #3"); or by a combination of both treatments (to detect only small RNAs bearing a 5´ polyphosphate or a 5´ hydroxyl, and a 3´ modification; "Library #4").

In the course of library preparation, it appeared that Libraries #4 contained very little ligated material, suggesting that small RNAs with a 3´ modification as well as $n \geqslant 0$ (with $n \neq 1$) phosphates on their 5´ end, are very rare in *Branchiostoma* regardless of developmental stage. This observation was confirmed by the annotation of the sequenced reads: most reads in Libraries #4 did not map on the *B. lanceolatum* genome, probably resulting from contaminating nucleic acids (see Supplementary Figure 2).

In Libraries #1 in each developmental stage, most sequenced reads fall in the 18–30 nt range as expected. Other libraries tend to be heavily contaminated with shorter or longer reads, and 18–30 nt reads only constitute a small fraction of the sequenced RNAs (see Figure 3**B** for adult male libraries; see Supplementary Data section 1 for other developmental stages). miRNA loci have been annotated in two other cephalochordate species, *B. floridae* and *B. belcheri* (156 pre-miRNA hairpins for *B. floridae* and 118 for *B. belcheri* in miRBase v. 22). We identified the *B. lanceolatum* orthologous loci for annotated pre-miRNA hairpins from *B. floridae* or *B. belcheri*. Mapping our libraries on that database allowed us to identify candidate *B. lanceolatum* miRNAs. These RNAs are essentially detected in our Libraries #1, implying that, like in most other metazoans, *B. lanceolatum* miRNAs are mostly 22 nt long, they bear a 5´ monophosphate and no 3´ methylation (see Figure 3**C** for adult male libraries; see Supplementary Data section 2 for other developmental stages). Among the *B. lanceolatum* loci homologous to known *B. floridae* or *B. belcheri* pre-miRNA loci, 56 exhibit the classical secondary structure and small RNA coverage pattern of pre-miRNAs (*i.e.*, a stable unbranched hairpin generating mostly 21–23 nt long RNAs from its arms). These 56 loci, the sequences of the miRNAs they produce, and their expression profile during development, are shown in Supplementary Table 1.

No evidence of RdRP-based siRNA amplification in *Branchiostoma*.

In an attempt to detect siRNAs, we excluded every sense pre-miRNA-matching read and searched for distinctive siRNA features in the remaining small RNA populations. Whether RdRPs generate long antisense RNAs which anneal to sense RNAs to form a substrate for Dicer, or whether they polymerize directly short single-stranded RNAs which are loaded on an Argonaute protein, the involvement of RdRPs in RNAi should result in the accumulation of antisense small RNAs for specific target genes. These small RNAs should exhibit characteristic features:

- a narrow size distribution (imposed either by the geometry of the Dicer protein, or by the processivity of the RdRP: Zhang et al., 2004 ; Aoki et al., 2007; the length of Argonaute-loaded RNAs can also be further refined by exonucleolytic trimming of 3´ ends protruding from Argonaute: Gu et al., 2009 ; Han et al., 2011 ; Liu et al., 2011 ; Feltzin et al., 2015 ; Wang et al., 2016 ; Hayashi et al., 2016);

- and possibly a sequence bias on their 5´ end; it is remarkable that the known classes of RdRP products in metazoans (nematode 22G and 26G RNAs) both display a strong bias for a guanidine at their 5´ end. RNA polymerases in general tend to initiate polymerization on a purine nucleotide (Jorgensen et al., 1969 ; Wu and Goldthwait, 1969a ; Wu and Goldthwait, 1969b ; Miller et al., 1986 ; Kuzmine et al., 2003 ; Juven-Gershon and Kadonaga, 2010 ; Hetzel et al., 2016) and it can be expected that primary RdRP products bear either a 5´A or a 5´G. Of note: loading on an Argonaute may also impose a constraint on the identity of the 5´ nucleotide, because of a sequence preference of either the Argonaute protein or its loading machinery (Mi et al., 2008 ; Montgomery et al., 2008 ; Takeda et al., 2008 ; Ghildiyal et al., 2010 ; Frank et al., 2010 ; Seitz et al., 2011).

The analysis of transcriptome-matching, non-pre-miRNA-matching small RNAs does not indicate that such small RNAs exist in *Branchiostoma* (see Figure 4 for adult males, and Supplementary Data, section 3, for the complete data set). In early embryos, 5´ monophosphorylated small RNAs exhibit the typical size distribution and sequence biases of piRNA-rich samples: a heterogeneous class of 23 to 30 nt long RNAs. Most of them tend to bear a 5´ uridine, but 23 to 26 nt long RNAs in the sense orientation to annotated transcripts tend to have an adenosine at position 10 (especially when the matched transcript exhibits a long ORF; see Supplementary Data, section 4). Vertebrate and *Drosophila* piRNAs display very similar size profiles and sequences biases (Saito et al., 2006 ; Lau et al., 2006 ; Girard et al., 2006 ; Aravin et al., 2006 ; Watanabe et al., 2006 ; Brennecke et al., 2007 ; Houwing et al., 2007). These 23–30 nt long RNAs may thus constitute the *Branchiostoma* piRNAs, but surprisingly, they do not appear to bear a 2´-*O*-methylation on their 3´ end (see Discussion).

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

In summary, transcriptome-matching small RNAs in our *Branchiostoma* libraries contain miRNA and piRNA candidates, but they do not contain any obvious class of presumptive secondary siRNAs that would exhibit a precise size distribution, and possibly a 5´ nucleotide bias. One could imagine that transcriptome-matching siRNAs were missed in our analysis, because of issues with the *Branchiostoma* transcriptome assembly. It is also conceivable that siRNAs exist in *Branchiostoma*, but they do not match its genome or transcriptome (they could match pathogen genomes, for example if they contribute to an anti-viral immunity). We therefore analyzed other potential siRNA types: **(i)** genome-matching reads that do not match abundant non-coding RNAs (rRNAs, tRNAs, snRNAs, snoRNAs or scaRNAs); **(ii)** reads that match transcripts exhibiting long ($\geqslant$ 100 codons, initiating on one of the three 5´-most AUG codons) open reading frames; **(iii)** reads that do not match the *Branchiostoma* genome, nor its transcriptome (potential siRNAs derived from pathogens). Once again, none of these analyses revealed any siRNA population in *Branchiostoma* (see detailed results in Supplementary Data, sections 1, 4 and 5). This is in striking contrast to *Cænorhabditis elegans*, where antisense transcriptome-matching siRNAs (mostly 22 nt long, starting with a G) are easily detectable in each of these categories (see Supplementary Data, section 7, for our analysis of publicly available *C. elegans* data; Gu et al., 2009).

Unambiguous RdRP activity is detectable, but unlikely to contribute to RNAi

Our failure to detect siRNA candidates may simply be due to the fact that they are poorly abundant in the analyzed developmental stages. In order to enrich for small RNA populations derived from RdRP activity, and exclude all the other types of small RNAs, we considered small RNAs mapping on exon-exon junctions in the antisense orientation. The antisense sequence of the splicing donor (GU) and acceptor (AG) sites does not constitute a donor/acceptor pair itself, implying that any RNA antisense to a spliced RNA must have originated from the action of an RdRP on the spliced RNA — it cannot derive from the splicing of an RNA transcribed in the antisense orientation.

We therefore selected all the small RNA reads that map on the transcriptome, but fail to map on the genome. These reads map predominantly on the sense strand of annotated transcripts, and they are hardly detected in libraries enriched for 3´ modification (Libraries #2 and 4; see Figure 5 for adult males, and Supplementary Data, section 6, for the complete data set). For several developmental stages, small RNAs mapping on the transcriptome in the sense orientation in Library #1 (total 5´ monophosphorylated small RNAs) exhibit a size distribution reminiscent of that of piRNAs. One possible interpretation is that piRNA-sized degradation products of spliced transcripts are loaded on Piwi proteins, and selectively stabilized. Small RNAs mapping in the sense orientation in Libraries #3 do not exhibit any consistent size preference, suggesting that these are mere degradation products. Importantly, antisense small RNAs mapping specifically on spliced transcripts tend to be extremely rare. Across development, the most abundant antisense small RNAs are 18 and 29 nt long in Libraries #1 (see Figure 5**A** and Supplementary Data, section 6) and 18 nt long in Libraries #3 (see Figure 5**C** and Supplementary Data, section 6), with Libraries #3 of 8h and 15h embryos exhibiting the highest antisense/sense read ratios (see Supplementary Data, section 6). Yet even these size classes do not exhibit any strong sequence bias (see Figure 5 for adult males, and Supplementary Data, section 6, for the complete data set). Because our library preparation focused on 18–30 nt RNAs, it is possible that the 18 nt peak is just a shoulder of a more prominent peak for shorter RNAs, that were not analyzed because they fall outside the 18–30 nt range. Yet even 17-mers (which could be analyzed, because the experimental preparation of an 18–30 nt library is always contaminated with other RNA lengths) fail to display any remarkable sequence bias (data not shown). We propose that the small RNAs that we observed after exclusion of piRNA candidates and pre-miRNA-matching reads are non-functional degradation products. It therefore appears unlikely that the *B. lanceolatum* RdRPs generate siRNAs or siRNA precursors.

Amphioxus RdRPs generate antisense RNAs for specific RNA templates

Selecting RNAs that map on exon-exon junctions in the antisense orientation (*i.e.*: unambiguously RdRP-derived RNA fragments) offers the possibility to identify the transcripts that serve as templates for RdRPs. We thus asked whether the transcripts matched by such small RNAs constitute a distinctive group of transcripts, or whether they are merely sampled randomly from the transcriptome. If the RdRPs generated antisense RNAs without any particular specificity, the abundance of unambiguously RdRP-derived small RNAs should correlate with transcript abundance, modulated by the number of exon-exon junctions. On the other hand, if RdRPs selected specific RNAs as templates, then some RNAs should generate disproportionate amounts of antisense RNA fragments.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

In order to normalize both for sense transcript abundance and for its number of exon-exon junctions, we computed the ratio of the number of antisense/sense reads mapping on a given transcript whithout mapping on the genome. We selected transcripts that generate high antisense/sense ratio in at least one developmental stage (ratio>10, with antisense reads for that transcript exceeding 10 reads per million of genome-matching reads that do not match abundant non-coding RNAs), and whose total number of antisense exon-exon reads in the pooled 24 libraries exceeds 100 raw reads. These stringent criteria identified 4 transcripts: BL67984 (which encodes an uncharacterized protein with a C-type lectin domain); BL15804 (which encodes a completely uncharacterized protein); BL28448 (which encodes a reverse-transcriptase of retroviral or retrotransposon origin) and BL16120 (which encodes a putative tyrosine kinase receptor). The first two transcripts generate mostly 5´ monophosphorylated antisense small RNAs (see Figure 6**A**) while the latter two generate mostly 5´ hydroxyl or polyphosphorylated small RNAs (see Figure 6**B**). Importantly, these antisense small RNAs do not tend to accumulate when the sense transcript is most abundant (see Figure 6**C**), but they tend to follow closely a peak of expression of candidate RdRPs with a putatively functional active site (BL09945, BL02069 and BL23385; see Figure 6**D**). This observation suggests that not only RdRPs select specific mRNA templates, but their activity on these templates is also developmentally regulated.

Because small RNAs mapping on exon-exon junctions in the antisense orientation do not exhibit any particular size distribution or sequence bias (see Supplementary Data, section 6), we conclude that RdRP activity on these transcripts is not involved in RNAi.

4 Discussion

In cellular organisms, the only known function for RdRPs is the generation of siRNAs or siRNA precursors. It is thus frequently assumed (Maida et al., 2009 ; Lewis et al., 2018) or hypothesized (Horie et al., 2016) that animal RdRPs participate in RNAi. In particular, it has recently been proposed that arthropod RdRPs are required for RNAi amplification, and arthropod species devoid of RdRPs may rather generate siRNA precursors through bidirectional transcription (Lewis et al., 2018). While this hypothesis would provide an elegant explanation to the sporadicity of RdRP gene distribution in the phylogenetic tree, the provided evidence remains disputable: it has been proposed that a high ratio of antisense over sense RNA is diagnostic of bidirectional transcription, yet it remains to be explained why RNA-dependent RNA polymerization would produce less steady-state antisense RNA than DNA-dependent polymerization.

Branchiostoma 5´ monophosphorylated small RNAs do not appear to bear a 2´-*O*-methyl on their 3´ end: Libraries #2 contain few genome-matching sequences, and their size distribution suggests they are mostly constituted of contaminating RNA fragments rather than miRNAs, piRNAs or siRNAs. In every animal model studied so far, piRNAs were shown to bear a methylated 3´ end (Vagin et al., 2006 ; Ruby et al., 2006 ; Kirino and Mourelatos, 2007a ; Houwing et al., 2007 ; Grimson et al., 2008 ; Fu et al., 2018). The enzyme responsible for piRNA methylation, Hen1 (also known as Pimet), has been identified in *Drosophila*, mouse and zebrafish (Horwich et al., 2007 ; Saito et al., 2007 ; Kirino and Mourelatos, 2007b ; Kamminga et al., 2010). In order to determine whether the absence of piRNA methylation in *Branchiostoma* could be due to an absence of the Hen1 enzyme, we searched for Hen1 orthologs in the predicted *Branchiostoma* proteome. Our HMMer search identified a candidate, BL03504. Its putative methyl-transferase domain contains every known important amino acid for Hen1 activity according to Huang et al., 2009 (see Supplementary Figure 4), suggesting that it is functional. Further studies will be required to investigate the biological activity of that putative enzyme, and to understand why it does not methylate *Branchiostoma* piRNAs.

We observed evidence of RdRP activity in *Branchiostoma*, which generates antisense RNAs from specific RNA templates in a developmentally-regulated manner. Yet these antisense RNAs do not appear to be processed into *bona fide* siRNAs, and most probably, the short antisense exon-exon junction RNAs that we observed are non-functional degradation products. We thus propose that *Branchiostoma* RdRPs are not involved in RNAi.

One could hypothesize that these RdRPs do not play any biological function. Yet at least two of them, BL02069 and BL23385, possess a full-length RdRP domain with a preserved catalytic site. The conservation of these two active genes suggests that they are functionally important. It can therefore be speculated that *Branchiostoma* RdRPs play a biological role, which is unrelated to RNAi. Such a function may involve the generation of double-stranded RNA (formed by the hybridization of template RNA with the RdRP product), but it could also involve single-stranded RdRP products. Future work will be needed to identify the biological functionality of these

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

enzymes. We also note that the fungus *Aspergillus nidulans*, whose genome encodes two RdRPs with a conserved active site, does not require any of those for RNAi (Hammond and Keller, 2005).

Animal RdRPs thus constitute an evolutionary enigma: not only have they been frequently lost independently in numerous animal lineages, but even in the clades where they have been conserved, their biological function seems to be variable. While RNAi is an ancient gene regulation pathway (Ghildiyal and Zamore, 2009), involving the deeply conserved Argonaute and Dicer protein families, the role of RdRPs in RNAi appears to be accessory. Even though RdRPs are strictly required for RNAi in very diverse extant clades (ranging from nematodes to plants), it would be misleading to assume that RNAi constitutes their only biological function. Given their functional lability, it is even questionable whether the last common ancestor to plants, animals and fungi used RdRPs for RNAi amplification: multiple, independent exaptation events of RdRPs for RNAi amplification may actually explain the current mechanistic diversity of their involvement in RNAi in current-day nematodes, plants and fungi.

Code availability

Source code, detailed instructions, and intermediary data files are accessible on GitHub (`https://github.com/HKeyHKey/Pinzon_et_al_2018`) as well as on `https://www.igh.cnrs.fr/en/research/departments/genetics-devel systemic-impact-of-small-regulatory-rnas/165-computer-programs`.

Data availability

Deep-sequencing data has been deposited at NCBI's Short Read Archive under accession #SRP125901. Sequences of the re-sequenced *B. lanceolatum* BL09945 locus have been deposited at GenBank under accession #MH261373 and #MH261374.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

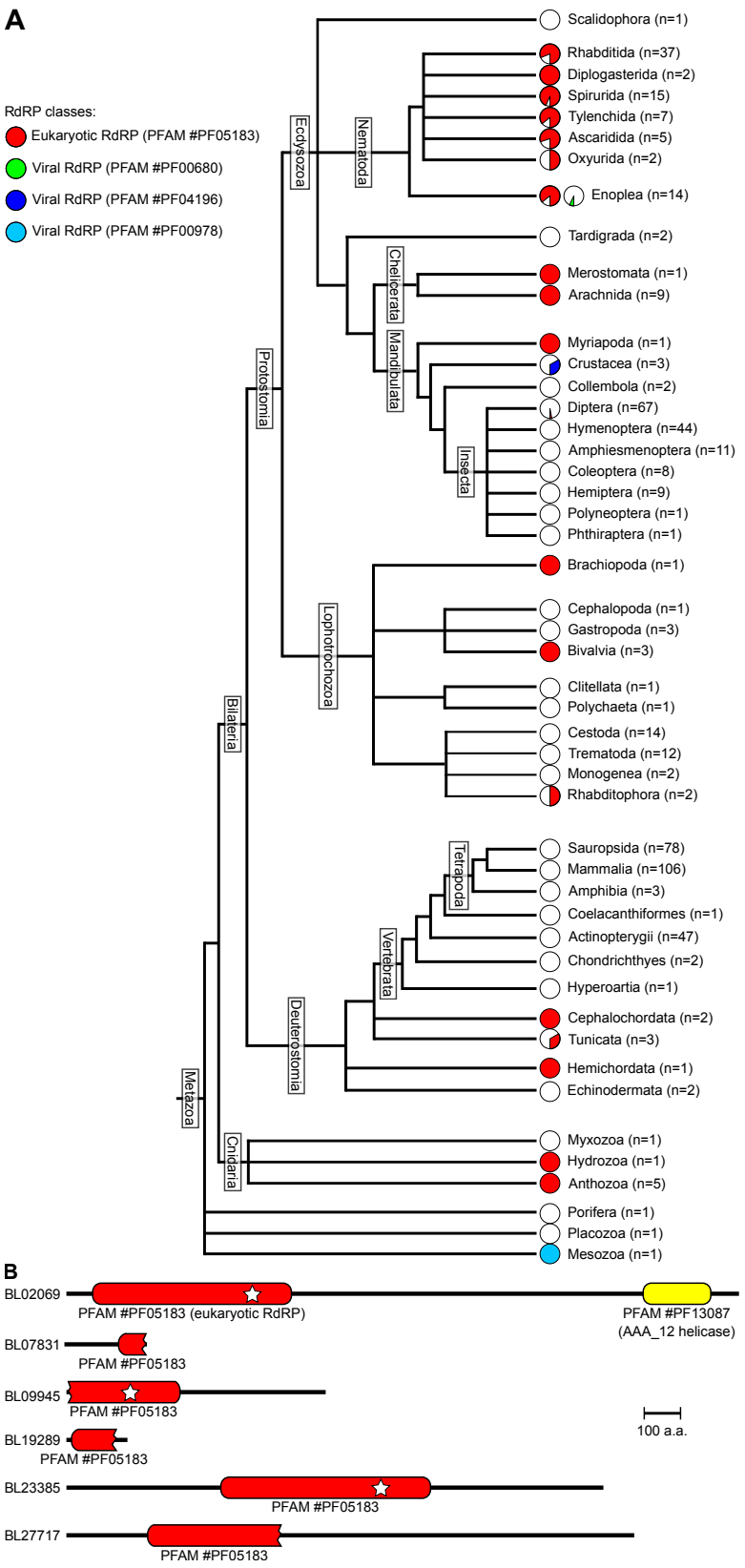


Figure 1: **Phylogenetic distribution of RdRP genes in metazoans. A.** Proteome sequences from 538 metazoans were screened for potential RdRPs. For each clade indicated on the right edge, n is the number of species analyzed in the clade, and piecharts indicate the proportion of species possessing RdRP genes (with each RdRP family represented by one piechart, according to the color code given at the top left). **B.** An HMMer search identifies 6 candidate RdRPs in the predicted *Branchiostoma lanceolatum* proteome. Only 2 candidates have a complete RdRP domain (represented by a red bar with round ends; note that apparent domain truncations may be due to defective proteome prediction). A white star indicates that every catalytic amino acid is present. Candidate BL02069 also possesses an additional known domain, AAA_12 (in yellow).

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

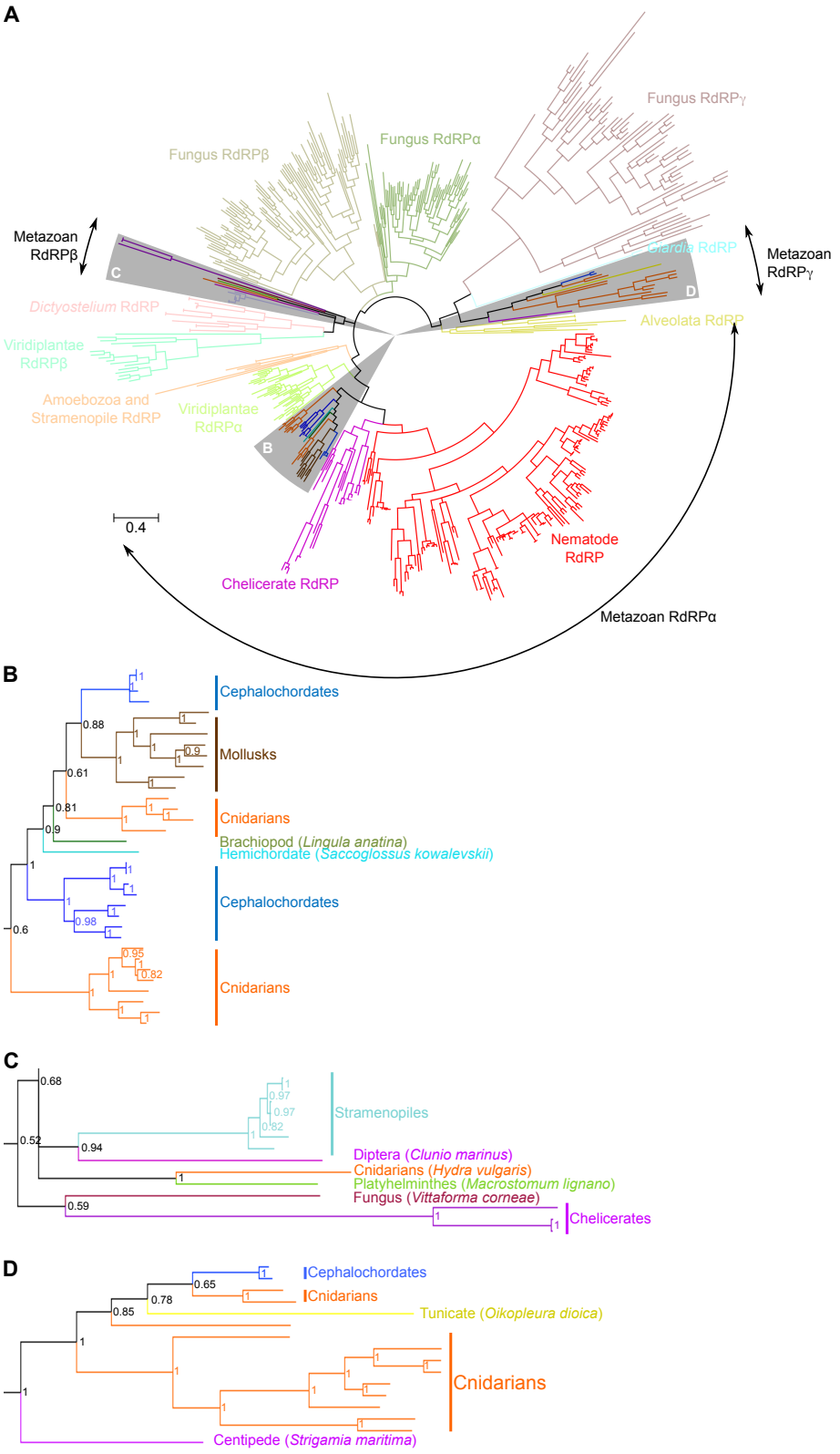


Figure 2: **Eukaryotic RdRP phylogeny supports the vertical transfer scenario.** **A**: Bayesian phylogenetic tree of the eukaryotic RdRP family. α, β and γ clades of eukaryotic RdRPs have been defined by Zong et al., 2009. Sectors highlighted in grey are detailed in panels **B**, **C** and **D** for clarity. Scale bar: 0.4 amino acid substitution per position. Posterior probability values are indicated for each node in panels **B**–**D**.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

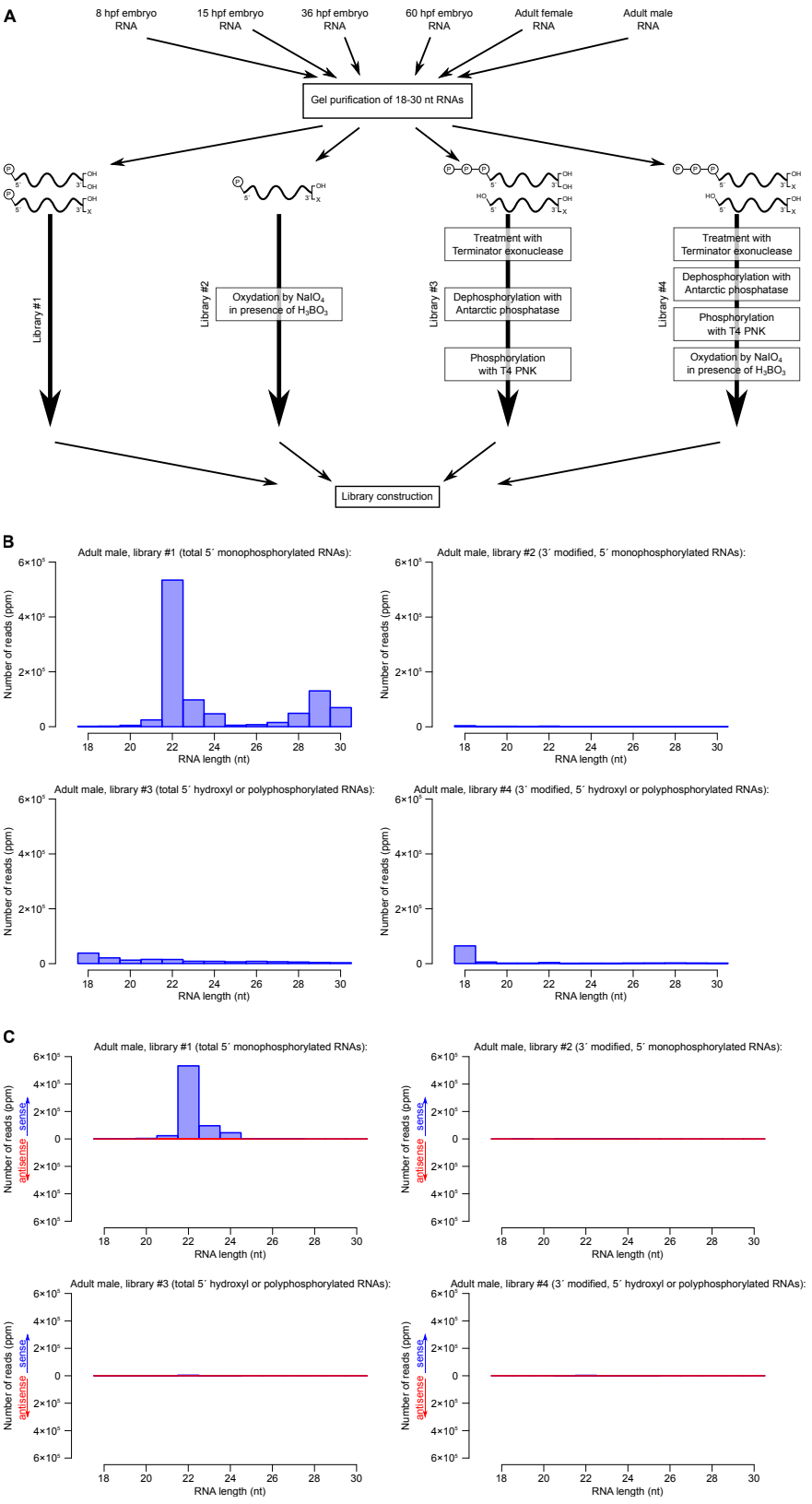


Figure 3: **Detection of *B. lanceolatum* small RNAs. A.** Four libraries were prepared for each biological sample, to detect small RNAs bearing either a single 5´ phosphate (Libraries #1 and 2) or any other number of phosphates (including zero; Libraries #3 and 4), and either a 2´-OH and 3´-OH 3´ end (Libraries #1 and 3), or a protected (*e.g.*, 2´-*O*-methylated) 3´ end (Libraries #2 and 4). hpf: hours post fertilization. **B.** Size distribution of genome-matching adult male small RNAs, excluding reads that match abundant non-coding RNAs (rRNAs, tRNAs, snRNAs, snoRNAs or scaRNAs). Read numbers are normalized by the total number of genome-matching reads (including <18 nt and >30 nt reads) that do not match abundant non-coding RNAs. **C.** Size distribution of adult male small RNAs matching pre-miRNA hairpins in the sense (blue) or antisense (red) orientation.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

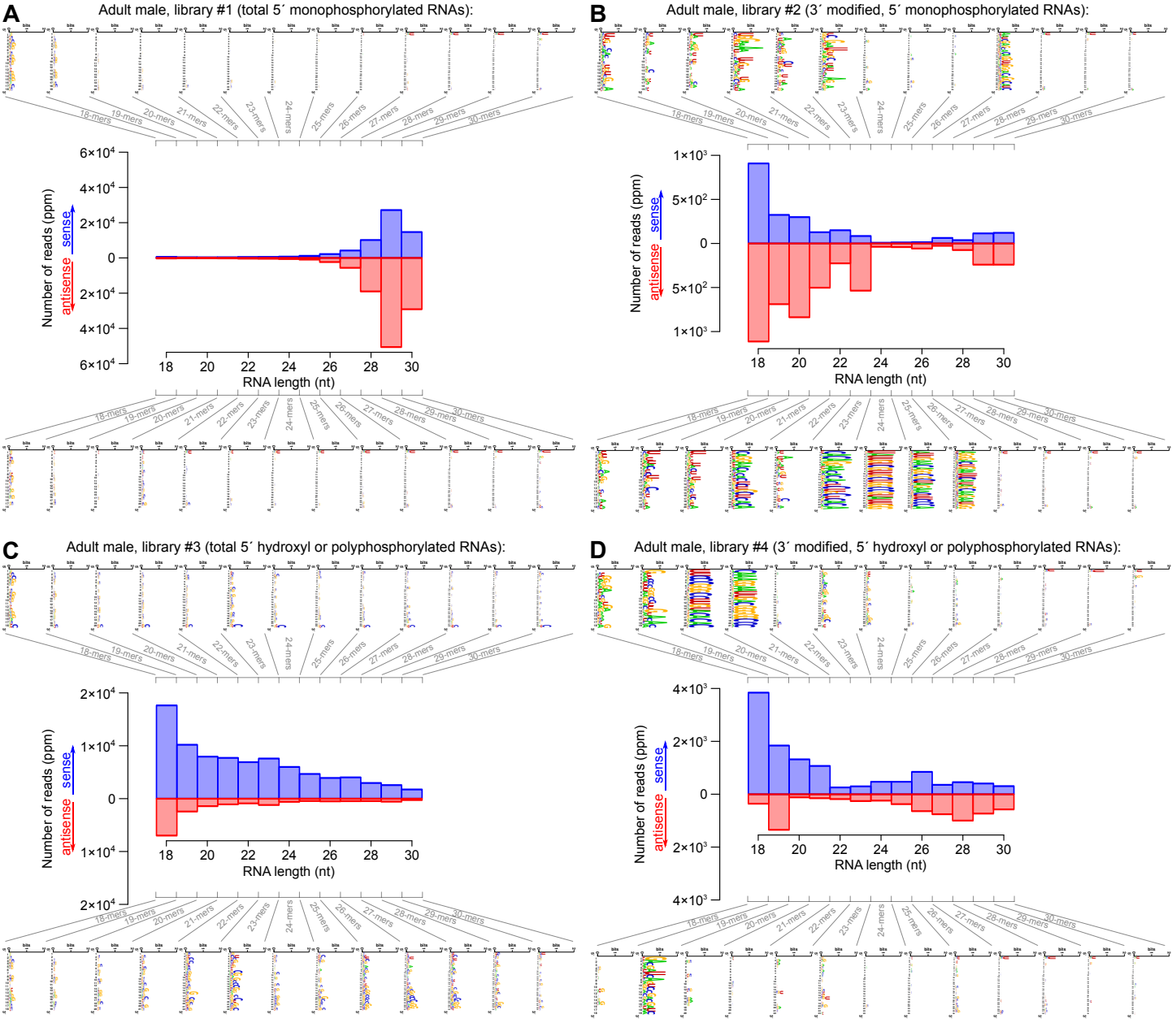


Figure 4: **Size distribution and sequence logos for transcriptome-matching small RNAs in adult males.** See Supplementary Data, section 3, for the other developmental stages. **A**: Library #1, **B**: Library #2, **C**: Library #3, **D**: Library #4. Numbers of reads are expressed as parts per million (ppm) after normalization to the total number of genome-matching reads that do not match abundant non-coding RNAs. For each orientation (sense or antisense-transcriptome-matching reads), a logo analysis was performed on each size class (18 to 30 nt long RNAs).

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

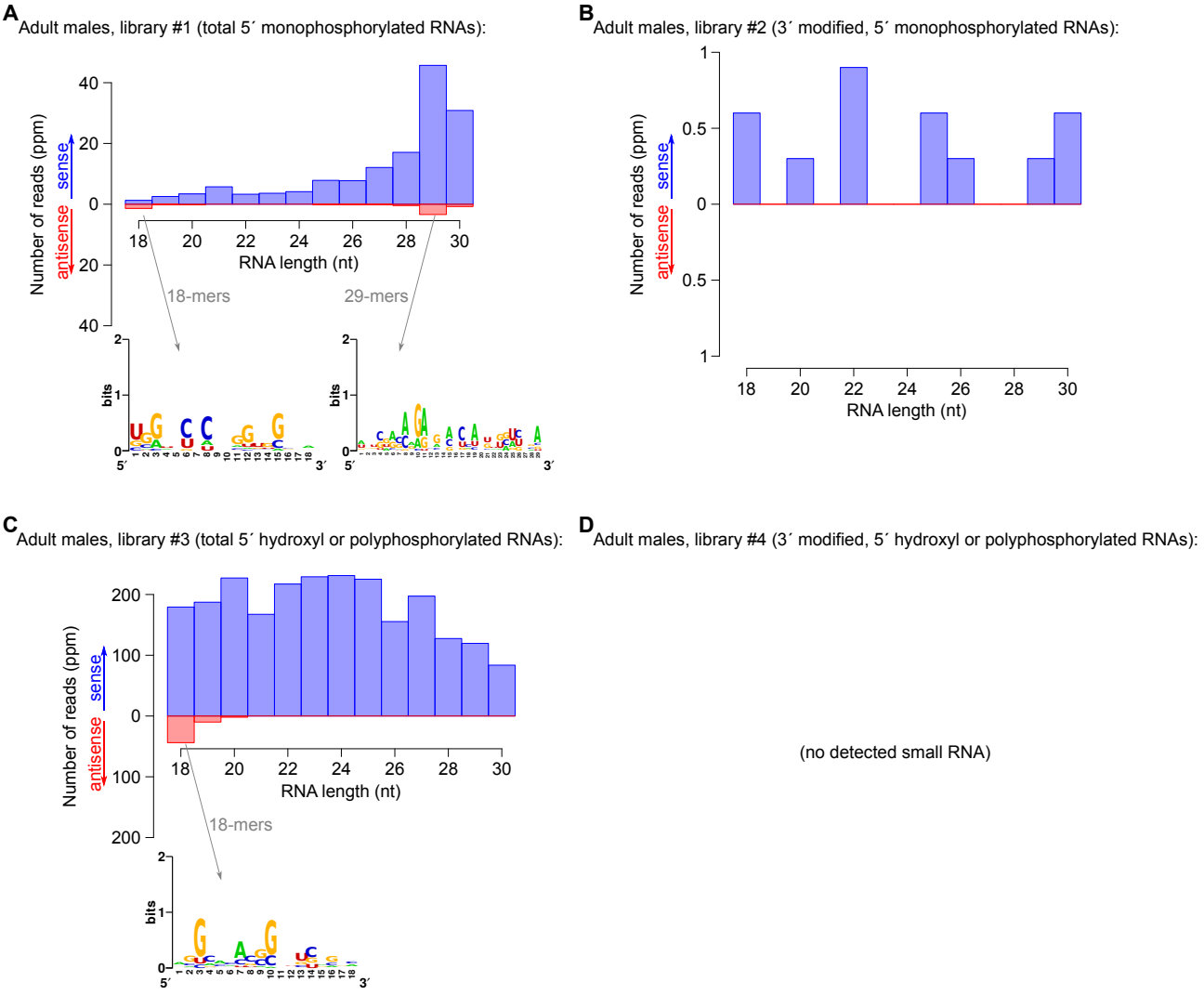


Figure 5: **Size distribution for small RNAs mapping specifically on the spliced transcriptome, in adult males.** See Supplementary Data, section 6, for the other developmental stages. **A**: Library #1, **B**: Library #2, **C**: Library #3, **D**: Library #4. Numbers of reads are expressed as parts per million (ppm) after normalization to the total number of genome-matching reads that do not match abundant non-coding RNAs. For antisense-transcriptome-matching reads, a logo analysis was performed on the prominent size class.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

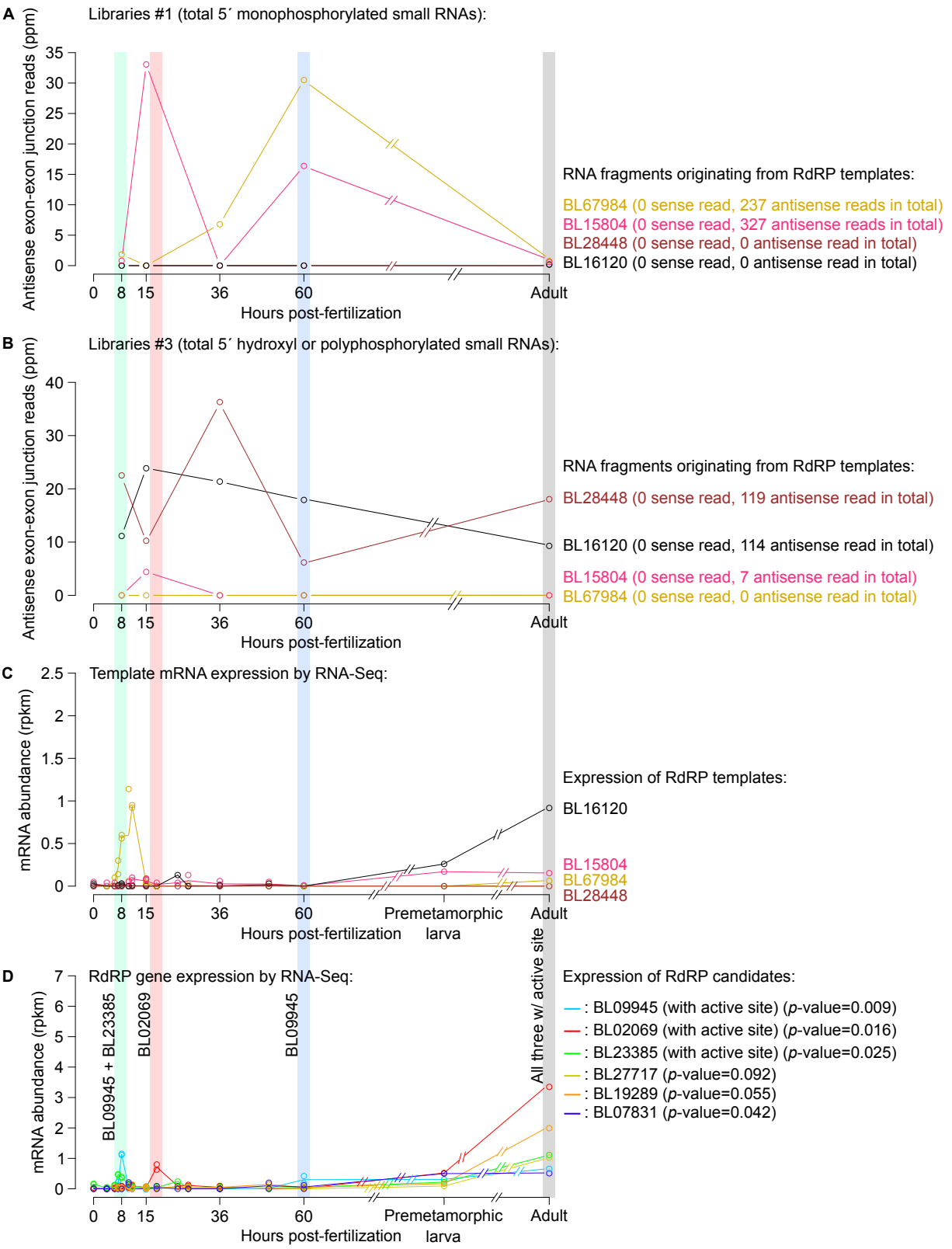


Figure 6: **Specific transcripts are preferentially used as RdRP templates.** Abundance of antisense exon-exon junction RNA fragments originating from the four identified RdRP template mRNAs, in Libraries #1 (panel **A**) and 3 (panel **B**). For each of the four identified RdRP template mRNAs (panel **C**) and the mRNAs for each of the six RdRP genes (panel **D**), their abundance in various developmental stages was measured by RNA-Seq, and reported as cRPKM (corrected-for-mappability reads per kb and per million mapped reads; Labbé et al., 2012). RdRP expression peaks are indicated by vertical color bars, and the name of peaking RdRPs is indicated in panel **D**. Adult male and female data were averaged (see Supplementary Figure 3 for separated sex analysis). In panel **D**, temporal regulation of RdRP expression in embryos and juveniles was assessed by the Kruskal-Wallis test (p-values are indicated in the legend for each RdRP).

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

REFERENCES

Allshire, R. (2002). Molecular biology. RNAi and heterochromatin–a hushed-up affair. *Science*, 297(5588):1818–1819.

Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K., and Tabara, H. (2007). In vitro analyses of the production and activity of secondary small interfering RNAs in *C. elegans*. *EMBO J*, 26(24):5007–5019.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J. J., Ju, J., Sheridan, R., Sander, C., Zavolan, M., and Tuschl, T. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, 442(7099):203–207.

Bertrand, S. and Escrivá, H. (2011). Evolutionary crossroads in developmental biology: amphioxus. *Development*, 138(22):4819–4830.

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, 128(6):1089–1103.

Burroughs, A. M., Ando, Y., and Aravind, L. (2014). New perspectives on the diversification of the RNA interference system: insights from comparative genomics and small RNA sequencing. *Wiley Interdiscip Rev RNA*, 5(2):141–181.

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15):1972–1973.

Cogoni, C. and Macino, G. (1999). Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature*, 399(6732):166–169.

Curaba, J. and Chen, X. (2008). Biochemical activities of Arabidopsis RNA-dependent RNA polymerase 6. *J Biol Chem*, 283(6):3059–3066.

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, 27(8):1164–1165.

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform*, 23(1):205–211.

Elbashir, S. M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev*, 15(2):188–200.

Feltzin, V. L., Khaladkar, M., Abe, M., Parisi, M., Hendriks, G. J., Kim, J., and Bonini, N. M. (2015). The exonuclease Nibbler regulates age-associated traits and modulates piRNA length in *Drosophila*. *Aging Cell*, 14(3):443–452.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5´-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature*, 465(7299):818–822.

Fu, Y., Yang, Y., Zhang, H., Farley, G., Wang, J., Quarles, K. A., Weng, Z., and Zamore, P. D. (2018). The genome of the Hi5 germ cell line from *Trichoplusia ni*, an agricultural pest and novel model for small RNA biology. *Elife*, 7:e31628.

Fuentes, M., Benito, E., Bertrand, S., Paris, M., Mignardot, A., Godoy, L., Jimenez-Delgado, S., Oliveri, D., Candiani, S., Hirsinger, E., D'Aniello, S., Pascual-Anaya, J., Maeso, I., Pestarino, M., Vernier, P., Nicolas, J. F., Schubert, M., Laudet, V., Genevière, A. M., Albalat, R., Garcia, Fernandez, J., Holland, N. D., and Escrivá, H. (2007). Insights into spawning behavior and development of the European amphioxus (*Branchiostoma lanceolatum*). *J Exp Zool B Mol Dev Evol*, 308(4):484–493.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L., Zapp, M. L., Weng, Z., and Zamore, P. D. (2008). Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science*, 320(5879):1077–1081.

Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., and Zamore, P. D. (2010). Sorting of *Drosophila* small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA*, 16(1):43–56.

Ghildiyal, M. and Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nat Rev Genet*, 10(2):94–108.

Girard, A., Sachidanandam, R., Hannon, G. J., and Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, 442(7099):199–202.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B. J., Chiang, H. R., King, N., Degnan, B. M., Rokhsar, D. S., and Bartel, D. P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*, 455(7217):1193–1197.

Gu, W., Shirayama, M., Conte, D. J., Vasale, J., Batista, P. J., Claycomb, J. M., Moresco, J. J., Youngman, E. M., Keys, J., Stoltz, M. J., Chen, C. C., Chaves, D. A., Duan, S., Kasschau, K. D., Fahlgren, N., Yates, J. R., Mitani, S., Carrington, J. C., and Mello, C. C. (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol Cell*, 36(2):231–244.

Hall, I. M., Shankaranarayana, G. D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S. I. (2002). Establishment and maintenance of a heterochromatin domain. *Science*, 297(5590):2232–2237.

Hammond, T. M. and Keller, N. P. (2005). RNA silencing in *Aspergillus nidulans* is independent of RNA-dependent RNA polymerases. *Genetics*, 169(2):607–617.

Han, B. W., Hung, J. H., Weng, Z., Zamore, P. D., and Ameres, S. L. (2011). The 3´-to-5´ exoribonuclease Nibbler shapes the 3´ ends of microRNAs bound to *Drosophila* Argonaute1. *Curr Biol*, 21(22):1878–1887.

Han, T., Manoharan, A. P., Harkins, T. T., Bouffard, P., Fitzpatrick, C., Chu, D. S., Thierry-Mieg, D., Thierry-Mieg, J., and Kim, J. K. (2009). 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in *Caenorhabditis elegans*. *Proc Natl Acad Sci USA*, 106(44):18674–18679.

Hayashi, R., Schnabl, J., Handler, D., Mohn, F., Ameres, S. L., and Brennecke, J. (2016). Genetic and mechanistic diversity of piRNA 3´-end formation. *Nature*, 539(7630):588–592.

Hetzel, J., Duttke, S. H., Benner, C., and Chory, J. (2016). Nascent RNA sequencing reveals distinct features in plant transcription. *Proc Natl Acad Sci USA*, 113(43):12316–12321.

Horie, M., , Kobayashi, Y., Honda, T., Fujino, K., Akasaka, T., Kohl, C., Wibbelt, G., Mühldorfer, K., Kurth, A., Müller, M. A., Corman, V. M., Gillich, N., Suzuki, Y., Schwemmle, M., and Tomonaga, K. (2016). An RNA-dependent RNA polymerase gene in bat genomes derived from an ancient negative-strand RNA virus. *Sci Rep*, 6:25873.

Horwich, M. D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P. D. (2007). The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol*, 17(14):1265–1272.

Houwing, S., Kamminga, L. M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D. V., Blaser, H., Raz, E., Moens, C. B., Plasterk, R. H., Hannon, G. J., Draper, B. W., and Ketting, R. F. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell*, 129(1):69–82.

Huang, Y., Ji, L., Huang, Q., Vassylyev, D. G., Chen, X., and Ma, J. B. (2009). Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Nature*, 461(7265):823–827.

Hutvágner, G., McLachlan, J., Pasquinelli, A. E., Bálint, E., Tuschl, T., and Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science*, 293(5531):834–838.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Iyer, L. M., Koonin, E. V., and Aravind, L. (2003). Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct Biol*, 3:1.

Jorgensen, S. E., Buch, L. B., and Nierlich, D. P. (1969). Nucleoside triphosphate termini from RNA synthesized *in vivo* by *Escherichia coli*. *Science*, 164(3883):1067–1070.

Juven-Gershon, T. and Kadonaga, J. T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol*, 339(2):225–229.

Kamminga, L. M., Luteijn, M. J., den Broeder, M. J., Redl, S., Kaaij, L. J., Roovers, E. F., Ladurner, P., Berezikov, E., and Ketting, R. F. (2010). Hen1 is required for oocyte development and piRNA stability in zebrafish. *EMBO J*, 29(21):3688–3700.

Kirino, Y. and Mourelatos, Z. (2007a). Mouse Piwi-interacting RNAs are 2´-O-methylated at their 3´ termini. *Nat Struct Mol Biol*, 14(4):347–348.

Kirino, Y. and Mourelatos, Z. (2007b). The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA*, 13(9):1397–1401.

Kuhlmann, M., Borisova, B. E., Kaller, M., Larsson, P., Stach, D., Na, J., Eichinger, L., Lyko, F., Ambros, V., Söderbom, F., Hammann, C., and Nellen, W. (2005). Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. *Nucleic Acids Res*, 33(19):6405–6417.

Kuzmine, I., Gottlieb, P. A., and Martin, C. T. (2003). Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. *J Biol Chem*, 278(5):2819–2823.

Labbé, R. M., Irimia, M., Currie, K. W., Lin, A., Zhu, S. J., Brown, D. D., Ross, E. J., Voisin, V., Bader, G. D., Blencowe, B. J., and Pearson, B. J. (2012). A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells*, 30(8):1734–1745.

Lau, N. C., Seto, A. G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D. P., and Kingston, R. E. (2006). Characterization of the piRNA complex from rat testes. *Science*, 313(5785):363–367.

Lee, S. R. and Collins, K. (2007). Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. *Nat Struct Mol Biol*, 14(7):604–610.

Lewis, S. H., Quarles, K. A., Yang, Y., Tanguy, M., Frézal, L., Smith, S. A., Sharma, P. P., Cordaux, R., Gilbert, C., Giraud, I., Collins, D. H., Zamore, P. D., Miska, E. A., Sarkies, P., and Jiggins, F. M. (2018). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat Ecol Evol*, 2(1):174–181.

Li, H., Bowling, A. J., Gandra, P., Rangasamy, M., Pence, H. E., McEwan, R. E., Khajuria, C., Siegfried, B. D., and Narva, K. E. (2018). Systemic RNAi in western corn rootworm, *Diabrotica virgifera virgifera*, does not involve transitive pathways. *Insect Sci*, 25(1):45–56.

Liu, N., Abe, M., Sabin, L. R., Hendriks, G. J., Naqvi, A. S., Yu, Z., Cherry, S., and Bonini, N. M. (2011). The exoribonuclease Nibbler controls 3´ end processing of microRNAs in *Drosophila*. *Curr Biol*, 21(22):1888–1893.

Louis, A., Roest Crollius, H., and Robinson-Rechavi, M. (2012). How much does the amphioxus genome represent the ancestor of chordates? *Brief Funct Genomics*, 11(2):89–95.

Maida, Y., Yasukawa, M., Furuuchi, M., Lassmann, T., Possemato, R., Okamoto, N., Kasim, V., Hayashizaki, Y., Hahn, W. C., and Masutomi, K. (2009). An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. *Nature*, 461(7261):230–235.

Makeyev, E. V. and Bamford, D. H. (2002). Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol Cell*, 10(6):1417–1427.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Marker, S., Le, Mouël, A., Meyer, E., and Simon, M. (2010). Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in *Paramecium tetraurelia*. *Nucleic Acids Res*, 38(12):4092–4107.

Martienssen, R. and Moazed, D. (2015). RNAi and heterochromatin assembly. *Cold Spring Harb Perspect Biol*, 7(8):a019323.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G. J., and Qi, Y. (2008). Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5´ terminal nucleotide. *Cell*, 133(1):116–127.

Miller, M. A., Schwartz, T., Pickett, B. E., He, S., Klem, E. B., Scheuermann, R. H., Passarotti, M., Kaufman, S., and O'Leary, M. A. (2015). A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. *Evol Bioinform Online*, 11:43–48.

Miller, W. A., Bujarski, J. J., Dreher, T. W., and Hall, T. C. (1986). Minus-strand initiation by brome mosaic virus replicase within the 3´ tRNA-like structure of native and modified RNA templates. *J Mol Biol*, 187(4):537–546.

Montgomery, T. A., Howell, M. D., Cuperus, J. T., Li, D., Hansen, J. E., Alexander, A. L., Chapman, E. J., Fahlgren, N., Allen, E., and Carrington, J. C. (2008). Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell*, 133(1):128–141.

Motamedi, M. R., Verdel, A., Colmenares, S. U., Gerber, S. A., Gygi, S. P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell*, 119(6):789–802.

Pak, J. and Fire, A. (2007). Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science*, 315(5809):241–244.

Pélisson, A., Sarot, E., Payen-Groschêne, G., and Bucheton, A. (2007). A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. *J Virol*, 81(4):1951–1960.

Qian, X., Hamid, F. M., El, Sahili, A., Darwis, D. A., Wong, Y. H., Bhushan, S., Makeyev, E. V., and Lescar, J. (2016). Functional evolution in orthologous cell-encoded RNA-dependent RNA polymerases. *J Biol Chem*, 291(17):9295–9309.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*, 61(3):539–542.

Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D. P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, 127(6):1193–1207.

Saito, K., Nishida, K. M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M. C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev*, 20(16):2214–2222.

Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M. C. (2007). Pimet, the *Drosophila* homolog of HEN1, mediates 2´-O-methylation of Piwi- interacting RNAs at their 3´ ends. *Genes Dev*, 21(13):1603–1608.

Sánchez, R., Serra, F., Tárraga, J., Medina, I., Carbonell, J., Pulido, L., de María, A., Capella-Gutíerrez, S., Huerta-Cepas, J., Gabaldón, T., Dopazo, J., and Dopazo, H. (2011). Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Res*, 39(Web Server issue):W470–474.

Schiebel, W., Haas, B., Marinković, S., Klanner, A., and Sänger, H. L. (1993). RNA-directed RNA polymerase from tomato leaves. II. Catalytic *in vitro* properties. *J Biol Chem*, 268(16):11858–11867.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Seitz, H., Tushir, J. S., and Zamore, P. D. (2011). A 5´-uridine amplifies miRNA/miRNA* asymmetry in *Drosophila* by promoting RNA-induced silencing complex formation. *Silence*, 2:4.

Sigova, A., Rhind, N., and Zamore, P. D. (2004). A single Argonaute protein mediates both transcriptional and posttranscriptional silencing in *Schizosaccharomyces pombe*. *Genes Dev*, 18(19):2359–2367.

Sijen, T., Fleenor, J., Simmer, F., Thijssen, K. L., Parrish, S., Timmons, L., Plasterk, R. H., and Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell*, 107(4):465–476.

Sijen, T., Steiner, F. A., Thijssen, K. L., and Plasterk, R. H. (2007). Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science*, 315(5809):244–247. (but note that this article's scientific integrity has been seriously questioned: `https://pubpeer.com/publications/2B00E5BEB5B75B499550D03C15EFA4`).

Smardon, A., Spoerke, J. M., Stacey, S. C., Klein, M. E., Mackin, N., and Maine, E. M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr Biol*, 10(4):169–178.

Sonnhammer, E. L., Eddy, S. R., Birney, E., Bateman, A., and Durbin, R. (1998). Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res*, 26(1):320–322.

Stein, P., Svoboda, P., Anger, M., and Schultz, R. M. (2003). RNAi: mammalian oocytes do it without RNA-dependent RNA polymerase. *RNA*, 9(2):187–192.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., and Watanabe, Y. (2008). The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. *Plant Cell Physiol*, 49(4):493–500.

Tang, G., Reinhart, B. J., Bartel, D. P., and Zamore, P. D. (2003). A biochemical framework for RNA silencing in plants. *Genes Dev*, 17(1):49–63.

Tapial, J., Ha, K. C., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., Quesnel-Vallières, M., Permanyer, J., Sodaei, R., Marquez, Y., Cozzuto, L., Wang, X., Gómez-Velázquez, M., Rayon, T., Manzanares, M., Ponomarenko, J., Blencowe, B. J., and Irimia, M. (2017). An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res*, 27(10):1759–1768.

Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, 313(5785):320–324.

Vasale, J. J., Gu, W., Thivierge, C., Batista, P. J., Claycomb, J. M., Youngman, E. M., Duchaine, T. F., Mello, C. C., and Conte, Jr., D. (2010). Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. *Proc Natl Acad Sci USA*, 107(8):3582–3587.

Voinnet, O. (2008). Use, tolerance and avoidance of amplified RNA silencing by plants. *Trends Plant Sci*, 13(7):317–328.

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I., and Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, 297(5588):1833–1837.

Wang, H., Ma, Z., Niu, K., Xiao, Y., Wu, X., Pan, C., Zhao, Y., Wang, K., Zhang, Y., and Liu, N. (2016). Antagonistic roles of Nibbler and Hen1 in modulating piRNA 3´ ends in *Drosophila*. *Development*, 143(3):530–539.

Wassenegger, M. and Krczal, G. (2006). Nomenclature and functions of RNA-directed RNA polymerases. *Trends Plant Sci*, 11(3):142–151.

Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N., and Imai, H. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev*, 20(13):1732–1743.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Wu, C. W. and Goldthwait, D. A. (1969a). Studies of nucleotide binding to the ribonucleic acid polymerase by a fluoresence technique. *Biochemistry*, 8(11):4450–4458.

Wu, C. W. and Goldthwait, D. A. (1969b). Studies of nucleotide binding to the ribonucleic acid polymerase by equilibrium dialysis. *Biochemistry*, 8(11):4458–4464.

Yigit, E., Batista, P. J., Bei, Y., Pang, K. M., Chen, C. C., Tolia, N. H., Joshua-Tor, L., Mitani, S., Simard, M. J., and Mello, C. C. (2006). Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell*, 127(4):747–757.

Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell*, 118(1):57–68.

Zong, J., Yao, X., Yin, J., Zhang, D., and Ma, H. (2009). Evolution of the RNA-dependent RNA polymerase (RdRP) genes: duplications and possible losses before and after the divergence of major eukaryotic groups. *Gene*, 447(1):29–39.

Acknowledgments

The authors are grateful to Dr. Darryl Conte for helpful discussions and to Kazufumi Mochizuki and Phillip D. Zamore for critical reading of the manuscript. We thank the *B. lanceolatum* genome consortium for the assembly and annotation of the *B. lanceolatum* genome, Dr. Manuel Irimia for assistance in transcriptomics analyses, and Dr. Ferdinand Marlétaz for sharing unpublished data. This research was supported by an ATIP-Avenir grant from CNRS and Sanofi (to H.S.) and a post-doctoral fellowship from La Ligue contre le cancer (to N.P.). H.E.'s laboratory was supported by the CNRS and the ANR16-CE12-0008-01 and S.B. by the Institut Universitaire de France.

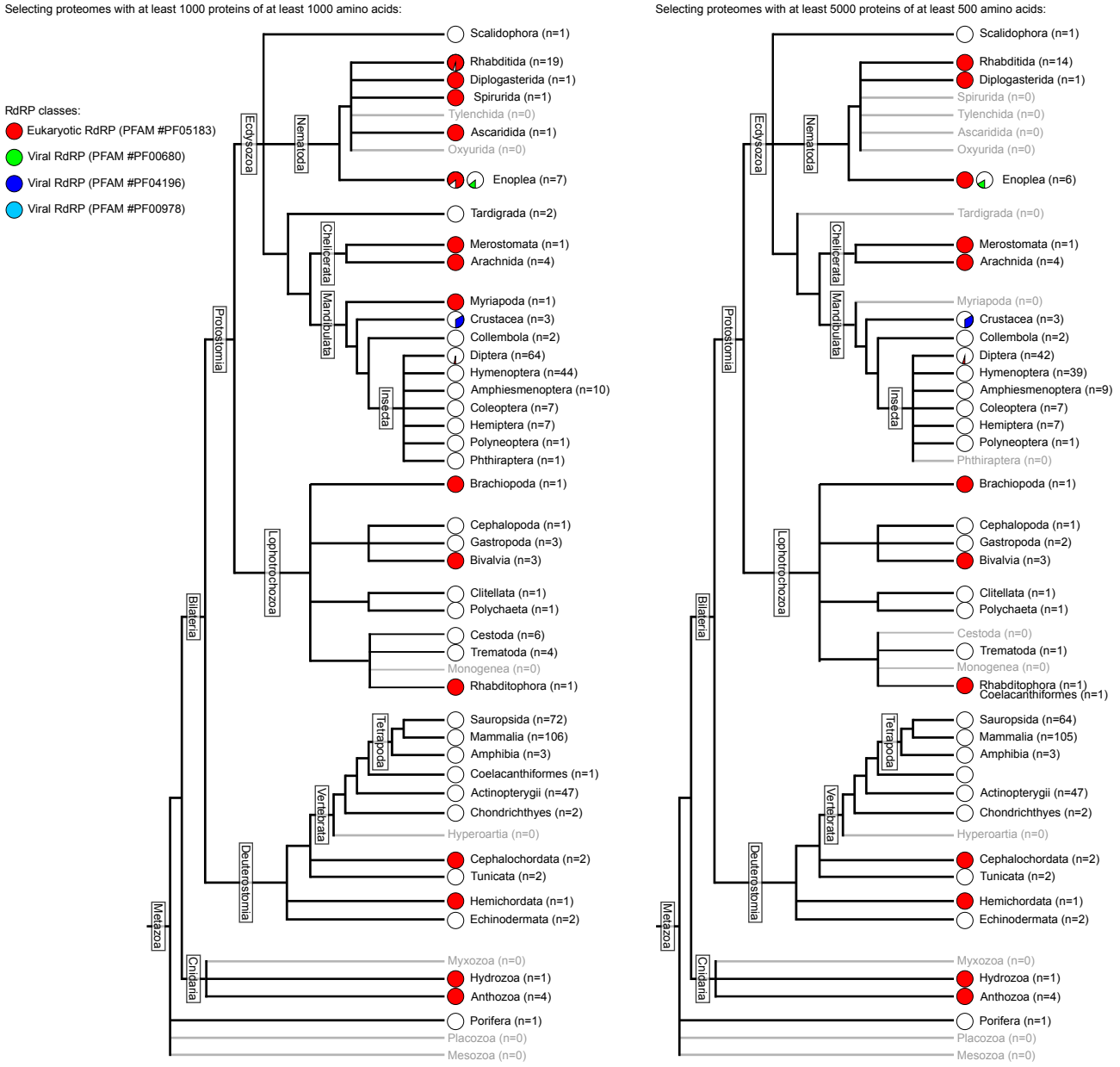
Author contributions

N.P. and I.B. performed the experiments, L.S. prepared the samples, S.B. and H.S. performed computational analyses, H.E. and H.S. supervised the project, and H.S. wrote the manuscript.

Competing interests

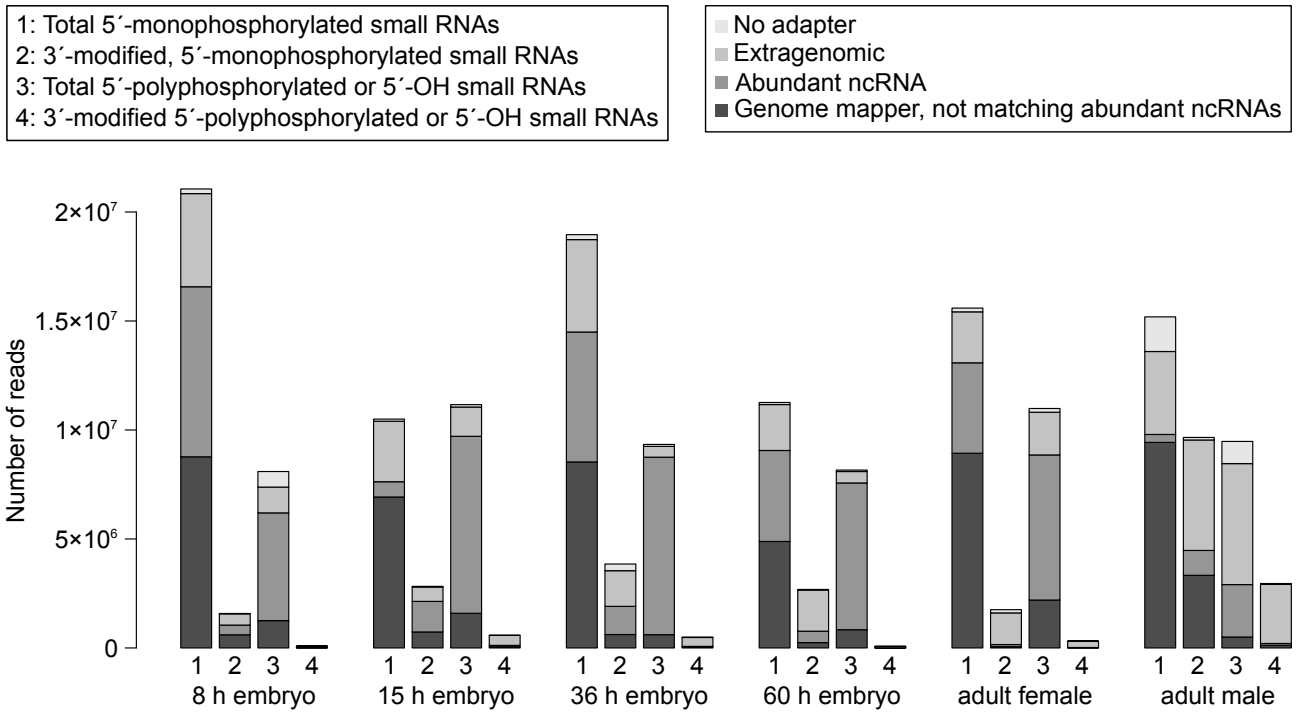
The authors declare no competing financial interests.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



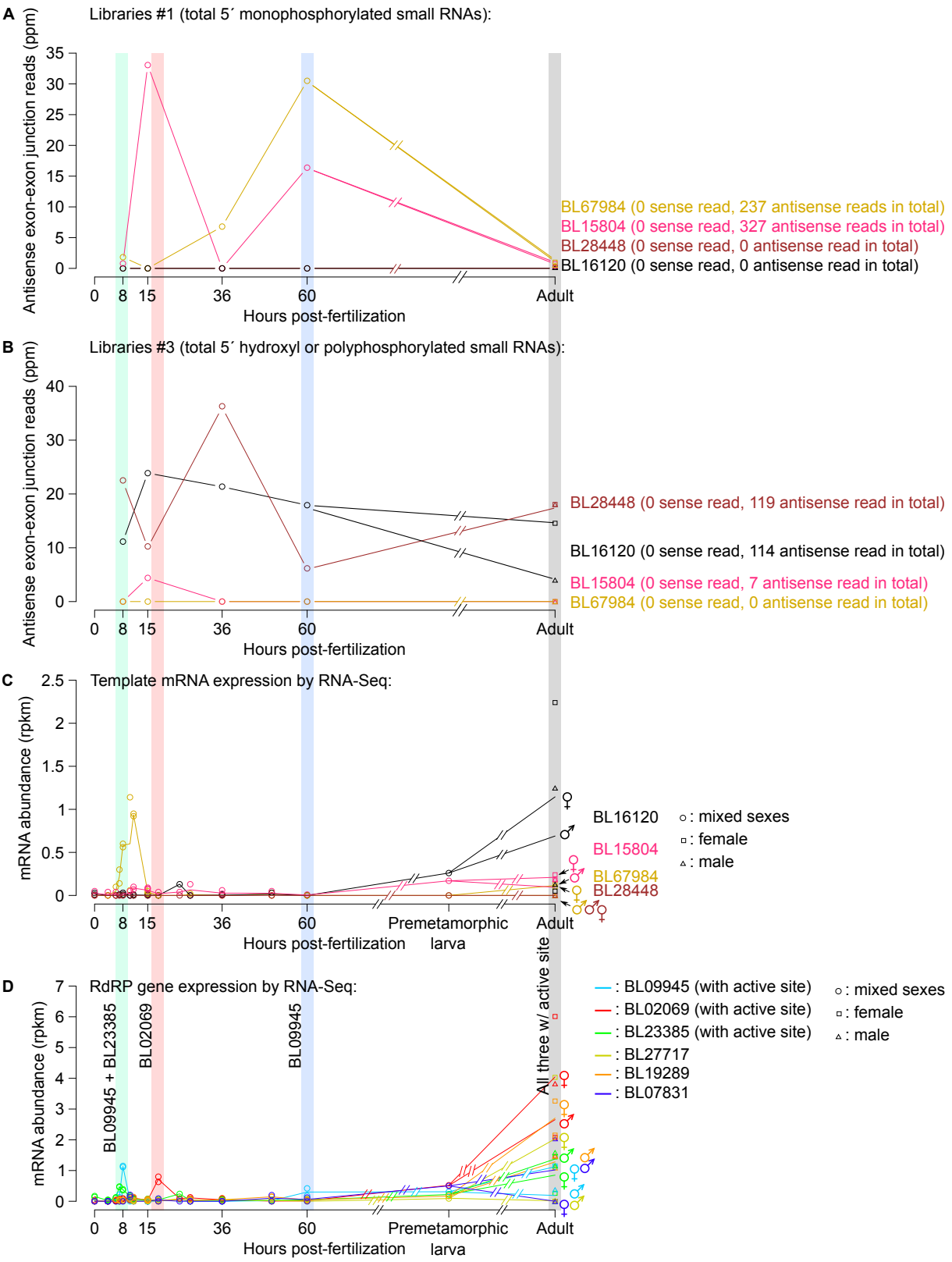
Supplementary Figure 1: **Exclusion of dubious proteomes still indicates many independent RdRP losses.** Among the 538 analyzed proteomes, 442 contain at least 1,000 proteins of at least 1,000 amino acids (left panel) and 383 contain at least 5,000 proteins of at least 500 amino acids (right panel). Selective analysis of these species does not fundamentally change the results shown in Figure 1**A**. Same conventions than in Figure 1**A**. Some clades analyzed in Figure 1**A** could not be analyzed here after proteome exclusion: they are shown in grey.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



Supplementary Figure 2: **Size and quality of the Small RNA-Seq libraries.** "No adapter" indicates that the 3´ adapter was not detected in the read. "Extragenomic" means that the adapter-trimmed read does not match on the *B. lanceolatum* genome assembly. "Abundant ncRNA" means that it maps on the genome assembly, on one of the genes for known abundant non-coding RNAs (rRNAs, tRNAs, snRNAs, snoRNAs, scaRNAs). "Genome mapper, not matching abundant ncRNAs" means that it maps elsewhere in the genome assembly.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



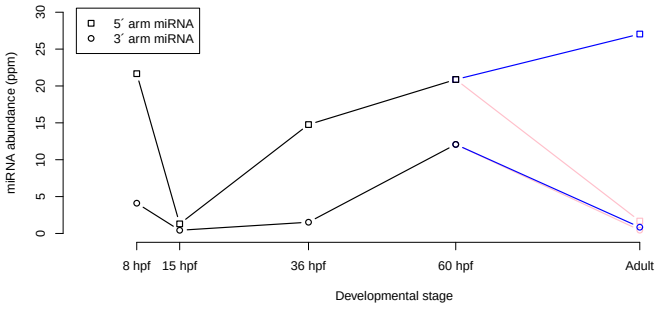
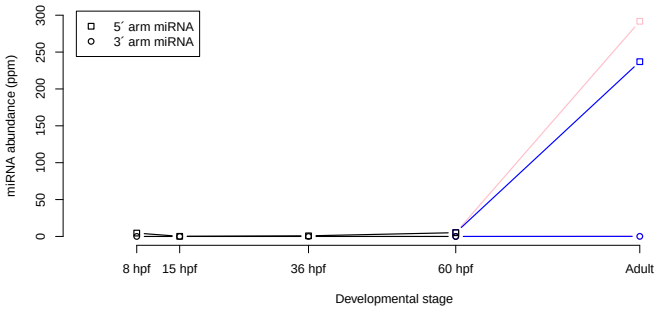
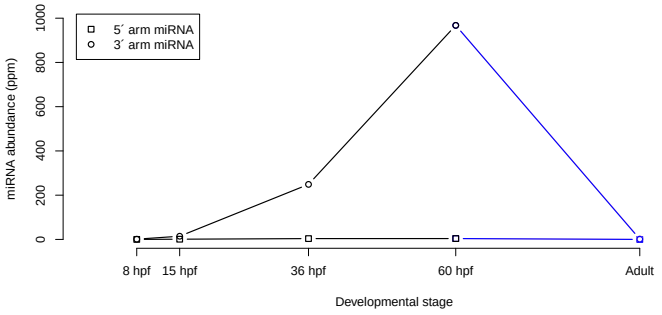
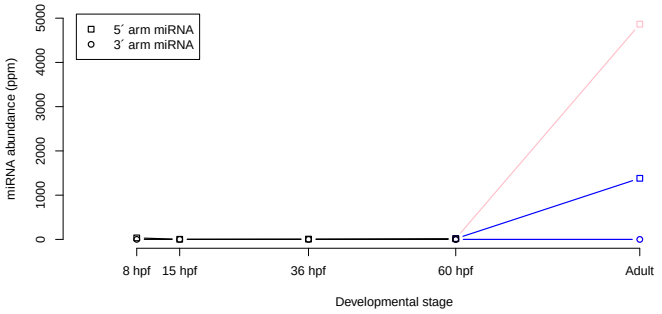
Supplementary Figure 3: **Specific transcripts are preferentially used as RdRP templates (separated sexes).** Same conventions than for Figure 6, but without averaging adult male and female data.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

```
                                                                R701
B. lanceolatum BL03504     D-----------------------EEEFGGPVFSPPLYRQRYQTVADLVK----KYRPKR
N. vectensis AGW15602      --------------------------REQLGPKFDPPVYRQRYHRVIEVVK----EHKAKR
D. rerio NP_001017842      ------------------------------ATPFSPPLYMQRYQFVIDYVK----TYRPRK
M. musculus NP_001072114   E-----------------------VSPEKVIRFKPPLYKQRYQFVRDLVD----RHEPKK
A. thaliana NP_567616      IRSLLSERPCLNYNILLLGVKGPSEERMEAAFFKPPLSKQRVEYALKHIR----ESSAST
D. melanogaster NP_610732  -----------------------KMTETGITFDPPVYEQRYCATIQILEDARWKDQIKK

B. lanceolatum BL03504     LVDFGCAEGKLIRFLK-PEESLEQLTGIDLEGEVLESIRGIIKPLLSDYVQPRPRPFTVS
N. vectensis AGW15602      VLDFGCAEAKMLRSLINSTTNIEELVGVDIDRDLLEDSIFRIRPLTTDYLTPRPHPLAVS
D. rerio NP_001017842      VIDFGCAECCLLKKLKFHRNGIQLLVGVDINSVVLLKRMHSLAPLVSDYLQPSDGPLTIE
M. musculus NP_001072114   VADLGCGDAKLLKLLKI-YPCIQLLVGVDINEEKLHSNGHRLSPYLGEFVKPRDLDLTVT
A. thaliana NP_567616      LVDFGCGSGSLLDSLLDYPTSLQTIIGVDISPKGLARAAKMLHVKLN---KEACNVKSAT
D. melanogaster NP_610732  VVEFGCAEMRFFQLMR-RIETIEHIGLVDIDKSLLMRNLTSVNPLVSDYIRSRASPLKVQ
                                          E796      E799, H800
B. lanceolatum BL03504     LYQGSIAECDDRFKSYDMVTCVEVIEHLDPPVLDAMPSNVFGHMRPSVVVVTTPNSEFNV
N. vectensis AGW15602      LYQGSISKADDRFCDFDVVACIEIVEHLVPEHLEAMPAVLLGQLSPLVAIVTTPNADFNV
D. rerio NP_001017842      LYQGSVMEREPCTKGFDLVTCVELIEHLELEEVERFSEVVFGYMAPGAVIVTTPNAEFNP
M. musculus NP_001072114   LYHGSVVERDSRLLGFDLITCIELIEHLDSDDLARFPDVVFGYLSPAMVVISTPNAEFNP
A. thaliana NP_567616      LYDGSILEFDSRLHDVDIGTCLEVIEHMEEDQACEFGEKVLSLFHPKLLIVSTPNYEFNT
D. melanogaster NP_610732  ILQGNVADSSEELRDTDAVIAIELIEHVYDDVLAKIPVNIFGFMQPKLVVFSTPNSDFNV
                                                   H860
B. lanceolatum BL03504     LFPN--------------F-----SGFRNADHRFEWTRQEFQTWAEGVAQRF-SYDVTFH
N. vectensis AGW15602      LFPD--------------L-----VGFRHWDHKFEWTRAEFKDWATSQADKF-GYSVTFE
D. rerio NP_001017842      LLPG--------------L-----RGFRNYGHKFEWTRAEFQTWAHRVCREH-GYSVQFT
M. musculus NP_001072114   LFPT--------------V------TLRDADHKFEWSRMEFQTWALHVANCY-NYRVEFT
A. thaliana NP_567616      ILQRSTPETQEENNSEPQL-----PKFRNHDHKFEWTREQFNQWASKLGKRH-NYSVEFS
D. melanogaster NP_610732  IFTR--------------FNPLLPNGFRHEDHKFEWSRDEFKNWCLGIVEKYPNYMFSLT

B. lanceolatum BL03504     GIGTGPEGTEHLGCCTQMAIFERKQTPYDE--------N-STVLWGTPYELIAEAVFPYR
N. vectensis AGW15602      GIGSGPSGTEHLGCCSQMALFIKQNTAPA---------G-RQTGFGEPYNLIARVEHPYR
D. rerio NP_001017842      GVGEAAGHWRDVGFCTQIAVFQRNFDGVNRSMS-----N-AEHLEPSVYRLLYRVVYPSL
M. musculus NP_001072114   GVGTPPAGSEHVGYCTQIGVFTKNGGKLSK-PS-----V-SQQCDQHVYKPVYTTSYPSL
A. thaliana NP_567616      GVGGS--GEVEPGFASQIAIFRREASSVE---------N-VAESSMQPYKVIWEWKKEDV
D. melanogaster NP_610732  GVGNPPKEYESVGPVSQIAIFVRKDMLEMQ-LVNPLVSKPNIDKESIPYKLIHTVEYPFY

B. lanceolatum BL03504     ENTLSKEQQILQEVQYYIRQIM--QRRVHGEDEEKD---NADDDDSAPE-----------
N. vectensis AGW15602      KCTLTEEEKILIELDRTLWFLS--QPSA-YEDDEISD--SEDLGD--DK-----------
D. rerio NP_001017842      CDNNIYQKTLINEVLYEAQHLR--QQWL-IRENMNNN----AHFYS---P-PLMEALHHG
M. musculus NP_001072114   QQEKVLKFVLVGELLIQVDRLRLRYQRM-LRDREKDRGPKPGDMDSCPAPHLLLGAVFTE
A. thaliana NP_567616      ----------------------------------EKKK-----------------------
D. melanogaster NP_610732  VDTRTEKEKLWTEVQIELQRFK--RQF---ESSEIEE---GTYQDT--------------
```

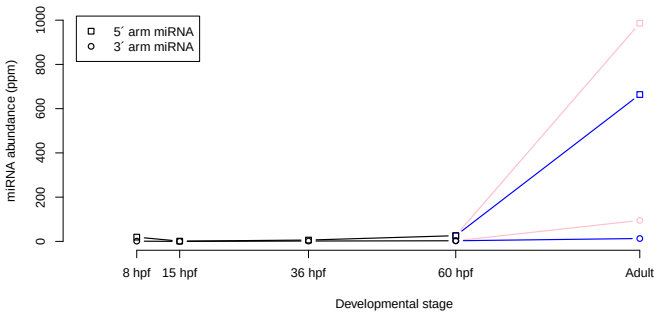
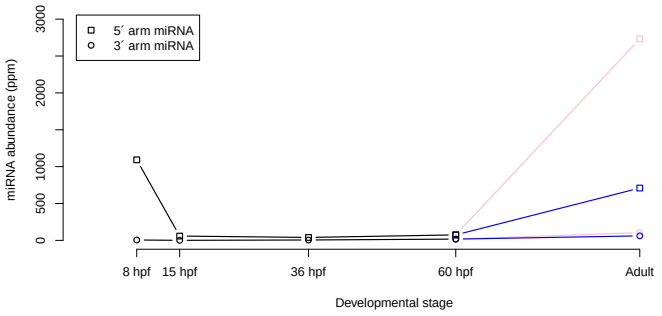
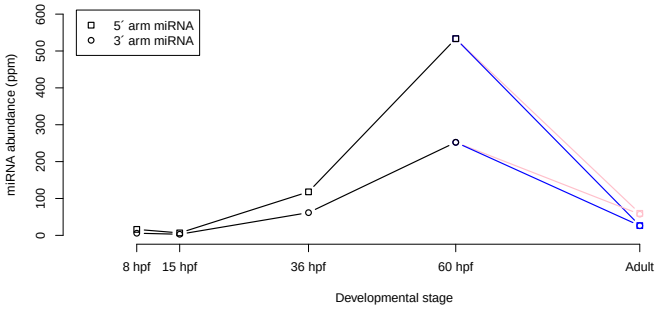
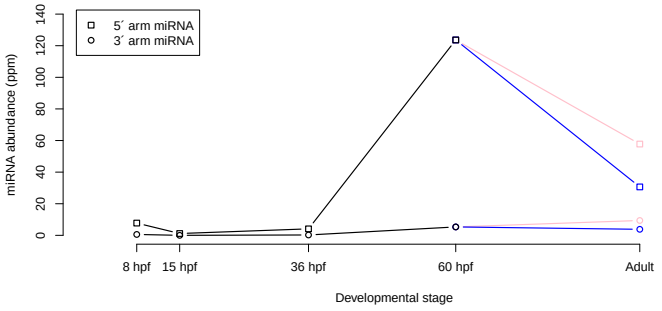
Supplementary Figure 4: **A *Branchiostoma* Hen1 candidate contains the known essential amino acids for Hen1 activity.** Sequences of 5 known Hen1 proteins (from *Nematostella vectensis*, *Danio rerio*, *Mus musculus*, *Arabidopsis thaliana* and *Drosophila melanogaster*) were aligned with the identified *Branchiostoma lanceolatum* Hen1 candidate (only the part of the alignment spanning amino acids 661–939 of the *Arabidopsis* protein is shown). Alignment was performed with t-coffee (version 11.00.8cbe486); other alignment programs (Clustal Omega v.1.2.4, t-coffee v.8.93, Kalign v.2.03, MAFFT v.7.215, but not muscle v.3.8.31) give the same main result: amino acids and amino acid combinations required for Hen1 catalytic activity (Huang et al., 2009) are conserved in the *Branchiostoma* candidate. Amino acids boxed in red were shown to be essential for *Arabodipsis* Hen1 activity; in orange: amino acids whose absence affects Hen1 activity without abolishing it entirely. Amino acid numbering is based on the *Arabidopsis* sequence.

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bfl-mir-71 ortholog at Sc0000005 bp 2992213-2992115 (- strand)	5´arm: UGAAAGACAUGGGUAGUGAGAU 3´arm: CCCAUUUCCCUGUCUUUCAAC	
bfl-mir-4856b ortholog at Sc0000022 bp 2102339-2102425 (+ strand)	5´arm: UCGCAUUGACGUCAGCGCCGUU 3´arm: (low abundance)	
bfl-mir-4876 ortholog at Sc0000005 bp 2219860-2219951 (+ strand)	5´arm: (low abundance) 3´arm: CUUACGUGCCACGUGAGGACUCU	
bfl-mir-10c ortholog at Sc0000000 bp 2027699-2027621 (- strand)	5´arm: UACCCUGUAGAUCCGGACUUGUGA 3´arm: (low abundance)	

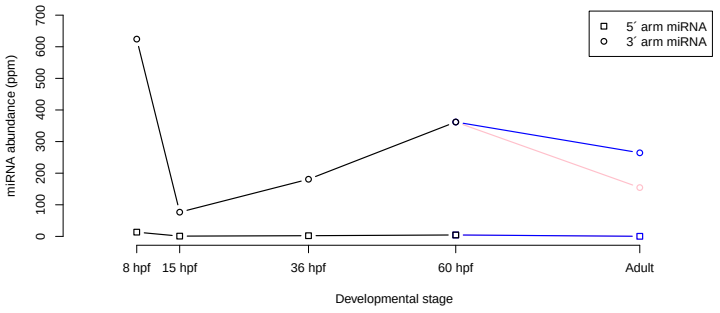
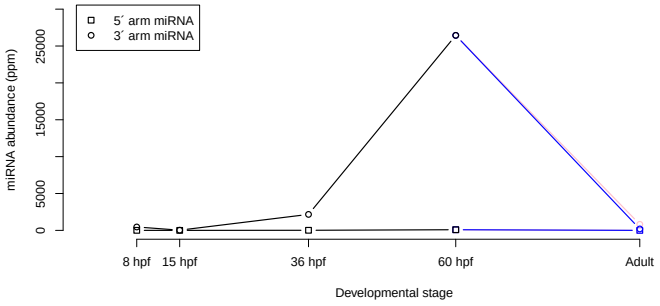
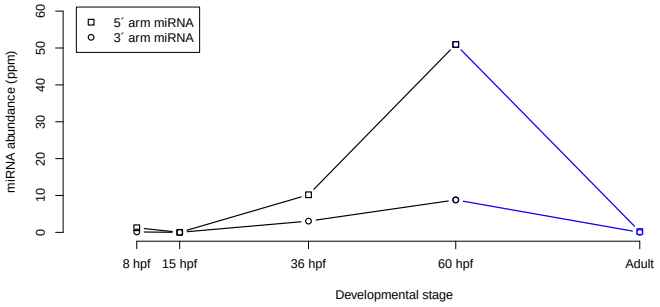
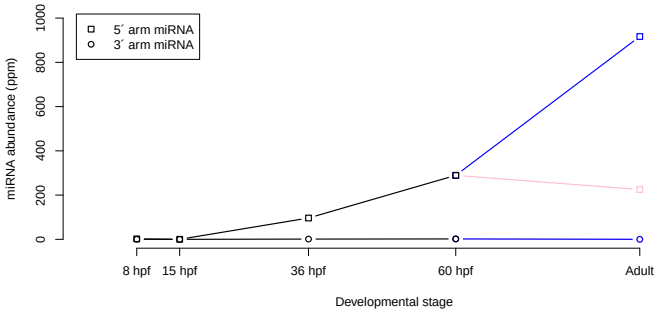
(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bfl-mir-4869 ortholog at Sc0000022 bp 2099003-2099080 (+ strand)	5´arm: ACGAUGUUGACUUCCGCUCCUCU 3´arm: GGGACCUGUAGUCAACACGAGA	
bbe-mir-125a ortholog at Sc0000265 bp 98448-98548 (+ strand)	5´arm: UCCCUGAGACCCUAACUUGUGA 3´arm: ACAGGUUAGGAUCUUGGGAGCU	
bbe-mir-4874 ortholog at Sc0000043 bp 1147175-1147086 (- strand)	5´arm: AGUUUGUAGCAUCCAGCUGAGCU 3´arm: CUGGCUGGUUGGUGCAAACAGG	
bbe-mir-281 ortholog at Sc0000000 bp 9083742-9083658 (- strand)	5´arm: AGGAGAGCCGUUCUGUGACUGU 3´arm: (low abundance)	

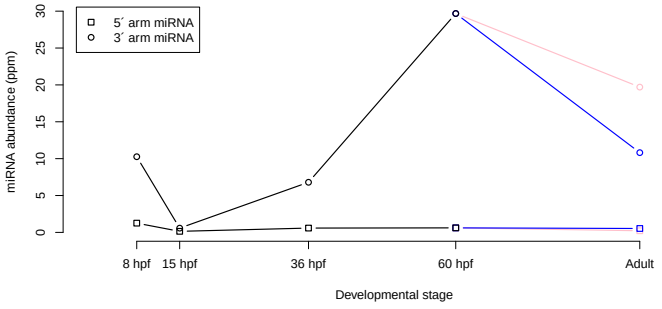
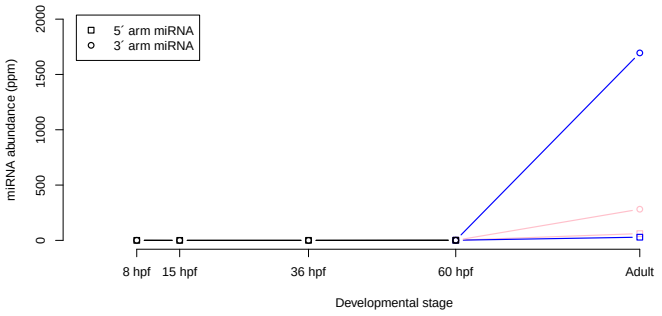
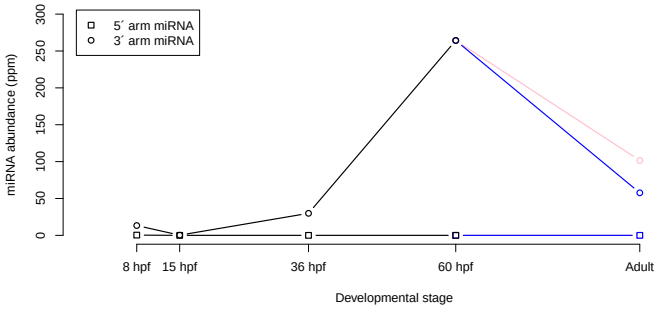
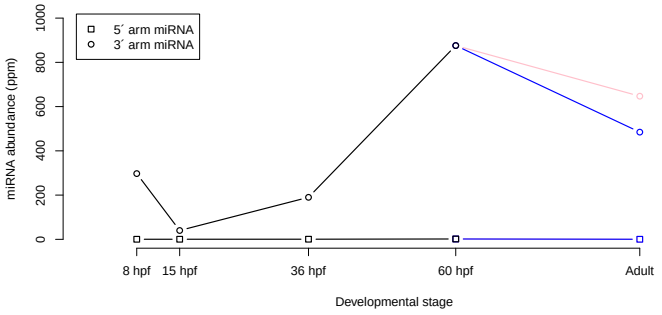
(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bbe-mir-92a-2 ortholog at Sc0000007 bp 4735603-4735697 (+ strand)	5´arm: AGGCCAGGAUUGGUGGCAAUGCC 3´arm: UAUUGCACUUGUCCCGGCCUUU	
bbe-mir-2056 ortholog at Sc0000086 bp 322130-322049 (- strand)	5´arm: CAGGUAUGUCUGCGGUGAGGCU 3´arm: UUUCACUGUAGAUCUACCUGCG	
bbe-mir-4880 ortholog at Sc0000057 bp 1180856-1180942 (+ strand)	5´arm: UUUGCUAUUCGAUGACCAGUGG 3´arm: (low abundance)	
bbe-mir-31 ortholog at Sc0000399 bp 174399-174498 (+ strand)	5´arm: UGGCAAGAUGUUGGCAUAGCUG 3´arm: (low abundance)	

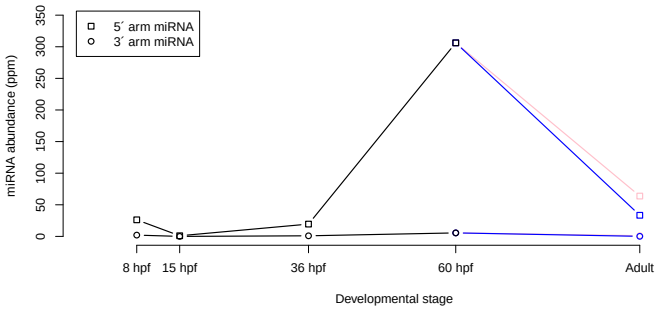
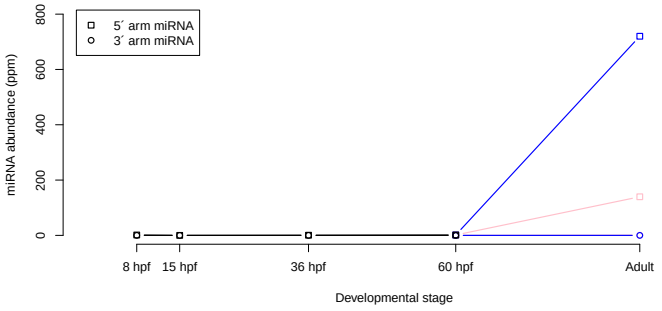
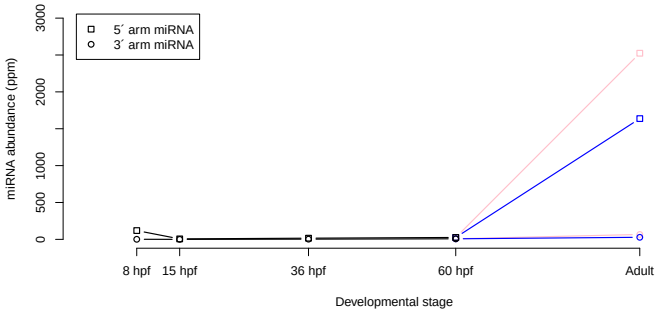
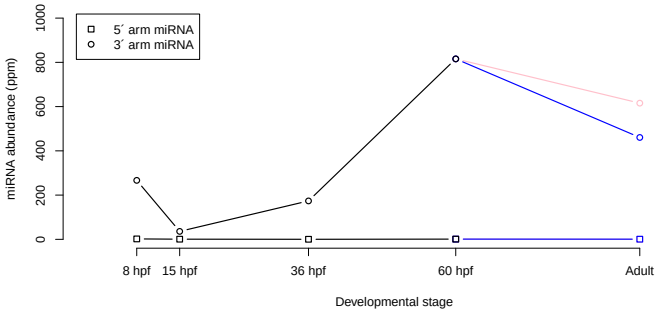
(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bfl-mir-200b ortholog at Sc0000010 bp 3877902-3877999 (+ strand)	5´arm: `(low abundance)` 3´arm: `UAAUACUGUCUGGUAAUGAUGUU`	
bfl-mir-2071 ortholog at Sc0000288 bp 188813-188893 (+ strand)	5´arm: `AUGCGGUGCGGUGGUAGCAACCG` 3´arm: `AUUGUUACACCGCGCCGCAAAAG`	
bfl-mir-4865 ortholog at Sc0000063 bp 315717-315798 (+ strand)	5´arm: `(low abundance)` 3´arm: `UGUAGAGAGAGUGACAGGUUGU`	
bbe-mir-200c ortholog at Sc0000010 bp 3875272-3875359 (+ strand)	5´arm: `(low abundance)` 3´arm: `UAACACUGUCUGGUAAUGAUG`	

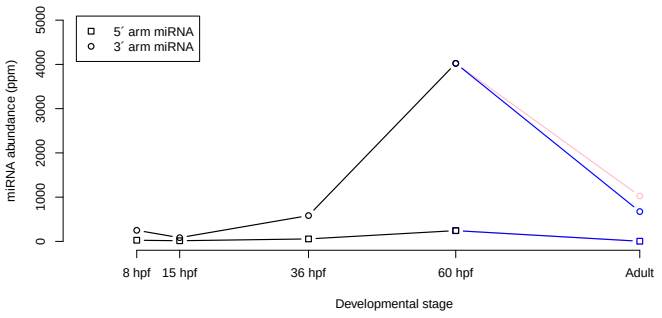
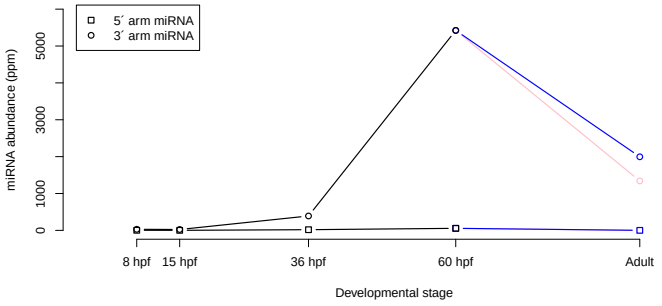
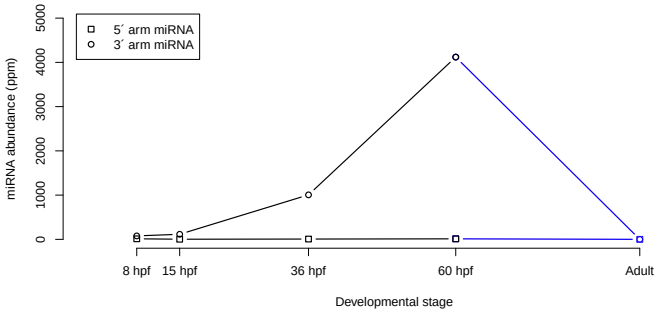
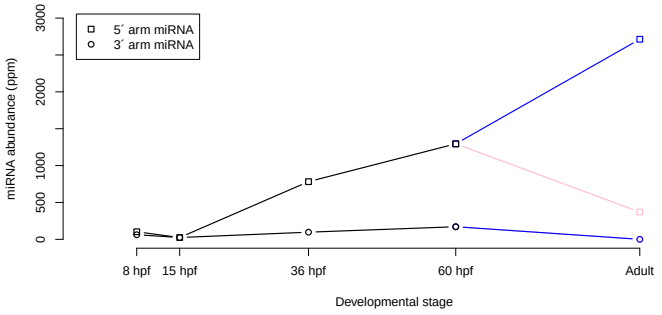
(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bfl-mir-29a ortholog at Sc0000043 bp 1849481-1849575 (+ strand)	5´arm: `GCUGAUUUCAGUUGGUGCUAGA` 3´arm: `(low abundance)`	
bbe-mir-2058 ortholog at Sc0000110 bp 921627-921549 (- strand)	5´arm: `UGAGAAGUAAGACUACCAUCCCGU` 3´arm: `(low abundance)`	
bfl-let-7a-2 ortholog at Sc0000265 bp 98758-98852 (+ strand)	5´arm: `UGAGGUAGUAGGUUGUAUAGUU` 3´arm: `CUGUGCAACCUGCUAGCUCUCC`	
bfl-mir-200c ortholog at Sc0000010 bp 3875267-3875358 (+ strand)	5´arm: `(low abundance)` 3´arm: `UAACACUGUCUGGUAAUGAUG`	

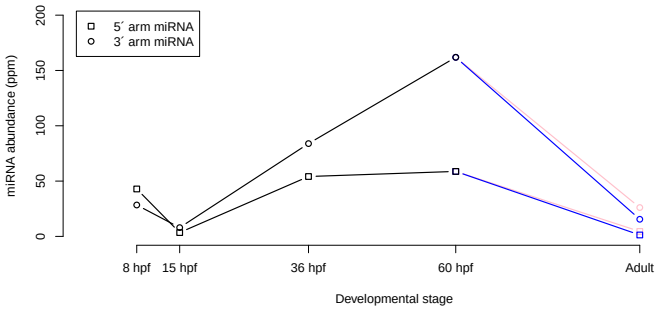
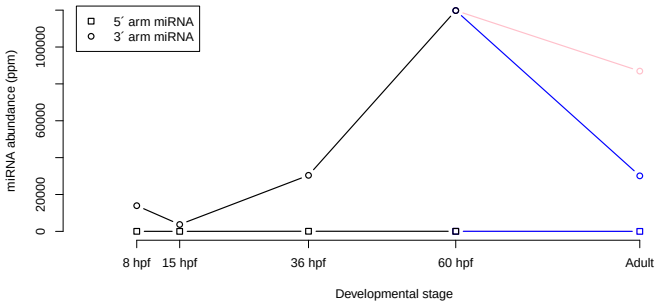
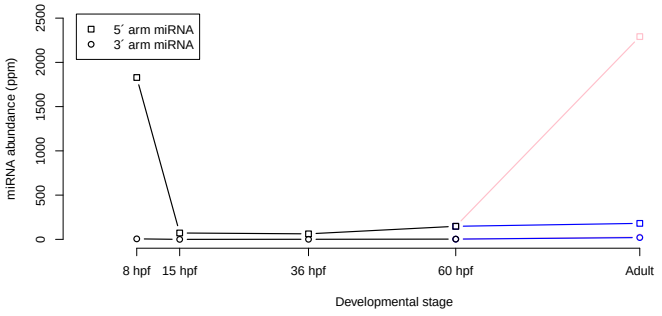
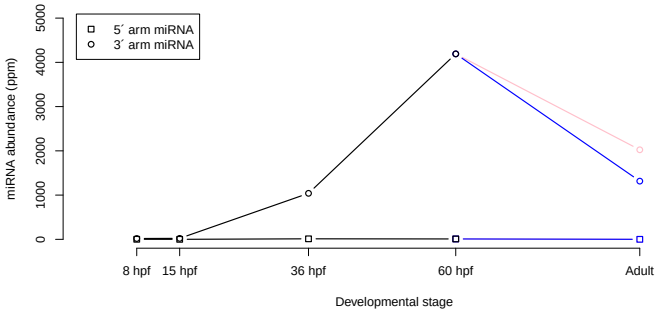
(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bfl-mir-4860 ortholog at Sc0000063 bp 316331-316404 (+ strand)	5´arm: UGCCUGUCAACGUCUCUGUACA 3´arm: UGUAGAGAUUGUGUGACGGGUAGU	
bfl-mir-4864 ortholog at Sc0000184 bp 79034-79113 (+ strand)	5´arm: AGGGAGAUCGUCUCGGGCAUACA 3´arm: UAGCCAGACCUGAUCUCCCUGC	
bfl-mir-4861 ortholog at Sc0000005 bp 3994270-3994362 (+ strand)	5´arm: AGCCAAUGCGGCAUGUAAAAGGC 3´arm: UUUACGUGCCACAUUGUCUCCU	
bfl-mir-2062 ortholog at Sc0000099 bp 755622-755703 (+ strand)	5´arm: UGCAACAAUAUUUCAUCAGUGG 3´arm: ACUGGUGAAAUGUAGUUGCGUA	

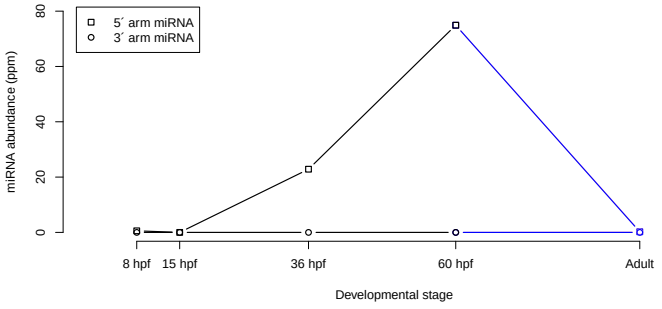
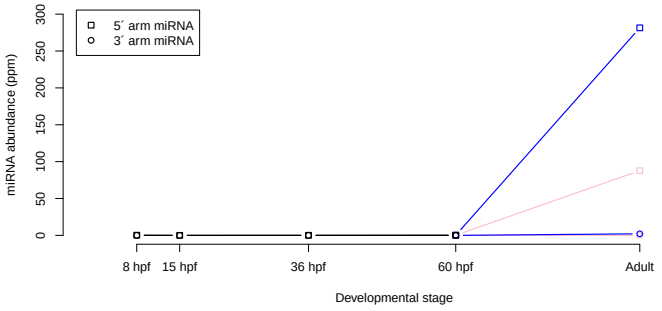
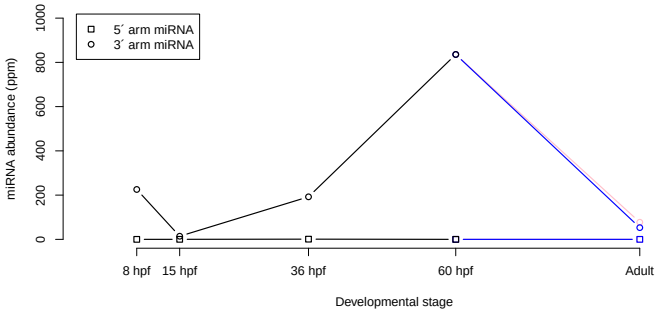
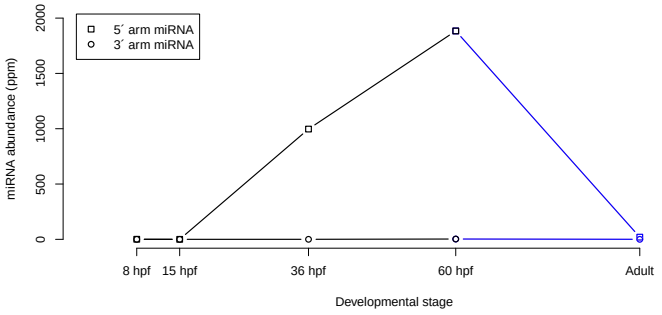
(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bbe-mir-4859 ortholog at Sc0000221 bp 303532-303447 (- strand)	5´arm: AGCAGCGAGCAUUACGGUCAUU 3´arm: UGACAGUAAUGCCCGCUGACUU	
bfl-mir-4871 ortholog at Sc0000001 bp 4527403-4527475 (+ strand)	5´arm: UCUGAAGUACCUGUUGCCAAAGG 3´arm: UUUGGCACUGGUACUUUGGAGU	
bbe-mir-100 ortholog at Sc0000265 bp 96682-96781 (+ strand)	5´arm: AACCCGUAGAUCCGAACUUGUGU 3´arm: CAAGCUCGUGUCUAUGGGUCU	
bbe-mir-22 ortholog at Sc0000015 bp 676722-676797 (+ strand)	5´arm: AGCUCUUCACUCGGUAGCUCUG 3´arm: AAGCUGCCAGAUGAAGAGCUGU	

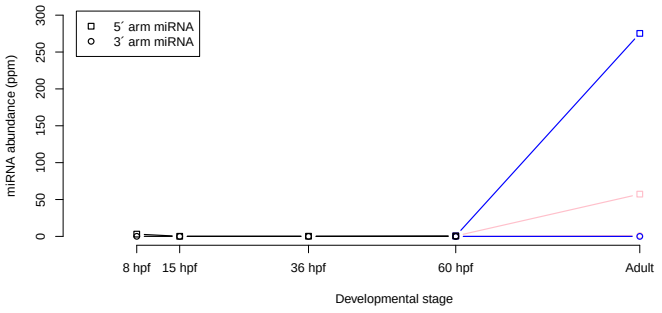
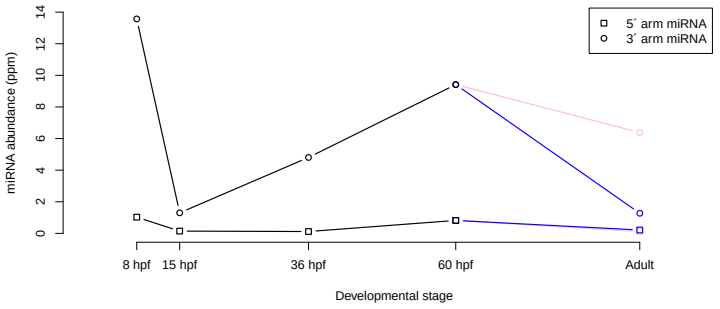
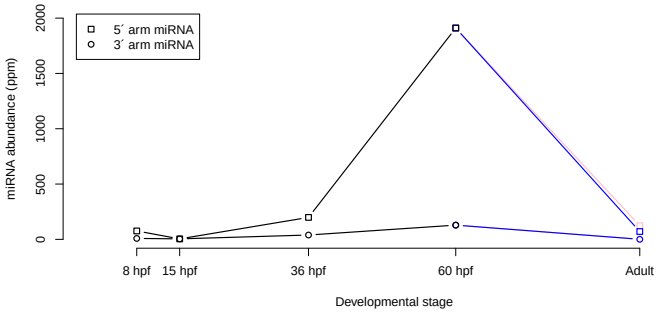
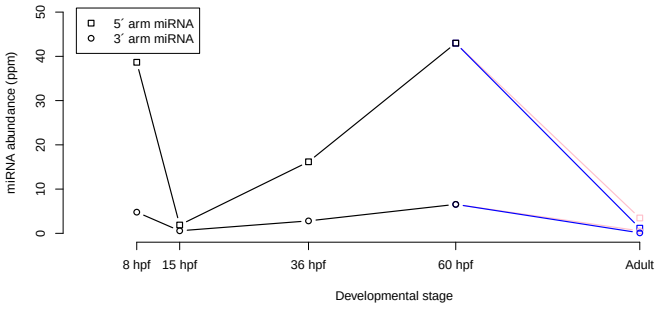
(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bfl-mir-190 ortholog at Sc0000166 bp 638463-638558 (+ strand)	5´arm: `UGAUAUGUUUGAUAUUUGGUUG` 3´arm: `(low abundance)`	
bbe-mir-2057 ortholog at Sc0000110 bp 922600-922519 (- strand)	5´arm: `UGAGAAGUUAGCCAACCAUCCGG` 3´arm: `(low abundance)`	
bbe-mir-2061 ortholog at Sc0000005 bp 2547710-2547629 (- strand)	5´arm: `(low abundance)` 3´arm: `UUGCAUAGGUACAUUGGUCAGU`	
bfl-mir-124 ortholog at Sc0000076 bp 963692-963792 (+ strand)	5´arm: `AGUGUUCACGGCGGUCCUUAAU` 3´arm: `(low abundance)`	

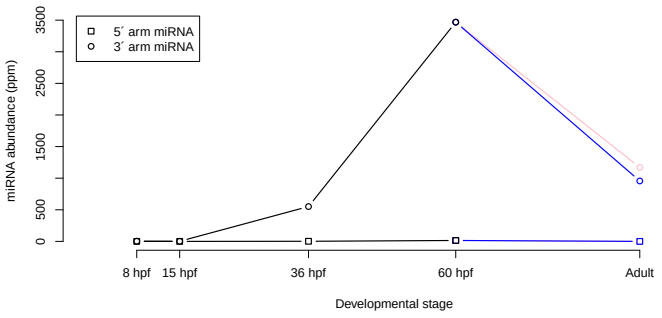
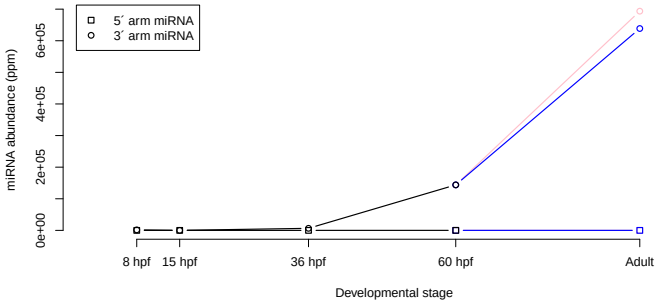
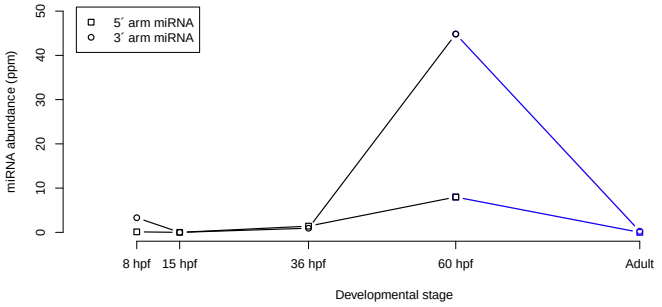
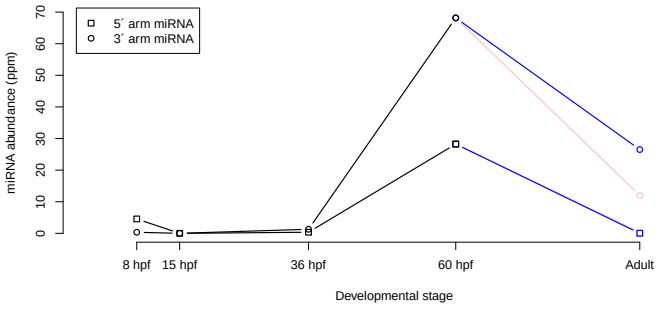
(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bfl-mir-4873 ortholog at Sc0000015 bp 1919187-1919110 (- strand)	5´arm: UGUUCCCACCUUCUGAUGUUGUU 3´arm: (low abundance)	
bfl-mir-4891 ortholog at Sc0000002 bp 2469330-2469421 (+ strand)	5´arm: (low abundance) 3´arm: CGUACCAGGACGCUCGUCUGCC	
bfl-mir-2070 ortholog at Sc0000079 bp 1008448-1008530 (+ strand)	5´arm: UUUCCACAGCCUCUACACAUGU 3´arm: AUGUGCAUAAGCUGUGGGAGCA	
bfl-mir-2076 ortholog at Sc0000004 bp 4642959-4643036 (+ strand)	5´arm: AAUUGCACUAGAGUGAUUUGUU 3´arm: (low abundance)	

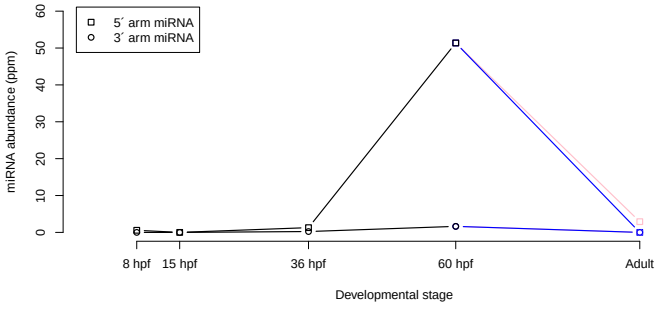
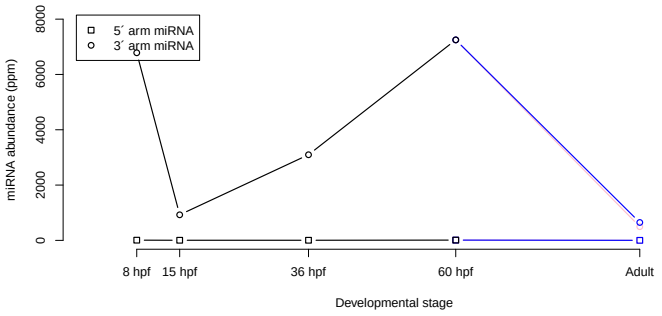
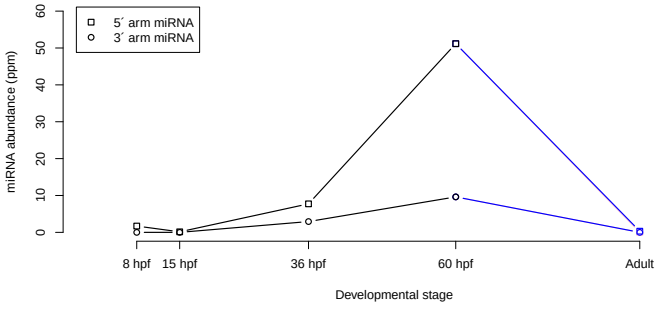
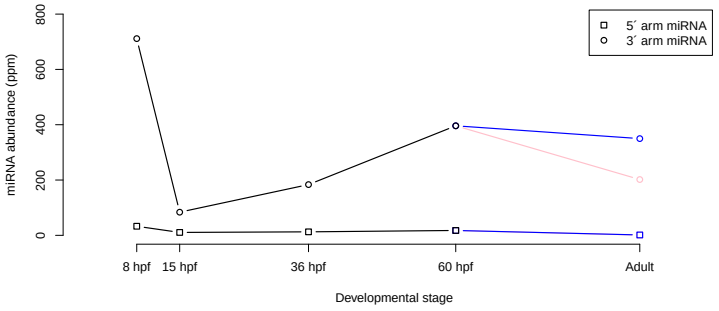
(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bbe-mir-133 ortholog at Sc0000092 bp 690262-690162 (- strand)	5´arm: AAAGCUGGUAAAUUGGAACCA 3´arm: UUGGUCCCCUUCAACCAGCUGU	
bfl-mir-92b ortholog at Sc0000007 bp 2890460-2890560 (+ strand)	5´arm: AGGUCUGGACAGUUGCAAUCUU 3´arm: CAUUGCACUCGUCCCGGCCUGA	
bfl-mir-137 ortholog at xpSc0039671 bp 309878-309778 (- strand)	5´arm: (low abundance) 3´arm: UAUUGCUUGAGAAUACACGUGA	
bfl-mir-4868b ortholog at Sc0000017 bp 1346380-1346293 (- strand)	5´arm: CUCAUCACACCGGAAGCUGUUA 3´arm: UCAGCUCCAGCUGUGAUGAGUG	

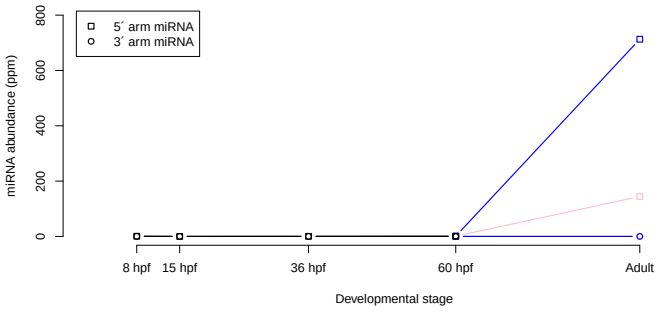
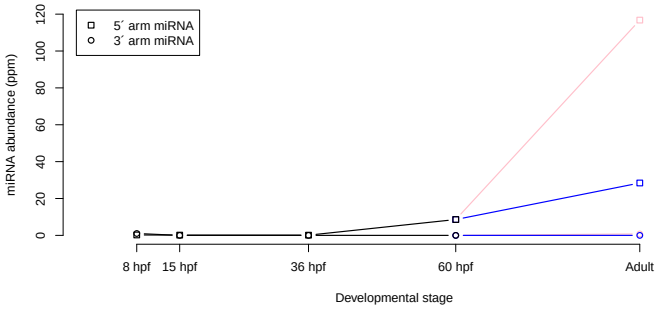
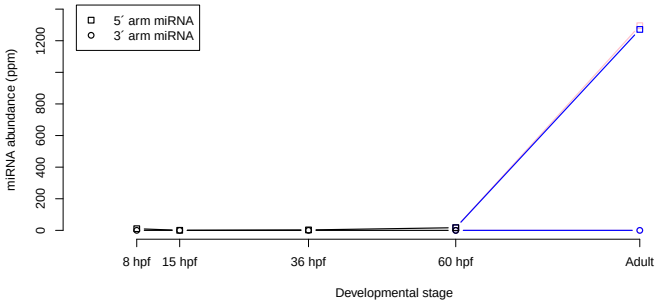
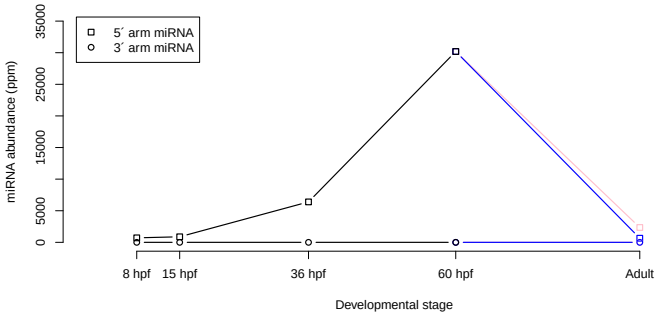
(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bfl-mir-242 ortholog at Sc0000023 bp 500019-499937 (- strand)	5´arm: `UUGCGUAGGCGUUGUGCACACU` 3´arm: `(low abundance)`	
bfl-mir-92d ortholog at Sc0000062 bp 529331-529413 (+ strand)	5´arm: `(low abundance)` 3´arm: `UAUUGCACUUAUCCUGGCCUGU`	
bfl-mir-4880 ortholog at Sc0000057 bp 1180852-1180940 (+ strand)	5´arm: `UUUGCUAUUCGAUGACCAGUGG` 3´arm: `(low abundance)`	
bbe-mir-92a-1 ortholog at Sc0000007 bp 4735908-4735986 (+ strand)	5´arm: `AGGUCGUGAUAGGCGGCAAUGUU` 3´arm: `UAUUGCACUUGUCCCGGCCUUU`	

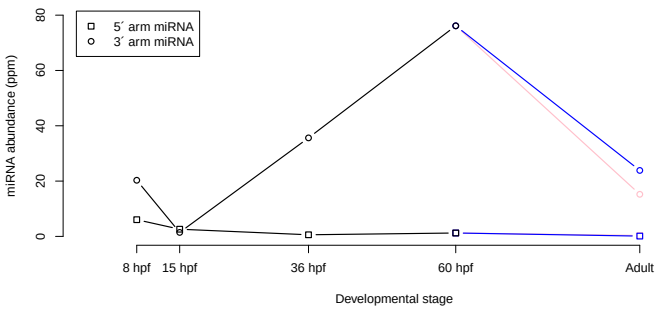
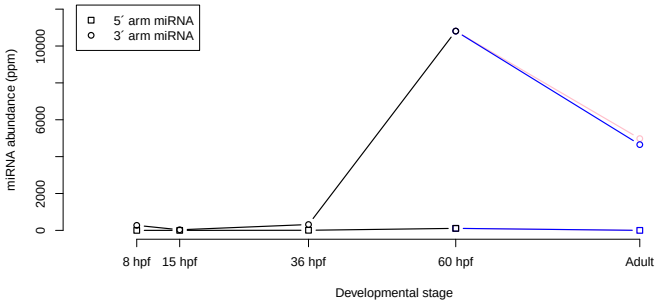
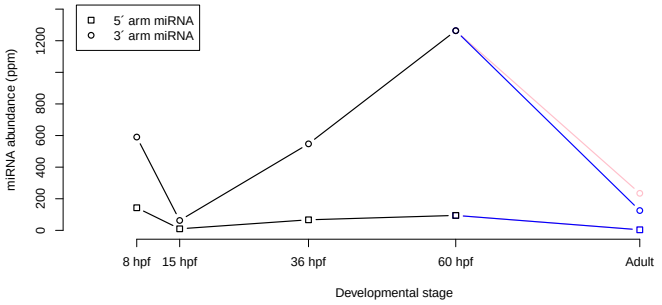
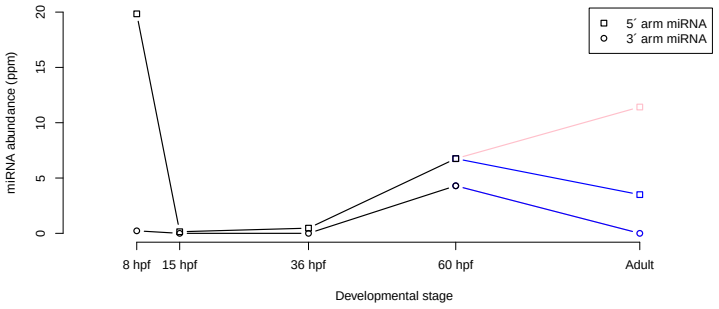
(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bfl-mir-2058 ortholog at Sc0000110 bp 921624-921546 (- strand)	5´arm: `UGAGAAGUAAGACUACCAUCCCGU` 3´arm: `(low abundance)`	
bfl-mir-10b ortholog at Sc0000000 bp 1938354-1938254 (- strand)	5´arm: `AACCCUGUGGAUCCGAUCUUGUG` 3´arm: `(low abundance)`	
bfl-mir-4856a ortholog at Sc0000022 bp 2103002-2103085 (+ strand)	5´arm: `ACGCAGUGACGUCAGCGCCUCU` 3´arm: `(low abundance)`	
bfl-mir-183 ortholog at Sc0000043 bp 398562-398462 (- strand)	5´arm: `UAUGGCACUGGUAGAAUUCACUGA` 3´arm: `(low abundance)`	

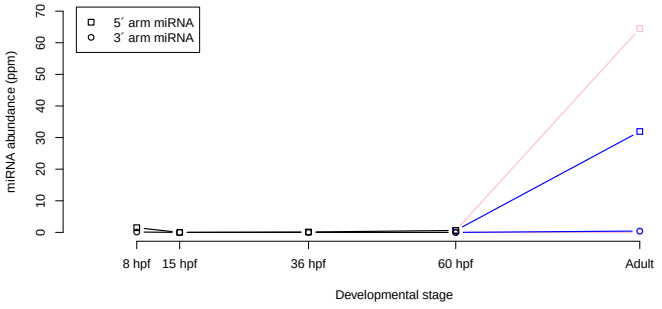
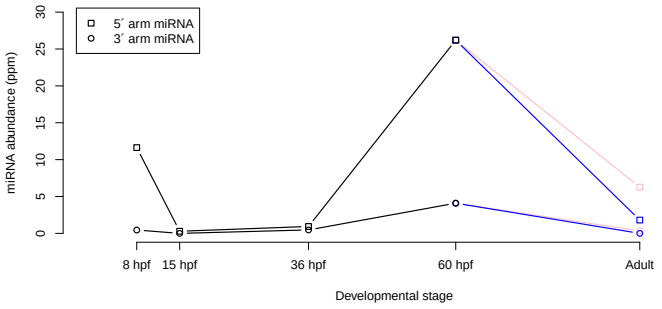
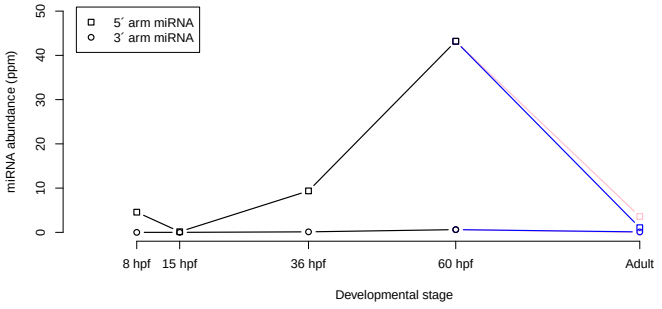
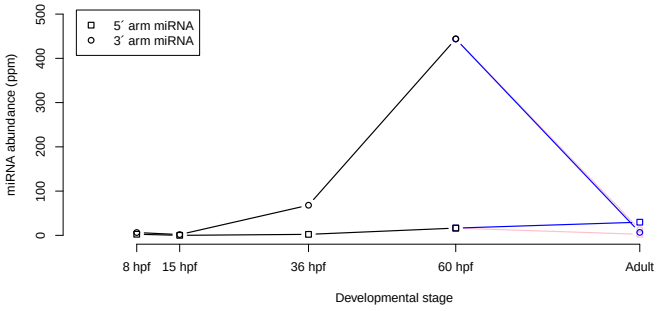
(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bbe-mir-2067 ortholog at Sc0000082 bp 843448-843367 (- strand)	5´arm: `(low abundance)` 3´arm: `AAGCAGCAGUAUGCAAUGGUGA`	
bfl-mir-184 ortholog at Sc0000017 bp 1306111-1306211 (+ strand)	5´arm: `CUUAUCACUUCUUCCGCCGAGC` 3´arm: `UGGACGGAGAACUGAUAAGGGCC`	
bbe-mir-34a ortholog at Sc0000034 bp 210967-210880 (- strand)	5´arm: `CGUUCCCUGUGUGCUGCUG` 3´arm: `AGCCACUGUACACACUCCCGUA`	
bfl-mir-4904 ortholog at Sc0000028 bp 432156-432080 (- strand)	5´arm: `UCCCGGAGUCGUUCUAUACGCCG` 3´arm: `(low abundance)`	

(continued on next page)

bioRxiv preprint doi: https://doi.org/10.1101/339820; this version posted June 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Pre-miRNA	miRNA sequences	Abundance profile in development
bfl-mir-4870 ortholog at Sc0000009 bp 4682265-4682351 (+ strand)	5´arm: `GAUGUUUGUACUGUCUGUCUGUU` 3´arm: `(low abundance)`	
bfl-mir-4866 ortholog at Sc0000043 bp 1850204-1850295 (+ strand)	5´arm: `UCACACUUGUACUUCUUAGCAU` 3´arm: `(low abundance)`	
bbe-mir-9 ortholog at Sc0000019 bp 2253983-2254064 (+ strand)	5´arm: `UCUUUGGUUAUCUAGCUGUAUGA` 3´arm: `(low abundance)`	
bfl-mir-217 ortholog at Sc0000230 bp 138098-138165 (+ strand)	5´arm: `UACUGCAUCAGGAACUGAUUGG` 3´arm: `AAUCUGUUCCUCAUGCAUGGCU`	

Supplementary Table 1: **Detection of conserved miRNAs**. *Branchiostoma lanceolatum* orthologs for *B. floridae* or *B. belcheri* pre-miRNA hairpins (as described in miRBase v.22) were screened for their predicted secondary structure and the abundance of the small RNAs they generate. Only those hairpins that comply with these rules are shown in this table. First column: name of orthologous pre-miRNA, and genomic coordinates in *B. lanceolatum*. Second column: sequences of the major forms of the 5´ arm and 3´ arm miRNAs, if expressed at ⩾10 ppm in at least one developmental stage (miRNAs that do not meet that criterion are flagged "low abundance"). Third column: abundance of the 5´ arm and 3´ arm miRNAs in Libraries #1 along development. Embryonic stages contain mixed sexes; adult stages are shown in blue and pink for males and females, respectively. Trimming (up to 3 nt) and templated extension of miRNA 3´ ends were considered when measuring read counts.