



The miRNA biogenesis in marine bivalves

Umberto Rosani¹, Alberto Pallavicini² and Paola Venier¹

¹Department of Biology, University of Padova, Padova, Italy

²Department of Life Sciences, University of Trieste, Trieste, Italy

ABSTRACT

Small non-coding RNAs include powerful regulators of gene expression, transposon mobility and virus activity. Among the various categories, mature microRNAs (miRNAs) guide the translational repression and decay of several targeted mRNAs. The biogenesis of miRNAs depends on few gene products, essentially conserved from basal to higher metazoans, whose protein domains allow specific interactions with dsRNA. Here, we report the identification of key genes responsible of the miRNA biogenesis in 32 bivalves, with particular attention to the aquaculture species *Mytilus galloprovincialis* and *Crassostrea gigas*. In detail, we have identified and phylogenetically compared eight evolutionary conserved proteins: DROSHA, DGCR8, EXP5, RAN, DICER TARBP2, AGO and PIWI. In mussels, we recognized several other proteins participating in the miRNA biogenesis or in the subsequent RNA silencing. According to digital expression analysis, these genes display low and not inducible expression levels in adult mussels and oysters whereas they are considerably expressed during development. As miRNAs play an important role also in the antiviral responses, knowledge on their production and regulative effects can shed light on essential molecular processes and provide new hints for disease prevention in bivalves.

Subjects Aquaculture, Fisheries and Fish Science, Evolutionary Studies, Genomics, Marine Biology

Keywords miRNA biogenesis, *Mytilus galloprovincialis*, *Crassostrea gigas*, Bivalves, RNAi

Submitted 5 September 2015

Accepted 15 February 2016

Published 7 March 2016

Corresponding author

Paola Venier, paola.venier@unipd.it

Academic editor

Ana Pavasovic

Additional Information and
Declarations can be found on
page 18

DOI 10.7717/peerj.1763

© Copyright

2016 Rosani et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

INTRODUCTION

Different types of non-coding RNAs (ncRNAs) have gained attention for their powerful regulatory action on eukaryotic genes and other genetic elements (*Carninci et al., 2005; Mortimer, Kidwell & Doudna, 2014*). The process known as RNA interference (RNAi) exemplifies an evolutionary conserved mechanism of gene silencing based on small guide RNAs and specific interacting proteins (*Tomoyasu et al., 2008; Gammon & Mello, 2015*). Silencing RNAs (siRNAs) and microRNAs (miRNAs) take part to the same control machinery whereas Piwi-interacting RNAs (piRNAs) peculiarly silence germ-line transposons, among other roles (*Théron et al., 2014; Iwasaki, Siomi & Siomi, 2015*). Long noncoding RNAs (lncRNAs) in their normal or mutated forms can widely influence physiological and pathological processes, as multiple lines of evidence indicate their involvement in chromosome inactivation and epigenetic modifications, control of mRNA decay and translation, and DNA sequestration of transcription factors (*Huarte, 2015; Ruan, 2015*). More recently, circular RNAs have been identified as a group of competing

endogenous RNAs whose effects in the miRNA function and transcriptional/post-transcriptional regulation are now matter of study (Qu *et al.*, 2015).

miRNAs are single-stranded RNA molecules of around 22 nucleotides, presenting conserved structural features and able to modulate the expression of eukaryotic genes by inhibition of mRNA translation or enhancement of mRNA decay (Ambros, 2003; Bartel, 2004; Tarver, Donoghue & Peterson, 2012). Up to now, diversified sets of miRNAs have been detected in five eukaryotic taxa (eumetazoans, silicisponges, vascular plants, *Clamydomonas* and *Ectocarpus* spp.) while they are apparently absent in protists (Grimson *et al.*, 2008; Tarver *et al.*, 2015). Depending on the annotation procedure, the number of human miRNAs varies from 523 to 1,881 miRNA precursors, as reported in MirGeneDB (Fromm *et al.*, 2015) or in miRBase v. 21 (Kozomara, Griffiths & Jones, 2014), respectively. Overall, human miRNAs could target 30–60% of the transcribed genes (John *et al.*, 2004; Sand *et al.*, 2012), with implications in cell differentiation (Berezikov *et al.*, 2005), cell death (Xu *et al.*, 2015), stress responses (Mendell & Olson, 2012) and diseases (Huang *et al.*, 2014; Min & Chan, 2015).

The miRNA biogenesis starts from pri-miRNA transcripts, mostly generated from RNA polymerase II in form of long non-coding RNAs and able to form a hairpin subsequently recognized by the so called microprocessor complex. DROSHA, a double-stranded RNA-specific ribonuclease III, and the RNA binding protein *Di-George syndrome Critical Region gene 8* (DGCR8) are the microprocessor's core proteins which allow interactions with the DDX5 helicase, the RNA binding protein Lin-28 and hnRNP A1, among other elements (Jean-Philippe, Paz & Caputi, 2013; Hong *et al.*, 2013). During the recognition of pri-miRNAs at the dsRNA-ssRNA junction, DGCR8 acts as a crucial molecular anchor and directs DROSHA to cleave 11 bp away from the junction, with consequent release of hairpin-shaped pre-miRNAs (Denli *et al.*, 2004). Pre-miRNAs are firstly exported to the cytoplasm via the *Exportin5* (XPO5) by interaction with the small GTPase RAN; then, they are further processed by the RISC loading complex, composed by the endoribonuclease DICER, the RNA binding protein TARBP2 and Argonaute proteins (MacRae *et al.*, 2008; Miyoshi *et al.*, 2009). The evolutionary conserved Argonaute proteins are specialized in binding small RNAs and exist in several isoforms, with AGO and PIWI representing two distinct subclades (Tolia & JoshuaTor, 2007; Ender & Meister, 2010).

AGOs select the 'guide' miRNA strand necessary for targeted gene silencing and, therefore, are responsible for final miRNA maturation. Several other proteins have been demonstrated to cooperate in miRNA processing and functions (Ender & Meister, 2010). In fact, AGOs operate transcriptional repression and cause mRNA decay by interacting with the GW-rich N-terminal region of GW182, a protein associated with cellular P-bodies (Van Kouwenhove, Kedde & Agami, 2011). Other proteins involved in the mRNA turnover (CAF1, PABPC1, eIF4G; CCR4-NOT and PAN2-PAN3 deadenylation complexes; in human somatic cells, also the decapping complex DCP1-DCP2 and at least four helicases, DDX5, DDX6, DDX17 and DDX42) may cooperate with the AGO-GW182 complex to reduce the mRNA translation efficiency (Nottrott, Simard & Richter, 2006; Fabian & Sonenberg, 2012).

Unlike AGOs, the PIWI proteins specifically interact with piRNAs to participate in the germline specification, gametogenesis, transposon silencing and in the maintenance of genome integrity (Carmell et al., 2007; Malone & Hannon, 2009; Ghildiyal & Zamore, 2009; Siomi et al., 2011). The piRNA mechanism of action is not so well defined but probably it involves the *arginine methyl-transferase* PRMT5, *tudor domain-containing proteins* (TDRDs) and the *Maelstrom* protein (MAEL) (Sokolova et al., 2011).

With the widespread and cost-effective use of Next Generation Sequencing (NGS) technologies, miRNAs have been deeply explored in non-model organisms, including bacteria (Xu et al., 2014), plants (Rhee, Chae & Kim, 2015) and viruses (Kincaid & Sullivan, 2012; Diebel et al., 2015). The basic set of genes involved in the miRNA biogenesis, and related protein interactions, are well known in mammals (Lau & MacRae, 2009), and also in other metazoans like *Cnidaria* (Moran et al., 2013), *Platyhelminthes* (Resch & Palakodeti, 2012) and insects (Lucas & Raikhel, 2013; Hussain & Asgari, 2014). Regarding mollusks, lists of miRNAs have been reported for a few species (Jiao et al., 2014; Chen et al., 2014; Martín-Gómez et al., 2014; Zhou et al., 2014), miRNA families have been investigated in the limpet genome (Kenny et al., 2015) and one study has considered bivalve DICER sequences for phylogenetic analysis (Gao et al., 2014). A general overview on the bivalve miRNA biogenesis complements is still lacking, so we took advantage of several genomic and transcriptomic datasets available for *Lophotrochozoa* (GIGA Community of Scientists, 2014) to identify and characterize the core elements involved in the miRNA formation pathway in *Mytilus* and *Crassostrea* spp. and other bivalves.

MATERIALS & METHODS

Sequences coding for proteins centrally involved in the miRNA pathway, namely DROSHA, DGCR8, XPO5, RAN, DICER, TARBP2, AGO and PIWI, have been methodically identified in the genomes and transcriptomes of *M. galloprovincialis* (Mg) and *C. gigas* (Cg) as well as in other bivalve and non-bivalve species (66 species, listed in Table 1).

Sequence retrieval and analysis

The Mg WGS project (ID APJB000000000.1 (Nguyen, Hayes & Ingram, 2014)) and the Cg genome draft (GCA_000297895 (Zhang et al., 2012)) were retrieved from GenBank, whereas the oyster genome annotations were obtained from Ensembl Metazoa release 29 (http://metazoa.ensembl.org/Crassostrea_gigas/Info/Index). A Mg reference transcriptome was produced using 18,788 ESTs of mixed tissues previously obtained by Sanger sequencing (Venier et al., 2009) and 453 million reads obtained by paired-end (2 × 100 bp) Illumina Hiseq2000 sequencing of digestive gland from North Adriatic Sea mussels (ID: PRJNA88481) (Gerdol et al., 2014), and from haemocytes, gills, mantle and muscle of Spanish mussels (ID: SRP033481) (Moreira et al., 2015). The quality of the sequencing readout was evaluated by the FastQC suite (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) discarding the reads with PHRED quality below 20 and presenting more than two ambiguous nucleotides. De-novo assembly was performed with Trinity, release 2013-08-14 (Grabherr et al., 2011), setting the minimum contig length at 200 bp and using default settings. Subsequently,

Table 1 Organisms included in the present work. Phylum, organism name, sequence origin and reference, ID used in phylogenetic trees and identified sequences are reported. Protostomia (green), Deuterostomia (orange) and novel protein sequences (numbers in bold) are well discernible.

Phylum	Species	Sequence origin	Ref	Used ID	DROSHA	DGCR8	XPO5	RAN	DICER	TARBP2	AGO or PIWI
Ctenophora	<i>Pleurobrachia bachei</i>	G	EM	Ple_bac	No	No	1	No	1	No	4
	<i>Mnemiopsis leidyi</i>	G	EM	Mne_lei	No	No	1	1	1	No	4
Porifera	<i>Amphimedon queenslandica</i>	G	EM	Aq	No	1	1	1	2	No	2
Placozoa	<i>Trichoplax adhaerens</i>	G	EM	Tri_adh	No	No	1	1	2	No	1
	<i>Nematostella vectensis</i>	G	EM	Nvec	1	1	1	1	2	No	4
Cnidaria	<i>Porites australiensis</i>	T	TSA	Por_au	1	1	1	1	1	No	6
	<i>Anthopleura elegantissima</i>	T	TSA	Ant_ele	1	1	1	1	1	No	6
Nematoda	<i>Caenorhabditis elegans</i>	G	M	Ce	1	1		1	1	1	23
	<i>Daphnia pulex</i>	G	EM	Dap_pul	1	1	1	1	2	1	10
Arthropoda	<i>Culex quinquefasciatus</i>	G	EM	Cq	1	1	1	1	2	1	4
	<i>Drosophila melanogaster</i>	G	EM	Dm	1	1	1	1	2	1	4
	<i>Nasonia vitripennis</i>	G	M	Nv	1	1	1	1	2	1	4
	<i>Tribolium castaneum</i>	G	M	Tc	1	1	1	1	2	1	4
	<i>Apis mellifera</i>	G	EM	Am	1	1	1	1	2	1	4
	<i>Lasioglossum albipes</i>	G	K	La	1	1	1	1	2	1	4
	<i>Acyrtosiphon pisum</i>	G	A	Ap	1	1	1	1	2	1	15
Platyhelminthes	<i>Schistosoma mansoni</i>	G	GD	Sch_man	1	1	2	1	2	1	3
	<i>Schmidtea mediterranea</i>	G	SG	Sch_med	1	1		1	2	1	4
Rotifera	<i>Adineta vaga</i>	G	V	Adi_vag	1	1	1	1	1	1	4
Brachiopoda	<i>Lingula anatina</i>	G	L	Lin_ana	1		1	1	1	1	3
Annelida	<i>Capitella telata</i>	G	EM	Ct	1	1	1	1	1	1	3
	<i>Helobdella robusta</i>	G	EM	Hel_rob	1	1	1	1	1	1	4
Cephalopoda	<i>Octopus bimaculoides</i>	G	M	Oct_bim	1	1	1	1	1	1	4
	<i>Aplysia californica</i>	G	B	Ac	1	1	1	3	1	1	4
	<i>Lottia gigantea</i>	G	M	Lg	1	1	1	1	1	1	3
	<i>Mytilus galloprovincialis</i>	T	Local	Mg	1	1	1	1	1	1	3
	<i>Mytilus edulis</i>	T	Local	Me	1	1	1	1	1	1	2
	<i>Mytilus californianus</i>	T	Local	Mc				1		1	2
	<i>Anadara trapezia</i>	T	Local	At				1			1
	<i>Tegillarca granosa</i>	T	Local	Tg				1			
	<i>Bathymodiolus azoricus</i>	T	Local	Ba				1			
	<i>Perna viridis</i>	T	Local	Pv			1	1	1	1	2
	<i>Ennucula tenuis</i>	T	Local	Et				1		1	1
	<i>Crassostrea corteziensis</i>	T	Local	Cc		1	1	1	1		4
	<i>Crassostrea gigas</i>	G	EM	Cg	1	1	1	1	1	1	4

(continued on next page)

Table 1 (continued)

Phylum	Species	Sequence origin	Ref	Used ID	DROSHA	DGCR8	XPO5	RAN	DICER	TARBP2	AGO or PIWI
Mollusca	<i>Crassostrea hongkongensis</i>	T	Local	Ch		1		1			3
	<i>Crassostrea virginica</i>	T	Local	Cv	1						4
	<i>Crassostrea angulata</i>	T	local	Ca	1	1	1	1			2
	<i>Ostrea chilensis</i>	T	Local	Oc				1			
	<i>Ostrea edulis</i>	T	Local	Oe				1			2
	<i>Ostrea lurida</i>	T	local	Ol				1			1
	<i>Ostreola stentina</i>	T	Local	Os				1			
	<i>Saccostrea glomerata</i>	T	Local	Sg				1			
	<i>Argopecten irradians</i>	T	Local	Ai				1			1
	<i>Mizuhopecten yessoensis</i>	T	Local	My	1		1	1		1	2
	<i>Pecten maximus</i>	T	Local	Pm				1			2
	<i>Pinctada fucata</i>	G	F	Pf			1	1	1	1	
	<i>Solemya velum</i>	T	Local	Sv	1	1	1	1		1	3
	<i>Elliptio complanata</i>	T	Local	Ec	1		1	1		1	1
	<i>Pyganodon grandis</i>	T	Local	Pg			1	1			2
	<i>Unio merus tetralasmus</i>	T	Local	Ut			1	1			3
	<i>Villosa lienosa</i>	T	Local	Vl				1			1
	<i>Corbicula fluminea</i>	T	local	Cf				1			1
<i>Meretrix meretrix</i>	T	local	Mm				1			2	
<i>Ruditapes decussatus</i>	T	local	Rd				1			1	
<i>Ruditapes philippinarum</i>	T	local	Rp				1			1	
Echinodermata	<i>Strongylocentrotus purpuratus</i>	G	M	Sp	1	1	1	1	1	1	3
Hemichordata	<i>Saccoglossus kowalevskii</i>	G	M	Sk	1	1	1	1	1	1	1
Chordata	<i>Homo sapiens</i>	G	M	Hs	1	1	1	1	1	1	8
	<i>Ciona intestinalis</i>	G	M	Ci	1	1	1	1	1	1	3
	<i>Branchiostoma floridae</i>	G	M	Bf	1	1	1	1	1	1	7
	<i>Oncorhynchus mykiss</i>	G	O	Om	1	1	1	1	1	1	5
	<i>Danio rerio</i>	G	M	Dr	1	1	1	1	1	1	5
Streptophyta	<i>Arabidopsis thaliana</i>	G	P	At	No	No	1	1	4	No	10
	<i>Populus trichocarpa</i>	G	P	Pt	No	No	1	1	5	No	11

Notes.

Abbreviations: A, Aphidbase; B, broadinstitute.org/ftp/pub/assemblies/invertebrates/aplysia/; EM, Ensembl Metazoa v.29; F, Takeuchi et al. (2012) DNA Res. 19(2): 117–130; G, Genome; GD, GeneDB; K, Kocher et al. (2013) Genome Biology 14 (12): R142; L, Lou et al. (2015) Nat Commun. 8; 6:8301; M, Metazome v3.0; O, Berthelot et al. (2014) Nat Commun. 22; 5: 3657; P, Phytosome 11; SG, SmedGD; T, Transcriptome; TSA, NCBI Transcriptome shotgun assembly; V, Genoscope.

Table 2 Key proteins of the miRNA biogenesis with their structural domains.

Process step	Protein name	Domains
Microprocessor complex	DROSHA	2× RIBO III + 1× DSRM
	DGCR8	1× WW + 2× DSRM
Moving to cytoplasm	XPO5	1× XPO1
	RAN	1× RAN
RISC loading complex	DICER	PDB + Helicase + DICER + PAZ + 2× RIBO + DSRM
	TARBP2	3× DSRM
Final miRNA maturation	AGO	DUF + PAZ + PIWI
	PIWI-like	PAZ + PIWI

protein coding sequences (cds) were predicted with Transdecoder ([Grabherr et al., 2011](#)). Transcriptomic reads of 30 bivalve *spp.* (Cg plus other 29 species) were retrieved from the SRA archive and assembled as described above (details in [File S1](#)). The protein predictions of further 33 organisms were directly retrieved from public repositories or extracted from the corresponding genome releases. The NCBI transcriptome shotgun assembly (TSA) database was interrogated to retrieve hits for two additional cnidarians, *Porites australiensis* and *Anthopleura elegantissima* ([Table 1](#)).

Protein domain searches

To investigate the presence of eight key proteins of miRNA biogenesis (DROSHA, DGCR8, XPO5, RAN, DICER, TARBP2, AGO and PIWI), we downloaded their predictive HMM from PFAM v.27 (listed in [Table 2](#)) and we scanned the sequence datasets with HMMer v3.1 ([Eddy, 2011](#)) applying a cut-off E-value of 0.01. To achieve a meaningful comparison of proteins from different organisms, we retained only hits presenting all diagnostic domains. Moreover, we identified several mussel transcripts related to protein interactions occurring in the miRNA biogenesis. To identify such proteins, we retrieved from PFAM the diagnostic domains of human homologs (listed in [Table 3](#)) and we scanned their presence in the Mg transcriptome as described above. Protein domain organization was reconstructed using SMART ([Letunic, Doerks & Bork, 2012](#)).

Gene structure analysis

We used the transcript sequences of DROSHA, DGCR8, XPO5, DICER and TARBP2 as blast queries against all Mg genomic contigs (*blastn*) in order to recover the related gene structures. Positive hits having *e*-value lower than 10^{-20} were extracted and assembled on the corresponding transcript, used as backbone. RNA-seq read mappings with adapted parameters (CLC Genomic Workbench *large gap mapping tool*, with similarity and length fraction set at 0.9) allowed us to ascertain the correct gene assembly. Homolog gene structures were retrieved by interrogating genomic browsers, like Metazome v.3 (for *C. intestinalis*, *B. floridae*, *D. rerio*, *S. kowalevskii*, *S. purpuratus*, *N. vectensis*, *T. castaneum*, *L. gigantea*, *O. bimaculoides*, *C. elegans* and *H. sapiens*) and Ensembl Metazoa v.29 (for *C. gigas*, *C. quinquefasciatus*, *D. melanogaster*, *N. vitripennis*, *A. mellifera*, *A. queenslandica*,

Table 3 miRNA biogenesis proteins of *Mytilus galloprovincialis*. Protein name, GenBank ID, transcript (bp) and protein length (aa), identified domains and annotation (first hit, *e*-value and percentage of similarity) are reported.

	Protein name	GenBank ID	Transcript length (bp)	Protein length (aa)	Identified domain(s)	Annotation (first hit)	<i>E</i> -value (e^{\wedge})	Similarity (%)
Key miRNA biogenesis proteins	MgDROSHA	KT447251	4,384	1,377	2× RIBO III + 1× DSRM	Ribonuclease 3-like (<i>Crassostrea gigas</i>)	0	67
	MgDGCR8	KT447252	2,483	728	1× WW + 2× DSRM	Microprocessor complex subunit DGCR8-like (<i>Crassostrea gigas</i>)	0	50
	MgXPO5	KT447259	3,875	1,201	XPO1	Exportin-5-like (<i>Crassostrea gigas</i>)	0	55
	MgRAN	KT447254	1,113	214	RAN	GTP-binding nuclear protein Ran (<i>Crassostrea gigas</i>)	-143	93
	MgDICER	KT447258	6,013	1,850	PDB + Helicase + DICER + PAZ + 2× RIBO + DSRM	Endoribonuclease Dicer-like (<i>Crassostrea gigas</i>)	0	58
	MgTARBP2	KT447253	7,583	321	3× DSRM	Probable RISC-loading complex subunit (<i>Crassostrea gigas</i>)	-143	69
	MgAGO	KT447257	3,337	892	DUF + PAZ + PIWI	Protein argonaute-2-like (<i>Crassostrea gigas</i>)	0	84
	MgPIWIa	KT447255	2,686	867	PAZ + PIWI	Piwi-like protein 1 (<i>Crassostrea gigas</i>)	0	75
	MgPIWIb	KT447256	3,603	948	PAZ + PIWI	Piwi-like protein 2 (<i>Hydra vulgaris</i>)	0	59
	MgGW182	KT447250	3,825	1,274	UBA + RRM	Trinucleotide repeat-containing gene 6C protein-like (<i>Crassostrea gigas</i>)	0	45
	MgCNOT1	KT694355	5,373	1,791	DUF3819 + NOT1	CCR4-NOT transcription complex subunit 1-like (<i>Crassostrea gigas</i>)	0	69
	MgCNOT2	KT694357	864	288	NOT2_3_5	CCR4-NOT transcription complex subunit 2 (<i>Pinctada fucata</i>)	-156	82
	MgCNOT3	KT694358	2,142	714	NOT3 + NOT2_3_5	CCR4-NOT transcription complex subunit 3-like (<i>Crassostrea gigas</i>)	0	97
	MgCNOT6	KT694359	2,592	864	Exo_endo_phos	Uncharacterized protein LOC105348954 isoform X1 (<i>Crassostrea gigas</i>)	0	71
	MgCNOT7	KT694360	897	299	CAF1	CCR4-NOT transcription complex subunit 7-like (<i>Crassostrea gigas</i>)	0	84
	MgCNOT9	KT694361	927	309	RCD1	Cell differentiation protein RCD1 homolog (<i>Crassostrea gigas</i>)	0	93
	MgCNOT10	KT694356	2,133	711	TPR_1	CCR4-NOT transcription complex subunit 10-like (<i>Crassostrea gigas</i>)	0	71
	MgDDX5	KT694371	1,740	538	DEAD + Helic	ATP-dependent RNA helicase DDX5 (<i>Crassostrea gigas</i>)	0	75
	MgDDX6	KT694372	1,332	443	DEAD + Helic	ATP-dependent RNA helicase me31b (<i>Crassostrea gigas</i>)	0	88

(continued on next page)

Table 3 (continued)

	Protein name	GenBank ID	Transcript length (bp)	Protein length (aa)	Identified domain(s)	Annotation (first hit)	E-value (e [^])	Similarity (%)
Other interacting proteins	MgDDX20	KT694373	1,836	612	DEAD + Helic	ATP-dependent RNA helicase DDX20 (<i>Crassostrea gigas</i>)	0	77
	MgDDX42	KT694374	2,196	731	DEAD + Helic	ATP-dependent RNA helicase DDX42 (<i>Crassostrea gigas</i>)	0	72
	MgPABP	KT694365	1,881	627	4 × RRM + PABP	polyadenylate-binding protein 4 (<i>Hydra vulgaris</i>)	0	74
	MgeIF4G	KT694364	5,019	1,672	MIF4G + MA3 + W2	eukaryotic translation initiation factor 4 gamma (<i>Crassostrea gigas</i>)	0	57
	MgPAN2	KT694367	3,606	1,202	UCH_1 + RNase_T	PAB-dependent poly(A)-specific ribonuclease sub-unit PAN2 (<i>Lingula anatina</i>)	0	72
	MgPAN3	KT694368	2,334	778	None	PAB-dependent poly(A)-specific ribonuclease sub-unit PAN3 (<i>Lingula anatina</i>)	0	67
	MgDCP1	KT694362	1,611	536	DCP1	mRNA-decapping enzyme 1A-like (<i>Crassostrea gigas</i>)	−126	73
	MgDCP2	KT694363	1,313	385	DCP2 + NUDIX	m7GpppN-mRNA hydrolase (<i>Lingula anatina</i>)	−117	67
	MgPRMT5	KT694369	1,893	631	PRMT5	protein arginine N-methyltransferase 5-like (<i>Crassostrea gigas</i>)	0	72
	MgTudor-11	KT694370	2,682	894	4 × Snc + TUDOR	Hypothetical protein mRNA (<i>Lottia gigantea</i>)	0	73
MgMaelstrom	KT694366	1,321	404	HMG + MAEL	Protein maelstrom (<i>Crassostrea gigas</i>)	−155	62	

P. bachei, *M. leidy*, *T. adhaerens*, *N. vectensis*, *D. pulex*, *S. mansoni*, *S. mediterranea*, *A. vaga*, *L. anatine*, *H. robusta* and *C. telata*) or by local *blastn* against the downloaded genomes (*A. pisum* and *L. albipes*).

Phylogenetic analysis

The inferred protein sequences were aligned using MUSCLE, release 2014-05-29 (Edgar, 2004). Subsequently, the fasta alignments were analyzed using Gblocks v.0.91 (Castresana, 2000) to extract conserved positions (positions common to 51% of the locally aligned sequences). Trees were built using neighbor joining or maximum likelihood clustering methods with 1,000 bootstrap replicates. Bayesian phylogenies were reconstructed using MrBayes v.3.2.5 (Ronquist et al., 2012), with GTR substitution evolutionary model with gamma-distributed rate variation across sites, evaluating the convergence after 1,000,000 runs (0.5 was considered as cut-off value). Trees were visualized and edited with FigTree v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Digital expression analysis

To analyze the expression of the selected genes in Cg and Mg RNA datasets, we retrieved all available RNA-seq samples from the NCBI SRA archive. For Cg, we analyzed 123 Illumina RNA-seq samples related to adult tissues or developmental stages. For Mg, we analyzed 13 RNA samples from gills (1), digestive gland (6), haemocytes (2), mantle (2) and muscle (2). Overall, we included in the expression analysis 2,271 and 453M reads for Cg and Mg, respectively (File S2). The trimmed reads were mapped to Cg and Mg genes using the CLC Genomics Workbench v.8.0 (Qiagen, Hilden, Germany) mapping tool, with length and similarity fractions set at 0.75 and 0.95, respectively, and mismatch/insertion/deletion penalties at 3/3/3. The number of uniquely mapped reads of each dataset were counted and used to calculate digital expression values as TPM (Transcripts Per Kilobase Million mapped reads) as described by (Wagner, Kin & Lynch, 2013), considering 3 TPMs as lower detection limit.

RESULTS

Mussel transcripts related to the miRNA biogenesis

We identified *Mytilus galloprovincialis* transcripts involved in the miRNA biogenesis by systematic searches of diagnostic domains (Table 2) in a transcriptome assembly produced from 453 million Illumina reads. Thus, we recovered nine transcripts coding for DROSHA, DGCR8, XPO5, RAN, DICER, TARBP2 and for three Argonaute genes (one Ago and two Piwi-like proteins, Table 3). We also identified 21 mussel proteins expected to play a role in the miRNA maturation or involved in RNAi processes (File S3).

Figure 1 relates the general process of eukaryotic miRNA biogenesis to the mussel proteins identified in this work. MgDROSHA and MgDGCR8 are expected to start the maturation of pri-miRNAs produced by RNA polymerase II. MgDROSHA codes for a 1,377 aa length protein containing all the canonical domains (2 RIBOc domains in positions 959–1,093 and 1,139–1,271 and one DSRM domain in position 1,278–1,351) whereas MgDGCR8 is a 728 aa length protein having one WW domain in position 229–258,

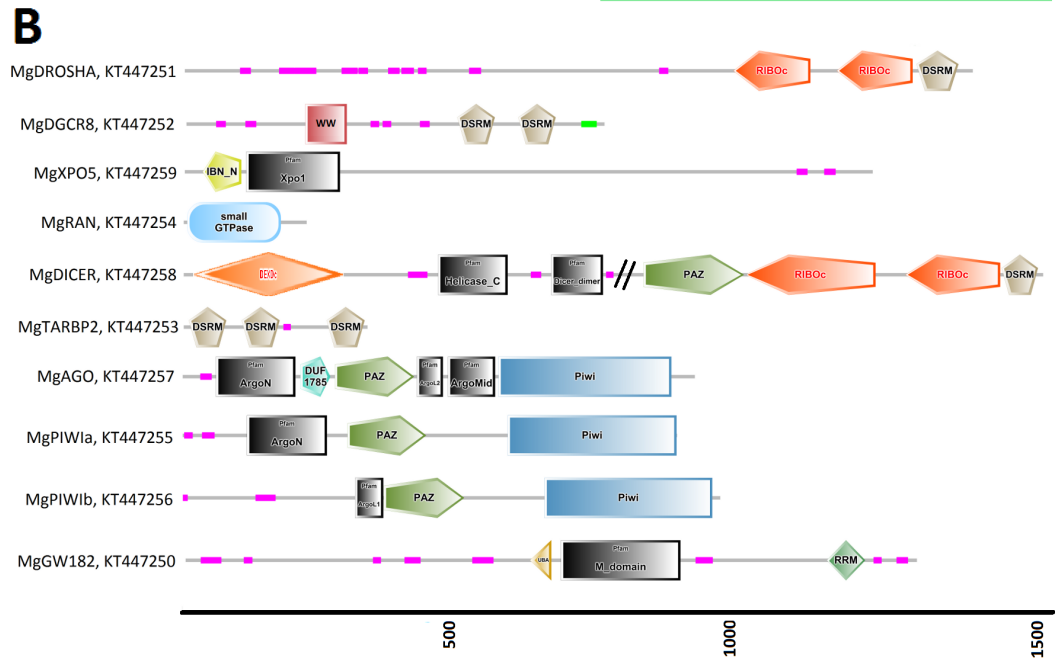
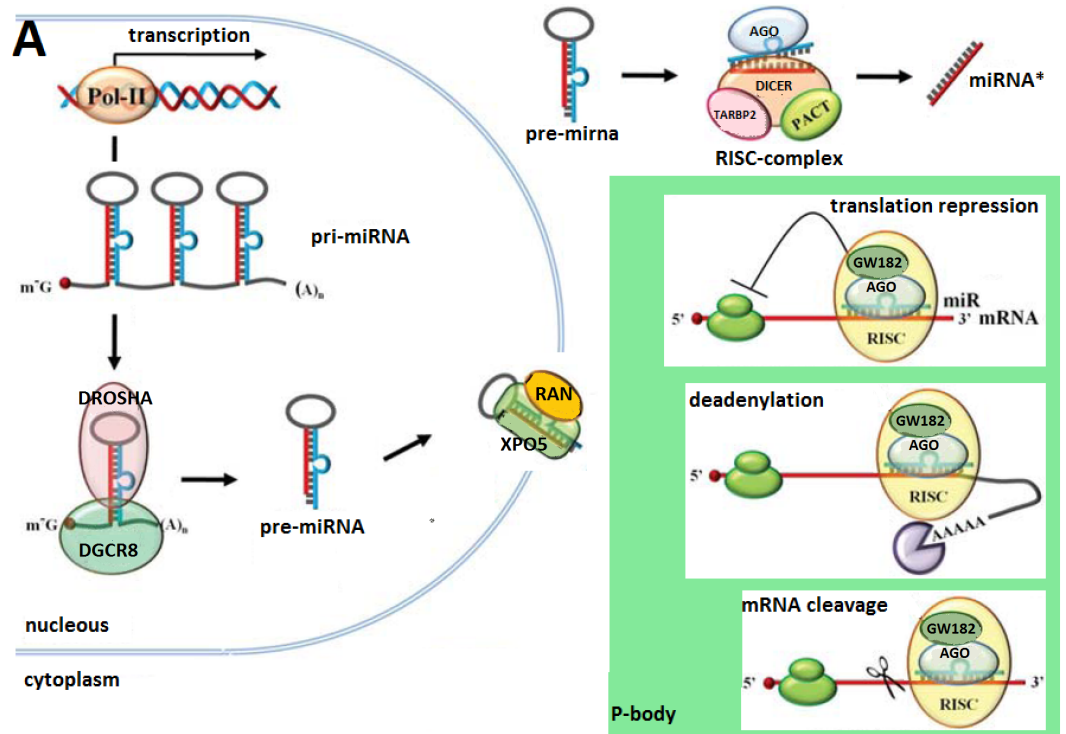


Figure 1 (A) Graphical reconstruction of mussel miRNA biogenesis process. (Modified from *Kapinas & Delany, 2011*). (B) Conserved domains of the mussel miRNA complements.

necessary for the interaction with DROSHA, and two DSRM domains (positions 472–536 and 578–642) necessary for pri-miRNA binding. MgXPO5 is expected to cooperate with MgRAN in the pre-miRNA cytoplasmic translocation. MgRAN encodes a 214 aa protein whereas MgXPO5 has a length of 1,201 aa and includes two 5' conserved domains (IBN_N and Xpo1) and one conserved region necessary for the interaction with interleukin enhancer-binding factor 3 (position 525–562). In mussels, the RISC complex uploading pre-miRNAs is defined by the endoribonuclease MgDICER (1,850 aa) and MgTARBP2 (321 aa). Like in *Lophotrochozoa*, mussel DICER is encoded by a unique gene and contains the seven canonical domains, namely two helicase domains, one DICER-dimer domain, one PAZ, two RIBOc and a final DSRM domain. MgTARBP2 displays three DSRM domains in positions 9–73, 101–166 and 249–314. Moreover, *M. galloprovincialis* possess three argonaute proteins ranging from 861 to 941 aa in length and representative of one AGO (DUF1785, PAZ and PIWI domains) and two PIWI-like proteins (PAZ and PIWI domains). We considered the above mentioned gene products as the key complement of the miRNA biogenesis.

Among the possible interacting proteins, we identified MgGW182, a transcript encoding a protein shorter than the human counterparts but holding all the features considered significant for its interaction with AGOs and the CCR4-NOT complex. In fact, MgGW182 possesses 19 N-terminal GW stretches, followed by one UBA domain, a Q-rich region (M domain) and a C-terminal RNA recognition motif (RRM domain). Moreover, we recognized a C-terminal conserved site known as PAM2 (*Kozlov et al., 2010*), expected to interact with the poly(A) binding protein 1 (MgPABPC1) through the MLL motif and inhibit the mRNA translation by interfering with the mRNA circularization process (*Piao et al., 2010; Van Kouwenhove, Kedde & Agami, 2011*). In the mussel transcriptome, we also found putative homologs for a number of CNOT complex proteins (CNOT1, 2, 3, 6, 7, 9, and 10), for the eukaryotic translation initiation factor 4 gamma, 1 eIF4G, PAB-dependent poly(A)-specific ribonuclease subunits PAN2, PAN3, the decapping complex proteins DCP1 and DCP2, and several RNA helicases demonstrated to be crucial in the miRNA maturation (DDX5) and RNAi (DDX5- 6- 20 and 42). Finally, we recognized the putative mussel homologs of protein arginine methyltransferase 5 (MgPRMT5), tudor domain containing protein (MgTDRD-11) and maelstrom spermatogenic transposon silencer (MgMAEL).

Mussel genes related to the miRNA biogenesis

Taking advantage of mussel WGS data (*Nguyen, Hayes & Ingram, 2014*) we investigated the organization of the main genes involved in the mussel miRNA biogenesis. Fragmentation of the genomic mussel assembly (2.3 million contigs; 700 bp on average) and considerable dimension of the analyzed genes (9.6–17.6 kbp gene size in the case of Cg) prevented the recovery of the full gene sequences. Nevertheless, we can describe the complete gene structures of DROSHA, DGCR8, EXP5, DICER and TARBP2 (i.e., five of eight searched sequences) whose length varies between 7.5 and 27 kbp, confirmed by the back-mapping of 115,377 Illumina paired reads (*Fig. 2, File S4*). Moreover, these mussel genes showed

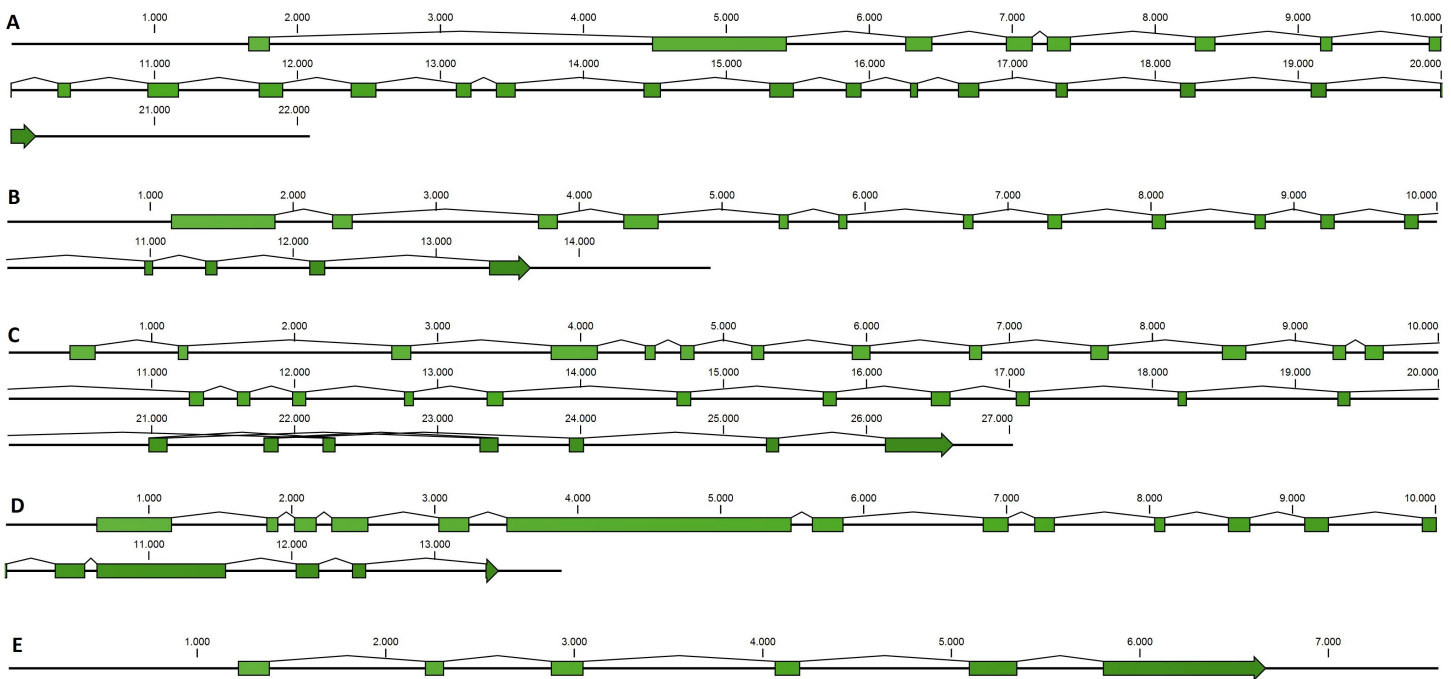


Figure 2 Mussel gene structures of DROSHA (A), DGCR8 (B), EXP5 (C), DICER (D) and TARBP2 (E). Green boxes represent exons, length is reported as base pair scale.

a remarkable conservation in terms of exon number when compared with a selection of homolog genes from deuterostome and protostome organisms (Table 4).

Transcripts related to the miRNA biogenesis in bivalve *spp*

To identify the miRNA biogenesis complements in marine mollusks, we used homologous genes retrieved from the genomes of *C. gigas*, *L. gigantea* and *A. californica*. Since the *C. gigas* genome includes annotations only for the cds regions, we exploited full-length transcripts obtained from a locally assembled oyster transcriptome to expand the genome annotations in this species. In particular, we updated the annotation of CgDGCR8 and CgDICER and we added new annotations for CgPIWI-1 (CGI_10008757: genomic contig JH815696, position 184178–187825) and CgTARBP2 (JH818440, 414703–419857).

Since many marine bivalve *spp.* do not have at present a sequenced genome, we used publicly available RNA-seq data to build 29 specie-specific transcriptome assemblies and retrieve the homologous sequences of interest. After domain searching, we carefully considered the high number of positive hits to retain only proteins including all the expected protein features. Thus, we retrieved 132 complete hits from marine mollusks: 10 DROSHAs, 9 DGCR8s, 14 XPO5s, 34 RANs, 7 DICERs, 13 TARBP2s and 45 Argonaute-like proteins, the latter classified in 13 AGO and 32 PIWI proteins by phylogenetic analysis (Table 1, File S5).

Phylogenetic analysis of the miRNA biogenesis proteins

The inferred sequences of single miRNA biogenesis proteins were aligned together with those retrieved from 34 sequenced genomes. Here, we report the phylogenetic analysis of

Table 4 Number of exons of five key miRNA biogenesis genes. Metazome 3.0 and Ensembl Metazoa v.29 genome browsers were interrogated with the previously analyzed hits for each organism. La and Ap genomes were downloaded and analyzed locally. Mg gene structures were retrieved as described in Methods. In green are reported Protostomia; in orange Deuterostomia.

Species	<i>Homo sapiens</i>	<i>Ciona intestinalis</i>	<i>Branchiostoma floridae</i>	<i>Danio rerio</i>	<i>Saccoglossus kowalevskii</i>	<i>Strongylocentrotus purpuratus</i>	<i>Nematostella vectensis</i>	<i>Amphimedon queenslandica</i>	<i>Caenorhabditis elegans</i>	<i>Capitella telata</i>	<i>Culex quinquefasciatus</i>	<i>Drosophila melanogaster</i>	<i>Nasonia vitripennis</i>	<i>Tribolium castaneum</i>	<i>Apis mellifera</i>	<i>Lasioglossus albipes</i>	<i>Acyrtosiphon pisum</i>	<i>Lotia gigantea</i>	<i>Crassostrea gigas</i>	<i>Mytilus galloprovincialis</i>
ID	Hs	Ci	Bf	Dr	Sk	Sp	Nvec	Aq	Ce	Ct	Cq	Dm	Nv	Tc	Am	La	Ap	Lg	Cg	Mg
DROSHA	27	24	29	17	24	20	13	14	6	28	3	3	11	9	13	23	1	23	30	23
DGCR8	14	10	15	10	15	13	7	No	11	18	4	5	11	6	6	6	2	11	18	16
XPO5	32	4	8	21	29	31	2	28	No	30	9	2	11	10	9	10	1	32	34	31
DICER	27	23	26	17	14	19	12	10	26	13	7	8	5	9	29	33	19	16	19	18
TARBP2	9	1	6	9	6	2	No	No	11	8	4	5	7	7	7	6	7	6	7	6

the five proteins centrally involved in the miRNA biogenesis, namely DROSHA, DGCR8, DICER, TARBP2 and AGOs (File S5 includes all protein sequences). We back-traced the presence of a canonical DROSHA up to Cnidaria, although we found only incomplete hits in Porifera and Placozoa and the genomes of *Ctenophora spp.* lack of both DROSHA and DGCR8, as reported by other authors (Maxwell et al., 2012). The DROSHA sequences from Cnidaria's appeared as general outgroup whereas those of Chordata clustered as outgroup of the other protostomes. DROSHAs from Mollusca and Arthropoda clustered consistently with the different taxa whereas those from Platyhelminthes, Rotifera, Brachiopoda and Annelida grouped together, with DROSHA from *Caenorhabditis elegans* (Nematoda) being the most far-related (Fig. 3A). Contrary to DROSHA, we identified a complete DGCR8 also in the Porifera *Amphimedon queenslandica*, suggesting that also DROSHA should be present in this taxa. Following phylogenetic analysis, we highlighted Cnidaria and Porifera proteins as outgroup, with mollusks (and Annelida) clustering with Arthropoda and more distantly *Platyhelminthes* and *Rotifera* hits. The *Chordata* sequences clustered as a separate group (Fig. 3B).

The finding of putative DICER sequences in *Ctenophora spp.* supports the presence of this gene through the whole Opisthokonta evolution (Maxwell et al., 2012). Also plants possess DICER homologues which occur in different copy number among taxa: two genes in Porifera, Placozoa, Cnidaria, Platyhelminthes and Arthropoda (with the exception of *D. pulex* that possess three genes); four genes in plants like *A. thaliana* and *P. trichocarpa* and one gene in Ctenophora, Rotifera, Cephalopoda Mollusca and Chordata. Moreover, the presence of DICER was reported in some Protozoa and fungi (Mukherjee, Campos & Kolaczowski, 2013). Phylogenetic analyses, separate insect DICER-2, plant DICERs from DICER-1. DICER-1 clade shows a consistent clustering of Arthropoda, Mollusca

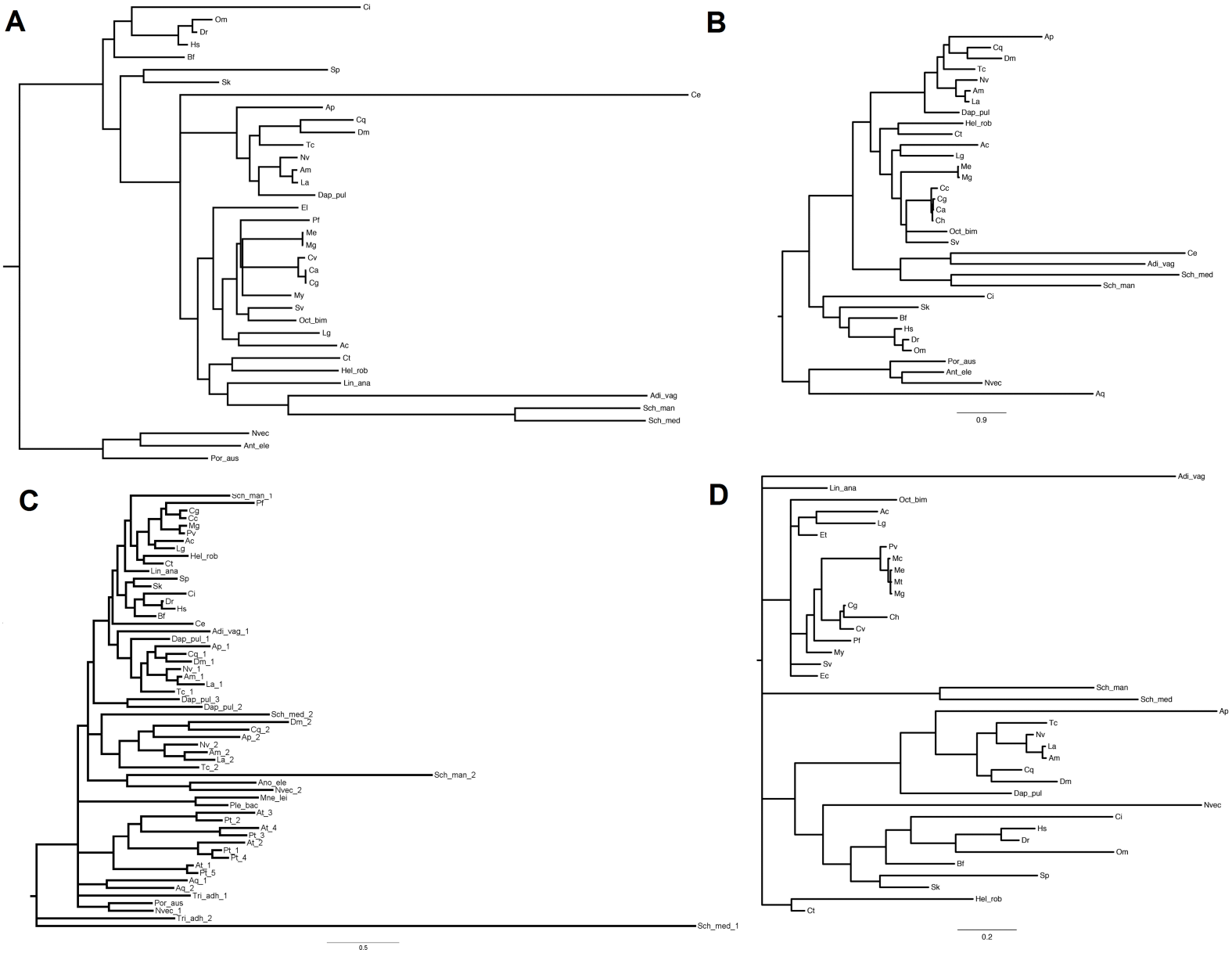


Figure 3 Phylogenetic relationships of four miRNA biogenesis proteins. (A) DROSHA, (B) DGCR8, (C) DICER and (D) TARBP2. Inferred protein sequences were aligned using MUSCLE, conserved positions were extracted using Gblocks and subjected to MrBayes analysis.

and Chordata hits, whereas some branches of basal metazoans and Platyhelminthes are not well resolved (Fig. 3C). Likewise, the phylogenetic tree regarding TARBP2 displays a clear cut-off between the proteins of mollusks, chordates and arthropods (Fig. 3D). We back-traced the miRNA cytoplasm export complex composed by RAN and XPO5 in all analyzed metazoans. Both RAN and XPO5 represent widely expressed sequences since we found them also in transcriptome assemblies, although with suboptimal sequence coverage.

Several AGO and Piwi proteins can be present in individual organisms and, in fact, we identified a total of 235 proteins. Whereas humans possess eight proteins, we found four proteins in the majority of the analyzed insect *spp.* (with the exception of 15 proteins in *A. pisum*) and three or four different proteins in bivalve *spp.* Also, basal Metazoa possess Argonaute-like sequences: four in the genomes of Ctenophora and Cnidaria *spp.*, one in the Placozoa *T. adhaerens* and two in *A. queenslandica*. The case of *C. elegans* is remarkable since it holds several Argonaute gene families and at least 24 proteins (Hoogstrate *et al.*, 2014). In agreement with other phylogenetic studies (Swarts *et al.*, 2014), the Argonaute proteins from plants and the majority of those from *C. elegans* formed distinct clades and, moreover, a clear separation was evident between AGO and PIWI proteins. Bivalve protein sequences clustered always separately forming one cluster for AGO-like hits and two clusters for PIWI-like proteins (Fig. 4).

Digital expression analysis of mussel and oyster miRNA biogenesis genes

We used the 13 Mg and 124 Cg RNA-seq samples to evaluate the expression levels of miRNA biogenesis genes in different tissues and conditions. Based on total mapped reads, we computed TPM values and we used elongation factor 1 α (El1 α) as normalizer housekeeping gene to compare the expression level of the different genes in each sample.

For Mg, the sequence analysis indicated a scarce basal expression of the genes mentioned above in five adult tissues: gill, digestive gland, haemolymph, muscle and mantle (below 2% of El1 α , except for DDX5, RAN and CNOT9). Mantle and muscle appeared the most responsive tissues whereas haemolymph was the least responsive one. In particular, the genes that we considered as the core components of miRNA biogenesis were expressed at levels below 0.5% of El1 α (File S6).

For Cg, we analyzed a considerable number of RNA-seq libraries representative of adult tissues (85) and developmental stages (39) (File S6). In adult oysters we observed low basal expression, as detected in the mussel samples. In fact, none of the experimental conditions reported for the analyzed RNA-seq samples influenced substantially the expression of the core miRNA pathway genes (expression levels below 2% of El1 α), with the exception of the high levels of CgPIWI-1 levels in male and female gonads (around 3.5%, Fig. 5). Conversely, most of the miRNA biogenesis genes were expressed at remarkable levels during the early stages of the oyster development: mainly from two cells to the rotary movement and, for some genes, also in the next developmental stages until *D-shaped* larvae, with no detectable signals afterward in spat and juveniles. Hence, these genes are particularly active in the early development, in particular one AGO (CGI_10020511) and two PIWI transcripts from the egg to trocophora (Fig. 5). In the same developmental stages we also noticed a

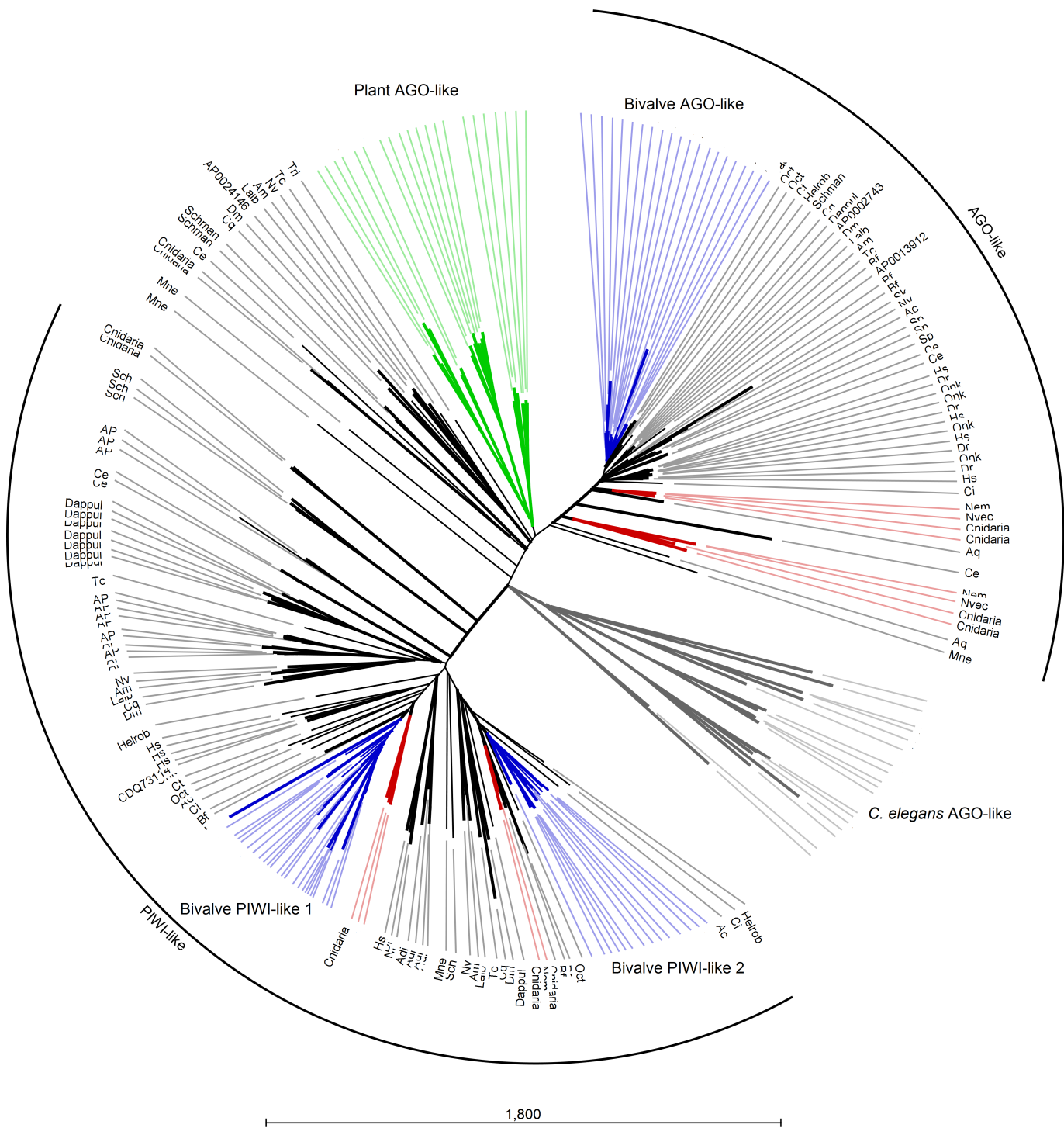


Figure 4 Phylogenetic relationships of Argonaute-like proteins. Proteins were aligned using MUSCLE and tree was generated using Neighbor Joining algorithm with 1,000 bootstrap replicates. Plant proteins are highlighted in green, whereas *C. elegans* hits are reported in grey. Blue lines represent mollusk hits, red lines represent hits from basal metazoans.

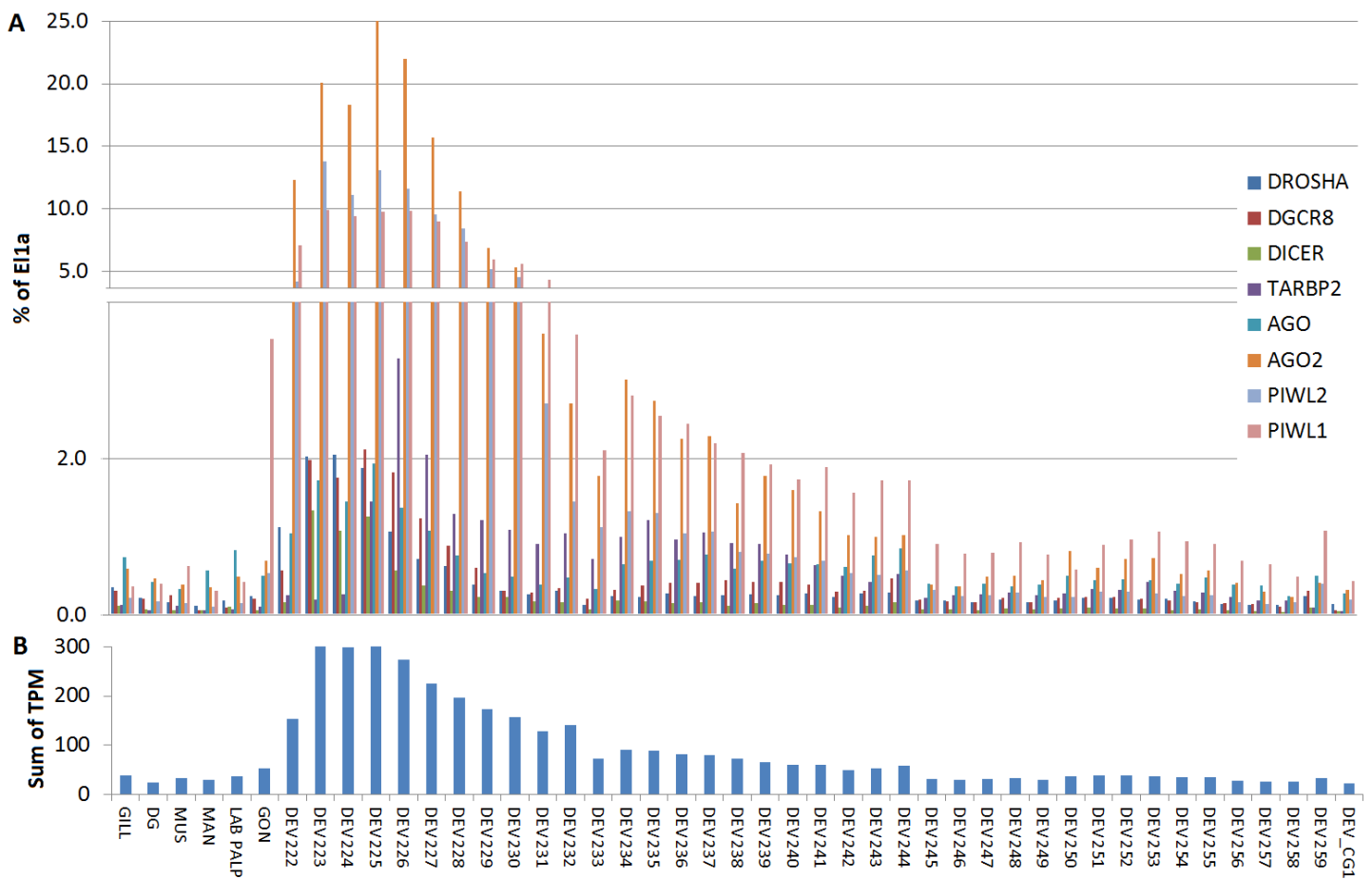


Figure 5 Digital expression analysis in oyster. The expression of the 8 miRNA biogenesis genes were computed in tissue-specific RNA libraries and in RNA libraries from different developmental phases. (A) Expression values represented as percentage of E11 α . (B) Cumulative TPM expression values of the 8 genes in the same samples.

remarkable expression of the key miRNA genes, with the co-expression of DROSHA and DGCR8 evident in all the analyzed samples.

DISCUSSION

Small RNAs are important regulators of the gene expression, as recognized in various model and non-model organisms (*Kim et al., 2014a; Kim et al., 2014b; Martini et al., 2014; Hussain & Asgari, 2014; Sahoo et al., 2014; Britton et al., 2014; Poole et al., 2014; Solofoharivelo et al., 2014*), including some bivalves (*Jiao et al., 2014; Zhou et al., 2014*). In addition to the identification of miRNAs, a general comprehension of the miRNA biogenesis in itself is also significant (*Grimson et al., 2008; Wu et al., 2011; Moran et al., 2013*). However, the main genes involved in miRNA formation in bivalves have not been described and characterized so far. In this study, we have provided an overview on the miRNA biogenesis complements in bivalves *spp.*, with particular attention to *M. galloprovincialis* and *C. gigas*. To the best of our knowledge, we report for the first

time the presence of a complete miRNA biogenesis pathway in *M. galloprovincialis*, the full-length transcript sequences of DICER, DGCR8, XPO5, RAN, DROSHA, TARBP2, three Argonaute genes and the identification of many other components that are candidate miRNA complement-interacting proteins such as MgGW182. By using local transcriptome assemblies, we identified these genes also in many other marine bivalves. The general low expression levels of these transcripts in the adult tissues of both *M. galloprovincialis* and *C. gigas*, and the considerable gene size, have probably prevented a previous identification of full-length sequences in not-well-covered bivalve transcriptomes. In fact, we obtained complete transcript sequences only from sequenced genomes or highly-covered transcriptomes whereas in other transcriptome assemblies we retrieved only few complete sequences. Overall, we have analyzed 523 miRNA complement sequences, 145 of them belonging to marine mollusks and displaying a consistent sequence clustering (*Ostreoida* and *Mytiloida* proteins generated two distinct clades, located always as sister group of arthropods).

However, the copy number of Argonaute genes somewhat differs among bivalves, as *C. gigas* and *A. californica* genomes coding for four proteins (2 AGO and 2 PIWI proteins) whereas *M. galloprovincialis* and *L. gigantea* possess three proteins (1 AGO and 2 PIWIs). We also highlighted the over-expression of the miRNA biogenesis genes during the first phases of the oyster development. A genome protection mechanism based on piRNA expression during early developmental stages is well known in mammals ([Malone & Hannon, 2009](#); [Kim et al., 2014a](#); [Kim et al., 2014b](#)) but such mechanism has not been reported in bivalves and additional investigations are necessary.

Finally, the identification of several mussel proteins either necessary or cooperative in the miRNA biogenesis, supports the existence of a complete and functional miRNA pathway in mussels and, probably, in other bivalves. Up to now, protein-protein or protein-RNA interaction data are not available for bivalve *spp.* and these topics may represent a direction of work in the future. Meanwhile, the expression analyses of miRNA biogenesis genes coupled with the identification of the miRNAs expressed in naturally infected and laboratory-treated bivalves could provide both validation and new insights on these interesting processes.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work is supported by PRIN 2010-11 (20109XZEPR) and FP7-KBBE-2010-4-266157 (BIVALIFE). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

PRIN2010-11: 20109XZEPR.

BIVALIFE: FP7-KBBE-2010-4-266157.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Umberto Rosani conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables.
- Alberto Pallavicini conceived and designed the experiments, wrote the paper, reviewed drafts of the paper.
- Paola Venier conceived and designed the experiments, contributed reagents/materials/-analysis tools, wrote the paper, reviewed drafts of the paper.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

Genbank: [KT447251](#), [KT447252](#), [KT447259](#), [KT447254](#), [KT447258](#), [KT447253](#), [KT447257](#), [KT447255](#), [KT447256](#), [KT447250](#), [KT694355](#), [KT694356](#), [KT694357](#), [KT694358](#), [KT694359](#), [KT694360](#), [KT694361](#), [KT694362](#), [KT694363](#), [KT694364](#), [KT694365](#), [KT694366](#), [KT694367](#), [KT694368](#), [KT694369](#), [KT694370](#), [KT694371](#), [KT694372](#), [KT694373](#), [KT694374](#).

Data Availability

The following information was supplied regarding data availability:

The research in this article did not generate any raw data.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.1763#supplemental-information>.

REFERENCES

- Ambros V. 2003.** MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell* **113**(6):673–676 DOI [10.1016/S0092-8674\(03\)00428-8](https://doi.org/10.1016/S0092-8674(03)00428-8).
- Bartel DP. 2004.** MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**(2):281–297 DOI [10.1016/S0092-8674\(04\)00045-5](https://doi.org/10.1016/S0092-8674(04)00045-5).
- Berezikov E, Guryev V, Van de Belt J, Wienholds E, Plasterk RHA, Cuppen E. 2005.** Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* **120**(1):21–24 DOI [10.1016/j.cell.2004.12.031](https://doi.org/10.1016/j.cell.2004.12.031).
- Britton C, Winter AD, Gillan V, Devaney E. 2014.** microRNAs of parasitic helminths—identification, characterization and potential as drug targets. *International Journal for Parasitology: Drugs and Drug Resistance* **4**(2):85–94 DOI [10.1016/j.ijpddr.2014.03.001](https://doi.org/10.1016/j.ijpddr.2014.03.001).
- Carmell MA, Girard A, Van de Kant HJG, Bourc’his D, Bestor TH, De Rooij DG, Hannon GJ. 2007.** MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental Cell* **12**(4):503–514 DOI [10.1016/j.devcel.2007.03.001](https://doi.org/10.1016/j.devcel.2007.03.001).
- Carninci P, The FANTOM Consortium. 2005.** The transcriptional landscape of the mammalian genome. *Science* **309**(5740):1559–1563 DOI [10.1126/science.1112014](https://doi.org/10.1126/science.1112014).

- Castresana J. 2000.** Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17(4):540–552 DOI [10.1093/oxfordjournals.molbev.a026334](https://doi.org/10.1093/oxfordjournals.molbev.a026334).
- Chen G, Zhang C, Jiang F, Wang Y, Xu Z, Wang C. 2014.** Bioinformatics analysis of hemocyte miRNAs of scallop *Chlamys farreri* against acute viral necrobiosis virus (AVNV). *Fish & Shellfish Immunology* 37(1):75–86 DOI [10.1016/j.fsi.2014.01.002](https://doi.org/10.1016/j.fsi.2014.01.002).
- Denli AM, Tops BBJ, Plasterk RHA, Ketting RF, Hannon GJ. 2004.** Processing of primary microRNAs by the microprocessor complex. *Nature* 432(7014):231–235 DOI [10.1038/nature03049](https://doi.org/10.1038/nature03049).
- Diebel KW, Oko LM, Medina EM, Niemeyer BF, Warren CJ, Claypool DJ, Tibbetts SA, Cool CD, Clambey ET, Van Dyk LF. 2015.** Gammaherpesvirus small noncoding RNAs are bifunctional elements that regulate infection and contribute to virulence *in vivo*. *mBio* 6(1):e01670–e01614 DOI [10.1128/mBio.01670-14](https://doi.org/10.1128/mBio.01670-14).
- Eddy SR. 2011.** Accelerated profile HMM searches. *PLoS Computational Biology* 7(10):e1002195 DOI [10.1371/journal.pcbi.1002195](https://doi.org/10.1371/journal.pcbi.1002195).
- Edgar RC. 2004.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5):1792–1797 DOI [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340).
- Ender C, Meister G. 2010.** Argonaute proteins at a glance. *Journal of Cell Science* 123(11):1819–1823 DOI [10.1242/jcs.055210](https://doi.org/10.1242/jcs.055210).
- Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL, Newcomb JM, Sempere LF, Flatmark K, Hovig E, Peterson KJ. 2015.** A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. *Annual Review of Genetics* 49(1):213–242 DOI [10.1146/annurev-genet-120213-092023](https://doi.org/10.1146/annurev-genet-120213-092023).
- Fabian MR, Sonenberg N. 2012.** The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nature Structural & Molecular Biology* 19(6):586–593 DOI [10.1038/nsmb.2296](https://doi.org/10.1038/nsmb.2296).
- Gammon DB, Mello CC. 2015.** RNA interference-mediated antiviral defense in insects. *Current Opinion in Insect Science* 8(April):111–120 DOI [10.1016/j.cois.2015.01.006](https://doi.org/10.1016/j.cois.2015.01.006).
- Gao Z, Wang M, Blair D, Zheng Y, Dou Y. 2014.** Phylogenetic analysis of the endoribonuclease Dicer family. *PLoS ONE* 9(4):e95350 DOI [10.1371/journal.pone.0095350](https://doi.org/10.1371/journal.pone.0095350).
- Gerdol M, De Moro G, Manfrin C, Milandri A, Riccardi E, Beran A, Venier P, Pallavicini A. 2014.** RNA Sequencing and de novo assembly of the digestive gland transcriptome in *Mytilus galloprovincialis* fed with toxinogenic and non-toxic strains of *Alexandrium minutum*. *BMC Research Notes* 7:722 DOI [10.1186/1756-0500-7-722](https://doi.org/10.1186/1756-0500-7-722).
- Ghildiyal M, Zamore PD. 2009.** Small silencing RNAs: an expanding universe. *Nature Reviews Genetics* 10(2):94–108 DOI [10.1038/nrg2504](https://doi.org/10.1038/nrg2504).
- GIGA Community of Scientists. 2014.** The global invertebrate genomics alliance (GIGA): developing community resources to study diverse invertebrate genomes. *The Journal of Heredity* 105(1):1–18 DOI [10.1093/jhered/est084](https://doi.org/10.1093/jhered/est084).
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011.**

- Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* **29**(7):644–652 DOI 10.1038/nbt.1883.
- Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP. 2008.** Early origins and evolution of microRNAs and piwi-interacting RNAs in animals. *Nature* **455**(7217):1193–1197 DOI 10.1038/nature07415.
- Hong S, Noh H, Chen H, Padia R, Pan ZK, Su S-B, Jing Q, Ding H-F, Huang S. 2013.** Signaling by p38 MAPK stimulates nuclear localization of the microprocessor component p68 for processing of selected primary microRNAs. *Science Signaling* **6**(266):ra16 DOI 10.1126/scisignal.2003706.
- Hoogstrate SW, Volkers RJM, Sterken MG, Kammenga JE, Snoek LB. 2014.** Nematode endogenous small RNA pathways. *Worm* **3**(March) DOI 10.4161/worm.28234.
- Huang J-T, Wang J, Srivastava V, Sen S, Liu S-M. 2014.** MicroRNA machinery genes as novel biomarkers for cancer. *Frontiers in Oncology* **4**:13 DOI 10.3389/fonc.2014.00113.
- Huarte M. 2015.** The emerging role of lncRNAs in cancer. *Nature Medicine* **21**(11):1253–1261 DOI 10.1038/nm.3981.
- Hussain M, Asgari S. 2014.** MicroRNAs as mediators of insect host–pathogen interactions and immunity. *Journal of Insect Physiology* **70**:151–158 DOI 10.1016/j.jinsphys.2014.08.003.
- Iwasaki YW, Siomi MC, Siomi H. 2015.** PIWI-interacting RNA: its biogenesis and functions. *Annual Review of Biochemistry* **84**:405–433 DOI 10.1146/annurev-biochem-060614-034258.
- Jean-Philippe J, Paz S, Caputi M. 2013.** hnRNP A1: the Swiss army knife of gene expression. *International Journal of Molecular Sciences* **14**(9):18999–19024 DOI 10.3390/ijms140918999.
- Jiao Y, Zheng Z, Du X, Wang Q, Huang R, Deng Y, Shi S, Zhao X. 2014.** Identification and characterization of MicroRNAs in pearl oyster *pinctada martensii* by solexa deep sequencing. *Marine Biotechnology* **16**(1):54–62 DOI 10.1007/s10126-013-9528-x.
- John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. 2004.** Human microRNA targets. *PLoS Biology* **2**(11):e363 DOI 10.1371/journal.pbio.0020363.
- Kapinas K, Delany AM. 2011.** MicroRNA biogenesis and regulation of bone remodeling. *Arthritis Research & Therapy* **13**(3):Article 220 DOI 10.1186/ar3325.
- Kenny NJ, Namigai EKO, Marlétaz F, Hui JHL, Shimeld SM. 2015.** Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans *patella vulgata* (Mollusca, Patellogastropoda) and *spirobranchus (Pomatoceros) lamarcki* (Annelida, Serpulida). *Marine Genomics* **24**:139–146 DOI 10.1016/j.margen.2015.07.004.
- Kim S, Günesdogan U, Zylicz JJ, Hackett JA, Cougot D, Bao S, Lee C, Dietmann S, Allen GE, Sengupta R, Surani MA. 2014b.** PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. *Molecular Cell* **56**(4):564–579 DOI 10.1016/j.molcel.2014.10.003.

- Kim M-C, Lee S-W, Ryu D-Y, Cui F-J, Bhak J, Kim Y. 2014a.** Identification and characterization of microRNAs in normal equine tissues by next generation sequencing. *PLoS ONE* **9(4)**:e93662 DOI [10.1371/journal.pone.0093662](https://doi.org/10.1371/journal.pone.0093662).
- Kincaid RP, Sullivan CS. 2012.** Virus-encoded microRNAs: an overview and a look to the future. *PLoS Pathogens* **8(12)**:e1003018 DOI [10.1371/journal.ppat.1003018](https://doi.org/10.1371/journal.ppat.1003018).
- Kozlov G, Safaee N, Rosenauer A, Gehring K. 2010.** Structural basis of binding of P-body-associated proteins GW182 and Ataxin-2 by the Mllet domain of poly(A)-binding protein. *The Journal of Biological Chemistry* **285(18)**:13599–13606 DOI [10.1074/jbc.M109.089540](https://doi.org/10.1074/jbc.M109.089540).
- Kozomara A, Griffiths-Jones S. 2014.** miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* **42**:D68–D73 DOI [10.1093/nar/gkt1181](https://doi.org/10.1093/nar/gkt1181).
- Lau P-W, MacRae IJ. 2009.** The molecular machines that mediate microRNA maturation. *Journal of Cellular and Molecular Medicine* **13(1)**:54–60 DOI [10.1111/j.1582-4934.2008.00520.x](https://doi.org/10.1111/j.1582-4934.2008.00520.x).
- Letunic I, Doerks T, Bork P. 2012.** SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research* **40(D1)**:D302–D305 DOI [10.1093/nar/gkr931](https://doi.org/10.1093/nar/gkr931).
- Lucas K, Raikhel AS. 2013.** Insect microRNAs: biogenesis, expression profiling and biological functions. *Insect Biochemistry and Molecular Biology* **43(1)**:24–38 DOI [10.1016/j.ibmb.2012.10.009](https://doi.org/10.1016/j.ibmb.2012.10.009).
- MacRae IJ, Ma E, Zhou M, Robinson CV, Doudna JA. 2008.** *In vitro* reconstitution of the human RISC-loading complex. *Proceedings of the National Academy of Sciences of the United States of America* **105(2)**:512–517 DOI [10.1073/pnas.0710869105](https://doi.org/10.1073/pnas.0710869105).
- Malone CD, Hannon GJ. 2009.** Small RNAs as guardians of the genome. *Cell* **136(4)**:656–668 DOI [10.1016/j.cell.2009.01.045](https://doi.org/10.1016/j.cell.2009.01.045).
- Martín-Gómez L, Villalba A, Kerkhoven RH, Abollo E. 2014.** Role of microRNAs in the immunity process of the flat oyster *ostrea edulis* against bonamiosis. *Infection, Genetics and Evolution* **27(October)**:40–50 DOI [10.1016/j.meegid.2014.06.026](https://doi.org/10.1016/j.meegid.2014.06.026).
- Martini P, Sales G, Brugiolo M, Gandaglia A, Naso F, De Pittà C, Spina M, Gerosa G, Chemello F, Romualdi C, Cagnin S, Lanfranchi G. 2014.** Tissue-specific expression and regulatory networks of pig microRNAome. *PLoS ONE* **9(4)**:e89755 DOI [10.1371/journal.pone.0089755](https://doi.org/10.1371/journal.pone.0089755).
- Maxwell EK, Ryan JF, Schnitzler CE, Browne WE, Baxevanis AD. 2012.** MicroRNAs and essential components of the microRNA processing machinery are not encoded in the genome of the ctenophore *Mnemiopsis leidyi*. *BMC Genomics* **13**:714 DOI [10.1186/1471-2164-13-714](https://doi.org/10.1186/1471-2164-13-714).
- Mendell JT, Olson EN. 2012.** MicroRNAs in stress signaling and human disease. *Cell* **148(6)**:1172–1187 DOI [10.1016/j.cell.2012.02.005](https://doi.org/10.1016/j.cell.2012.02.005).
- Min P-K, Chan SY. 2015.** The biology of circulating microRNAs in cardiovascular disease. *European Journal of Clinical Investigation* **45**:860–874 DOI [10.1111/eci.12475](https://doi.org/10.1111/eci.12475).

- Miyoshi K, Okada TN, Siomi H, Siomi MC. 2009.** Characterization of the miRNA-RISC loading complex and miRNA-RISC formed in the Drosophila miRNA pathway. *RNA* **15**(7):1282–1291 DOI [10.1261/rna.1541209](https://doi.org/10.1261/rna.1541209).
- Moran Y, Praher D, Fredman D, Technau U. 2013.** The evolution of microRNA pathway protein components in Cnidaria. *Molecular Biology and Evolution* **30**(12):2541–2552 DOI [10.1093/molbev/mst159](https://doi.org/10.1093/molbev/mst159).
- Moreira R, Pereiro P, Canchaya C, Posada D, Figueras A, Novoa B. 2015.** RNA-Seq in *mytilus galloprovincialis*: comparative transcriptomics and expression profiles among different tissues. *BMC Genomics* **16**(1):728 DOI [10.1186/s12864-015-1817-5](https://doi.org/10.1186/s12864-015-1817-5).
- Mortimer SA, Kidwell MA, Doudna JA. 2014.** Insights into RNA structure and function from genome-wide studies. *Nature Reviews. Genetics* **15**(7):469–479 DOI [10.1038/nrg3681](https://doi.org/10.1038/nrg3681).
- Mukherjee K, Campos H, Kolaczowski B. 2013.** Evolution of animal and plant dicers: early parallel duplications and recurrent adaptation of antiviral RNA binding in plants. *Molecular Biology and Evolution* **30**(3):627–641 DOI [10.1093/molbev/mss263](https://doi.org/10.1093/molbev/mss263).
- Nguyen TTT, Hayes BJ, Ingram BA. 2014.** Genetic parameters and response to selection in blue mussel (*mytilus galloprovincialis*) using a SNP-based pedigree. *Aquaculture* **420–421**(January):295–301 DOI [10.1016/j.aquaculture.2013.11.021](https://doi.org/10.1016/j.aquaculture.2013.11.021).
- Nottrott S, Simard MJ, Richter JD. 2006.** Human Let-7a miRNA blocks protein production on actively translating polyribosomes. *Nature Structural & Molecular Biology* **13**(12):1108–1114 DOI [10.1038/nsmb1173](https://doi.org/10.1038/nsmb1173).
- Piao X, Zhang X, Wu L, Belasco JG. 2010.** CCR4-NOT deadenylates mRNA associated with RNA-induced silencing complexes in human cells. *Molecular and Cellular Biology* **30**(6):1486–1494 DOI [10.1128/MCB.01481-09](https://doi.org/10.1128/MCB.01481-09).
- Poole CB, Gu W, Kumar S, Jin J, Davis PJ, Bauche D, McReynolds LA. 2014.** Diversity and expression of microRNAs in the filarial parasite, *brugia malayi*. *PLoS ONE* **9**(5):e96498 DOI [10.1371/journal.pone.0096498](https://doi.org/10.1371/journal.pone.0096498).
- Qu S, Yang X, Li X, Wang J, Gao Y, Shang R, Sun W, Dou K, Li H. 2015.** Circular RNA: a new star of noncoding RNAs. *Cancer Letters* **365**:141–148 DOI [10.1016/j.canlet.2015.06.003](https://doi.org/10.1016/j.canlet.2015.06.003).
- Resch AM, Palakodeti D. 2012.** Small RNA pathways in *schmidtea mediterranea*. *The International Journal of Developmental Biology* **56**(1–3):67–74 DOI [10.1387/ijdb.113436ar](https://doi.org/10.1387/ijdb.113436ar).
- Rhee S, Chae H, Kim S. 2015.** PlantMirnaT: miRNA and mRNA integrated analysis fully utilizing characteristics of plant sequencing data. *Methods* **83**:80–87 DOI [10.1016/j.ymeth.2015.04.003](https://doi.org/10.1016/j.ymeth.2015.04.003).
- Ronquist F, Teslenko M, Van der Mark P, Ayres D, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012.** MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**(3):539–542 DOI [10.1093/sysbio/sys029](https://doi.org/10.1093/sysbio/sys029).
- Ruan X. 2015.** Long noncoding RNA central of glucose homeostasis. *Journal of Cellular Biochemistry* Epub ahead of print Nov 26 2015 DOI [10.1002/jcb.25427](https://doi.org/10.1002/jcb.25427).

- Sahoo G, Ansari MD, Dikhit M, Gupta N, Rana S, Das P. 2014.** Computational identification of microRNA-like elements in leishmania major. *MicroRNA* 2(3):225–230 DOI 10.2174/2211536602666131203232422.
- Sand M, Skrygan M, Georgas D, Sand D, Gambichler T, Altmeyer P, Bechara FG. 2012.** The miRNA machinery in primary cutaneous malignant melanoma, cutaneous malignant melanoma metastases and benign melanocytic nevi. *Cell and Tissue Research* 350(1):119–126 DOI 10.1007/s00441-012-1446-0.
- Siomi MC, Sato K, Pezic D, Aravin AA. 2011.** PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Reviews Molecular Cell Biology* 12(4):246–258 DOI 10.1038/nrm3089.
- Sokolova OA, Iakushev EI, Stoliarenko AD, Mikhaleva EA, Gvozdev VA, Klenov MS. 2011.** The interplay of transposon silencing genes in the *Drosophila melanogaster* germline. *Molekuliarnaia Biologiia* 45(4):633–641.
- Solofoharivelo MC, Van der Walt AP, Stephan D, Burger JT, Murray SL. 2014.** MicroRNAs in fruit trees: discovery, diversity and future research directions. *Plant Biology* 16(5):856–865 DOI 10.1111/plb.12153.
- Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, Van der Oost J. 2014.** The evolutionary journey of argonaute proteins. *Nature Structural & Molecular Biology* 21(9):743–753 DOI 10.1038/nsmb.2879.
- Tarver JE, Cormier A, Pinzón N, Taylor RS, Carré W, Strittmatter M, Seitz H, Coelho SM, Cock JM. 2015.** microRNAs and the evolution of complex multicellularity: identification of a large, diverse complement of microRNAs in the brown alga ectocarpus. *Nucleic Acids Research* 43:6384–6398 DOI 10.1093/nar/gkv578.
- Tarver JE, Donoghue PCJ, Peterson KJ. 2012.** Do miRNAs have a deep evolutionary history? *BioEssays* 34(10):857–866 DOI 10.1002/bies.201200055.
- Théron E, Dennis C, Brasset E, Vaury C. 2014.** Distinct features of the piRNA Pathway in somatic and germ cells: from piRNA cluster transcription to piRNA processing and amplification. *Mobile DNA* 5(1):28 DOI 10.1186/s13100-014-0028-y.
- Tolia NH, Joshua-Tor L. 2007.** Slicer and the argonautes. *Nature Chemical Biology* 3(1):36–43 DOI 10.1038/nchembio848.
- Tomoyasu Y, Miller SC, Tomita S, Schoppmeier M, Grossmann D, Bucher G. 2008.** Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in tribolium. *Genome Biology* 9(1):Article R10 DOI 10.1186/gb-2008-9-1-r10.
- Van Kouwenhove M, Kedde M, Agami R. 2011.** MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nature Reviews Cancer* 11(9):644–656 DOI 10.1038/nrc3107.
- Venier P, De Pittà C, Bernante F, Varotto L, De Nardi B, Bovo G, Roch P, Novoa B, Figueras A, Pallavicini A, Lanfranchi G. 2009.** MytiBase: a knowledgebase of mussel (*M. galloprovincialis*) transcribed sequences. *BMC Genomics* 10(February):72 DOI 10.1186/1471-2164-10-72.
- Wagner GP, Kin K, Lynch VJ. 2013.** A model based criterion for gene expression calls using RNA-seq data. *Theory in Biosciences Theorie in Den Biowissenschaften* 132(3):159–164 DOI 10.1007/s12064-013-0178-3.

- Wu B, Jin M, Gong J, Du X, Bai Z. 2011.** Dynamic evolution of CIKS (TRAF3IP2/Act1) in metazoans. *Developmental & Comparative Immunology* **35(11)**:1186–1192 DOI [10.1016/j.dci.2011.03.027](https://doi.org/10.1016/j.dci.2011.03.027).
- Xu M-J, Qiu S-B, Nisbet AJ, Fu J-H, Shao C-C, Zhu X-Q. 2014.** Global characterization of microRNAs in trichomonas gallinae. *Parasites & Vectors* **7**:Article 99 DOI [10.1186/1756-3305-7-99](https://doi.org/10.1186/1756-3305-7-99).
- Xu C, Zhang L, Li H, Liu Z, Duan L, Lu C. 2015.** MiRNA-1469 promotes lung cancer cells apoptosis through targeting STAT5a. *American Journal of Cancer Research* **5(3)**:1180–1189.
- Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PWH, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A, Hedgecock D, Xu Z, Liu Y, Domazet-Lošo T, Du Y, Sun X, Zhang S, Liu B, Cheng P, Jiang X, Li J, Fan D, Wang W, Fu W, Wang T, Wang B, Zhang J, Peng Z, Li Y, Li N, Wang J, Chen M, He Y, Tan F, Song X, Zheng Q, Huang R, Yang H, Du X, Chen L, Yang M, Gaffney PM, Wang S, Luo L, She Z, Ming Y, Huang W, Zhang S, Huang B, Zhang Y, Qu T, Ni P, Miao G, Wang J, Wang Q, Steinberg CEW, Wang H, Li N, Qian L, Zhang G, Li Y, Yang H, Liu X, Wang J, Yin Y, Wang J. 2012.** The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490(7418)**:49–54 DOI [10.1038/nature11413](https://doi.org/10.1038/nature11413).
- Zhou Z, Wang L, Song L, Liu R, Zhang H, Huang M, Chen H. 2014.** The identification and characteristics of immune-related microRNAs in haemocytes of oyster *Crassostrea gigas*. *PLoS ONE* **9(2)**:e88397 DOI [10.1371/journal.pone.0088397](https://doi.org/10.1371/journal.pone.0088397).