https://doi.org/10.17816/ecogen172113-125

DIVERSITY OF RNA INTERFERENCE PATHWAYS IN REGULATION OF ENDOGENOUS AND EXOGENOUS SEQUENCES EXPRESSION IN CILIATES *TETRAHYMENA* AND *PARAMECIUM*

© I.V. Nekrasova, A.A. Potekhin

Saint Petersburg State University, Saint Petersburg, Russia

For citation: Nekrasova IV, Potekhin AA. Diversity of RNA interference pathways in regulation of endogenous and exogenous sequences expression in ciliates *Tetrahymena* and *Paramecium*. *Ecological genetics*. 2019;17(2):113-125. https://doi.org/10.17816/ecogen172113-125.

Received: *26.10.2018* Revised: *04.02.2019* Accepted: *17.06.2019*

❀ RNA interference plays a major role in biology of ciliates. Diverse small RNAs regulate many processes in vegetative cells of ciliates *Tetrahymena* and *Paramecium*. Different types of endogenous and exogenous nucleotide sequences induce different RNAi pathways resulting in silencing of the homologous sequences in the macronuclear genome. Likely this way ciliates are able to quickly inactivate heterogeneous sequences and to adapt efficiently to the environmental conditions and external stimuli.

❀ **Keywords:** ciliates; RNA interference; small RNAs; Dicer; Piwi; RNA-dependent RNA polymerases; gene silencing.

РАЗНООБРАЗИЕ ПУТЕЙ ИНТЕРФЕРЕНЦИИ РНК В РЕГУЛЯЦИИ ЭКСПРЕССИИ ЭНДОГЕННЫХ И ЭКЗОГЕННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ У ИНФУЗОРИЙ *TETRAHYMENA* И *PARAMECIUM*

© И.В. Некрасова, А.А. Потехин

ФГБОУ ВО «Санкт-Петербургский государственный университет», Санкт-Петербург

Для цитирования: Некрасова И.В., Потехин А.А. Разнообразие путей интерференции РНК в регуляции экспрессии эндогенных и экзогенных последовательностей у инфузорий *Tetrahymena* и *Paramecium* // Экологическая генетика. – 2019. – Т. 17. – № 2. – С. 113–125. https://doi.org/10.17816/ecogen172113-125.

Поступила: *26.10.2018* Одобрена: *04.02.2019* Принята: *17.06.2019*

❀ Интерференция РНК играет огромную роль в биологии инфузорий. Разнообразные малые РНК регулируют многие процессы в жизни вегетативных клеток инфузорий *Tetrahymena* и *Paramecium*. Разные типы эндогенных и экзогенных нуклеотидных последовательностей запускают разные пути интерференции РНК, приводящие к сайленсингу гомологичных последовательностей в геноме макронуклеуса. Вероятно, благодаря этим молекулярным механизмам инфузории способны быстро инактивировать чужеродные гены и эффективно адаптироваться к условиям среды обитания и внешним стимулам.

❀ **Ключевые слова:** инфузории; интерференция РНК; малые РНК; Dicer; Piwi; РНК-зависимые РНК-полимеразы; сайленсинг генов.

Gene silencing due to the activity of small RNAs plays important roles in the regulation of different processes in major eukaryotes [1]. The role of double-stranded RNA (dsRNA) in gene inactivation was demonstrated for the first time in the nematode *Caenorhabditis elegans*, via a mechanism called "RNA interference" [2]. Currently, the term RNA interference defines the silencing of any gene, including dsRNA molecules of any type [3] (Fig. 1). Dicer enzymes cut dsRNA molecules into small interfering RNAs (siRNAs), which then form complexes with proteins of the Argonaute/Piwi family. In such complexes, small RNAs become single-stranded, whereas the second strand undergoes degradation. Two types of such complexes have been distinguished: RNAinduced silencing complex (RISC) and RNAinduced transcriptional silencing (RITS). RISC complexes detect mRNA targets homologous to small RNAs included within them and cause their degradation. Small RNAs included within RITS complexes direct them to homologous regions of

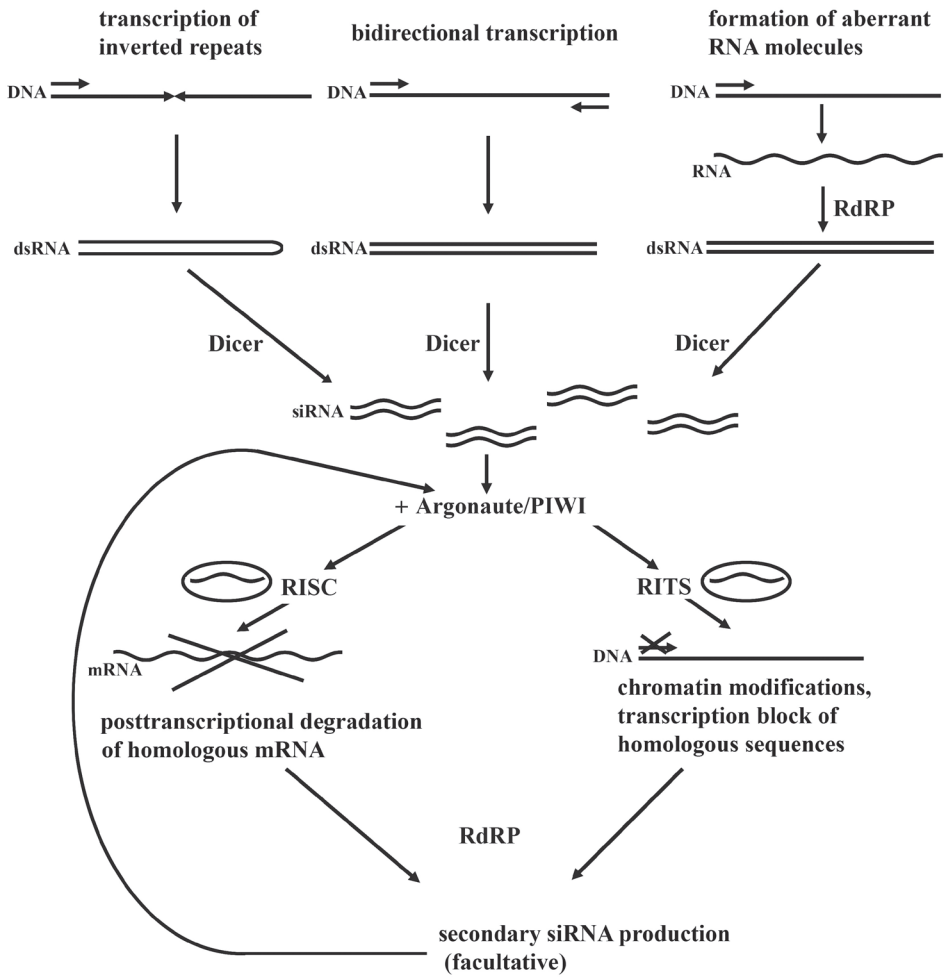

Fig. 1. General scheme of RNA interference pathways. дцРНК – double-stranded RNA; RdRP – RNA-dependent RNA polymerase; complexes of small RNAs and proteins: RISC – RNA-Induced Silencing Complex, RITS – RNA-Induced Transcriptional Silencing

DNA, which results in chromatic modification at these sites (see Fig. 1). Thus, silencing can be performed at both transcriptional and post-transcriptional levels by means of mRNA [3].

It is supposed that originally the main function of RNA interference was protection against the replication and expansion of extraneous sequences such as transposons and viruses [1]. The main proteins involved in RNA interference probably first arose in the common ancestor of all eukaryotes, which had a minimum of one Argonaute protein, one Piwi protein, one nuclease Dicer, and one RNA-dependent RNA polymerase (RDRP) [1]. Proteins of these families have been detected in major eukaryotes and are characterized by high conservation. In the most studied ciliates, *Paramecium tetraurelia* and *Tetrahymena thermophila* (belonging to the same class Oligohymenophorea), the RNA interference pathways differ but are served by sets of proteins of these families; only Argonaute proteins are lacking in ciliates. The most studied and effective option for achieving RNA interference in the cells of these protists is genome scanning, namely, formation of a somatic genome in the sexual process, when almost all noncoding DNA sequences are removed from the developing macronucleus [4]. Cores of two types are available in ciliates cells: (i) the multi-genome macronucleus (MAC), a somatic core, the genes of which are expressed, and (ii) the generative, usually diploid micronucleus (MIC) responsible for the storage of genetic information; it is transferred to the offspring during the sexual process. If in MIC genome of *Paramecium* and *Tetrahymena* approximately 30% of DNA is presented by noncoding repeats and their derivatives, transposons and their derivatives, and the genes contained there are not expressed, then the MAC genome has DNA sequences only required for the life of a ciliate. In the sexual process, the MAC of ciliates is destroyed and MIC meiosis products form a zygotic nucleus (synkaryon), the division of which results in formation of the anlages of new MICs and MACs. The formation of the MAC genome is controlled by complex epigenetic mechanisms, which are

based on RNA interference [5, 6]. While the majority of the noncoding sequences, transposons and their derivatives, internal eliminated sequences (IES) are removed from the formed somatic genome, with further stitching of their flanking areas, and fragmentation of chromosomes occurred into short "mini-chromosomes" without centromeres, to which telomere repeats are added *de novo* [4]. The removal of all noncoding sequences upon formation of the somatic genome during the sexual process is the first stage of ciliates protection against "excessive" sequences, which accumulate in the silent genome of the generative nucleus and are not selected against; they are effectively removed from the somatic genome.

However, different pathways of RNA interference function outside the sexual process. Different proteins of Dicer and Piwi families are involved in these pathways and in the mechanisms of genome scanning in both *Paramecium* and *Tetrahymena* [7–9]. This review is devoted to those biological phenomena of two model species, *P. tetraurelia* and *T. thermophila*, which are connected to the operation of RNA interference in vegetative cells.

PRIMARY AND SECONDARY SMALL INTERFERING RNAS OF CILIATES

siRNAs are continuously synthesized and present in the vegetative cells of ciliates, which can be primary or secondary and differ in their size and modifications. Their functions are not yet clear, but it is obvious that siRNAs are among the major regulators of gene expression in ciliates. siRNAs are always available in tetrahymena cells, having a length of 23–24 nt, which are antisense for MAC genes [10]. The introduction of a transgene sequence into the cells of *T. thermophila* using electroporation, with which the RNA molecules homologous to any gene and forming double-stranded hairpins are transcribed, causes the accumulation of 23–24 nt siRNAs and significant reduction of the level of mRNA homologous endogenous sequences [11]. As the major categories of siRNAs, microRNAs, and piRNAs of the other eukaryotes, these small RNAs almost always contain U at the 5'-end, and for the half of these siRNAs the 3'-end nucleotide, usually U also, is probably added outside the template [10] (Fig. 2).

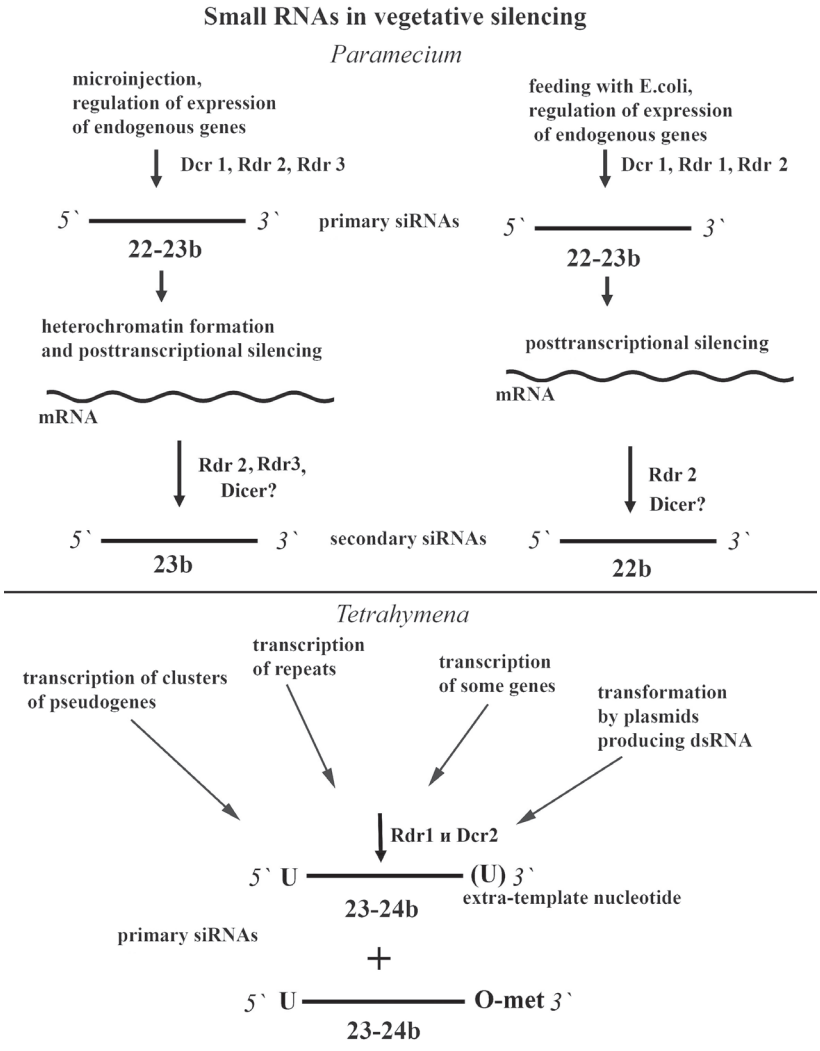

Fig. 2. Small RNAs involved in gene silencing in *Paramecium* and *Tetrahymena* vegetative cells. In *Paramecium* such small RNAs are subdivided into primary siRNAs produced by Dcr1p and RNA dependent RNA polymerases Rdr1p, Rdr2p, and Rdr3p, and secondary siRNAs. Induced by feeding with dsRNA-producing *E. coli* secondary siRNAs are synthesized by Rdr2p, which utilizes as template mRNA interacting with primary siRNAs. When silencing is induced by transgene microinjection, secondary siRNAs are produced not only by Rdr2p and Rdr3p, but also Dicer proteins are probably involved. Both pathways are likely to function when *Paramecium* regulates expression of endogenous genes. In *Tetrahymena* vegetative silencing may be induced by endogenous sequences, such as pseudogenes, repeats of different kinds, some genes, and also by exogenous vectors. In all cases Rdr1p and Dcr2p produce small RNAs, which possess methylated ribose or non-template U at 3'-ends

The pathway of RNA interference in vegetative cells of paramecia can be artificially initiated using several methods: (i) micro-injection of a large number of copies of a transgene sequence, which results in the appearance of aberrant transcripts from both chains of the introduced transgene; (ii) micro-injection of plasmids, from which untranslated dsRNA is transcribed [12, 13]; or (iii) ciliates feeding on the bacteria *Escherichia coli* producing dsRNAs [14, 15]. In the last case, the required sequence is cloned into the vector between promoters T7 directed toward each other and used to transform *E. coli*, which then produce dsRNAs during IPTG induction. If ciliates are fed with such *E. coli*, then dsRNAs stimulate the silencing of paramecia complementary sequences, although the mechanism behind this is unknown [16]. In the same manner as during silencing caused by injection, the silencing caused by feeding form siRNAs from dsRNAs under effect of protein Dcr1, one of the Dicer enzymes of *P. tetraurelia* (see below). Detailed study of the siRNAs formed during the micro-injection of purified transgene sequences or plasmids with a transgene as inserts and during feeding demonstrated that their properties and mechanism of synthesis differ depending on the way in which the synthesis is activated [14, 17, 18] (see Fig. 2). For example, the proteins Rdr3 and Ptiwi14 are involved in transgene-induced silencing rather than in dsRNA-induced silencing [7, 14].

During feeding, *P. tetraurelia* forms primary siRNAs 23 nt long from both chains of introduced RNA duplex. Primary siRNAs amount to approximately of all small RNAs formed during feeding [17], and RNA-dependent RNA polymerases Rdr1 and Rdr2 are involved in their accumulation (see below). The role of RDRP in this process is unclear, as it is supposed that dsRNAs are immediately decomposed by Dicer ferment forming siRNAs. Its role probably involves increasing the amount of dsRNAs. It is also possible that RNAs are derived from digestive vacuoles in the cytoplasm in a single-stranded state and are converted to dsRNAs as a result of RDRP activity [17]. After transgene introduction, primary siRNAs 22–23 nt long are formed (see Fig. 2), which results in the silencing of homologous endogenous genes [12, 14]. Interestingly, during transgene introduction, their double-way transcription takes place, different ways of siRNA synthesis are activated on the antiparallel transcripts: proteins Rdr2 and Cid2 (see below) are responsible for synthesis of antisense RNAs on the template of silenced mRNAs, the synthesis of siRNAs from antisense transcripts is probably produced by Rdr3 [18]. Some primary siRNAs are homologous to the vestor fragments. The share of primary siRNAs during micro-injection amounts to 95%–99%. It is known that the proteins Dcr1 and Rdr3 are also involved in the silencing of genes of the surface antigens; that is, they participate in regulation of the expression of endogenous sequences not connected with transgene injection [18].

Then, secondary siRNAs appear to be synthesized on the template of endogenous sequences. Secondary siRNAs are formed as continuation of cascade started with appearance of primary siRNAs, this is not necessarily connected with transgene overexpression. Even a small amount of aberrantly untranslated transcripts can be sufficient to initiate this mechanism; theoretically, such a cascade can start as a background by transcribed pseudogenes or any other endogenous sequences. Secondary siRNAs of *P. tetraurelia* are 22–23 nt long [14], and RDPPs are responsible for their synthesis. There is no simple answer to the question of how secondary siRNAs are synthesized in *Paramecium*. Similarity in the structures of primary and secondary siRNAs during transgene-induced silencing may indicate that the mechanisms of their synthesis are very similar [18]. It is likely that, in the cases of both transgene-induced silencing and silencing caused by feeding, the secondary siRNAs are formed on the mRNA template as they are homologous to the areas of sequence outside the introduced transgene or insert in the plasmid used for feeding [17, 18], but the specific mechanism behind their formation is still unknown. The distinctive feature of transgene-induced secondary siRNAs is methylation of the 3'-end ribose and the availability of monophosphate at the 5'-end [14], which means that their synthesis involves both RDRP and Dicer proteins [14]. Thus, it can be supposed that long dsRNAs are synthesized on the mRNA template, which further serves as a substrate for Dicer protein. At the same time, the level of secondary siRNAs during feeding was shown not to be reduced in the case of Dcr1 protein dysfunction [17], which supports the existence of a Dicer-independent method of secondary siRNA synthesis on the mRNA template. Secondary siRNAs amount to only 1%–5% of all siRNAs, but they are homologous to the entire sequence of their mRNA target, including 3'-UTR, rather than only to a transgene introduced short sequence. The specific function of secondary siRNAs in paramecium cells is not completely clear. Their level is very low in comparison with that of primary siRNAs, and their role in establishing artificially induced silencing is insignificant [17]. The issue of secondary siRNA synthesis during the induction of silencing in vegetative cells of tetrahymena has not been studied yet.

For transgene-induced silencing, it has been demonstrated that siRNAs ensure post-transcriptional silencing, which requires the proteins Dcr1, Rdr2, Ptiwi13, and Cid2, as well as co-transcriptional silencing connected with the establishment of heterochromatic marks (which requires proteins Rdr3 and Ptiwi14, as well as proteins involved in post-transcriptional silencing). It is likely that both types of silencing partially overlap with each other in terms of their induction [18].

PROTEINS OF DICER FAMILY IN VEGETATIVE CELLS OF CILIATES

Proteins of the Dicer family are assigned to type III RNAases and may have up to six domains; some domains of these proteins are lacking in different eukaryotes. However, two domains having RNAase III activity are a conserved feature in almost all eukaryotes [1]. These proteins cut dsRNA forming short (21–28 nt) duplex molecules with terminal double-nucleotide protrusions and monophosphate residue at the 5'-end [19]. Three proteins of the Dicer family were detected in *T. thermophila*. The gene *DCR1* is expressed continuously, but *DCR1* expression peaks during cell starvation preceding the sexual process of conjugation [10]. Dysfunction of the protein Dcr1 shows no phenotypic effects [20]; both domains of RNAase III in Dcr1 are severely modified and Dcr1 in *Tetrahymena* is catalytically inactive. The main function of Dicerlike protein of tetrahymena Dcl1 involves the establishment of small RNAs responsible for formation of the somatic genome during the sexual process [6, 20, 21]. Its gene is expressed at the highest level in conjugated cells [10, 20]. The protein Dcr2 contains a helicase domain and two domains of RNAase III [10, 20]; its gene is expressed constitutively, and the protein is required for vegetative cell growth [20]. The Dcr2 protein has been shown to be involved in the silencing of genes in the vegetative cells of *T. thermophila*.

The evolution of the *Paramecium* genus included a minimum of three whole genome duplications; the two later ones took place directly before the divergence of sibling species of the *P. aurelia* complex [22], in which connection many genes in genome of *P. tetraurelia* are presented by the groups of ohnologs (copies resulted from whole genome duplications) having maximum up to eight members. In particular, eight genes encoding proteins with RNAase III domains were detected in the genome of *P. tetraurelia* [23, 24]; it is likely that all genes of the Dicer family in *P. tetraurelia* originated from one primeval gene [22]. Three out of eight genes (*DCR1*, *DCR2*, and *DCR3*) contain two domains each of RNAase III and two domains of helicase subdomain; that is, they encode typical proteins of Dicer [8]. The only catalytically active protein of Dicer in paramecia is Dcr1. Two other Dcr proteins have RNAase domains that are significantly modified; their gene knockout is associated with phenotypic effects. The gene *DCR1* in *P. tetraurelia* is expressed during the entire cell cycle, and its knockout results in the disordered synthesis of small RNAs with a size of 23 nt participating in the vegetative silencing of paramecia [8]. The genome of *P. tetraurelia* also has five more genes encoding Dicerlike proteins containing only RNAase domain: Dcl1–Dcl5 [8]. The functions of two of them, Dcl1 and Dcl4, have been determined; their genes *DCL1* and *DCL4* are expressed at low levels during the whole of vegetative cell growth [23]. The Dicerlike proteins Dcl2, Dcl3, and Dcl5 are responsible for the production of small RNAs providing rearrangement of the core structure during the sexual process [6, 8, 23]. Thus, similar to the case for the protein Dcr2 of tetrahymena, the only protein of the Dicer family responsible for gene silencing in vegetative cells of *P. tetraurelia* is Dcr1.

ROLE OF RNA-DEPENDENT RNA POLYMERASES IN VEGETATIVE SILENCING

RNA-dependent RNA polymerases exist in RNA-containing viruses. It was shown that, in different eukaryotes, these ferments ensure the strengthening of silencing via siRNA amplification [25]. The genome of *T. thermophila* contains only one RDR gene, which is constitutively expressed throughout the entire life cycle [26]. This gene encodes polymerase Rdr1 and is vitally important [26]. This RDRP synthesizes ling dsRNAs, which are then cut by ferment Dcr2 into small RNAs with a size of 23 and 24 nt [26]. Rdr1 and Dcr2 physically associate with each other, which stimulates the activity of Dcr2 [26]. Besides, Rdr1 in *T. thermophila*, as well as all RDRPs, forms complexes (RDRC, RNAdirected RNApolymerase complexes) interacting with different supplementary proteins (Table 1 and Fig. 3).

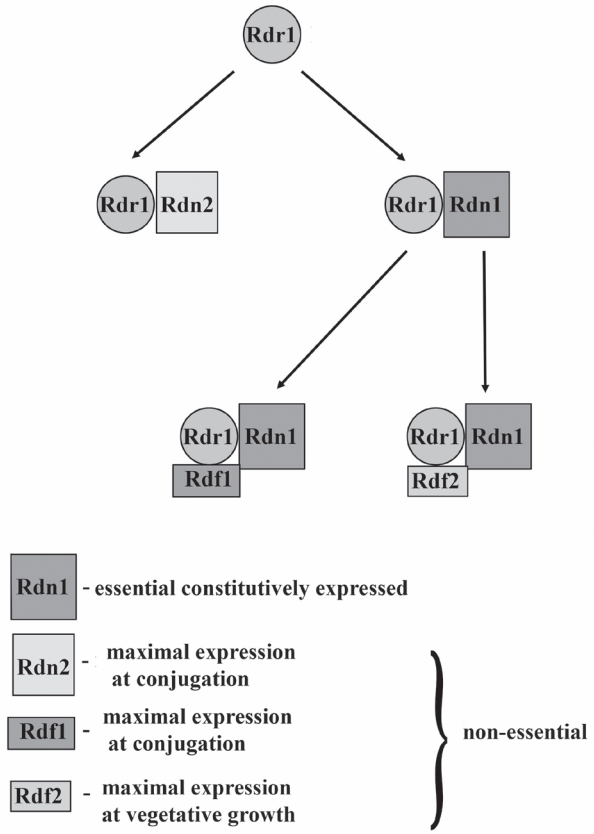

Fig. 3. Possible complexes between RDRP of *T. thermophila* Rdr1p and accessory proteins Rdn1p, Rdn2p, Rdf1p, and Rdf2p

Table 1

RDRPs and Dicer proteins involved in vegetative silencing in *Tetrahymena* and *Paramecium* [по 8, 14, 17, 20, 24, 26, 27]

Homolog in *Tetrahymena*	Function in *Tetrahymena*	Interactions with other proteins	Homolog in *Paramecium*	Functions in *Paramecium*	Interactions with other proteins
Rdr1p	Production of siRNAs; essential	Dcr2, Rdn1 and Rdn2, Rdf1 and Rdf2	Rdr1	Production of primary siRNAs in feeding with *E. coli*; non-essential	Dcr1 (presumably)
			Rdr2	Production of siRNAs in feeding with *E. coli* and in transgene microinjection; essential	unknown
–	–	–	Rdr3	Production of siRNAs in transgene microinjection	unknown
–	–	–	Rdr4	Pseudogene (presumably)	–
Rdn1p	Accumulation of siRNAs; uridinilation of RNA (presumably); essential	Rdr1; Rdf1, Rdf2	Cid2	Accumulation of siRNAs; essential	unknown
Rdn2p	Accumulation of antisense siRNAs for clusters of pseudogenes; uridinilation of RNA (presumably); non-essential	Rdr1	Cid1	Accumulation of siRNAs; non-essential	unknown
Rdf1p	Accumulation of siRNAs; non-essential during vegetative growth; when absent, nuclear differentiation at conjugation is blocked	Rdr1, Rdn1	–	–	–
Rdf2p	Accumulation of siRNAs produced from DNA hairpins; non-essential	Rdr1, Rdn1	–	–	–
Dcr2p	Production of siRNAs in vegetative cells; essential	Rdr1	Dcr1	Production of siRNAs in vegetative cells; essential	Rdr1 (presumably)

Four major proteins that can be contained within these complexes were identified [27]. Two protein paralogs, Rdn1 and Rdn2 (Rdr1associated nucleotidyl transferases), have homology with poly(A)-polymerase [27] and catalyze the addition of uridine to RNA substrates *in vitro* outside of the template [27]. Interestingly, the patterns of their gene expression differ: *RDN1* is expressed throughout the entire life cycle in both the vegetative phase and during conjugation, whereas the level of expression of *RDN2* peaks during conjugation, when the level of *RDN1* expression is rather low. Two other proteins associated with Rdr1 lacking domains with certain functions were identified and called Rdf1 and Rdf2 (Rdr1associated factors). The genes encoding them are arranged in the genome in tandem, which indicates that they originated as a result of duplication. Their expression patterns also differ: the level of expression of *RDF2* peaks during vegetative growth and is reduced during conjugation, whereas *RDF1* is expressed at its maximum level during conjugation and is rather weakly expressed in vegetative cells.

It turned out that the proteins Rdn1 and Rdn2 are contained within different RDRCs, and only complexes containing Rdn1 can recruit the proteins Rdf1 and Rdf2 [27]. Thus, the protein Rdr1 of *T. thermophila* can form several different complexes via the mutually exclusive interaction of Rdr1 with one of two nucleotidyl transferases. The *in vitro* function of all of these complexes cannot be differentiated, but the knockout of some components affects the phenotype in different ways [27]. *In vitro* complexes containing Rdn1 or Rdn2 have similar activity: both complexes can interact with protein Dcr2 and promote the cutting of RNA, forming siRNAs 24 nt long. The loss of a subunit of Rdn2 results *in vivo* in disturbance of the accumulation of certain siRNAs; loss of Rdf2 prevents the accumulation of other siRNAs. Loss of Rdf1 and Rdf2 affects the distribution of DNA during cell division, whereas the knockout of *RDN2* or *RDF1* blocks conjugation. It is clear that, despite *in vitro* Rdn1 and Rdn2 having similar biochemical activities, their roles *in vivo* are significantly different. One of the potential functions of the proteins Rdn1 and Rdn2 of *T. thermophila* is uridiniliration of the single-stranded RNA template for Rdr1, which can enhance specificity via the stabilization of 3'-end RNA; alternatively, Rdn1 and Rdn2 can operate on the duplex of siRNAs as the half of small RNAs 23–24 nt has uridine at the 3'-end outside the template [27].

In *P. tetraurelia*, there are four genes encoding RDRP. It was demonstrated that polymerase Rdr1 is involved in silencing caused by feeding [14], and ferment Rdr2 is required for the synthesis of siRNA during silencing caused by feeding, as well as during silencing in response to transgene injection [17, 24]. Meanwhile, Rdr3 functions in the synthesis of siRNAs during silencing caused by plasmid injection, from which dsRNA is read [14, 24] (see Fig. 2); this ferment participates in the regulation of gene expression of the surface antigens [14]. *RDR4* is probably a pseudogene, as indicated by the very low level of its expression and the complete absence of phenotypic features upon its silencing. The genes *RDR1* and *RDR4* are paralogs that appeared after the second complete duplication of the genome [14]. Detailed analysis of the sequences of amino acids of four RDRPs of paramecia demonstrated that, in comparison with well-known RDRPs of other organisms, Rdr1 and Rdr2 of paramecia are more conserved and have very close identity to the protein Rdr1 of *T. thermophila* (see Table 1). The genes *RDR1* and *RDR2* are expressed at a low level within the vegetative phase of the life cycle. The protein Dcr1 is immediately responsible for the synthesis of primary siRNAs. Severe reduction of the level of Dcr1 or proteins Rdr1 and Rdr2 results in disturbance of the accumulation of siRNAs 23 nt long [8]. The reasons why two different RDRPs besides Dicer are needed for their formation remain unclear. It is likely that RDRPs are required for amplification of the trigger dsRNA if it comes from the digestive vacuoles to the cytoplasm in very small quantities. Alternatively, bacterial RNAs are imported from digestive vacuoles in single-stranded form; then, the role of RDRP can consist of the synthesis of a complementary second chain for the activation of RNA interference [17]. It is also possible that Rdr1 plays a structural role in the interaction with Dcr1, for example, as occurs in *Tetrahymena* (see above).

A gene encoding a protein of the family of the nonclassic poly(A/U)-nucleotidyl transferases that was detected in *P. tetraurelia* [24] was also found in *S. pombe*, in which one of these ferments, Cid1, is involved in RNA interference. This gene in *P. tetraurelia*, called *CID1*, is an ortholog of the gene *RDN2* in *T. thermophila*. *CID1* is constitutively expressed throughout the entire cell cycle of paramecium [24]. Its knockout removes the ability of cells to undergo RNA interference induced by exogenous RNA, and null mutants of *CID1* are viable throughout the entire cell cycle, with no phenotypic anomalies being detected [24]. Further analysis of the genome of *P. tetraurelia* enabled the detection of 22 Cid1like poly(A)nucleotidyl transferases, five of which, including Cid1, are allied to the proteins Rdn1 and Rdn2 of *T. thermophila* (see Table 1). Among them, Cid2 is also involved in dsRNA-induced silencing and is probably of vital importance [24].

PROTEINS OF PIWI FAMILY IN VEGETATIVE CELLS OF CILIATES

Proteins of the Argonaute/Piwi family are the most conserved participants in RNA interference. Two paralogous groups can be distinguished in this family: Piwilike

proteins functioning in certain stages of animal development and Argonautelike (Ago) proteins interacting with different classes of small RNA in most eukaryotes. Ciliates have only Piwi proteins [7, 28]. The main domains of Piwi proteins are PAZ and Piwi [29]. The Piwi domain exerts RNAase activity [7]. The 5'-end of small RNAs interacts with the Piwi domain site, and the PAZ domain binds to their 3'-end [7]. The formation of a complex with small RNAs results in Piwi proteins cutting one of the chains, and it dissociates from the complex [30].

Proteins of the Piwi family of *T. thermophila* are called Twi (T*etrahymena* Piwi) [1]. In total, 12 Twi proteins (Twi1–Twi12) are encoded in the genome of *T. thermophila*, and their major role is involved in RNA interference in vegetative cells [28]. Exceptions are found for the proteins Twi1 and Twi11, which participate in the nuclear rearrangements in the sexual process [9]. The genes encoding the proteins Twi2, Twi8, and Twi12 are actively expressed during vegetative growth. It was demonstrated that the proteins Twi2 and Twi12 are localized in the cytoplasm, whereas Twi8 is localized in MAC [28, 31]. None of the Twi proteins was detected in transcriptionally silent MIC. Only Twi12 is required for the vegetative growth of cells. Several more Twi proteins are available in vegetative cells at low levels. It was revealed that every Twi protein is connected with a population of small RNAs in the vegetative phase of the life cycle of tetrahymena (Table 2).

Interestingly, siRNAs of 18–22 nt long that bind Twi12 protein are formed not via Dicer activity but are cut out from the 3'-end of mature tRNAs, which is indicated by the presence of the CCA sequence on their 3'-end outside of the template. The main identified function of Twi12 is activation of the exoribonuclease Xrn2 in MAC required for mRNA processing. The import to the nucleus from the cytoplasm requires Twi12 to be loaded with a molecule of small RNA; fragments of numerous tRNAs are optimal for this as their 3'-end are split out by RNAase T2 [32]. In general, the similar structure of 23–24 nt siRNAs of tetrahymena indicates that the specificity of their interaction with different Twi proteins depends not on modification of their 5'- and 3'-end but on the sequences of which they consist. Many small RNAs that bind with Twi2 protein are antisense relative to pseudogenes forming clusters in genome MAC. Knockout of the genes *TWI2* and *RDN2* and partial knockout of the gene *TWI8* result in the loss of these small RNAs [28]. Despite the fact that genome MAC of ciliates does not contain any noncoding sequences, a number of repeats still exist, to which sequences the homologous siRNAs were detected in vegetative cells. It turned out that the majority of siRNAs homologous to high-copy-number repeats are also connected with the Twi2 protein. However, during knockout of the *TWI8* gene, such siRNAs do not accumulate in the cells, and the knockout of *TWI2* and *TWI6* genes and any components of RDRC results in reduction of their quantity [28]. Homologous low-copy-number repeats of siRNAs are connected with the Twi7 protein, and siRNAs 22–24 nt long or 33–36 nt long homologous to telomeric repeats are associated with Twi10. Ultimately, siRNAs associated with Twi8 protein are, as a rule, homologous to both strands of some protein-coding genes; namely, these siRNAs switch off gene expression. Interestingly, tandem *TWI2–TWI6* is located within these genes. The quantity of mRNA for *TWI2* is increased in clones with knockout of the *TWI8* gene and is reduced in clones with the overexpression of this gene. Besides, seven loci were identified, which are not coding sequences, pseudogenes, or repeats, for which the availability of homologous siRNAs 23–24 nt long was detected [28]. In the absence of siRNAs homologous to these loci (e. g., as a result of knockout of the gene encoding factor Rdf2 required for their accumulation), RNA synthesis starts with these loci, which indirectly confirms the necessity of their silencing. The distribution of these siRNAs in their sense and antisense chains of the template is asymmetric, and in all cases, the majority of siRNAs are complementary to DNA chains forming in these loci a loop-shaped structure 50–100 nt long. Hairpins formed in DNA probably serve as primers for RDRP initiating siRNA synthesis. Such DNA structure is probably a conservative marker demonstrating that transcripts with this template shall be silenced, for example, as virus-like [28].

Proteins of the Piwi subfamily in *P. tetraurelia* are called Ptiwi (P*aramecium* t*etraurelia* Piwi**)**; in total, 17 genes encoding these proteins have been identified in the genome of this infusorium, 12 of which are paired ohnologs [7, 24]. The proteins of the Ptiwi family can be split into three subfamilies [7] (Fig. 4).

The first subfamily includes the proteins Ptiwi13/14 and Ptiwi02/04/05, the genes of which are uniformly expressed throughout the entire life cycle, as well as protein Ptiwi08. Proteins Ptiwi14 and Ptiwi08 are paralogs, but gene *PTIWI8* is activated only after meiosis, in the period of development of a new MAC [7] (see Fig. 4). The second subfamily is formed by a single protein, Ptiwi07, which has diverged the most from the rest of the Piwi proteins of paramecia. This protein is an ortholog of Twi8 tetrahymena; however, their functions are probably different. As the *TWI8* gene is expressed at a high level in the vegetative phase and the gene *PTIWI07* is expressed at later stages of MAC development [7]. In different periods of the sexual process, the majority of genes of the third subfamily are expressed (see Fig. 4). The *PTIWI03/01/09* genes are activated during miosis, the genes *PTIWI06/10/11* are expressed simultaneously with *PTIWI07* at the final

Table 2

Small RNAs and Piwi proteins in vegetative cells of *Tetrahymena* [28, 31, 32]

Characteristics of small RNAs	Homologous regions in MAC genome	Interacting Piwi proteins	Localization of Piwi proteins	Expession period of Piwi genes	Proteins necessary for small RNAs accumulation
23-24 nt., U at 3'-end	Clusters of pseudogenes	Twi2; less frequent – Twi8	Twi2 – in cytoplasm; Twi8 – in MAC	During vegetative growth and at conjugation	Twi2; Rdn2; Rdr1; partly – Twi8
23-24 nt., no U at 3'-end	Multicopied repeats	Twi2	In cytoplasm	During vegetative growth and at conjugation; non-essential	Twi8; Twi2, Twi6; Rdr1; Rdn1; Rdf1
23-24 nt., contain methylated ribose at 3'-end (as scanRNAs)	Some coding sequences	Twi8	In MAC at vegetative growth; in old MAC and then in new MAC during conjugation	Non-essential during vegetative growth; during conjugation in absence if Twi8p cells cannot complete formation of new MAC genome	Twi8; RDRC involving Rdf1
23-24 nt., U at 3'-end; 32-34 nt.	Low copy repeats	Twi7	unknown	During vegetative growth and at conjugation	Twi8; Twi2, Twi6; RDRC components
23-24 nt., U at 3'-end; 33-36 nt.	Telomeric repeats	Twi10	unknown	During vegetative growth; at conjugation expression increases	Twi8; Twi2, Twi6p; RDRC components
23-24 nt., asymmetric position at DNA template chains	Loci, where DNA forms loop-like structures 50-100 bp in size	Twi2 and Twi8	Twi2 – in cytoplasm; Twi8 – in MAC	During vegetative growth and at conjugation	Rdf2
18-22 nt., non-template CCA at 3'-end	Formed from 3'-ends of mature tRNAs	Twi12	In cytoplasm	During vegetative growth; essential	Xrn2 interacts with Twi12 in MAC

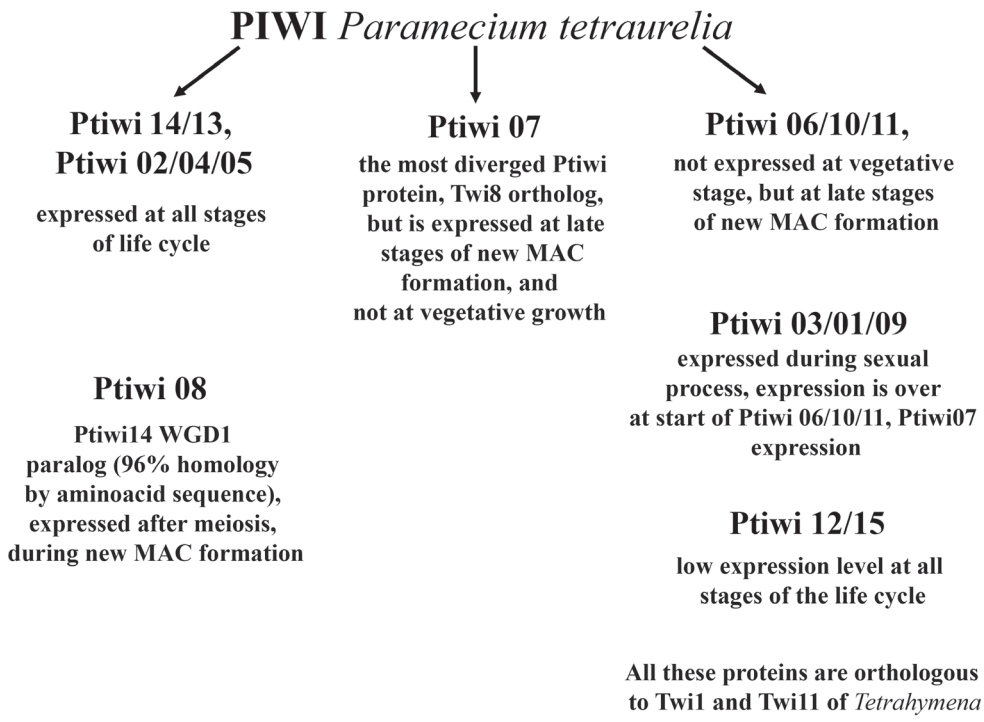

Fig. 4. Characteristic features of three subfamilies of *P. tetraurelia* Piwi proteins

stage of MAC genome formation, and only the genes *PTIWI12* and *PTIWI15* are constitutively expressed at a low level throughout the entire life cycle. Two IESs are included in the *PTIWI10* gene sequence, one of which is placed next to the start codon, and the second one is inside the coding sequence. This feature of the *PTIWI10* gene specifies its expression at the later stages of the sexual process; its expression becomes possible only after removal of the appropriate IES [33]. As IESs are included in the coding sequences and regulatory areas of the *PTIWI11*, *PTIWI06*, and *PTIWI07* genes, it is possible that regulation of their expression is performed in a similar way [33].

The exact protein functions were determined for only eight Ptiwi proteins. The proteins Ptiwi01, Ptiwi09, Ptiwi10, and Ptiwi11 are involved in genomic scanning in the sexual process [6, 7, 33]. Ptiwi13 participates in silencing caused by the micro-injection of transgenes, as well as in the silencing induced by feeding of *E. coli*. Ptiwi14 is involved only in silencing caused by the micro-injection of transgenes, and Ptiwi12 and Ptiwi15 are involved in silencing caused by feeding [7].

In the process of gene silencing of paramecia induced by feeding, the protein called Pds1 (*Paramecium* dsRNAinduced RNAispecific protein) is also involved [24], the gene of which is expressed throughout the entire life cycle [24]. No genes associated with the *PDS1* gene were found in the genome of *P. tetraurelia*, as well as no obvious homologs being detected in the genomes of other ciliates; however, this gene is present in all species of the *Paramecium* genus. Reduction of the level of this protein similar to Cid1 results in loss of the ability to perform RNA interference in response to the appearance of exogenous RNAs but does not affect vitality. The exact function of this gene remains unclear.

In general, the screening of *P. tetraurelia* mutants unable to start RNA interference in response to dsRNA produced by food bacteria demonstrated that this pathway is not required and includes genes *RDR1*, *CID1*, and *PDS1* that are not of vital importance, as well as vitally important genes *DCR1* [8], *RDR2* [14], and *CID2*. Besides, this pathway of RNA interference includes three Piwi proteins (Ptiwi12, Ptiwi15, and Ptiwi13) [17]. It was demonstrated that *P. tetraurelia* can synthesize small RNAs, antisense to rRNA and mRNA of the food bacteria, and this ability depends on the protein Rdr2 [17]. Transgene-induced RNA interference includes not only the proteins Dcr1, Rdr2, Cid2, and Ptiwi13 but also Rdr3 and Ptiwi14 [7, 14]. Interestingly, the vital genes are involved in both pathways, which can indicate that originally the mechanism of RNA interference of ciliates was used for the neutralization of foreign genes, which could penetrate into the core with viruses or transposons. The ability to synthesize siRNA for the neutralization of RNA arriving with food is not required, at least in laboratory conditions, as clones with mutations in the genes specific for this

pathway are supported in culture for years while being fed with the same "safe" bacteria. However, the genomes of all types of complex *P. aurelia*, as well as of *P. multimicronucleatum* and *P. caudatum*, have at least one copy of the genes *RDR1*, *CID1*, and *PDS1*, and the key factors of RNA interference starting in response to exogenous RNAs are highly conserved within the *Paramecium* genus [24]. This indicates that this pathway can play important roles in paramecia survival in natural conditions, where any microorganisms can penetrate into ciliates cells as food or together with food bacteria.

CONCLUSIONS

Ciliates are a goldmine for the examination of cell processes that are based on RNA interference, which plays a significant biological role in these protists. In the ciliates *Tetrahymena* and *Paramecium*, small RNAs are the basis of the formation of the complicated core mechanism during the sexual process [6], which results in the occurrence of physical and functional differences between the genomes of generative and somatic nuclei [5]. In vegetative cells, siRNAs regulate the expression of their own genes and do not permit the display of extraneous DNA sequences. No viruses have been detected in ciliates so far, which is astonishing, given the level of knowledge about many representatives of *Ciliophora*. This indirectly indicates the effectiveness of the antiviral protective systems in ciliates cells; RNA interference is the oldest mechanism of protection against viruses. It is tempting to suppose that RNA interference could be involved in the other rapid responses of ciliates to external stimuli. For example, it is obvious that siRNAs are involved in switching of the surface antigens of *P. tetraurelia* [34], which means that they can provide the regulation of gene expression during adaptation to environmental changes. It is known that noncoding RNAs of bacterial origin can affect expression of the genes of *Caenorhabditis elegans* [35]. The transcriptomes of *P. tetraurelia* containing symbiotic bacteria *Caedibacter taeniospiralis* in the cytoplasm differ from those of uninfected cells [36]. Moreover, data about the occurrence of small RNA in paramecia cells, which is antisense relative to the transcripts of food bacteria [17], suggest that RNA interference can regulate the fate of exogenous transcripts and the interaction of ciliates with symbionts.

The study was sponsored by grant RSF 16-14-10157.

REFERENCES

1. Cerutti H, Casas-Mollano JA. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet*. 2006;50(2):81-99. https://doi.org/10.1007/s00294-006-0078-x.
2. Fire A, Xu S, Montgomery MK, et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. 1998;391(6669):806-811. https://doi.org/10.1038/35888.
3. Epigenetics. Ed. by C.D. Allis, M.L. Caparros, T. Jenuwein, D. Reinberg. New York: Cold Spring Harbor Laboratory Press; 2015. 984 p.
4. Coyne RS, Chalker DL, Yao MC. Genome downsizing during ciliate development: nuclear division of labor through chromosome restructuring. *Annu Rev Genet*. 1996;30:557-578. https://doi.org/10.1146/annurev.genet.30.1.557.
5. Betermier M, Duharcourt S. Programmed rearrangement in Ciliates: *Paramecium*. *Microbiol Spectr*. 2014;2(6). https://doi.org/10.1128/microbiolspec.MDNA3-0035-2014.
6. Некрасова И.В., Потехин А.А. Интерференция РНК в формировании соматического генома у инфузорий *Paramecium* и *Tetrahymena* // Экологическая генетика. – 2018. – Т. 16. – № 4. – С. 5–22. [Nekrasova IV, Potekhin AA. RNA interference in formation of the somatic genome of ciliates *Paramecium* and *Tetrahymena*. *Ecological genetics*. 2018;16(4):5-22. (In Russ.)]. https://doi.org/10.17816/ecogen1645-22.
7. Bouhouche K, Gout JF, Kapusta A, et al. Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res*. 2011;39(10):4249-64. https://doi.org/10.1093/nar/gkq1283.
8. Lepère G., Nowacki M., Serrano V. et al. Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res*. 2009;37(3):903-915. https://doi.org/10.1093/nar/gkn1018
9. Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell*. 2002;110(6):689-699. https://doi.org/10.1016/s0092-8674(02)00909-1.
10. Lee SR, Collins K. Two classes of endogenous small RNAs in *Tetrahymena thermophila*. *Genes Dev*. 2006;20(1):28-33. https://doi.org/10.1101/gad.1377006.
11. Howard-Till RA, Yao MC. Induction of gene silencing by hairpin RNA expression in *Tetrahymena thermophila* reveals a second small RNA pathway. *Mol Cell Biol*. 2006;26(23):8731-8742. https://doi.org/10.1128/MCB.01430-06.
12. Galvani A, Sperling L. Transgene-mediated post-transcriptional gene silencing is inhibited by 3' noncoding sequences in *Paramecium*. *Nucleic Acids Res*. 2001;29(21):4387-4394. https://doi.org/10.1093/nar/29.21.4387.
13. Ruiz F, Vayssié L, Klotz C, et al. Homology-dependent gene silencing in *Paramecium*. *Mol Biol*

Cell. 1998;9(4):931-943. https://doi.org/10.1091/mbc.9.4.931.
14. Marker S, Le Mouël A, Meyer E, Simon M. Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in *Paramecium tetraurelia*. *Nucleic Acids Res*. 2010;38(12):4092-4107. https://doi.org/10.1093/nar/gkq131.
15. Galvani A, Sperling L. RNA interference by feeding in *Paramecium*. *Trends Genet*. 2002;18(1):11-12. https://doi.org/10.1016/s0168-9525(01)02548-3.
16. Garnier O, Serrano V, Duharcourt S, Meyer E. RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. *Mol Cell Biol*. 2004;24(17):7370-7379. https://doi.org/10.1128/MCB.24.17.7370-7379.2004.
17. Carradec Q, Gotz U, Arnaiz O, et al. Primary and secondary siRNA synthesis triggered by RNAs from food bacteria in the ciliate *Paramecium tetraurelia*. *Nucleic Acids Res*. 2015;43(3):1818-1833. https://doi.org/10.1093/nar/gku1331.
18. Gotz U, Marker S, Cheaib M, et al. Two sets of RNAi components are required for heterochromatin formation in trans triggered by truncated transgenes. *Nucleic Acids Res*. 2016;44(12):5908-5923. https://doi.org/10.1093/nar/gkw267.
19. Elbashir SM, Harborth J, Lendeckel W, et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*. 2001;411(6836):494-498. https://doi.org/10.1038/35078107.
20. Mochizuki K, Gorovsky MA. A Dicer-like protein in *Tetrahymena* has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes Dev*. 2005;19(1):77-89. https://doi.org/10.1101/gad.1265105.
21. Malone CD, Anderson AM, Motl JA, et al. Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of *Tetrahymena thermophila*. *Mol Cell Biol*. 2005;25(20):9151-9164. https://doi.org/10.1128/MCB.25.20.9151-9164.2005.
22. Aury JM, Jaillon O, Duret L, et al. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*. 2006;444(7116):171-178. https://doi.org/10.1038/nature05230.
23. Sandoval PY, Swart EC, Arambasic M, Nowacki M. Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting. *Dev Cell*. 2014;28(2):174-188. https://doi.org/10.1016/j.devcel.2013.12.010.
24. Marker S, Carradec Q, Tanty V, et al. A forward genetic screen reveals essential and non-essential RNAi factors in *Paramecium tetraurelia*. *Nucleic Acids Res*. 2014;42(11):7268-7280. https://doi.org/10.1093/nar/gku223.
25. Baulcombe DC. Molecular biology. Amplified silencing. *Science*. 2007;315(5809):199-200. https://doi.org/10.1126/science.1138030.
26. Lee SR, Collins K. Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. *Nat Struct Mol Biol*. 2007;14(7):604-10. https://doi.org/10.1038/nsmb1262.
27. Lee SR, Talsky KB, Collins K. A single RNA-dependent RNA polymerase assembles with mutually exclusive nucleotidyl transferase subunits to direct different pathways of small RNA biogenesis. *RNA*. 2009;15(7):1363-1374. https://doi.org/10.1261/rna.1630309.
28. Couvillion MT, Lee SR, Hogstad B, et al. Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of *Tetrahymena thermophila*. *Genes Dev*. 2009;23(17):2016-2032. https://doi.org/10.1101/gad.1821209.
29. Cerutti L, Mian N, Bateman A. Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends Biochem Sci*. 2000;25(10):481-482. https://doi.org/10.1016/s0968-0004(00)01641-8.
30. Matranga C, Tomari Y, Shin C, et al. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell*. 2005;123(4):607-620. https://doi.org/10.1016/j.cell.2005.08.044.
31. Farley BM, Collins K. Transgenerational function of *Tetrahymena* Piwi protein Twi8p at distinctive noncoding RNA loci. *RNA*. 2017;23(4):530-545. https://doi.org/10.1261/rna.060012.116.
32. Couvillion MT, Bounova G, Purdom E, et al. A *Tetrahymena* Piwi bound to mature tRNA 3' fragments activates the exonuclease Xrn2 for RNA processing in the nucleus. *Mol Cell*. 2012;48(4):509-520. https://doi.org/10.1016/j.molcel.2012.09.010.
33. Furrer DI, Swart EC, Kraft MF, et al. Two sets of Piwi proteins are involved in distinct sRNA pathways leading to elimination of germline-specific DNA. *Cell Rep*. 2017;20(2):505-20. https://doi.org/10.1016/j.celrep.2017.06.050.
34. Baranasic D, Oppermann T, Cheaib M, et al. Genomic characterization of variable surface antigens reveals a telomere position effect as a prerequisite for RNA interference-mediated silencing in *Paramecium tetraurelia*. *MBio*. 2014;5(6):e01328. https://doi.org/10.1128/mBio.01328-14.
35. Liu H, Wang X, Wang HD, et al. *Escherichia coli* noncoding RNAs can affect gene expression and physiology of *Caenorhabditis elegans*. *Nat Commun*. 2012;3:1073. https://doi.org/10.1038/ncomms2071.

36. Grosser K, Ramasamy P, Amirabad AD, et al. More than the "killer trait": infection with the bacterial endosymbiont *Caedibacter Taeniospiralis* causes transcriptomic modulation in *Paramecium* host. *Genome Biol Evol*. 2018;10(2):646-656. https://doi.org/10.1093/gbe/evy024.

❀ Information about the authors

Irina V. Nekrasova – PhD, Associate Professor, Faculty of Biology. Saint Petersburg State University, St. Petersburg, Russia. E-mail: ne-irina@yandex.ru.

Alexey A. Potekhin – PhD, Full-Professor, Faculty of Biology. Saint Petersburg State University, St. Petersburg, Russia. E-mail: alexey.potekhin@spbu.ru.

❀ Информация об авторах

Ирина Владимировна Некрасова — канд. биол. наук, доцент кафедры микробиологии биологического факультета. ФГБУ ВПО «Санкт-Петербургский государственный университет», Санкт-Петербург. E-mail: ne-irina@yandex.ru.

Алексей Анатольевич Потехин — канд. биол. наук, профессор кафедры микробиологии биологического факультета. ФГБУ ВПО «Санкт-Петербургский государственный университет», Санкт-Петербург. E-mail: alexey.potekhin@spbu.ru.