piRNA biogenesis in ciliates – Investigating the slicer activity of *Oxytricha* PIWI proteins

Final research report for Marshall Plan Scholarship

Florian Pühringer, Landweber Lab, Columbia University, New York

Time of research project:

1.8.2019 – 30.4.2020

Supervisor University of Vienna:

Isabella Moll, Ph.D
Professor, Molecular Bacteriology

Supervisor Columbia University:

Laura Landweber, Ph.D
Professor, Biological Sciences and Systems Biology

Summary

It is central to the viability of every life-form to ensure the stability and integrity of its collective DNA, the genome. If genome integrity is compromised by selfish genetic elements or parasitic DNA, cells lose control over key biological regulations with deleterious effects on their fitness. One of the most extraordinary and enigmatic ways to ensure genome integrity has been uncovered in the ciliate Oxytricha triffalax. In order to silence parasitic DNA elements, Oxytricha undergoes sophisticated genome rearrangements over the course of its life cycle. Research done by the Landweber lab could demonstrate that genome rearrangements in Oxytricha are regulated by piRNAs, a class of evolutionarily conserved small RNAs. Guided by Argonaute proteins of the PIWI clade, piRNAs, in an inverse fashion, mark parasitic DNA elements for elimination. Previous studies established a coherent model for DNA elimination in Oxytricha and found that this process is dependent on the PIWI protein Otiwi1. Based on these findings, my research project aimed to understand the molecular mechanisms underlying Otiwi1 function. Using Oxytricha cell culture and RT-qPCR experiments, I correlated mRNA expression levels of Otiwi1 with selected candidate piRNA-biogenesis factors. I immunoprecipitated Otiwi1 from Oxytricha cells and identified its binding partners during genome rearrangement. Furthermore, I planned to biochemically characterize Otiwi1 and assay its RNA or DNA binding affinity, as well as its predicted endonuclease activity. Therefore, I expressed and purified Otiwi1 wildtype and mutant proteins from Tetrahymena thermophila.

Table of Contents

INTRODUCTION

Transposable elements

All life-forms rely on storing their genetic information in DNA molecules. The full set of DNA molecules is called the genome and encodes RNA and proteins that guide and regulate cellular processes. If genome integrity is compromised by mutations or unwanted DNA rearrangements, cells lose control over key biological regulations and may either commit suicide or enter a path towards cancer development. It is therefore central to the viability of every life-form to ensure the stability and integrity of its DNA.

One major threat to genome stability and integrity are transposable elements (TEs). TEs are genetic parasites that are found in all three domains of life, making up close to 50% of the human genome sequence [1]. Their emergence dates back to primordial cells, when TEs were inevitable co-evolving partners in the evolution of all genomes. Also called selfish genetic elements, TEs aim to maximally propagate and replicate themselves with a host genome. They do this by either excising their DNA and inserting it at a new place within the genome (DNA transposons, cut and paste), or by copying themselves via an RNA intermediate and subsequently inserting the copy at a new place (retrotransposons, copy and paste). Both modes of action disrupt regulatory sequences or cause deleterious chromosomal rearrangements within the host genome [2]. Although TEs reside within host genomes, the host fitness does not select for their persistence. Therefore TEs must continuously propagate and amplify to avoid extinction [3]. Propagation relies on transmitting TE copies to the next generation, hence TE amplification must happen in the germline cells and indeed, most TEs are primarily active within the germline [4].

TE amplifications needs to defend against anti-amplification mechanisms by the host cells that need to defend against TE amplification. These mechanisms have created an ever on-going genetic conflict, where genome defense mechanisms function to counterbalance TE activity [5]. Genome defense mechanisms work by selectively silencing TEs in the germline to sustain the fitness of future generations and often rely on small RNA systems [6].

Small RNA systems and the piRNA pathway

Small RNA systems are generally defined by two components: a small RNA at short length (20-30 nucleotides) that associates with a member of the Argonaute protein family. The small RNA hereby guides the Argonaute protein to a regulatory target, resulting in a change of gene expression or the silencing of TEs [7]. Besides Argonaute-sRNA complexes as the central defining feature, small RNA pathways are quite diverse regarding the mode of small RNA biogenesis, the mechanism of gene regulation and the involved effector proteins. Argonaute proteins are conserved and abundant throughout the eukaryotic lineage and can be subdivided into the Ago and the Piwi subfamilies. The Ago subfamily proteins are usually guided by siRNAs and miRNAs which are generated from double-stranded precursors using and RNase III-type enzyme called Dicer. While miRNAs, like transcription factors, regulate endogenous cellular pathways in plants and animals, siRNAs make up a defense system against viruses (exo-siRNAs) as well as selfish genetic elements (endo-siRNAs) [7].

piRNAs in contrast, are bound by the Piwi subfamily of proteins and were found to be longer (~25-30 nucleotides) than known small RNAs at the time of their discovery. They are known to silence TEs in the germline of animals and unlike miRNAs and siRNAs, were defined to not rely on processing from double-stranded precursors by Dicer, but to arise from single-stranded RNA precursors in a Dicer-independent manner [8]. Furthermore, all known piRNAs are 2'-O methylated at their 3'ends, unlike miRNAs but similar to most siRNAs [8,9]. While the piRNA pathway has been mostly and intensively studied in the germline of *Drosophila melanogaster*, more and more piRNA research in different organisms is conducted nowadays. In *Drosophila,* the piRNA-pathway has mostly been defined as an adaptive immune system that silences TEs, specifies and regulates the germline. Studies from other organisms, however, highlight the flexible architecture of the pathway, that besides its core function, also allows to regulate gene function and expression [6].

Non-animal Piwi proteins

Because the piRNA-pathway has been characterized in animals, it is often still considered as an animal specific genome defense mechanism. However, non-animal Piwi orthologues can be found in many lineages involving slime molds and ciliates. These lineages are over 1 billion years old and were present long before eukaryotic diversity arose [10]. Ciliate Piwi proteins, like their animal counterparts, were shown to be functional and associate with small RNAs. In fact, the first Piwi protein shown to bind small RNAs was TWI1 from *Tetrahymena thermophila* [11]. Although these complexes exist and are known to be essential in the three studied ciliates *Tetrahymena thermophila, Paramecium tetraurelia* and *Oxytricha trifallax,* their function varies a bit from the animal piRNA-pathway function [12]. This variation is a result of the unique and complex genome architecture of ciliates as well as of their life cycle. The following section describes general ciliate biology and subsequently explains the role of the piRNA-pathway to regulate the ciliate life cycle.

Ciliates and nuclear dimorphism

Ciliates belong to Protozoa and are characterized by harboring hair-like organelles, the so-called cilia. They are an important group of protists and can be found in different watery environments like ponds, oceans, rivers or soil [13]. They are heterotrophs and feed on mostly bacteria and algae that are taken up by their oral-apparatus. Some ciliates are known to be animal parasites, however, only one species is known to cause human disease [14]. About 4500 unique species have been described and in most taxonomic classifications, ciliates are ranked as an own phylum under the protist kingdom [15].

What sets ciliates apart from almost all other eukaryotes is their nuclear dimorphism. This means that they harbor two nuclei within one unicellular organism. The macronucleus (MAC) is the larger one, which, in high copy numbers, encodes all genetic information required for everyday-life of a ciliate. The smaller nucleus is termed micronucleus (MIC) and its gene expression is kept silent throughout everyday-life [12]. Everyday-life refers to

vegetative growth where cells, when they are well-fed, divide asexually. Here, the MIC undergoes mitosis while the larger MAC, that doesn't have spindle fibers, divides amitotically with its chromosomes probably segregating randomly [16]. In contrast to this, ciliates can also grow sexually where they undergo non-replicative conjugation. Conjugation is thought to get activated by poor growing conditions in the wild, and can be initiated by starvation under lab conditions. Ciliates of compatible mating types form pairs which activates meiosis of the MIC. Subsequently a cascade of events leads to the formation of a zygotic MIC in each of the two cells, that further undergoes mitosis. At this point, both cells harbor two identical zygotic MICs as the old MAC starts to degrade. One of the two zygotic MICs is retained as MIC whereas the other MIC, in a process called genome rearrangement, differentiates into a new MAC [16,17]. This process is of remarkable complexity and accuracy. For a short period of time the newly formed MAC and the degrading MAC co-exist what allows the direct, cytoplasmic transmission of epigenetic factors from parent to offspring [18].

Although ciliates are unicellular, the MIC mirrors the germline of animals while the MAC represents the soma. Thus, ciliates show a simplified and unique mode of germline-soma differentiation of higher eukaryotes. This is especially interesting in the light of ciliate genome rearrangement, where the MIC (germline) undergoes meiosis, fertilization and is immortal. The MAC (soma) is replaced and rebuilt in every generation [12]. The ciliate germline is comparable to the germlines of higher eukaryotes in that it always remains totipotent and can develop into somatic nuclei. At the same time, it must cope with risks of totipotency, like TE invasion and uncontrolled replication [12]. Indeed, ciliate germlines contain large proportions of repetitive sequences and TEs that need to be strictly controlled. Therefore, as mentioned earlier, the MIC is completely silent during vegetative growth while the MAC, which is transcriptionally active is devoid of TEs. This mode of TE control is maintained during the genome rearrangement process, where all TE derived sequences are precisely excised from the MIC that develops into the new MAC [19]. By that, the newly formed MAC in each generation resembles a "clean" and streamlined genome, consisting of only genes and regulatory elements. Taken together, ciliates and

their nuclear dimorphism represents a unique and probably ancient form of genome architecture that ensures efficient gene expression and TE control.

Genome rearrangement in ciliates and the piRNA pathway

The formation of a new MAC from a zygotic MIC is a complicated process that involves DNA excision and repair. Besides TEs, ciliate genomes contain large numbers of non-genic sequences termed internal eliminated sequences (IES). These are remnants of TEs and interrupt coding genes, therefore need to be eliminated. In the three well-studied ciliates, a different amount of IESs is eliminated: 34% in *Tetrahymena*, 25% in *Paramecium*, and 95% in *Oxytricha* [20,21]. The whole rearrangement process happens over a period of 12h in *Tetrahymena* to 72h in *Oxytricha* and in addition to DNA elimination involves the re-joining, formation and amplification of telomer-capped MAC chromosomes [12].

Genetic experiments in *Paramecium tetraurelia* established that mutations in the parental MAC are transmitted to the new MAC via a cytoplasmic factor during genome rearrangement [22]. It was speculated that this factor guides the removal of IESs based on their absence and presence in the parental MAC. Later it was shown that this cytoplasmic factor is noncoding RNA. Injection of an RNA molecule with a specific IES sequence into the parental MAC of a cell, could inhibit the elimination of this specific IES [23]. By now, a model in all three well-studied ciliates emerged, where noncoding RNA determines which MIC sequences are retained (termed MDS = macronuclear destined sequence) and which MIC sequences are eliminated (IES) in order to build up a new MAC genome [16]. Thus, noncoding RNAs act as an essential information transfer system that defines the genomic content of the next generation.

Amongst the most important class to guide and regulate ciliate genome rearrangement are piRNAs. Similar to their anciently attributed function of TE defense in animals, they are known to protect ciliates against TE expansion [24]. Precisely, the Piwi-bound small RNAs direct the elimination of IESs during the development of the new MAC in ciliates. While they therefore completely remove TE-derived sequences or TEs, their counterparts

in animal germlines, initiate silencing of such sequences – an analogous mechanism that connects the evolutionary origin of the pathway to TE silencing. Given this unifying theme, it's interesting to note, that piRNA target selection differs in ciliates. In *Tetrahymena* and *Paramecium* piRNAs target and mark IESs for degradation, while in *Oxytricha* piRNAs do the opposite, and target MDSs for retention [12]. Ultimately, this difference doesn't affect the final outcome of IES elimination and MDS retention during rearrangement, it rather demonstrates the adaptivity and flexibility of the pathway. The following section will focus on the piRNA pathway in *Oxytricha*, subject of the conducted research, and outline the core questions of investigation.

Genome rearrangement in Oxytricha

Oxytricha trifallax, one of the three well-studied ciliate organisms described above harbors a MAC genome made of over 16.000 telomer-capped nanochromosomes that are 3.2kb on average in length. The MAC genome represents only 5% of the complete genomic content, the MIC genome, thus, during sexual reproduction and genome rearrangement, 95% of the zygotic MIC are eliminated [16]. This massive DNA elimination step relies on degradation of IES elements followed by descrambling and polytenization, two processes that are necessary to form mature nanochromosomes – they are, however, not topic of this research project. In contrast, I want to discuss how *Oxytricha* MDSs are marked for retention and thereby protected from degradation.

Research in *Tetrahymena* and *Paramecium* established the so-called scan RNA model that guides IES degradation in these ciliates [23,25]. Here, MIC transcripts are processed into small RNAs that are in turn loaded into Piwi proteins that in turn scan the parental MAC for sequence homology. Scan RNA (scnRNA) complexes that find a sequence match are recycled. The remaining complexes, consisting of IESs matching (= non MAC sequences) scnRNAs, are transported to the developing MAC where they mark IESs for elimination. A signal cascade starts that in an RNAi-like process leads to the degradation of IESs. The remaining sequences (MDSs) are joined and telomers are added to form new and functional MAC chromosomes [26].

Like the other ciliates, *Oxytricha* produces a huge amount of conjugation-specific small RNAs, however, they don't follow the scan RNA model, and are therefore simply termed piRNAs [27].

These are completely absent in vegetative cells, are 27 nucleotides long and show a strong 5'U bias, similar to their animal counterparts [27]. They are bound to the protein Otiwi1, an *Oxytricha* orthologue of Piwi and in contrast to scnRNAs they originate from the MAC and are therefore specific for MDSs and not IESs. Otiwi1 localizes to the parental MAC nucleus at the beginning of conjugation and re-localizes to the developing MAC at later stages of conjugation. Otiwi1 is essential for conjugation, as a knockdown of it results in nonviable offspring [27]. These data build up a model, where *Oxytricha* piRNAs are transcribed from the parental MAC, and after processing, transported to the developing MAC where they, by an unknown mechanism, target MDSs for retention (Figure1, image taken from Landweber lab paper [27]).

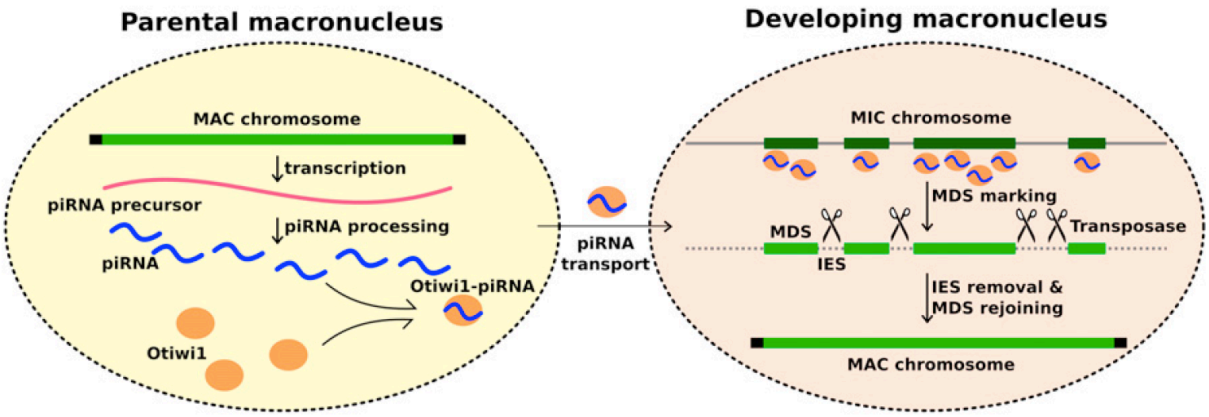


Figure 1. Model for piRNA-guided genome rearrangement in Oxytricha (image courtesty [27])

The above-mentioned model inspires many important questions about the molecular mechanisms underlying genome rearrangement in *Oxytricha*. In the following part I will thematically as well as experimentally focus on two mechanistic question that are both directly connected to the protein function of Otiwi1:

- **How are piRNA precursors processed in *Oxytricha*?**
    - **Is Otiwi1 catalytically active and is its activity required to produce mature piRNAs?**
- **How does an Otiwi1-piRNA complex target MDSs for their retention?**

How are piRNA precursors processed in Oxytricha?

piRNA precursor processing is well studied in animals as well as in *Tetrahymena* and *Paramecium*. A defining feature for animal piRNAs is that their synthesis and processing, in contrast to miRNAs and siRNAs, is independent of Dicer [6]. This definition doesn't hold true for *Tetrahymena* and *Paramecium* scnRNAs. These are exclusively bound by Argonaute proteins of the Piwi clade and are therefore categorized as piRNAs, however, their processing relies on Dicer proteins. In brief, MIC transcribed single-stranded precursors are processed by an RNA-dependent RNA polymerase (RdRP) and the generated double strand is then cleaved by Dicer. The created small RNA-duplex is in turn loaded into the respective Piwi protein that cleaves and thereby releases one of the two strands. The processing to form a mature scnRNA-Piwi complex is completed by the methyltransferase Hen1 that installs a 2'-O-methylation on the scnRNA 3' end [26]. Although scnRNAs are piRNAs as they are bound to Piwi proteins, their processing mechanism rather reflects that of siRNAs in animals. The coupling of RdRP and Dicer to produce smallRNA Argonaute complexes is thought to be anciently conserved to the last eukaryotic common ancestor. Moreover, protein domains of Dicer and RdRP are conserved in prokaryotes as well [28].

Given the strong conservation and predominance of this biogenesis mechanism in other ciliates, it is tempting to speculate that *Oxytricha* piRNA biogenesis occurs in a similar fashion. Nevertheless, experimental evidence is lacking, and *Oxytricha* piRNAs could be produced similar to animal piRNAs. Here, a single-stranded transcript is cleaved to create a 5'phosphate and in a process that directly relies on the endonuclease activity of the Piwi protein and the assistance of other cofactors, mature piRNAs are generated in a

head-to-tail fashion. This mechanism is a combination of the “ping-pong cycle” and “phasing” and was recently shown to be conserved in all animals [29].

While both discussed biogenesis modes are mechanistically diverging and rely on a different set of protein factors, they both depend on the slicing activity of a central Piwi protein. Slicing, in general, is the endonucleolytic cleavage of a single stranded RNA substrate by an Argonaute protein. The reaction is directed by a small RNA, which guides the Argonaute to its target by complementary base-pairing. Upon target recognition, the Argonaute cleaves the RNA substrate via its PIWI domain [30]. To characterize the potential slicing activity of Otiwi1, the central Piwi protein of *Oxytricha*, I plan to biochemically assay slicing in an in vitro setup. In addition, I aim to characterize Otiwi slicing in vivo. To further investigate the mode of piRNA biogenesis in *Oxytricha*, I want to identify potentially active Dicer proteins in the *Oxytricha* genome and measure and compare their mRNA expression levels at different stages of development to Otiwi1 mRNA expression.

How does an Otiwi1-piRNA complex target MDSs for their retention?

While research by the Landweber lab established that Otiwi1 targets MDSs for retention, it remains to be seen how targeting works at the mechanistic level [27]. Piwi proteins of *Tetrahymena* and *Paramecium* mark IESs for degradation by depositing the heterochromatic H3K9me3 mark on these sequences. Orthologues of the heterochromatin protein 1 (HP1) in turn bind these marks and are reported to form degradation bodies to turn over IESs [12]. *Oxytricha’s* orthogonally different piRNA pathway presents challenges to this model as piRNAs mark sequences for retention rather than deletion. Furthermore both MDSs and IESs in *Oxytricha* are shorter than one nucleosome unit – the length that might be required for H3K9 modifications [16]. An alternative mode of marking MDSs could be the physical binding of Otiwi1-piRNA complexes. A similar mechanism is reported for the *Stylonichia lemnae*, a ciliate species that is closely related to *Oxytricha*. Here, MDS DNA is reported to be directly targeted by

a Piwi-piRNA complex [31]. These complexes then constitute a “physical roadblock” and thereby prevent the degradation of IESs. Examples of RNA-guided Argonaute proteins that target DNA rather than RNA can be found in bacteria as well [32]. To investigate the possibility that an Otiwi1-piRNA complex directly targets and physically protects IES DNA from degradation, I aim to investigate the DNA binding activity of Otiwi1-piRNA complexes in vitro.

RESULTS

3D modeling of Otiwi1 predicts a conserved Piwi architecture and catalytic activity

Research by the Landweber lab identified 13 Piwi orthologues in *Oxytricha* that were named Otiwi1-Otiwi13. Phylogenetic analyses of these proteins showed that some cluster closely to the Ago subfamily of Argonautes and others to the Piwi subfamily. This large quantity of Piwi proteins reflects the RNA-rich regulation of ciliates and is similar to the Piwi landscape in *Tetrahymena* and *Paramecium*. Otiwi1 was shown to be essential for genome rearrangement and was furthermore implicated in piRNA biogenesis as well as MDS targeting [27]. I focused my research on Otiwi1 in order to characterize its mechanistic function regarding these two pathways. In addition to Otiwi1, Otiwi2 was shown to associate with piRNAs as well (unpublished data, Landweber lab), therefore I also conceived analysis to investigate Otiwi2 function in an exploratory manner.

It’s well established that slicing of Piwi proteins, a key step in piRNA biogenesis, relies on the catalytic activity of four amino acid residues in the PIWI domain. These residues make up a DEDX motif, where X is either a histidine or arginine [28,33,34]. To predict the slicing activity of Otiwi1 and Otiwi2, I generated multiple sequence alignments (MSA) of their PIWI domains and PIWI domains of Piwi proteins with known slicing activity. Both Otiwi1 and Otiwi2, harbor a DEDH motif (Figure S2). To analyze the DEDH motif in a structural context I then modelled the 3D structure of Otiwi1 and Otiwi2 using the MODELLER software with standard parameters and the Siwi structure (PDB 5duh) as a template [35].

In contrast to de novo modeling software, MODELLER relies on at least one template structure of a similar protein to accurately evaluate the predicted 3D fold of the model. The generated 3D model predicted a conserved architecture of Otiwi1 and Otiwi2 (data not shown) compared to Siwi with a more flexible but shorter N-terminal region and slight differences in the PAZ domain (Figure 2a, 2c). Investigation of the DEDX motif predicted a functional catalytic pocket in Otiwi1/2 that mirrors that of Siwi (Figure 2d). Siwi was biochemically shown to harbor slicing activity [34]. In summary, Otiwi1 and Otiwi2 were strongly predicted to be catalytically active. These predictions were the basis to biochemically investigate slicing.

Expression and purification of Otiwi1 from Tetrahymena cell lines

To biochemically measure slicing, I aimed to adapt an assay published by Gunawardane et al. [36]. Here, a purified protein is loaded with a guide RNA followed by incubation with a compatible and radiolabeled target RNA. Cleavage of the target RNA is followed via autoradiography on a UREA-PAGE gel.

Trials by the Landweber lab to express *Oxytricha* proteins using conventional expression systems like *E. coli* or *Hi5* cells were not successful. Due to the time-consuming work of cultivating *Oxytricha* cells, purification of endogenous Otiwi1/Otiwi2 is not feasible. Furthermore, both proteins are specifically expressed during sexual conjugation so that cells would need to be starved and mated in order to purify them. To overcome those technical barriers, I chose to express and purify Otiwi1/Otiwi2 from *Tetrahymena* cells that can be cultivated in larger scales in a feasible manner. I designed expression constructs to N or C terminally tag Otiwi1/Otiwi2 with FLAG-TEV-ZZ, a tag combination that is widely used in *Tetrahymena* genetics [37].

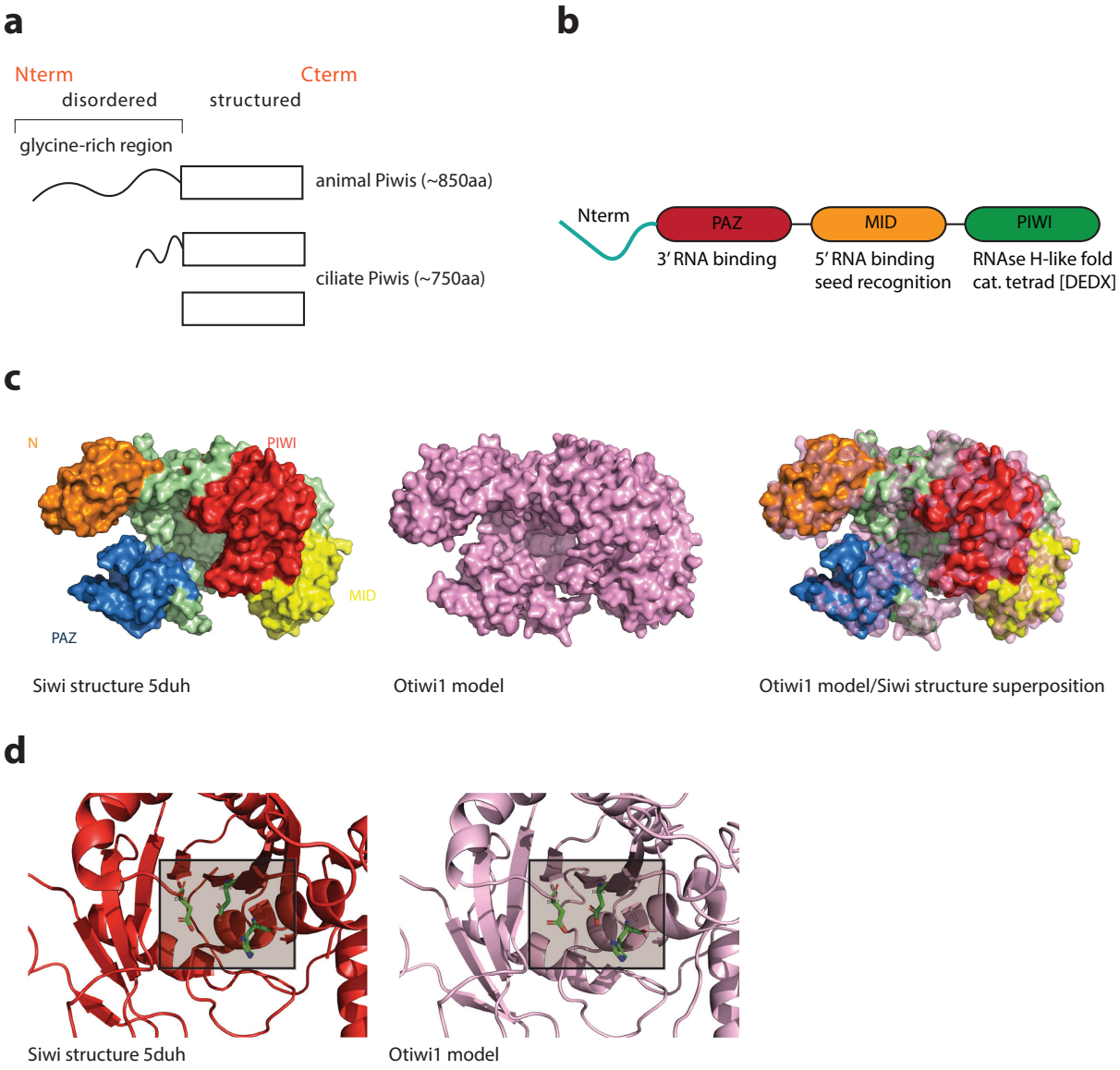


Figure 2. 3D Structural Modeling of Otiwi1 based on Siwi (PDB 5duh)
a, Depiction of the general architecture of Piwi proteins; extracted from multiple sequence alignments of ~30 different Piwi proteins from animals and ciliates
b, Conserved domain architecture of Argonaute proteins; PAZ domain harbors a pocket that binds the 3' end of the guide RNA, MID domain binds 5' end of guide RNA which recognizes the seed region of the target RNA, PIWI domain harbors the catalytic activity needed for slicing excetuted by a catalytic tetrad composed of DEDX motif (X=histidine or arginine)
c, 3D modeling of Otiwi1 using MODELLER software based on the published Siwi structure; third image shows a 3D superposition of the Siwi structure and the Otiwi1 model
d, Catalytic tetrad comparisson of Otiwi1 model with Siwi structure; E residue could not be crystallized therefore is also not present in the model

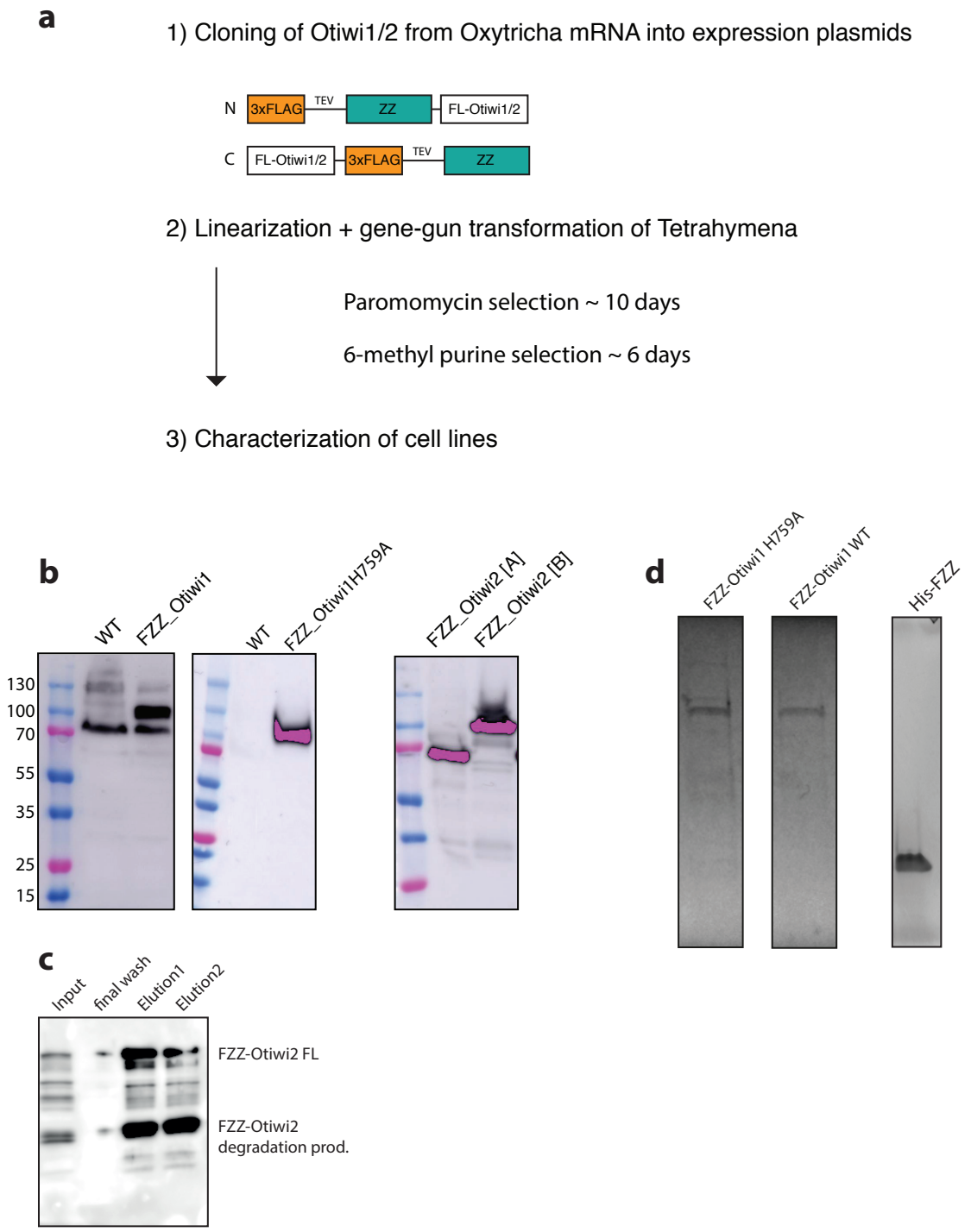


Figure 3. Expression and purification of Oxytricha proteins from Tetrahymena and FZZ protein from E.Coli
a, Depiction of the work-flow that was used to generate Tetrahymena cell lines expressing tagged (N and C term) Otiwi1 and Otiwi2 proteins
b, Characterization of the generated cell lines by western blot; For Otiwi1 WT and H759A and Otiwi2 WT, cell lines that stably express the proteins were created
c, Western blot showing Otiwi2 purification trials; FL protein and degradation/cleavage product are indicated
d, SDS-PAGE gel showing purified proteins

Using the gene-gun transformation method, I stably integrated the constructs into the *APRT1* locus of *Tetrahymena*. ∆*APRT1* cells were shown to be resistant to 6-Methylpurine which therefore can be used for genetic assortment and selection of the cell lines [38]. In addition, positive clones can be selected with Paromomycin because of a resistance cassette in the expression construct. Using gene-gun transformation and a double selection process with 6-Methylpurine and Paromomycin (Figure 3a), I succeeded in generating *Tetrahymena* cell lines that stably express tagged versions of Otiwi1/2. Western blot analyses of whole protein cell extracts showed that C-terminally tagged proteins completely failed to be expressed. Similar behavior was previously seen for the *Tetrahymena* Piwi protein Twi1 (personal communication), and might be because of interference of the tag with the C-terminal Piwi domain. Nevertheless, Otiwi1 as well as Otiwi2 were expressed well when the tag was added as an N-terminal fusion (Figure 3b). MSA indicate that the N term of Otiwi1/2 is of disordered nature what might serve as a flexible linker for adding a tag.

All expression constructs harbor a copper-inducible promotor to switch on protein expression in order to collect cells and purify the proteins. Several trials of purifying the N-terminally tagged Otiwi2 protein were unsuccessful due to protein degradation (Figure 3c). Different buffer adaptions and expression conditions couldn't circumvent this issue. I therefore generated a codon-optimized expression construct for *E. coli* and initiated expression and purification trials by using an N-terminal His-FZZ tag. In general, FZZ is known to enhance protein solubility, however, Otiwi2 could not be expressed in a soluble manner. Given the technical difficulties to purify Otiwi2, I then focused my efforts on Otiwi1. I used the previously generated knowledge and created an additional cell line that expresses a putative catalytically dead version of Otiwi1. Hereby, I mutated histidine 759 that is part of the catalytic tetrad, to alanine (H759A). Although not in large quantities, FZZ-Otiwi1 WT and FZZ-Otiwi1 H759A could be purified from *Tetrahymena* by FLAG affinity purification (Figure 3d). I planned to use those proteins in slicing assays as well as in electrophoretic mobility shift assays (EMSA). Therefore, I additionally purified a His-FZZ protein from *E. coli*, intended to be used as negative control in the biochemical assays.

Generation of transgenic *Oxytricha* cell lines to express Otiwi1-variant proteins

Besides characterizing the slicing activity of Otiwi1 in vitro, I aimed to study piRNA biogenesis in vivo. Precisely, I planned to assay the mating-phenotype of Otiwi1 H759A compared to Otiwi1 WT and also measure the impact this mutation has on the global small RNA production of mated *Oxytricha* cells. Moreover, in vivo experiments with Otiwi1 H759A could be helpful to position Otiwi1 slicing within the piRNA pathway. As discussed in the introduction, slicing is essential for piRNA biogenesis in animals and other ciliates. We thus hypothesized that Otiwi1 slicing plays a critical role in *Oxytricha* piRNA biogenesis. To investigate, given that the hypothesis is true, if slicing is used in a Dicer dependent pathway to remove the passenger strand of the RNA duplex, or rather in a direct way so slice an RNA precursor, I aimed to adapt an experiment by Noto et. al [25]. In this experiment, Otiwi WT and Otiwi H759A are immunoprecipitated to analyze the properties of their bound RNA by native gel electrophoresis. This approach allows to determine if the bound RNA is double or single-stranded, whereas double-stranded RNA in the H759A sample and single-stranded RNA in the WT sample would indicate a Dicer dependent processing mechanism. Single-stranded RNA in both samples would argue for a direct mechanism related to “phasing”, as discussed in the introduction. Taken together, three different in vivo experiments including phenotypic characterization, smallRNA sequencing and immunoprecipitation should be carried out. All experiments would ideally be done across three genotypes: Otiwi1 KO, Otiwi1 WT and Otiwi1 H759A. Genome editing by CRISPR is not yet established in *Oxytricha.* However, previous work in the Landweber lab established an Otiwi1 KO cell line based on a whole chromosome deletion of its corresponding nanochromosome (Fig 4a). I therefore decided to use this cell line as the basis to express Otiwi1 WT and Otiwi1 H759A. I cloned constructs that mirror the endogenous Otiwi1 nanochromosome but include tags for immunoprecipitation and are devoid of introns (Figure 3c). Prior to injection of these constructs, I genotyped the Otiwi1 KO cell line to control for the presence of the Otiwi1 chromosome. As mentioned in the underlying publication of this cell line [39], I could still detect N-terminal fragments or shortened Otiwi1 nanochromosomes by single cell PCR. However, no full-

length chromosomes could be detected (Fig 4b). Using a developed protocol from the Landweber lab, a colleague injected the constructs to the MAC of single *Oxytricha* cells and clonally amplified them. The experiment is currently on-going.

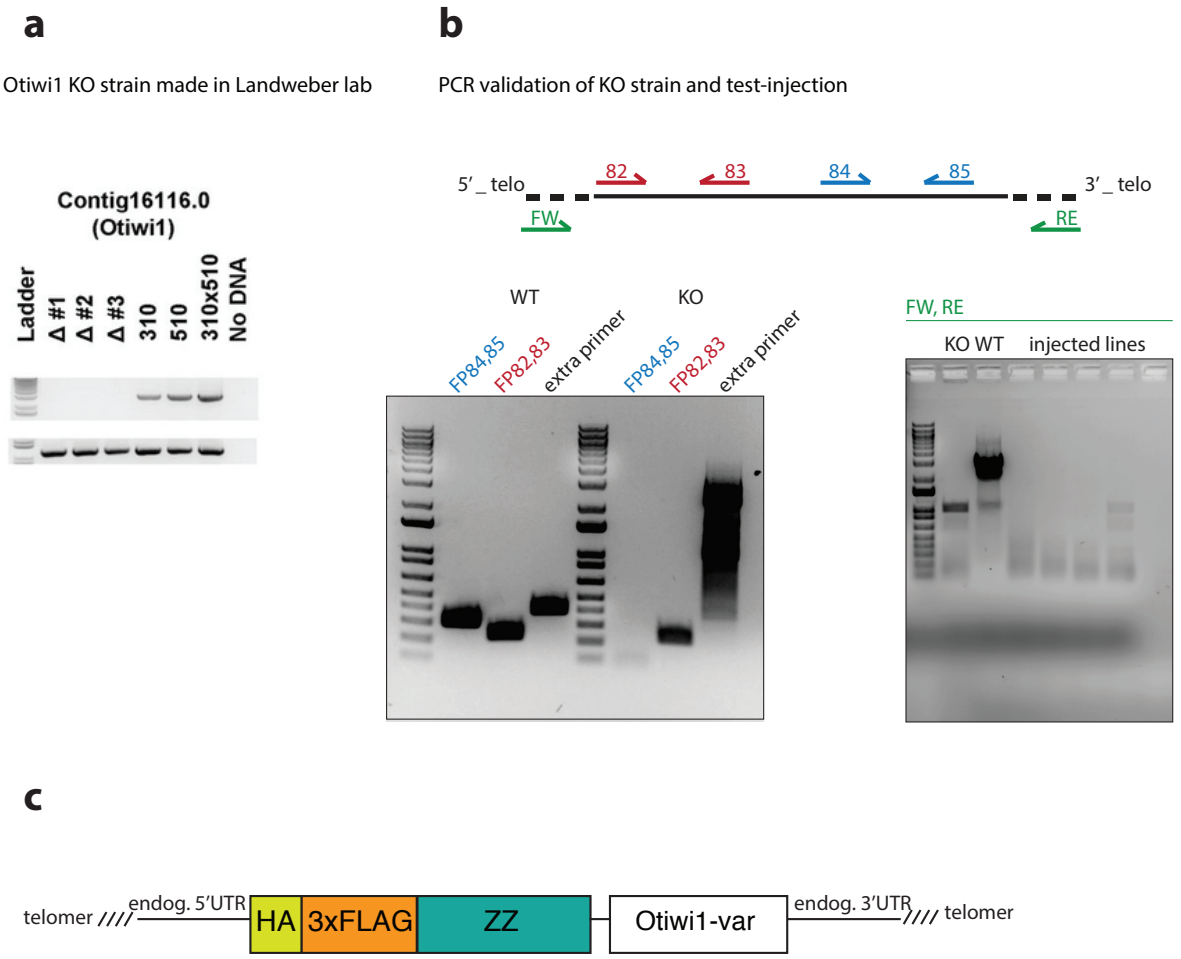


Figure 4. Generation of transgenic Oxytricha cell lines to endogenously express tagged Otiwi1 variants
a, Whole chromosome deletion KO line for Otiwi1 that was previously generated in the Landweber lab
b, Genotyping of the Otiwi1 KO cell line; primer binding sites are indicated in schematic illustration
c, Depiction of expression construct that was used to inject Otiwi1 KO cell line; the construct harbors the endogenous nanochromosome organisation of Otiwi1 but includes tags at the Nterm; Otiwi1-var that are expressed this way include the WT version and different mutants including DEDX mutants

DCR1 and DCR2 are specifically expressed during *Oxytricha* development

In *Tetrahymena* and *Paramecium,* Dicer proteins are essential for the production of scnRNAs during genome rearrangement. To investigate the potential involvement of Dicer proteins in *Oxytricha* genome rearrangement and piRNA production, I took an exploratory approach. I used BLAST to mine the *Oxytricha* proteome for orthologues of human Dicer and identified 11 confident hits requiring an evalue < 1x10^-4 (Fig 5a). It's known from the literature that functional Dicers are RNAse III enzymes that recognize and cleave double stranded RNA [30]. This allowed me to further confine the hit list to DCR1, DCR2 and R3D1a that harbor RNAse III as well as dsRNA-binding domains (Figure 5a). While attempts to immunoprecipitate *Oxytricha* Dicer by using commercial antibodies against human Dicer1 failed (data not shown), I aimed to measure the mRNA expression of the three selected Dicers. Precisely, I sought to compare their expression between vegetative growth and sexual development where piRNAs are produced. Moreover, I planned to compare Dicer expression levels to those of Otiwi1 to probe a potential correlation. To do so, I initially collected total RNA from vegetative cells as well as from mated cells throughout development (0-60hrs) and validated qPCR primers for DCR1, DCR2, R3D1a and Otiwi1 via PCR and agarose gel analysis. I then performed RT-qPCR using the expression of 23S rRNA as an internal normalization control because of its stable expression levels throughout development [40]. All calculated expression levels were additionally normalized against their respective expression during vegetative growth. By that, I only display enriched mRNA expression during sexual development. Interestingly, DCR1 and DCR2 expression is upregulated during development while R3D1a expresses at basal/vegetative growth levels (Figure 5b). Furthermore, the qPCR data shows that DCR1 and DCR2 expression can be detected at 12h post mating and increases at 24h post mating, although only slightly. This pattern can as well be seen for Otiwi1 expression indicating weak positive correlation. Expression levels of Otiwi1 at 12h and 24h are close to 2fold higher than DCR1 and DCR2 expression levels (Figure 5b). Similar results were previously shown for *Paramecium* Piwi proteins and Dicers during development [41]. An explanation for this difference is probably the sheer quantity of Otiwi1

molecules required to associate with the whole MAC genome split up in 27 nucleotide piRNAs. In general, the RT-qPCR results are in line with RNA-sequencing results from the Landweber lab. Further evaluation of these results by measuring protein expression via western blot was not yet possible, due to the lack of specific antibodies for *Oxytricha* DCR1, DCR2 and R3D1a.

a

geneID	chr	gene name (Oxydb)	previous name	length	pred MW (kD)	Rnase III, dsRNAb	additional domains
g6905	OXYTRI_MAC_15171	DCR2	Dcl1	3545	130,0	y	
g2569	OXYTRI_MAC_119	DCR1	Dcl2	4492	164,7	y	CBS domain
g16184	OXYTRI_MAC_22569			1614	59,2	n	DEAD helicase, AAA, helicase C
g10029	OXYTRI_MAC_17740	R3D1a	Dicer	8391	307,7	y	helicase, ATPase
g24401	OXYTRI_MAC_7520			2925	107,3	n	helicase, ATPase
g14435	OXYTRI_MAC_2122			1668	61,2	n	helicase, ATPase
g789	OXYTRI_MAC_10586			1475	54,1	n	helicase, ATPase, DNA binding
g23438	OXYTRI_MAC_6599			1424	52,2	n	helicase, ATPase, DNA binding
g26802	OXYTRI_MAC_9705			2504	91,8	n	glycine rich, RRN, Mtase, helC, DNA binding
g14555	OXYTRI_MAC_21312			2387	87,5	n	helicase, ATPase, DNA binding
g3176	OXYTRI_MAC_12363			1549	56,8	n	helicase, ATPase, DNA binding

b

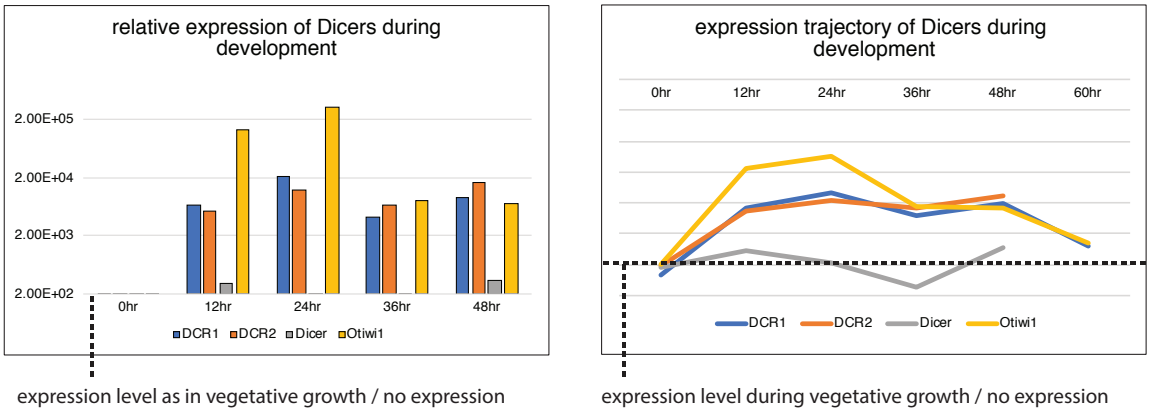


Figure 5. Identification of Dicer proteins in the Oxytricha genome and validation of mRNA expression over a developmental time course

a, Potential Dicer proteins were identified by BlastP; Proteins were further selected based on their domain architecture assuming that active Dicer harbors at least an RNAse III domain as well as a dsRNAb domain

b, Expression profiling of selected Dicer proteins via RT-qPCR over a developmental time course; primers were selected to be specific for each Dicer and experiments were carried out in duplicates; mRNA expression was normalized to rRNA expression as well as to mRNA expression of each assayed protein during Oxytricha vegetative growth

Otiwi1 interacts with a phosphatase complex harboring ANTAR domains during *Oxytricha* genome rearrangement

Research by the Landweber lab established that Otiwi1 binds 27 nucleotide long piRNAs that almost exclusively match MDS sequences while a KO of Otiwi1 leads to arrest of genome arrangement resulting in cell death [27]. These results implicate Otiwi1 function in the targeting and selection of MDSs to be passed on to the developing MAC. How this selection (see Figure 1) works mechanistically remains unclear. To investigate potential protein complexes and interaction partners that aid Otiwi1 in this process, I sought to perform Otiwi1 immunoprecipitation experiments and further use mass spectrometry (MS) to identify its interaction partners during genome rearrangement. Western blot analysis and RT-qPCR shows that Otiwi1 mainly is expressed during early rearrangement at 12h and 24h post mating of *Oxytricha* cells (Figure 5 and [27]), whereas its expression strongly decreases at later time points. Because of that and the assumption that Otiwi1 is essential for 1) piRNA biogenesis (12-24h) as well as for 2) MDS targeting and selection (24-36h), Otiwi1 probably interacts with different proteins at these different time points. I focused on the 2) MDS selection function of Otiwi1 and therefore harvested cells 24h post mating for IP. Using different buffer and sonication conditions, western blot analysis to estimate the general IP efficiency showed that Otiwi1 mostly remained in the unbound (UB) fraction. To control for efficient antibody binding, I repeated the IP with purified Otiwi1 (from Figure 3) as input and noticed poor antibody binding to Otiwi1 (Figure 6a). The reason for this low binding affinity could be that the published binding site of the commercial antibody is partially buried within the Piwi domain of Otiwi1 (inferred from model in Figure 2). Because of the above-mentioned difficulties, the IP-MS results could potentially only reflect a subpopulation of Otiwi1 that, due to biophysical properties, is accessible to antibody binding. MS analysis of the IP fraction revealed Otiwi1 as the highest enriched identified protein, validating the IP approach. Highly enriched Otiwi1 interaction partners involve components of a phosphatase complex with ANTAR domains, an RNA helicase as well as the DNA helicase MCM5. ANTAR domains and MCM5 are implicated in DNA replication that is thought to play a critical role for genome rearrangement in *Stylonichia*. Here, Piwi is thought to physically mark MDS sequences

and thereby block DNA replication. Replicated sequences are then thought to be recognized and eliminated by an unknown degradation machinery [31]. While it's tempting to speculate that a similar mechanism occurs in *Oxytricha*, further investigation and more-efficient IP experiments need to be done. Only by collecting more experimental data, a model for MDS selection can be built up.

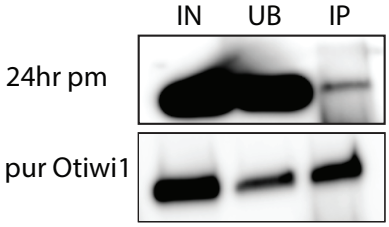
Biochemical characterization of Otiwi1

To answer parts of the two core research questions formulated in the introduction I sought to characterize the purified proteins from Figure 3. In brief I planned EMSA assays to determine the DNA vs. RNA binding preferences of Otiwi1. I sought to use the catalytic mutant Otiwi1 H759A and incubate the protein with ssRNA vs ssDNA of the exact same sequence, a 27nt piRNA mimic adapted from [25]. Furthermore I sought to incubate the protein with RNA-DNA hybrids, where only one strand is phosphorylated and thereby bound as the "guide" strand by Otiwi1 as previously established for Argonaute proteins [42]. These binding assays investigate the mode of Otiwi1 MDS targeting, specifically the possibility that Otiwi1 is an RNA-guided but DNA targeting Piwi protein.

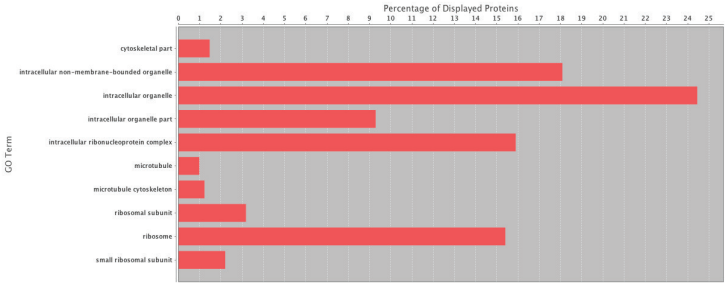
The second biochemical avenue is made up by slicing assays following protocols by [36]. These assays characterize the catalytic activity of Otiwi1 and are an important part of establishing Otiwi1's function in piRNA biogenesis.

Both these biochemical characterizations are on-going at the moment.

a



b



c

Identified proteins; top hits based on total spectral counts	Accession Number	MW (kDa)	24h_Otiwi1	12h_Control
Macronuclear development protein 1 OS=Oxytricha trifallax	J9I8K4_9SPIT	89 kDa	751	
Serine/threonine-protein phosphatase OS=Oxytricha trifallax	J9FC96_9SPIT	36 kDa	182	
Tubulin beta chain OS=Oxytricha trifallax	J9IAZ2_9SPIT	49 kDa	119	24
Tubulin beta chain OS=Oxytricha trifallax	J9FW75_9SPIT	49 kDa	118	26
Serine/threonine-protein phosphatase OS=Oxytricha trifallax	J9II79_9SPIT	36 kDa	92	
Tubulin alpha chain OS=Oxytricha trifallax	J9IP51_9SPIT	50 kDa	59	9
Piwi-like protein 1 OS=Oxytricha trifallax	J9ILC9_9SPIT	104 kDa	59	0
Uncharacterized protein OS=Oxytricha trifallax	J9I5H7_9SPIT	99 kDa	58	0
tRNA splicing protein SPL1 OS=Oxytricha trifallax	J9I6E4_9SPIT	51 kDa	51	
Serine/threonine-protein phosphatase OS=Oxytricha trifallax	J9HZT6_9SPIT	35 kDa	50	
60S ribosomal protein L4 OS=Oxytricha trifallax	J9IQ31_9SPIT	45 kDa	49	0
Uncharacterized protein OS=Oxytricha trifallax	J9HWG8_9SPIT	92 kDa	48	0
Uncharacterized protein OS=Oxytricha trifallax	J9HIY0_9SPIT	38 kDa	44	4
60S ribosomal protein L3 OS=Oxytricha trifallax	J9HN47_9SPIT	45 kDa	44	1
DEAD/DEAH box helicase family protein OS=Oxytricha trifallax	J9I2J3_9SPIT	144 kDa	44	0
Uncharacterized protein OS=Oxytricha trifallax	J9HMN7_9SPIT	109 kDa	40	6
Uncharacterized protein OS=Oxytricha trifallax	J9I8M5_9SPIT	70 kDa	38	
Uncharacterized protein OS=Oxytricha trifallax	J9ICB9_9SPIT	64 kDa	37	0
Uncharacterized protein OS=Oxytricha trifallax	J9FZ54_9SPIT	93 kDa	35	1
60S ribosomal protein L7A, putative OS=Oxytricha trifallax	J9ICG8_9SPIT	29 kDa	35	2
Histone H4 OS=Oxytricha trifallax	J9J0Q6_9SPIT	12 kDa	33	
60S ribosomal protein L7 OS=Oxytricha trifallax	J9I0E1_9SPIT	28 kDa	32	2
Heat shock protein 90 OS=Oxytricha trifallax	J9HY07_9SPIT	81 kDa	32	0
Actin, cytoplasmic OS=Oxytricha trifallax	J9HWC4_9SPIT	42 kDa	30	1
Ribonucleoside-diphosphate reductase OS=Oxytricha trifallax	J9HVP5_9SPIT	98 kDa	30	0
Lipoxygenase, putative OS=Oxytricha trifallax	J9INR4_9SPIT	94 kDa	30	0
Uncharacterized protein OS=Oxytricha trifallax	J9IXD2_9SPIT	77 kDa	30	0
Kinesin-like protein OS=Oxytricha trifallax	J9EQP9_9SPIT	106 kDa	29	
40S ribosomal protein S4 OS=Oxytricha trifallax	J9IGC5_9SPIT	30 kDa	28	1

Otiwi1 (bait)
ribosomal subunits
protein phosphatase complex, harboring ANTAR domains
DEAD box RNA helicase

Figure 6. Native IP-MS experiment from Oxytricha cells at 24h post mating
a, IP efficiency using anti Otiwi1 antibody and 3 million Oxytricha cells; lower panel depicts IP using purified Otiwi1 as to estimate efficacy of antibody
b, GO term overview of Otiwi1 interactors
c, Top 30 hits (from ~400 total hits) based on total spectral counts; coloring sub-classifies protein interactors

DISCUSSION

Expression of Oxytricha proteins in Tetrahymena

Protein expression is an empirical and iterative process and only large, collective efforts in the biochemistry community lead to generalized and efficient expression protocols. As the ciliate community is small and conventional expression methods often don't work for ciliate proteins, only few strategies to express and purify ciliate proteins are available. Here, I successfully used a method spearheaded by my lab colleague Takahiko Akematsu that allows to express and purify *Oxytricha* Otiwi1 in *Tetrahymena*. The methodology is based on previous publications that establish well-tolerated tag combinations for ciliate proteins as well as guidelines for *Tetrahymena* cell line generation [37,38]. Cell line generation is an efficient process due to the combined transformant selection with 6-methyl purine and paromomycin. The end result is cells that are 100% genetically assorted and carry a copper inducible construct to express the protein of interests. Although protein expression and extraction efficiency are not comparable to conventional *E. coli* methods and are more labor-intensive, *Tetrahymena* is a well-suited and practical system to express and purify ciliate proteins.

A diverging functional mechanism of genome rearrangements in ciliates

Oxytricha belongs to the group of ciliates and shares many molecular characteristics with *Tetrahymena* and *Paramecium.* Despite similarities in genome rearrangement, *Oxytricha,* in that piRNAs target MDSs instead of IESs, carries out this process in the opposite manner. This difference represents an evolutionary sign change that occurred in an early lineage of these ciliates. A potential explanation for this change could be massive TE invasion and radiation in *Oxytricha* that forced the organism to eliminate 95% of its genomic sequence during genome rearrangement. This high burden of sequences to eliminate, makes it economically more efficient to specifically select the remaining 5% for

retention. A similar mechanism is found in *Stylonichia*, that belongs to the same class of ciliates as *Oxytricha* and is closely related. The evolutionary distance between *Oxytricha* and *Tetrahymena* is as long as that between humans and fungi, what makes the orthogonally different pathways less surprising.

The differences in the small RNA pathways further raise the question of how MDSs in *Oxytricha* are marked for retention. If the reason for the orthogonal pathway was only economics in the production of small RNAs, a natural hypothesis would be that H3K9me3 marks MDSs. H3K9me3 controls gene expression in all domains of life and is well established to silence TE expression in plants, animals and fungi. More importantly, H3K9me3 marks IESs in *Tetrahymena* and *Paramecium* for degradation. What speaks against this model is two things. First, many MDSs in *Oxytricha* are too small to form a nucleosome unit that is required for the deposition of histone modifications. Second, research from *Stylonichia* suggests alternative model where the piRNA-Piwi complex by itself constitutes the mark. Here, RNA/DNA base-pairing leads to the physical binding of a piRNA-Piwi complex to all MDSs in the developing MAC. This binding then leads to temporal stalling of replication during polytene chromosome formation and only replicated sequences are in turn degraded by a specific mechanism [31]. Taking this model into account, it's interesting that Otiwi1 strongly interacts with replication specific factors during genome rearrangement (Figure 6). It's unclear if an Otiwi1-piRNA complex preferably interacts with DNA rather than RNA. The proposed EMSA experiments as described above will deliver insights. In summary, there are promising hypotheses of the molecular mechanism to mark MDSs in *Oxytricha* but convincing evidence is yet lacking.

METHODS

Cell culture and *Oxytricha* mating

Oxytricha trifallax strains JRB310 and JRB510 were cultured in Pringsheim media (0.11 mM Na2HPO4, 0.08 mM MgSO4, 0.85 mM Ca(NO3)2, 0.35 mM KCl, pH 7.0) with algae *Chlamydomonas reinhardtii* and *Klebsiella* bacteria as food source. To induce mating, JRB310 and JRB510 cells were mixed right after food depletion. Cells started to pair 2–3 hr post-mixing, and conjugation efficiency was between 70% and 95%. Post-injection cells were raised in Volvic water.

Tetrahymena cells were grown at 30 °C in super proteose peptone (SPP) medium containing 1% proteose peptone (Becton Dickinson, Sparks, MD, USA), 0.1% yeast extract (Becton Dickinson), 0.2% glucose (Sigma-Aldrich, St. Louis, MO, USA), and 0.003% EDTA-Fe (Sigma-Aldrich). 6mp (Sigma-Aldrich) dissolved in distilled water at 15 mg/mL was used as a stock solution.

Cloning of constructs for *Tetrahymena* injection

Otiwi1 and Otiwi2 CDS were cloned from total RNA extracted via standard phenol-chloroform extraction from mated JRB310xJRB510 *Oxytricha* cells. Total RNA was reverse-transcribed using Superscript III Reverse Transcriptase (ThermoFisher Scientific) and oligo dT primers following manufacturer's instructions. Generated cDNA was used as template to amplify Otiwi1 and Otiwi2 via PCR. Quality and correct size of the PCR products were checked by agarose gel electrophoresis followed by Gibson assembly into previously digested plasmid pFZZ-NEO4 [37] using NEBuilder® HiFi DNA Assembly (NEB) following manufacturer's instructions. Assembled DNA was transformed into One Shot™ TOP10 Chemically Competent *E. coli* cells (Invitrogen) by heat-shock transformation and cells were grown at 37°C overnight. Plasmids were isolated by Miniprep using QIAprep® Miniprep kit (Quiagen) and the sequences were validated by sanger-sequencing (Genewiz).

Cloning of expression constructs for protein expression in *E. coli*

pET_His6 was a kind gift from the Sternberg lab and was used as the base vector to clone in different sequences. To clone in FZZ, pET_His6 was digested with SspI and EcoRI-HF and the digestion product was gel-extracted and purified using QIAquick Gel Extraction Kit (Quiagen). FZZ was amplified by PCR using previously generated plasmids as template and the construct was assembled by Gibson assembly and further processed as described above.

To generate a plasmid to express His-FZZ-Otiwi2 in *E. coli*, the Otiwi2 sequence was initially codon-optimized for *E. coli* expression (GenScript) and amplified, gel-extracted and purified as described above. pET_His6 was digested with SspI and EcoRI-HF and Gibson assembly with two inserts (FZZ and Otiwi2) was carried out as described above. All generated plasmids were validated by sanger-sequencing (Genewiz).

Cloning of constructs for *Oxytricha* microinjection

pL018_pOtiwi1_ex_FZZ was a kind gift from my lab colleague Takahiko Akematsu and harbors the FZZ coding sequence flanked by the 5 and 3' UTR of the endogenous Otiwi1 nanochromosome. The plasmid was digested with NdeI and SpeI and subjected to a Gibson assembly reaction with two inserts that were previously generated by PCR – 1) HA-FZZ (HA tag was added via FWD primer), 2) Otiwi1 WT or Otiwi1 H759A CDS. The constructs were transformed, miniprepped and validated by sanger-sequencing (Genewiz) as described above. Validated plasmids were then used as template for PCR to add full *Oxytricha* telomer sequences to the CDS to mimic the complete Otiwi1 nanochromosome. PCR products were verified by gel electrophoresis and DNA was extracted by ethanol precipitation. DNA pellets were resuspended in nuclease free water and run through an ultra-free MC column (Millipore). DNA brought to a final concentration of 1-3 mg/mL for microinjection.

Site directed mutagenesis

For construct mutagenesis, primers were designed following a one-step mutagenesis protocol (Zheng, Baumann, & Reymond, 2004). PCR was carried out using high-fidelity Phusion polymerase (NEB) with elongation time of 45sec/kb of the plasmid.
The PCR reaction was incubated with DpnI enzyme for 2h before purification and transformation.

Tetrahymena transformation

Prior to transformation, 30mL of cells were grown to saturation in SPP media (1% proteose peptone (Becton Dickinson, Sparks, MD, USA), 0.1% yeast extract (Becton Dickinson), 0.2% glucose (Sigma-Aldrich, St. Louis, MO, USA), and 0.003% EDTA-Fe (Sigma-Aldrich)), spun down at 1000g, washed once in starvation media (10mM Tris-Hcl) and incubated in starvation media for 2 days.
The injected plasmids were linearized by PCR amplification of the whole injection cassette. PCR reactions were quickly verified by agarose gel electrophoresis and purified using the QIAquick PCR purification kit (Quiagen). Gold particles (BioRad) were washed in 70% ethanol followed by three washes in H_2O and resupension in 50% glycerol (storage solution). 25uL of gold/glycerol was spun at 10 000g and glycerol was carefully pipetted out. Purified DNA was mixed with gold particles followed by addition of 50uL 2.5M $CaCl_2$ and 20uL of 0.1M spermidine, the mixture was vortexed well and stored at 4°C overnight. Next day, the gold particles were centrifuged, supernatant was decanted and the particles were washed twice with 70% Ethanol followed by one wash with 100% Ethanol. Ethanol was taken out and particles were transferred to plastic tubes for gene-gun injection (BioRad) and dried for 10 minutes. Starved cells were subsequently centrifuged and injected using the Helios® Gene Gun System under sterile conditions. Injected cells were recovered in 20mL SPP media supplemented with antibiotics for 3-4h at 30°C while gently shaking. After recovering, paromomycin (100ug/mL) was added and the 20mL SPP were transferred to a 96well plate. Cells were grown for 2 days at 30°C and positive clones were selected and genetically assorted using 500ug/mL of paromomycin and 15ug/mL 6-methylpurine.

Protein expression in Tetrahymena

100mL cells were grown to saturation for 2 days at 30°C and used to inoculate 900mL of SPP media. After ~5h of growth, 1ug/mL of $CdCl_2$ was added and cells were grown for another 20h at 30°C while gently shaking. Cells were harvested at 2500g, the pellet was washed once in 10mM Tris-Hcl and snap frozen in liquid nitrogen for protein purification.

Purification of FZZ-Otiwi1 and FZZ-Otiwi1 H759A

Tetrahymena cell pellets from expression culture were lysed in lysis buffer (LB) (20mM Tris-HCL pH 7.5, 300mM NaCl, 10% glycerol, 2mM $MgCl_2$, 1% NP-40) supplemented with 1mM TCEP, 30U/uL Benzonase and Complete protease inhibitor tables (Roche) and incubated at 4°C while gently shaking for 45min. Lysates were sonicated with 3 cycles to each 10sec ON, 20sec OFF for 2:30min at amplitude 30% with 1min pause-time between the cycles, on a Fisherbrand™ Ultra-homogenizer 500W with a big tip. Following sonication, lysates were cleared by centrifugation with 16.000g at 4°C for 30 minutes. Supernatants were subsequently filtered through two layers of cheesecloth and incubated with pre-equilibrated anti FLAG M2 beads (Sigma) for 3h at 4°C. Samples were spun down for 2min at 500g at 4°C, supernatant was carefully pipetted off and beads were resuspended in 1mL wash buffer (WB) (20mM Tris-HCL pH 7.5, 300mM NaCl, 10% glycerol, 2mM $MgCl_2$, 1% NP-40, 1mM ATP, 100mM KCL, 10mM $Mg(OAc)_2$ and washed six times. Beads were then rinsed three times in rinse buffer (RB) (20mM Tris-HCL pH 7.5, 150mM NaCl, 10% glycerol, 2mM $MgCl_2$, 1% NP-40). Proteins were pre-eluted in RB supplemented with 0.25mg/mL 3xFLAG peptide (Sigma) for 30 minutes at 4°C while rotating. Afterwards proteins were eluted three times in RB supplemented with 1mg/mL 3xFLAG peptide. Proteins were analyzed by SDS-PAGE.

Protein expression in *E. coli* and purification of FZZ

Prior to protein expression, the expression plasmid was transformed to BL21 competent cells (Agilent). A single transformed colony was then used to incubate a 10mL pre-culture to further incubate a 50mL main culture to an OD600 of 0.05. The culture was grown at 37°C until it reached OD600 of 0.8, induced with 1mM IPTG, cooled down to RT and

incubated at 20°C for 20h while shaking. Cells were harvested and separated into 4x 12.5mL cultures and snap frozen for future purification. FZZ was purified using a Ni-NTA Spin Kit (Quiagen) following manufacturer's instructions. FZZ was eluted by 500mM imidazole, dialyzed against FZZ dialysis buffer (10mM Tris-Hcl pH 7.5, 150mM NaCl, 1mM EDTA, 50% glycerol, 1mM PMSF, 5mM B-Mercaptoethanol), concentrated and snap-frozen at -80°C.

Western blot analysis

Western blot analysis was carried out following standard protocols. Samples were resolved by SDS-polyacrylamide gel electrophoresis and transferred to PVDF membrane by a semi-dry blotting apparatus (BioRad) in transfer buffer (48 mM Tris base, 39 mM glycine, 20 % methanol) at 10V for 40 minutes. The membrane was blocked in 5% milk in PBS-T (PBS + 0.05% Tween 20) for 1h and incubated with primary antibody in the same solution for overnight. Following antibody incubation, the membrane was washed three times in PBS-T and incubated with secondary antibody for 2h in 5% milk in PBS-T. Membrane was washed three times in PBS-T again, covered with Clarity Western ECL blotting substrate (BioRad) and imaged using a ChemiDoc MP imaging system.

Immunoprecipitation of Otiwi1 from *Oxytricha*

1.5 million mated cells were resuspended in 1.5mL lysis buffer (LB) (50mM Tris-HCL pH 7.5, 150mM NaCl, 2mM EDTA, 1% NP-40) supplemented with 1mM TCEP, 0.1% sodium deoxycholate and Complete protease inhibitors (Roche). 2uL of Benzonase (30U/mL) were added and samples were lysed for 1h at 4°C while rotating. Samples were sonicated with 2 cycles to each (1sec ON, 9sec OFF, 10x) with amplitude 20% on a Fisherbrand™ Ultra-homogenizer 500W and a small tip. Lysates were cleared by centrifugation at 18.000g for 15 minutes and the total protein concentrations were measured using standard Bradford assay to normalize the samples to each other. A small aliquot per sample was stored on ice as Input (IN). Samples were incubated with 2uL of antibody each (Anti-PIWIL1, abcam12337) for 3h at 4°C. After 3h, 6uL of equilibrated dynabeads protein G were added and samples were further incubated overnight. An unbound (UB)

aliquot was taken and stored on ice. Beads were washed four times for 10 minutes with 1mL of LB. A small immunoprecipitation (IP) aliquot was taken. The rest of the sample was rinsed four times in LB without 1% NP-40. Beads were stored at -20°C, IN, UB, IP aliquots were analyzed by SDS-PAGE. Beads were analyzed by mass spectrometry.

Mass spectrometry analysis

Mass spectrometry analysis was carried out as an on-bead digestion service provided by the Proteomics and Macromolecular Crystallography center at HICCC shared resources, Columbia University. The proteomics data was then visualized and inspected using Scaffold Viewer software.

RT-qPCR

Total RNA was extracted from frozen 310x510 mated *Oxytricha* cells using TRIzol reagent and standard RNA extraction protocols. 1ug of total RNA was digested with RQ1 RNase-free DNase (Promega) and reverse transcribed using Superscript III, oligo dT primers and standard protocols. cDNA was used as template for RT-qPCR to quantify the mRNA expression of Dicer proteins in *Oxytricha*. RT-qPCR was carried out using Maxima SYBR Green Master Mix (ThermoFisher Scientific) and the PCR protocol suggested by the manufacturer.

REFERENCES

[1] I. H. G. S. Consortium, *Nature* **2001**, *409*, 860–921.

[2] D. A. Petrov, Y. T. Aminetzach, J. C. Davis, D. Bensasson, A. E. Hirsh, *Mol. Biol. Evol.* **2003**, *20*, 880–892.

[3] R. N. McLaughlin, H. S. Malik, H. S. Malik, *J. Exp. Biol.* **2017**, *220*, 6–17.

[4] D. Haig, *BioEssays* **2016**, *38*, 1158–1166.

[5] E. V. Koonin, K. S. Makarova, Y. I. Wolf, M. Krupovic, *Nat. Rev. Genet.* **2019**, 1–13.

[6] D. M. Ozata, I. Gainetdinov, A. Zoch, D. O'Carroll, P. D. Zamore, *Nat. Rev. Genet.* **2019**, *20*, 89–108.

[7] M. Ghildiyal, P. D. Zamore, *Nat. Rev. Genet.* **2009**, *10*, 94–108.

[8] V. V. Vagin, A. Sigova, C. Li, H. Seitz, V. Gvozdev, P. D. Zamore, *Science (80-.).* **2006**, *313*, 320–324.

[9] A. Pelisson, E. Sarot, G. Payen-Groschene, A. Bucheton, *J. Virol.* **2007**, *81*, 1951–1960.

[10] V. T. Yerlici, L. F. Landweber, *Microbiol. Spectr.* **2014**, *2*, DOI 10.1128/microbiolspec.MDNA3-0025-2014.

[11] A. G. Seto, R. E. Kingston, N. C. Lau, *Mol. Cell* **2007**, *26*, 603–609.

[12] S. E. Allen, M. Nowacki, *J. Mol. Biol.* **2020**, *432*, 4186–4198.

[13] W. Foissner, **2007**, pp. 1–8.

[14] D. H. Lynn, *The Ciliated Protozoa : Characterization, Classification, and Guide to the Literature*, Springer **2008**.

[15] Z. Yi, W. Song, J. C. Clamp, Z. Chen, S. Gao, Q. Zhang, *Mol. Phylogenet. Evol.* **2009**, *50*, 599–607.

[16] V. T. Yerlici, L. F. Landweber, **2014**, DOI 10.1128/microbiolspec.MDNA3-0025-2014.

[17] S. Duharcourt, G. Lepère, E. Meyer, *Trends Genet.* **2009**, *25*, 344–50.

[18] M. Nowacki, L. F. Landweber, *Curr. Opin. Microbiol.* **2009**, *12*, 638–643.

[19] X. Chen, J. R. Bracht, A. D. Goldman, E. Dolzhenko, D. M. Clay, E. C. Swart, D. H. Perlman, T. G. Doak, A. Stuart, C. T. Amemiya, R. P. Sebra, L. F. Landweber, **2014**, DOI 10.1016/j.cell.2014.07.034.

[20] E. Hamilton, P. Bruns, C. Lin, V. Merriam, E. Orias, L. Vong, D. Cassidy-Hanley, *Genetics* **2005**, *170*, 1611–21.

[21] F. Guérin, O. Arnaiz, N. Boggetto, C. Denby Wilkes, E. Meyer, L. Sperling, S. Duharcourt, *BMC Genomics* **2017**, *18*, 327.

[22] L. M. Epstein, J. D. Forney, *Mol. Cell. Biol.* **1984**, *4*, 1583–90.

[23] K. Mochizuki, N. A. Fine, T. Fujisawa, M. A. Gorovsky, *Cell* **2002**, *110*, 689–99.

[24] E.-M. Weick, E. A. Miska, *Development* **2014**, *141*, 3458–71.

[25] T. Noto, H. M. Kurth, K. Kataoka, L. Aronica, L. V DeSouza, K. W. M. Siu, R. E. Pearlman, M. A. Gorovsky, K. Mochizuki, *Cell* **2010**, *140*, 692–703.

[26] U. E. Schoeberl, K. Mochizuki, *J. Biol. Chem.* **2011**, *286*, 37045–52.

[27] W. Fang, X. Wang, J. R. Bracht, M. Nowacki, L. F. Landweber, *Cell* **2012**, *151*, 1243.

[28] D. C. Swarts, K. Makarova, Y. Wang, K. Nakanishi, R. F. Ketting, E. V Koonin, D. J. Patel, J. Van Der Oost, *Nat. Struct. Mol. Biol.* **2014**, *21*, DOI 10.1038/nsmb.2879.

[29] I. Gainetdinov, C. Colpan, A. Arif, K. Cecchini, P. D. Zamore, *Mol. Cell* **2018**, *71*, 775-790.e5.

[30] S. M. Hammond, *FEBS Lett.* **2005**, *579*, 5822–5829.

[31] J. Postberg, F. Jönsson, P. P. Weil, A. Bulic, S. A. Juranek, H.-J. Lipps, *Epigenetics Chromatin* **2018**, *11*, 31.

[32] K. W. Doxzen, J. A. Doudna, *PLoS One* **2017**, *12*, e0177097.

[33] J.-J. Song, S. K. Smith, G. J. Hannon, L. Joshua-Tor, *Science* **2004**, *305*, 1434–7.

[34] N. Matsumoto, H. Nishimasu, K. Sakakibara, H. Siomi, M. C. Siomi, O. N. Correspondence, K. M. Nishida, T. Hirano, R. Ishitani, O. Nureki, *Cell* **2016**, *167*, 484–497.

[35] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper, A. Sali, *Curr. Protoc. Bioinforma.* **2006**, *Chapter 5*, Unit-5.6.

[36] L. S. Gunawardane, K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, M. C. Siomi, *Science (80-.).* **2007**, *315*, 1587–1590.

[37] K. Kataoka, U. E. Schoeberl, K. Mochizuki, *J. Microbiol. Methods* **2010**, *82*, 342–6.

[38] T. Akematsu, A. Findlay, Y. Fukuda, R. E. Pearlman, J. Loidl, E. Orias, E. P Hamilton, *Genes (Basel).* **2018**, *9*, DOI

10.3390/genes9040179.

[39] D. M. Clay, V. T. Yerlici, D. J. Villano, L. F. Landweber, *G3: GenesIGenomesIGenetics* **2019**, *9*, 3105–3118.

[40] J. S. Khurana, X. Wang, X. Chen, D. H. Perlman, L. F. Landweber, *Genetics* **2014**, *197*, 839–49.

[41] D. I. Furrer, E. C. Swart, M. F. Kraft, P. Y. Sandoval, M. Nowacki, *Cell Rep.* **2017**, *20*, 505–520.

[42] T. Miyoshi, K. Ito, R. Murakami, T. Uchiumi, *Nat. Commun.* **2016**, *7*, 11846.

SUPPLEMENT

Constructs used in this study:

p1_pFZZ_Otiwi1_APRT1locus_NTERM

p2_Otiwi1_FZZ_APRT1locus_CTERM

p3_FZZ_Otiwi2_APRT1locus_NTERM

p4_Otiwi2_FZZ_APRT1locus_CTERM

p5_pETHis6TEV_FZZ

p7_pFZZ_Otiwi1_H759A_APRT1locus_NTERM

p8_pETHis6TEV_Otiwi2_FZZ

p9_pETHis6TEV_FZZ_Otiwi2

p10_HA_FZZ_otiwi1_endog_pl018

p12_HA_FZZ_otiwi1_H759A_endog_pl018

RT-qPCR primers used in this study:

FP45_qPCR_Otiwi1_FWD;	CAACGGTATCTGCGTTCCTG
FP46_qPCR_Otiwi1_REV;	GCGAACGGGAGATGAAAGAC
FP62_DCR1_qPCR FWD;	ACTCAATCGCCCAGAGTTAAG
FP63_DCR1_qPCR REV;	CTGGTCTTCAGTAGGCAAAGAT
FP68_DCR2_qPCR FWD;	CTACTGCACCAGGTTTGTATGT
FP69_DCR2_qPCR REV;	TCTCCTTCTCCTGCCTTCTT
FP74_Dicer_qPCR FWD;	CCAACGTCATAGAGGAAGGATTAG
FP75_Dicer_qPCR REV;	CGTTGTTCTTACGAGCACATAAAT

Fig S2. MSA of different Piwi domains
catalytic residues are indicated