Characterisation of the RNA-interference pathway as a Tool for Genetics in the Nascent Phototrophic Endosymbiosis, *Paramecium bursaria*

Benjamin H. Jenkins[1,2] *†, Finlay Maguire[3] *, Guy Leonard[1,2], Joshua D. Eaton[1], Steven West[1], Benjamin E. Housden[1], David, S. Milner[1,2], & Thomas A. Richards[1,2] †

[1]Living Systems Institute and Biosciences, University of Exeter, Devon EX4 4QD, UK

[2]Department of Zoology, University of Oxford, 11a Mansfield Road, Oxford OX1 3SZ, UK

[3]Faculty of Computer Science, Dalhousie University, 6050 University Ave, Halifax NS B3H 1W5, Canada

* These authors contributed equally

† Correspondence: ben.jenkins@zoo.ox.ac.uk & thomas.richards@zoo.ox.ac.uk

KEYWORDS

symbiosis, interaction, algae, RNAi, Dicer, siRNA

ABSTRACT

Endosymbiosis was fundamental for the evolution of eukaryotic complexity. Endosymbiotic interactions can be dissected through forward and reverse-genetic experiments, such as RNA-interference (RNAi). However, distinguishing small (s)RNA pathways in a eukaryote-eukaryote endosymbiotic interaction is challenging. Here, we investigate the repertoire of RNAi pathway protein-encoding genes in the model nascent endosymbiotic system, *Paramecium bursaria–Chlorella* spp. Using comparative genomics and transcriptomics supported by phylogentics, we identify essential proteome components of the small interfering (si)RNA, scan (scn)RNA, and internal eliminated sequence (ies)RNA pathways. Our analyses reveal that copies of these components have been retained throughout successive whole genome duplication (WGD) events in the *Paramecium* clade. We then validate feeding-induced siRNA-based RNAi in *P. bursaria* via knock-down of the splicing factor, *u2af1*, which we show to be crucial to host growth. Finally, using simultaneous knock-down paradox controls to rescue the effect *u2af1* knock-down, we demonstrate that feeding-induced RNAi in *P. bursaria* is dependent upon a core pathway of host-encoded *Dcr1*, *Piwi* and *Pds1* components. Our experiments confirm the presence of a functional, host-derived RNAi pathway in *P. bursaria* that generates 23-nt siRNA, validating use of the *P. bursaria-Chlorella* spp. system to investigate the genetic basis of a nascent endosymbiosis.

INTRODUCTION

Endosymbiosis was fundamental for the evolution of eukaryotic cellular complexity[1–4]. In order to investigate the genetic basis of an emergent endosymbiotic system, we must develop experimentally tractable endosymbiotic model species[5–7]. *Paramecium bursaria* is a ciliate protist which harbours several hundred cells of the green algae, *Chlorella* spp., in a nascent facultative photo-endosymbiosis[8–12]. The algae provide sugar and oxygen derived from photosynthesis, in exchange for amino acids, CO_2, divalent cations, and protection from viruses and predators[5,6,13–19]. While the interaction is heritable, the *P. bursaria-Chlorella* spp. system is described as a 'nascent' or 'facultative' endosymbiosis, as both host and endosymbiont can typically survive independently[10,20–23]. *P. bursaria* therefore represents a potentially tractable model system with which to investigate the genetic basis of a nascent endosymbiotic cell-cell interaction.

RNA-interference (RNAi) is a form of post-transcriptional gene silencing that is dependent upon conserved small (s)RNA processing pathways[24–26]. The principal components of a functional RNAi pathway are conserved in many eukaryotes[27–29], though loss in some lineages suggests a mosaic pattern of pathway retention[30]. Typically, these pathways rely on size-specific sRNA processing via an endoribonuclease Dicer[26,27], targeted RNA cleavage activity of an Argonaute (AGO-Piwi) containing effector complex[28,31], and RNA-dependent RNA polymerase (RdRP) amplification of either primary or secondary sRNA triggers[32–37]. These triggers may include partially degraded mRNA cleavage products[36], exogenous sRNA[36], or full length mRNA transcripts[32], suggesting that RdRPs may have broader functions in some systems.

In ciliates, multiple whole genome duplication (WGD) events have led to the rapid expansion of gene families encoding RNAi components[38–40], resulting in a subsequent diversification of protein function. Example functions include transposon elimination, nuclear rearrangement, and transcriptional regulation[32,35,41–45]. Elegant investigation of the non-photo-endosymbiotic model system, *Paramecium tetraurelia,* has identified three distinct classes of RNAi pathway in *Paramecium*. The ciliate-specific scan (scn)RNA (25-nt) and internal eliminated sequence (ies)RNA (~28-nt) pathways are endogenous, and function primarily to eliminate the bulk of noncoding DNA present in the germline micronuclear genome during development of the somatic macronucleus[35,41–43,46,47]. *Paramecium* also encodes a short-interfering (si)RNA (23-nt) pathway capable of processing both exogenously[48–51] and endogenously[32,44] derived RNA precursors. Although siRNA is believed to have evolved to protect against foreign genetic elements (such as viruses, transposons and transgenes)[52], some siRNA-based RNAi factors have also been implicated in the regulation of endogenous transcriptome expression in the non-photo-endosymbiotic model system, *P. tetraurelia*[32].

In the photo-endosymbiotic *P. bursaria-Chlorella* spp. system, the existence of a functional siRNA-based RNAi pathway would provide an experimental approach to knock-down gene expression via the delivery of exogenously derived dsRNA homologous to a target transcript[33,48]. Preliminary evidence suggests that siRNA-based RNAi can be induced in *P. bursaria* (110224 strain)[53]; however, a comprehensive analysis with appropriate controls has yet to be conducted. To demonstrate direct evidence of an RNAi-mediated effect, one would need to rescue a putative phenotype through targeted inhibition of the RNAi-knock-down machinery. Such controls are of paramount importance when conducting genetic knock-down experiments in a complex endosymbiotic system, and the presence of a eukaryotic green algal endosymbiont in *P. bursaria* necessitates caution. RNAi has been reported in some green-algal species[54], and thus it is important that controlled experimental characterisation of these distinct pathways be conducted before genetic knock-down in *P. bursaria* can be inferred.

Here, we elucidate a cognate repertoire of predicted RNAi component-encoding genes present in *P. bursaria*, confirming that the host genome encodes essential proteome constituents of the siRNA-, scnRNA-, and iesRNA- based RNAi pathways. These include multiple paralogues/orthologues of the pathway components; Dicer, Dicer-like, Piwi (AGO-Piwi), Rdr (RdRP), Cid and Pds1, which have been identified in the non-photo-endosymbiotic model system, *P. tetraurelia*[33,34,36]. We trace the occurrence of RNAi protein-encoding genes in the *Paramecium* clade using comparative genomics combined with transcriptomics and supported by phylogenetic analysis, and demonstrate that these genes have been retained throughout successive WGDs. Using an *E. coli* vector feeding-based approach for RNAi induction, we demonstrate functional siRNA-based RNAi in *P. bursaria* via knock-down of a conserved ciliate splicing factor, *u2af1,* which we show to be similar to the *u2af* (65 kDa) constitutive splicing factor present in humans[55,56]. We demonstrate that RNAi-mediated knock-down of *u2af1* results in significant culture growth retardation in *P. bursaria*, suggesting that this gene has a critical function. Finally, we corroborate the function of several siRNA-based RNAi factors in *P. bursaria*; including *Dcr1*, two unduplicated AGO-Piwi factors (*PiwiA1* and *PiwiC1*), and a *Paramecium*-specific *Pds1*, via simultaneous component knock-down to rescue *u2af1* culture growth retardation. Collectively, these data support the presence of a functional, host derived, exogenously-induced siRNA-based RNAi pathway in the *P. bursaria-Chlorella* spp. endosymbiotic system, dependent on *Dcr1*, *Piwi* and *Pds1* protein function.

RESULTS

Bioinformatic Identification of a Putative RNAi Pathway in *P. bursaria*

A feeding-induced siRNA-based RNAi pathway has been validated as a tool for gene knock-down in the non-photo-endosymbiotic ciliate, *P. tetraurelia*[33,34,36,48]. To establish the presence of a comparable pathway in *P. bursaria*, combined genomic- and transcriptomic-analyses were employed to identify putative homologues for all previously characterised *P. tetraurelia* RNAi protein components[33,45] (**Figure 1**). We found that *P. bursaria* encodes a total of five Dicer or Dicer-like endonucleases (*Dcr1*, *Dcr2/3* – **Dataset S1**; *Dcl1/2*, *Dcl3/4* and *Dcl5* – **Dataset S2**), three RdRPs (*Rdr1/4*, *Rdr2* and *Rdr3* – **Dataset S3**), six AGO-Piwi components (*PiwiA1, PiwiA2, PiwiB, PiwiC1, PiwiC2* and *PiwiD*– **Dataset S4**), a single *Paramecium*-specific Pds1 (*Pds1* – **Dataset S5**), and two nucleotidyl transferase (*Cid1/3* and *Cid2* – **Dataset S6**) genes. Among those identified are homologues of the essential feeding-induced siRNA pathway components present in *P. tetraurelia*. In *P. bursaria,* these are; *Dcr1*, *Pds1, Rdr1/4*, *Cid1/3, Cid2,* and putative *PiwiA1* and *PiwiC1* homologues, although we were unable to accurately identify the precise relationship of these *Piwi* paralogues due to lack of phylogenetic resolution (**Dataset S4**). Sequences corresponding to each of these RNAi protein-encoding genes were present in our *P. bursaria* transcriptome dataset, indicating that these host-derived RNAi components are transcriptionally active. These data reveal that *P. bursaria* encodes a putative functional feeding-induced siRNA pathway, indicating that an experimental approach to knock-down gene expression is tractable in this system. Additionally, we show that *P. bursaria* encodes homologues for components of the transgene-induced siRNA pathway, as well as the endogenous ciliate-specific scnRNA and iesRNA pathways involved in nuclear reorganisation and development. For a full list of identified RNAi components, and predicted associated pathways in *P. bursaria*, see **Table 1.**

Further analyses were conducted to identify the presence of comparable RNAi pathway components in the algal endosymbiont. For a full overview of the host and endosymbiont transcriptome dataset binning process, please refer to the **Methods**. Using both *P. tetraurelia* and *C. reinhardtii* query sequences, our analyses identified a putative homologue for *Dcl1* that clustered with strong support to known green algal Dicer sequences. (**Dataset S7**). No algal homologue for AGO-Piwi or RdRP could be detected in any of the algal-endosymbiont datasets sampled, suggesting that these components are either not transcriptionally active, or are absent altogether in the algal endosymbiont of *P. bursaria*. The absence of RdRP is consistent with its absence in most green algal species sampled[54]. To explore the possible function of a *Dcl* homologue in the sampled endosymbiotic green algae of *P. bursaria*, we conducted sRNA sequencing of algae isolated from the host under standard growth conditions. Our sRNA sequencing data demonstrated that the isolated algal endosymbiont of *P. bursaria* was not actively generating sRNA >20-nt (**Figure S1**), confirming that the mode length of endosymbiont Dicer-derived sRNA does not resemble those of siRNA (23-nt), scnRNA (25-nt) or iesRNA (~28-nt) known to be generated by the non-photo-endosymbiotic model system, *P. tetraurelia*[33–

[35,41,42]. This is an important distinction, as it would allow one to ensure that any genetic knock-down approach in the *P. bursaria-Chlorella* spp. system could be attributed to the *Paramecium* host, based on the size of sRNAs generated.

Validation of Feeding-Based RNAi in *P. bursaria*

To demonstrate the activity of the putative siRNA-based RNAi pathway in *P. bursaria*, identified in **Figure 1**, we targeted the conserved splicing factor encoding gene, *u2af* [55,56]. Many ciliates genomes are intron-rich and dependent upon splicing for both macronuclear generation and transcription[57], thus it was predicted that knock-down of *u2af* would considerably impact *P. bursaria* growth. Transcriptome analysis revealed that *P. bursaria* encodes three paralogues with sequence similarity to the *u2af* (65 kDa) constitutive splicing factor present in humans[55,56], and indicates that these paralogues likely diverged prior to the radiation of the ciliate clade (**Figure 2a**). Interestingly, the *u2af1* orthologue has been subject to gene duplication prior to diversification of the *Paramecium aurelia* species complex, with >85% copy retention across all species sampled ([n = 39 retained / 11*4 predicted], showing 5 putative gene losses in 11 taxa; **Dataset S8**).

Phylogenetic analysis indicates that *u2af1* is highly conserved in ciliates, supporting the hypothesis that it may have an essential function (**Figure 2b**). Using an *E. coli* vector-based feeding approach for RNAi induction, delivery of a 500-nt dsRNA fragment corresponding to *u2af1* resulted in significant *P. bursaria* culture growth retardation compared to an empty vector control, consistent with an RNAi effect (**Figure 2c**). Interestingly, retardation to culture growth in response to *u2af1* dsRNA exposure was greater under constant darkness (D 24h), with a mean cell number after 12 days that was significantly less (-84.4%; ***) compared to parallel cultures maintained under standard light-dark (LD 12:12h) conditions. This is consistent with an increased rate of *P. bursaria* feeding in the dark resulting in greater *E. coli* vector uptake[58]. Using mRNA extracted from *P. bursaria* during *u2af1*-RNAi feeding, qPCR revealed a significant reduction in *u2af1* gene expression in response to complementary dsRNA exposure (**Figure 2d**).

Next, we designed a 450-nt 'scramble' dsRNA control using a 'DNA shuffle' tool (https://www.bioinformatics.org/sms2/shuffle_dna.html) to randomly shuffle a 450-nt nonsense sequence, ensuring that the resultant 'scramble' dsRNA bore no significant sequence similarity to any *P. bursaria* host or algal endosymbiont transcripts present in the transcriptome datasets. For confirmation of the null effect of 'scramble' dsRNA exposure compared to an empty vector control, see **Figure S2**. Following 'scramble' dsRNA exposure, sRNA isolated from *P. bursaria* was sequenced and mapped against the original 'scramble' DNA template. This allowed us to demonstrate a distinct abundance of sense and antisense 23-nt reads in *P. bursaria* (**Figure 2e**). These results are consistent

with previous studies demonstrating *Dicer*-dependent cleavage of dsRNA into 23-nt fragments in *C. elegans* and *P. tetraurelia*[26,35]. Collectively, these data confirm the presence of a Dicer-mediated siRNA-based RNAi pathway capable of processing exogenously-derived dsRNA into 23-nt siRNA, and which is induced through the consumption of bacterial cells via phagocytosis.

Investigating *Dcr1* Function

Having demonstrated feeding-based RNAi induction, we investigated putative Dicer function in *P. bursaria*. Further dsRNA constructs (Dcr1A, Dcr1B) were designed to specifically target two regions of the *Dcr1* transcript present in *P. bursaria*. A BLASTn search against the *P. bursaria-Chlorella* spp. host and endosymbiont transcript datasets confirmed that the identified dsRNA template from these constructs was predicted to target only *Dcr1,* accounting for all possible 23-nt fragments and allowing for ≤2-nt mismatches. Using mRNA extracted from *P. bursaria* during *Dcr1*-RNAi feeding, qPCR revealed knock-down of *Dcr1* in response to Dcr1A and Dcr1B exposure (**Figure 3a**). Importantly, we found that knock-down was only detected in *Dcr1* when the qPCR amplicon was directly adjacent to the dsRNA target site, with detectable mRNA reduction less evident as the qPCR target amplicon was moved further along the transcript towards the 5' end. This finding suggests that the transcript is only partially degraded upon dsRNA-mediated knock-down – an important consideration for the design of effective RNAi reagents for future experiments.

We next checked for the occurrence of any off-target effects arising from Dcr1A/Dcr1B dsRNA exposure, as these may result in knock-down of additional Dicer or Dicer-like components in *P. bursaria*. An additional set of qPCR amplicons was designed to target each of the *Dcr2/3*, *Dcl1/2*, *Dcl3/4* and *Dcl5* transcripts identified from our host transcript dataset (**Figure 1**). Full-length sequences were derived from genomic data and compared to respective transcriptome data to ensure that each transcript encompassed the entire open reading frame, allowing us to assess expression from approximately the same relative position on each transcript. A further BLASTn search against the *P. bursaria-Chlorella* spp. host and endosymbiont transcript datasets confirmed that each qPCR amplicon site was specific to the host. Using mRNA extracted from *P. bursaria* during *Dcr1*-RNAi feeding, and qPCR amplicons adjacent to the equivalent position of the Dcr1A/B dsRNA target site, qPCR revealed no significant knock-down in *Dcr2/3*, *Dcl1/2*, *Dcl3/4*, or *Dcl5* transcripts in response to Dcr1A and Dcr1B dsRNA exposure (**Figure 3b**). Indeed, we noted an increase in *Dcr2/3, Dcl1/2* and *Dcl3/4* expression suggesting that these transcripts are potentially being up-regulated to compensate for reduced *Dcr1* expression. These data confirm that exposure to Dcr1A/Dcr1B dsRNA results in specific knock-down of host *Dcr1* in *P. bursaria*.

To understand the effect of *Dcr1* knock-down on endogenously triggered *P. bursaria* RNAi function, a size distribution of global host-derived sRNA abundance was compared between cultures exposed to Dcr1A and Dcr1B dsRNA, or a non-hit 'scramble' dsRNA control (**Figure 3c**). A significant reduction in 23-nt sense and antisense sRNA reads was observed upon Dcr1A/Dcr1B dsRNA exposure, accompanied by no significant reduction in any other sRNA read size between 20-30 nt. These data demonstrate that delivery of Dcr1A/Dcr1B dsRNA results in a specific reduction in endogenous 23-nt siRNA abundance, indicative of disruption of predicted *Dcr1* function. An increase in all ≥25-nt sRNA reads upon Dcr1A/Dcr1B dsRNA exposure (**Figure 3c**) may correspond to the increased expression of *Dcr2/3*, *Dcl1/2* and *Dcl3/4* transcripts observed in **Figure 3b** (see also **Table 1**), further corroborating that these additional *Dicer* or *Dicer-like* components are potentially being up-regulated in *P. bursaria* to compensate for disruption of *Dcr1* function. Nonetheless, the absence of significant reduction in all other sRNA sizes (with the exception of 23-nt) indicates that knock-down in response to Dcr1A/Dcr1B dsRNA exposure is effective in specifically reducing host *Dcr1* function in *P. bursaria.*

Validation of *Dcr1*, *Piwi* and *Pds1* Function

Finally, we sought to corroborate the function of a putative feeding-induced siRNA-based RNAi pathway in *P. bursaria*. In an attempt to disrupt *P. bursaria* siRNA-based RNAi function, we exposed cultures to Dcr1A/Dcr1B dsRNA during *u2af1*-RNAi feeding. Simultaneous knock-down of *Dcr1* during *u2af1*-RNAi feeding gave rise to an 'RNAi rescue' phenotype, restoring *P. bursaria* culture growth in *u2af1* dsRNA exposed cultures (**Figure 4a**). Importantly, this effect was significantly greater than the same relative simultaneous delivery of an empty vector control during *u2af1*-RNAi feeding, indicating that rescue of *P. bursaria* culture growth wasn't due to dilution of *u2af1* dsRNA template.

We next aimed to corroborate the function of three feeding-induced siRNA-based components that were least supported by phylogenetic inference in *P. bursaria*: *PiwiA1*, *PiwiC1* and *Pds1*. As for *Dcr1*, simultaneous knock-down of either *Pds1*, *PiwiA1*, or *PiwiC1* during *u2af1*-RNAi feeding each gave rise to an 'RNAi recue' phenotype, restoring *P. bursaria* culture growth in *u2af1* dsRNA exposed cultures (**Figure 4b**). *Pds1* is a Paramecium-specific component of feeding-induced siRNA-based RNAi first discovered in *P. tetrauelia*[33]. Sequence homology searches of known functional protein domains could not ascribe a function to Pds1, however it was suggested that this protein may play a role in the export of dsRNA from the digestive vacuole into the host cytoplasm[33,36]. Our confirmation that feeding-induced siRNA-based RNAi in *P. bursaria* is dependent upon *Pds1* is important. As the sampled green algae do not encode an identifiable homologue of *Pds1*, this re-iterates that the RNAi effect we have observed is derived from the *P. bursaria* host, and not from the algal endosymbiont.

Delivery of *Dcr1*, *PiwiA1*, *PiwiC1* or *Pds1* dsRNA to perturb siRNA-based RNAi function will never provide a complete 'RNAi rescue', since they are themselves important for cellular function. Indeed, partial knock-down of *Dcr1* in this manner may explain why reduction in *Dcr1* transcript expression was not deemed to be statistically significant via qPCR (**Figure 3b**). Mutagenesis screens in *P. tetraurelia* have previously revealed that *Dcr1* null alleles typically result in lethality[33], suggesting that these pathway components have essential functions in *Paramecium*. We propose that partial knock-down of *Dcr1* via an *E. coli* feeding vector-based paradox-approach is therefore preferable to total silencing that would otherwise kill the cell. Perturbation of RNAi through disruption of *Dcr1* (knock-down, rather than knock-out) is sufficient to attenuate the RNAi effect, and thereby provide an appropriate control for inferring a *bona fide* RNAi effect through feeding-induced siRNA-mediated gene knock-down. We have demonstrated that disruption of these essential RNAi components is effective at perturbing both background endogenous (*Dcr1*; **Figure 3c**) and exogenously triggered (*Dcr1, PiwiA1, PiwiC1* and *Pds1*; **Figure 4a/b**) siRNA pathways in *P. bursaria*. Taken together, these data confirm that feeding-induced siRNA-based RNAi in *P. bursaria* is dependent upon host *Dcr1, PiwiA1*, *PiwiC1* and *Pds1* function.

DISCUSSION

Here, we have identified the repertoire of cognate RNAi components present in *P. bursaria,* including essential proteome constituents of the siRNA, scnRNA, and iesRNA RNAi pathways. These include orthologues of the pathway components; Dicer, Dicer-like, Piwi, Rdr, Cid and Pds1 that are present in the non-photo-endosymbiotic model system, *P. tetraurelia*. Our comparison across the *Paramecium* clade (**Figure 1**) reveals that many of these components likely originated from the whole genome duplication (WGD) event that occurred prior to the radiation of the *Paramecium aurelia* species complex, which diverged separately from the *P. bursaria* lineage. Importantly, an unusually large number of copies of RNAi-component encoding genes have been retained in the *Paramecium* clade (>80% retention for all components; **Table 2**), exceeding the 40-60% retention rate observed in paralogues of this WGD event across the *Paramecium aurelia* species complex[40]. This observation suggests that these RNAi components are either highly expressed, and thus retention is enforced by gene dosage constraints, and/or have undergone significant neo- or sub-functionalisation driving retention of these paralogues following initial duplication[59]. Our phylogenetic analysis of Dicer, RdRP and AGO-Piwi components (**Datasets S1-4**) supports the occurrence of at least three WGD events within the ciliate group[38–40]. These are hypothesised to have occurred i) after the divergence of the CONThreeP clade (Colpodea, Oligohymenophorea, Nassophorea, Prostomatea, Plagiopylea and Phyllopharyngea)[60,61] from *Oxytricha trifallax* (Spirotrichea) and the broader ciliates, ii) after the divergence of *Paramecium* from *Tetrahymena thermophilia* and remaining Oligohymenophorea, and

iii) after the divergence of the *Paramecium aurelia* species complex from the remainder of the *Paramecium* clade (*Paramecium caudatum* and *Paramecium bursaria*).

Using an *E. coli* vector feeding-based approach to RNAi induction, we have demonstrated that knock-down of a conserved splicing factor, *u2af1*, results in *P. bursaria* culture growth retardation. Segregation of host germline and somatic nuclei in *Paramecium* makes long term stable [conventional] transformation methods inconsistent and therefore unfeasible for systematic functional genomic profiling. *Paramecium* species are known to conjugate through sexual reproduction approximately every 200 generations[62,63]. This means that a growing library of somatic transformants (featuring RNAi deficient mutations) would need to be maintained and propagated through mitosis to prevent these genetic changes from being lost upon regeneration of the somatic macronucleus. We therefore propose that delivery of exogenously derived dsRNA complementary to a target transcript, in the manner conducted in this study and others[48–51], remains the optimal experimental approach for large-scale gene knock-down surveys in this, and possibly other, ciliate systems.

Finally, we have corroborated the function of several RNAi components; including *Dcr1*, two unduplicated Piwi factors (*PiwiA1* & *PiwiC1*), and *Pds1*, via simultaneous component knock-down to rescue *P. bursaria* cell growth, supporting the hypothesis that these factors are required for exogenously-induced siRNA-based RNAi induction in this system. We have demonstrated that, though the algal endosymbiont encodes a putative RNAi pathway including a *Dcl1* homologue, these do not appear to generate sRNAs in the same size range as the host (**Figure S1**). This, together with our assessment of the *Paramecium*-specific factor, *Pds1,* further reinforces that any RNAi effect initiated through a feeding-based approach is likely host-derived. The data presented in this study have allowed us to de-convolute a functional exogenously-inducible siRNA-based RNAi pathway in the endosymbiotic ciliate, *P. bursaria*. We hope that these results will further promote the use of the *P. bursaria-Chlorella* spp. endosymbiosis as a key model system to investigate the genetic basis of a nascent endosymbiotic cell-cell interaction.

AUTHOR CONTRIBUTIONS

B.H.J., D.S.M., and T.A.R. conceived and designed the experiments. F.M. conducted transcriptome assembly and binning. B.H.J., D.S.M., and T.A.R. wrote the manuscript. B.H.J., D.S.M., and G.L. conducted experimental work and analysed the data. B.E.H., S.W. and J.D.E aided in conceptual and experimental design, and in conducting experimental work.

ACKNOWLEDGEMENTS

295 This work was primarily supported by an EMBO YIP award and a Royal Society University Research
296 Fellowship (UF130382) and latterly by an ERC Consolidator Grant (CELL-in-CELL) to T.A.R. S.W. and
297 J.D.E. were supported by awards from the Wellcome Trust (WT107791/Z/15/Z) and the Lister institute.
298 We thank Karen Moore and the University of Exeter Sequence Service for support with the various
299 sequencing projects. We thank Éric Meyer, Institut de Biologie de l'Ecole Normale Supérieure, Paris,
300 for advice during set up of the *P. bursaria* RNAi approach.

301 **DECLARATION OF INTERESTS**

302 The authors declare no competing interests.

303

304 **FIGURES**

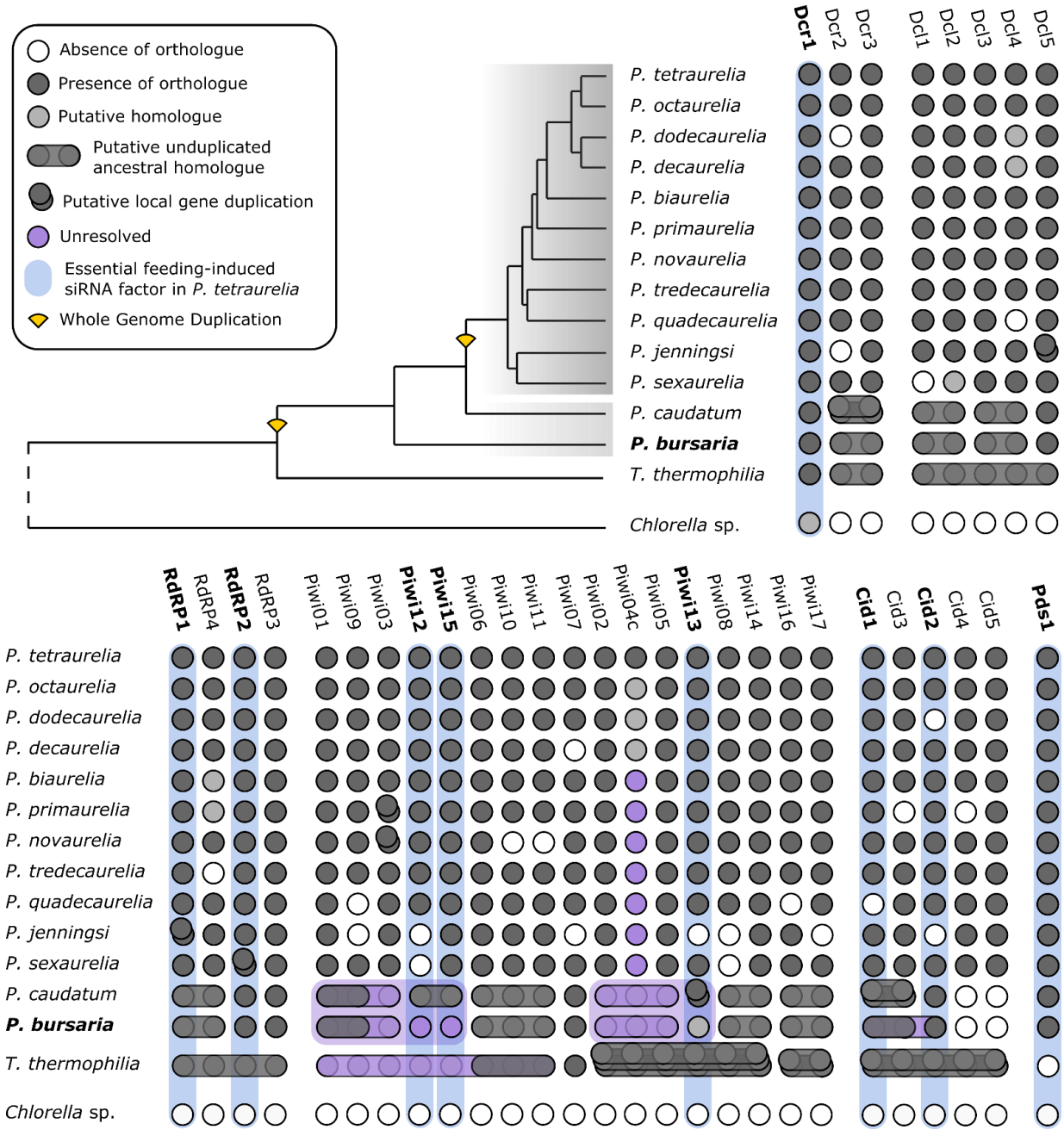


Figure 1. Identifying a Putative RNAi Pathway in *P. bursaria*. Coulson plot showing the presence / absence of putative RNAi pathway component-encoding genes, identified from *Paramecium* genome/transcriptome sequence surveys based on shared sequence identity. Genes highlighted in blue represent components essential for feeding-induced siRNA-based RNAi in *P. tetraurelia*. Genes highlighted in purple represent duplicated components with unclear paralogue/orthologue resolution. Stacked genes (single or unduplicated orthologues) represent putative local gene duplications. Phylogeny schematic based on *Dcr1* amino acid alignment (**Dataset S1**), with shaded regions indicating species hypothesised to share the same number of ancestral whole gene duplications (WGDs). For all phylogenies, see **Datasets S1-6**. Nucleotide sequence

(https://doi.org/10.6084/m9.figshare.13387811.v1) and amino acid alignment data (https://doi.org/10.6084/m9.figshare.13387631.v1) for putative *P. bursaria* homologues are available at Figshare.

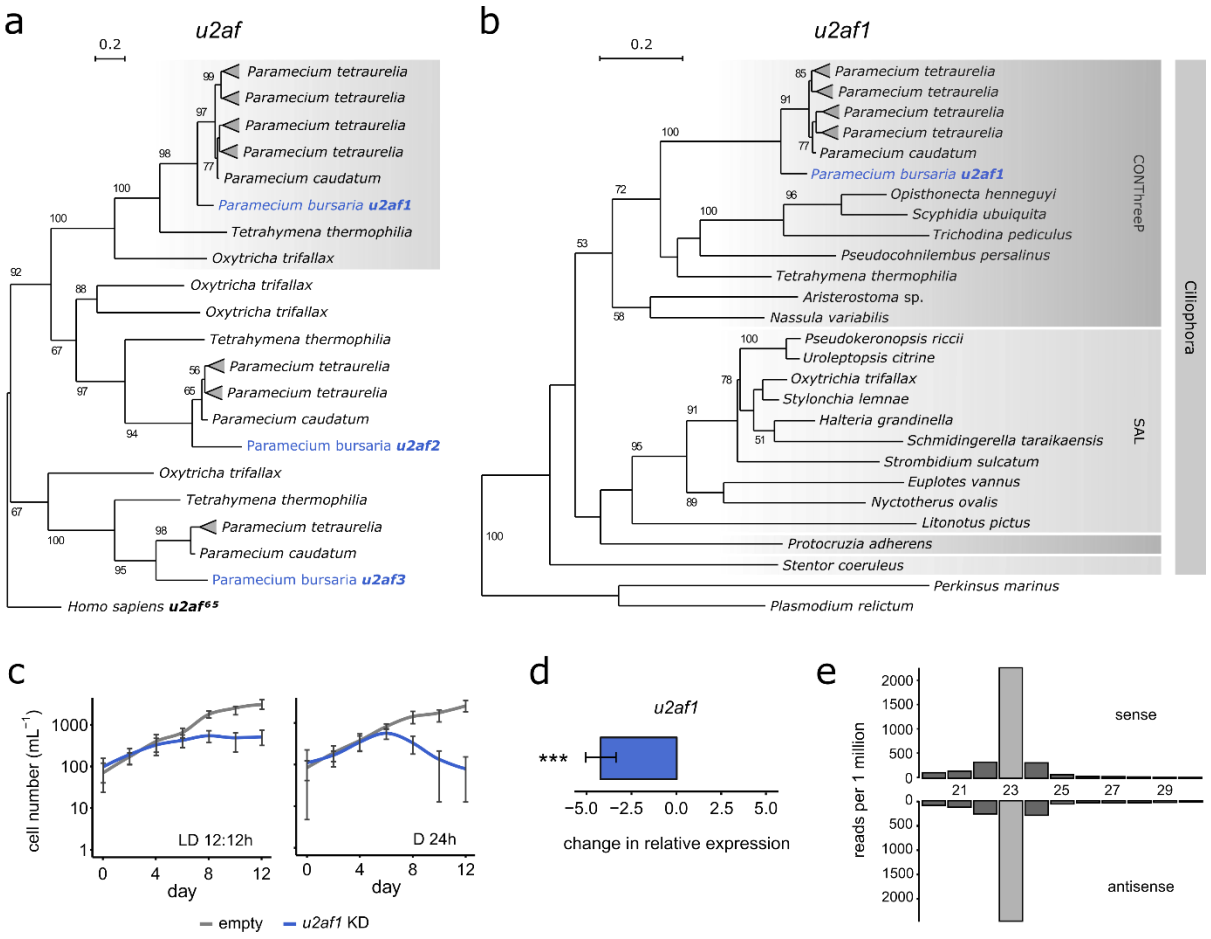


Figure 2. Validation of Feeding-Based RNAi in *P. bursaria* (**a**) *U2af* phylogeny (based on 166 sampled aligned amino acid sites) calculated using iqtree with an rtREV+G4 best fit substitution model chosen according to BIC, and with 1,000 non-rapid, non-parametric bootstrap replicates. This phylogeny highlights the three orthologues of *u2af* (65kDa) encoded by *Paramecium*. Note the shaded branch corresponding to the *u2af1* orthologue targeted in this study. (**b**) Specific *u2af1* phylogeny (based on 225 sampled aligned amino acid sites) calculated using iqtree with an LG+G4 best fit substitution model chosen according to BIC, and with 1,000 non-rapid, non-parametric bootstrap replicates. This phylogeny shows the distribution of *u2af1* across the ciliates. Ciliate clades CONThreeP (Colpodea, Oligohymenophorea, Nassophorea, Prostomatea, Plagiopylea and Phyllopharyngea) and SAL (Spirotrichea, Armaphorea and Listomatea) are defined according to Lynn[61,64] and Adl[61,65]. For all phylogenies, bootstrap values above 50 are shown. Amino acid alignment data for putative *P. bursaria* homologues used in the above datasets are available on Figshare (https://doi.org/10.6084/m9.figshare.13387631.v1). (**c**) *P. bursaria* cell number in cultures fed with HT115 *E. coli* expressing *u2af1* dsRNA (blue) or containing an empty vector control (grey). *P. bursaria* cells were resuspended daily into fresh feeding media for 12 days under standard light-dark (LD 12:12h) or constant darkness (D 24h) conditions. Note that the effect of *u2af1* dsRNA exposure was

more potent when feeding was conducted under constant darkness, giving rise to a mean cell number after 12 days that was 84.4% less compared to parallel cultures maintained under standard light-dark conditions (***; calculated as $p \leq 0.001$ using a generalized linear model with quasi-Poisson distribution). Data are represented as mean ± SD of five biological replicates. Here and elsewhere, the term 'KD' is used in figure to denote 'knock-down'. (**d**) qPCR of mRNA extracted from day 3 of *u2af1*-RNAi feeding, revealing potent gene knock-down in *P. bursaria* in response to *u2af1* dsRNA exposure. Change in relative expression (ddCT) was calculated for treated (*u2af1* dsRNA) vs untreated (empty vector) control cultures, and normalised against the standardised change in expression of a *GAPDH* housekeeping gene. Data are represented as mean ± SEM of three biological replicates, per treatment. Un-paired dCt values were pooled and averaged prior to calculation of ddCt. Error bars for ddCt values were propagated from the SEM of dCt values. Raw Ct, dCt and error propagation calculations for all ddCt values are available in **Table S3**. Significance for qPCR data calculated as ***$p \leq 0.001$ using a paired t-test. (**e**) Size distribution of sense and antisense sRNAs mapping to a 450-nt 'scramble' dsRNA construct, expressed via transformed *E. coli* fed to *P. bursaria* for three days prior to sRNA extraction and sequencing. Scramble dsRNA presented no significant hit to the identified *P. bursaria* host or endosymbiont transcriptome to ensure that the 23-nt sRNA detected was of definitive exogenous origin.

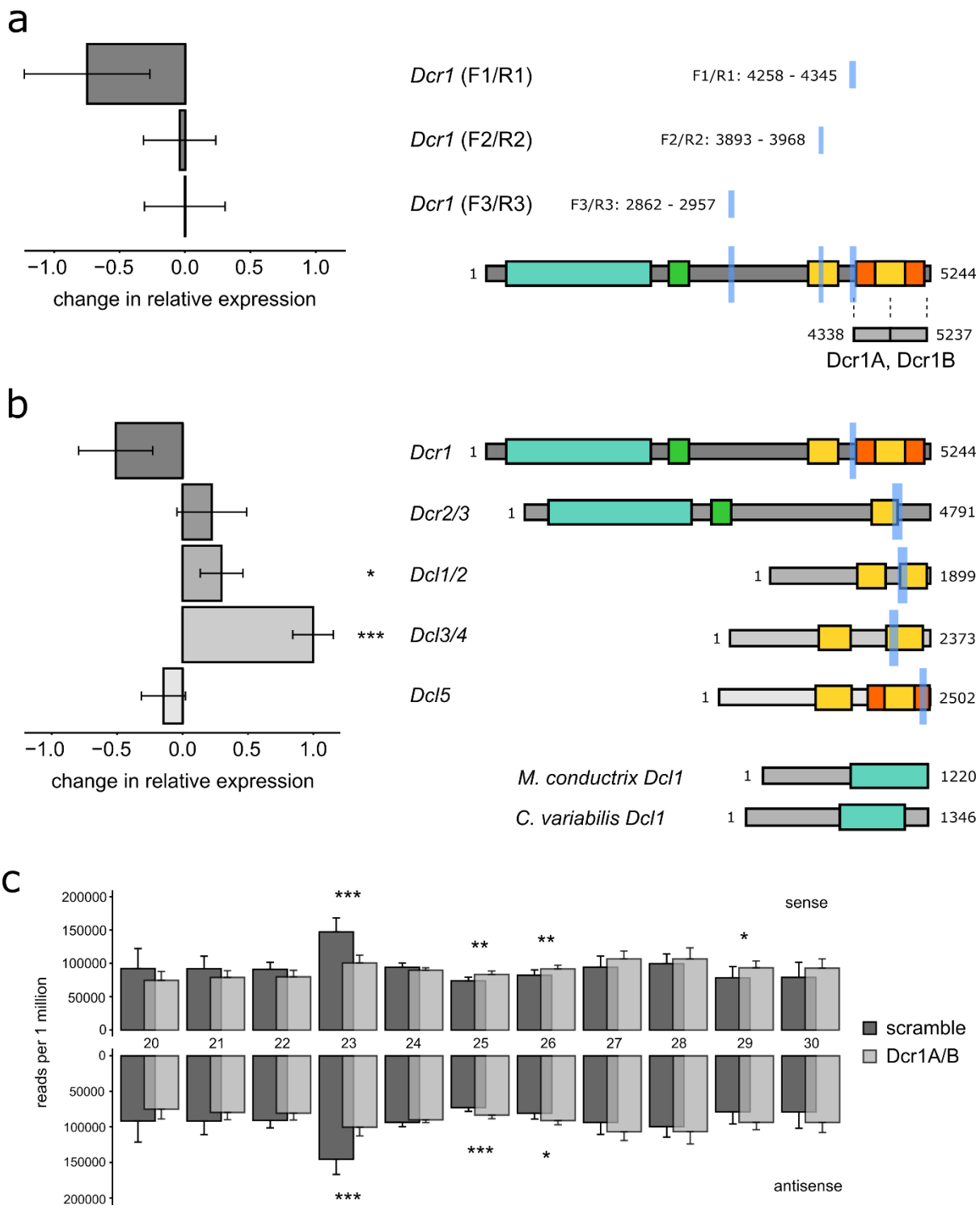
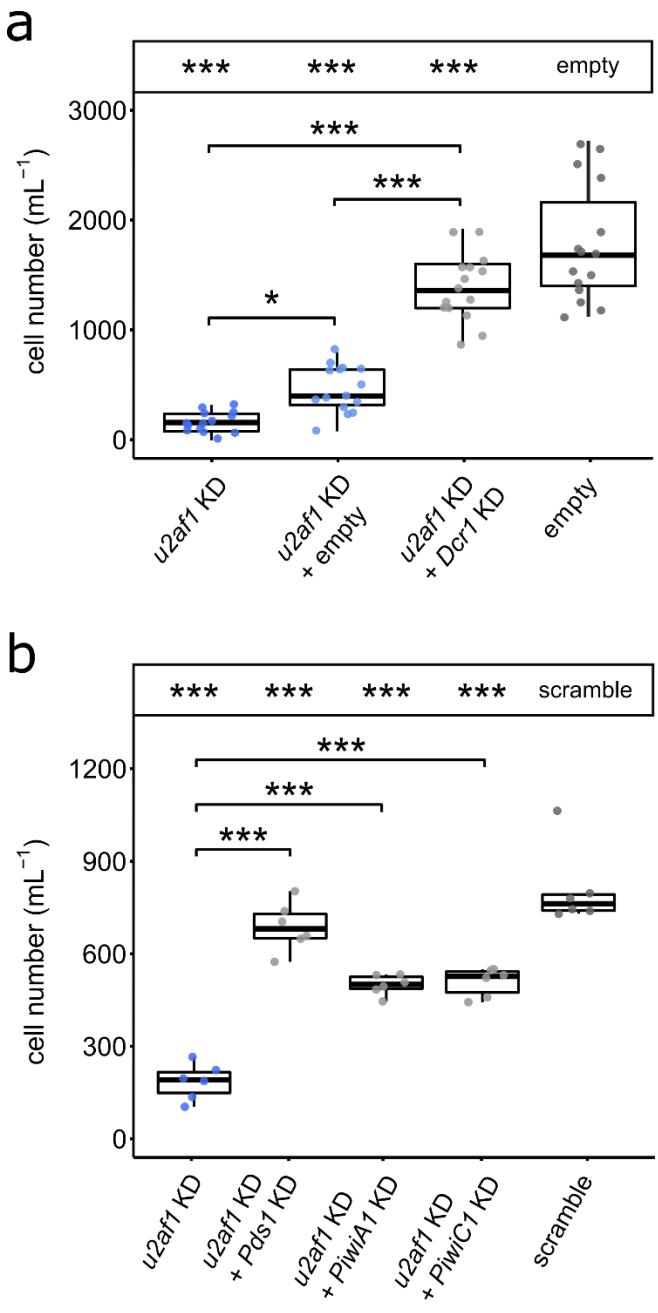


Figure 3. Investigating *Dcr1* Function (**a**) qPCR of mRNA extracted from day 3 of *Dicer*-RNAi feeding, revealing *Dcr1* gene knock-down in *P. bursaria* in response to Dcr1A and Dcr1B dsRNA exposure. A schematic representation of *P. bursaria Dcr1* shows the target sites of tandem 500-nt Dcr1A and Dcr1B dsRNA constructs, and amplicon sites of respective *Dcr1* F1/R1, F2/R2 and F3/R3 qPCR primers (light blue). Note the proximity of the amplicon site to the dsRNA target site, relative to the degree of knock-down detected via qPCR. (**b**) Additional qPCR of mRNA extracted from day 3 of *Dicer*-RNAi feeding.

Knock-down was not observed in *Dcr2/3*, *Dcl1/2* or *Dcl3/4*, and is inconclusive in *Dcl5*. A schematic representation of all *P. bursaria Dicer* or *Dicer*-like transcripts shows functional domain homology, and amplicon sites of respective qPCR primers (light blue) for each transcript. *M. conductrix Dcl1* demonstrates the divergence of *Dicer* homologues in the *P. bursaria* algal endosymbiont. For all qPCR data, change in relative expression (ddCt) was calculated for treated (Dcr1A+Dcr1B dsRNA) vs untreated (scramble dsRNA) control cultures, and normalised against the standardised change in expression of a *GAPDH* housekeeping gene. Data are represented as mean ± SEM of six un-paired biological replicates, per treatment. Un-paired dCt values were pooled and averaged prior to calculation of ddCt. Error bars for ddCt values were propagated from the SEM of dCt values. Raw Ct, dCt and error propagation calculations for all ddCt values are available in **Tables S4-5**. Significance for qPCR calculated as *p ≤ 0.05, ***p ≤ 0.001 using a paired t-test. For all schematic domains: turquoise = Helicase; green = Dicer dimer; yellow = RIBOc; orange = RNC. Amino acid alignment data for putative *P. bursaria* homologues used in the above datasets are available on Figshare (https://doi.org/10.6084/m9.figshare.13387631.v1). For phylogenetic analysis confirming the identity of these *Dicer* and *Dicer*-like components in *P. bursaria*, see **Datasets S1** & **S2.** (**c**) Overlaid size distribution of sense and antisense sRNAs mapping to *P. bursaria* host transcripts during exposure to Dcr1A and Dcr1B dsRNA (light grey), or a non-hit 'scramble' dsRNA control (dark grey). sRNA was sequenced from *P. bursaria* after 7, 8 and 9 days of *E. coli* vector-based RNAi feeding to deliver respective dsRNA. Note the reduction in 23-nt sense and antisense reads upon Dcr1A/Dcr1B dsRNA exposure, accompanied by an increase in ≥25-nt sense and antisense reads. Data are presented as mean ± SD of nine biological replicates (three per time point), and normalised to reads per 1 million 20-30 nt reads, per sample. Significance for sRNA abundance calculated as *p ≤ 0.05, **p ≤ 0.01, and ***p ≤ 0.001 by one-way analysis of variance (ANOVA).



386
387

388 **Figure 4. Validation of *Dcr1*, *Piwi* and *Pds1 Function*** (**a**) *P. bursaria* cell number after 10 days of
389 feeding with HT115 *E. coli* expressing either: *u2af1* dsRNA (dark blue); *u2af1* dsRNA mixed with empty
390 vector (light blue) or *Dcr1* dsRNA (grey); or an empty vector control (white). (**b**) *P. bursaria* cell number
391 after 12 days of feeding with HT115 *E. coli* expressing either: *u2af1* dsRNA (dark blue); *u2af1* dsRNA
392 mixed with *Pds1, PiwiA1,* or *PiwiC1* dsRNA (grey); or a 'scramble' control (dark grey). Multiple vector
393 delivery was conducted at a 50:50 ratio during feeding. Significance calculated as ***p ≤ 0.001 using
394 a generalized linear model with quasi-Poisson distribution. Boxplot data are represented as max,
395 upper quartile (Q3), mean, lower quartile (Q1) and min values of five biological replicates. Asterisks in
396 the box above each plot correspond to significance compared to empty vector, or 'scramble' dsRNA

controls, respectively. Individual data points are shown. Confirmation of the consistent effect of empty vector compared to 'scramble' dsRNA exposure are shown in **Figure S2**.

TABLES

Table 1. Full list of identified RNAi components and predicted associated pathways in *P. bursaria*. See also attached file.

***P. bursaria* homologue**	**Accession no.[1]**	**Corresponding *P. tetraurelia* homologue[2]**	**Associated pathway in *P. tetraurelia***	**Reference**
Dcr1	pending	*Dcr1*	Feeding-induced siRNA (23-nt), Transgene-induced siRNA (23-nt)	33
Dcr2/3	pending	*Dcr2*	-	-
		Dcr3	-	-
Dcl1/2	pending	*Dcl1*	-	35,41,42
		Dcl2	scnRNA (25-nt)	35,41,42
Dcl3/4	pending	*Dcl3*	scnRNA (25-nt)	35,41,42
		Dcl4	-	-
Dcl5	pending	*Dcl5*	iesRNA (~28-nt)	41,42
Pds1	pending	*Pds1*	Feeding-induced siRNA (23-nt)	33
Rdr1/4	pending	*RDR1*	Feeding-induced siRNA (23-nt)	33
		RDR4	-	-
Rdr2	pending	*RDR2*	Feeding-induced siRNA (23-nt), Transgene-induced siRNA (23-nt), Endogenous siRNA (23-nt)	32–34
Rdr3	pending	*RDR3*	Transgene-induced siRNA (23-nt), Endogenous siRNA (23-nt)	32–34
PiwiA1	pending	*Ptiwi01*	scnRNA (25-nt)	66
		Ptiwi09	scnRNA (25-nt)	66
		Ptiwi03	-	-
		Ptiwi12	Feeding-induced siRNA (23-nt)	33
		Ptiwi15	Feeding-induced siRNA (23-nt)	33
PiwiA2	pending	*Ptiwi06*	-	-
		Ptiwi10	-	-
		Ptiwi11	-	-
PiwiB	pending	*Ptiwi07*	-	-
PiwiC1	pending	*Ptiwi02*	-	-
		Ptiwi04c	-	-
		Ptiwi05	-	-
		Ptiwi13	Feeding-induced siRNA (23-nt), Transgene-induced siRNA (23-nt)	33

PiwiC2	pending	*Ptiwi08*	-	-
		Ptiwi14	Transgene-induced siRNA (23-nt)	[33]
PiwiD	pending	*Ptiwi16*	-	[33]
		Ptiwi17	-	[45]
Cid1/3	pending	*CID1*	Feeding-induced siRNA (23-nt)	[33]
		CID3	Transgene-induced siRNA (23-nt)	[33]
Cid2	pending	*CID2*	Feeding-induced siRNA (23-nt), Transgene-induced siRNA (23-nt)	[33]

[1] Pending NCBI submission, nucleotide (https://doi.org/10.6084/m9.figshare.13387811.v1) and amino acid (https://doi.org/10.6084/m9.figshare.13387631.v1) sequence data for *P. bursaria* RNAi component-encoding genes are available on Figshare

[2] For phylogenetic identification of Paramecium homologues, see **Datasets S1-6**

Table 2. Calculation of RNAi component-encoding gene retention rate across the *Paramecium aurelia* species complex. See also attached file.

***P. bursaria* homologue**	**Putative *P. tetraurelia* homologues[1]**	**Predicted copies in *P. aurelia* species complex**	**Actual copies in *P. aurelia* species complex[2]**	**Copy loss in *P. aurelia* species complex**	**Gene retention rate (%)[3]**
Dcr1	1	11	11	0	100.00
Dcr2/3	2	22	20	-2	90.91
Dcl1/2	2	22	21	-1	95.45
Dcl3/4	2	22	21	-1	95.45
Dcl5	1	11	11	0	100.00
Pds1	1	11	11	0	100.00
Rdr1/4	2	22	21	-1	95.45
Rdr2	1	11	11	0	100.00
Rdr3	1	11	11	0	100.00
PiwiA1	5	55	51	-4	92.73
PiwiA2	3	33	31	-2	93.94
PiwiB	1	11	9	-2	81.82
PiwiC1[4]	3	33	32	-1	96.97
PiwiC2	2	22	20	-2	90.91
PiwiD	2	22	20	-2	90.91
Cid1/3	2	22	20	-2	90.91
Cid2	1	11	9	-2	81.82

[1] For phylogenetic identification of Paramecium homologues, see **Datasets S1-6**

[2] Actual copy number taken from **Figure 1**

[3] Not including locally duplicated copies of *Dcl5*, *Rdr1*, *Rdr2* and *Piwi03*

[4] *Piwi04c* excluded due to poor phylogenetic resolution

METHODS

Culture conditions and media

In all RNAi experiments, *Paramecium bursaria* 186b (CCAP 1660/18) strain was used. For genome analysis, *P. bursaria* 186b strain was used. For transcriptome analysis, *P. bursaria* 186b, CCAP 1660/12 and Yad1g1N strains were used.

P. bursaria cells were cultured in New Cereal Leaf – Prescott Liquid media (NCL). NCL media was prepared by adding 4.3 mgL^{-1} $CaCl_2.2H_2O$, 1.6 mgL^{-1} KCl, 5.1 mgL^{-1} K_2HPO_4, 2.8 mgL^{-1} $MgSO_4.7H_2O$ to deionised water. 1 gL^{-1} wheat bran was added, and the solution boiled for 5 minutes. Once cooled, media was filtered once through Whatman Grade 1 filter paper and then through Whatman GF/C glass microfiber filter paper. Filtered NCL media was autoclaved at 121°C for 30 mins to sterilise prior to use.

NCL medium was bacterized with *Klebsiella pneumoniae* SMC and supplemented with 0.8 mgL^{-1} β-sitosterol prior to propagation. *P. bursaria* cells were sub-cultured 1:9 into fresh bacterized NCL media once per month, and maintained at 18°C with a light-dark (LD) cycle of 12:12h.

Transcriptome analysis

RNA was extracted from *P. bursaria* 186b for transcriptome analysis, using ~10^6 host cells from five replicates at three time points over the 12:12 hr LD cycle (6 hr L, 1.5 hr D, and 10.5 hr D). RNA extraction was performed using the RNA PowerSoil Total RNA Isolation Kit (MoBio) following the manufacturers' protocol. Samples were checked for quality using an Agilent TapeStation (High Sensitivity RNA ScreenTape) and a NanoDrop ND-1000, resulting in four low-quality samples which were discarded (RNA Integrity Number <1, NanoDrop concentration <15 $ngμL^{-1}$, or TapeStation <450 ng total). RNA for the remaining 11 samples (four 6 hr L, four 1.5 hr D, and three 10.5 hr D) was matched to 400 ng, and library preparation performed using the TruSeq Stranded Total RNA Kit (Illumina) following the manufacturers' protocol. Prepared libraries of 11 samples were then sequenced using a paired-end 120-bp rapid run across two lanes on an Illumina HiSeq 2500, yielding ~1,112 million reads (with a mean of 101 million reads per sample [S.E.M. of 2.9 million reads]). For

details of additional transcriptome sequence acquisition from *P. bursaria* CCAP 1660/12, which also included 'single-cell' transcriptome analyses, please refer to the **Supplementary Methods**.

Raw reads were trimmed at Q5 in Trimmomatic (v0.32)[67]. Reads were then error corrected using rcorrector (v1.0.0) and digitally normalised using Khmer v1.4.1[68]at a k-mer size of 20 and average coverage of 20. The remaining reads were then assembled using rnaSPAdes (v3.11.1)[69] and Trinity (v2.0.2)[70]. On the basis of RSEM (v1.2.24)[71] and assembly statistics, the Trinity assembly was selected for further analysis.

ORFs were called from Trinity assembled transcripts using Transdecoder, using both ciliate (*Tetrahymena*) and universal encodings. The longest peptide sequences were retained for each. The remaining ORFs were then annotated via a BLASTX (v2.2.31) search against a genome database consisting of: *Arabidopsis thaliana*, *Aspergillus nidulans*, *Bacillus cereus* ATCC 14579, *Burkholderia pseudomallei* K96243, *Candidatus Korarchaeum cryptofilum* OPF8, *Chlamydomonas reinhardtii*, *Chlorella variabilis* NC64A, *Chlorella vulgaris* C-169, *Escherichia coli* str. K-12 substr. MG1655, *Homo sapiens*, *Methanococcus maripaludis* S2, *Oxytricha trifallax*, *Paramecium biaurelia*, *Paramecium caudatum*, *Paramecium multimicronucleatum*, *Paramecium primaurelia*, *Paramecium sexaurelia*, *Paramecium tetraurelia*, *Saccharomyces cerevisiae* S288C, *Streptomyces coelicolor* A32, *Sulfolobus islandicus* M.14.25, *Tetrahymena borealis*, *Tetrahymena elliotti*, *Tetrahymena malaccensis*, *Tetrahymena thermophila* macronucleus, *Tetrahymena thermophila* micronucleus, and *Ustilago maydis*.

Assembled transcripts were subsequently binned into either 'host', 'endosymbiont', 'food' or 'other' datasets, using a phylogeny-based machine-learning approach (https://github.com/fmaguire/dendrogenous, see **Supplementary Methods**). Binned sequences were further annotated using SignalP (v4.0), TMHMM (v2.0), and BLAST2GO (v4). Each dataset was filtered to remove any sequences with a predicted peptide sequence shorter than 30 amino acids.

P. bursaria Yad1g1N[22] transcriptome reads were downloaded from DDBJ (Submission DRA000907), and processed using the same approach to assembly and binning as *P. bursaria* 186b dataset described above.

Phylogenetic analysis

Assembled datasets of ciliate encoded predicted proteins ('host' bin) and universally encoded predicted proteins ('endosymbiont' bin) were searched using BLASTp and a minimum expectation of 1e-05, to identify homologues of annotated protein sequences that are putatively encoded by both host and endosymbiont. Proteins predicted from genomic data were downloaded from ParameciumDB[72] for *Paramecium biaurelia, Paramecium caudatum, Paramecium decaurelia,*

Paramecium dodecaurelia, Paramecium jenningsi, Paramecium novaurelia, Paramecium octaurelia, Paramecium primaurelia, Paramecium quadecaurelia, Paramecium sexaurelia, *Paramecium tetraurelia* and *Paramecium tredecaurelia*. These were added to a curated dataset of genomic and transcriptomic data from a further 41 ciliate species[73] to assess for homologues throughout the ciliates. Identified homologues were checked against the NCBI non-redundant protein sequences (nr) database via reciprocal BLASTp search.

Protein sequences were aligned using MAFFT[74] (v7.471) and masked using TrimAL[75] (v1.4.rev15) allowing for no gaps. Sequences were manually checked in SeaView[76] (v5.0.4), and highly-divergent or identical sequences from the same genomic source were removed. Phylogenies were generated using IQ-TREE (v2.0.3) with 1,000 non-parametric non-rapid bootstraps, using the best fit substitution model calculated with IQ-TREE's inbuilt ModelFinder implementation and according to the Bayesian Inference Criterion (BIC). The models chosen for tree generation are listed in the respective figure legends.

Gene synthesis and construct design

Sequences for plasmid constructs were synthesised *de novo* by either Genscript or SynBio Technologies, and cloned into an L4440 plasmid vector (Addgene plasmid #1654). Sequences and cloning sites for each plasmid construct are detailed in **Table S1**. All modified constructs were confirmed by Sanger sequencing (Eurofins Genomics).

RNAi feeding

P. bursaria was fed with *E. coli* transformed with an L4440 plasmid construct with paired IPTG-inducible T7 promoters, facilitating targeted gene knock-down through the delivery of complementary double-stranded (ds)RNA. L4440 plasmid constructs were transformed into *E. coli* HT115 competent cells and grown overnight on LB agar (50 μgmL^{-1} Ampicillin and 12.5 μgmL^{-1} Tetracycline) at 37°C. Positive transformants were picked and grown overnight in LB (50 μgmL^{-1} Ampicillin and 12.5 μgmL^{-1} Tetracycline) at 37°C with shaking (180 rpm). Overnight pre-cultures were back-diluted 1:25 into 50 mL of LB (50 μgmL^{-1} Ampicillin and 12.5 μgmL^{-1} Tetracycline) and incubated for a further 2 hours under the same conditions, until an OD_{600} of between 0.4 and 0.6 was reached. *E. coli* cultures were then supplemented with 0.4 mM IPTG to induce template expression within the L4440 plasmid, and incubated for a further 3 hours under the same conditions. *E. coli* cells were pelleted by centrifugation (3100 x *g* for 2 mins), washed with sterile NCL media, and pelleted once more. *E. coli* cells were then re-suspended in NCL media supplemented with 0.4 mM IPTG, 100 μgmL^{-1} Ampicillin, and 0.8 μgmL^{-1} β-sitosterol, and adjusted to a final OD_{600} of 0.1.

P. bursaria cells were pelleted by gentle centrifugation in a 96-well plate (10 mins at 800 x *g*), taking care not to disturb the cell pellet by leaving 50 µl of supernatant, and re-suspended 1:4 into 200 µl of induced *E. coli* culture media (to make 250 µl total). Feeding was conducted daily for up to 14 days using freshly prepared bacterized media.

qPCR analysis

RNA was extracted from *P. bursaria* 186b for gene expression analysis after three days of RNAi feeding. *P. bursaria* cells (~10^3 per culture) were pelleted by gentle centrifugation (800 x *g* for 10 mins), snap-frozen in liquid nitrogen, and stored at -80°C. RNA extraction was performed using TRIzol reagent (Invitrogen), following the manufacturer's protocol after re-suspending each pellet in 900 µl TRIzol reagent. RNA was precipitated using GlycoBlue Co-precipitant (Invitrogen) to aid RNA pellet visualisation, and then cleared of residual DNA using the TURBO DNA-*free* Kit (Ambion), following the manufacturer's protocol for routine DNase treatment.

RNA was reverse transcribed into single stranded cDNA using the SuperScript® III First-Strand Synthesis SuperMix (Invitrogen), following the manufacturer's protocol. Quantitative PCR (qPCR) was performed in a StepOnePlus Real-Time PCR system (Thermo Fisher Scientific). Reaction conditions were optimised using a gradient PCR, with a standard curve determined using 10-fold dilutions of *P. bursaria* cDNA: *u2af1* (slope: -3.525; R^2: 0.994; efficiency: 92.157%), *dcr1* (slope: -3.400; R^2: 0.998; efficiency: 96.862%), *dcr2/3* (slope: -3.395; R^2: 0.996; efficiency: 97.050%), *dcl1/2* (slope: -3.494; R^2: 0.999; efficiency: 93.281%), *dcl3/4* (slope: -3.280; R^2: 0.999; efficiency: 101.767%), *dcl5* (slope: -3.411; R^2: 0.999; efficiency: 96.416%), and *GAPDH* (slope: -3.427; R^2: 1.000; efficiency: 95.802%), using StepOne software v2.3. Each 20 µL reaction contained 10 µL PowerUp SYBR Green Master Mix (Thermo Fisher Scientific), 500 mM each primer and 1 µL (50 ng) cDNA. Primers pairs for each reaction are listed in **Table S2**. Each reaction was performed in duplicate for each of 3 biological replicates, alongside a 'no-RT' (i.e. non-reverse transcribed RNA) control to detect any genomic DNA contamination. Cycling conditions were as follows: UDG activation, 2 mins at 50°C and DNA polymerase activation, 2 mins at 95°C, followed by 40 cycles of 15 secs, 95°C and 1 min at 55-65°C (*u2af1* (57°C), *dcr1* (60°C), *dcr2/3* (60°C), *dcl1/2* (60°C), *dcl3/4* (60°C), *dcl5* (60°C), and *GAPDH* (60°C)). Each reaction was followed by melt-curve analysis, with a 60-95°C temperature gradient (0.3°C s^{-1}), ensuring the presence of only a single amplicon, and ROX was used as a reference dye for calculation of C_T values. C_T values were then used to calculate the change in gene expression of the target gene in RNAi samples relative to control samples, using a derivation of the $2^{-\Delta\Delta CT}$ algorithm[77].

sRNA isolation and sequencing

Total RNA for sRNA sequencing was extracted from *P. bursaria* (or free-living algal) cultures using TRIzol reagent (Invitrogen), as detailed above. To isolate sRNA from total RNA, samples were size separated on a denaturing 15% TBE-Urea polyacrylamide gel. Gels were prepared with a 15 mL mix with final concentrations of 15% Acrylamide/Bis (19:1), 8M Urea, TBE (89 mM Tris, 89 mM Borate, 2 mM EDTA), and the polymerisation started by the addition of 150 µL 10% APS (Sigma-Aldrich) and 20 µL TEMED (Sigma-Aldrich). Gels were pre-equilibrated by running for 15 mins (200 V, 30 mA) in TBE before RNA loading. The ladder mix consisted of 500 ng ssRNA ladder (50-1000nt, NEB#N0364S), and 5-10 ng of each 21 & 26-nt RNA oligo loaded per lane. The marker and samples were mixed with 2X RNA loading dye (NEB) and heat denatured at 90 °C for 3 mins before snap cooling on ice for 2 min prior to loading. Blank lanes were left between samples/replicates to prevent cross-contamination during band excision. Gels were then run for 50 mins (200V, 30 mA).

Once run, gels were stained by shaking (60 rpm) for 20 mins at RT in a 40 mL TBE solution containing 4 µL SYBR® Gold Nucleic Acid Gel Stain. Bands of the desired size range (~15-30 nt) were visualised under blue light, excised and placed into a 0.5 mL tube pierced at the bottom by a 21-gauge needle, resting within a 1.5 mL tube, and centrifuged (16,000 x *g* for 1 min). 400 µL of RNA elution buffer (1M Sodium acetate pH 5.5 and 1mM EDTA) was added to the 1.5 mL tube containing centrifuged gel slurry, and the empty 0.5 mL tube discarded. Gel slurry was manually homogenized until dissolved using a 1 mL sterile plunger and incubated at RT for 2 hours with shaking at 1,400 rpm.

Solutions containing RNA elution buffer and gel slurry were transferred to a Costar Spin-X 0.22 µm filter column and centrifuged (16,000 x *g* for 1 min). The filter insert containing acrylamide was discarded. 1 mL of 100% EtOH was added to each solution, alongside 15 µg of GlycoBlue™ Coprecipitant (Invitrogen) to aid sRNA pellet visualisation, and stored overnight at -80°C to precipitate. Precipitated solutions were centrifuged at 4°C (12,000 x *g* for 30 mins), and the supernatant discarded. sRNA pellets were washed with 500 µL of cold 70% EtOH (12,000 x *g* for 15 mins at 4°C), and air dried in a sterile PCR hood for 10 mins, before re-suspending in 15 µL of RNAse-free water and storage at -80°C.

sRNA-seq and read processing

sRNA concentrations were determined using an Agilent 2100 Bioanalyzer, following the Agilent Small RNA kit protocol, and all samples matched to 0.7 $ngmL^{-1}$ prior to sequencing. Library preparation and subsequent RNA-seq was performed for 54 samples using 50-bp paired-end, rapid run across four lanes on an Illumina HiSeq 2500, yielding ~120-150 million paired-end reads per lane (~9-11 million paired-end reads per sample).

The raw paired-end reads from the RNA-seq libraries were trimmed using Trim Galore in order to remove barcodes (4-nt from each 3'- and 5'- end) and sRNA adaptors, with additional settings of a phred-score quality threshold of 20 and minimum length of 16-nt. Result were subsequently checked with FastQC.

Trimmed reads were mapped against the 'host' or 'endosymbiont' dataset of assembled transcripts using the HISAT2 alignment program with default settings. Post-mapping, the BAM files were processed using SAMTOOLS and a set of custom scripts (https://github.com/guyleonard/paramecium) to produce a table of mapped read accessions and their respective read lengths. Size distributions of sRNA abundance for each sample were plotted using the R programming language packages; tidyverse, grid.extra and ggplot2 in R Studio (v.1.3.1073).

DATA AND SOFTWARE AVAILABILITY

The raw reads generated during transcriptome and sRNA sequencing are available on the NCBI Sequence Read Archive (accessions: SAMN14932981, SAMN14932982). All other datasets are available on Figshare (https://doi.org/10.6084/m9.figshare.c.5241983.v1), under the relevant headings. Custom scripts for host and endosymbiont transcript binning (https://github.com/fmaguire/dendrogenous) and sRNA read processing (https://github.com/guyleonard/paramecium) are available on GitHub.

REFERENCES

1. Archibald, J. M. Endosymbiosis and eukaryotic cell evolution. *Curr. Biol.* **25**, R911-921 (2015).
2. Keeling, P. J. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annual Review of Plant Biology* **64**, 583–607 (2013).
3. Howe, C. J., Barbrook, A. C., Nisbet, R. E. R., Lockhart, P. J. & Larkum, A. W. D. The origin of plastids. *Philos Trans R Soc Lond B Biol Sci* **363**, 2675–2685 (2008).
4. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135 (2004).
5. Esteban, G. F., Fenchel, T. & Finlay, B. J. Mixotrophy in ciliates. *Protist* **161**, 621–641 (2010).
6. Johnson, M. D. Acquired phototrophy in ciliates: a review of cellular interactions and structural adaptations. *J Eukaryot Microbiol* **58**, 185–95 (2011).
7. Pröschold, T., Darienko, T., Silva, P. C., Reisser, W. & Krienitz, L. The systematics of Zoochlorella revisited employing an integrative approach. *Environmental Microbiology* **13**, 350–364 (2011).

8. Fujishima, M. & Kodama, Y. Endosymbionts in *Paramecium*. *Eur J Protistol* **48**, 124–37 (2012).
9. Siegel, R. W. Hereditary endosymbiosis in *Paramecium bursaria*. *Exp Cell Res* **19**, 239–52 (1960).
10. Achilles-Day, U. E. & Day, J. G. Isolation of clonal cultures of endosymbiotic green algae from their ciliate hosts. *J Microbiol Methods* **92**, 355–7 (2013).
11. Hoshina, R. & Kusuoka, Y. DNA analysis of algal endosymbionts of ciliates reveals the state of algal integration and the surprising specificity of the symbiosis. *Protist* **167**, 174–184 (2016).
12. Zagata, P., Greczek-Stachura, M., Tarcz, S. & Rautian, M. The evolutionary relationships between endosymbiotic green algae of *Paramecium bursaria* syngens originating from different geographical locations. *Folia Biol (Krakow)* **64**, 47–54 (2016).
13. Brown, J. A. & Nielsen, P. J. Transfer of photosynthetically produced carbohydrate from endosymbiotic *Chlorellae* to *Paramecium bursaria*. *J. Protozool.* **21**, 569–570 (1974).
14. Kato, Y. & Imamura, N. Amino acid transport systems of Japanese *Paramecium* symbiont F36-ZK. *Symbiosis* **47**, 99–107 (2009).
15. Kato, Y. & Imamura, N. Effect of sugars on amino acid transport by symbiotic *Chlorella*. *Plant Physiol Biochem* **46**, 911–7 (2008).
16. Kawakami, H. & Kawakami, N. Behavior of a virus in a symbiotic system, *Paramecium bursaria—Zoochlorella*. *The Journal of Protozoology* **25**, 217–225 (1978).
17. Summerer, M., Sonntag, B., Hörtnagl, P. & Sommaruga, R. Symbiotic ciliates receive protection against UV damage from their algae: a test with *Paramecium bursaria* and *Chlorella*. *Protist* **160**, 233–243 (2009).
18. Parker, R. C. Symbiosis in *Paramecium bursaria*. *Journal of Experimental Zoology* **46**, 1–12 (1926).
19. Ziesenisz, E., Reisser, W. & Wiessner, W. Evidence of de novo synthesis of maltose excreted by the endosymbiotic *Chlorella* from *Paramecium bursaria*. *Planta* **153**, 481–485 (1981).
20. Kodama, Y. & Fujishima, M. Cycloheximide induces synchronous swelling of perialgal vacuoles enclosing symbiotic *Chlorella vulgaris* and digestion of the algae in the ciliate *Paramecium bursaria*. *Protist* **159**, 483–94 (2008).
21. Tanaka, M. *et al.* Complete elimination of endosymbiotic algae from *Paramecium bursaria* and its confirmation by diagnostic PCR. *Acta Protozool* **41**, 255–261 (2002).
22. Kodama, Y. & Fujishima, M. Symbiotic *Chlorella variabilis* incubated under constant dark conditions for 24 hours loses the ability to avoid digestion by host lysosomal enzymes in digestive vacuoles of host ciliate *Paramecium bursaria*. *FEMS Microbiol. Ecol.* **90**, 946–955 (2014).

634 23. Omura, G. *et al.* A bacteria-free monoxenic culture of *Paramecium bursaria*: its growth
635 characteristics and the re-establishment of symbiosis with *Chlorell*a in bacteria-free conditions.
636 *Jpn J Protozool* **37**, (2004).
637 24. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis*
638 *elegans*. *Nature* **391**, 806–11 (1998).
639 25. Timmons, L., Court, D. L. & Fire, A. Ingestion of bacterially expressed dsRNAs can produce
640 specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**, 103–112 (2001).
641 26. Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in
642 developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
643 27. Carmell, M. A. & Hannon, G. J. RNase III enzymes and the initiation of gene silencing. *Nat. Struct.*
644 *Mol. Biol.* **11**, 214–218 (2004).
645 28. Peters, L. & Meister, G. Argonaute proteins: mediators of RNA silencing. *Mol Cell* **26**, 611–23
646 (2007).
647 29. Cerutti, H. & Casas-Mollano, J. A. On the origin and functions of RNA-mediated silencing: from
648 protists to man. *Curr Genet* **50**, 81–99 (2006).
649 30. Shabalina, S. A. & Koonin, E. V. Origins and evolution of eukaryotic RNA interference. *Trends*
650 *Ecol. Evol. (Amst.)* **23**, 578–587 (2008).
651 31. Ender, C. & Meister, G. Argonaute proteins at a glance. *J Cell Sci* **123**, 1819–23 (2010).
652 32. Karunanithi, S. *et al.* Exogenous RNAi mechanisms contribute to transcriptome adaptation by
653 phased siRNA clusters in *Paramecium*. *Nucleic Acids Res* **47**, 8036–8049 (2019).
654 33. Marker, S., Carradec, Q., Tanty, V., Arnaiz, O. & Meyer, E. A forward genetic screen reveals
655 essential and non-essential RNAi factors in *Paramecium tetraurelia*. *Nucleic Acids Res* **42**, 7268–
656 80 (2014).
657 34. Marker, S., Le Mouël, A., Meyer, E. & Simon, M. Distinct RNA-dependent RNA polymerases are
658 required for RNAi triggered by double-stranded RNA versus truncated transgenes in
659 *Paramecium tetraurelia*. *Nucleic Acids Res* **38**, 4092–107 (2010).
660 35. Lepère, G. *et al.* Silencing-associated and meiosis-specific small RNA pathways in *Paramecium*
661 *tetraurelia*. *Nucleic Acids Res* **37**, 903–915 (2009).
662 36. Carradec, Q. *et al.* Primary and secondary siRNA synthesis triggered by RNAs from food bacteria
663 in the ciliate *Paramecium tetraurelia*. *Nucleic Acids Res* **43**, 1818–33 (2015).
664 37. Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K. & Tabara, H. In vitro analyses of the production
665 and activity of secondary small interfering RNAs in *C. elegans*. *EMBO J* **26**, 5007–19 (2007).
666 38. Aury, J. M. *et al.* Global trends of whole-genome duplications revealed by the ciliate
667 *Paramecium tetraurelia*. *Nature* **444**, 171–8 (2006).

668 39. McGrath, C. L., Gout, J.-F., Doak, T. G., Yanagi, A. & Lynch, M. Insights into three whole-genome
669 duplications gleaned from the *Paramecium caudatum* genome sequence. *Genetics* **197**, 1417–
670 1428 (2014).
671 40. Gout, J.-F. *et al.* Universal trends of post-duplication evolution revealed by the genomes of 13
672 *Paramecium* species sharing an ancestral whole-genome duplication.
673 http://biorxiv.org/lookup/doi/10.1101/573576 (2019) doi:10.1101/573576.
674 41. Hoehener, C., Hug, I. & Nowacki, M. Dicer-like enzymes with sequence cleavage preferences.
675 *Cell* **173**, 234-247.e7 (2018).
676 42. Sandoval, P. Y., Swart, E. C., Arambasic, M. & Nowacki, M. Functional diversification of Dicer-like
677 proteins and small RNAs required for genome sculpting. *Dev Cell* **28**, 174–88 (2014).
678 43. Malone, C. D., Anderson, A. M., Motl, J. A., Rexer, C. H. & Chalker, D. L. Germ line transcripts are
679 processed by a Dicer-like protein that is essential for developmentally programmed genome
680 rearrangements of *Tetrahymena thermophila*. *Mol Cell Biol* **25**, 9151–9164 (2005).
681 44. Nekrasova, I. V., Владимировна, Н. И., Potekhin, A. A. & Анатольевич, П. А. Diversity of RNA
682 interference pathways in regulation of endogenous and exogenous sequences expression in
683 ciliates *Tetrahymena* and *Paramecium*. *Ecological genetics* **17**, 113–125 (2019).
684 45. Drews, F. *et al.* Two Piwis with Ago-like functions silence somatic genes at the chromatin level.
685 *bioRxiv* 2020.08.24.263970 (2020) doi:10.1101/2020.08.24.263970.
686 46. Chalker, D. L., Fuller, P. & Yao, M.-C. Communication between parental and developing genomes
687 during *Tetrahymena* nuclear differentiation is likely mediated by homologous RNAs. *Genetics*
688 **169**, 149–160 (2005).
689 47. Mochizuki, K. & Gorovsky, M. A. Conjugation-specific small RNAs in *Tetrahymena* have predicted
690 properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev.* **18**, 2068–2073
691 (2004).
692 48. Galvani, A. & Sperling, L. RNA interference by feeding in *Paramecium*. *Trends Genet* **18**, 11–2
693 (2002).
694 49. Paschka, A. G. *et al.* The use of RNAi to analyze gene function in spirotrichous ciliates. *European*
695 *Journal of Protistology* **39**, 449–454 (2003).
696 50. Slabodnick, M. M. *et al.* The kinase regulator Mob1 acts as a patterning protein for *Stentor*
697 morphogenesis. *PLoS Biol* **12**, (2014).
698 51. Sobierajska, K., Joachimiak, E., Bregier, C., Fabczak, S. & Fabczak, H. Effect of phosducin silencing
699 on the photokinetic motile response of *Blepharisma japonicum*. *Photochem Photobiol Sci* **10**,
700 19–24 (2011).

52. Carthew, R. W. & Sontheimer, E. J. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).
53. He, M. *et al.* Genetic basis for the establishment of endosymbiosis in *Paramecium*. *ISME J* **13**, 1360–1369 (2019).
54. Cerutti, H., Ma, X., Msanne, J. & Repas, T. RNA-mediated silencing in algae: biological roles and tools for analysis of gene function ▿. *Eukaryot Cell* **10**, 1164–1172 (2011).
55. Shao, C. *et al.* Mechanisms for U2AF to define 3′ splice sites and regulate alternative splicing in the human genome. *Nat Struct Mol Biol* **21**, 997–1005 (2014).
56. Cho, S. *et al.* Splicing inhibition of U2AF65 leads to alternative exon skipping. *PNAS* **112**, 9926–9931 (2015).
57. Saudemont, B. *et al.* The fitness cost of mis-splicing is the main determinant of alternative splicing patterns. *Genome Biology* **18**, 208 (2017).
58. Miura, T., Moriya, H. & Iwai, S. Assessing phagotrophy in the mixotrophic ciliate *Paramecium bursaria* using GFP-expressing yeast cells. *FEMS Microbiol Lett* **364**, (2017).
59. Ohno, S. Evolution by Gene Duplication. (Springer-Verlag, 1970). doi:10.1007/978-3-642-86659-3.
60. Adl, S. M. *et al.* The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal of Eukaryotic Microbiology* **52**, 399–451 (2005).
61. Gao, F. *et al.* The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). *Scientific Reports* **6**, 24874 (2016).
62. Beisson, J. *et al.* DNA microinjection into the macronucleus of *Paramecium*. *Cold Spring Harb Protoc* **2010**, pdb.prot5364 (2010).
63. Jahn, C. L. & Klobutcher, L. A. Genome remodeling in ciliated protozoa. *Annu Rev Microbiol* **56**, 489–520 (2002).
64. Lynn, D. The ciliated protozoa: characterization, classification, and guide to the literature. (Springer Netherlands, 2008). doi:10.1007/978-1-4020-8239-9.
65. Adl, S. M. *et al.* The revised classification of eukaryotes. *J Eukaryot Microbiol* **59**, 429–493 (2012).
66. Bouhouche, K., Gout, J.-F., Kapusta, A., Bétermier, M. & Meyer, E. Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res.* **39**, 4249–4264 (2011).
67. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

734 68. Crusoe, M. R. *et al.* The khmer software package: enabling efficient nucleotide sequence
735 analysis. *F1000Res* **4**, 900 (2015).
736 69. Bushmanova, E., Antipov, D., Lapidus, A. & Prjibelski, A. D. rnaSPAdes: a de novo transcriptome
737 assembler and its application to RNA-Seq data. *GigaScience* **8**, (2019).
738 70. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity
739 platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512 (2013).
740 71. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or
741 without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
742 72. Arnaiz, O., Meyer, E. & Sperling, L. ParameciumDB 2019: integrating genomic data across the
743 genus for functional and evolutionary biology. *Nucleic Acids Research* gkz948 (2019)
744 doi:10.1093/nar/gkz948.
745 73. Irwin, N. A. T. *et al.* The function and evolution of motile DNA replication systems in ciliates.
746 *Current Biology* (2020) doi:10.1016/j.cub.2020.09.077.
747 74. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence
748 alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
749 75. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment
750 trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
751 76. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user interface
752 for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).
753 77. Zhang, J. D., Biczok, R. & Ruschhaupt, M. ddCt: The ddCt algorithm for the analysis of
754 quantitative real-time PCR (qRT-PCR). (Bioconductor version: Release (3.11), 2020).
755 doi:10.18129/B9.bioc.ddCt.
756