Emergent RNA-RNA interactions can promote stability in a nascent phototrophic endosymbiosis

Benjamin H. Jenkins[1,2]*, Finlay Maguire[3], Guy Leonard[1,2], Joshua D. Eaton[1], Steve 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

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

KEYWORDS

Paramecium bursaria, algae, mutualism, symbiosis, RNAi, Dicer, RDR, siRNA

ABSTRACT

Eukaryote-eukaryote endosymbiosis was responsible for the spread of chloroplast (plastid) organelles. Stability is required for the metabolic and genetic integration that drives the establishment of new organelles, yet the mechanisms which act to stabilise nascent endosymbioses – between two fundamentally selfish biological organisms – are unclear. Theory suggests that enforcement mechanisms, which punish misbehaviour, may act to stabilise such interactions by resolving conflict. However, how such mechanisms can emerge in a nascent endosymbiosis has yet to be explored. Here, we propose that endosymbiont-host RNA-RNA interactions, arising from digestion of the endosymbiont population, can result in a cost to host growth for breakdown of the endosymbiosis. Using the model nascent endosymbiosis, *Paramecium bursaria* – *Chlorella* spp., we demonstrate that this mechanism is dependent on the host RNA-interference (RNAi) system. We reveal through small RNA (sRNA) sequencing that endosymbiont-derived mRNA released upon endosymbiont digestion can be processed by the host RNAi system into 23-nt sRNA. We predict multiple regions of shared sequence identity between endosymbiont and host mRNA, and demonstrate through delivery of synthetic endosymbiont sRNA that exposure to these regions can knock-down expression of complementary host genes, resulting in a cost to host growth. This process of host gene knock-down in response to endosymbiont-derived RNA processing by host RNAi factors, which we term 'RNAi-collisions', represents a

mechanism which can promote stability in a nascent eukaryote-eukaryote endosymbiosis. By imposing a cost for breakdown of the endosymbiosis, endosymbiont-host RNA-RNA interactions may drive maintenance of the symbiosis across fluctuating ecological conditions and symbiotic status.

SIGNIFICANCE STATEMENT

Stable endosymbiosis between eukaryotic microbes has driven the evolution of further cellular complexity. Yet the mechanisms which can act to stabilise a nascent eukaryote-eukaryote endosymbiosis are unclear. Using the model nascent endosymbiotic system, *Paramecium bursaria–Chlorella*, we demonstrate that endosymbiont-host RNA-RNA interactions can drive a cost to host growth upon endosymbiont digestion, punishing the host for misbehaviour. These RNA-RNA interactions are facilitated by the host RNA-interference system. For endosymbiont mRNA sharing a high-level of sequence identity with host transcripts, this process can result in host gene knock-down. We propose that these endosymbiont-host RNA-RNA interactions–'RNAi collisions'–represent a viable enforcement mechanism to sanction the host for breakdown of the endosymbiosis, promoting the stability of a nascent endosymbiotic interaction.

MAIN TEXT

A nascent endosymbiosis between eukaryotes was a pre-requisite for the spread of photosynthetic organelles, such as plastids[1–5]. This transition from transiently engulfed cells to obligate organelles is driven by metabolic and genetic integration. Yet, to become manifest, a stable intermediary state must exist upon which this process of integration can proceed[1,3,5–8]. Conflict is an inevitable outcome of all symbioses, the resolution of which can significantly impact the stability of an interaction. Enforcement mechanisms which punish misbehaviour can act to stabilise symbioses[9,10], yet we know little about how these can emerge in a nascent endosymbiotic interaction. *Paramecium bursaria* – a ciliate protist which harbours a clonal population of intracellular green algae, *Chlorella* spp.[11–13] – represents a tractable model system to study emergent mechanisms in a nascent endosymbiosis[14]. The interaction is facultative[11,14–17] and based on two-way metabolic exchange[18–25]. Endosymbiotic algae are housed within modified host phagosomes called perialgal vacuoles, which may be fused with host lysosomes to trigger digestion[26] allowing *P. bursaria* to maintain control over the interaction in the event of conflict[27–29]. However, it is

unclear how this endosymbiotic system is protected from over-exploitation by the host which would ultimately lead the interaction to collapse[30–35].

RNA-RNA interactions can play a role in host-pathogen symbiotic systems[36–40], whereby RNA can 'hi-jack' the small interfering RNA (siRNA) pathway of the symbiotic partner to modulate expression of genes involved in virulence or alternatively resistance. Whether analogous RNA-RNA interactions could occur in an endosymbiotic system has yet to be elucidated. The presence of a functional siRNA pathway in *Paramecium*, and its role in RNA-interference (RNAi), has been validated as a tool for gene silencing in both *Paramecium bursaria*[41] and non-photo-endosymbiotic congener, *Paramecium tetraurelia*[42–46]. RNAi can be initiated through the provision of bacterial food transformed to express double-stranded RNA (dsRNA) with high sequence similarity to a target transcript[41,43]. *Paramecium* can also process single-stranded RNA (ssRNA) – including ribosomal RNA (rRNA) and messenger RNA (mRNA) – derived from a prokaryotic cell acquired through phagotrophy[42]. This processing is mediated by conserved RNAi protein components which also function in endogenous transcriptome regulation[41,42,44,47]. Significantly, these studies propose that mRNA, derived both exogenously[41,42] and endogenously[41,44], can act as substrates for siRNA generation in *Paramecium*.

We demonstrate that RNA released upon digestion of the algal endosymbiont is processed by the host RNAi system in *P. bursaria*. For endosymbiont-derived mRNA sharing a high-level of sequence identity with host transcripts, this processing may interfere with endogenous host gene expression resulting in a cost to host growth (**Fig. S1**). We track the interaction through sRNA sequencing, recapitulate the effect through exposure to synthetic endosymbiont RNA, and demonstrate that this mechanism is mediated by host Dicer, Piwi, Pds1 and RdRP proteins. This process of host gene knock-down in response to endosymbiont-derived RNA processing by host RNAi factors, which we term 'RNAi-collisions', represents a candidate enforcement mechanism which can promote stability in a nascent eukaryote-eukaryote endosymbiosis. By imposing a cost for breakdown of the endosymbiosis, endosymbiont-host RNA-RNA interactions may drive maintenance of the symbiosis across fluctuating ecological conditions and symbiotic status.

Endosymbiont digestion in P. bursaria results in an RNAi-mediated 'physiological cost' to the host

P. bursaria can be purged of endosymbiotic algae via treatment with the ribosomal translational inhibitor, cycloheximide[14]. A comparison of ribosomal protein (RP) L29A predicted protein sequences, the active site of cycloheximide function, confirmed that *Paramecium* possess a specific nucleotide polymorphism identified as a determinant of cycloheximide resistance in other species (**Fig. S2**). This substitution was absent in all green algal species assessed, including known algal endosymbionts of *P. bursaria*. Upon treatment with cycloheximide, a significant reduction in algal-chlorophyll fluorescent intensity per host cell was observed after two-to-three days (**Fig. 1A**). A clear de-coupling of host *P. bursaria* cell number and algal fluorescence was also observed, consistent with translational inhibition in the algae but not the host (**Fig. S2**), demonstrating that loss of a photosynthetic endosymbiont population does not immediately result in a decline in host cell number. Additional staining with LysoTracker Green to identify acidic vacuoles, including lysosomes, revealed that this loss of algal fluorescence was linked to increased lysosomal activity in the host cytoplasmic environment (**Fig. 1B**). These data support that elimination of the endosymbiotic algal population during cycloheximide treatment[14,26] is triggered by host digestion.

Continued treatment with cycloheximide resulted in a significant retardation to *P. bursaria* culture growth, however this same effect was not observed in the non-photo-endosymbiotic congener species, *P. tetraurelia* (**Fig. 1C**). Both *Paramecium* species harbour the same cycloheximide resistance conferring substitution in RPL29A (**Fig. S2**), suggesting that elimination of endosymbiotic algae is costly to *P. bursaria* culture growth. Loss of algal derived photosynthate and plastid derived metabolites represents an obvious cost. However, we sought to explore whether a part of this cost may be attributed to host exposure to endosymbiont-derived RNA during digestion of the endosymbiont population, and whether this cost was mediated by the host RNAi system.

Knock-down of *Dcr1* (a host-encoded endoribonuclease Dicer required for siRNA generation) through complementary dsRNA exposure significantly rescued the cost to *P. bursaria* culture growth associated with cycloheximide treatment (**Fig. 1D** and Fig. S3). This was consistent with *Dcr1* (Dicer) knock-down in a prior study attenuating the effect of *E. coli* vector-based RNAi feeding in *P. bursaria*[41]. In the aforementioned study, knock-down of host-encoded *PiwiA1, PiwiC1* (AGO-Piwi effectors required for targeted RNA cleavage) and

Pds1 (a *Paramecium*-specific component with an unknown but essential role in exogenously induced RNAi) also attenuated an *E. coli*-vector feeding-based RNAi effect[41]. Here, knock-down of host-encoded *PiwiA1* and *Pds1* similarly rescued the cost to *P. bursaria* culture growth associated with cycloheximide treatment (**Fig. 1E** and Fig. S3). Involvement of *Pds1* confirms that the observed RNAi effect is mediated by the host, as no identifiable homologue of *Pds1* could be identified in the green algal genomes and transcriptomes sampled[41,48]. Furthermore, it excludes the possibility that off-target effects arising from host Dicer knock-down, including potential compensatory function of additional Dicer or Dicer-like paralogues in *P. bursaria*[41], could be responsible for the RNAi effect observed. The near complete rescue observed upon partial knock-down of Dicer suggests that host Dicer processing, rather than loss of algal derived metabolites, is a more significant factor relating to growth retardation under these conditions. Taken together, these data suggest that the physiological cost to *P. bursaria* growth incurred during cycloheximide treatment, in which the endosymbiotic algae are being broken down and digested by the host, is mediated by host-encoded RNAi components. These data support the occurrence of RNAi-mediated RNA-RNA interactions between endosymbiont and host.

Endosymbiont breakdown triggers an abundance of Dicer-dependent endosymbiont derived sRNA within *P. bursaria*

To investigate the occurrence of RNA-RNA interactions between endosymbiont and host, we tracked the abundance of endosymbiont-derived, host-processed sRNA in *P. bursaria* during endosymbiont digestion. Disruption of host RNAi was achieved through partial knock-down of Dicer, allowing us to directly test for an increase in endosymbiotic algal-derived 23-nt sRNA (the size associated with host Dicer processing[41,42,49]) resulting from cycloheximide induced endosymbiont breakdown. Upon treatment with cycloheximide, we identified an increased abundance in all 21-29 nt reads mapping to endosymbiont-derived mRNA over two-to-three days (**Fig. 2A** and **Fig. S4**). This same trend was observed for reads mapping to algal endosymbiont rRNA-derived sRNA (**Fig. S5**). Reads were mapped with 100% identity, allowing no mismatches. Any reads that additionally mapped to the host with 100% identity were removed, to ensure that the subset of sRNA detected was of definitive algal origin. A significant increase in algal mRNA-derived 23-nt sense and antisense sRNA demonstrates a greater abundance of potential RNAi substrates during endosymbiont digestion[41,42,49].

Partial knock-down of host Dicer during cycloheximide treatment significantly ablated this endosymbiont mRNA-derived 23-nt antisense abundance after three days of cycloheximide treatment (**Fig. 2B** and **Fig. S4**). This suggests that an increase in endosymbiont mRNA-derived 23-nt antisense sRNA during endosymbiont digestion is dependent on host Dicer function. This is consistent with knock-down of host Dicer specifically reducing 23-nt sRNA abundance in *Paramecium*[41,42,49], and the observation that *Paramecium* RNAi factors are capable of processing both endogenously and exogenously derived mRNA[42,44]. It is important to note that this effect likely under-represents the full extent of host Dicer processing, due to the requirement for a paradoxical and therefore incomplete Dicer perturbation through Dicer-dependent RNAi-based knock-down[41]. Furthermore, these data represent only a subset of the sRNA potentially present due to the stringent mapping approach used to identify endosymbiont-derived sRNA, as any reads that additionally mapped to host RNA template with 100% sequence identity in either orientation were excluded from this analysis. While these reads with high shared sequence identity are the most important subset of sRNA for the identification of putative mRNA-mRNA interactions between endosymbiont and host, their exclusion here has allowed us to identify host RNAi processing of definitively endosymbiont-derived transcripts. Interestingly, one of the host-derived transcripts that mapped with 100% identity to an excluded endosymbiont-derived 23-nt sRNA was *P. bursaria* heat shock protein 90 (*HSP90*). An investigation of regions of high shared sequence identity between endosymbiont and host is explored below.

Next, we conducted a series of control observations. Firstly, we assessed whether cycloheximide treatment was altering general host-derived sRNA production, despite the inferred host resistance discussed above (**Fig. S2**). Reads were mapped to a dataset of 20 host transcripts that contained no potential 23-nt overlap with any identified algal transcripts (allowing for ≤2-nt mismatches), to ensure that these host transcripts were unaffected by an increased rate of putative RNA-RNA interactions derived from the endosymbiont (**Fig. S6**). No significant increase in 23-nt abundance was observed for reads mapping to this subset of host mRNA transcripts during cycloheximide treatment, consistent with host resistance to cycloheximide not altering the host-derived population of sRNAs with low sequence identity to algal mRNA. Secondly, we assessed whether endosymbiotic

algal strains cultured under free-living conditions would also generate an abundance of 23-nt sRNA upon cycloheximide treatment (**Fig. S7**). Importantly, no clear increase in 21-25 nt algal-derived sRNA abundance was observed during algal treatment with cycloheximide when grown outside of the host cytoplasmic environment, nor was a 23-nt peak in algal derived sRNAs evident. This is consistent with data from a previous study which found that the algal endosymbiont of *P. bursaria* was not actively generating sRNA >20-nt[41]. These results suggest that increased algal endosymbiont-derived 23-nt sRNA abundance upon treatment with cycloheximide is dependent on host sRNA processing within the endosymbiotic system. Taken together, these data indicate that an abundance of endosymbiont-derived RNAs are released during endosymbiont digestion, which then act as substrate for processing by the host RNAi system, resulting in 23-nt sRNAs. Importantly, processing of endosymbiont-derived sRNA in a host Dicer-dependent manner indicates that these interactions are occurring in the host cytoplasm, supporting the hypothesis that endosymbiont-host mRNA-mRNA interactions are possible.

Comparison of transcriptome data reveals the potential for host transcript interaction by endosymbiont-derived RNAs

To understand the extent of possible RNA-RNA interactions between endosymbiont and host, we built a bioinformatic transcriptome processing tool, eDicer (https://github.com/fmaguire/eDicer), which allows identification of all possible Dicer-generated sense and antisense oligonucleotides produced from a given transcriptome dataset. By mapping 23-nt reads to a second ('host') dataset allowing for ≤2-nt mismatches, potential RNA-RNA interactions between an input (endosymbiont, vector, or food) and a host RNA population can be identified. Inclusion of reads with ≤2-nt mismatches were based on the tolerance for mismatching complementarity reported during RNAi-mediated knock-down of gene expression in multiple systems[50–54]. Confirmation of a similar mismatch tolerance in *P. bursaria* is shown in **Fig. 3E** & **Fig. S12**. For a full overview of the eDicer comparative analysis process, please refer to the **Supplementary Methods**.

Using a dataset consisting of transcripts binned as either 'endosymbiont' or 'host' (using a curated *P. bursaria* transcriptome[55]), we identified 35,703 distinct 23-nt putative mRNA-mRNA interactions between the *P. bursaria* 'host' and 'endosymbiont' RNA populations, representing 0.121% of the total inventory of distinct host 23-nt k-mers

identified (**Fig. S8A**; see also **Table S2** & **Supplementary Methods**). This was found to be 120-fold greater than the number of putative 23-nt mRNA-mRNA interactions predicted between bacterial food and host transcripts (from two different bacterial sources). Furthermore, the ratio of total:'potentially-lethal' putative 23-nt mRNA-mRNA interactions (determined by cross-referencing genes known to be conditionally essential in *Saccharomyces cerevisiae*[56]) was found to be greater between endosymbiont-and-host transcripts (1:1.4) than between bacterial food-and-host transcripts (1:0.16/0.17). Similar patterns were observed for predicted rRNA-rRNA interactions (**Fig. S8B,** see also **Table S2** & **Supplementary Methods**). These *in silico* analyses demonstrate that there is far greater potential for the occurrence of both mRNA-mRNA and rRNA-rRNA interactions between the algal endosymbiont and host (eukaryote-eukaryote) RNA populations, than there is between bacterial food and host (prokaryote-eukaryote) RNA populations.

In vivo exposure to a synthetic endosymbiont mRNA-derived chimera generates knock-down of high-identity host transcripts

Having identified the occurrence of putative endosymbiont-host RNA-RNA interactions *in silico*, we assessed whether exposure to synthetic fragments of algal mRNA could recapitulate the cost to host growth observed during endosymbiont digestion (**Figure 1E**). We identified ten endosymbiont mRNA interaction fragments that shared >91% (or 21-nt) sequence identity with host transcripts across a 23-nt region, selected at random from the analysis reported in **Fig. S8A**. Each mRNA interaction fragment was chosen to contain at least 1 SNP specific to the endosymbiont to ensure that any identifiable effect could be attributed to an endosymbiont-like transcript, rather than the host version. These interaction fragments were predicted to still be effective RNAi templates for *P. bursaria* based on the tolerance for mis-matching complementarity observed in RNAi-mediated knock-down in other systems[50–54] (and confirmed below in *P. bursaria*; **Fig. 3E** & **Fig. S12**). Six mRNA interaction fragments showed putative homology to 'non-lethal' yeast genes (including *EF1-α* and *HSP90*); two showed putative homology to 'lethal' yeast genes (including *tub-β*); and the remaining two had no identifiable homologues in yeast (**Table S1**). Significantly, *HSP90* was the host transcript previously identified as a candidate for putative endosymbiont-host RNA-RNA interaction in the sRNA analysis of **Fig. 2**. However, to ensure that any identifiable effect could be attributed to the endosymbiont-derived transcript, the

newly identified interaction fragment was chosen from a different region of this gene containing 2 SNPs specific to the endosymbiont. All ten mRNA interaction fragments were composed of the predicted interacting 23-nt sRNA sequence (**Fig. S8A**) flanked by 11-nt of contiguous 'non-hit' endosymbiont transcript as a filler on each side (**Table S3**). These ten 45-nt fragments were combined into a single 450-nt synthetic endosymbiotic algal chimera (**Fig. 3A**), cloned into an L4440 plasmid, and transformed into *E. coli* for feeding-induced RNAi.

Exposure to endosymbiotic algal chimera dsRNA resulted in significant retardation of *P. bursaria* culture growth (**Fig. 3B**). Once more, this effect was attenuated by knock-down of host Dicer, demonstrating an RNAi-mediated response to synthetic endosymbiont-host mRNA-mRNA interactions that resembled the host response to endosymbiont digestion in **Fig. 1C-E**. It is important to note that synthetic exposure to endosymbiont-derived RNA via an *E. coli* feeding vector would likely over-represent these putative RNA interactions. However, to address this issue, we designed a non-hit 'nonsense' control composed of a tandem assembly of the 11-nt regions of contiguous 'non-hit' algal transcript present in the chimera. A relative dilution of endosymbiont chimera dsRNA delivery alongside the 'nonsense' control (1, 1:1 and 1:3) also resulted in *P. bursaria* culture growth retardation (**Fig. S9**), demonstrating that a reduced relative delivery of synthetic endosymbiont RNA can also result in a cost to host growth. Similar experiments were conducted to investigate the possibility of endosymbiont-host rRNA-rRNA interactions, but showed that there was no equivalent effect on host growth, suggesting that ribosomal RNA is shielded from such effects (**Fig. S10**, see also **Table S4**).

To identify which putative mRNA-mRNA interactions were capable of facilitating a cost to *P. bursaria* culture growth, we identified the ten 450-nt host transcripts predicted to be hit by each endosymbiont derived 23-nt mRNA interaction fragment used in the chimera (**Figure 3C** - three examples are shown). Individual exposure to dsRNA corresponding to each individual but longer *P. bursaria* transcript-form identified three which resulted in a significant cost to host growth, relative to Dicer knock-down controls (**Fig. 3D**; see **Fig. S11** for all 10). These *P. bursaria* transcripts correspond to elongation factor-1α (*EF-1α*), heat shock protein 90 (*HSP90*; the host transcript identified as a candidate for putative endosymbiont-host RNA-RNA interaction in the sRNA analysis of **Fig. 2**), and tubulin-β chain

(*tub-β*). It can therefore be inferred that these are the 23-nt mRNA interaction fragments (**Fig. 3A/C**) that likely resulted in a cost to host growth observed during synthetic endosymbiont chimera dsRNA exposure (**Fig. 3B**). Using mRNA extracted from *P. bursaria* during chimera-RNAi feeding described above (**Fig. 3B**), qPCR revealed a reduction in host transcript expression of *EF-1α* and *tub-β* in response to endosymbiont chimera dsRNA exposure (**Fig. 3E**). Host transcript expression of *HSP90* appears inconclusive, however, expression of all three host transcripts was partially rescued upon knock-down of host Dicer (**Fig. S12**). From these data we can therefore conclude that 23-nt of >91% complementary 'endosymbiont' transcript is sufficient to facilitate detectable knock-down of a corresponding host transcript, demonstrating that exposure to endosymbiont-derived RNA originating from phagosomes is capable of impacting host gene expression via host RNAi knock-down.

It is important to consider that delivery of a synthetic endosymbiont-derived RNA chimera represents only an approximation of the putative 23-nt RNA-RNA interactions occurring in *P. bursaria*. Exposure to endosymbiont RNA via an *E. coli* feeding vector would likely over-represent these putative interactions. However, the random generation of 23-nt Dicer substrates across the dsRNA chimera (only three regions of which resulted in a detectable cost to host growth; **Fig. 3A/D** & **Fig. S11**) represents only part of the wider *E. coli*-derived RNA population processed by the host RNAi system during chimera-RNAi feeding. Together with dilution of endosymbiont chimera dsRNA delivery (**Fig. S9**), these observations support the hypothesis that only a relatively small number of 23-nt RNA-RNA interactions can induce a cost to host growth in *P. bursaria.* Furthermore, these data confirm that sRNAs with ≤2-nt mismatches are effective templates for RNAi-mediated knock-down of gene expression in *P. bursaria*, suggesting that the sRNA analysis in **Fig. 2** may under-represent the true scale of endosymbiont-derived sRNAs that can result in RNA-RNA interactions between endosymbiont and host. Importantly, we note that the occurrence of putative endosymbiont-host RNA-RNA interactions does not exclude a wider cost to host growth arising from processing a large population of algal derived RNAs with no defined host target, a process which must pose a cost to host cellular economics and transcriptional control/fidelity. For further justification of this 'synthetic' approach, and a consideration of the results that we can reliably draw from these data, see **Discussion S1**.

Single-stranded delivery of synthetic endosymbiont-derived RNA, analogous to endosymbiont mRNA, results in a cost to *P. bursaria* growth

As RNA derived naturally from the endosymbiont was unlikely to be double-stranded, two further constructs were designed to assess the efficacy of synthetic endosymbiont chimera single-stranded RNA (ssRNA) exposure in *P. bursaria* (**Fig. 4** & **Fig. S13**). Exposure to endosymbiont chimera ssRNA also resulted in a significant cost to *P. bursaria* culture growth, however this effect was only observed when delivered in the sense orientation ([+]ssRNA). Notably, chimera [+]ssRNA exposure resulted in a greater cost to host growth than chimera dsRNA exposure. A similar orientation bias was observed upon Dicer knock-down during cycloheximide treatment in a prior experiment, in which delivery of Dicer [+]ssRNA rescued culture growth to a greater extent than delivery of Dicer [-]ssRNA or Dicer dsRNA (**Fig. S14**). Importantly, the orientation of [+]ssRNA represents the same orientation as the mRNA transcripts from which each of these target templates were identified.

Knock-down of host Dicer (dsRNA delivery) was able to attenuate the cost to host growth associated with endosymbiont chimera [+]ssRNA exposure (**Fig. 4**). Importantly, simultaneous knock-down of host *Rdr1* and *Rdr2* (RNA-dependent RNA polymerases involved in amplification of primary or secondary sRNA triggers during RNAi[42,46]) also significantly rescued this cost to host growth. This is consistent with the role of RdRP proteins in processing exogenous sRNA, partially degraded mRNA cleavage products, or full-length mRNA transcripts in *Paramecium*[42,44–46]. As for Dicer, knock-down of host RdRP through RNAi represents a potential paradox, and so constitutes only partial or transient loss of function[41]. However, we are again confident that disruption in this manner would be sufficient to perturb RdRP-mediated RNAi in *P. bursaria* after 14 days, as seen in **Fig. 4**. While the mechanism of down-stream Dicer processing of RdRP-generated dsRNA substrates remains unknown (see **Discussion S1**), we can nonetheless infer from these data that single-stranded RNA-induced RNAi knock-down in *P. bursaria* is partially dependent upon host RdRP function.

Irrespective of the mechanistic basis for a sense oriented ssRNA knock-down bias, it is important to note that delivery of chimera [+]ssRNA demonstrated here represents the same orientation as the endosymbiont derived mRNA interaction fragments from which these putative endosymbiont-host RNA-RNA interactions were identified. This is consistent

with the observation that *Paramecium* RNAi factors are capable of processing both endogenously and exogenously derived mRNA[42,44]. The ability to facilitate a cost to host growth through delivery of single-stranded RNA, in the same orientation as the host target mRNA, therefore supports the hypothesis that a cost to host growth upon endosymbiont digestion can be facilitated by host RNAi-mediated RNA-RNA interactions between endosymbiont and host mRNA populations.

DISCUSSION

Through manipulation of the *P. bursaria* – *Chlorella* spp. endosymbiotic system, we have demonstrated that RNA released upon digestion of the algal endosymbiont is processed by the host RNAi system. For endosymbiont-derived mRNA sharing a high-level of sequence identity with host transcripts, this processing can result in knock-down of endogenous host gene expression, resulting in a cost to host growth (**Fig. S1**). We therefore postulate that these RNA-RNA interactions are of importance in eukaryote-eukaryote endosymbioses where partners are likely to share greater sequence similarity, especially among conserved transcripts which tend to be more highly represented among lethal or conditionally essential genes[56]. Due to the inherent difficulty in characterising such mechanisms directly, we have relied on a multiple experimental approach to demonstrate the viability of putative RNA-RNA interactions at each stage of the process. We have tracked the interaction through sRNA sequencing, recapitulated the effect through exposure to synthetic endosymbiont-derived RNA – including sense ssRNA analogous to endosymbiont mRNA – and demonstrated that this mechanism is mediated by host Dicer, Piwi, Pds1 and RdRP proteins. This process of host gene knock-down in response to endosymbiont-derived RNA processing by host RNAi factors, which we term 'RNAi-collisions', represents a candidate enforcement mechanism to sanction the host for breakdown of the interaction, a factor that would promote stability in a nascent eukaryote-eukaryote endosymbiosis.

The long-term maintenance of symbiotic interactions represents a quandary for evolutionary theory[30–35]. How do such relationships avoid over-exploitation by one partner that would ultimately lead the interaction to collapse? Partner switching is one option, however the result is typically a reduced pattern of co-evolution between symbiont and host that can inhibit the process of metabolic and genetic integration, stalling the evolution of stable interactions which are needed if a system is to move towards evolution of an

organelle[13,57–60]. Previous studies have suggested that when selfish behaviours arise, the evolution of enforcement to punish or suppress the exploitative partner can act to restore cooperation[9,10]. Enforcement mechanisms have been identified in diverse biological systems[61,62], and are argued to be one of the most effective drivers of cooperation in egalitarian alliances between different species, such as that between *P. bursaria* and its algal endosymbiont[9]. The results presented here suggest that 'RNAi collisions' between endosymbiont and host, which are capable of imposing a cost to host growth for breakdown of the symbiosis, could provide an emergent enforcement mechanism to discourage over-exploitation of the endosymbiont population by the host.

Interestingly, the emergence of this mechanism appears to be a by-product of pre-existing biological features that are already likely to be under strong selective pressure. For instance, the widely functional RNAi system of the host[42,43,45], or the conserved gene repertoire and sequence composition of the host and endosymbiont transcriptomes from which some 'RNAi collisions' have here been identified (including many transcripts which are conditionally essential in other systems[56]). Unlike comparable mechanisms of RNA-RNA interactions that have been studied in host-pathogen symbioses[36–40], these 'RNAi collisions' appear to be untargeted and, hence, emergent. In the aforementioned host-pathogen systems, targeted RNA is passed from one partner to the other in order to modulate expression of transcripts involved in virulence or resistance[36,39,40]. However, in order for such systems to evolve, they must first exist in an untargeted form upon which selection is able to act, allowing the emergence of specific RNA factors[37,38]. Identification of undirected 'RNAi collisions' in the *P. bursaria* – *Chlorella* system represents one such intermediary state, emergent in nature and untargeted, upon which sustained cellular interaction coupled with the potential for host-symbiont conflict could drive the selection of targeted RNA-RNA interactions.

We therefore propose that 'RNAi collisions' represent a putative mechanism to discourage over-exploitation of the endosymbiont population by the host. Here we have used the example of mass endosymbiont digestion in response to drug treatment to simulate this effect in the extreme. In natural interactions between *P. bursaria* and its algal endosymbiont, such a cost would only need to occur in the drastic occurrence of mass endosymbiont digestion in order to drive stability of the interaction. Importantly, the

403 endosymbiotic algal population within *P. bursaria* is largely composed of closely related or
404 clonal lineages[11–13], and as such, the fate of the algal population should be considered as a
405 collective unit. This cost need only act to supress large-scale, rapid destruction by the host in
406 order to drive the maintenance of a surviving subsection of the endosymbiont population.
407 Previous studies have demonstrated how *P. bursaria* is capable of manipulating
408 endosymbiont load in response to varying light conditions to better suit its own ends[27,28],
409 however, in these examples, reduction of endosymbiont number through digestion is slow
410 and partial. By providing a system that selects against rapid and near-complete digestion of
411 the endosymbiont population, 'RNAi collisions' effectively buffer the nascent endosymbiotic
412 interaction against total breakdown. We suggest that this has allowed the relationship to be
413 maintained across time and varying ecological conditions, even in the event of host-
414 symbiont conflict[27,28] and fluctuating endosymbiont numbers[14,26,27]. As an alternate route to
415 conflict resolution that avoids partner switching, we propose that such a mechanism would
416 facilitate greater co-evolution between endosymbiont and host. Over time, this would allow
417 the metabolic and genetic integration that drives the formation of obligate symbioses to
418 become manifest. We therefore present 'RNAi collisions' as a new mechanism in this
419 endosymbiotic system – a factor which can promote stability in the face of conflict in an
420 emergent endosymbiotic eukaryote-eukaryote cell-cell interaction.

421 **AUTHOR CONTRIBUTIONS**

422 B.H.J., D.S.M., and T.A.R. conceived and designed the experiments. F.M. and T.A.R.
423 conceived and designed the eDicer computational analysis. B.H.J., D.S.M., and T.A.R. wrote
424 the manuscript. B.H.J., D.S.M., F.M. and G.L. conducted experimental work and analysed the
425 data. B.E.H., S.W. and J.D.E aided in conceptual and experimental design, and in conducting
426 experimental work.

427 **ACKNOWLEDGEMENTS**

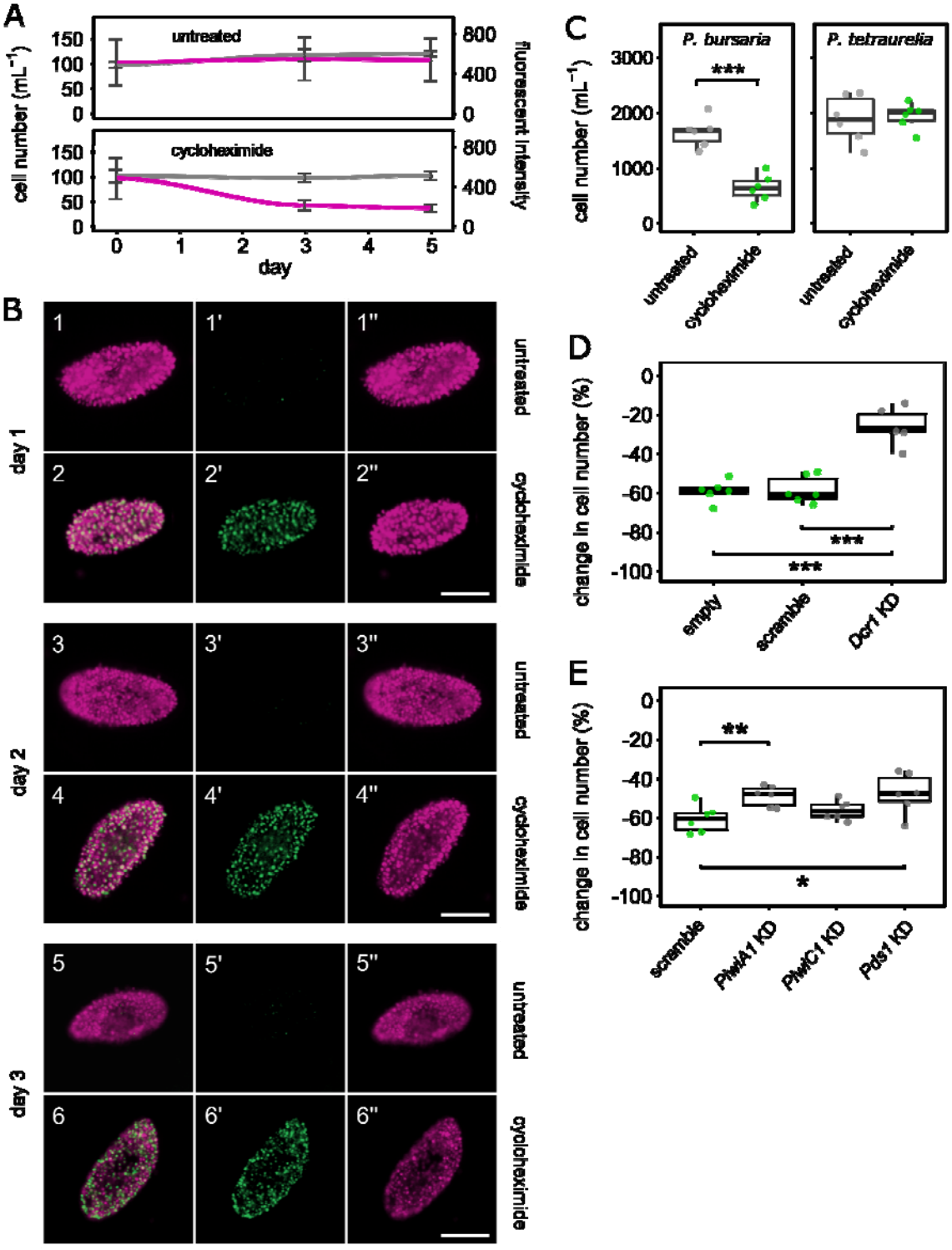
428 This work was primarily supported by an EMBO YIP award and a Royal Society University
429 Research Fellowship (UF130382) and latterly by an ERC Consolidator Grant (CELL-in-CELL) to
430 T.A.R. S.W. and J.D.E. were supported by awards from the Wellcome Trust
431 (WT107791/Z/15/Z) and the Lister institute. F.M. was supported by a Donald Hill Family
432 Fellowship in Computer Science. We thank Karen Moore and the University of Exeter
433 Sequence Service for support with the various sequencing projects. We thank Éric Meyer,

Institut de Biologie de l'Ecole Normale Supérieure, Paris, for advice during set up of the *P. bursaria* RNAi approach.

DECLARATION OF INTERESTS

The authors declare no competing interests.

438 **FIGURES**



439

440

Figure 1. Endosymbiont digestion in *P. bursaria* results in an RNAi-mediated 'physiological cost' to the host. (**A**) *Paramecium* cell number (grey) vs algal chlorophyll fluorescent intensity (pink) in stationary phase *P. bursaria* cultures treated with cycloheximide (50 μgmL^{-1}), compared to untreated controls. Loss of algal fluorescent intensity indicates algal death in response to cycloheximide treatment. Data are represented as mean ± SD of six biological replicates. (**B**) Stages of endosymbiont elimination in *P. bursaria* cells over three days of cycloheximide treatment (50 μgmL^{-1}). Algal chlorophyll fluorescence (pink) highlights endosymbiotic algae within the *P. bursaria* cell. LysoTracker Green fluorescence (green) indicates increased host lysosomal activity in response to cycloheximide treatment. Scale bar – 50 µm. (**C**) *Paramecium* cell number in *P. bursaria* or *P. tetraurelia* cultures after 8 days of treatment with cycloheximide (50 μgmL^{-1}; green), compared to untreated controls (light grey). *Paramecium* cultures were fed with *E. coli* transformed with an empty RNAi vector. (**D**) Percentage change in *P. bursaria* cell number in cultures treated with cycloheximide (50 μgmL^{-1}), compared to untreated controls, after 12 days of feeding with *E. coli* expressing; Dicer (*Dcr1*) dsRNA (dark grey) to induce knock-down (KD), non-hit 'scramble' dsRNA or an empty vector control (green). The relative effect of Dicer dsRNA exposure indicates partial rescue of *P. bursaria* culture growth retardation in response to cycloheximide induced endosymbiont digestion. (**E**) Percentage change in *P. bursaria* cell number in cultures treated with cycloheximide (50 μgmL^{-1}), compared to untreated controls, after 12 days of feeding with *E. coli* expressing; *PiwiA1, PiwiC1* or *Pds1* dsRNA (dark grey) to induce knock-down (KD), or a non-hit 'scramble' dsRNA control (green). (**C-E**) Feeding was conducted daily for four days prior to cycloheximide treatment, and continued throughout. Boxplot data are represented as max, upper quartile (Q3), mean, lower quartile (Q1) and min values of six biological replicates. Individual data points are shown. Significance calculated as *$p \leq 0.05$, **$p \leq 0.01$, and ***$p \leq 0.001$, using a generalized linear model with quasi-Poisson distribution. See also **Figure S3** for the raw count data used to calculate the percentage change in cell number presented in **1D-E**.

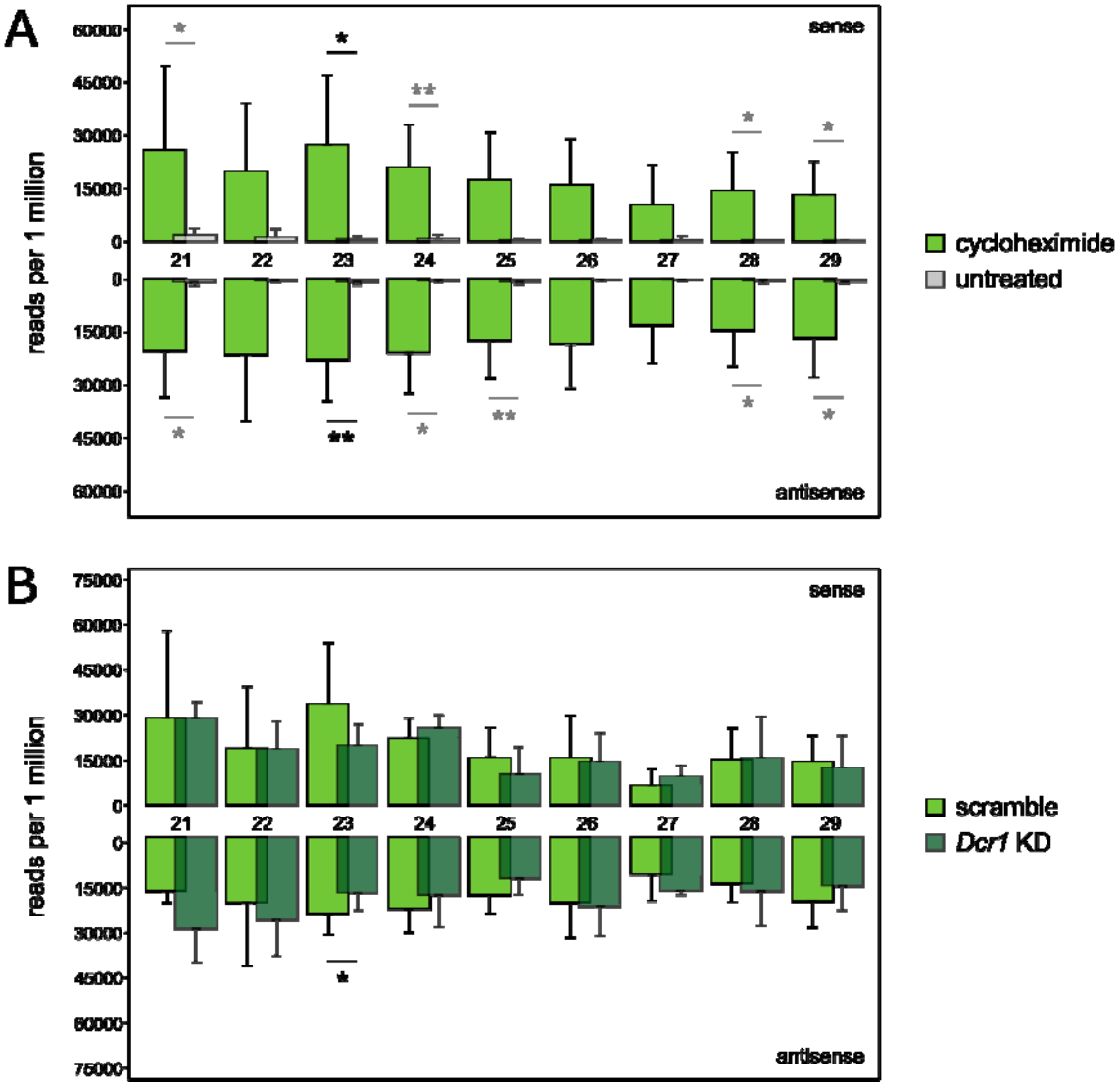


Figure 2. Endosymbiont breakdown triggers an influx of Dicer-dependent endosymbiont-derived sRNA within *P. bursaria*. **(A)** Size distribution (nt) of sRNA mapped to endosymbiont-derived cytoplasmic mRNA. sRNA was extracted from *P. bursaria* cultures over two days (day 2 and 3) of cycloheximide treatment (50 μgmL^{-1}), or from untreated controls. Note the relative increase in 23-nt abundance between untreated and cycloheximide treated cultures. *P. bursaria* cultures were fed with *E. coli* transformed to express non-hit 'scramble' dsRNA. Data are represented as mean ± SD of six biological replicates, and normalised against total endosymbiont mRNA-mapping 21-29-nt reads per dataset. (**B**) Size distribution (nt) of sRNA mapped to endosymbiont-derived cytoplasmic mRNA. sRNA was extracted from *P. bursaria* cultures on day 3 of cycloheximide treatment

(50 µgmL^{-1}). *P. bursaria* cultures were fed with *E. coli* transformed to express Dicer (*Dcr1*) dsRNA to induce knock-down (KD), or a non-hit 'scramble' dsRNA control. Note the relative increase in 23-nt abundance during cycloheximide treatment in *P. bursaria* cultures exposed to scramble dsRNA, compared to Dicer dsRNA. Data are represented as mean ± SD of three biological replicates, and normalised against total endosymbiont mRNA-mapping 21-29-nt reads per dataset. (**A**-**B**) Feeding was conducted daily for four days prior to cycloheximide treatment, and continued throughout. Significance calculated as *p ≤ 0.05, **p ≤ 0.01, and ***p ≤ 0.001 using a generalised linear model with quasi-Poisson distribution. All curated 'endosymbiont' mRNA transcript bins used for sRNA mapping are available on Figshare (10.6084/m9.figshare.12301736). See also **Figure S4** for sRNA abundance at each individual day.

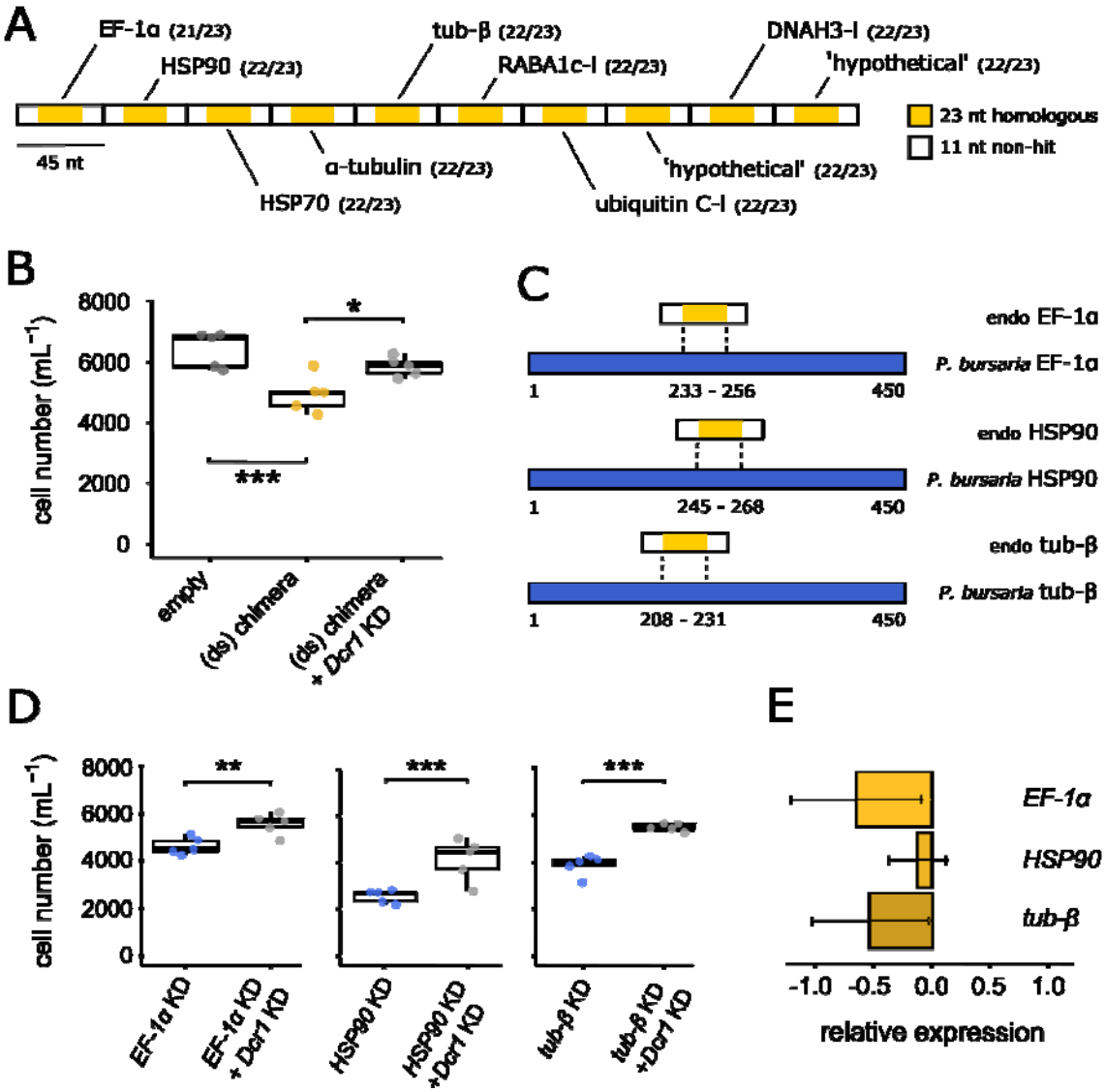


Figure 3. ***In vivo*** **exposure to a synthetic endosymbiont-derived RNA chimera generates simultaneous knock-down of homologous host genes in** ***P. bursaria*****.** (**A**) Schematic showing chimeric construct design. Ten endosymbiont-derived 45-nt transcript sequences (**Table S3**), featuring a 23-nt region with >91% sequence identity between endosymbiont and host (yellow) flanked by 11-nt of 'non-hit' algal transcript (white). Numbers in brackets denote sequence identity between endosymbiont and host. (**B**) *P. bursaria* cell number after 12 days of feeding with *E. coli* expressing: chimera dsRNA (yellow); chimera dsRNA mixed with Dicer (*Dcr1*) dsRNA (light grey; rescue); or an empty vector control (dark grey). See **Figure S9** for an experimental dilution of dsRNA chimera delivery. (**C**) Schematic demonstrating the

503 degree of overlap between three endosymbiont (endo) derived transcripts from the
504 chimeric construct (white/yellow) and respective 450-nt homologous region of the host (*P.*
505 *bursaria*) transcript (blue). Each region of 23-nt overlap (yellow) represents putative RNAi
506 'collisions' between endosymbiont and host. (**D**) *P. bursaria* cell number after 12 days of
507 feeding with *E. coli* expressing: host *EF1-α, HSP90* or *tub-β* 450-nt dsRNA (blue) to induce
508 host knock-down (KD), compared to Dicer (*Dcr1*) dsRNA mixed controls (grey; rescue
509 phenotype). See also **Figure S11** for individual knock-down (KD) of each of the ten broader
510 host targets of the endosymbiont-derived dsRNA chimera. (**E**) qPCR of mRNA extracted from
511 day 3 of chimera-RNAi feeding (**3B**), revealing knock-down of *EF1-α, HSP90* and *tub-β* host
512 gene expression in *P. bursaria* in response to endosymbiont derived chimera dsRNA
513 exposure. Standardised expression of an Actin housekeeping gene was used for
514 normalisation. Data are represented as mean ± SD of three biological replicates. These three
515 genes were assessed for knock-down as a result of endosymbiont derived chimera
516 exposure, based on evidence that directed knock-down of the wider corresponding host
517 gene led to retardation of *P. bursaria* culture growth (**3D**). For an extended figure showing
518 the effect of Dicer knock-down on the expression of these three host genes, see **Figure S12**.
519 (**B**/**D**) Multiple vector delivery was conducted at a 50:50 ratio during feeding. Boxplot data
520 are represented as max, upper quartile (Q3), mean, lower quartile (Q1) and min values of
521 five biological replicates. Individual data points are shown. Significance for boxplot data
522 calculated as *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, using a generalized linear model with
523 quasi-Poisson distribution.

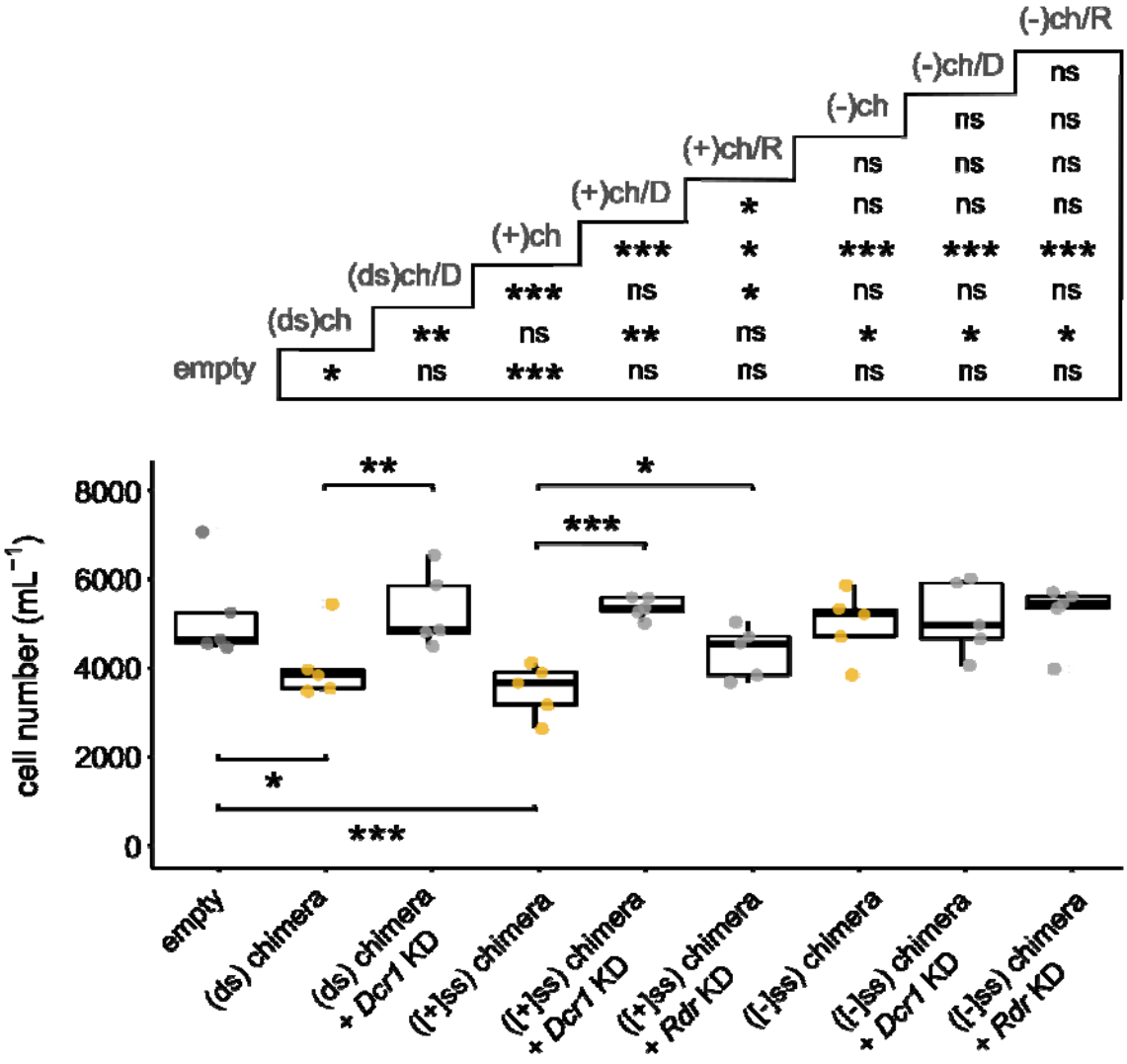


Figure 4. Single-stranded delivery of synthetic endosymbiont-derived RNA, analogous to endosymbiont mRNA, results in a cost to *P. bursaria* growth. *P. bursaria* cell number after 14 days of feeding with *E. coli* expressing; chimera RNA in ds, [+]ss or [-]ss orientation (yellow); chimera RNA in ds, [+]ss or [-]ss orientation mixed with Dicer (*Dcr1*) or RdRP (*Rdr*) dsRNA (light grey; rescue); or an empty vector control (dark grey). Multiple vector delivery was conducted at a 50:50 ratio during feeding. Asterisks displayed in the grid above denote pairwise significance values. Boxplot data are represented as max, upper quartile (Q3), mean, lower quartile (Q1) and min values of five biological replicates. All significance for boxplot data calculated as *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001, 'ns' no significance, using a generalized linear model with quasi-Poisson distribution.

METHODS

Culture conditions and media

In all RNAi experiments, *Paramecium bursaria* 186b (CCAP 1660/18) strain was used. For experiments requiring a non-photo-endosymbiotic *Paramecium* system for comparison, *Paramecium tetraurelia* nd7 strain was used. For eDicer analysis, *Paramecium bursaria* 186b and Yad1g1N strains were used[41,55].

Paramecium 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. *Paramecium* cells were sub-cultured 1:9 into fresh bacterized NCL media once per month for *Paramecium bursaria* 186b, and once every two weeks for *Paramecium tetraurelia* nd7. *Paramecium* cultures were maintained at 18°C with a light-dark (LD) cycle of 12:12h.

LysoTracker staining and fluorescent imaging

Paramecium cells were fixed using 0.5% paraformaldehyde and incubated for 20 mins at RT. Cells were then treated with 2 µM LysoTracker Green DND-26 (Invitrogen) to stain acidic vesicles such as the lysosome, and incubated in constant darkness for 2 hours at RT prior to imaging. Stained *Paramecium* cells were imaged on an ImageXpress Pico Automated Cell Imaging System at 10x magnification, using the Cy5 (absorbance – 630/40 nm, emission – 695/45 nm) and FITC (absorbance – 465/40 nm, emission – 525/30 nm) channels to capture algal chlorophyll autofluorescence and LysoTracker Green DND-26 fluorescence respectively.

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. 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

Paramecium 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 RNA (dsRNA). 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.

Paramecium 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.

Single-stranded construct design and confirmation

To create a version of L4440 which expressed only single-stranded (ssRNA), L4440 was digested with KpnI and PvuII (Promega), gel-purified (Wizard SV Gel and PCR Clean-Up System, Promega) and blunted using *PfuUltra* HF DNA polymerase (Agilent Technologies). The blunt vector was re-ligated using T4 DNA ligase (Thermo Scientific) and confirmed by sequencing (Eurofins Genomics), generating plasmid pDM004 which contains only a single T7 promoter. Fragments were then excised from their respective L4440 plasmids, or amplified by PCR (Q5 Polymerase; New England Biolabs) to swap the restriction sites, digested, and ligated into pDM004 to generate plasmids containing inserted fragments in

595 sense [+] or antisense [-] orientation. Generated plasmids were transformed into HT115 *E.*
596 *coli* for use in RNAi feeding experiments.

597 To confirm that these plasmid constructs generated only ssRNA, cultures of *E. coli*
598 HT115 containing pDM005-1 ([-]ssRNA chimera) or pDM005-2 ([+]ssRNA chimera) were
599 grown overnight at 37°C, 180 rpm in LB supplemented with 50 μgmL^{-1} ampicillin and 12.5
600 μgmL^{-1} tetracycline. Cultures were diluted 1:25 in fresh medium and grown to an OD_{600} of
601 0.4-0.6. Expression was then induced with 400 µM IPTG for 3 hrs, after which 1 mL of
602 culture was pelleted by centrifugation for 2 mins at 3,100 x *g*. RNA was extracted using an
603 RNeasy Mini Kit (Qiagen), following the manufacturer's protocol for Total RNA Purification
604 from Animal Cells. 20 µl of RNA was then treated with 10 μgmL^{-1} RNase A (Sigma-Aldrich) for
605 1 hour at 30°C in the presence of 300 mM NaCl (stabilising dsRNA and allowing RNase A to
606 degrade only ssRNA[63]). A separate aliquot was left untreated, with 300 mM NaCl added to
607 facilitate precipitation. All samples were then extracted with 1:1 phenol:chloroform and
608 precipitated with 2 volumes of ethanol. Pellets were washed twice with 80% ethanol and re-
609 suspended in nuclease-free water. RNA samples were then cleared of residual genomic DNA
610 using the TURBO DNA-*free* Kit (Ambion), following the manufacturer's protocol for routine
611 DNase treatment. RT-PCR was then performed using the Qiagen OneStep RT-PCR kit
612 following the manufacturer's instructions, with 0.5 µL template RNA and 0.6 µM each
613 primer (pDM005_RT_F: 5'-ACTTCAATGATTCGCAGCGG-3' and pDM005_RT_R: 5'-
614 AAGTAGCTGCTGTTCTCGGT-3'), generating an 85-nt PCR product. Cycling conditions were as
615 detailed in the manufacturer's protocol (30 cycles), with 1 min annealing at 50°C. PCR
616 products were then resolved on a 2% agarose gel to assess for the presence/absence of
617 amplification in each sample (**Figure S11**).

618 ***qPCR analysis***

619 RNA was extracted from *P. bursaria* 186b for gene expression analysis after three days of
620 RNAi feeding. *Paramecium* cells (~10^3 per culture) were pelleted by gentle centrifugation
621 (800 x *g* for 10 mins), snap-frozen in liquid nitrogen, and stored at -80°C. RNA extraction was
622 performed using TRIzol reagent (Invitrogen), following the manufacturer's protocol after re-
623 suspending each pellet in 900 µl TRIzol reagent. RNA was precipitated using GlycoBlue Co-
624 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: *EF-1α* (slope: -3.353; R^2: 0.999; efficiency 98.740%), *HSP90* (slope: -3.319; R2: 0.998; efficiency 100.131%), *tub-β* (slope: -3.378; R^2: 0.992; efficiency 97.692%), and *actin* (slope: -3.349; R^2: 0.983; efficiency 98.866%), using StepOne software v2.3. Each 20 µL reaction contained 10 µL PowerUp SYBR Green Master Mix (Thermo Fisher Scientific), 500 nM each primer (300 nM for *tub-β*) and 1 µL (50 ng) cDNA. 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 (*EF-1α* (58°C), *HSP90* (58°C), *tub-β* (57°C) and *actin* (58°C)). Primers pairs for each reaction are listed in **Table S5**. 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[64].

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.

Assigning sRNAs to the algal endosymbiont transcript bins

Trimmed reads were mapped against the '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. Using these exported count tables, of mapped reads per read length, transcripts with >10 hits for read lengths between 21-25 nt were searched using reciprocal BLASTX against the NCBI non-redundant 'nr' proteins sequence database, in order to assign taxonomic identity to each transcript. This allowed filtering of the main 'endosymbiont' dataset into subsets corresponding to either: algal cytoplasmic mRNA, algal cytoplasmic rRNA, algal plastid RNA or algal mitochondrial RNA; host RNA contamination; bacterial RNA contamination; or vector RNA contamination. All identified algal cytoplasmic rRNA (28 transcripts), plastid RNA (56 transcripts) or mitochondrial RNA (18 transcripts) sequences were sorted into new datasets representing each RNA species. Host (5 transcripts) and bacterial (34 transcripts) contamination were sorted into respective 'host' or 'bacterial' datasets. Vector (1 transcript) and all unidentifiable sequences (20 transcripts) were separated into a dataset labelled as 'other'. All remaining transcripts with >10 sRNA hits for read sizes 21-25 nt in the original dataset now represented putative algal mRNA transcripts (i.e. for 23-nt reads this represented 3,659 sRNA reads mapping to 148 transcripts). In this algal mRNA dataset, we also included transcripts that fell below the >10 sRNA hit threshold for manual curation, and which therefore may correspond to reads mapping to all of the above bins (i.e. for 23-nt reads this represented 1,949 reads mapping to 605 transcripts in total, of which 1,408 reads were mapping to 468 transcripts which were not manually curated, so could be affected by contamination). This manual binning process was carried out in order to double check the original automated binning post transcriptome assembly, and to check for possible chimeric host-algal transcript sequences produced as a by-product of cDNA synthesis and transcriptome sequencing and assembly. This process collectively allowed accurate segregation of the existing 'endosymbiont' transcript bin into discrete subsets based on RNA 'species'.

Using the HISAT2 alignment program with default settings, trimmed reads were mapped once more against the newly filtered algal mRNA, rRNA, plastid and mitochondrial datasets from above. Count tables of mapped reads per read length were once again

717 generated from the BAM files, and used to plot a size distribution of 21-29 nt endosymbiont
718 derived sRNA abundance per RNA species. Size distributions of sRNA abundance for each
719 sample were plotted using the R programming language packages; tidyverse, grid.extra and
720 ggplot2 in R Studio.

721 ***eDicer methods for identifying putative RNA-RNA interactions***

722 To predict putative mRNA-mRNA interactions using eDicer
723 (https://github.com/fmaguire/eDicer), both the 'host' and 'endosymbiont' transcript bins
724 processed from transcriptome data for *P. bursaria* Yad1g1N[55] were further filtered to
725 minimise the risk of false positives for host and endosymbiont cross-comparisons. For the
726 'host' bins, any transcript with >90% ID BLASTN hit to the following genome assemblies
727 were removed: *Chlamydomonas reinhardtii* cc503 cw92 mt, *Chlamydomonas reinhardtii* v3,
728 *Chlorella sorokiniana* 1228 v2, *Chlorella sorokiniana* DOE1412 v3, or *Chlorella sorokiniana*
729 utex1230 lanl v2 assemblies from Los Alamos National Labs Greenhouse genome database.
730 Similarly, for the 'endosymbiont' bins, any transcript with >90% ID BLASTN hit to the
731 following ParameciumDB genome assemblies were removed: *P. biaurelia* V1-4 v1, *P.*
732 *caudatum* 43c3d v1, *P. bursaria* MAC 110224 v1, *P. decaurelia* MAC 223 v1, *P. dodecaurelia*
733 MAC 274 v1, *P. jenningsi* MAC M v1, *P. novaurelia* MAC TE v1, *P. octaurelia* K8 CA1, *P.*
734 *quadecaurelia* MAC NiA v1, *P. primaurelia* Ir4-2 v1, *P. tetraurelia* MAC 51 with and without
735 IES, or *P. tetraurelia* MAC. Additionally, the reference (refseq) genome of *Escherichia coli*
736 MG1655 (NCBI acc.: NZ_CP012868.1) and *Klebsiella pneumoniae* HS11286 (NCBI acc.:
737 NC_016845.1) were used as 'food' comparison datasets.

738 The filtered 'endosymbiont' transcript bin, *E. coli* CDS and *K. pneumoniae* CDS were
739 decomposed into all possible 21-23 nt reads with jellyfish v2.2.10 and aligned with 95%
740 identity to the 'host' bin using Bowtie v1.2.3 (recommended for short alignments). These *in-*
741 silico RNA-RNA interaction simulations were performed using a wrapper tool we created
742 named eDicer (v1.0.0) (https://github.com/fmaguire/eDicer). The number of distinct
743 aligning k-mers (i.e. putative interactions of unique sequences) for each pair of bins were
744 then normalised by dividing by the total number of distinct k-mers in both bins, and x 100,
745 to calculate a Jaccard Index % (i.e. normalised set similarity).

746 To predict putative 'lethal' RNA-RNA interactions, a separate analysis was conducted
747 using a subset of each dataset that was putatively homologous to a yeast 'lethal' gene

748 database[56]. This 'lethal' dataset contained genes known to be conditionally essential in
749 *Saccharomyces cerevisiae*. These putative 'lethal' homologues for each dataset were
750 identified using a tBLASTx search of the yeast 'lethal' database[56] with a gathering threshold
751 set at 1e-10 with a minimum of 50% sequence identity. Once curated, each 'lethal' dataset
752 was subject to eDicer analysis as described above.

753 To predict putative rRNA-rRNA interactions using eDicer, a further analysis was
754 conducted using a dataset consisting of full-length ribosomal RNA (rRNA) clusters for
755 *Microactinium conductrix* [NCBI acc.: ASM224581v2][65], *K. pneumoniae* [NCBI acc.:
756 NC_016845.1][66], *E. coli* [NCBI acc.: NZ_CP012868.1], *P. bursaria* Yad1g1N[55] and *P. bursaria*
757 186b. Once again, each dataset was subject to eDicer analysis as described above. All *in-*
758 *silico* RNA-RNA interaction predictions were plotted using the R programming language
759 packages; tidyverse, grid.extra and ggplot2 in R Studio.

760 ***Manual curation of additional host transcript bins***

761 In order to identify host transcripts that could be impacted by putative RNA-RNA
762 interactions as a result of endosymbiont derived sRNA exposure, trimmed Illumina reads for
763 all sRNA sequencing samples were mapped to the 'endosymbiont mRNA' transcript dataset.
764 Resulting mapping files were filtered with custom scripts
765 (https://github.com/guyleonard/paramecium) producing tables of mapped hits with their
766 respective read lengths. Reads from all tables, of length 23-nt only, were then extracted with
767 SEQTK (resulting in 3,690 total). A BLASTn search of these 23-nt endosymbiont reads was
768 conducted against the 'host' transcript dataset, to identify host transcripts with ≥ 95%
769 identity over a 23-nt region. This resulted in three candidate host transcripts. These three
770 host transcripts were searched using BLASTx against the NCBI non-redundant 'nr' protein
771 database, resulting in one transcript with 100% sequence identity to both endosymbiont
772 and host HSP90. For the remaining two transcripts, SNPs present in the identified 23-nt
773 mapped sequence indicate that these are putative host sequences with 2-nt of mismatch
774 compared to the respective endosymbiont sequence, and were therefore unlikely to
775 represent a product of RNA-RNA interactions between endosymbiont and host as a result of
776 endosymbiont derived 23-nt sRNA exposure.

777 In order to identify host transcripts that could be classified as 'non-hit' transcripts (i.e.
778 host transcripts sharing a low level of sequence identity with algal transcripts over 23-nt

regions), the 'host' dataset was searched using an organism specific BLASTn (Chlorellaceae [NCBI: taxid35461]) against the NCBI nucleotide 'nr/nt' database with a minimum expectation of 1e-05. Any host transcripts sharing >20-nt sequence identity with algal transcripts over a 23-nt region were rejected. This process was repeated until a dataset of 20 'non-hit-algal' host transcripts were identified. These putative 'non-hit-algal' host transcripts were then searched using BLASTx against the NCBI non-redundant 'nr' protein database, to confirm that these 20 transcripts were derived from *Paramecium* and were not bacterial contamination.

DATA AND SOFTWARE AVAILABILITY

The raw reads generated during 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.4978160.v1), under the relevant headings. Custom scripts for sRNA read processing (https://github.com/guyleonard/paramecium, https://doi.org/10.5281/zenodo.4638888) and eDicer comparative analysis (https://github.com/fmaguire/eDicer, https://doi.org/10.5281/zenodo.4659378) are available on GitHub and archived within the Zenodo repository.

BIBLIOGRAPHY

1. Archibald, J. M. Endosymbiosis and eukaryotic cell evolution. *Curr. Biol.* **25**, R911-921 (2015).
2. Baurain, D. *et al.* Phylogenomic evidence for separate acquisition of plastids in Cryptophytes, Haptophytes, and Stramenopiles. *Molecular biology and evolution* **27**, 1698–709 (2010).
3. Keeling, P. J. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annual Review of Plant Biology* **64**, 583–607 (2013).
4. Archibald, J. M. The puzzle of plastid evolution. *Curr Biol* **19**, R81-8 (2009).
5. 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).

806 6. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer:
807 organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135 (2004).
808 7. Gray, M. W. Mitochondrial evolution. *Cold Spring Harb Perspect Biol* **4**, (2012).
809 8. Moran, N. A. Symbiosis as an adaptive process and source of phenotypic complexity.
810 *Proc. Natl. Acad. Sci. U.S.A.* **104 Suppl 1**, 8627–8633 (2007).
811 9. Ågren, J. A., Davies, N. G. & Foster, K. R. Enforcement is central to the evolution of
812 cooperation. *Nat Ecol Evol* **3**, 1018–1029 (2019).
813 10. West, S. A., Kiers, E. T., Simms, E. L. & Denison, R. F. Sanctions and mutualism
814 stability: why do rhizobia fix nitrogen? *Proceedings of the Royal Society of London.*
815 *Series B: Biological Sciences* **269**, 685–694 (2002).
816 11. Achilles-Day, U. E. & Day, J. G. Isolation of clonal cultures of endosymbiotic green
817 algae from their ciliate hosts. *J Microbiol Methods* **92**, 355–7 (2013).
818 12. Hoshina, R. & Kusuoka, Y. DNA Analysis of algal endosymbionts of ciliates reveals the
819 state of algal integration and the surprising specificity of the symbiosis. *Protist* **167**, 174–
820 184 (2016).
821 13. Zagata, P., Greczek-Stachura, M., Tarcz, S. & Rautian, M. The evolutionary relationships
822 between endosymbiotic green algae of *Paramecium bursaria* syngens originating from
823 different geographical locations. *Folia Biol (Krakow)* **64**, 47–54 (2016).
824 14. Kodama, Y. & Fujishima, M. Cycloheximide induces synchronous swelling of perialgal
825 vacuoles enclosing symbiotic *Chlorella vulgaris* and digestion of the algae in the ciliate
826 *Paramecium bursaria*. *Protist* **159**, 483–94 (2008).
827 15. Kodama, Y. & Fujishima, M. Symbiotic *Chlorella variabilis* incubated under constant
828 dark conditions for 24 hours loses the ability to avoid digestion by host lysosomal
829 enzymes in digestive vacuoles of host ciliate *Paramecium bursaria*. *FEMS Microbiol.*
830 *Ecol.* **90**, 946–955 (2014).

16. Omura, G. *et al.* A bacteria-free monoxenic culture of *Paramecium bursaria*: its growth characteristics and the re-establishment of symbiosis with *Chlorella* in bacteria-free conditions. *Jpn J Protozool* **37**, (2004).
17. Tanaka, M. *et al.* Complete elimination of endosymbiotic algae from *Paramecium bursaria* and its confirmation by diagnostic PCR. *Acta Protozool* **41**, 255–261 (2002).
18. Brown, J. A. & Nielsen, P. J. Transfer of photosynthetically produced carbohydrate from endosymbiotic Chlorellae to *Paramecium bursaria*. *J. Protozool.* **21**, 569–570 (1974).
19. Esteban, G. F., Fenchel, T. & Finlay, B. J. Mixotrophy in ciliates. *Protist* **161**, 621–641 (2010).
20. Johnson, M. D. Acquired phototrophy in ciliates: a review of cellular interactions and structural adaptations. *J Eukaryot Microbiol* **58**, 185–95 (2011).
21. Kato, Y. & Imamura, N. Amino acid transport systems of Japanese *Paramecium* symbiont F36-ZK. *Symbiosis* **47**, 99–107 (2009).
22. Kato, Y. & Imamura, N. Effect of sugars on amino acid transport by symbiotic *Chlorella*. *Plant Physiol Biochem* **46**, 911–7 (2008).
23. Kawakami, H. & Kawakami, N. Behavior of a virus in a symbiotic system, *Paramecium bursaria*—Zoochlorella. *The Journal of Protozoology* **25**, 217–225 (1978).
24. Parker, R. C. Symbiosis in *Paramecium bursaria*. *Journal of Experimental Zoology* **46**, 1–12 (1926).
25. 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).
26. Kodama, Y. & Fujishima, M. Infectivity of *Chlorella* species for the ciliate *Paramecium bursaria* is not based on sugar residues of their cell wall components, but on their ability

855 to localize beneath the host cell membrane after escaping from the host digestive vacuole
856 in the early infection process. *Protoplasma* **231**, 55–63 (2007).
857 27. Lowe, C. D., Minter, E. J., Cameron, D. D. & Brockhurst, M. A. Shining a light on
858 exploitative host control in a photosynthetic endosymbiosis. *Curr Biol* **26**, 207–211
859 (2016).
860 28. Sørensen, M. E. S. *et al.* The role of exploitation in the establishment of mutualistic
861 microbial symbioses. *FEMS Microbiol Lett* **366**, (2019).
862 29. Sørensen, M. E. S., Wood, A. J., Cameron, D. D. & Brockhurst, M. A. Rapid
863 compensatory evolution can rescue low fitness symbioses following partner-switching.
864 *bioRxiv* 2020.11.06.371401 (2020) doi:10.1101/2020.11.06.371401.
865 30. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396
866 (1981).
867 31. Foster, K. R. & Wenseleers, T. A general model for the evolution of mutualisms. *J. Evol.*
868 *Biol.* **19**, 1283–1293 (2006).
869 32. Frank, S. A. *Foundations of Social Evolution*. (Princeton University Press, 1998).
870 33. Herre, E. A., Knowlton, N., Mueller, U. G. & Rehner, S. A. The evolution of mutualisms:
871 exploring the paths between conflict and cooperation. *Trends in Ecology & Evolution* **14**,
872 49–53 (1999).
873 34. Szathmáry, E. & Smith, J. M. The major evolutionary transitions. *Nature* **374**, 227–232
874 (1995).
875 35. West, S. A., Fisher, R. M., Gardner, A. & Kiers, E. T. Major evolutionary transitions in
876 individuality. *PNAS* **112**, 10112–10119 (2015).
877 36. Buck, A. H. *et al.* Exosomes secreted by nematode parasites transfer small RNAs to
878 mammalian cells and modulate innate immunity. *Nat Commun* **5**, 5488 (2014).

879 37. Guo, Z., Li, Y. & Ding, S. W. Small RNA-based antimicrobial immunity. *Nat Rev*
880 *Immunol* **19**, 31–44 (2019).
881 38. Huang, C. Y., Wang, H., Hu, P., Hamby, R. & Jin, H. Small RNAs - big players in plant-
882 microbe interactions. *Cell Host Microbe* **26**, 173–182 (2019).
883 39. Wang, M. *et al.* Bidirectional cross-kingdom RNAi and fungal uptake of external RNAs
884 confer plant protection. *Nat Plants* **2**, 16151 (2016).
885 40. Weiberg, A. *et al.* Fungal small RNAs suppress plant immunity by hijacking host RNA
886 interference pathways. *Science* **342**, 118–23 (2013).
887 41. Jenkins, B. H. *et al.* Characterization of the RNA-interference pathway as a tool for
888 reverse genetic analysis in the nascent phototrophic endosymbiosis, *Paramecium*
889 *bursaria*. *Royal Society Open Science* **8**, 210140.
890 42. Carradec, Q. *et al.* Primary and secondary siRNA synthesis triggered by RNAs from food
891 bacteria in the ciliate *Paramecium tetraurelia*. *Nucleic Acids Res* **43**, 1818–33 (2015).
892 43. Galvani, A. & Sperling, L. RNA interference by feeding in *Paramecium*. *Trends Genet*
893 **18**, 11–2 (2002).
894 44. Karunanithi, S. *et al.* Exogenous RNAi mechanisms contribute to transcriptome
895 adaptation by phased siRNA clusters in *Paramecium*. *Nucleic Acids Res* **47**, 8036–8049
896 (2019).
897 45. Marker, S., Carradec, Q., Tanty, V., Arnaiz, O. & Meyer, E. A forward genetic screen
898 reveals essential and non-essential RNAi factors in *Paramecium tetraurelia*. *Nucleic*
899 *Acids Res* **42**, 7268–80 (2014).
900 46. Marker, S., Le Mouël, A., Meyer, E. & Simon, M. Distinct RNA-dependent RNA
901 polymerases are required for RNAi triggered by double-stranded RNA versus truncated
902 transgenes in *Paramecium tetraurelia*. *Nucleic Acids Res* **38**, 4092–107 (2010).

903 47. Karunanithi, S. *et al.* Feeding exogenous dsRNA interferes with endogenous sRNA
904 accumulation in *Paramecium*. *DNA research : an international journal for rapid*
905 *publication of reports on genes and genomes* **27**, (2020).
906 48. Cerutti, H., Ma, X., Msanne, J. & Repas, T. RNA-mediated silencing in algae: biological
907 roles and tools for analysis of gene function . *Eukaryot Cell* **10**, 1164–1172 (2011).
908 49. Hoehener, C., Hug, I. & Nowacki, M. Dicer-like enzymes with sequence cleavage
909 preferences. *Cell* **173**, 234-247.e7 (2018).
910 50. Ahmed, F. & Raghava, G. P. S. Designing of highly effective complementary and
911 mismatch siRNAs for silencing a gene. *PLoS One* **6**, (2011).
912 51. Du, Q., Thonberg, H., Wang, J., Wahlestedt, C. & Liang, Z. A systematic analysis of the
913 silencing effects of an active siRNA at all single-nucleotide mismatched target sites.
914 *Nucleic Acids Res* **33**, 1671–1677 (2005).
915 52. Jackson, A. L. *et al.* Widespread siRNA "off-target" transcript silencing mediated by
916 seed region sequence complementarity. *RNA* **12**, 1179–1187 (2006).
917 53. Doench, J. G., Petersen, C. P. & Sharp, P. A. siRNAs can function as miRNAs. *Genes*
918 *Dev* **17**, 438–442 (2003).
919 54. Ameres, S. L., Martinez, J. & Schroeder, R. Molecular basis for target RNA recognition
920 and cleavage by human RISC. *Cell* **130**, 101–112 (2007).
921 55. Kodama, Y. *et al.* Comparison of gene expression of *Paramecium bursaria* with and
922 without *Chlorella variabilis* symbionts. *BMC Genomics* **15**, 183 (2014).
923 56. Cotton, J. A. & McInerney, J. O. Eukaryotic genes of archaebacterial origin are more
924 important than the more numerous eubacterial genes, irrespective of function.
925 *Proceedings of the National Academy of Sciences* **107**, 17252–17255 (2010).
926 57. Fujishima, M. & Kodama, Y. Endosymbionts in *Paramecium*. *Eur J Protistol* **48**, 124–37
927 (2012).

928 58. Matthews, J. L. *et al.* Partner switching and metabolic flux in a model cnidarian-
929 dinoflagellate symbiosis. *Proc Biol Sci* **285**, (2018).
930 59. Matsuura, Y. *et al.* Recurrent symbiont recruitment from fungal parasites in cicadas.
931 *PNAS* **115**, E5970–E5979 (2018).
932 60. Husnik, F. & McCutcheon, J. P. Repeated replacement of an intrabacterial symbiont in
933 the tripartite nested mealybug symbiosis. *PNAS* **113**, E5416–E5424 (2016).
934 61. Hickey, D. A. Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* **101**, 519–
935 531 (1982).
936 62. Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the
937 legume–rhizobium mutualism. *Nature* **425**, 78–81 (2003).
938 63. Hansen, D. R., Van Alfen, N. K., Gillies, K. & Powell, W. A. Naked dsRNA associated
939 with hypovirulence of *Endothia parasitica* is packaged in fungal vesicles. *Journal of*
940 *General Virology*, **66**, 2605–2614 (1985).
941 64. Zhang, J. D., Biczok, R. & Ruschhaupt, M. ddCt: The ddCt Algorithm for the Analysis of
942 Quantitative Real-Time PCR (qRT-PCR). (Bioconductor version: Release (3.11), 2020).
943 doi:10.18129/B9.bioc.ddCt.
944 65. Arriola, M. B. *et al.* Genome sequences of *Chlorella sorokiniana* UTEX 1602 and
945 *Micractinium conductrix* SAG 241.80: implications to maltose excretion by a green alga.
946 *The Plant Journal* **93**, 566–586 (2018).
947 66. Liu, P. *et al.* Complete genome sequence of *Klebsiella pneumoniae* subsp. pneumoniae
948 HS11286, a multidrug-resistant strain isolated from human sputum. *J. Bacteriol.* **194**,
949 1841–1842 (2012).
950
951