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 photosynthetic organelles. Interaction 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 eukaryote-eukaryote endosymbiosis has yet to be explored. Here, we propose that endosymbiont-host RNA-RNA interactions, arising from digestion of endosymbionts, 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) pathway. 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 additionally identify 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 the host, 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 a symbiosis across fluctuating ecologies and symbiotic states.

SIGNIFICANCE STATEMENT

Stable endosymbiosis between eukaryotic microbes was a pre-requisite for the spread of photosynthetic organelles. Yet the mechanisms which act to stabilise a nascent eukaryote-eukaryote endosymbiosis are unclear. Using the model nascent endosymbiosis, *Paramecium bursaria* – *Chlorella* spp., we demonstrate that endosymbiont-host RNA-RNA interactions can drive a cost to host growth for 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 processing 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.

INTRODUCTION

The formation of a nascent endosymbiotic interaction between eukaryotes was a pre-requisite for the spread of photosynthetic organelles[1–5]. While textbook endosymbiotic theory typically depicts a prokaryote-eukaryote endosymbiosis, the occurrence of stable eukaryote-eukaryote interactions that result in the acquisition of organelles (specifically plastids), are more numerous[2–4,6]. This transition from transiently engulfed cells to established 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[1,3,5,7–9]. Conflict is an inevitable outcome of all symbioses, the resolution of which can significantly impact the stability of an interaction. Partner switching offers a simple yet effective escape from conflict. However, this often results in 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 to produce new organelles[10–14]. Enforcement offers a viable alternative, whereby cellular mechanisms act to

punish misbehaviour and therefore stabilise an interaction[15,16]. However, we know little about how enforcement mechanisms can emerge in a nascent endosymbiotic interaction.

The ciliate protist, *Paramecium bursaria*, and its intracellular green algae, *Chlorella* spp., represents a model phototrophic endosymbiosis[10,17]. Hundreds of alga cells, many of which are clonal[11,18,19], are harboured within modified host phagosomes called perialgal vacuoles[20]. The algae provide sugar and oxygen derived from photosynthesis, in exchange for amino acids, CO_2, divalent cations, and protection from viruses and a range of predators[6,21–27]. While the interaction is heritable, the *P. bursaria* – *Chlorella* system is described as a 'nascent' or 'facultative' endosymbiosis as both host and endosymbiont can typically survive independently[18,28–31]. Disruption of the endosymbiosis can lead to fusion of host lysosomes with the perialgal vacuoles and lysis of the algal endosymbiont within[29]. Manipulation of the endosymbiont population, including partner switching, is believed to have allowed *P. bursaria* to maintain control over the interaction[32–34]. Yet it is unclear how this endosymbiotic system is protected from over-exploitation by the host which would ultimately lead the interaction to collapse[35–40].

RNA-RNA interactions can play a role in host-pathogen symbiotic systems[41–45], whereby RNA is passed between interacting partners in order to modulate the expression of genes involved in virulence or resistance[41,44,45]. Many of these interactions utilise or 'hijack' the small interfering RNA (siRNA) pathway of the symbiotic partner[46]. These systems represent antagonistic symbioses[47,48]; 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*[49] and non-photo-endosymbiotic congener, *Paramecium tetraurelia*[50–54]. In both species, 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[49,51]. Evidence suggests that *Paramecium* can additionally process single-stranded RNA (ssRNA) – including ribosomal RNA (rRNA) and messenger RNA (mRNA) – derived from a prokaryotic cell acquired through phagocytosis[50]. This processing is mediated by conserved RNAi protein components which also function in endogenous transcriptome regulation[49,50,52,55].

Significantly, these studies propose that mRNA, derived both exogenously[50] and endogenously[49,52], can act as substrates for siRNA generation in *Paramecium*.

Here, we demonstrate that in *P. bursaria*, RNA released upon digestion of the algal endosymbiont is processed by the host RNAi system, giving rise to endosymbiont-derived sRNAs. For endosymbiont-derived mRNA sharing a high-level of sequence identity with host transcripts, these generated sRNAs may interfere with endogenous host gene expression, resulting in a cost to host growth (**Figure S1**). We track the interaction through sRNA sequencing, recapitulate the effect through *in vivo* experiments, and demonstrate that this mechanism is mediated by host Dicer, Piwi (AGO-Piwi), Pds1 and Rdr (RdRP) proteins. By demonstrating a physiological cost to the host for mass digestion of the endosymbiont population, we reveal a viable enforcement mechanism in the *P. bursaria-Chlorella* spp. endosymbiotic system. Such sanctions to the host would impose a fitness cost for the rapid and near-complete breakdown of the endosymbiosis, driving maintenance of the interaction across time and fluctuating ecological conditions. This process of host gene knock-down in response to endosymbiont-derived RNA processing by host RNAi factors, which we term 'RNAi-collisions', therefore represents a mechanism which can promote stability in a nascent eukaryote-eukaryote endosymbiotic cell-cell interaction.

RESULTS

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 cycloheximide: an RNA translational inhibitor to which *Paramecium* is known to exhibit a high degree of resistance[20,29,56]. Exposure to cycloheximide to disrupt algal function results in swelling of the perialgal vacuole and fusion with host lysosomes, triggering digestion of the algal endosymbionts[29]. A specific single nucleotide polymorphism in the ribosomal protein (RP) L29A has previously been identified as a determinant of cycloheximide resistance[57,58]. In order to corroborate cycloheximide resistance in both *P. bursaria* and non-photo-endosymbiotic *P. tetraurelia*, we compared predicted protein sequences of RPL29A in several species exhibiting cycloheximide resistance or susceptibility. All *Paramecium* strains, including *P. bursaria*, encoded leucine (L) in place of glutamine (Q) at this position (amino acid 38), inferring host resistance. This amino acid residue substitution is absent from the equivalent position in all

algal RPL29A proteins compared (including broader green algal species *Micractinium conductrix, Chlorella sorokiniana,* and *Chlamydomonas reinhardtii*), inferring that the endosymbiotic *Chlorella* spp. has selective susceptibility to cycloheximide (**Figure S2**).

Upon treatment with cycloheximide, *P. bursaria* cultures experienced a significant reduction in average algal-chlorophyll fluorescent intensity per host cell after only two-to-three days (**Figure 1A**). Quantification via live cell imaging revealed a clear de-coupling of *P. bursaria* cell number and loss of algal-chlorophyll fluorescence in cycloheximide treated cultures. This response is consistent with translational inhibition in the algae, but not in the host, and demonstrates that loss of a photosynthetic endosymbiont population does not immediately result in a decline in host Paramecia cell number. *P. bursaria* cells were additionally stained with LysoTracker Green to mark acidic vacuoles such as lysosomes. Assessment of LysoTracker fluorescence within the *P. bursaria* cytoplasmic environment during the first three days of cycloheximide treatment demonstrates that the loss of algal-chlorophyll fluorescence is linked to increased lysosomal activity in the host (**Figure 1B**). This indicates that the host is using phagocytic-like functions to break down the algae, confirming that elimination of the endosymbiont population in response to cycloheximide treatment[20,29] is triggered by host digestion.

To investigate the effect of continued cycloheximide treatment on *Paramecium* culture growth, both *P. bursaria* and non-photo-endosymbiotic *P. tetraurelia* were treated daily with cycloheximide over 8 days of exponential growth. A significant retardation to *P. bursaria* culture growth, but not *P. tetraurelia* culture growth, was observed during extended cycloheximide treatment (**Figure 1C**). Importantly, both *Paramecium* species harbour the same cycloheximide resistance conferring leucine-38 substitution in RPL29A (**Figure S2**), suggesting that elimination of endosymbiotic algae is costly to *P. bursaria* culture growth during continued cycloheximide treatment. Loss of algal derived photosynthate and plastid derived metabolites represents an obvious cost of endosymbiont elimination in *P. bursaria*. 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.

To investigate whether a cost to host growth upon endosymbiont digestion was mediated by the host RNAi system, we performed RNAi knock-down of host components known to be essential for RNAi function. Knock-down of *Dcr1* (a host-encoded

153 endoribonuclease Dicer required for siRNA generation) had previously been shown to
154 attenuate the effect of *E. coli* vector-based RNAi feeding in *P. bursaria*[49]. We therefore sought
155 to assess whether disruption of Dicer-mediated RNAi function could similarly attenuate a
156 putative RNAi effect arising from endosymbiont digestion. *P. bursaria* were fed with *E. coli*
157 expressing *Dcr1* (Dicer) dsRNA to compare the change in cell number between untreated and
158 cycloheximide treated conditions (**Figure 1D**; see also **Figure S3A**/**B**). We observed a
159 significant rescue of *P. bursaria* culture growth upon treatment with cycloheximide in cultures
160 exposed to Dicer dsRNA, compared to a 450-nt non-hit 'scramble' dsRNA[49] or empty vector
161 control. This indicates that partial knock-down of host Dicer was able to attenuate the cost to
162 *P. bursaria* growth associated with endosymbiont digestion. Importantly, the near complete
163 rescue observed suggests that host Dicer processing, rather than loss of algal derived
164 metabolites, may be a more significant factor in preventing host over-exploitation of the
165 endosymbiotic interaction under these conditions.

166 To further assess whether a physiological cost to host growth upon endosymbiont
167 digestion is mediated by the host RNAi machinery, *P. bursaria* cells were fed with *E. coli*
168 expressing *PiwiA1, PiwiC1* or *Pds1* dsRNA[49] in both untreated and cycloheximide treated
169 conditions. Knock-down of host-encoded *PiwiA1, PiwiC1* (AGO-Piwi effectors required for
170 targeted RNA cleavage) and *Pds1* (a *Paramecium*-specific component with an unknown but
171 essential role in exogenously induced RNAi) have previously been shown to attenuate the
172 effect of *E. coli* feeding-based RNAi in *P. bursaria*[49]. Here, knock-down of *PiwiA1* and *Pds1* also
173 attenuated the cost to *P. bursaria* culture growth upon treatment with cycloheximide,
174 compared to a non-hit 'scramble' dsRNA control (**Figure 1E**; see also **Figure S3C**/**D**). No
175 significant rescue of *P. bursaria* culture growth was observed in cultures exposed to *PiwiC1*
176 dsRNA, suggesting either that this component does not function in endosymbiont-digestion
177 induced RNA processing in this system, or that individual component knock-down was not
178 sufficient to attenuate this effect. These data suggest that the physiological cost to *P. bursaria*
179 growth incurred during cycloheximide treatment, in which the endosymbiotic algae are being
180 broken down and digested by the host, is dependent upon host-encoded *Dcr1* (Dicer), *PiwiA1*
181 and *Pds1* function. Importantly, the involvement of *Pds1* confirms that the observed RNAi
182 effect is mediated by the host, as the green algae sampled do not encode an identifiable
183 homologue of this RNAi pathway component[49,59]. Furthermore, it excludes the possibility that

184 off-target effects arising from host Dicer knock-down, including potential compensatory
185 function of broader Dicer or Dicer-like paralogues in *P. bursaria*[48], could be responsible for
186 the RNAi effect observed. Taken together, these data demonstrate a cost for endosymbiont
187 digestion that is mediated by host-encoded RNAi components, supporting the occurrence of
188 putative RNAi-mediated RNA-RNA interactions between endosymbiont and host.

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

191 To investigate the occurrence of RNA-RNA interactions between endosymbiont and host, we
192 tracked the abundance of endosymbiont-derived, host-processed sRNA in *P. bursaria* during
193 endosymbiont digestion. As above, *P. bursaria* cultures were treated with cycloheximide to
194 trigger endosymbiont elimination, compared to untreated control cultures, and fed with *E.*
195 *coli* expressing Dicer dsRNA (to disrupt host RNAi function) or a 'scramble' dsRNA control.
196 Dicer was chosen to disrupt RNAi function as knock-down of this component resulted in the
197 greatest rescue to *P. bursaria* culture growth during cycloheximide treatment (**Figure 1D**).
198 Total sRNA from each condition were isolated and sequenced, and the resulting forward and
199 reverse reads mapped to a manually curated dataset of transcripts identified as either algal-
200 derived cytoplasmic mRNA (excluding organellar mRNA - **Figure 2**; see also **Figure S4**) or algal-
201 derived cytoplasmic rRNA (excluding organellar rRNA - **Figure S5**). Reads were mapped with
202 100% identity, allowing no mismatches, to ensure that the subset of sRNA detected was of
203 definitive algal origin. This allowed us to directly test for an increase in endosymbiotic algal-
204 derived 23-nt sRNA (the size associated with host Dicer processing[49,50,60]) as a result of
205 cycloheximide induced endosymbiont breakdown, and assess the effect of knocking down
206 Dicer-mediated RNAi function on relative sRNA abundance.

207 We identified an increased abundance in all endosymbiont mRNA-derived 21-29 nt
208 reads upon treatment with cycloheximide, compared to the untreated control cultures
209 (**Figure 2A**; see also **Figure S4**). This same trend was observed for reads mapping to algal
210 endosymbiont rRNA-derived sRNA (**Figure S5**). This general increase in total algal sRNA
211 abundance upon treatment with cycloheximide indicates that this may be a result of
212 endosymbiont breakdown and degradation by the host during digestion. While not all of
213 these values were deemed to be statistically significant due to the high variance associated
214 with the sRNA datasets, a significant increase in endosymbiont mRNA-derived 23-nt sense

215 and antisense sRNA demonstrates an increased abundance of potential RNAi substrates
216 during endosymbiont digestion[49,50,60]. Importantly, the peak in endosymbiont mRNA-derived
217 23-nt antisense sRNA was significantly ablated in *P. bursaria* cultures exposed to Dicer dsRNA
218 during cycloheximide treatment, compared to a 'scramble' dsRNA control (**Figure 2B**; see also
219 **Figure S4**). This observation is supported by results from a previous study, which had
220 demonstrated that exposure to Dicer dsRNA was effective in specifically reducing both host
221 *Dcr1* expression and host-derived 23-nt sRNA abundance in *P. bursaria*[48]. Here, our results
222 suggest that an increase in endosymbiont mRNA-derived 23-nt antisense sRNA abundance
223 during endosymbiont digestion is also dependent on host-Dicer function, supporting the
224 hypothesis that endosymbiont-derived mRNA can be processed by the host RNAi machinery.
225 This is consistent with the observation that *Paramecium* RNAi factors are capable of
226 processing both endogenously and exogenously derived mRNA[50,52].

227 It should be noted that this Dicer-dependent increase in endosymbiont mRNA-derived
228 23-nt antisense sRNA was observed at both day two and day three of cycloheximide
229 treatment (**Figure 2** & **S4**), coinciding with the changes in algal chlorophyll / LysoTracker
230 Green fluorescence observed in an earlier experiment (**Figure 1A**/**B**). However, due to the
231 high variance associated with these sRNA datasets, a significant effect of 23-nt antisense sRNA
232 ablation upon Dicer knock-down could only be detected after three days of cycloheximide
233 treatment. This is consistent with a gradual increase in Dicer processed amplification products
234 following exposure to algal sRNA over multiple days. It is also important to note that this
235 effect likely under-represents the full extent of host Dicer processing, due to the paradoxical
236 and incomplete nature of host Dicer perturbation through RNAi knock-down[49]. Furthermore,
237 these data represent only a subset of the sRNA potentially present due to the stringent
238 mapping approach used to identify algal endosymbiont-derived sRNA, as any reads that
239 additionally mapped to the host with 100% sequence identity were excluded from this
240 analysis. While these excluded reads with a high level of shared sequence identity are
241 undoubtedly the most important subset of sRNA for the identification of putative mRNA-
242 mRNA interactions between endosymbiont and host, their exclusion here has allowed us to
243 comprehensively identify host RNAi processing of definitively algal endosymbiont-derived
244 transcripts. Interestingly, one of these host-derived transcripts that mapped with 100%
245 identity to an endosymbiont-derived 23-nt sRNA was *P. bursaria* heat shock protein 90

(*HSP90*). The importance of this putative endosymbiont-host RNA-RNA interaction, and a broader investigation of these regions of high shared sequence identity between endosymbiont and host RNA, is explored below.

To corroborate these findings, we conducted a series of control observations. Firstly, we explored the possibility that cycloheximide treatment was altering host derived sRNA production despite inferred host resistance (**Figure S2**). To ensure that host transcripts were unaffected by an increased rate of putative RNA-RNA interactions derived from the endosymbiont, we manually identified 20 host transcripts that contained <87% sequence similarity to any identified algal endosymbiont transcripts, and checked that they had no potential 23-nt overlap (allowing for ≤2-nt mismatches). All 21-25 nt forward and reverse sRNA reads were then mapped to these control host transcripts (**Figure S6**). These data demonstrate that for host mRNA transcripts which do not share a high-level of sequence identity with endosymbiont RNA, 23-nt sRNA abundance does not increase significantly during cycloheximide treatment. This is consistent with host resistance to cycloheximide not altering the general host derived population of sRNAs. Secondly, we sought to investigate whether free-living endosymbiont-derived algal strains would produce excess 23-nt sRNAs as a consequence of cycloheximide treatment. Two algal strains cultured under free-living conditions (*Chlorella variabilis* F36-ZK and *Chlorella variabilis* NC64A: both known endosymbionts of *P. bursaria*[61]) were cultured with and without cycloheximide treatment and subject to sRNA isolation and sequencing. (**Figure S7**). Importantly, no clear increase in 21-25 nt algal sRNA abundance was observed, with no 23-nt sRNA peak evident during treatment with cycloheximide under standard free-living growth conditions. 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[49].

These results suggest that increased algal endosymbiont-derived 23-nt sRNA abundance upon treatment with cycloheximide is dependent on host sRNA processing within the *P. bursaria*-algal endosymbiotic system. Taken together, these data indicate that an abundance of endosymbiont-derived sRNA are released during endosymbiont digestion, and that these may act as substrates for host RNAi processing. 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 as a result.

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 other systems[62–66]. Confirmation of a similar mismatch tolerance in *P. bursaria* is shown later in **Figures 3E** & **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[67]), 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 (**Figure 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*[68]) 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 (**Figure 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.

307 ***In vivo* exposure to a synthetic endosymbiont-derived RNA chimera generates simultaneous**
308 **knock-down of homologous host genes in *P. bursaria***

309 Having identified the occurrence of putative host transcript interactions by endosymbiont-
310 derived mRNA *in silico*, we sought to investigate whether *in vivo* exposure to endosymbiont
311 RNA could recapitulate the cost to host growth observed during endosymbiont digestion
312 (**Figure 1E**). We identified ten endosymbiotic mRNA interaction fragments that shared >91%
313 (or 21-nt) sequence identity with host transcripts across a 23-nt region, selected at random
314 from the analysis reported in **Figure S8A**. Each mRNA interaction fragment was chosen to
315 contain at least 1 SNP specific to the endosymbiont to ensure that any identifiable effect could
316 be attributed to the endosymbiont-derived transcript, rather than the host. We predicted
317 that these interaction fragments should still act as effective templates for RNAi-mediated
318 knock-down in *P. bursaria* based on the degree of tolerance for mismatching
319 complementarity observed in RNAi-mediated knock-down in other eukaryotic systems[62–66]
320 (and confirmed below in *P. bursaria*; **Figures 3E** & **S12**). Six of these mRNA interaction
321 fragments showed putative homology to 'non-lethal' yeast genes (including *EF1-*α and
322 *HSP90*); two showed putative homology to 'lethal' yeast genes (including *tub-*β); and the
323 remaining two had no identifiable homologues in yeast (**Table S1**). Significantly, heat shock
324 protein 90 (*HSP90*) was the host transcript previously identified as a candidate for putative
325 endosymbiont-host RNA-RNA interaction in the sRNA analysis of **Figure 2**. However, since the
326 region identified in **Figure 2** shared 100% sequence identity with the host, this newly
327 identified interaction fragment (with 2 SNPs specific to the endosymbiont) allowed us to
328 ensure that any identifiable effect could be attributed to endosymbiont-derived transcript.
329 All ten endosymbiont transcript fragments were composed of the predicted interacting 23-nt
330 sRNA sequence (**Figure S8A**) flanked by 11-nt of contiguous 'non-hit' endosymbiont transcript
331 as a filler on each side (**Table S3**). These ten 45-nt fragments were combined into a single 450-
332 nt synthetic endosymbiotic algal chimera (**Figure 3A**), cloned into an L4440 plasmid, and
333 transformed into *E. coli* for RNAi.

334 Using the same *E. coli* vector feeding-based approach to induce RNAi, exposure to
335 endosymbiotic algal chimera dsRNA resulted in significant retardation of *P. bursaria* culture
336 growth (**Figure 3B**). Once more, this effect was partially rescued by knock-down of host Dicer,
337 demonstrating an RNAi effect as a result of synthetic endosymbiont-host mRNA-mRNA

338 interactions that partially resembled the host response to endosymbiont elimination
339 observed in **Figure 1C-E**. It is important to note, however, that synthetic exposure to
340 endosymbiont-derived RNA via an *E. coli* feeding vector could be expected to over-represent
341 these putative RNA interactions. To address this issue, we designed a non-hit 'nonsense'
342 control composed of a tandem assembly of the 11-nt filler regions of contiguous 'non-hit'
343 algal transcript present in the chimera. A relative dilution of endosymbiont chimera dsRNA
344 delivery alongside the 'nonsense' control from 1, to 1:1 and 1:3, also resulted in *P. bursaria*
345 culture growth retardation (**Figure S9**), demonstrating that a reduced relative delivery of
346 synthetic endosymbiont RNA can also result in a cost to host growth. Similar experiments
347 were conducted to investigate the possibility of endosymbiont-host rRNA-rRNA interactions,
348 but showed that there was no equivalent effect on host growth (**Figure S10**, see also **Table**
349 **S4**).

350 In order to explore which of these identified putative mRNA-mRNA interactions were
351 capable of facilitating the cost to host growth observed in **Figure 3B**, we identified the ten
352 450-nt host transcripts predicted to be hit by each endosymbiont derived 23-nt mRNA
353 interaction fragment used in the chimera (**Figure 3C** - three examples are shown). Using the
354 same feeding-based approach to induce RNAi, individual exposure to dsRNA corresponding
355 to each wider *P. bursaria* transcript identified three which resulted in significant *P. bursaria*
356 culture growth retardation, relative to Dicer knock-down controls (**Figure 3D**; see **Figure S11**
357 for all 10). These *P. bursaria* transcripts correspond to elongation factor-1α (*EF-1α*), heat
358 shock protein 90 (*HSP90*; the host transcript identified as a candidate for putative
359 endosymbiont-host RNA-RNA interaction in the sRNA analysis of **Figure 2**), and tubulin-β
360 chain (*tub-β*). From these findings, it can be inferred that these are the 23-nt mRNA
361 interaction fragment targets of the chimera (**Figure 3A/C**) that were likely to result in the cost
362 to host growth observed upon synthetic endosymbiont chimera dsRNA exposure (**Figure 3B**).

363 We next sought to discern whether feeding on *E. coli* expressing synthetic
364 endosymbiont chimera dsRNA (in **Figure 3B**) led to a reduction of host *EF-1α, HSP90* and *tub-*
365 *β* mRNA levels. Using mRNA extracted from *P. bursaria* during chimera-RNAi feeding (**Figure**
366 **3B**), qPCR revealed a reduction in host transcript expression of *EF-1α* and *tub-β* in response
367 to endosymbiont chimera dsRNA exposure (**Figure 3E**). Host transcript expression of *HSP90*
368 appears inconclusive, however reduction in mRNA levels of all three host transcripts was

369 partially rescued upon simultaneous knock-down of host Dicer (**Figure S12**). From these data
370 we can conclude that only 23-nt of >91% complementary 'endosymbiont' transcript is
371 required to facilitate detectable knock-down of the corresponding host transcript, suggesting
372 that sRNAs with ≤2-nt mismatches are effective during RNAi-mediated knock-down of gene
373 expression in *P. bursaria* (see rationale for **Figure S8**). Importantly, this also demonstrates
374 that exposure to endosymbiont derived RNA containing a 23-nt region of overlap with host
375 transcripts is capable of impacting host gene expression.

376 It is important to consider that delivery of a synthetic endosymbiont-derived RNA
377 chimera in this manner represents only an approximation of the putative 23-nt RNA-RNA
378 interactions that may be occurring in the *P. bursaria – Chlorella spp.* system. As we have
379 already mentioned, exposure to endosymbiont RNA via an *E. coli* feeding vector would likely
380 over-represent these putative RNA interactions. Nonetheless, the random generation of 23-
381 nt Dicer substrates across the dsRNA chimera (only three regions of which were shown to
382 result in a detectable cost to host growth; **Figure 3A/D** & **Figure S11**) represents only part of
383 the wider *E. coli*-derived RNA population delivered to the host during RNAi-feeding. Coupled
384 with evidence of the efficiency of 'diluted' endosymbiont chimera dsRNA delivery (**Figure S9**),
385 these observations support the hypothesis that only a relatively small quantity of 23-nt RNA-
386 RNA interactions may be sufficient to induce a cost to host growth in *P. bursaria.* Importantly,
387 the observation that a 23-nt fragment with >91% sequence identity to a target transcript is
388 capable of facilitating detectable knock-down in *P. bursaria* indicates that the sRNA analysis
389 in **Figure 2** likely under-represents the true scale of endosymbiont-derived sRNAs that may
390 result in RNA-RNA interactions between endosymbiont and host. We also note that the
391 occurrence of identified endosymbiont-host RNA-RNA interactions does not exclude a wider
392 cost to the host arising from mass processing of a large population of non-target algal derived
393 RNAs, a process which must surely pose a cost to host cellular economics and transcriptional
394 fidelity. For further justification of this 'synthetic' approach, and a consideration of the results
395 that we can reliably draw from these data, see **Discussion S1**.

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

398 As RNA derived naturally from the endosymbiont was unlikely to be double-stranded, two
399 further constructs were designed to assess the efficacy of synthetic endosymbiont-derived

chimera single-stranded RNA (ssRNA) exposure in *P. bursaria* (**Figure 4**; for confirmation of ssRNA expression see also **Figure S13**). Using an *E. coli* vector feeding-based approach to induce RNAi, exposure to endosymbiont chimera ssRNA also resulted in significant retardation to *P. bursaria* culture growth when delivered in the sense orientation ([+]ssRNA). Notably, chimera [+]ssRNA exposure resulted in greater retardation to *P. bursaria* culture growth than chimera dsRNA exposure. However, delivery of chimera ssRNA in the antisense orientation ([-]ssRNA) did not result in significant retardation to *P. bursaria* culture growth. This [+]ssRNA orientation bias was also observed during Dicer knock-down of cycloheximide treated cultures in a prior experiment, in which delivery of Dicer [+]ssRNA resulted in a stronger rescue phenotype ($p > 0.05$) compared to delivery of Dicer [-]ssRNA or Dicer dsRNA (**Figure S14**). Importantly, the orientation of [+]ssRNA represents the same orientation as the mRNA transcripts from which each of these target templates was identified.

Simultaneous knock-down of host Dicer (dsRNA delivery) was able to partially rescue *P. bursaria* culture growth in endosymbiont chimera [+]ssRNA exposed cultures (**Figure 4**). In an additional attempt to disrupt host RNAi function, two further constructs were designed to target two RdRP encoding transcripts (*Rdr1, Rdr2*) present in *P. bursaria* (**Dataset S1B**). RdRP is an RNA-dependent RNA polymerase known to be essential for triggering a siRNA-based RNAi response through the generation of dsRNA from ssRNA[50,54]. RdRP proteins are implicated in the production of 'secondary siRNAs' from cleaved mRNA degradation products, and have recently been suggested to function in endogenous transcriptome regulation by generating mRNA-associated primary siRNAs[50,52–54]. These studies indicate that RdRP processes single-stranded (sense) RNA precursors into dsRNA substrates for Dicer. Here, simultaneous knock-down of *Rdr1* and *Rdr2* was able to rescue *P. bursaria* culture growth in endosymbiont chimera [+]ssRNA exposed cultures, though the effect (while still significant) was less than the near-complete rescue achieved through Dicer knock-down (**Figure 3B** & **Figure 4**). As for Dicer, knock-down of host RdRP through RNAi represents a potential paradox, and so constitutes only partial or transient loss of function[49]. 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 **Figure 4**. While the mechanism of down-stream Dicer processing of RdRP-generated dsRNA substrates remains unknown (see **Discussion S1**), we

430 can nonetheless infer from these data that single-stranded RNA-induced RNAi knock-down in
431 *P. bursaria* is partially dependent upon host RdRP function.

432 Irrespective of the mechanistic basis for a sense oriented ssRNA knock-down bias, it is
433 important to note that delivery of chimera [+]ssRNA demonstrated here represents the same
434 orientation as the endosymbiont derived mRNA interaction fragments from which these
435 putative endosymbiont-host RNA-RNA interactions were identified. Once more, this is
436 consistent with the observation that *Paramecium* RNAi factors are capable of processing both
437 endogenously and exogenously derived mRNA[50,52]. The ability to facilitate a cost to host
438 growth through delivery of single-stranded RNA in the same orientation as the host target
439 mRNA, supports the hypothesis that a cost to host growth upon endosymbiont digestion can
440 be facilitated by host RNAi-mediated RNA-RNA interactions between endosymbiont and host
441 mRNA populations.

442 **DISCUSSION**

443 Through manipulation of the model endosymbiotic system, *P. bursaria-Chlorella* spp., we
444 have demonstrated that endosymbiotic algal-derived RNA is released upon digestion of the
445 endosymbiont population. In a mechanism mediated by the host-encoded protein, *Dcr1*
446 (Dicer) these endosymbiont derived RNAs are processed into sRNAs, which can subsequently
447 act as substrates for the host RNAi system. This mechanism is additionally mediated by host-
448 encoded *PiwiA1* (AGO-Piwi), *Pds1* and *Rdr* (RdRP) components. In cases where a high level of
449 sequence identity exists between endosymbiont and host transcripts across a 23-nt region,
450 we propose that these processed endosymbiont derived sRNAs are capable of inducing knock-
451 down of complementary host transcripts. We postulate that these RNA-RNA interactions are
452 of importance in eukaryote-eukaryote endosymbioses where partners are likely to share
453 greater sequence similarity, especially among conserved transcripts. Through processing of
454 'exogenous' endosymbiont-derived RNA that results in knock-down of endogenous host gene
455 expression, a physiological cost to the host for large-scale elimination of the endosymbiont
456 population may be imposed. We demonstrate that these endosymbiont-host RNA-RNA
457 interactions can be recapitulated *in vivo* using an *E. coli* RNAi feeding vector expressing small
458 fragments of synthetic endosymbiotic algal mRNA. We also show that this effect is most
459 potent when the substrate for RNAi is delivered in a single-stranded sense orientation, in a
460 mechanism mediated by the host-encoded protein, RdRP. This is important, because it is

analogous to the manner of RNA exposure (mRNA) experienced by the host upon interaction with the endosymbiont. Due to the inherent difficulty in characterising such mechanisms directly, we have relied on this multiple experimental approach to demonstrate the viability of putative RNA-RNA interactions at each stage of the process. Collectively, this mechanism demonstrates the capacity for putative endosymbiont-host mRNA-mRNA interactions as a result of host RNAi processing – a mechanism which we term RNAi mediated RNA-RNA collisions (or 'RNAi collisions'; defined in **Figure S1**).

The long-term maintenance of symbiotic interactions represents a quandary for evolutionary theory[35–40]. 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 to produce new organelles[10–14]. 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[15,16]. These mechanisms have been identified in systems that span vast biological scales, from selfish elements in genomes[69] to conflict between partners in legume-rhizobium symbioses[70]. Indeed, enforcement is 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, by encouraging maintenance of the symbiosis that acts to stabilise the interaction[15]. 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 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[50,51,53], or the conserved gene repertoire and sequence composition of the host and endosymbiont transcriptomes (which include many transcripts which are conditionally essential in model systems[68]) from which some putative 'RNAi collisions' have so far been identified. In addition, unlike comparable mechanisms of RNA-RNA interactions that have been studied in host-pathogen symbioses[41–45], these 'RNAi collisions' appear to be untargeted and, hence, emergent. In the

492 aforementioned host-pathogen systems, targeted RNA is passed from one partner to the
493 other in order to modulate expression of transcripts involved in virulence or resistance[41,44,45].
494 However, in order for such targeted systems to evolve, they must first exist in an untargeted
495 form upon which selection is able to act, allowing the emergence of specific RNA factors[42,43].
496 Identification of undirected 'RNAi collisions' in the *P. bursaria* – *Chlorella* system represents
497 one such intermediary state, emergent in nature and untargeted, upon which sustained
498 cellular interaction coupled with the potential for host-symbiont conflict could drive the
499 selection of targeted RNA-RNA interactions. The reduced pattern of co-evolution between
500 endosymbiont and host in the *P. bursaria* – *Chlorella* system, resulting from the potential for
501 repeated partner switching by the host[11,34], offers a possible explanation for why this
502 mechanism has remained emergent.

503 We propose that 'RNAi collisions' represent a putative mechanism to discourage over-
504 exploitation of the endosymbiont population by the host. In a nascent eukaryote-eukaryote
505 endosymbiotic system this process would stabilise the interaction and, therefore, promote
506 cooperation. Here we have used the example of mass endosymbiont digestion in response to
507 drug treatment to simulate this effect in the extreme. In natural interactions between *P.*
508 *bursaria* and its algal endosymbiont, such a cost would only need to occur in the drastic
509 occurrence of mass endosymbiont digestion in order to drive stability of the interaction.
510 Importantly, the endosymbiotic algal population within *P. bursaria* is largely composed of
511 closely related or clonal lineages[11,18,19], and as such, the fate of the algal population should be
512 considered as a collective unit. This cost need only act to supress large-scale, rapid destruction
513 by the host in order to drive the maintenance of a surviving subsection of the endosymbiont
514 population. Previous studies have demonstrated how *P. bursaria* is capable of manipulating
515 endosymbiont load in response to varying light conditions to better suit its own ends[32,33],
516 however, in these examples, reduction of endosymbiont number through digestion is slow
517 and partial. By providing a system that selects against rapid and near-complete digestion of
518 the endosymbiont population, 'RNAi collisions' effectively buffer the endosymbiotic
519 interaction against total breakdown. We suggest that this has allowed the relationship to be
520 maintained across time and varying ecological conditions, even in the event of host-symbiont
521 conflict[32,33] and fluctuating endosymbiont numbers[20,29,32]. As an alternate route to conflict
522 resolution that avoids partner switching, we propose that such a mechanism would facilitate

greater co-evolution between endosymbiont and host. Over time, this would allow the metabolic and genetic integration that drives the formation of obligate symbioses to become manifest. We therefore present 'RNAi collisions' as a new mechanism in this endosymbiotic system – a factor which can promote stability in the face of conflict in an endosymbiotic eukaryote-eukaryote cell-cell interaction.

AUTHOR CONTRIBUTIONS

B.H.J., D.S.M., and T.A.R. conceived and designed the experiments. F.M. and T.A.R. conceived and designed the eDicer computational analysis. 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.

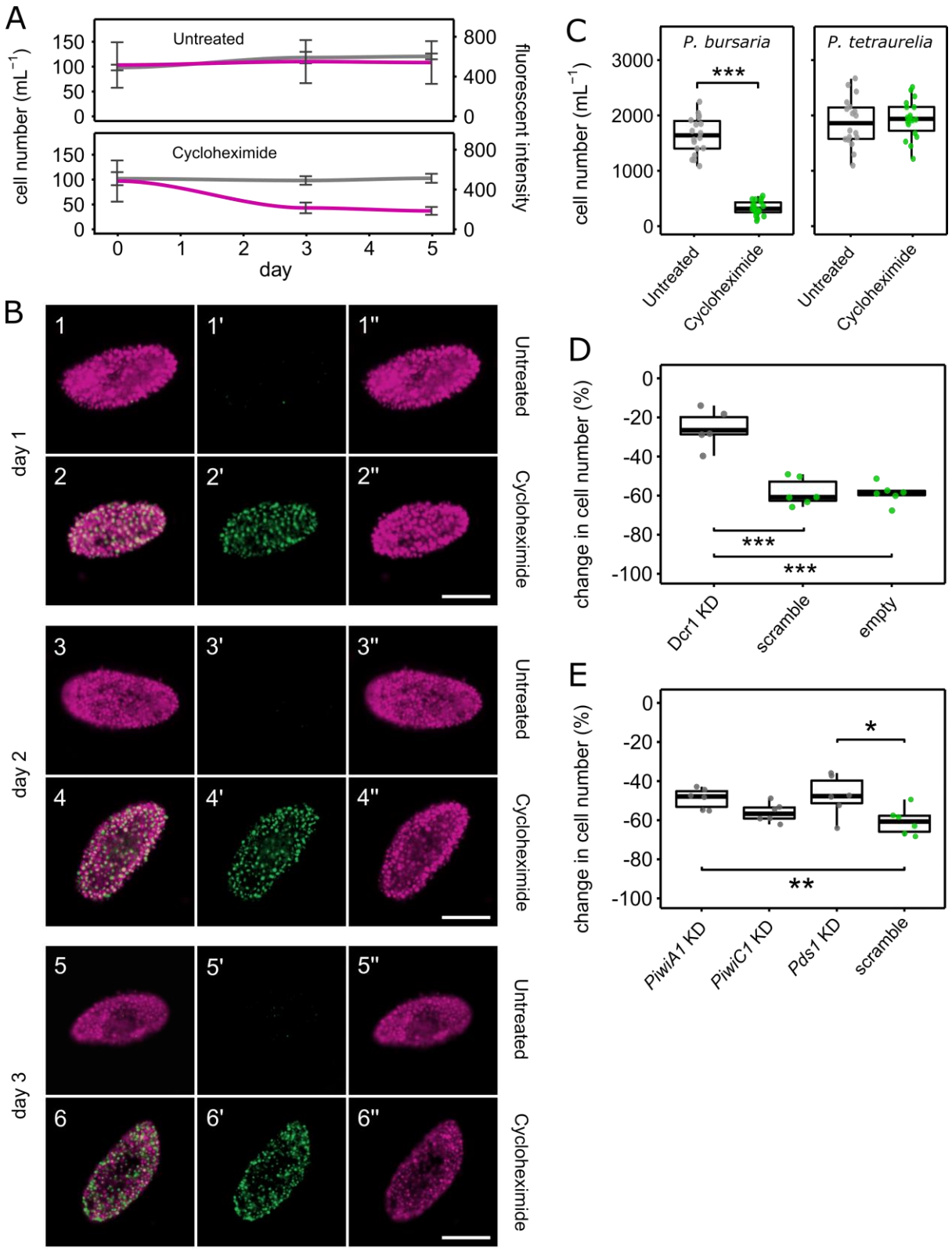
ACKNOWLEDGEMENTS

This work was primarily supported by an EMBO YIP award and a Royal Society University Research Fellowship (UF130382) and latterly by an ERC Consolidator Grant (CELL-in-CELL) to T.A.R. S.W. and J.D.E. were supported by awards from the Wellcome Trust (WT107791/Z/15/Z) and the Lister institute. F.M. was supported by a Donald Hill Family Fellowship in Computer Science. We thank Karen Moore and the University of Exeter 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.

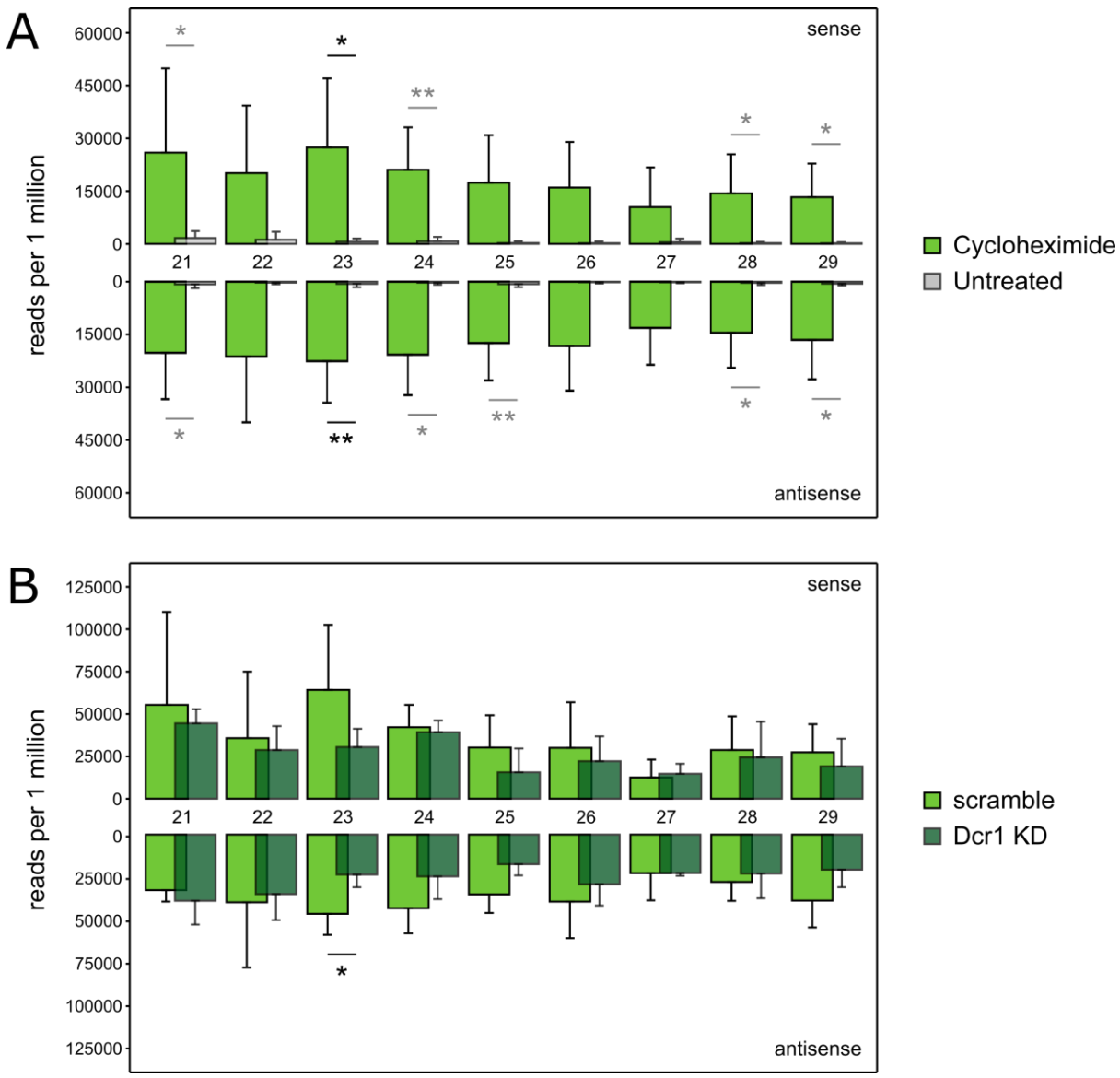
DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURES


A
Untreated
Cycloheximide
cell number (mL^{-1})
fluorescent intensity
day
B
day 1
day 2
day 3
Untreated
Cycloheximide
C
P. bursaria
P. tetraurelia
cell number (mL^{-1})
Untreated
Cycloheximide
D
change in cell number (%)
Dcr1 KD
scramble
empty
E
change in cell number (%)
PiwiA1 KD
PiwiC1 KD
Pds1 KD
scramble

548 **Figure 1. Endosymbiont digestion in *P. bursaria* results in an RNAi-mediated 'physiological**
549 **cost' to the host.** (**A**) *Paramecium* cell number (grey) vs algal chlorophyll fluorescent intensity
550 (pink) in *P. bursaria* cultures treated with cycloheximide (50 µgmL^{-1}) compared to an
551 untreated control. Cultures were maintained in the stationary phase. Loss of algal fluorescent
552 intensity indicates algal death in response to cycloheximide treatment. Data are represented
553 as mean ± SD of six biological replicates. (**B**) Stages of endosymbiont elimination in *P. bursaria*
554 cells at days 1, 2 and 3 of cycloheximide treatment (50 µgmL^{-1}). Algal chlorophyll fluorescence
555 (pink) highlights the presence of endosymbiotic algae within the *P. bursaria* cell. LysoTracker
556 Green fluorescence (green) indicates increased host lysosomal activity in response to
557 cycloheximide treatment. Scale bar – 50 µm. (**C**) *Paramecium* cell number in *P. bursaria* or *P.*
558 *tetraurelia* cultures after 8 days of treatment with cycloheximide (50 µgmL^{-1}), compared to
559 untreated controls. *Paramecium* cultures were fed with *E. coli* transformed with an empty
560 RNAi vector. (**D**) Percentage change in *P. bursaria* cell number in cultures treated with
561 cycloheximide (50 µgmL^{-1}), compared to untreated control cultures, after 12 days of feeding
562 with *E. coli* expressing; Dicer dsRNA (dark grey) to induce knockdown (KD), non-hit 'scramble'
563 dsRNA or an empty vector control (green). The relative effect of Dicer dsRNA exposure
564 compared to scramble and empty vector controls indicates partial rescue of the growth
565 retardation induced by cycloheximide treatment and loss of endosymbiotic algae. (**E**)
566 Percentage change in *P. bursaria* cell number in cultures treated with cycloheximide (50 µgmL^{-}
567 ^{1}), compared to untreated control cultures, after 12 days of feeding with *E. coli* expressing;
568 *PiwiA1, PiwiC1* or *Pds1* dsRNA (dark grey) to induce transcript knockdown (KD), or a non-hit
569 'scramble' dsRNA control (green). *P. bursaria* cultures were simultaneously treated with
570 cycloheximide (50 µgmL^{-1}) for 8 days (starting four days into feeding). (**C-E**) All boxplot data
571 are represented as max, upper quartile (Q3), mean, lower quartile (Q1) and min values of six
572 biological replicates. Individual data points are shown. All significance calculated as *$p \leq 0.05$,
573 **$p \leq 0.01$, and ***$p \leq 0.001$, using a generalized linear model with quasi-Poisson distribution.
574 See also **Figure S3** for the raw count data used to calculate the percentage change in cell
575 number presented in **1D**-**E**.



576

577

578 **Figure 2. Endosymbiont breakdown triggers an influx of dicer-dependent endosymbiont-**
579 **derived sRNA within *P. bursaria.* (A)** Size distribution of sRNA mapped to endosymbiont-
580 derived cytoplasmic mRNA. sRNA was extracted from *P. bursaria* cultures over two days (day
581 2 and 3) of cycloheximide treatment (50 µgmL^{-1}), or from an untreated control. Note the
582 relative increase in 23-nt abundance between untreated and cycloheximide treated cultures.
583 *P. bursaria* cultures were fed with HT115 *E. coli* transformed to express 'scramble' dsRNA.
584 Data are represented as mean ± SD of six biological replicates, and normalised against total
585 endosymbiont mRNA-mapping 21-29-nt reads per dataset. (**B**) Size distribution of sRNA
586 mapped to endosymbiont-derived cytoplasmic mRNA. sRNA was extracted from *P. bursaria*
587 cultures on day 3 of cycloheximide treatment (50 µgmL^{-1}). *P. bursaria* cultures were fed with

588 HT115 *E. coli* transformed to express 'scramble' or *Dcr1* (Dicer) dsRNA. Note the relative
589 increase in 23-nt abundance during cycloheximide treatment in *P. bursaria* cultures exposed
590 to scramble dsRNA, compared to Dicer dsRNA. Data are represented as mean ± SD of three
591 biological replicates, and normalised against total endosymbiont mRNA-mapping 21-29-nt
592 reads per dataset. (**A-B**) For all cultures, feeding was conducted daily for four days prior to
593 cycloheximide treatment, and continued throughout. Significance calculated as *p ≤ 0.05, **p
594 ≤ 0.01, and ***p ≤ 0.001 using a generalised linear model with quasi-Poisson distribution.
595 Here and elsewhere, the term 'KD' is used in figure to denote 'knock-down'. All curated
596 'endosymbiont' mRNA transcript bins used for sRNA mapping are available on Figshare
597 (10.6084/m9.figshare.12301736).

598

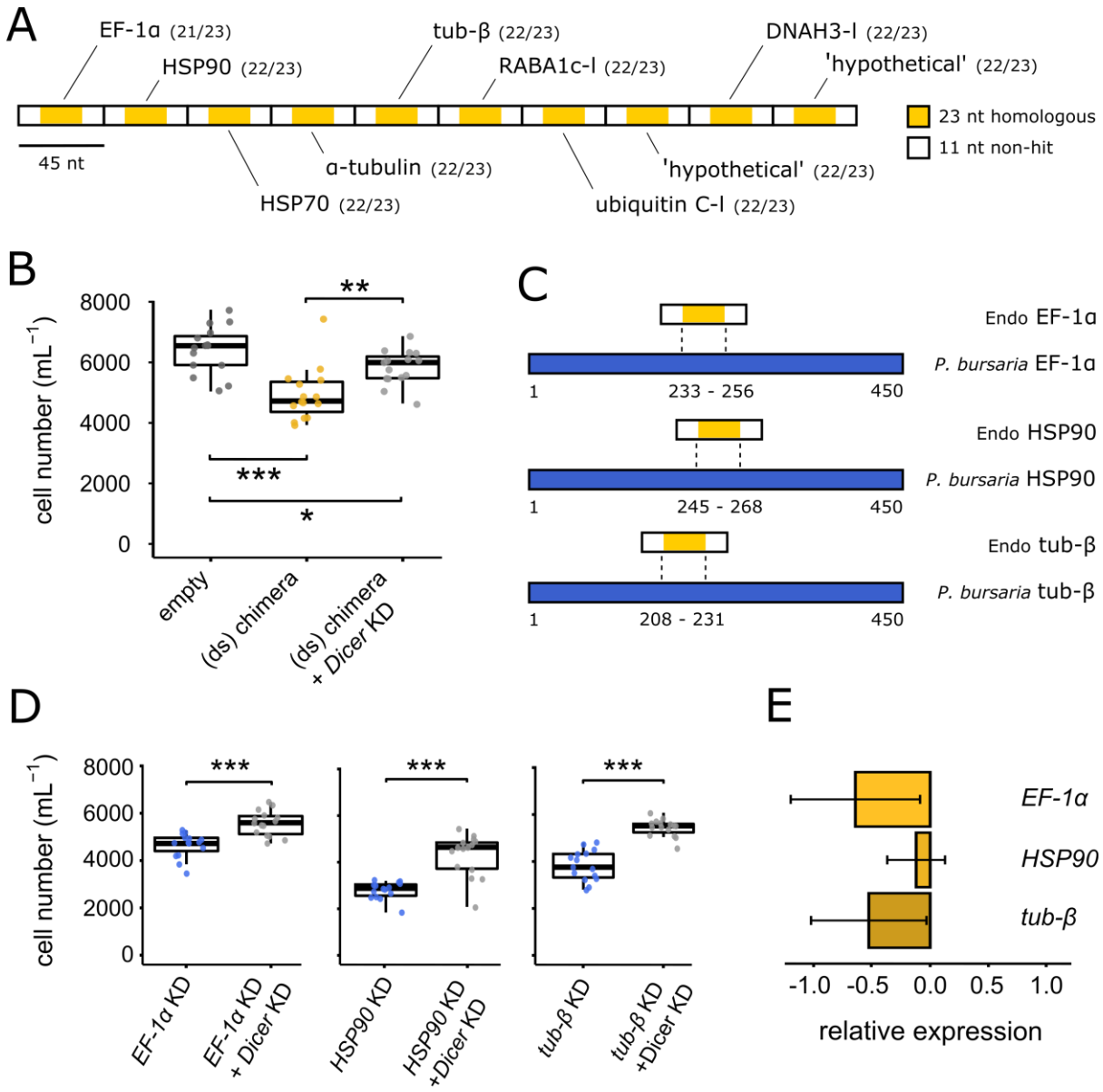


Figure 3. *In Vivo* exposure to a synthetic endosymbiont-derived RNA chimera generates simultaneous knock-down of homologous host genes in *P. bursaria*. (**A**) Schematic showing chimeric construct design. Ten endosymbiont-derived 45-nt transcript sequences (**Table S3**), featuring a 23-nt region with >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 dsRNA (light grey; rescue); or an empty vector control (dark grey). Multiple vector delivery was conducted at a 50:50 ratio during feeding. See **Figure S9** for an experimental dilution of dsRNA chimera delivery. Here and elsewhere, the term 'KD' is used in figure to denote 'knock-down'. (**C**) Schematic figure demonstrating the degree of overlap between three

endosymbiont (endo) derived transcripts from the chimeric construct (white/yellow) and respective 450-nt homologous region of the host (*P. bursaria*) transcript (blue). Each region of 23-nt overlap (yellow) represents putative RNAi 'collisions' between endosymbiont and host. The remaining 11-nt sections are endosymbiont specific and were duplicated for all 10 collision fragments and used to construct the 'nonsense' template. (**D**) *P. bursaria* cell number after 12 days of feeding with *E. coli* expressing 450-nt dsRNA corresponding to *EF1-α*, *HSP90* and *tub-β* host homologues from the *P. bursaria* transcriptome (blue), to determine which dsRNA chimera targeted regions (**3A**/**C**) could facilitate a detectable phenotype using the wider corresponding host gene as an RNAi substrate, compared to Dicer dsRNA mixed controls (grey; rescue phenotype). For individual knock-down of all ten chimera constituent targets using longer regions of the *P. bursaria* homologues, see **Figure S11**. Multiple vector delivery was conducted at a 50:50 ratio during feeding. (**E**) qPCR of mRNA extracted from day 3 of chimera-RNAi feeding (**3B**), revealing knockdown of *EF1-α*, *HSP90* and *tub-β* host gene expression in *P. bursaria* in response to endosymbiont derived chimera dsRNA exposure. Standardised expression of an Actin housekeeping gene was used for normalisation. Data are represented as mean ± SD of three biological replicates. These three genes were assessed for knock-down as a result of endosymbiont derived chimera exposure, based on evidence that directed knock-down of the wider corresponding host gene led to retardation of *P. bursaria* culture growth (**3D**). For an extended figure showing the effect of Dicer knock-down on the expression of these three host genes, see **Figure S12**. All 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 for boxplot data calculated as $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$, using a generalized linear model with quasi-Poisson distribution.

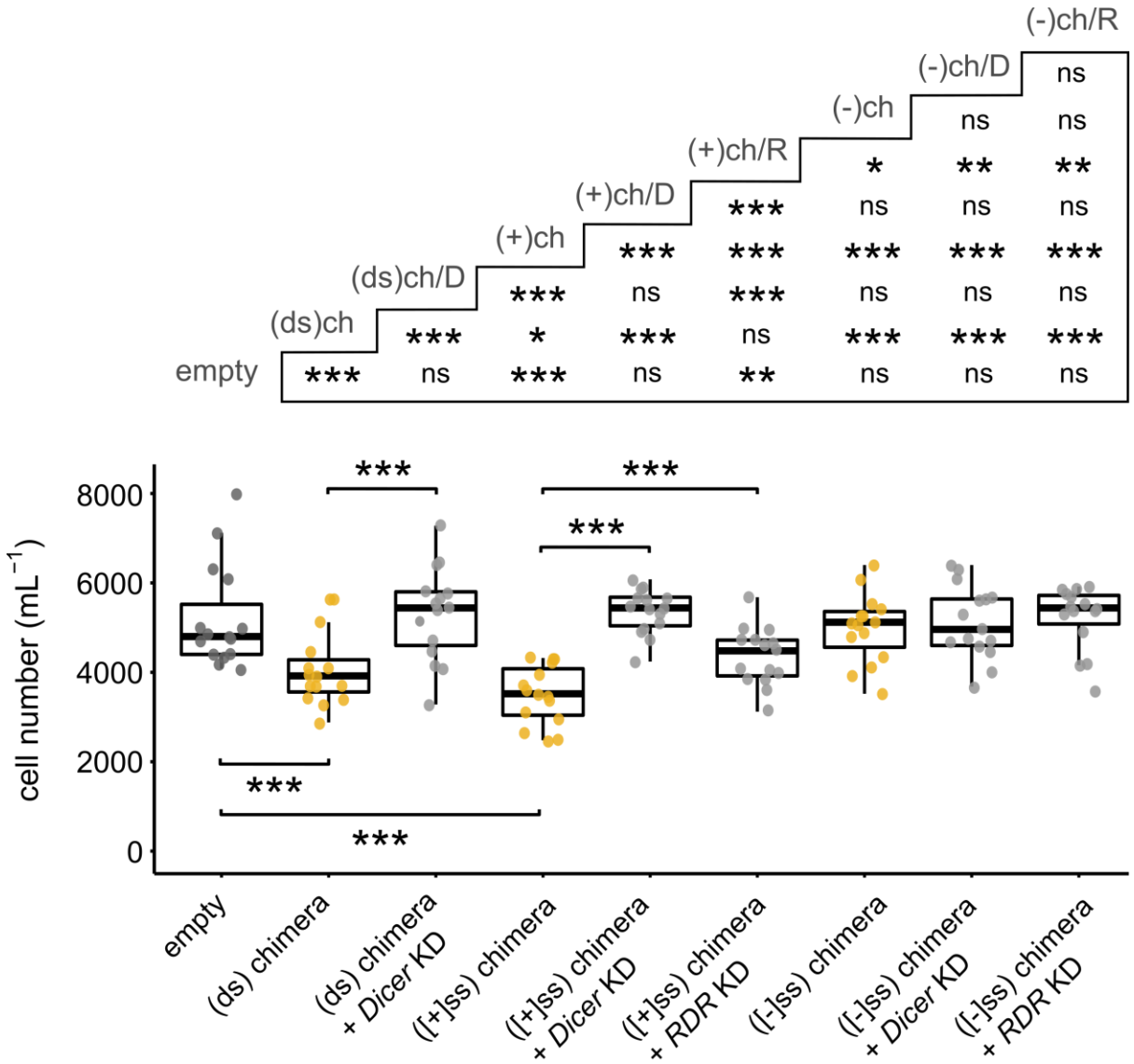


Figure 4. Single-stranded delivery of synthetic endosymbiont-derived RNA, analogous to endosymbiont mRNA, results in a cost to *P. bursaria* growth. *P. bursaria* cell number after 14 days of feeding with *E. coli* expressing; chimera RNA in ds, [+]ss or [-]ss orientation (yellow); chimera RNA in ds, [+]ss or [-]ss orientation mixed with Dicer or RdRP dsRNA (light grey; rescue); or an empty vector control (dark grey). Multiple vector delivery was conducted at a 50:50 ratio of *E. coli* strain exposure during feeding. Asterisks displayed in the grid above denote pairwise significance values. All boxplot data are represented as max, upper quartile (Q3), mean, lower quartile (Q1) and min values of six 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. Here and elsewhere, the term 'KD' is used in figure to denote 'knock-down'.

646 **METHODS**

647 ***Culture conditions and media***

648 In all RNAi experiments, *Paramecium bursaria* 186b (CCAP 1660/18) strain was used. For
649 experiments requiring a non-photo-endosymbiotic *Paramecium* system for comparison,
650 *Paramecium tetraurelia* nd7 strain was used. For eDicer analysis, *Paramecium bursaria* 186b
651 and Yad1g1N strains were used[49].

652 *Paramecium* cells were cultured in New Cereal Leaf – Prescott Liquid media (NCL). NCL
653 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^{-}
654 ^{1} $MgSO_4.7H_2O$ to deionised water. 1 gL^{-1} wheat bran was added, and the solution boiled for 5
655 minutes. Once cooled, media was filtered once through Whatman Grade 1 filter paper and
656 then through Whatman GF/C glass microfiber filter paper. Filtered NCL media was autoclaved
657 at 121°C for 30 mins to sterilise prior to use.

658 NCL medium was bacterized with *Klebsiella pneumoniae* SMC and supplemented with
659 0.8 mgL^{-1} β-sitosterol prior to propagation. *Paramecium* cells were sub-cultured 1:9 into fresh
660 bacterized NCL media once per month for *Paramecium bursaria* 186b, and once every two
661 weeks for *Paramecium tetraurelia* nd7. *Paramecium* cultures were maintained at 18°C with a
662 light-dark (LD) cycle of 12:12h.

663 ***LysoTracker staining and fluorescent imaging***

664 *Paramecium* cells were fixed using 0.5% paraformaldehyde and incubated for 20 mins at RT.
665 Cells were then treated with 2 µM LysoTracker Green DND-26 (Invitrogen) to stain acidic
666 vesicles such as the lysosome, and incubated in constant darkness for 2 hours at RT prior to
667 imaging. Stained *Paramecium* cells were imaged on an ImageXpress Pico Automated Cell
668 Imaging System at 10x magnification, using the Cy5 (absorbance – 630/40 nm, emission –
669 695/45 nm) and FITC (absorbance – 465/40 nm, emission – 525/30 nm) channels to capture
670 algal chlorophyll autofluorescence and LysoTracker Green DND-26 fluorescence respectively.

671 ***Gene synthesis and construct design***

672 Sequences for plasmid constructs were synthesised *de novo* by either Genscript or SynBio
673 Technologies, and cloned into an L4440 plasmid vector. Sequences and cloning sites for each

674 plasmid construct are detailed in **Table S1**. All modified constructs were confirmed by Sanger
675 sequencing (Eurofins Genomics).

676 ***RNAi feeding***

677 *Paramecium* was fed with *E. coli* transformed with an L4440 plasmid construct with paired
678 IPTG-inducible T7 promoters, facilitating targeted gene knock-down through the delivery of
679 complementary double-stranded RNA (dsRNA). L4440 plasmid constructs were transformed
680 into *E. coli* HT115 competent cells and grown overnight on LB agar (50 μgmL^{-1} Ampicillin and
681 12.5 μgmL^{-1} Tetracycline) at 37°C. Positive transformants were picked and grown overnight in
682 LB (50 μgmL^{-1} Ampicillin and 12.5 μgmL^{-1} Tetracycline) at 37°C with shaking (180 rpm).
683 Overnight pre-cultures were back-diluted 1:25 into 50 mL of LB (50 μgmL^{-1} Ampicillin and 12.5
684 μgmL^{-1} Tetracycline) and incubated for a further 2 hours under the same conditions, until an
685 OD_{600} of between 0.4 and 0.6 was reached. *E. coli* cultures were then supplemented with 0.4
686 mM IPTG to induce template expression within the L4440 plasmid, and incubated for a further
687 3 hours under the same conditions. *E. coli* cells were pelleted by centrifugation (3100 x *g* for
688 2 mins), washed with sterile NCL media, and pelleted once more. *E. coli* cells were then re-
689 suspended in NCL media supplemented with 0.4 mM IPTG, 100 μgmL^{-1} Ampicillin, and 0.8
690 μgmL^{-1} β-sitosterol, and adjusted to a final OD_{600} of 0.1.

691 *Paramecium* cells were pelleted by gentle centrifugation in a 96-well plate (10 mins at
692 800 x *g*), taking care not to disturb the cell pellet by leaving 50 µl of supernatant, and re-
693 suspended 1:4 into 200 µl of induced *E. coli* culture media (to make 250 µl total). Feeding was
694 conducted daily for up to 14 days using freshly prepared bacterized media.

695 ***Single-stranded construct design and confirmation***

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

704 antisense [-] orientation. Generated plasmids were transformed into HT115 *E. coli* for use in
705 RNAi feeding experiments.

706 To confirm that these plasmid constructs generated only ssRNA, cultures of *E. coli*
707 HT115 containing pDM005-1 ([-]ssRNA chimera) or pDM005-2 ([+]ssRNA chimera) were
708 grown overnight at 37°C, 180 rpm in LB supplemented with 50 μgmL^{-1} ampicillin and 12.5
709 μgmL^{-1} tetracycline. Cultures were diluted 1:25 in fresh medium and grown to an OD_{600} of 0.4-
710 0.6. Expression was then induced with 400 µM IPTG for 3 hrs, after which 1 mL of culture was
711 pelleted by centrifugation for 2 mins at 3,100 x *g*. RNA was extracted using an RNeasy Mini
712 Kit (Qiagen), following the manufacturer's protocol for Total RNA Purification from Animal
713 Cells. 20 µl of RNA was then treated with 10 μgmL^{-1} RNase A (Sigma-Aldrich) for 1 hour at 30°C
714 in the presence of 300 mM NaCl (stabilising dsRNA and allowing RNase A to degrade only
715 ssRNA[71]). A separate aliquot was left untreated, with 300 mM NaCl added to facilitate
716 precipitation. All samples were then extracted with 1:1 phenol:chloroform and precipitated
717 with 2 volumes of ethanol. Pellets were washed twice with 80% ethanol and re-suspended in
718 nuclease-free water. RNA samples were then cleared of residual genomic DNA using the
719 TURBO DNA-*free* Kit (Ambion), following the manufacturer's protocol for routine DNase
720 treatment. RT-PCR was then performed using the Qiagen OneStep RT-PCR kit following the
721 manufacturer's instructions, with 0.5 µL template RNA and 0.6 µM each primer
722 (pDM005_RT_F: 5'-ACTTCAATGATTCGCAGCGG-3' and pDM005_RT_R: 5'-
723 AAGTAGCTGCTGTTCTCGGT-3'), generating an 85-nt PCR product. Cycling conditions were as
724 detailed in the manufacturer's protocol (30 cycles), with 1 min annealing at 50°C. PCR
725 products were then resolved on a 2% agarose gel to assess for the presence/absence of
726 amplification in each sample (**Figure S11**).

727 ***qPCR analysis***

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

734 the TURBO DNA-*free* Kit (Ambion), following the manufacturer's protocol for routine DNase
735 treatment.

736 RNA was reverse transcribed into single stranded cDNA using the SuperScript® III First-
737 Strand Synthesis SuperMix (Invitrogen), following the manufacturer's protocol. Quantitative
738 PCR (qPCR) was performed in a StepOnePlus Real-Time PCR system (Thermo Fisher Scientific).
739 Reaction conditions were optimised using a gradient PCR, with a standard curve determined
740 using 10-fold dilutions of *P. bursaria* cDNA: *EF-1α* (slope: -3.353; R^2: 0.999; efficiency
741 98.740%), *HSP90* (slope: -3.319; R2: 0.998; efficiency 100.131%), *tub-β* (slope: -3.378; R^2:
742 0.992; efficiency 97.692%), and *actin* (slope: -3.349; R^2: 0.983; efficiency 98.866%), using
743 StepOne software v2.3. Each 20 µL reaction contained 10 µL PowerUp SYBR Green Master
744 Mix (Thermo Fisher Scientific), 500 nM each primer (300 nM for *tub-β*) and 1 µL (50 ng) cDNA.
745 Each reaction was performed in duplicate for each of 3 biological replicates, alongside a 'no-
746 RT' (i.e. non-reverse transcribed RNA) control to detect any genomic DNA contamination.
747 Cycling conditions were as follows: UDG activation, 2 mins at 50°C and DNA polymerase
748 activation, 2 mins at 95°C, followed by 40 cycles of 15 secs, 95°C and 1 min at 55-65°C (*EF-1α*
749 (58°C), *HSP90* (58°C), *tub-β* (57°C) and *actin* (58°C)). Primers pairs for each reaction are listed
750 in **Table S5**. Each reaction was followed by melt-curve analysis, with a 60-95°C temperature
751 gradient (0.3°C s^{-1}), ensuring the presence of only a single amplicon, and ROX was used as a
752 reference dye for calculation of C_T values. C_T values were then used to calculate the change
753 in gene expression of the target gene in RNAi samples relative to control samples, using a
754 derivation of the $2^{-\Delta\Delta CT}$ algorithm[72].

755 ***sRNA isolation and sequencing***

756 Total RNA for sRNA sequencing was extracted from *P. bursaria* (or free-living algal) cultures
757 using TRIzol reagent (Invitrogen), as detailed above. To isolate sRNA from total RNA, samples
758 were size separated on a denaturing 15% TBE-UREA polyacrylamide gel. Gels were prepared
759 with a 15 mL mix with final concentrations of 15% Acrylamide/Bis (19:1), 8M UREA, TBE (89
760 mM Tris, 89 mM Borate, 2 mM EDTA), and the polymerisation started by the addition of 150
761 µL 10% APS (Sigma-Aldrich) and 20 µL TEMED (Sigma-Aldrich). Gels were pre-equilibrated by
762 running for 15 mins (200 V, 30 mA) in TBE before RNA loading. The ladder mix consisted of
763 500 ng ssRNA ladder (50-1000nt, NEB#N0364S), and 5-10 ng of each 21 & 26-nt RNA oligo
764 loaded per lane. The marker and samples were mixed with 2X RNA loading dye (NEB) and heat

765 denatured at 90 °C for 3 mins before snap cooling on ice for 2 min prior to loading. Blank lanes
766 were left between samples/replicates to prevent cross-contamination during band excision.
767 Gels were then run for 50 mins (200V, 30 mA).

768 Once run, gels were stained by shaking (60 rpm) for 20 mins at RT in a 40 mL TBE
769 solution containing 4 µL SYBR® Gold Nucleic Acid Gel Stain. Bands of the desired size range
770 (~15-30 nt) were visualised under blue light, excised and placed into a 0.5 mL tube pierced at
771 the bottom by a 21-gauge needle, resting within a 1.5 mL tube, and centrifuged (16,000 x *g*
772 for 1 min). 400 µL of RNA elution buffer (1M Sodium acetate pH 5.5 and 1mM EDTA) was
773 added to the 1.5 mL tube containing centrifuged gel slurry, and the empty 0.5 mL tube
774 discarded. Gel slurry was manually homogenized until dissolved using a 1 mL sterile plunger
775 and incubated at RT for 2 hours with shaking at 1,400 rpm.

776 Solutions containing RNA elution buffer and gel slurry were transferred to a Costar Spin-
777 X 0.22 µm filter column and centrifuged (16,000 x *g* for 1 min). The filter insert containing
778 acrylamide was discarded. 1 mL of 100% EtOH was added to each solution, alongside 15 µg of
779 GlycoBlue™ Coprecipitant (Invitrogen) to aid sRNA pellet visualisation, and stored overnight
780 at -80°C to precipitate. Precipitated solutions were centrifuged at 4°C (12,000 x *g* for 30 mins),
781 and the supernatant discarded. sRNA pellets were washed with 500 µL of cold 70% EtOH
782 (12,000 x *g* for 15 mins at 4°C), and air dried in a sterile PCR hood for 10 mins, before re-
783 suspending in 15 µL of RNAse-free water and storage at -80°C.

784 ***sRNA-seq and read processing***

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

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

794 ***Assigning sRNAs to the algal endosymbiont transcript bins***

795 Trimmed reads were mapped against the 'endosymbiont' dataset of assembled transcripts
796 using the HISAT2 alignment program with default settings. Post-mapping, the BAM files were
797 processed using SAMTOOLS and a set of custom scripts
798 (https://github.com/guyleonard/paramecium) to produce a table of mapped read accessions
799 and their respective read lengths. Using these exported count tables, of mapped reads per
800 read length, transcripts with >10 hits for read lengths between 21-25 nt were searched using
801 reciprocal BLASTX against the NCBI non-redundant 'nr' proteins sequence database, in order
802 to assign taxonomic identity to each transcript. This allowed filtering of the main
803 'endosymbiont' dataset into subsets corresponding to either: algal cytoplasmic mRNA, algal
804 cytoplasmic rRNA, algal plastid RNA or algal mitochondrial RNA; host RNA contamination;
805 bacterial RNA contamination; or vector RNA contamination. All identified algal cytoplasmic
806 rRNA (28 transcripts), plastid RNA (56 transcripts) or mitochondrial RNA (18 transcripts)
807 sequences were sorted into new datasets representing each RNA species. Host (5 transcripts)
808 and bacterial (34 transcripts) contamination were sorted into respective 'host' or 'bacterial'
809 datasets. Vector (1 transcript) and all unidentifiable sequences (20 transcripts) were
810 separated into a dataset labelled as 'other'. All remaining transcripts with >10 sRNA hits for
811 read sizes 21-25 nt in the original dataset now represented putative algal mRNA transcripts
812 (i.e. for 23-nt reads this represented 3,659 sRNA reads mapping to 148 transcripts). In this
813 algal mRNA dataset, we also included transcripts that fell below the >10 sRNA hit threshold
814 for manual curation, and which therefore may correspond to reads mapping to all of the
815 above bins (i.e. for 23-nt reads this represented 1,949 reads mapping to 605 transcripts in
816 total, of which 1,408 reads were mapping to 468 transcripts which were not manually curated,
817 so could be affected by contamination). This manual binning process was carried out in order
818 to double check the original automated binning post transcriptome assembly, and to check
819 for possible chimeric host-algal transcript sequences produced as a by-product of cDNA
820 synthesis and transcriptome sequencing and assembly. This process collectively allowed
821 accurate segregation of the existing 'endosymbiont' transcript bin into discrete subsets based
822 on RNA 'species'.

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

826 generated from the BAM files, and used to plot a size distribution of 21-25 nt endosymbiont
827 derived sRNA abundance per RNA species. Size distributions of sRNA abundance for each
828 sample were plotted using the R programming language packages; tidyverse, grid.extra and
829 ggplot2 in R Studio.

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

831 To predict putative mRNA-mRNA interactions using eDicer
832 (https://github.com/fmaguire/eDicer), both the 'host' and 'endosymbiont' transcript bins
833 processed from transcriptome data for *P. bursaria* Yad1g1N[67] were further filtered to
834 minimise the risk of false positives for host and endosymbiont cross-comparisons. For the
835 'host' bins, any transcript with >90% ID BLASTN hit to the following genome assemblies were
836 removed: *Chlamydomonas reinhardtii* cc503 cw92 mt, *Chlamydomonas reinhardtii* v3,
837 *Chlorella sorokiniana* 1228 v2, *Chlorella sorokiniana* DOE1412 v3, or *Chlorella sorokiniana*
838 utex1230 lanl v2 assemblies from Los Alamos National Labs Greenhouse genome database.
839 Similarly, for the 'endosymbiont' bins, any transcript with >90% ID BLASTN hit to the following
840 ParameciumDB genome assemblies were removed: *P. biaurelia* V1-4 v1, *P. caudatum* 43c3d
841 v1, *P. bursaria* MAC 110224 v1, *P. decaurelia* MAC 223 v1, *P. dodecaurelia* MAC 274 v1, *P.*
842 *jenningsi* MAC M v1, *P. novaurelia* MAC TE v1, *P. octaurelia* K8 CA1, *P. quadecaurelia* MAC
843 NiA v1, *P. primaurelia* Ir4-2 v1, *P. tetraurelia* MAC 51 with and without IES, or *P. tetraurelia*
844 MAC. Additionally, the reference (refseq) genome of *Escherichia coli* MG1655 (NCBI acc.:
845 NZ_CP012868.1) and *Klebsiella pneumoniae* HS11286 (NCBI acc.: NC_016845.1) were used as
846 'food' comparison datasets.

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

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

857 database[68]. This 'lethal' dataset contained genes known to be conditionally essential in
858 *Saccharomyces cerevisiae*. These putative 'lethal' homologues for each dataset were
859 identified using a tBLASTx search of the yeast 'lethal' database[68] with a gathering threshold
860 set at 1e-10 with a minimum of 50% sequence identity. Once curated, each 'lethal' dataset
861 was subject to eDicer analysis as described above.

862 To predict putative rRNA-rRNA interactions using eDicer, a further analysis was
863 conducted using a dataset consisting of full-length ribosomal RNA (rRNA) clusters for
864 *Microactinium conductrix* [NCBI acc.: ASM224581v2][73], *K. pneumoniae* [NCBI acc.:
865 NC_016845.1][74], *E. coli* [NCBI acc.: NZ_CP012868.1], *P. bursaria* Yad1g1N[67] and *P. bursaria*
866 186b. Once again, each dataset was subject to eDicer analysis as described above. All *in-silico*
867 RNA-RNA interaction predictions were plotted using the R programming language packages;
868 tidyverse, grid.extra and ggplot2 in R Studio.

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

870 In order to identify host transcripts that could be impacted by putative RNA-RNA interactions
871 as a result of endosymbiont derived sRNA exposure, trimmed Illumina reads for all sRNA
872 sequencing samples were mapped to the 'endosymbiont mRNA' transcript dataset. Resulting
873 mapping files were filtered with custom scripts
874 (https://github.com/guyleonard/paramecium) producing tables of mapped hits with their
875 respective read lengths. Reads from all tables, of length 23-nt only, were then extracted with
876 SEQTK (resulting in 3,690 total). A BLASTn search of these 23-nt endosymbiont reads was
877 conducted against the 'host' transcript dataset, to identify host transcripts with ≥ 95% identity
878 over a 23-nt region. This resulted in three candidate host transcripts. These three host
879 transcripts were searched using BLASTx against the NCBI non-redundant 'nr' protein
880 database, resulting in one transcript with 100% sequence identity to both endosymbiont and
881 host HSP90. For the remaining two transcripts, SNPs present in the identified 23-nt mapped
882 sequence indicate that these are putative host sequences with 2-nt of mismatch compared
883 to the respective endosymbiont sequence, and were therefore unlikely to represent a product
884 of RNA-RNA interactions between endosymbiont and host as a result of endosymbiont
885 derived 23-nt sRNA exposure.

886 In order to identify host transcripts that could be classified as 'non-hit' transcripts (i.e.
887 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).

6. Esteban, G. F., Fenchel, T. & Finlay, B. J. Mixotrophy in Ciliates. *Protist* **161**, 621–641 (2010).
7. 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).
8. Gray, M. W. Mitochondrial Evolution. *Cold Spring Harb Perspect Biol* **4**, (2012).
9. Moran, N. A. Symbiosis as an adaptive process and source of phenotypic complexity. *Proc. Natl. Acad. Sci. U.S.A.* **104 Suppl 1**, 8627–8633 (2007).
10. Fujishima, M. & Kodama, Y. Endosymbionts in paramecium. *Eur J Protistol* **48**, 124–37 (2012).
11. 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).
12. Matthews, J. L. *et al.* Partner switching and metabolic flux in a model cnidarian-dinoflagellate symbiosis. *Proc Biol Sci* **285**, (2018).
13. Matsuura, Y. *et al.* Recurrent symbiont recruitment from fungal parasites in cicadas. *PNAS* **115**, E5970–E5979 (2018).
14. Husnik, F. & McCutcheon, J. P. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *PNAS* **113**, E5416–E5424 (2016).
15. Ågren, J. A., Davies, N. G. & Foster, K. R. Enforcement is central to the evolution of cooperation. *Nat Ecol Evol* **3**, 1018–1029 (2019).
16. West, S. A., Kiers, E. T., Simms, E. L. & Denison, R. F. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**, 685–694 (2002).

17. Siegel, R. W. Hereditary endosymbiosis in Paramecium bursaria. *Exp Cell Res* **19**, 239–52 (1960).
18. 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).
19. 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).
20. 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 to localize beneath the host cell membrane after escaping from the host digestive vacuole in the early infection process. *Protoplasma* **231**, 55–63 (2007).
21. Brown, J. A. & Nielsen, P. J. Transfer of photosynthetically produced carbohydrate from endosymbiotic Chlorellae to Paramecium bursaria. *J. Protozool.* **21**, 569–570 (1974).
22. Johnson, M. D. Acquired phototrophy in ciliates: a review of cellular interactions and structural adaptations. *J Eukaryot Microbiol* **58**, 185–95 (2011).
23. Kato, Y. & Imamura, N. Amino acid transport systems of JapaneseParamecium symbiont F36-ZK. *Symbiosis* **47**, 99–107 (2009).
24. Kato, Y. & Imamura, N. Effect of sugars on amino acid transport by symbiotic Chlorella. *Plant Physiol Biochem* **46**, 911–7 (2008).
25. Kawakami, H. & Kawakami, N. Behavior of a Virus in a Symbiotic System, Paramecium bursaria—Zoochlorella. *The Journal of Protozoology* **25**, 217–225 (1978).
26. Parker, R. C. Symbiosis in Paramecium bursaria. *Journal of Experimental Zoology* **46**, 1–12 (1926).

27. 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).
28. 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).
29. 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).
30. 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).
31. Tanaka, M. *et al.* Complete elimination of endosymbiotic algae from Paramecium bursaria and its confirmation by diagnostic PCR. *Acta Protozool* **41**, 255–261 (2002).
32. Lowe, C. D., Minter, E. J., Cameron, D. D. & Brockhurst, M. A. Shining a Light on Exploitative Host Control in a Photosynthetic Endosymbiosis. *Curr Biol* **26**, 207–211 (2016).
33. Sørensen, M. E. S. *et al.* The role of exploitation in the establishment of mutualistic microbial symbioses. *FEMS Microbiol Lett* **366**, (2019).
34. Sørensen, M. E. S., Wood, A. J., Cameron, D. D. & Brockhurst, M. A. Rapid compensatory evolution can rescue low fitness symbioses following partner-switching. *bioRxiv* 2020.11.06.371401 (2020) doi:10.1101/2020.11.06.371401.
35. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).

36. Foster, K. R. & Wenseleers, T. A general model for the evolution of mutualisms. *J. Evol. Biol.* **19**, 1283–1293 (2006).
37. Frank, S. A. *Foundations of Social Evolution*. (Princeton University Press, 1998).
38. Herre, E. A., Knowlton, N., Mueller, U. G. & Rehner, S. A. The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends in Ecology & Evolution* **14**, 49–53 (1999).
39. Szathmáry, E. & Smith, J. M. The major evolutionary transitions. *Nature* **374**, 227–232 (1995).
40. West, S. A., Fisher, R. M., Gardner, A. & Kiers, E. T. Major evolutionary transitions in individuality. *PNAS* **112**, 10112–10119 (2015).
41. Buck, A. H. *et al.* Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nat Commun* **5**, 5488 (2014).
42. Guo, Z., Li, Y. & Ding, S. W. Small RNA-based antimicrobial immunity. *Nat Rev Immunol* **19**, 31–44 (2019).
43. Huang, C. Y., Wang, H., Hu, P., Hamby, R. & Jin, H. Small RNAs - Big Players in Plant-Microbe Interactions. *Cell Host Microbe* **26**, 173–182 (2019).
44. Wang, M. *et al.* Bidirectional cross-kingdom RNAi and fungal uptake of external RNAs confer plant protection. *Nat Plants* **2**, 16151 (2016).
45. Weiberg, A. *et al.* Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science* **342**, 118–23 (2013).
46. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806–11 (1998).
47. Keeling, P. J. & McCutcheon, J. P. Endosymbiosis: The feeling is not mutual. *J. Theor. Biol.* **434**, 75–79 (2017).

48. Oulhen, N., Schulz, B. J. & Carrier, T. J. English translation of Heinrich Anton de Bary's 1878 speech, 'Die Erscheinung der Symbiose' ('De la symbiose'). *Symbiosis* **69**, 131–139 (2016).

49. Jenkins, B. H. *et al. Characterisation of the RNA-interference pathway as a Tool for Genetics in the Nascent Phototrophic Endosymbiosis,* Paramecium bursaria. http://biorxiv.org/lookup/doi/10.1101/2020.12.16.423098 (2020) doi:10.1101/2020.12.16.423098.

50. Carradec, Q. *et al.* Primary and secondary siRNA synthesis triggered by RNAs from food bacteria in the ciliate Paramecium tetraurelia. *Nucleic Acids Res* **43**, 1818–33 (2015).

51. Galvani, A. & Sperling, L. RNA interference by feeding in Paramecium. *Trends Genet* **18**, 11–2 (2002).

52. Karunanithi, S. *et al.* Exogenous RNAi mechanisms contribute to transcriptome adaptation by phased siRNA clusters in Paramecium. *Nucleic Acids Res* **47**, 8036–8049 (2019).

53. Marker, S., Carradec, Q., Tanty, V., Arnaiz, O. & Meyer, E. A forward genetic screen reveals essential and non-essential RNAi factors in Paramecium tetraurelia. *Nucleic Acids Res* **42**, 7268–80 (2014).

54. Marker, S., Le Mouël, A., Meyer, E. & Simon, M. Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in Paramecium tetraurelia. *Nucleic Acids Res* **38**, 4092–107 (2010).

55. Karunanithi, S. *et al.* Feeding exogenous dsRNA interferes with endogenous sRNA accumulation in Paramecium. *DNA research : an international journal for rapid publication of reports on genes and genomes* **27**, (2020).

56. Ayala, A. & Weis, D. S. The Effect of Puromycin and Cycloheximide on the Infection of Algae-Free Paramecium bursaria by Symbiotic Chlorellae1. *The Journal of Protozoology* **34**, 377–381 (1987).

57. Stevens, D. R., Atteia, A., Franzén, L. G. & Purton, S. Cycloheximide resistance conferred by novel mutations in ribosomal protein L41 of Chlamydomonas reinhardtii. *Mol Gen Genet* **264**, 790–5 (2001).

58. Yao, M. C. & Yao, C. H. Transformation of Tetrahymena to cycloheximide resistance with a ribosomal protein gene through sequence replacement. *Proc Natl Acad Sci U S A* **88**, 9493–7 (1991).

59. 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).

60. Hoehener, C., Hug, I. & Nowacki, M. Dicer-like Enzymes with Sequence Cleavage Preferences. *Cell* **173**, 234-247.e7 (2018).

61. Hoshina, R., Kato, Y., Kamako, S. & Imamura, N. Genetic Evidence of "American" and "European" Type Symbiotic Algae of Paramecium bursaria Ehrenberg. *Plant biol (Stuttg)* **7**, 526–532 (2005).

62. Ahmed, F. & Raghava, G. P. S. Designing of Highly Effective Complementary and Mismatch siRNAs for Silencing a Gene. *PLoS One* **6**, (2011).

63. Du, Q., Thonberg, H., Wang, J., Wahlestedt, C. & Liang, Z. A systematic analysis of the silencing effects of an active siRNA at all single-nucleotide mismatched target sites. *Nucleic Acids Res* **33**, 1671–1677 (2005).

64. Jackson, A. L. *et al.* Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* **12**, 1179–1187 (2006).

65. Doench, J. G., Petersen, C. P. & Sharp, P. A. siRNAs can function as miRNAs. *Genes Dev* **17**, 438–442 (2003).

66. Ameres, S. L., Martinez, J. & Schroeder, R. Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* **130**, 101–112 (2007).

67. Kodama, Y. *et al.* Comparison of gene expression of Paramecium bursaria with and without Chlorella variabilis symbionts. *BMC Genomics* **15**, 183 (2014).

68. Cotton, J. A. & McInerney, J. O. Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of function. *Proceedings of the National Academy of Sciences* **107**, 17252–17255 (2010).

69. Hickey, D. A. Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* **101**, 519–531 (1982).

70. Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the legume–rhizobium mutualism. *Nature* **425**, 78–81 (2003).

71. Hansen, D. R., Van Alfen, N. K., Gillies, K. & Powell, W. A. Naked dsRNA Associated with Hypovirulence of Endothia parasitica Is Packaged in Fungal Vesicles. *Journal of General Virology*, **66**, 2605–2614 (1985).

72. Zhang, J. D., Biczok, R. & Ruschhaupt, M. *ddCt: The ddCt Algorithm for the Analysis of Quantitative Real-Time PCR (qRT-PCR)*. (Bioconductor version: Release (3.11), 2020). doi:10.18129/B9.bioc.ddCt.

73. Arriola, M. B. *et al.* Genome sequences of Chlorella sorokiniana UTEX 1602 and Micractinium conductrix SAG 241.80: implications to maltose excretion by a green alga. *The Plant Journal* **93**, 566–586 (2018).

74. Liu, P. *et al.* Complete genome sequence of Klebsiella pneumoniae subsp. pneumoniae HS11286, a multidrug-resistant strain isolated from human sputum. *J. Bacteriol.* **194**, 1841–1842 (2012).

1084