Immune-like glycan-sensing and horizontally-acquired glycan-processing orchestrate host control in a microbial endosymbiosis

Benjamin H. Jenkins[1, 2]*, Estelle S. Kilias[1], Fiona R. Savory[1], Megan E. S. Sørensen[3,4], Camille Poirier[1], Victoria Attah[1], Georgia C. Drew[1], Josephine Blagrave[1], Luis J. Galindo[1], Guy Leonard[1], Duncan D. Cameron[3,5,6], Michael A. Brockhurst[3,7], David S. Milner[1] & Thomas A. Richards[1]*

[1]Department of Biology, University of Oxford, 11a Mansfield Road, Oxford OX1 3SZ, UK

[2]Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, UK

[3]Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

[4]Institute of Microbial Cell Biology, Heinrich Heine University, 40225 Düsseldorf, Germany

[5]Department of Earth and Environmental Sciences, School of Natural Sciences, University of Manchester

[6]Manchester Institute of Biotechnology, University of Manchester, Manchester M1 7DN, UK

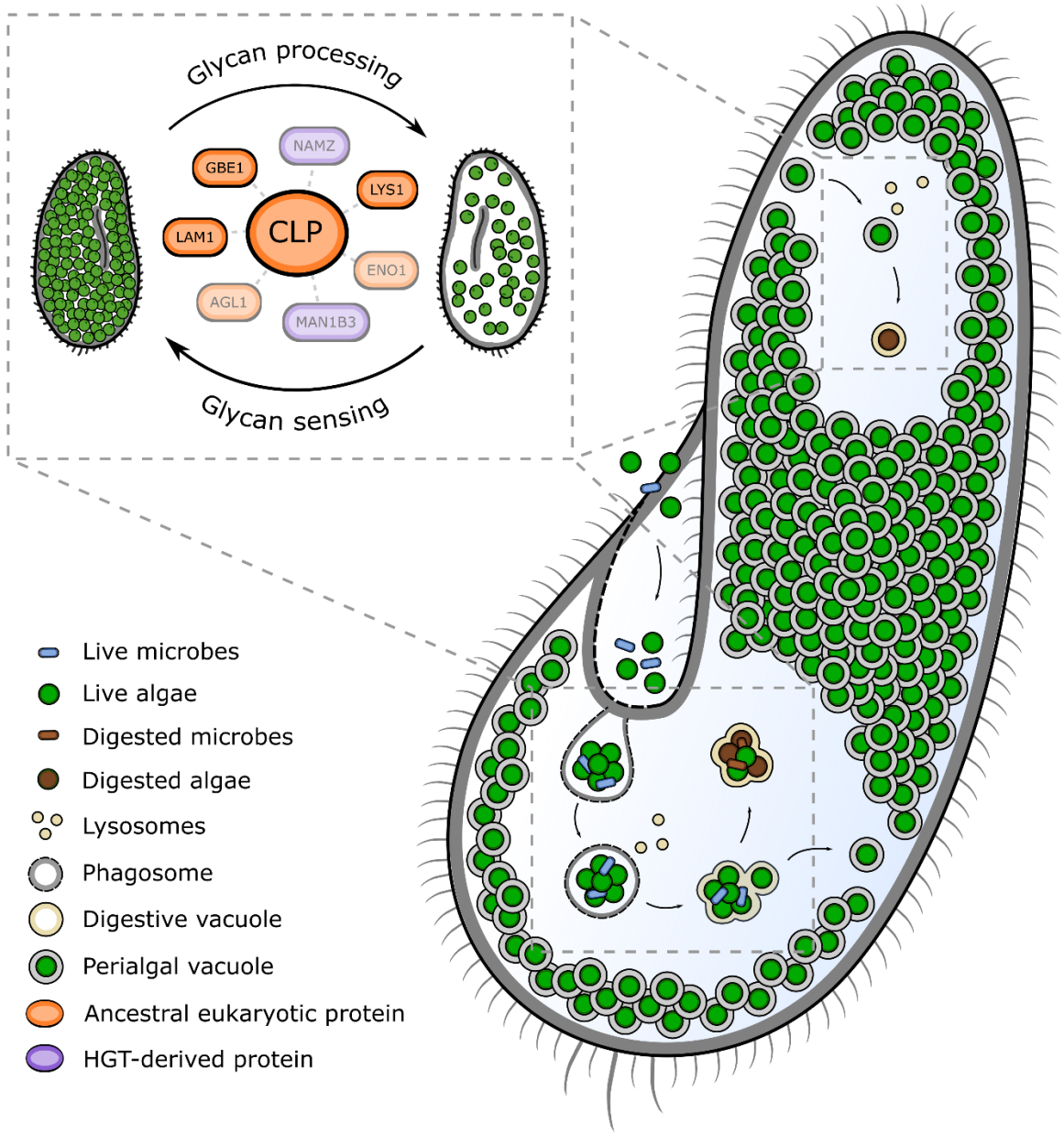
[7]Division of Evolution, Infection and Genomics, School of Biological Sciences, University of Manchester, Manchester M13 9PT, UK

Abstract

Endosymbiosis was a key factor in the evolution of eukaryotic cellular complexity. Yet the mechanisms that allow host regulation of intracellular symbionts, a pre-requisite for stable endosymbiosis and subsequent organelle evolution, are largely unknown. Here, we describe an immune-like glycan-sensing/processing network, partly assembled through horizontal gene-transfers (HGTs), that enables *Paramecium bursaria* to control its green algal endosymbionts. Using phylogenetics, RNA-interference (RNAi), and metabolite exposure experiments, we show that *P. bursaria* regulates endosymbiont destruction using glycan-sensing/processing – a system that includes a eukaryotic-wide chitin-binding chitinase-like protein (*CLP*) localized to the host phago-lysosome. RNAi of *CLP* alters expression of eight host glycan-processing genes, including two prokaryote-derived HGTs, during endosymbiont destruction. Furthermore, glycan-sensing/processing dynamically regulates endosymbiont number in *P. bursaria*, plasticity crucial to maximize host fitness across ecological conditions. *CLP* is homologous to a human phagocyte-associated innate immune factor, revealing how immune functions can be alternatively adapted and expanded, partly through HGT, enabling endosymbiotic control.

[151 words]

53 **Graphical Abstract**



54

55

56

57

58

59

60

61 **Main**

62 Endosymbiosis played a crucial role in the evolution of eukaryotic cellular complexity[1–4] and
63 underpins diverse ecologically and medically-significant host-microbe interactions (e.g., [5–11]).
64 Long-term maintenance of a stable endosymbiosis requires a network of host mechanisms
65 that regulate endosymbiont functions, allowing the host to maximize the fitness benefit of
66 the interaction under fluctuating ecological conditions[12–15]. Orchestrating this network
67 requires extensive genetic and cellular re-modelling by the host; however, the function and
68 evolution of these mechanisms are poorly understood.

69 The ciliate protist, *Paramecium bursaria* (*Pb*), is an RNAi-tractable emerging model
70 organism to study facultative endosymbiotic interactions[16–19]. Each cell maintains hundreds
71 of intracellular green algae from the *Chlorella*-clade (e.g., *Chlorella* & *Micractinium*[20]) in a
72 symbiosis based on photosynthesis-derived metabolic exchange[21–24]. Endosymbiotic algae
73 are housed in modified host phagosomes (perialgal vacuoles) localized beneath the host cell
74 cortex by microtubules[25–27]. *Pb* can regulate endosymbiont load through detachment and
75 lysosomal fusion of the perialgal vacuole to trigger endosymbiont digestion[26–29]. In this
76 manner, *Pb* dynamically controls the endosymbiont population in response to factors such as
77 light intensity to maximize host fitness, often at the expense of endosymbiont fitness[13,29–31].
78 *Chlorella* algae possess cell walls composed of complex glycans (predominantly
79 oligomannosides and chitin[32–34]) which a host must break-down to digest its endosymbionts.
80 Accordingly, the presence of *Chlorella* endosymbionts leads to increased expression of host
81 chitin-processing genes in both *Pb*[35] and *Hydra viridis*[36], a metazoan host which also forms
82 symbioses with *Chlorella* species[36].

83 Glycan processing and sensing requires functionally diverse glycosyl-hydrolases
84 (GHs)[37,38]. GHs are widespread but discontinuously distributed in eukaryotes, with these
85 enzymes performing important structural, nutritional and signaling functions in many
86 ecologically-significant interactions[39–42]. This includes break-down and sensing of glycans
87 derived from hosts, pathogens, symbionts, or food[35,36,39,42–45]. Chitin oligosaccharides, such as
88 the monosaccharide N-acetyl-glucosamine (GlcNAc), are common cues of innate immunity
89 that can trigger both pathogen defense and symbiotic maintenance[42,44–48]. The spread of GHs
90 in eukaryotes has been further facilitated by horizontal gene transfer (HGT), particularly
91 among pathogens of glycan-producing hosts[49–56].

92 Here, we investigate the glycan-sensing and processing network in *Pb*, focusing on a
93 GlcNAc oligosaccharide binding chitinase-like protein (*CLP*) which is homologous to an innate
94 immune-factor protein in humans. We demonstrate using RNAi that host *CLP* perturbation
95 can disrupt endosymbiont destruction and alters the expression of intracellular signaling
96 systems and GH genes, suggesting a role in host sensing cascades and down-stream glycan-
97 processing. We demonstrate that 9 of the 37 GH genes in *Pb* (24%) were acquired via
98 prokaryote-to-eukaryote HGT, including two (a lysozyme – *NAMZ*; and a mannosidase –
99 MAN1B3) with altered expression during *CLP* RNAi perturbation, indicating how HGT can
100 augment ancestral functions of preexisting genes in order to facilitate a host-microbe
101 endosymbiotic interaction. We show that RNAi perturbation of an additional GH gene (a
102 eukaryotic lysozyme – *LYS1*) also alters host control of endosymbiotic algae. We propose that
103 a functional network of immune-like glycan sensing, dependent upon *CLP*, and augmented by
104 HGT-derived glycan processing genes, has been repurposed as a mechanism of endosymbiont
105 control in *Pb* through the regulation of endosymbiont destruction. We discuss how such
106 networks represent a key innovation enabling hosts to control and regulate their intimate
107 associations with microbial endosymbionts.

108 **Results**

109 ***Horizontal gene-transfer has shaped glycan-processing in P. bursaria***

110 Host-microbe interactions are often governed by glycan-processing[44–48]. To identify the
111 repertoire of glycan-processing genes in *Pb* 186b[19], we designed a bioinformatic annotation
112 pipeline for the discovery of carbohydrate-active enzymes[37]
113 (https://github.com/benjaminhjenkins/CAZyme_survey). This identified a total of 37 putative
114 glycosyl-hydrolase (GH) (25 GH families; **Fig. S1a**) and 12 putative glycosyl-transferase (GT)
115 (10 GT families; **Fig. S1b**) encoding genes. Phylogenetic analyses revealed a mosaic
116 distribution of GH and GT genes across the ciliates, including multiple instances of gene
117 duplication and loss (**Supplementary Dataset S1**).

118 Phylogenetic analyses were then conducted to identify candidate horizontal gene-
119 transfers (HGT) among the *Pb* GH genes. These resolved tree topologies for nine of the 37 *Pb*
120 GH genes (24.3%) that were best explained by HGT[57] from within the prokaryotes into a
121 eukaryotic ancestor lineage of *Pb*, given current genome sampling (**Fig. 1** – four HGT events

122 involving five *Pb* genes, discussed further below; see **Supplementary Table S1** and **Fig. S2** for
123 details on the additional four HGTs identified). Four of the putative HGT-derived genes were
124 shown to be acquired specifically in the *Pb* stem lineage (**Fig. 1a/c** & **S2 b/c**) and were present
125 in all five *Pb* strains assessed, ruling out genome project contamination. Comparison of
126 trimmed mean M (TMM) values of transcript abundance in *Pb* 186b[19] (under scramble control
127 RNAi conditions, see below) demonstrates transcription of each putative HGT-derived gene
128 with the exception of *MAN1B2*, indicating possible pseudogenization of *MAN1B2* following
129 *MAN1B1/2* duplication. Each putative HGT-derived gene contained introns (**Supplementary**
130 **Table S1**), including short introns characteristic of *Paramecia*[19]. Furthermore, each proposed
131 open reading frame (ORF) contained between 9 and 43 UAG/UAA codons (representing 2.8%
132 to 5.6% of codons per ORF; **Supplementary Table S1**), which in *Paramecium*[58] code for
133 glutamine rather than stop codons as translated in the universal genetic code used by the
134 prokaryotic progenitor lineage of the HGT. Collectively, these data suggest that the putative
135 HGT-derived genes have been integrated to function within the host eukaryotic genome.

136 The proposed HGT-derived genes were annotated as encoding putative *β-*
137 *mannosidase* (*MAN1B1-3* and *MANEB*; **Fig. 1a** and **Fig. S2d**), *lysozyme* (*NAMZ* and *TTSA*; **Fig.**
138 **1b/c**), *β-glucosidase* (GAB; **Fig. S2a**), *cellulase* (CELA1; **Fig. S2b**), and *β-fructofuranosidase*
139 (INV; **Fig. S2c**) enzymes. Interestingly, the proposed functions of these HGT-derived genes are
140 enriched for cleavage of mannose (*MAN1B1-3*, *MANEB*) and GlcNAc (*NAMZ*, *TTSA*) containing
141 substrates – components of both bacterial food and the endosymbiotic algal cell wall[32–34] –
142 hinting that these HGT-acquired functions are important for mixotrophic functions and
143 endosymbiont control. These data reveal a mosaic distribution of glycan-processing genes in
144 *Pb* and the broader ciliate group that has been augmented by prokaryote-to-eukaryote HGT.

145 ***RNAi of glycan-processing genes perturbs host-mediated endosymbiont destruction, and***
146 ***perturbation of CLP determines expression of glycan-processing genes***

147 Next, we sought to understand whether *Pb* GH genes function in endosymbiont control. We
148 conducted an RNA-interference (RNAi) screen on *Pb* populations grown in microwell plates
149 and treated with cycloheximide to induce endosymbiont destruction[28,29]. Per-host
150 fluorescence (a proxy for endosymbiotic algal load[59]) was measured on an ImageXpress Pico
151 automated cell imaging system, and compared to 'non-hit' RNAi control (scramble)
152 populations to assess the effect of GH transcript perturbation on endosymbiont destruction

153 (**Fig. 2a**). RNAi of *u2af* –a splicing factor in *Pb* required for host growth– was used as a positive
154 RNAi control. The RNAi screen was conducted twice (two biological replicates per treatment
155 with six technical replicate populations each; n = 12) and mean per-host fluorescence data
156 were assessed with a mixed effects model to account for uncontrolled sources of variance.
157 This revealed that 22% of observable variance could be attributed to RNAi-based gene
158 perturbation, 24% could be attributed to random effects (biological replicate and plate
159 positional artifacts), and 54% remained unexplained. Due to the high unexplained variability
160 and considerable influence of confounding factors, we have been conservative in our
161 interpretation of these data and consider the results set out below as trends.

162 RNAi of one GH and one GH family (chimeric RNAi-constructs were designed to target
163 multiple GH paralogues belonging to the same family; **Fig. S3**) led to higher per-host
164 fluorescence compared to scramble controls (control fluorescence = 44.6 relative
165 fluorescence units [RFU]) (**Fig 2a-b**), suggesting that perturbation of these genes stalls
166 endosymbiont destruction. These putative GHs were annotated as a *1,4-α-glucan branching*
167 *enzyme* (*GBE1*) proposed to function in glycogen biosynthesis (fluorescence = 49.3 RFU,
168 increased 10.6%; t=2.2, df=228, p=0.028) and a family of *lysosomal-α-mannosidases* (*LAM*)
169 predicted to cleave mannose (fluorescence = 49.3 RFU, increased 10.6%; t=2.2, df=272,
170 p=0.030). RNAi of two other GHs led to greater reduction in per-host fluorescence compared
171 to scramble controls (**Fig. 2a-b**), suggesting that perturbation of these genes may increase
172 endosymbiont destruction. These included a putative *chitinase-like protein* (*CLP*) predicted to
173 bind GlcNAc oligosaccharides (fluorescence = 38.3 RFU, reduced 14.2%; t=-3.1, df=291,
174 p=0.002) and a putative *Lys1-like lysozyme* (*LYS1*) predicted to cleave GlcNAc-GlcNAc (e.g.,
175 chitin) or GlcNAc-MurNAc (e.g., peptidoglycan) containing substrates (fluorescence = 41.0
176 RFU, reduced 8.1%; t=-1.97, df=290, p=0.050). The greater reduction in fluorescence observed
177 upon *CLP* RNAi is intriguing because an orthologue of this gene binds to glycan moieties such
178 as those present in *Chlorella* cell walls, and has been suggested to play a role in the sensing
179 of chitin-containing microbes, amongst other functions[60–62] (described in more detail below).
180 We hypothesize that CLP may act as a binding protein which functions in a glycan
181 processing/sensing signaling cascade involved in the regulation of endosymbiont load and
182 therefore sought to further explore the role of *CLP* during endosymbiont destruction in *Pb*.

183 If *CLP* functions in a signaling pathway which orchestrates endosymbiont control, *CLP*
184 RNAi perturbation should influence the transcription of other *Pb* encoded GH genes. To
185 explore whether GH function was altered during endosymbiont destruction in a *CLP*-
186 dependent manner, we assessed GH expression in cycloheximide treated and untreated
187 control cultures, and compared changes in expression under *CLP* RNAi or scramble control
188 RNAi conditions. Ensemble RNA sequencing was conducted under each condition and reads
189 were mapped to the *Pb* 186b genome[19]. To confirm *CLP* perturbation in response to RNAi,
190 reads were mapped only to the 3' *CLP* transcript end to avoid cross-talk from the *CLP* RNAi
191 construct (**Fig. 2c**). Importantly, *CLP* RNAi in untreated cultures resulted in 2.5-fold over-
192 expression of *CLP* (**Fig. 2d**), suggesting either that compensatory gene expression was
193 occurring in response to increased *CLP* dsRNA and subsequent siRNA exposure[16,63], and/or
194 that the RNAi signal (siRNA generation) had spread along the *CLP* transcript due to RNA-
195 dependent RNA-polymerase amplification of cleaved mRNA transcripts that are over-
196 represented in the mapped reads[16,64]. Both outcomes are known phenomena in *Paramecium*
197 RNAi experiments[16,63,64] and could interact, producing complexity in terms of phenotypic
198 outcome. Regardless, *CLP* RNAi resulted in significant perturbation in *CLP* transcript
199 abundance (**Fig. 2d**).

200 *CLP* RNAi during endosymbiont destruction significantly altered the expression of 8
201 (24%) GH genes (Bonferroni corrected since these did not test an *a priori* hypothesis; $p <$
202 0.0015) (**Fig. 2e**; **Fig. S4**). These include two putative GlcNAc-processing genes (*LYS1*; **Fig. 2a**,
203 and *NAMZ*; **Fig. 1b**) and two putative HGT-derived genes (aforementioned *NAMZ*, and
204 *MAN1B3*; **Fig. 1a**). The most significant changes in directionality of expression fold-change
205 during endosymbiont destruction, resulting from *CLP* RNAi, were observed for *LYS1*
206 (putatively involved in endosymbiont destruction; **Fig. 2a**), *MAN1B3* (a putative HGT-derived
207 gene predicted to cleave mannose; **Fig. 1b**), and *LAM1* (within a GH family putatively involved
208 in endosymbiont destruction; **Fig. 2a**). Interestingly, all three additional genes implicated in
209 endosymbiont control upon RNAi (*LYS1, GBE1* and *LAM1;* **Fig. 2a**) were significantly altered
210 upon *CLP* RNAi during endosymbiont destruction. These data demonstrate that a functional
211 network of glycan-processing genes, at least in part controlled by *CLP*, acts in the regulation
212 of endosymbiont destruction in *Pb*.

213 ***CLP RNAi alters expression of transcription factors and intracellular signaling systems***

214 To explore the broader role of *CLP* in *Pb*, we investigated how perturbed *CLP* function
215 influenced additional pathways triggered during endosymbiont destruction. Untargeted
216 metabolomics of *Pb* during *CLP* RNAi, compared to scramble control RNAi, revealed
217 enrichment in metabolites belonging to cell signaling (inositol phosphate, and
218 phosphatidylinositol signaling) and carbohydrate metabolic pathways (**Fig. 3a**; see also
219 **Supplementary Table S3**). A significant reduction in relative abundance of several metabolites
220 was detected upon *CLP* RNAi, including inositol 1,3,4-trisphosphate, fructose 1,6-
221 bisphosphate, and unspecified glucan disaccharides, which may include maltose, the product
222 of photosynthesis released by the algal endosymbionts[21–23] (**Fig. 3b**; **Fig. S5**).

223 We then assessed the expression of all *Pb* transcripts during endosymbiont
224 destruction (cycloheximide treatment) and *CLP* RNAi, compared to both untreated and
225 scramble control RNAi conditions (as per **Fig. 2e**). This identified distinct transcriptional
226 profiles for each condition (**Fig. 3c**), with clear separation observed between untreated and
227 cycloheximide treated cultures under scramble control RNAi conditions. A less distinct
228 separation profile was observed between untreated and cycloheximide treated cultures
229 during *CLP* RNAi, consistent with our hypothesis that functions associated with endosymbiont
230 destruction may be coordinated by *CLP*. To explore transcriptional patterns associated with
231 each treatment, we performed differential expression analysis for each pairwise condition.
232 Differentially expressed transcripts ([$\log_2$ fold change] ≥ 1, adjusted p-value < 0.001) were
233 hierarchically clustered based on relative expression under each condition (**Fig. 3d; Fig. S6**),
234 yielding three distinct transcript clusters (**Fig. 3e**; see also **Supplementary Table S4**).

235 Cluster 1 contains transcripts that are up-regulated during cycloheximide treatment,
236 independently of *CLP* RNAi (**Fig. 3f**), and includes genes related to autophagy (ser/thr-protein
237 kinase Atg1), intracellular destruction and transport (lipocalin, lipase), carbohydrate
238 metabolism (fructose-1,6-bisphosphatase; consistent with the mass spectrometry analysis,
239 **Fig. 3b**), and oxidative stress (glutathione peroxidase). We propose that these are genes
240 involved in the phago-lysosomal processes activated during destruction of algal
241 endosymbionts[25–29].

242 Cluster 2 contains transcripts predominantly down-regulated during cycloheximide
243 treatment in scramble control cultures (**Fig. 3g**), and so represent host functions possibly
244 inactivated during endosymbiont destruction. Interestingly, these transcriptional patterns

245 were not observed upon cycloheximide treatment coupled with *CLP* RNAi (**Fig. S6c-d**),
246 supporting the proposition that *CLP* RNAi can perturb endosymbiont control (seen in **Fig. 2a**).
247 This cluster includes genes predicted to function in signal transduction, and two putative
248 phosphatidylinositol binding proteins (again consistent with the metabolomics data; **Fig. 3a**)
249 indicating that phosphatidylinositol signaling is altered during regulation of endosymbiont
250 destruction. Phosphoinositides – a broad class of membrane phospholipids – are key
251 regulators of endosomal processes and downstream signals of extracellular receptor
252 activation, acting in intracellular membrane trafficking, autophagy, and innate immunity[65–69].
253 Other genes in this cluster were predicted to be involved in intracellular vesicular trafficking
254 (vacuolar sorting-associated proteins 9 & 12, and adaptin-like proteins) and cytoskeletal
255 arrangement (dynein heavy chain, and tubulin-tyrosine ligase/polyglutamylase). We propose
256 that these genes function in the cellular re-arrangements that maintain endosymbiosis in *Pb*,
257 and may represent processes that are inactivated during endosymbiont destruction, such as
258 detachment of the perialgal vacuole from the host cell cortex and lysosomal fusion[25–29]. The
259 observed down-regulation of these genes during algal destruction is consistent with the
260 proposal that these systems function in maintenance of the endosymbiotic population.

261 Cluster 3 contains transcripts with significantly increased expression during *CLP*
262 perturbation (**Fig. 3h**), irrespective of cycloheximide-induced endosymbiont destruction (as
263 in Clusters 1 and 2). This includes genes predicted to function in signal transduction, including
264 phosphatidylinositol signaling (as in **Fig. 3g**), serine-threonine protein kinases, transcription
265 factors, and a transcriptional activator. We therefore propose that these may be components
266 of a signaling cascade in which *CLP* is a component, and suggest that disruption of this signal
267 through *CLP* RNAi may be responsible for the increased endosymbiont destruction trend
268 observed during cycloheximide treatment (**Fig. 2a**). Taken together, these findings indicate
269 that complex signaling mechanisms are involved in host-mediated endosymbiont destruction,
270 and suggest that *CLP* may act as a key regulator of these molecular processes.

271 ***CLP localizes proximate to the putative host phagosome and lysosome during endosymbiont***
272 ***destruction***

273 To understand the role of *CLP* in endosymbiont regulation, we performed
274 immunofluorescence microscopy to track CLP localization during cycloheximide-induced
275 endosymbiont destruction. Using a custom polyclonal antibody raised against *Pb* CLP (see

276 **Methods**), we observed CLP localization to the host phagosome and oral apparatus (**Fig. 4a;**
277 see also **Supplementary Table S5**). These cell systems are part of the ciliate feeding grove[70,71]
278 and are associated with both heterotrophic feeding and establishment/regulation of
279 endosymbiotic algae[17,18]. A large fraction of the cells exhibited phagosome-like CLP
280 localization (78.4% of cells in untreated conditions, 85.7% in cycloheximide treated).
281 However, some cells demonstrated an additional pattern of CLP localization consistent with
282 host lysosomes (32.4% of cells in untreated conditions, 40.3% in cycloheximide treated; **Fig.**
283 **4b**). These observations indicate that *CLP* is involved in at least two systems (phagosomal and
284 lysosomal) which have complex overlapping functions. These data also demonstrate that the
285 sampled population includes cells which are at mixed stages of response, as identified by
286 differential CLP localization patterns. This result is consistent with host cell and associated
287 endosymbiotic functions not being synchronized (e.g. only 32.4% of cells demonstrating CLP-
288 lysosome localization and 78.4% of cells demonstrating CLP-phagosome localization under
289 standard conditions; **Supplementary Table S5**).

290 To explore how CLP localization is adjusted during endosymbiont destruction, we
291 compared algal autofluorescence and CLP immunofluorescence in *Pb* during cycloheximide-
292 induced endosymbiont destruction. A significant reduction in per-host algal fluorescence
293 during cycloheximide treatment (**Fig. 4c**) was coupled with increased CLP fluorescence (**Fig.**
294 **4d**). This trend was observed for both phagosome and phago-lysosome-like exhibiting cells,
295 indicating that endosymbiont destruction is coupled to *CLP* dynamics in both cellular
296 processes. These data support the hypothesis that *CLP* is functioning as a sensor for the
297 detection of endosymbiont (**Fig. 4b**) or food-derived (**Fig. 4a**) glycan-substrates released
298 during endosymbiont destruction and heterotrophic feeding, and is consistent with data
299 suggesting the *LYS1* lysozyme (predicted to liberate these substrates and which is regulated
300 by *CLP*) is involved in endosymbiont destruction (**Fig. 3a, e**).

301 ***CLP is a chitin-binding protein homologous to an animal immune factor***

302 To further explore the function of *CLP* in *Pb,* we sought to understand the evolution of this
303 gene family across the eukaryotes. A single homologue of *Pb CLP*, *Stabilin-interacting*
304 *Chitinase-like Protein* (*SI-CLP*), was identified in the human genome. *SI-CLP* is an innate
305 immune-factor expressed in macrophages and neutrophils that activates signal-cascades
306 involved in inflammatory responses[62] (including phosphoinositide kinases[72]; **Fig. 3a**, **f-h**), and

307 is sorted into CD63+/LAMP1-positive lysosomes and p62lck-positive late endosomes in
308 mammals, indicating a role in secretory, digestive and autophagy-related intracellular
309 trafficking[61]. *SI-CLP* has broad substrate-binding capacity that includes GlcNAc
310 oligosaccharides and other glycan moieties, and may play a sensory role in host responses to
311 chitin-containing pathogens[60]. No homologue for *stabilin-1*, with which human *SI-CLP*
312 interacts, was identified in *Pb* or other ciliates (NCBI protein database; gathering threshold:
313 e-01; June 2024) and so this protein was not explored further.

314 Phylogenetic reconstruction of 1,445 candidate homologues reveals that *CLP* is
315 present in the majority of extant eukaryotic phyla searched (**Supplementary Dataset S2**)
316 demonstrating that this is an anciently-derived eukaryotic gene family. An amino acid
317 alignment of representative eukaryotic *CLP* homologues (**Fig. 5a**; **Fig. S7**) revealed strong
318 conservation of binding sites (all but one retained in 90% of the homologues aligned) and
319 predicts the presence of a 5' secretion signal peptide sequence in 75% of the sequences
320 sampled (DeepLoc2.0)[73]. Identifiable chitin cleavage sites[60,61] were not conserved, with
321 universal absence of the 2nd and 4th known catalytic residues, and 78% of species compared
322 lacking all but one catalytic residue. Putative binding and cleavage sites were mapped to the
323 predicted structure of *Pb* CLP (AlphaFold3; May 2024) revealing a pore-like structure in which
324 GlcNAc oligosaccharide binding is proposed to occur (**Fig. 5b**). Comparison with the crystal
325 structure of *H. sapiens (Hs)* SI-CLP[60] (which forms a dimer; **Fig. S8**) reveals positional similarity
326 of binding and cleavage sites. Western blot analysis of *Pb* CLP reveals two bands close to 37
327 kDa and 75 kDa (predicted molecular weight of *Pb* CLP, 40.24 kDa; see **Fig. S9**) indicating that
328 the *Pb* homologue may also form a dimer.

329 *Pb CLP* and *Hs SI-CLP* lack the conserved 'DxxDxDxE' motif associated with chitin-
330 cleavage (**Fig. 5c**). To understand catalytic function, *Pb* CLP engineered to carry a
331 *Saccharomyces* secretion peptide was expressed in a chitinase-deficient *Saccharomyces*
332 *cerevisiae* strain, and the secretome assayed using a Chitinase Activity Kit (**Fig. 5d**). We
333 observed low catalytic activity of *Pb* CLP on all substrates compared to *Pb* HEXB (predicted to
334 cleave terminal GlcNAc residues from glycan chains; **Fig. 2b**). It is important to note that low
335 activity of all enzymes, including HEXB, may be due to the required assay conditions (37°C)
336 being sub-optimal for *Pb* enzymes (~18-25°C). Nonetheless, we observed that directed
337 mutation of the 2nd and 4th active site residues of *Pb CLP* (*CLPΔ;* **Fig. 5c**) 'restored' cleavage

338 function and significantly increased activity on 2mer ($GlcNAc_2$ / $Chitin_2$) (t=3.141; df=6;
339 p=0.020) and 3mer ($GlcNAc_3$ / $Chitin_3$) (t=3.324; df=6; p=0.021) substrates (**Fig. 5d**),
340 supporting the hypothesis that these residues are important for catalytic activity. These data
341 indicate that *CLP* is an ancient eukaryotic gene which functions in GlcNAc oligosaccharide
342 (e.g., chitin) binding, but lacks potent cleavage activity, supporting our hypothesis that *CLP*
343 plays a role in glycan sensing.

344 ***Glycan break-down products alter host control of endosymbiotic algae***

345 Our genomic and RNAi data strongly point to glycan-sensing/processing being central to host-
346 mediated endosymbiont control. As such, we next explored how glycan break-down
347 substrates were altered during regulation of endosymbiosis in *Pb*, and the reciprocal effect of
348 these substrates on endosymbiont control. Here, we used altered light intensity to trigger
349 endosymbiont destruction[13,30] to simulate conditions experienced by *Pb* in nature. We
350 observed a reduction in *Pb* per-host fluorescence upon exposure to high-light (80 μ mol m^{-2} s^{-}
351 ^{1}; **Fig. 6a**). This was accompanied by no comparable reduction in host cell number (**Fig. 6b**)
352 indicating that these conditions did not have a negative effect on host viability during the
353 period of the experiment. Analysis of metabolite abundance in *Pb* under varying light
354 conditions (50 μ mol m^{-2} s^{-1} to 6 μ mol m^{-2} s^{-1})[30] revealed that reduction in endosymbiont load
355 was accompanied by increased abundance of GlcNAc and glucosamine (GlcN – a substrate
356 which also binds to *SI-CLP*[60]) (**Fig. 6c**; **Fig. S10**). These substrates are consistent with break-
357 down products of chitin-like glycans[32–34] (**Fig. 6d**) and support the hypothesis that host control
358 of endosymbiotic *Chlorella*[13,30] is facilitated through break-down of glycans present in the
359 algal cell wall. This is also consistent with the predicted function and involvement of *LYS1* (**Fig.
360 2a/e**) and *NAMZ* (**Fig. 2e**) in endosymbiont destruction under *CLP*-derived transcriptional
361 control.

362 Given the putative role of *CLP* as a glycan-sensor, we next tested how increased glycan
363 abundance influenced endosymbiont control in *Pb*. We observed increased per-host
364 fluorescence upon exposure to chitin, GlcNAc, and GlcN under standard growth conditions
365 (**Fig. 6e**). This effect was greatest upon exposure to GlcNAc and remained significant at lower
366 doses (0.1 mM; t=4.314; df=20; p=1.34^{-04}), indicating a heightened sensitivity to this
367 substrate. Supplementing free-living endosymbiosis-derived algae with GlcNAc did not enable
368 heterotrophic growth or improve autotrophic growth compared to nitrogen free conditions

369 (**Fig. S11a/b**), demonstrating that this was not due to an algal growth response. Host exposure
370 to chitin oligomers (t=18.860-24.805; df=50; $p < 2e^{-6}$), GlcNAc (t=12.896; df=50; $p < 2e^{-6}$), N-
371 acetyl-d-glucosamine-6-phosphate (GlcNAc-6P) (t=3.922; df=50; p=0.004), and GlcN (t=3.5;
372 df=50; p=0.010) during high-light treatment all significantly arrested the reduction in per-host
373 fluorescence associated with endosymbiont destruction (**Fig. 6f**). Exposures were conducted
374 at equal molar concentrations indicating that the length of the chitin polymer chain may be
375 an important factor. No response was observed for D-glucosamine-6-phosphate (GlcN-6P) or
376 glucose (Glc), indicating that this effect was not due to increased nutrient provision (carbon,
377 nitrogen, or phosphate) to the host. Importantly, these GlcNAc mono- and oligosaccharides
378 are consistent with substrates bound by *CLP* (**Fig. 5d**), and putative products of GlcNAc-
379 processing GHs (e.g., *LYS1*) implicated in *CLP*-dependent endosymbiont destruction (**Fig.
380 2a/e**).

381 Finally, we sought to explore how broader mixotrophic functions could influence
382 endosymbiont control. *Pb* is a mixotroph and phagocytoses bacteria in nature, therefore the
383 host must be exposed to additional sources of glycans. We observed that exposure to heat-
384 killed gram-positive bacteria, possessing exposed peptidoglycan-rich cell walls, also increased
385 algal fluorescence per-host (t=4.492; df=75 ; p=2.51e^{-05}) compared to heat-killed gram-
386 negative bacteria delivered at the same dose (**Fig. 6g**; **Fig. S12**). This suggests that broader
387 microbial trophic interactions, including mixotrophic feeding, may also influence the
388 dynamics of endosymbiosis in *Pb.*

389 These data have revealed that exposure to GlcNAc, and other break-down products of
390 glycans present in endosymbiotic algae and the cell walls of bacterial food, can alter
391 endosymbiont control in *Pb.* We propose that *CLP*, which functions to bind these substrates,
392 may act to regulate the broader cellular network that underpins this process. We therefore
393 postulate that glycan-sensing and processing represents an important mechanism of
394 endosymbiont control in *Pb* through the regulation of endosymbiont destruction.

395 **Discussion**

396 Here we show that an immune-like glycan-sensing/processing network involving a conserved
397 eukaryotic chitinase-like protein (*CLP*), and augmented by genes acquired through HGT,
398 orchestrates the regulation of endosymbiosis in *Pb*. We demonstrate that host-mediated

399 destruction of endosymbiotic algae releases glycan-substrates (**Fig. 6c**), specifically GlcNAc
400 and GlcN, and that environmental exposure to GlcNAc mono- and oligosaccharides stalls
401 endosymbiont destruction (**Fig. 6f**). Through comparative phylogenetics we reveal that *CLP* is
402 related to an innate immune-factor in humans (*SI-CLP*) involved in phago-lysosomal
403 inflammatory response pathways and possible molecular-sensing of pathogens[60–62], and
404 show that *CLP* in *Pb* serves a similar function to coordinate phago-lysosomal activity. *CLP*
405 localizes to the oral apparatus, phagosome, and lysosome of *Pb* with increasing localization
406 intensity during endosymbiont destruction (**Fig. 4d**), where GlcNAc-substrates arise as they
407 are liberated through endosymbiont destruction (**Fig. 6c**) or bacterial feeding (**Fig. 6g**). We
408 propose that these substrates are acted upon by at least six ancient eukaryotic genes (e.g.,
409 *LYS1* - a GH that directly determines the fate of algal endosymbionts*;* **Fig. 2a**) and two
410 horizontally-acquired glycan-processing genes (e.g., *MAN1B1, NAMZ;* **Fig. 1b**) which are
411 transcriptionally regulated by *CLP* function (**Fig. 2e**). We hypothesize that this mechanism of
412 glycan-sensing through *CLP* triggers a network of cellular processes involving transcriptional
413 regulation, intracellular signaling (**Fig. 3f**-**h**) and glycan-processing (**Fig. 2e**) to regulate
414 endosymbiont destruction in *Pb*.

415 Intriguingly, this host network of cellular processes includes genes acquired by *Pb*
416 through prokaryote-to-eukaryote HGT. Among the genes altered through *CLP* perturbation
417 during endosymbiont destruction are two genes proposed to be HGT-derived. *MAN1B3* (**Fig.**
418 **2a**), a predicted β-mannosidase, may be important for destruction of mannose-substrates in
419 the endosymbiont cell wall. However, *NAMZ* (**Fig. 2b**), a predicted lysozyme, is potentially the
420 most relevant, as this enzyme is proposed to cleave GlcNAc in chitin (i.e., from the
421 endosymbionts cell wall) or peptidoglycan (i.e., from the bacterial cell wall) substrates. The
422 interaction of predicted cleavage products of HGT-derived *NAMZ* (i.e., GlcNAc) with *CLP* (**Fig.**
423 **5d**) and their effect on endosymbiont destruction dynamics in the host (**Fig. 6e, f**), combined
424 with the finding that *CLP* perturbation influences subsequent expression of *NAMZ* (**Fig. 2e**),
425 suggests that this may represent part of the feedback loop of glycan-processing and sensing
426 proposed to occur during endosymbiont destruction. These observations demonstrate how
427 recently acquired gene functions (e.g., *NAMZ*) following HGT can augment ancestral functions
428 (*CLP*).

429 'You are what you eat' is a hypothesis which posits that phagocytosis can drive HGT,
430 from consumed cells to host nuclei, through establishment of a gene-transfer ratchet[74]. This
431 process has been posited as an important factor in endosymbiotic organelle acquisition during
432 eukaryotic evolution. The 'shopping bag' model went further, stating that the step-wise
433 acquisition of genes from former, phagocytosed cells (i.e., food), sometimes in transient
434 endosymbiosis, built the host genetic repertoire required for endosymbiotic functions that,
435 in turn, allows for stable symbiont integration[75,76]. This process is thought to be key for the
436 evolution of plastid (chloroplast) organelles (i.e., eukaryotic photosymbiosis). In *Pb*, we
437 observe data consistent with both hypotheses relating to the function of a facultative
438 endosymbiosis, but with a fundamental addition. Here, the genes acquired by *Pb* arose not
439 from transient interactions with eukaryotic algae, but from bacteria. The occurrence of
440 bacterial genes in eukaryotic genomes is not uncommon[54], though few cases have been
441 reported in ciliates[77–80] and also involve the transfer of GH genes, supporting the idea that
442 genes that function in glycan-processing are prime candidates for prokaryote-to-ciliate HGT.
443 *Paramecium* are well known for their ability to phagocytose or form stable symbiotic
444 interactions with a diverse consortium of bacterial species[81–84]. Thus, we hypothesize that
445 these genes likely originated from ancestral bacteria maintained either stably or transiently
446 in endosymbiosis[85,86], or ingested as food[74,76]. Regardless, our discovery that these genes of
447 bacterial origin may be involved in endosymbiotic algal destruction in *Pb* illuminates how
448 cross-kingdom HGT can shape host-microbe endosymbiotic interactions consistent with both
449 the 'you are what you eat' and the 'shopping bag' hypotheses.

450 The role of innate immunity in plant-microbe and animal-microbe symbioses is well-
451 understood[44,46,47,87], however such mechanisms have yet to be described in a single-celled
452 endosymbiotic system[88,89]. *Pb* lacks many components of a 'canonical' immune system
453 (**Supplementary Table S6**). Yet detection of self from non-self must be critical to allow the
454 host to defend against harmful microbial pathogens[90–92], and mechanisms of molecular
455 sensing that can determine friend from foe, or 'good' symbionts from 'bad', are essential for
456 a host maintaining a population of intracellular symbionts[5,7,12–15,87,93]. This is important when
457 the fitness interests of symbiont and host are not aligned, resulting in conflict between
458 partners – an inevitability in symbiosis[94]– and whereby detection and resolution of conflict
459 are needed for long-term stability of an interaction[13,29,31,95]. In *Pb*, maintenance of a

460 population of endosymbiotic algae requires significant re-modelling and co-opting of host
461 cellular machinery[18,25–27]. As such, a finely tuned network of molecular processes is required
462 that allows the host to detect context-dependent cues, and respond to, initiate, maintain, or
463 terminate the interaction. Here, we propose that this process occurs through the sensing of
464 glycan-substrates released during endosymbiont destruction or microbial feeding (**Fig. 6a-f**),
465 detected by *CLP*, and subsequent signaling that prompts the host to stall further
466 endosymbiont destruction. This process would act to minimize the costly effects of mass
467 endosymbiont destruction[29], perpetuating surviving endosymbionts, and promoting long-
468 term stability of the interaction. Our discovery highlights how a putatively ancient mechanism
469 of immune-like glycan sensing, shaped partly by functions acquired via prokaryote-to-
470 eukaryote HGT, can be co-opted for control of a facultative endosymbiotic interaction. We
471 propose that this functional network of glycan processing and sensing represents an
472 important factor in how stable host-microbe associations have evolved.

473 **Data Availability**

474 All data and script used for analysis and processing are available at:
475 https://github.com/benjaminhjenkins/Paramecium_bursaria. All sequence reads are
476 available on NCBI GenBank with the BioProject identifier PRJNA65904 and SRA accession
477 number SRR1251101, and at: https://github.com/guyleonard/paramecium and
478 https://zenodo.org/doi/10.5281/zenodo.4638887.

479 **Acknowledgements**

480 This work was primarily supported by an ERC Consolidator Grant (CELL-in-CELL) and a Royal
481 Society University Research Fellowship (UF130382) to T.A.R. We thank Prof. Neil Gow of the
482 University of Exeter for informative discussions on chitin sensing which inspired this work.
483 The authors gratefully acknowledge the Micron Advanced Bioimaging Facility (supported by
484 Wellcome Strategic Awards 091911/B/10/Z and 107457/Z/15/Z) for their support &
485 assistance in this work'.

486 **Methods**

487 ***Culture conditions and media***

488 In all experiments, *Paramecium bursaria* (*Pb*) 186b (CCAP 1660/18) strain was used. *Pb* cells
489 were cultured in New Cereal Leaf – Prescott Liquid media (NCL). NCL media was prepared by
490 adding 4.3 mgL^{-1} $CaCl_2.2H_2O$, 1.6 mgL^{-1} KCl, 5.1 mgL^{-1} K_2HPO_4, 2.8 mgL^{-1} $MgSO_4.7H_2O$ to
491 deionised water. 1 gL^{-1} wheat bran was added, and the solution boiled for 5 minutes. Once
492 cooled, media was filtered once through Whatman Grade 1 filter paper and then through
493 Whatman GF/C glass microfiber filter paper. Filtered NCL media was autoclaved at 121°C for
494 30 mins to sterilise prior to use.

495 NCL medium was bacterized with *Klebsiella pneumoniae* SMC and supplemented with
496 0.8 mgL^{-1} β-sitosterol prior to propagation. *Pb* 186b cells were sub-cultured 1:9 into fresh
497 bacterized NCL media every two months and maintained at 23°C at a light intensity of 25 = µ
498 mol m^{-2} s^{-1} (standard – e.g., **Fig. 6a/b**) or 80 µ mol m^{-2} s^{-1} (high – e.g., **Fig. 6a/b**) with a light-
499 dark (LD) cycle of 12:12 h. *Pb* cultures were maintained for at least three weeks following sub-
500 culture to ensure that cells were in stationary phase prior to experimentation.

501 In free-living algal assays, *Micractinium conductrix* SAG 241.80 (a member of the
502 *Chlorella* clade[20,22]) was used, as axenic algal endosymbionts isolated from *Pb* 186b could not
503 be reliably cultured. This strain was isolated from *Pb* SAG 27.96 and shares an identical ITS2
504 sequence with the *M. conductrix* endosymbiont of *Pb* 186b[96]. Algae were cultured in Modified
505 Bold's Basal Medium (MBBM) medium at 24°C with a 14:10 h LD cycle.

506 For bacterial feeding assays the following strains were used: *Escherichia coli* HT115
507 (Ec); *Klebsiella pneumoniae* SMC (Kp); *Serratia marascens* BS303 (Sm); *Leucobacter celer*
508 *subsp. astrifaciens* CBX151T (Ls); *Actinetobacter baumanii* (Ab); *Bacillus subtilis* 168 (Bs);
509 *Psuedomonas aeruginosa* PA14 (Pa); *Enterococcus faecalis* OG1RF (Ef); *Staphylococcus aureus*
510 MSSA476 (Sa); *Citrobacter freundii* (Cf); *Bifidobacterium breve* (Bb); and *Lactobacillus ruminis*
511 (Lr). Bacteria were cultured overnight (constant shaking, 30°C) in Luria-Bertani broth (Ec – Pa)
512 or Todd-Hewitt (Ef – Sa) broth based on standard cultivation methods, or in modified Gifu
513 Anaerobic Medium (mGAM; Nissui Pharmaceuticals) (Cf – Lr) under anaerobic conditions[97]
514 (5% H2, 5% CO2, 90% N2, <20 ppm O2).

515 For yeast secretion assays, *Saccharomyces cerevisiae* Y16947 (Euroscarf) was used.
516 This strain is a *S. cerevisiae* BY4742 CTSI deletion mutant (YLR286cΔ). *S. cerevisiae* was
517 cultured in YPD at 30°C and stored at -80°C in 25% glycerol solution.

518 ***Phylogenetic analysis***

519 Dataset generation and sampling for each phylogenetic analysis were conducted as follows.
520 For identification of CAZymes in *Pb*, predicted proteins[19] were searched with a custom
521 annotation tool (https://github.com/benjaminhjenkins/CAZyme_survey). For HGT analysis,
522 dataset curation was conducted as described by Irwin et al.[98]. Briefly, datasets consisting of
523 representative eukaryote proteins from all available major eukaryotic supergroups, and viral
524 proteins excluding over-represented Human Immuno-deficiency Virus-1, were clustered into
525 protein families using Diamond BLASTp and Markov Clustering. Hidden-Markov models
526 (HMMs) of clustered protein families were used to search a dataset of prokaryote proteins,
527 limited to 150-bacterial top hits per genus, and the resulting eukaryote, viral, and prokaryote
528 protein families were merged and re-clustered. For a full overview of HGT dataset curation,
529 including software versions and settings used, see Irwin et al.[98]. For a phylogeny of *CLP* in
530 eukaryotes, *Pb* 186b *CLP* and *Homo sapiens* (*Hs*) *SI-CLP* sequences were used to conduct
531 clade-specific searches of the NCBI protein database (gathering threshold: e-01, December
532 2023), including all available major eukaryotic supergroups and avoiding over-representation
533 from metazoan and plant taxa. These were added to an *SI-CLP* orthologue dataset (OrthoDB)
534 to give a final dataset of 1,445 candidate homologous amino acid sequences.

535 For all datasets, protein sequences were aligned using MAFFT v7.508[99] and masked
536 using TrimAL v1.4.15[100] allowing for no gaps. Highly divergent or identical sequence paralogs
537 from the same genomic source were manually removed. Phylogenies were generated using
538 IQ-TREE v2.2.0.3[101] with 1,000 ultra-fast bootstraps. Models for each tree were calculated
539 using IQ-TREE's ModelFinder implementation according to Bayesian inference criterion (BIC)
540 and are detailed in the relevant figure legends. Bootstrap methods and models chosen for
541 tree generation are listed in the respective figure legends.

542 ***RNAi feeding***

543 A full protocol for RNAi feeding and imaging in *Pb* is available at:
544 dx.doi.org/10.17504/protocols.io.8epv5jzm4l1b/v1. Briefly, *Pb* was fed with *E. coli*
545 transformed with an L4440 plasmid construct with paired IPTG-inducible T7 promoters,
546 facilitating targeted gene perturbation through the phagotrophic delivery of complementary
547 double-stranded (ds)RNA. L4440 plasmid constructs (**Supplementary Table S7**) were

548 transformed into *E. coli* HT115 competent cells and grown overnight on LB agar (50 µgmL^{-1}
549 Ampicillin and 12.5 µgmL^{-1} Tetracycline) at 37°C. Positive transformants were picked and
550 grown overnight in LB (50 µgmL^{-1} Ampicillin and 12.5 µgmL^{-1} Tetracycline) at 37°C with shaking
551 (180 rpm). Overnight pre-cultures were back-diluted 1:25 into 15 mL of LB (50 µgmL^{-1}
552 Ampicillin and 12.5 µgmL^{-1} Tetracycline) and incubated for a further 2 hours under the same
553 conditions, until an OD_{600} of between 0.4 and 0.6 was reached. *E. coli* cultures were then
554 supplemented with 0.4 mM IPTG to induce template expression within the L4440 plasmid,
555 and incubated for a further 3 hours under the same conditions. *E. coli* cells were pelleted by
556 centrifugation (3100 x *g* for 2 mins), washed with sterile NCL media, and pelleted once more.
557 *E. coli* cells were then re-suspended in NCL media supplemented with 0.4 mM IPTG, 100 µgmL^{-}
558 ^{1} Ampicillin, and 0.8 µgmL^{-1} β-sitosterol, 2% glycerol, and adjusted to a final OD_{600} of 3. *E. coli*
559 cells were split into single-use aliquots and stored at -20°C until feeding[102].

560 *Pb* cells were pelleted by gentle centrifugation in a 96-well plate (10 mins at 800 x *g*),
561 taking care not to disturb the cell pellet by leaving 100 µL of supernatant, and re-suspended
562 2:5 into 150 µL of diluted, induced *E. coli* culture media (to make 250 µL total; final OD 0.1).
563 Feeding was conducted daily for 6 days using frozen bacterial stocks and, where
564 cycloheximide was added alongside feeding, this was added on days 4 and 5 to a final
565 concentration of 50 µgmL^{-1}. The orientation of construct position on the microwell plates was
566 altered between experiments, ensuring different columns and rows were used for each
567 construct between experiments (see analysis below).

568 ***Live cell imaging***

569 High-throughput imaging was performed on an ImageXpress Pico Automated Cell Imaging
570 System (Molecular Devices) at 4x magnification, using the Cy5 (absorbance: 630/40 nm;
571 emission: 695/45 nm) channel to capture algal-chlorophyll autofluorescence as described
572 previously[29]. Images were automatically stitched together using the built-in
573 CellReporterXpress image acquisition software, generating a composite tiled image for each
574 well for downstream analysis. Individual cells in each well were masked using the
575 MetaExpress software and the average fluorescence of each object scored. Data for RNAi
576 screening (**Fig. 2a**) and glycan substrate / bacterial feeding (**Fig. 6a, e-g**) experiments were
577 gated by shape [0.6 < shape factor < 0.9], size [area < 10,000] and fluorescence [average

578 fluorescence intensity < 150] to select for single intact *Pb* cells during cycloheximide
579 treatment.

580 Average per-host fluorescence data for RNAi screening experiments (Fig. 2a) were
581 analysed using a linear mixed effects model using the lme4[103] and lmerTest[104] packages. The
582 model included separate random intercepts accounting for row position, column position,
583 and experimental block to control for batch and/or positional artefacts during imaging.
584 Degrees of freedom to calculate p-values were estimated using the Satterthwaite
585 approximation. Average per-host fluorescence data for glycan substrate / bacterial feeding
586 (**Fig. 6a, e-g**) experiments were analysed using a linear model on log-transformed data (**Fig.**
587 **6a, e-g**). All analyses were performed using R version 4.2.1.

588 ***RNA sequencing and differential expression analysis***

589 *Pb* cultures (6x replicates) were subject to RNAi feeding (as above) for 6 days prior to
590 harvesting, including a penultimate 2 days of cycloheximide treatment for the corresponding
591 samples. *Pb* cells (~50,000 per sample) were strained three times through a 40 µm cell strainer
592 to remove large debris, collected on an 11 µm filter, washed with NCL, and resuspended into
593 1 mL of TriZol reagent. RNA was extracted using the Zymoprep RNA extraction kit and stored
594 in nuclease-free water at -20°C. RNA was processed using the Illumina NovaSeq 6000 v1.5
595 workflow (Illumina™) with polyA selection. A 150 bp paired-end library was prepared and
596 resulted in 1,184,917,880 reads total consisting of 177,737,682,000 bp (SRA accession:
597 SRR1251101).

598 Paired-end reads were trimmed to remove adapters and primers using the TrimGalore
599 v0.6.7-1 wrapper tool of Cutadapt[105]. Trimmed reads were then aligned to our *Pb* 186b Pacbio
600 genome assembly[19] using BOWTIE2 v2.5.1[106], and contamination filtered using Samtools
601 v1.7[107] to keep only host *Pb* mapped sequences (80-90% of all reads). Normalized transcript
602 expression values for each read data set were calculated and cross-sample normalization
603 performed using Salmon v1.10.2[108]. Differential expression analysis was conducted using
604 DESeq2 v1.34.0[109] to assess change in relative transcript abundance between pairwise data
605 sets.

606 For four-way analysis of gene expression, pb186bvf_008968-T1 (corresponding to *Pb*
607 *CLP*) and 11 transcripts (with highly variable or over 75% zero read counts) were removed

608 prior to differential expression analysis to aid in downstream clustering of less differentially
609 expressed transcripts. These sequences were retained in each pair-wise analysis of gene
610 expression. For pairwise and four-way analysis of gene expression, differentially expressed
611 transcripts (|log2 fold change| ≥ 0.5, adjusted p-value < 0.001) were clustered based on
612 normalized read counts. Hierarchical clustering was performed using Pearson correlation and
613 a maximum height cut-off of 80%. Transcript annotation was conducted by manual curation
614 of differentially expressed transcripts using InterProScan.

615 ***Mass spectrometry***

616 For metabolomics following *CLP* RNAi, *Pb* cultures (6x replicates) were subject to RNAi feeding
617 (as above) for 7 days prior to harvesting. *Pb* cells (~40,000 per sample) were filtered onto 11
618 µm nylon filters (Merck Millipore NY1104700) to minimise residual bacteria. The cells were
619 washed with water, and the cells resuspended from the filter in 500 µL 80% v/v HPLC-grade
620 methanol. The samples were transferred to 1.5 mL microcentrifuge tubes, snap frozen in
621 liquid nitrogen, and stored at -80°C. Samples were subjected to LC/MS analysis using a
622 derivatised C18 method by the metabolomics facility at the University of Oxford, Department
623 of Chemistry. Data were subjected to sample-specific normalisation based on manual cell
624 counts, followed by median normalisation, log transformation, and Pareto scaling using
625 MetaboAnalyst v5.0[110] (https://genap.metaboanalyst.ca/). Data were then analysed using a
626 one-way ANOVA, followed by a Fisher's Least Significant Difference test with false discovery
627 rate (FDR) adjusted p-values (based on the Benjamini–Hochberg procedure).

628 For metabolomics of *Pb* 186b under differential symbiont load, cultures were first
629 grown at 50 µ mol m^{-2} s^{-1} to increase cell densities, and then split and acclimated to their
630 treatment conditions of low (6 µ mol m^{-2} s^{-1}) or moderate-high (50 µ mol m^{-2} s^{-1}) light intensity
631 for 3 days. They were then sampled and analyzed as described previously[24]. Briefly, *Pb* cells
632 were filtered on 11 µm nylon filters, washed in Volvic water, and then the cells were disrupted
633 by sonication (20% power for 10 s). The *Pb* fraction was collected by pushing 1 mL of the lysate
634 through a 1.6 µm filter, which caught the intact endosymbiont cells. The samples were
635 analyzed with a Synapt G2-Si with Acquity UPLC, recording in positive mode over a large
636 untargeted mass range (50 – 1000 Da). A 2.1x50mm Acuity UPLC BEH C18 column was used
637 with acetonitrile as the solvent. Masses of interest were investigated using the MarVis-Suite

638 2.0 software[111] (http://marvis.gobics.de/), using retention time and mass to compare against
639 KEGG[112] (https://www.genome.jp/kegg/) and MetaCyc[113] (https://biocyc.org/) databases.

640 ***Immunofluorescence microscopy and Western Blot of the Pb CLP proteins***

641 *Pb* cells were prepared for immunofluorescence microscopy using the protocol available at:
642 dx.doi.org/10.17504/protocols.io.yxmvm2ozng3p/v1. Primary rabbit polyclonal antibodies
643 for CLP were generated by Thermo Fisher Scientific, using the peptide sequence detailed in
644 **Supplementary Table S7**. Imaging was performed on a ScanR Olympus with a 20X objective
645 (refraction index = 1; emission = 698 [chlorophyll], 515 [Alexa Fluor 488]). Images were
646 collected using cellSens Dimension Software v3.2 and analysed using FIJI ImageJ2 version
647 2.3.0[114]. In brief, background noise was removed from the z-stack images using the 'rolling
648 ball background subtraction' (r=15). To count and analyse the fluorescence of the algae, a
649 smoothing gaussian blur filter was applied, and the z-stacks were combined to create a 2D
650 image. To resolve any overlapping cells, a Watershed segmentation was used after
651 thresholding the image with Otsu method. Processed images were false-coloured to aid
652 visualisation. Cells were manually scored based on whether they exhibited phagosome-like
653 fluorescence (localized to the oral groove and early endosome), lysosome-like fluorescence
654 (localized to small lysosome-like structures clustered around endosymbiotic algae), or both
655 (**Supplementary Table S5**). Algal fluorescence per-host cell and maximum *CLP* fluorescence
656 per host-cell were analysed using a linear model (**Fig. 4c/d**).

657 For Western Blot, *Pb* cells were first cleaned from debris by straining the culture
658 through a 40 µm cell strainer and washing the cells with autoclaved NCL media. The cells were
659 collected and concentrated via centrifugation (800 x *g* for 10min). Proteins were extracted for
660 30 min in 2x SDS buffer (55°C) supplemented with a mixture of protease-inhibitor (Roche;
661 Cat.4693116001), 2 µl of RNAse A (Qiagen; Cat. 19101) and 2 µl DNAse I (Fisher Scientific;
662 Cat. EN0521). The detailed protocol for the Western Blot of *Pb* is available at:
663 dx.doi.org/10.17504/protocols.io.ewov1q58kgr2/v1. Imaging was performed on an
664 iBrightCL1000 imager.

665 ***Yeast heterologous expression of Pb putative chitin interacting proteins***

666 *S. cerevisiae* Y16947 was transformed with a p426 GPD plasmid (ATCC 87361) containing a
667 gene expression insert modified with a 5' mating factor α (MFA) signal peptide sequence, or
668 an empty p426 GPD plasmid control, as described previously[49].

669 Transformants were grown for 7 days on SCM-URA at 30°C with shaking at 180 rpm,
670 then centrifuged at 3,200 x *g* for 5 mins and the supernatant collected. Supernatant from
671 each culture was centrifuged at 3,200 x *g* for 40 mins using a Spin-X UF concentrator 10,000
672 MWCO (Corning) to collect proteins over 10 kDa. Collected protein was washed in PBS and
673 matched to a concentration of 120 µg/mL prior to use. Protein activity assays were conducted
674 using 10 µL of sample (3x replication) on a Chitinase Activity Kit (Sigma-Aldrich #CS0980)
675 following the manufacturer's instructions, with incubation at 37°C for 3 hours. Comparisons
676 of catalytic function were analyzed using a generalized linear model with quasi-binomial
677 distribution (**Fig. 5d**).

678 ***Symbiont load assessment using flow cytometry***

679 *Pb* cultures to assess symbiont load in **Fig. 6c** and **Fig. S10** were grown under varying light
680 conditions (0, 6, 12, 24 and 50 µ mol m^{-2} s^{-1}) with a light-dark (LD) cycle of 14:10 h for seven
681 days. Symbiont load was estimated using a CytoFLEX S flow cytometer (Beckman Coulter Inc.,
682 CA, USA) by measuring the intensity of chlorophyll fluorescence for single *Pb* cells (excitation
683 488nm, emission 690/50nm) and gating cell size using forward side scatter[59].

684 ***Glycan substrate and bacterial feeding***

685 *Pb* cultures were washed twice with NCL and filtered using a 40 µm cell strainer to remove
686 large debris prior to feeding. For substrate feeding, *Pb* cells (~150 in 300 µL) in a 96-well plate
687 were supplemented with 0.01 mM, 0.1 mM, or 1 mM of glycan substrate in NCL media
688 (**Supplementary Table S8**). Concentrations of each substrate used are stated in the respective
689 figure panels. *Pb* cultures (6x replicates) were maintained under normal growth conditions at
690 either standard (25 µ mol m^{-2} s^{-1}) or high (80 µ mol m^{-2} s^{-1}) light intensity for up to 72 hours
691 prior to imaging.

692 For bacterial feeding, bacterial stocks were grown overnight for 48 hours, washed
693 twice in NCL, and fed to *Pb* cells (~75-100 in 250 µL) at a final OD_{600} of 0.1. *Pb* cultures (6x

694 replicates) were maintained under normal growth conditions for up to 48 hours prior to
695 imaging (see below).

696 ***Algal growth assays***

697 To determine if *M. conductrix* SAG 241.80 can utilize GlcNAc as a carbon source for
698 heterotrophic growth, growth assays were performed in continuous darkness over a 7-day
699 period in BBM+V media (CCAP) supplemented with 1 mM L-arginine using Biolog Phenotype
700 MicroArrays™ PM1 plates. Populations (3x replication) were initiated at a density of
701 approximately 5 x 106 cells/ml and changes in optical density (OD740) were regularly
702 recorded. Heterotrophic growth on GlcNAc was compared with negative (no carbon source)
703 and positive (α-D-glucose control treatments).

704 To determine if *M. conductrix* SAG 241.80 can utilize GlcNAc as a nitrogen source, growth
705 assays with 1 mM GlcNAc supplementation were performed over a 10-day period and
706 changes in optical density (OD740) were compared with negative (no nitrogen, -N) and
707 positive control (1 mM L-arginine supplementation) treatments. The algae were nitrogen
708 starved in BBM+V media (CCAP) lacking a nitrogen source (i.e., BBM+V-N) for 24 hours before
709 the growth assays commenced. Populations (6x replication) were then initiated in 96 well
710 plates at a density of approximately 5 x 106 cells/ml in BBM+V containing 100 mM MES and
711 100 mM MOPS buffer and, where applicable, 1 mM of the nitrogen source to be tested at pH
712 5.5, 6.5 and 7.5. Plates were incubated at 24°C with a 14/10 h LD cycle for 10 days and OD740
713 readings were regularly recorded.

714 **REFERENCES**

715 1. Bonen, L., Cunningham, R. S., Gray, M. W. & Doolittle, W. F. Wheat embryo mitochondrial 18S
716 ribosomal RNA: evidence for its prokaryotic nature. *Nucleic Acids Res* **4**, 663–671 (1977).
717 2. Bonen, L. & Doolittle, W. F. On the prokaryotic nature of red algal chloroplasts. *Proc Natl Acad*
718 *Sci U S A* **72**, 2310–2314 (1975).
719 3. Archibald, J. M. The puzzle of plastid evolution. *Curr Biol* **19**, R81-8 (2009).
720 4. Keeling, P. J. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution.
721 *Annual Review of Plant Biology* **64**, 583–607 (2013).

722 5. McFall-Ngai, M. J. & Ruby, E. G. Symbiont recognition and subsequent morphogenesis as early
723 events in an animal-bacterial mutualism. *Science* **254**, 1491–1494 (1991).
724 6. McFadden, G. I., Reith, M. E., Munholland, J. & Lang-Unnasch, N. Plastid in human parasites.
725 *Nature* **381**, 482–482 (1996).
726 7. Baker, A. C. Flexibility and specificity in coral-algal symbiosis: diversity, ecology, and
727 biogeography of Symbiodinium. *Annu. Rev. Ecol. Evol. Syst.* **34**, 661–689 (2003).
728 8. Curtis, B. A. *et al.* Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs.
729 *Nature* **492**, 59–65 (2012).
730 9. Husnik, F. *et al.* Horizontal gene transfer from diverse bacteria to an insect genome enables a
731 tripartite nested mealybug symbiosis. *Cell* **153**, 1567–1578 (2013).
732 10. Werner, G. D. A., Cornwell, W. K., Sprent, J. I., Kattge, J. & Kiers, E. T. A single evolutionary
733 innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. *Nat Commun* **5**,
734 4087 (2014).
735 11. Kwong, W. K., Del Campo, J., Mathur, V., Vermeij, M. J. A. & Keeling, P. J. A widespread coral-
736 infecting apicomplexan with chlorophyll biosynthesis genes. *Nature* **568**, 103–107 (2019).
737 12. Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the legume–
738 rhizobium mutualism. *Nature* **425**, 78–81 (2003).
739 13. Lowe, C. D., Minter, E. J., Cameron, D. D. & Brockhurst, M. A. Shining a light on exploitative
740 host control in a photosynthetic endosymbiosis. *Curr Biol* **26**, 207–211 (2016).
741 14. Foster, K. R., Schluter, J., Coyte, K. Z. & Rakoff-Nahoum, S. The evolution of the host
742 microbiome as an ecosystem on a leash. *Nature* **548**, 43–51 (2017).
743 15. López-Madrigal, S. & Duarte, E. H. Titer regulation in arthropod-*Wolbachia* symbioses. *FEMS*
744 *Microbiol Lett* **366**, fnz232 (2019).
745 16. Jenkins, B. H. *et al.* Characterization of the RNA-interference pathway as a tool for reverse
746 genetic analysis in the nascent phototrophic endosymbiosis, *Paramecium bursaria*. *R. Soc. open*
747 *sci.* **8**, rsos.210140, 210140 (2021).

748 17. Fujishima, M. & Kodama, Y. Mechanisms for establishing primary and secondary
749 endosymbiosis in Paramecium. *Journal of Eukaryotic Microbiology* **69**, e12901 (2022).
750 18. Jenkins, B. H. Mutualism on the edge: Understanding the *Paramecium–Chlorella* symbiosis.
751 *PLOS Biology* **22**, e3002563 (2024).
752 19. Leonard, G. *et al.* De novo genome sequence assembly of the RNAi-tractable endosymbiosis
753 model system *Paramecium bursaria* 186b reveals factors shaping intron repertoire.
754 2024.08.09.607295 Preprint at https://doi.org/10.1101/2024.08.09.607295 (2024).
755 20. Hoshina, R., Iwataki, M. & Imamura, N. *Chlorella variabilis* and *Micractinium reisseri* sp. nov.
756 (Chlorellaceae, Trebouxiophyceae): Redescription of the endosymbiotic green algae of
757 Paramecium bursaria (Peniculia, Oligohymenophorea) in the 120th year. *Phycological Research*
758 **58**, 188–201 (2010).
759 21. Ziesenisz, E., Reisser, W. & Wiessner, W. Evidence of de novo synthesis of maltose excreted by
760 the endosymbiotic *Chlorella* from *Paramecium bursaria*. *Planta* **153**, 481–485 (1981).
761 22. Arriola, M. B. *et al.* Genome sequences of *Chlorella sorokiniana* UTEX 1602 and *Micractinium*
762 *conductrix* SAG 241.80: implications to maltose excretion by a green alga. *The Plant Journal* **93**,
763 566–586 (2018).
764 23. Kato, Y. & Imamura, N. Effect of sugars on amino acid transport by symbiotic *Chlorella*. *Plant*
765 *Physiol Biochem* **46**, 911–7 (2008).
766 24. Sørensen, M. E. S. *et al.* Comparison of independent evolutionary origins reveals both
767 convergence and divergence in the metabolic mechanisms of symbiosis. *Current Biology* **30**,
768 328-334.e4 (2020).
769 25. Nishihara, N. *et al.* Microtubule-dependent movement of symbiotic algae and granules in
770 Paramecium bursaria. *Cell Motil Cytoskeleton* **43**, 85–98 (1999).
771 26. Takahashi, T., Shirai, Y., Kosaka, T. & Hosoya, H. Arrest of cytoplasmic streaming induces algal
772 proliferation in green Paramecia. *PLoS ONE* **2**, e1352 (2007).

773 27. Kodama, Y. & Fujishima, M. Synchronous Induction of detachment and reattachment of
774 symbiotic *Chlorella* spp. from the cell cortex of the host *Paramecium bursaria*. *Protist* **164**,
775 660–672 (2013).
776 28. Kodama, Y. & Fujishima, M. Cycloheximide induces synchronous swelling of perialgal vacuoles
777 enclosing symbiotic *Chlorella vulgaris* and digestion of the algae in the ciliate *Paramecium*
778 *bursaria*. *Protist* **159**, 483–94 (2008).
779 29. Jenkins, B. H. *et al.* Emergent RNA–RNA interactions can promote stability in a facultative
780 phototrophic endosymbiosis. *PNAS* **118**, (2021).
781 30. Sørensen, M. E. S., Wood, A. J., Cameron, D. D. & Brockhurst, M. A. Rapid compensatory
782 evolution can rescue low fitness symbioses following partner switching. *Current Biology* **31**,
783 3721-3728.e4 (2021).
784 31. Horas, E. L., Metzger, S. M., Platzer, B., Kelly, J. B. & Becks, L. Context-dependent costs and
785 benefits of endosymbiotic interactions in a ciliate–algae system. *Environmental Microbiology*
786 **24**, 5924–5935 (2022).
787 32. Gerken, H., Donohoe, B. & Knoshaug, E. Enzymatic cell wall degradation of *Chlorella vulgaris*
788 and other microalgae for biofuels production. *Planta* **237**, (2012).
789 33. Higuchi, R., Song, C., Hoshina, R. & Suzaki, T. Endosymbiosis-related changes in ultrastructure
790 and chemical composition of *Chlorella variabilis* (Archaeplastida, Chlorophyta) cell wall in
791 Paramecium bursaria (Ciliophora, Oligohymenophorea). *European Journal of Protistology* **66**,
792 149–155 (2018).
793 34. Hiramatsu, S., Ishihara, M., Fujie, M., Usami, S. & Yamada, T. Expression of a chitinase gene and
794 lysis of the host cell wall during *Chlorella virus* CVK2 infection. *Virology* **260**, 308–315 (1999).
795 35. Kodama, Y. *et al.* Comparison of gene expression of *Paramecium bursaria* with and without
796 *Chlorella variabilis* symbionts. *BMC Genomics* **15**, 183 (2014).
797 36. Hamada, M. *et al.* Metabolic co-dependence drives the evolutionarily ancient *Hydra–Chlorella*
798 symbiosis. *eLife* **7**, e35122 (2018).

799 37. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The
800 carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**, D490–D495
801 (2014).
802 38. Li, D., Wang, J., Liu, Y., Li, Y. & Zhang, Z. Expanded analyses of the functional correlations within
803 structural classifications of glycoside hydrolases. *Computational and Structural Biotechnology*
804 *Journal* **19**, 5931–5942 (2021).
805 39. Gooday, G. W. The ecology of chitin degradation. in *Advances in Microbial Ecology* (ed.
806 Marshall, K. C.) 387–430 (Springer US, Boston, MA, 1990). doi:10.1007/978-1-4684-7612-5_10.
807 40. Morozov, A. A. & Likhoshway, Y. V. Evolutionary history of the chitin synthases of eukaryotes.
808 *Glycobiology* **26**, 635–639 (2016).
809 41. Hamid, R. *et al.* Chitinases: An update. *J Pharm Bioallied Sci* **5**, 21–29 (2013).
810 42. Steinfeld, L., Vafaei, A., Rösner, J. & Merzendorfer, H. Chitin prevalence and function in
811 bacteria, fungi and protists. *Adv Exp Med Biol* **1142**, 19–59 (2019).
812 43. Gerbracht, J. V., Harding, T., Simpson, A. G. B., Roger, A. J. & Hess, S. Comparative
813 transcriptomics reveals the molecular toolkit used by an algivorous protist for cell wall
814 perforation. *Current Biology* **32**, 3374-3384.e5 (2022).
815 44. Zhang, J. *et al.* A receptor required for chitin perception facilitates arbuscular mycorrhizal
816 associations and distinguishes root symbiosis from immunity. *Current Biology* **34**, 1705-1717.e6
817 (2024).
818 45. Malolepszy, A. *et al.* A plant chitinase controls cortical infection thread progression and
819 nitrogen-fixing symbiosis. *eLife* **7**, e38874 (2018).
820 46. Desaki, Y., Miyata, K., Suzuki, M., Shibuya, N. & Kaku, H. Plant immunity and symbiosis signaling
821 mediated by LysM receptors. *Innate Immun* **24**, 92–100 (2018).
822 47. Altura, M. A., Stabb, E., Goldman, W., Apicella, M. & McFall-Ngai, M. J. Attenuation of host NO
823 production by MAMPs potentiates development of the host in the squid-*Vibrio* symbiosis. *Cell*
824 *Microbiol* **13**, 527–537 (2011).

825 48. Aschtgen, M.-S. *et al.* Insights into flagellar function and mechanism from the squid–*Vibrio*
826 symbiosis. *npj Biofilms and Microbiomes* **5**, 1–10 (2019).
827 49. Attah, V. *et al.* Duplication and neofunctionalization of a horizontally transferred xyloglucanase
828 as a facet of the Red Queen coevolutionary dynamic. *Proceedings of the National Academy of*
829 *Sciences* **121**, e2218927121 (2024).
830 50. Hughes, A. L. & Friedman, R. Genome-Wide Survey for Genes Horizontally Transferred from
831 Cellular Organisms to Baculoviruses. *Molecular Biology and Evolution* **20**, 979–987 (2003).
832 51. Blanc, G. *et al.* The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis,
833 coevolution with viruses, and cryptic sex. *The Plant Cell* **22**, 2943–2955 (2010).
834 52. Richards, T. A. *et al.* Horizontal gene transfer facilitated the evolution of plant parasitic
835 mechanisms in the oomycetes. *Proceedings of the National Academy of Sciences* **108**, 15258–
836 15263 (2011).
837 53. Van Etten, J. L. *et al.* Chloroviruses have a sweet tooth. *Viruses* **9**, E88 (2017).
838 54. Husnik, F. & McCutcheon, J. P. Functional horizontal gene transfer from bacteria to eukaryotes.
839 *Nat. Rev. Microbiol.* **16**, 67–79 (2018).
840 55. Gonçalves, I. R. *et al.* Genome-wide analyses of chitin synthases identify horizontal gene
841 transfers towards bacteria and allow a robust and unifying classification into fungi. *BMC*
842 *Evolutionary Biology* **16**, 252 (2016).
843 56. Savory, F., Leonard, G. & Richards, T. A. The role of horizontal gene transfer in the evolution of
844 the oomycetes. *PLOS Pathogens* **11**, e1004805 (2015).
845 57. Richards, T. A., Hirt, R. P., Williams, B. A. P. & Embley, T. M. Horizontal gene transfer and the
846 evolution of parasitic protozoa. *Protist* **154**, 17–32 (2003).
847 58. Caron, F. & Meyer, E. Does *Paramecium primaurelia* use a different genetic code in its
848 macronucleus? *Nature* **314**, 185–188 (1985).
849 59. Kadono, T., Kawano, T., Hosoya, H. & Kosaka, T. Flow cytometric studies of the host-regulated
850 cell cycle in algae symbiotic with green paramecium. *Protoplasma* **223**, 133–141 (2004).

851 60. Meng, G. *et al.* Structure of human stabilin-1 interacting chitinase-like protein (SI-CLP) reveals a
852 saccharide-binding cleft with lower sugar-binding selectivity. *J Biol Chem* **285**, 39898–39904
853 (2010).
854 61. Kzhyshkowska, J. *et al.* Novel stabilin-1 interacting chitinase-like protein (SI-CLP) is up-
855 regulated in alternatively activated macrophages and secreted via lysosomal pathway. *Blood*
856 **107**, 3221–3228 (2006).
857 62. Xiao, W. *et al.* Human secreted stabilin-1−interacting chitinase-like protein aggravates the
858 inflammation associated with rheumatoid arthritis and is a potential macrophage inflammatory
859 regulator in rodents. *Arthritis & Rheumatology* **66**, 1141–1152 (2014).
860 63. Karunanithi, S. *et al.* Exogenous RNAi mechanisms contribute to transcriptome adaptation by
861 phased siRNA clusters in Paramecium. *Nucleic Acids Res* **47**, 8036–8049 (2019).
862 64. Carradec, Q. *et al.* Primary and secondary siRNA synthesis triggered by RNAs from food
863 bacteria in the ciliate *Paramecium tetraurelia*. *Nucleic Acids Res* **43**, 1818–33 (2015).
864 65. Marat, A. L. & Haucke, V. Phosphatidylinositol 3-phosphates—at the interface between cell
865 signalling and membrane traffic. *EMBO J* **35**, 561–579 (2016).
866 66. Raiborg, C., Schink, K. O. & Stenmark, H. Class III phosphatidylinositol 3-kinase and its catalytic
867 product PtdIns3P in regulation of endocytic membrane traffic. *FEBS J* **280**, 2730–2742 (2013).
868 67. Tan, J. X. & Finkel, T. A phosphoinositide signalling pathway mediates rapid lysosomal repair.
869 *Nature* **609**, 815–821 (2022).
870 68. Okkenhaug, K. Signalling by the phosphoinositide 3-kinase family in immune cells. *Annu Rev*
871 *Immunol* **31**, 675–704 (2013).
872 69. O'Donnell, V. B., Rossjohn, J. & Wakelam, M. J. O. Phospholipid signaling in innate immune
873 cells. *J Clin Invest* **128**, 2670–2679.
874 70. Wichterman, R. The Biology of Paramecium. (Springer Science & Business Media, 2012).
875 71. Siegel, R. W. Hereditary endosymbiosis in *Paramecium bursaria*. *Exp Cell Res* **19**, 239–52
876 (1960).

877 72. Hollmén, M., Figueiredo, C. R. & Jalkanen, S. New tools to prevent cancer growth and spread: a
878 'Clever' approach. *Br J Cancer* **123**, 501–509 (2020).
879 73. Thumuluri, V., Almagro Armenteros, J. J., Johansen, A. R., Nielsen, H. & Winther, O. DeepLoc
880 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids*
881 *Research* **50**, W228–W234 (2022).
882 74. Doolittle, W. F. You are what you eat: a gene transfer ratchet could account for bacterial genes
883 in eukaryotic nuclear genomes. *Trends Genet.* **14**, 307–311 (1998).
884 75. Howe, C. J., Barbrook, A. C., Nisbet, R. E. R., Lockhart, P. J. & Larkum, A. W. D. The origin of
885 plastids. *Philos Trans R Soc Lond B Biol Sci* **363**, 2675–2685 (2008).
886 76. Larkum, A. W. D., Lockhart, P. J. & Howe, C. J. Shopping for plastids. *Trends Plant Sci* **12**, 189–
887 195 (2007).
888 77. Ricard, G. *et al.* Horizontal gene transfer from bacteria to rumen ciliates indicates adaptation to
889 their anaerobic, carbohydrates-rich environment. *BMC Genomics* **7**, 22 (2006).
890 78. Shaw, F. L. *et al.* Evolution and multifarious horizontal transfer of an alternative biosynthetic
891 pathway for the alternative polyamine sym-homospermidine. *J Biol Chem* **285**, 14711–14723
892 (2010).
893 79. Archibald, J. M. Plastid evolution: remnant algal genes in ciliates. *Current Biology* **18**, R663–
894 R665 (2008).
895 80. Feng, J.-M. *et al.* Single-cell transcriptome sequencing of rumen ciliates provides insight into
896 their molecular adaptations to the anaerobic and carbohydrate-rich rumen microenvironment.
897 *Molecular Phylogenetics and Evolution* **143**, 106687 (2020).
898 81. Görtz, H.-D. & Fokin, S. I. Diversity of endosymbiotic bacteria in Paramecium. in *Endosymbionts*
899 *in Paramecium* (ed. Fujishima, M.) 131–160 (Springer, Berlin, Heidelberg, 2009).
900 doi:10.1007/978-3-540-92677-1_6.
901 82. Flemming, F. E., Grosser, K. & Schrallhammer, M. Natural shifts in endosymbionts' occurrence
902 and relative frequency in their ciliate host population. *Front Microbiol* **12**, 791615 (2022).

903 83. Mironov, T. & Sabaneyeva, E. A robust symbiotic relationship between the ciliate *Paramecium*
904 *multimicronucleatum* and the bacterium Ca. *Trichorickettsia mobilis*. *Frontiers in Microbiology*
905 **11**, (2020).
906 84. Castelli, M. *et al.* Deianiraea, an extracellular bacterium associated with the ciliate
907 Paramecium, suggests an alternative scenario for the evolution of Rickettsiales. *ISME J* (2019).
908 85. Husnik, F. & Keeling, P. J. The fate of obligate endosymbionts: reduction, integration, or
909 extinction. *Curr. Opin. Genet. Dev.* **58–59**, 1–8 (2019).
910 86. Husnik, F. & McCutcheon, J. P. Repeated replacement of an intrabacterial symbiont in the
911 tripartite nested mealybug symbiosis. *PNAS* **113**, E5416–E5424 (2016).
912 87. Jacobovitz, M. R. *et al.* Dinoflagellate symbionts escape vomocytosis by host cell immune
913 suppression. *Nat Microbiol* **6**, 769–782 (2021).
914 88. Richter, D. J., Fozouni, P., Eisen, M. B. & King, N. Gene family innovation, conservation and loss
915 on the animal stem lineage. *eLife* **7**, e34226 (2018).
916 89. Woznica, A. *et al.* STING mediates immune responses in the closest living relatives of animals.
917 *eLife* **10**, e70436 (2021).
918 90. Buchmann, K. Evolution of innate immunity: clues from invertebrates via fish to mammals.
919 *Frontiers in Immunology* **5**, (2014).
920 91. Newman, M.-A., Sundelin, T., Nielsen, J. & Erbs, G. MAMP (microbe-associated molecular
921 pattern) triggered immunity in plants. *Frontiers in Plant Science* **4**, (2013).
922 92. Mogensen, T. H. Pathogen recognition and inflammatory signaling in innate immune defenses.
923 *Clin Microbiol Rev* **22**, 240–273 (2009).
924 93. Flemming, F. E., Potekhin, A., Pröschold, T. & Schrallhammer, M. Algal diversity in *Paramecium*
925 *bursaria*: species identification, detection of *Choricystis parasitica*, and assessment of the
926 interaction specificity. *Diversity* **12**, 287 (2020).
927 94. Douglas, A. E. Conflict, cheats and the persistence of symbioses. *New Phytologist* **177**, 849–858
928 (2008).

95. Ågren, J. A., Davies, N. G. & Foster, K. R. Enforcement is central to the evolution of cooperation. *Nat Ecol Evol* **3**, 1018–1029 (2019).

96. Spanner, C., Darienko, T., Filker, S., Sonntag, B. & Pröschold, T. Morphological diversity and molecular phylogeny of five *Paramecium bursaria* (Alveolata, Ciliophora, Oligohymenophorea) syngens and the identification of their green algal endosymbionts. *Sci Rep* **12**, 18089 (2022).

97. Spragge, F. *et al.* Microbiome diversity protects against pathogens by nutrient blocking. *Science* (2023) doi:10.1126/science.adj3502.

98. Irwin, N. A. T., Pittis, A. A., Richards, T. A. & Keeling, P. J. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nat Microbiol* **7**, 327–336 (2022).

99. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).

100. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

101. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268–274 (2015).

102. Irwin, N. A. T. *et al.* The function and evolution of motile DNA replication systems in ciliates. *Current Biology* (2020) doi:10.1016/j.cub.2020.09.077.

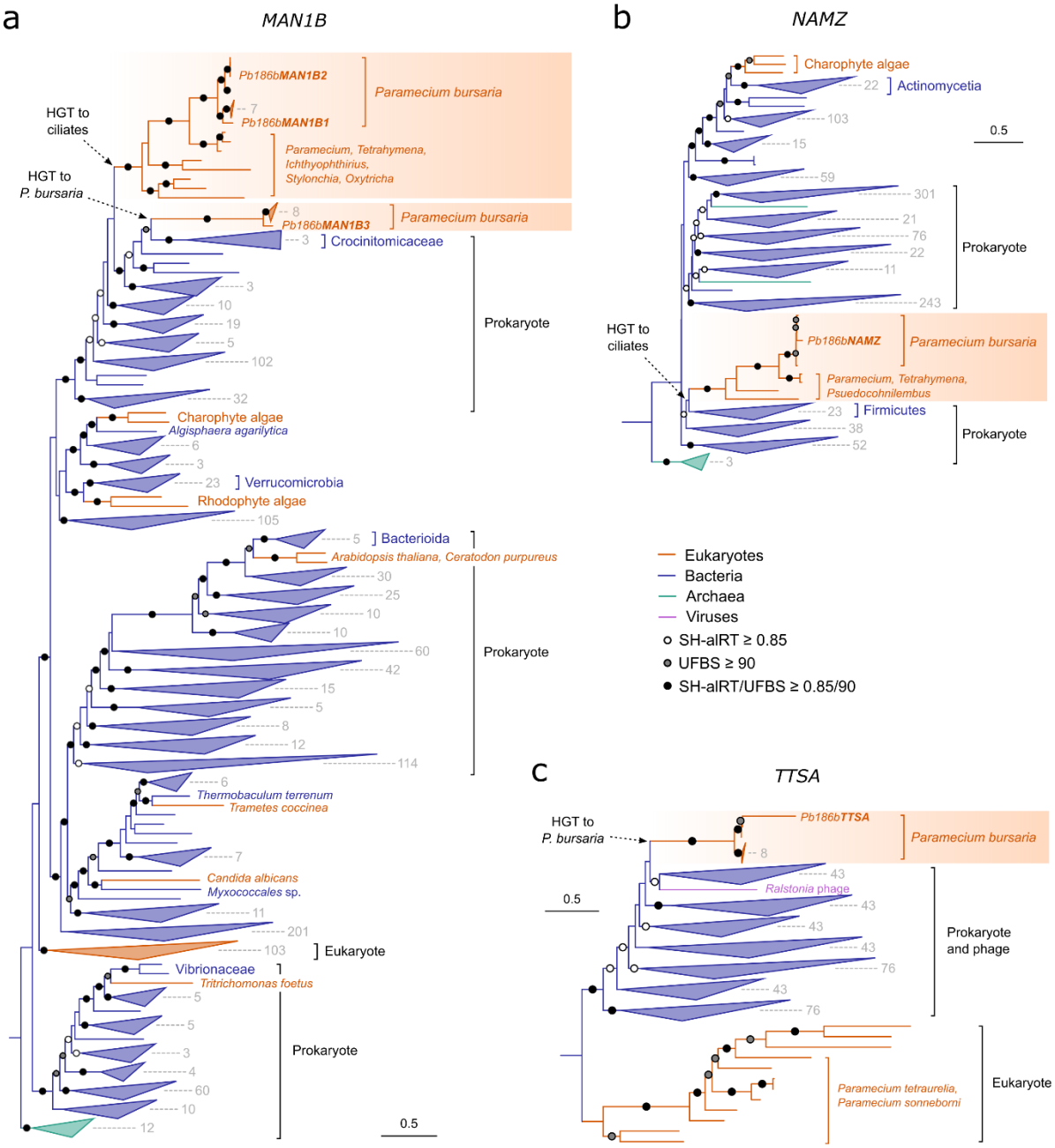
103. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48 (2015).

104. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* **82**, 1–26 (2017).

105. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

954 106. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–
955 359 (2012).
956 107. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079
957 (2009).
958 108. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-
959 aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).
960 109. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
961 RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
962 110. Pang, Z. *et al.* MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional
963 insights. *Nucleic Acids Research* **49**, W388–W396 (2021).
964 111. Kaever, A. *et al.* MarVis: a tool for clustering and visualization of metabolic biomarkers. *BMC*
965 *Bioinformatics* **10**, 92 (2009).
966 112. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for
967 understanding genome variations in KEGG. *Nucleic Acids Research* **47**, D590–D595 (2019).
968 113. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids*
969 *Research* **46**, D633–D639 (2018).
970 114. Rueden, C. T. *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC*
971 *Bioinformatics* **18**, 529 (2017).
972
973
974
975
976
977
978
979
980

981 **Figures**



982

983 **Figure 1: Horizontal gene-transfer has shaped glycan-processing in *P. bursaria*.** Maximum likelihood phylogenies of *Pb β-*
984 *mannosidase* (*MAN1B*; ***a***), *NamZ-like lysozyme* (*NAMZ*; ***b***), and *TtsA-like lysozyme* (*TTSA*; ***c***) genes. Topologies of these trees
985 provide strong support for prokaryote-to-eukaryote HGT, with two independent incidences of HGT and a subsequent gene
986 duplication of *MAN1B1/2* predicted to have occurred in *Pb*. We observed no evidence of *MAN1B2* transcription
987 (**Supplementary Table S1**) indicating possible pseudogenization of *MAN1B2* in *Pb* 186b. Trees were generated in IQ-TREE
988 using Q.pfam+R10 (***a***), Q.pfam+I+R10 (***b***), or Q.pfam+R7 (***c***) substitution models selected by Model Finder, and statistical
989 support assessed using SH-aLRT and ultra-fast bootstraps (UFBS; *n* = 1000). This figure shows four cases of HGT of GH gene
990 families, for maximum likelihood phylogenies of four additional predicted HGT-derived GH encoding genes in *Pb* see **Fig. S2.**

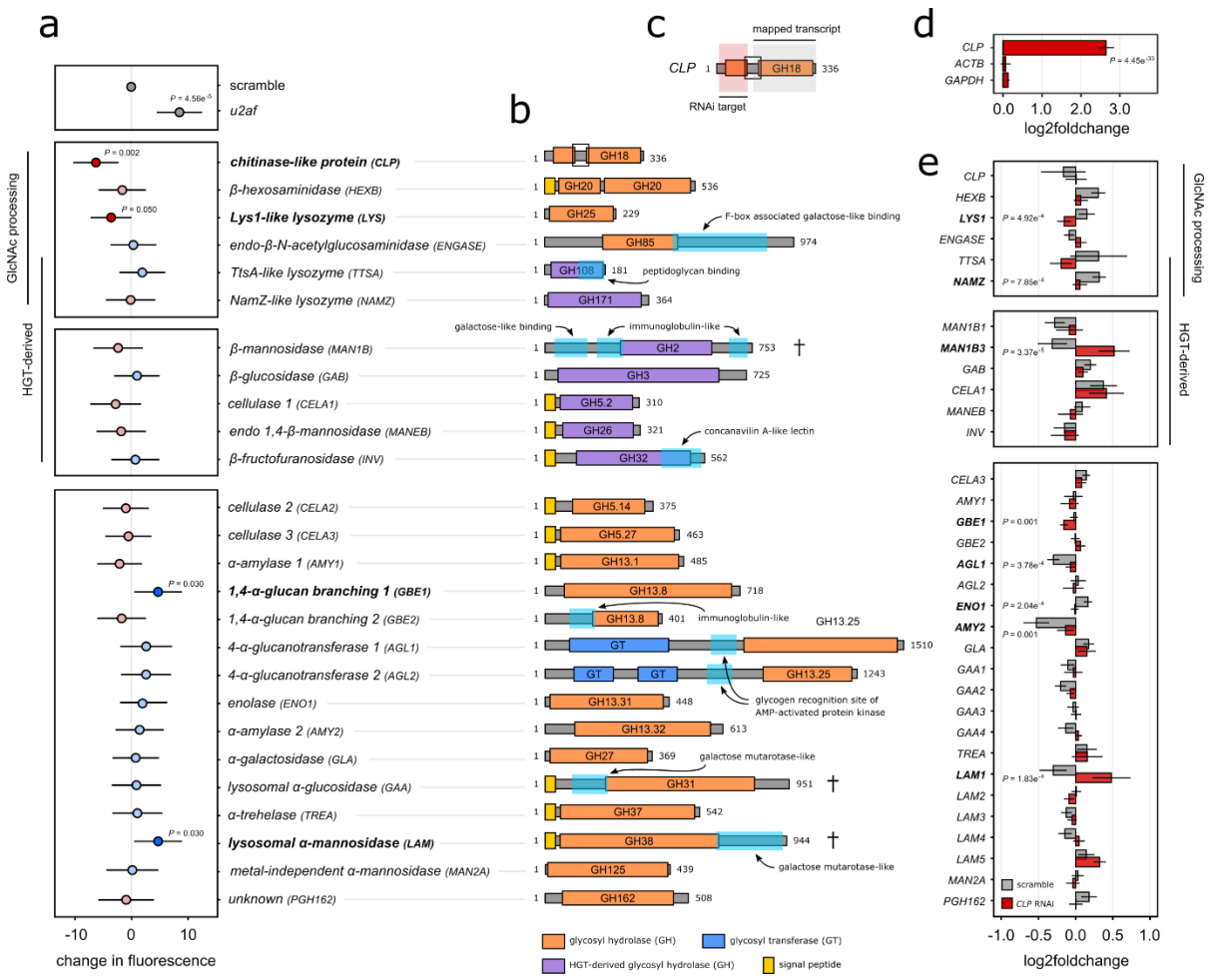


Figure 2: RNAi of glycan-processing genes perturbs host-mediated endosymbiont destruction, and perturbation of *CLP* determines expression of glycan-processing genes. All experiments were conducted using *Pb* cultures containing 75-100 cells (***a***) or 250,000 (***d, e***) cells. Cultures were fed daily with HT115 *E. coli* expressing target dsRNA to induce RNAi, or a non-hit 'scramble' dsRNA control. Per-host fluorescence is used throughout as a proxy for endosymbiotic algal load. ***a***, Change in per host fluorescence, compared to scramble treatment, in RNAi treated *Pb* cultures (n = 12) during endosymbiont break-down. Cells were fed under standard conditions for 4 days, followed by 2 days of simultaneous cycloheximide treatment (50 μgmL^{-1}) to trigger endosymbiont digestion. ***b***, Schematic illustration of functional domain architecture of target host genes. For chimeric constructs (†) targeting multiple genes a single representative is shown (see **Fig. S3** for all genes). GH, glycosyl hydrolase; GT, glycosyl transferase; additional domains are highlighted in cyan. ***c***, Schematic illustration of *CLP* RNAi construct design including dsRNA target (red) and mapped transcript region for differential expression analysis (grey). ***d***, Change in *CLP* gene expression (***c***) during RNAi treatment in untreated cultures, compared to Actin and GAPDH housekeeping genes. ***e***, Change in GH transcript expression in scramble (grey) or *CLP* RNAi treated (red) *Pb* cultures (n = 6) during cycloheximide induced endosymbiont break-down, compared to respective scramble or *CLP* RNAi cultures without cycloheximide treatment (see also **Supplementary Table S2**). Statistical significance is shown for difference in log2foldchange values between scramble and *CLP* RNAi conditions. For all data, statistical significance was calculated using a linear mixed effects model (***a***) or an un-paired two-tailed t-test (***d***, ***e***)**.** Conditions where (***a***) $p < 0.05$ or (***e***) Bonferroni-corrected $p < 0.0015$ are shown.

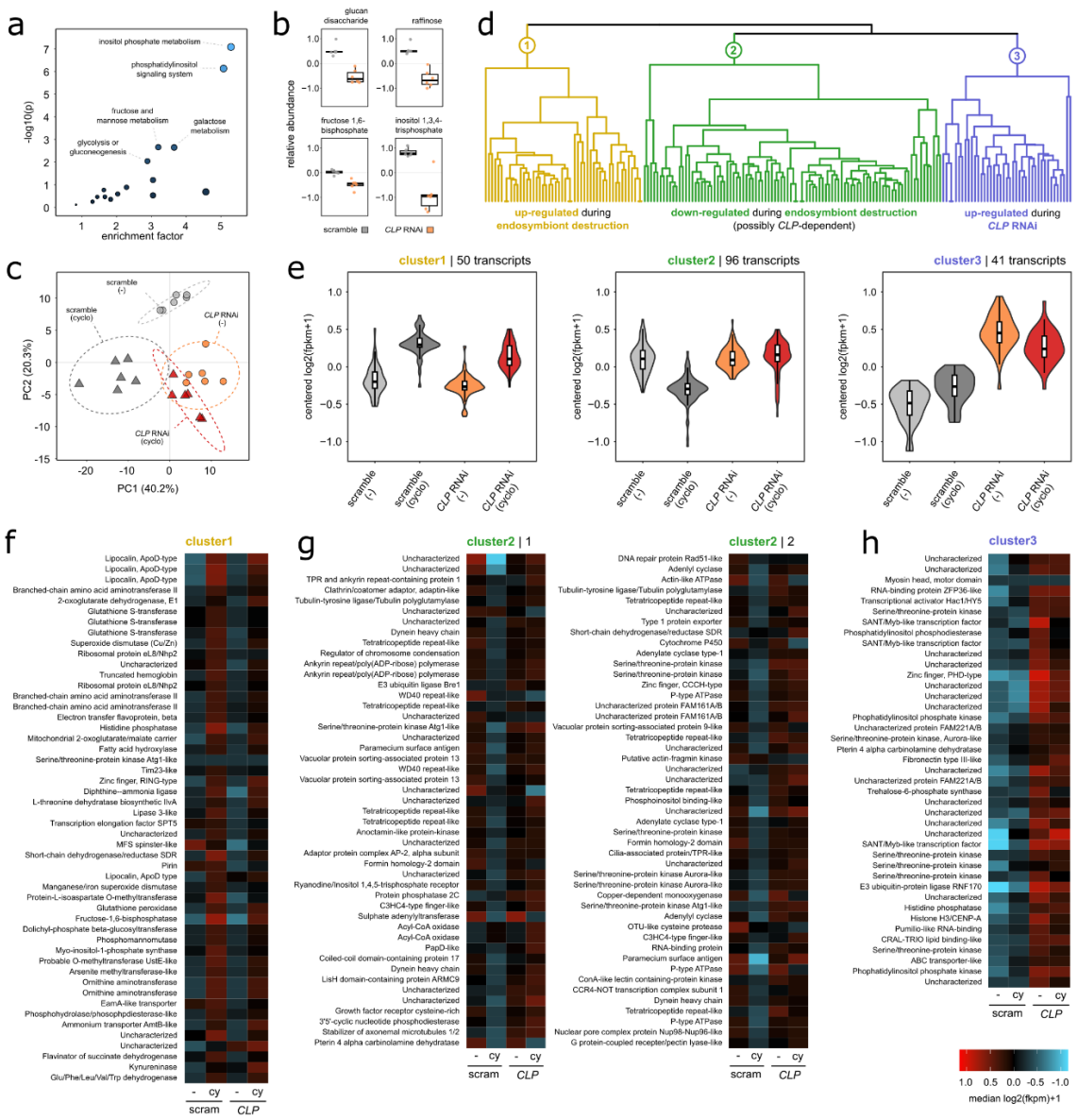


Figure 3: *CLP* RNAi alters expression of transcription factors and intracellular signalling systems. All experiments were conducted using *Pb* cultures (*n* = 6) containing approximately 250,000 cells. Cultures were fed daily with HT115 *E. coli* expressing target dsRNA to induce *CLP* knock-down, or 'scramble' dsRNA control with no hits to sampled *Pb* genes. ***a,*** Enrichment plot of significantly altered pathways during *CLP* RNAi, compared to a scramble control, detected by untargeted metabolomics (see also **Supplementary Table S3**). ***b,*** Relative abundance of significantly altered metabolites detected during *CLP* RNAi. ***c,*** PCA plot showing variance of transcriptional profiles in *Pb* cultures during *CLP* RNAi (orange, red) compared to a scramble control (grey, dark grey). Samples were treated with 50 µgmL^{-1} cycloheximide (triangle; cyclo) to trigger endosymbiont break-down, or remained untreated (circle; -). The percentage of explained variance for each PC is displayed. ***d,*** Hierarchical clustering of differentially expressed transcripts ([log2 fold change] ≥ 1, adjusted p value < 0.001) in each pairwise comparison. Clusters were defined through a dendogram maximum height cut-off of 80%. ***e,*** Three clusters of transcripts with similar expression patterns resulting from the cut dendogram (see also **Supplementary Table S4**) Data represented as median values (*n* = 6) of all transcripts from each cluster with overlaid distribution density. ***f-h,*** Heatmap of read-centred expression in each sample condition, and transcript annotation (InterProScan), for each of the identified transcript clusters. For relative abundance of all identified significantly altered metabolites during *CLP* RNAi, see **Fig. S4**. For individual heatmaps of each pairwise comparison across replicates during differential expression analysis, see **Fig. S6**.

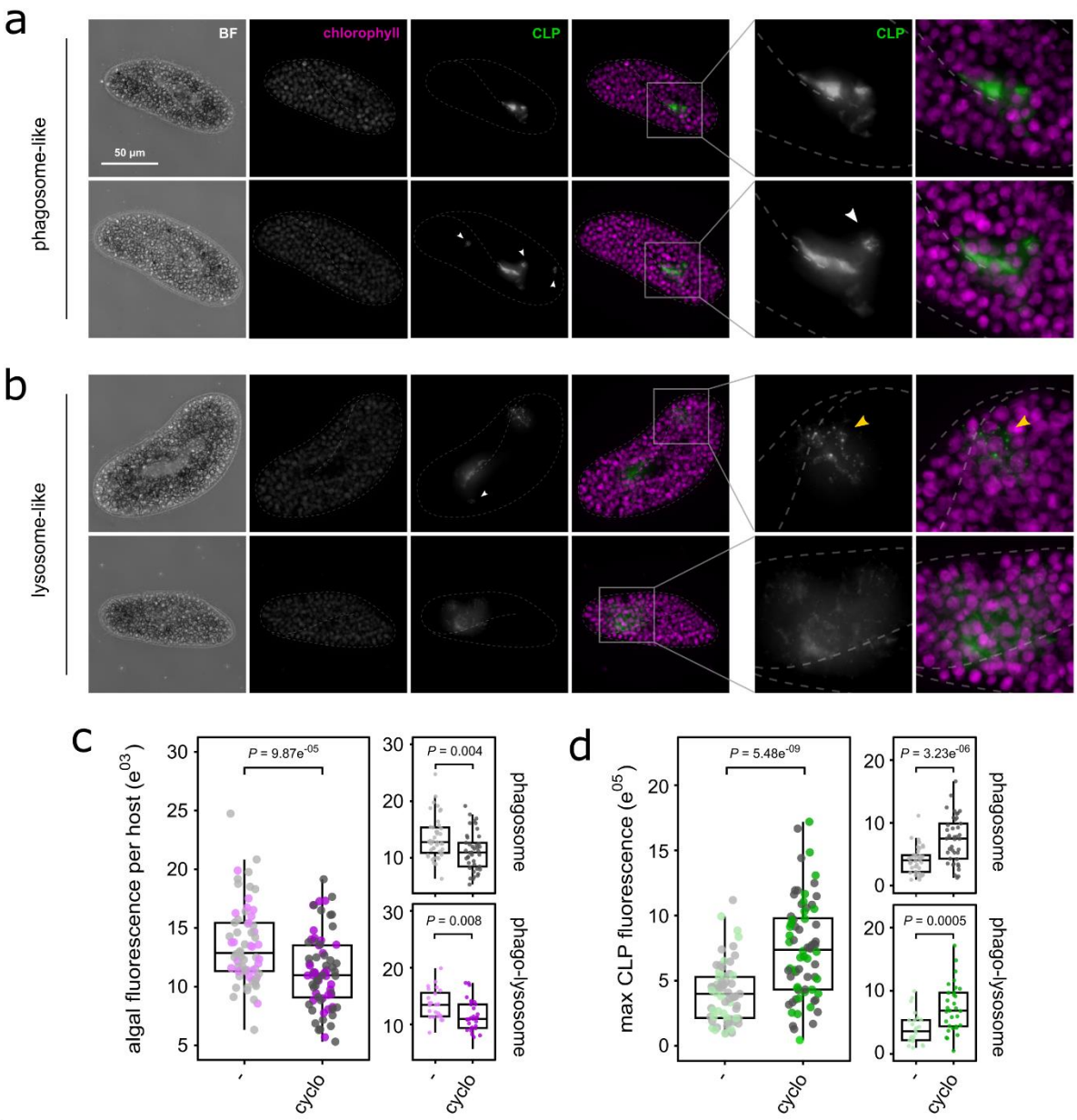


Figure 4: CLP localizes proximate to the putative host phagosome and lysosome during endosymbiont destruction. All experiments were conducted using *Pb* cells grown under standard conditions (- ; *n* = 70), or treated with 50 µgmL^{-1} cycloheximide (cyclo ; *n* = 69). All cells were treated a primary αCLP antibody followed by an Alexa Fluor 488-conjugated secondary antibody for immunofluorescence imaging. ***a-b***, Fluorescence imaging of *Pb* cells exhibiting phagosome-like (***a***) or lysosome-like (***b***) CLP localization. Magenta indicates endosymbiotic algal autofluorescence, green indicates Alexa Fluor 488 secondary antibody fluorescence. ***c-d*** Mean algal fluorescence (***c***) and max CLP fluorescence (***d***) per host cell in untreated (-) or cycloheximide treated (cyclo) cells. Cells were scored based on whether they exhibited phagosome- or phagosome/lysosome-like *CLP* localization. Boxplot data are represented as max, upper quartile (Q3), mean, lower quartile (Q1) and min values, and individual data points for each biological replicate are shown. Statistical significance was calculated using a linear model.

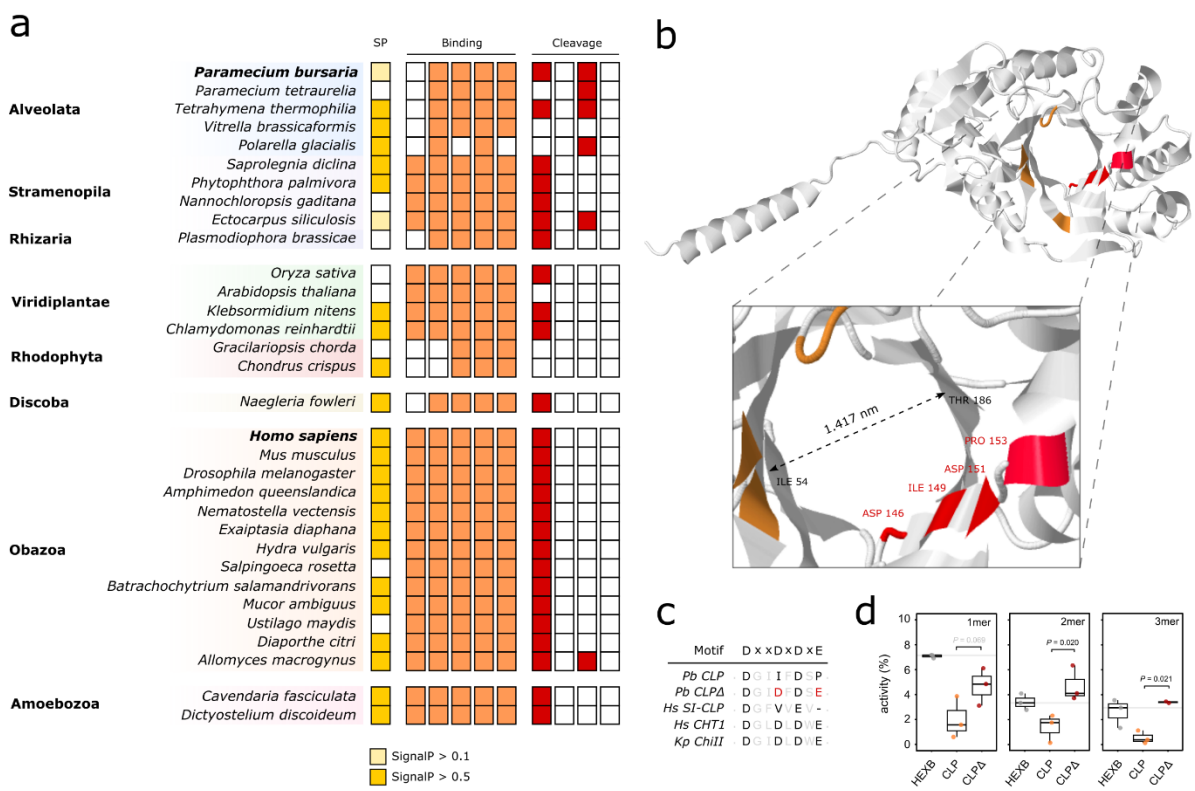


Figure 5: CLP is a chitin-binding protein homologous to an animal immune factor. a, Conserved functional domain residues of *CLP* across eukaryotes (see also **Fig. S7**) including predicted signal peptide (yellow) and characterised binding (orange) and cleavage (red) site amino acids. ***b,*** Predicted 3D structure of *Pb* CLP generated with AlphaFold3 (see also **Fig. S8**). Characterised binding (orange) and cleavage sites (red) are highlighted. ***c,*** Trimmed amino acid alignment of *Pb CLP* and *Hs SI-CLP* sequences, compared to known chitin-cleaving enzymes *Hs CHT1* and *K. pneumoniae ChiII*. Residues in bold are sites associated with a conserved chitin-cleavage motif. Residues in red indicate synthetically mutated sites to 'restore' chitin-cleaving function. ***d.*** Percentage activity of purified wild-type (CLP) and mutated (CLPΔ) *Pb* enzymes incubated for 3 hours in 1mer (GlcNAc), 2mer ($GlcNAc_2$; $Chitin_2$), or 3mer ($GlcNAc_3$; $Chitin_3$) chitin substrates. Enzymes were synthesized to encode a 5' yeast secretion signal, transformed into chitinase-deficient *Saccharomyces cerevisiae* strain Y16947, and expressed under selection for 7 days. Supernatant containing the expressed enzyme was concentrated, washed, and protein concentrations matched prior to the assay. Activity was standardised against supernatant from a yeast empty vector control. *Pb β-hexosaminidase* (**Fig. 2b**), an enzyme predicted to cleave terminal GlcNAc residues from glycan chains, was used as a positive control (HEXB). Boxplot data are represented as max, upper quartile (Q3), mean, lower quartile (Q1) and min values, and individual data points for each biological replicate are shown. Statistical significance was calculated using generalized linear models with quasi-binomial error distribution.

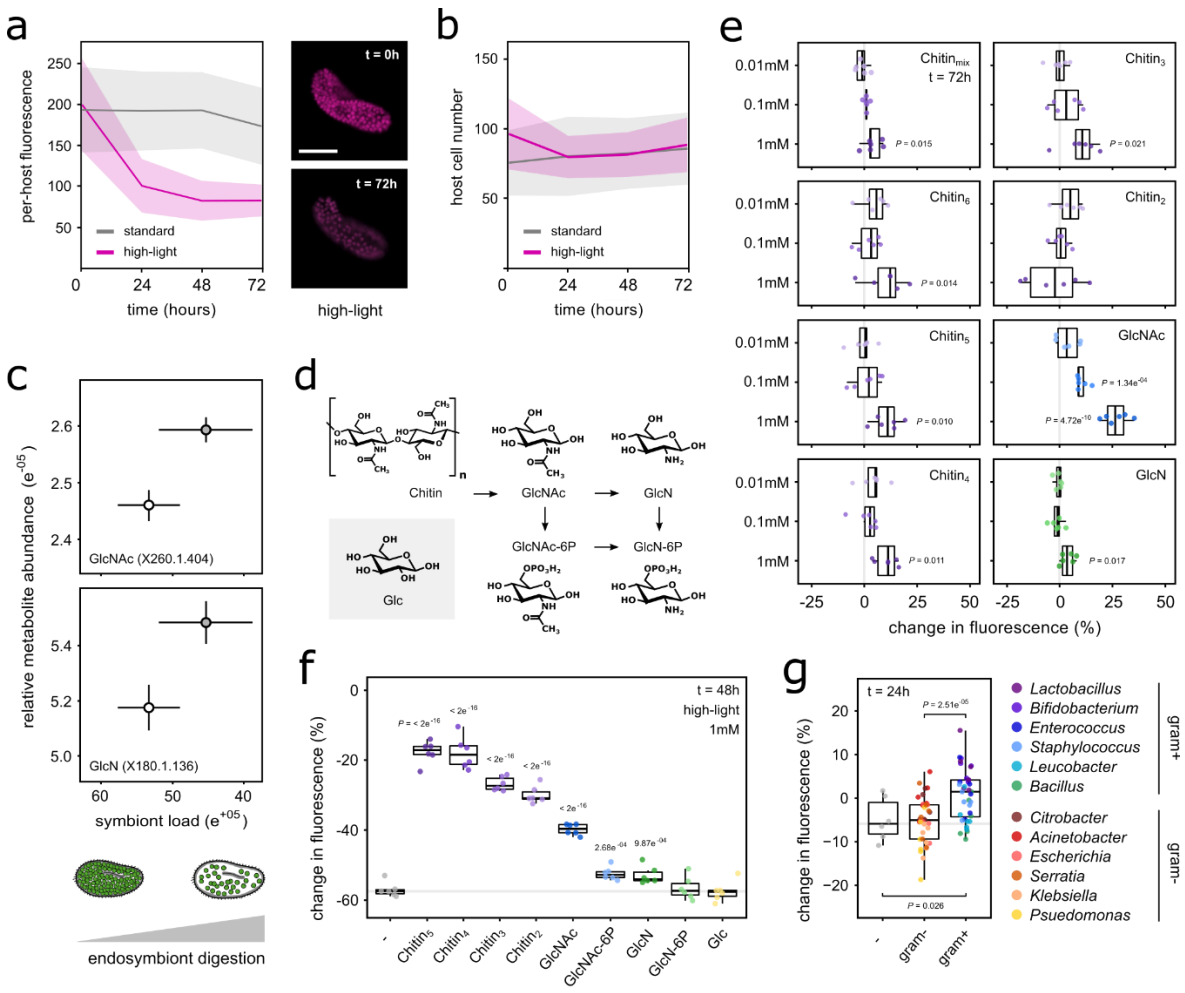


Figure 6: Glycan break-down products alter host control of endosymbiotic algae. All experiments were conducted using pre-starved *Pb* cultures (n = 6) grown under standard conditions (see **Methods**), unless stated otherwise. Growth under standard light was conducted at 25 µ mol m^{-2} s^{-1}, high-light at 80 µ mol m^{-2} s^{-1}, moderate high-light at 50 µ mol m^{-2} s^{-1}, and low-light at 6 µ mol m^{-2} s^{-1}. Per host fluorescence is used throughout as a proxy for endosymbiotic algal load. ***a***, Mean per host fluorescence in *Pb* cultures grown under high-light. Data are presented as mean ± sd. Scale bar = 50µm, magenta indicates endosymbiotic algal autofluorescence. ***b***, Host cell number in *Pb* cultures grown under high-light for 72 hours. Data are presented as mean ± sd. ***c***, Relative metabolite abundance in *Pb* cultures under variant light conditions, in which differential endosymbiont digestion is known to occur. Cells were grown under moderate high-light (white) or low-light (grey) for 3 days prior to metabolite extraction (**Fig. S10**). ***d***, Schematic metabolic pathway of chitin break-down, including the structure of glucose (Glc) for comparison. ***e***, Percentage change in per host fluorescence following exposure to chitin and chitin break-down products, compared to an untreated control. Cells were exposed to chitin (mixed oligosaccharides, or oligosaccharide lengths 6 to 2), GlcNAc, or GlcN for 72 hours. ***f***, Percentage change in per host fluorescence upon exposure to chitin break-down products during high-light initiated endosymbiont digestion. Cells were grown under high-light and exposed to chitin (oligosaccharide length 5 to 2), chitin break-down products (see ***d***), or Glc for 48 hours. Each substrate was added to a final concentration of 1mM. ***g***, Pooled percentage change in per host fluorescence following bacterial feeding in standard growth conditions. *Pb* cells were grown for 24 hours, and bacteria were heat-killed and matched to OD 0.1 prior to feeding (**Fig. S12**). All boxplot data are represented as max, upper quartile (Q3), mean, lower quartile (Q1) and min values, and individual data points for each biological replicate are shown. Statistical significance was calculated using a linear model on log-transformed data (**e**, **f**, **g**). Conditions where $p < 0.05$ are shown. GlcNAc, N-acetyl-D-glucosamine; GlcN, D-glucosamine; GlcNAc-6P, N-acetyl-D-glucosamine-6-phosphate; GlcN-6P, D-glucosamine-6-phosphate; Glc, Glucose.