Evolution of resilience against heat stress in a red-tide dinoflagellate

Katherine Dougan[1*], Zhi-Luo Deng[2*], Lars Wöhlbrand[3*], Carsten Reuse[4*], Boyke Bunk[5], Yibi Chen[1], Juliane Hartlich[4], Karsten Hiller[4], Uwe John[6], Jana Kalvelage[3], Johannes Mansky[4], Meina Neumann-Schaal[5], Jörg Overmann[5], Jörn Petersen[5], Selene Sanchez[4], Kerstin Schmidt-Hohagen[4], Sarah Shah[1], Cathrin Spröer[5], Helena Sztajer[4], Hui Wang[4], Debashish Bhattacharya[7], Ralf Rabus[3], Dieter Jahn[4], Cheong Xin Chan[1§#], Irene Wagner-Döbler[4§#]

[1]The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia
[2]Helmholtz-Center for Infection Research (HZI), Inhoffenstr. 7, 38124 Braunschweig, Germany
[3]Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg (HIFMB), Ammerländer Heerstraße 231, 26129 Oldenburg, Germany
[4]Braunschweig Center for Systems Biology (BRICS) at the Technical University of Braunschweig, Rebenring 56, 38106 Braunschweig, Germany
[5]German Culture Collection for Microorganisms and Cell Cultures (DSMZ), Inhoffenstraße 7B, 38124 Braunschweig, Germany
[6]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany
[7]Rutgers University, 59 Dudley Road, Foran Hall 102, New Brunswick, New Jersey, 08901 USA

*These authors contributed equally (data acquisition, analysis and interpretation)
§These authors contributed equally (design of the study, integration of all findings)

#corresponding authors: Cheong Xin Chan, The University of Queensland, Brisbane, Australia, email c.chan1@uq.edu.au and Irene Wagner-Döbler, Technical University of Braunschweig, Germany, email I.Wagner-Doebler@tu-braunschweig.de.

Abstract

Dinoflagellates are a large, ecologically important phylum of marine unicellular algae. Their huge genomes make it highly challenging to decipher the genetic basis of key processes such as harmful algal bloom (HAB) formation and response to warming oceans. To address these issues, we generated a high quality genome assembly from *Prorocentrum cordatum*, a globally abundant, HAB forming dinoflagellate. Our analyses demonstrate massive expansion of the gene inventory to 85,849 predicted genes, primarily driven by unusually long and frequent introns and dispersed duplicates enriched for bloom relevant functions. We find that cell yield is reduced at higher culture temperatures. To understand this response, we integrated transcriptome, proteome and metabolome data and identified both a global and a temperature specific heat-stress response. The underlying metabolic changes reflect damage to photosynthesis and central metabolism. The transcriptome data show that ~25% of genes are differentially expressed under heat stress, with concomittant extensive RNA editing and alternative exon usage. Multi-codon genes and transcripts for HSP70 and RuBisCo suggest a polycistronic gene organisation. Our work represents the first genome based analysis of a red tide dinoflagellate and demonstrates that temperature resilience in *P. cordatum* is mediated by a unique genome structure and multi-level transcriptional regulation.

Introduction

Harmful algal blooms (HABs) are extreme increases of algal abundance triggered by a combination of environmental factors, among them temperature, light intensity and nutrients, HABs are harmful to surrounding organisms and ecosystems through a combination of oxygen depletion and toxicity [1][2] and negatively impact fisheries and aquacultures worldwide, causing ~USD 8B per year in losses [3]. Often referred to as "red tides", due to the color of the algal populations, HABs are predominantly caused by bloom-forming species of dinoflagellates, such as *Alexandrium minutum* [4] and *Prorocentrum minimum* [5] recently renamed to *P. cordatum* [6] that have expanded their habitat and caused increasingly frequent red tides, linked to warming oceans and global climate change [7][8][9]. However, the molecular basis of the heat response in dinoflagellates has not been systematically investigated. Genomic studies of these species remain challenging even today due to their immense genome sizes (up to ~200 Gbp) [10] and their idiosyncratic genome features that are atypical of eukaryotes [11][12].

Dinoflagellates represent a diverse group of phytoplankton encompassing a broad spectrum of ecological niches, including free living, bloom forming, parasitic and symbiotic lineages, the latter comprising essential symbionts in coral reefs [13]. About one half of the studied taxa are are mixotrophic, i.e. they combine phototrophic and heterotrophic modes of energy generation, and the other half are heterotrophic [14][15]. The dinoflagellate photosynthetic machinery [16] powers the biological carbon pump, HABs and coral symbiosis, all of which are deeply affected by increased ocean temperatures. However, key questions of dinoflagellate biology remain unresolved. Transcriptional regulation is considered to be minimal in dinoflagellates [11], since few transcriptional regulators have been identified [17], and the chromosomes remain in a liquid crystalline state throughout the cell cycle [18]. Typically, very few genes show differential expression under temperature stress (e.g. [19]). Editing of the mRNA of a small fraction of nuclear genes has been discovered and suggested as an alternative pathway of generating physiological flexibility [20]. These interpretation are however complicated by the fact that most studies were performed on complex symbiotic systems, e.g. the coral holobiont, or dinoflagellate cultures with their associated bacterial microbiome, whose physiological interactions could mask mechanisms of transcriptional regulation [21]. The recent discovery of transcription associated domains (TADs) offers a

new approach for understanding transcriptional regulation in dinoflagellates [22][23]. Whereas transcriptome data are accumulating, much less proteome and metabolome analyses have been performed (e.g. [18][24]), and therefore an integrated understanding of heat stress response of dinoflagellates is lacking.

Among bloom-forming dinoflagellates, *P. cordatum* is an invasive, potentially toxic species [25][26][27]. Initially described in Asia, *P. cordatum* has expanded its habitat to coastal ecosystems worldwide [26], including the North Atlantic, where it is now regularly detected [28]. *P. cordatum* is tolerant of a wide range of salinities and temperatures, facilitating the increased frequency of its HABs [7][5].

Here, we present the *de novo* assembled genome of *P. cordatum*, the largest genome of a dinoflagellate sequenced to date, and provide transcriptome, proteome and metabolome data specific to the heat-stress response in distinct growth phases. Integrating the multi-omics datasets with genome structure, we demonstrate how genome evolution drives heat-stress resilience of a HAB-forming dinoflagellate.

Results

P. cordatum genome reveals hallmark of free-living dinoflagellates

We generated a *de novo* genome assembly from *P. cordatum* CCMP1329 of size 4.15 Gb (scaffold N50 length 349.2 Kb; Table S1). The assembled haploid genome (Figure S1) shares a similar level of completeness to other dinoflagellate genomes based on the recovery of core alveolate genes (60.2%; Extended data Table 1). Most (61.0%) assembled genome bases were annotated as repetitive elements, 46.0% represent unknown repeats, and 7.8% are simple repeats (Figure 1a). To assess the impact of genome and gene features on ecology and evolution, we examined them in *P. cordatum* relative to five representative taxa chosen based on the quality of available genome data and ecological niches: the free-living *Polarella glacialis* CCMP1383 [29] and *Symbiodinium natans* CCMP2548 [30], the symbiotic *Durusdinium trenchii* CCMP2556 [31] and *Cladocopium goreaui* SCF055 [32], and the parasitic *Amoebophrya ceratii* AT5.2 [33]. Except for the Order *Suessiales* (largely symbiotic) and the early-branching *Syndiniales* taxa (parasitic), the other dinoflagellates in the phylogenetic tree are free-living and bloom-forming (Figure 1b).

104 Among the six genomes, we observed the highest overall G+C content in *P. cordatum* (mean
105 59.7%; Figure 1c and Extended data Table 2) and in the 85,849 protein-coding genes (mean
106 65.9%; Figure 1d and Extended data Table 2; see Methods). Among the latter we found 47
107 genes potentially acquired via horizontal transfer from uncultivated marine taxa
108 (Supplementary Note, Table S1 and Figure S2). High G+C contributes to sequence stability
109 due to the more thermally stable hydrogen bonds in the G-C pairing. This observation may
110 reflect an adaptive mechanism in *P. cordatum* to ensure higher fidelity of transcription when
111 the cells are exposed to rapidly fluctuating temperatures, e.g. during diurnal vertical
112 migration or in spring blooms and red tides. The haploid genome size of *P. cordatum*
113 (estimated at 4.75 Gb; Figure S1) is the largest (e.g. 3.2-fold larger than that of the free-living
114 *Po. glacialis*, 39.6-fold larger than that of the parasitic *A. ceratii*; Figure 1e), and encodes the
115 largest number of protein-coding genes yet described in dinoflagellates (e.g. 4.3-fold greater
116 than those in *A. ceratii*; Figure 1e). *P. cordatum* also has the longest genes (mean length 24.5
117 Kb; Extended data Table 2) and introns (mean length 4.7 Kb) among the six studied taxa
118 (Extended data Table 2 and Figure 1e). All five dinoflagellate species had a high intron
119 frequency (between 9.8/gene in *P. cordatum* up to 16.1/gene in *Cladocopium goreaui*),
120 except for the parasitic *A. ceratii* (2.4/gene) (Extended data Table 2). Introns of *P. cordatum*
121 and *Po. glacialis* are enriched in introner elements, i.e. introns that contain simple inverted
122 repeats at both ends (Table S2 and Figure 1e), suggesting a prevalence of these elements in
123 the genomes of free-living dinoflagellates (Table S3 and Supplementary Note). The *P.*
124 *cordatum* genome has the largest proportion of helitron repeats (1.2% of the assembled bases
125 compared to <0.8% in the other genomes; Figure 1e) whereas long interspersed nuclear
126 elements and long-terminal repeats are expanded in the genomes of *Suessiales* taxa including
127 *Po. glacialis*. These elements are largely conserved in the *P. cordatum* genome (Figure S1).
128 See Supplementary Note for a detailed analysis of tandemly repeated genes, unidirectionally
129 encoded genes (Figure S3) and introner elements in *P. cordatum.*

130 Based on analysis of colinear gene blocks (see Methods), we found no evidence of segmental
131 duplication. This stands in stark contrast to the symbiotic *D. trenchii* for which 49.4% of the
132 genes were implicated suggesting whole-genome duplication [31]. Instead, 55,664 (64.8% of
133 85,849) genes of *P. cordatum* are dispersed duplicates (Table S4 and Figure 1e), contributing
134 to the large genome size. Our results suggest that most duplication events occurred
135 independently, or alternatively, colinearity of duplicated blocks was disrupted by extensive

136 rearrangements due in part to the abundant transposable elements. High gene content, long
137 introns, and extensive genetic duplication represent the signature of the large genome of this
138 bloom-forming dinoflagellate.

139 **Duplication enhanced diversity of gene functions in a bloom-forming dinoflagellate**

140 We annotated putative functions of predicted proteins of *P. cordatum* based on shared
141 sequence similarity covering ≥ 70% of full-length sequences in the UniProt database (see
142 Methods, Table S5). This stringent approach ensures high confidence of the annotated
143 functions given the scarcity of dinoflagellate sequences in public repositories and the
144 expected extensive genomic divergence among dinoflagellate genes [30]. Of the 85,849
145 proteins of *P. cordatum*, 41,078 (47.8%) were annotated with a putative function, 20,897
146 (24.3%) with Gene Ontology (GO) terms; similar proportions were observed in other
147 genomes (Table S6), with large proportions (e.g. 52.15% for *P. cordatum* and 62.83% for *A.*
148 *ceratii*) coding for functions yet to be discovered [34] and/or are remote homologs to
149 sequences in UniProt.

150 We assessed gene functions in *P. cordatum* relative to the other five genomes based on
151 relative abundance of GO terms annotated per genome (Figure 2a) and the total count of GO
152 terms (Figure S4; see Methods). Biological processes that were relatively more abundant in
153 *P. cordatum* compared to all other genomes were lipid metabolic process, DNA
154 recombination, DNA integration, carbohydrate metabolic process, proteolysis, signal
155 transduction, and DNA repair (Figure 2a). Molecular functions related to activities of
156 transmembrane transporter, methyltransferase, oxidoreductase, DNA binding, nucleic acid
157 binding, protein kinase, serine-type endopeptidase and hydrolase were likewise relatively
158 more abundant in *P. cordatum* compared to all other genomes (Figure 2a). For cellular
159 components, the dominant functions relate to extracellular region and integral component of
160 membrane (Figure 2a). The free-living, but non-bloom forming, cold adapted *Po. glacialis*
161 exhibited pronounced but complementary functional enrichments, indicating that the highly
162 abundant biological processes, molecular functions and cellular components found in *P.*
163 *cordatum* are typical of a bloom forming dinoflagellate (Figure 2a). The parasitic (non
164 photosynthetic) *A. ceratii* has only 2,799 genes with annotated GO terms, compared to
165 20,897 in *P. cordatum*, therefore the relative abundance of its functions is not comparable to
166 the other genomes (Figure S4).

167 To assess the impact of gene duplication on function, we assessed, in each genome, the
168 enrichment of functions among genes of distinct duplication modes (i.e. dispersed, proximal
169 and tandem) for each genome (see Methods). Figure 2b shows the significantly enriched ($p \leq$
170 0.01) GO terms in each category and the associated number of genes. In *P. cordatum*, gene
171 functions related to transmembrane transport, homeostasis and organelle assembly are
172 enriched among the dispersed duplicates, whereas those related to immune response and
173 photosynthesis are enriched among proximal duplicates, and those implicated in metabolic
174 processes (e.g. tricarboxylic acid cycle) and binding of biomolecules/ions are enriched among
175 tandem duplicates (Table S7). Interestingly, these terms are not significantly enriched ($p >$
176 0.05) in the distinct duplication modes for the other genomes (Figure 2b), although GO terms
177 related to transmembrane and cation transports are also abundant (implicating >100 genes per
178 genome) in the free-living *Po. glacialis* and *S. natans* (Figure 2b). These results demonstrate
179 the biological significance of these duplicated gene functions in *P. cordatum* and how distinct
180 duplication modes have shaped the evolution of genes that are central to adaptation.

181 **Integrated multi-omics analysis of heat stress in *P. cordatum***

182 To investigate the heat-stress response of *P. cordatum*, the axenic strain was grown in
183 defined mineral media at its optimal temperature of 20°C and from there it was shifted to
184 26°C and 30°C (Figure 3a). These temperatures are already common today in the Red Sea
185 [35][19] or at the equator in the summer [36] and are well within the range predicted for the
186 future ocean [37]. *P. cordatum* grew at a similar rate at all three temperatures. However, heat
187 stress reduced algal biomass already at 26°C, and more so at 30°C, with 62% and 41% of
188 final cell density observed, respectively, compared to 20°C (Figure 3b). Stable cell numbers
189 over two weeks at stationary phase indicate resilience of *P. cordatum* to the elevated
190 temperatures. Replicate cultures were harvested at exponential (Ex) and stationary (St)
191 growth phase and samples for transcriptome, proteome and metabolome analysis were
192 collected (Tables S7-S10).

193 We generated 1.964 Gb of transcriptome data (~110 million reads per sample; Table S8)
194 from 18 samples (6 conditions × 3 replicates). These data support 72,839 (84.8%) of the
195 85,849 predicted gene models (Extended data Table 2) based on shared sequence similarity
196 of assembled transcripts. We found 62,580 genes to be expressed (av. read count ≥1 for all
197 samples). The ternary plot displays the sub-transcriptome of highly expressed genes (average

198 count per million >32), showing that gene expression in the exponential phase was similar at
199 all three temperatures, while in the stationary phase, whereas major differences were
200 observed between the 26°C and 20°C treatments (Figure 3c). PCA based on gene expression
201 also showed large separation between 20°C stationary phase and higher temperatures (Figure
202 S5).

203 Our analyses of soluble and membrane protein fractions by LC-ESI-iontrap MS/MS yielded
204 2,098 identified proteins. Of these, 1,032 were of unknown function including 244 that are
205 unique to *P. cordatum*. The 68 chromophore-binding (antennae) proteins of the photosystem
206 comprised the largest group accounting for 70.6% of detected peptides. Proteins related to
207 photosynthesis, energy generation, and central metabolism were among the most abundant
208 proteins under all conditions. We found 779 proteins with significantly changed abundance at
209 26°C or 30°C compared to 20°C, which clustered according to growth phase and temperature
210 (Figs. 3d, Figure S6 and Table S10).

211 The metabolome assessed by GC-MS analyses yielded 173 compounds, of which 73
212 metabolites could be identified (Extended data Figure 3). Fifty-four metabolites displayed
213 significantly changed abundances ($p<0.05$) in response to growth stage and temperature,
214 particularly those involved in central metabolism and amino acid biosynthesis (Table S11).

215 Integration of the transcriptome and proteome data using DIABLO [38] allowed us to
216 distinguish two types of temperature responses: A generalized heat stress response
217 (component 1) with abundance changes common in both increased temperatures (26°C and
218 30°C) compared to 20°C (Figure 3e, left panel), and a temperature-specific heat stress
219 response (component 2) with abundance changes specific for 26°C or 30°C relative to 20°C
220 (Figure 3e, right panel). Transcripts and proteins implicated in those two components are
221 shown in Table S12. KEGG pathways that are enriched in each component or in both are
222 shown in Figure 3f. KEGG pathways for carbon metabolism, pentose phosphate pathway,
223 glyoxylate and dicarboxylate metabolism, biosynthesis of amino acids, and cysteine and
224 methionine metabolism were enriched in generalized heat stress response (component 1,
225 Extended data Figure 1). In contrast, oxidative phosphorylation, protein processing in
226 endoplasmic reticulum, vitamin digestion and absorption, and pyruvate metabolism pathways
227 were enriched according to the level of heat stress (Extended data Figure 2). The correlation
228 between the transcript and protein expression within component 1 is shown in Fig. 3g (detail

229 in Figure S7; similar correlation for component 2 is shown in Figure S8). Both positive and
230 negative correlations of expression between transcripts and proteins were observed. For
231 instance, the expression of the transcript encoding glutamine synthetase was positively
232 correlated to the expression of chlorophyll *a-b* binding protein and the light-harvesting
233 complex I LH38 proteins, while it negatively correlated to expression of pyruvate
234 dehydrogenase, sulfate and formate transporters.

235 **Temperature-related modulation of central metabolism in *P. cordatum***

236 A global visualisation of the data revealed that the associated sub-transcriptome (753
237 transcripts) was differently impacted according to growth stage and temperature, whereas the
238 associated sub-proteome (278 proteins) featured a marked increase in photosynthesis proteins
239 at higher temperatures, paralleled by a decreased abundance of proteins belonging to central
240 metabolism and oxidative phosphorylation (Fig. 4a). The 38 metabolites associated with
241 central metabolism followed largely the temperature-dependent profiles of the sub-proteome.

242 The detected components of the photosynthetic electron transport chain (PETC) showed
243 varying abundance profiles (Figure 4b). Whereas chlorophyll-binding proteins (CPs), light
244 harvesting complex (LHC), stress related proteins and photosystem I (PS I) proteins
245 increased at high temperature, detected subunits of the chloroplast ATP synthase generally
246 remained constant. Notably, PS II and the associated oxygen evolving complex (OEC) and
247 enzyme components of the ROS response (Tables S13 and S14) appeared unchanged, but
248 complex IV and mitochondrial ATP synthase of oxidative phosphorylation displayed reduced
249 abundances at higher temperatures (Figure 4d). Taken together, during heat stress *P.*
250 *cordatum* faces energy deprivation arising from a less efficient PETC, which should
251 immediately affect protein synthesis and central metabolism. Accordingly, amino acid
252 synthesis decelerated, as evident from decreased levels of involved enzymes (e.g. PSPH,
253 GLS) and amino acids (e.g. serine, glutamate) as well as TCA-cycle intermediates (e.g.
254 succinate) (Figure 4c). A detailed reconstruction of central metabolism is provided in
255 Extended data Figure 4. These results predict lower levels of enzymes involved in
256 concentrating CO_2 (PEPC) and in ATP-consuming reactions of the Calvin cycle (PRK, PGK)
257 (Figure 4c). The mostly unchanged profile of ribulose 1,5-bisphosphate carboxylase (Calvin
258 cycle) may be misleading in this context, because this key enzyme has decreased activity at

259 higher temperatures [39]. Enzymes involved in the synthesis of amylose remained stable (SS,
260 GBSS), possibly to maintain storage formation during imbalance in energy generation (Fig.
261 6b). In contrast to the declining levels of TCA metabolites and amino acids, some
262 carbohydrates and fatty acids increased with rising temperature, suggesting their participation
263 in protection of the cell against heat stress or as countermeasure against increased membrane
264 fluidity [40]. These results demonstrate the impact of heat stress on a broad range of
265 metabolic processes in *P. cordatum* such that its growth performance is severely affected.

266 **The transcriptome landscape of *P. cordatum* under heat stress**

267 The transcriptome landscape of *P. cordatum* was dominated by differential expression (DE)
268 of genes, which occurred for 2,142 genes in the exponential phase of growth, and 22,924
269 genes in the stationary growth phase (Figure S5b, Table S15) which is concordant with the
270 PCA (Figure S5a) and ternary plot (Figure 3c). The largest difference (14,322 DE genes) was
271 observed between 20°C and 26°C in the stationary phase (Figure S5b).

272 We analyzed the expression patterns of dispersed duplicated genes in all ortho-groups
273 between 20°C and 26°C in the stationary phase. The majority were regulated similarly. In
274 total, 368 ortho-groups had both genes being DE genes among 2,368 groups with two
275 dispersed gene members, of which 341 were regulated similarly (Figure 5a). For groups with
276 three dispersed gene members, 982 out of 994 groups had at least two genes regulated
277 similarly (Figure 5a). More generally, 813 out of 949 ortho-groups (85.7%) with at least three
278 DE dispersed genes had more than 80% of the DE genes regulated similarly (Figure 5a).

279 **Functional enrichment analysis of transcriptional response to temperature stress**

280 DE genes clustered clearly according to temperature and growth condition (Figure 5b). GO
281 term enrichment analysis based on DE genes was performed for both the Ex and St phase of
282 growth. At 26°C in the St phase, GO terms including inositol oxygenase activity, cellulase
283 activity, ATP-binding and metal ion binding were upregulated (Extended data Figure 5a,
284 Table S16), whereas ribosome, rRNA binding, translation, transmembrane transporter,
285 photosystem II and photosynthesis light reaction were downregulated (Extended data Figure
286 5b). In the Ex phase, structural constituents of cytoskeleton and microtubule were in the set
287 of downregulated genes (Table S16). The findings of a KEGG pathway enrichment analysis
288 based on KO annotation of genes were consistent with the GO term enrichment Extended

289 data Figure 5c-d, Table S16, S17, S18). A clustering analysis based on gene expression
290 showed 8 (Ex growth phase) or 7 (St growth phase) clusters with clear patterns of up- and
291 down-regulation under the different treatment conditions (Figure 5b, Extended data Figure
292 6a-b, Table S19 and S20) and GO term enrichment confirmed those reported above
293 (Extended data Figure 6c). Among 13 polyketide synthase I (PKS I) genes, which are the
294 major dinoflagellate toxin producing genes [41] five were upregulated under heat stress while
295 none was downregulated (Table S9).

296 **Posttranscriptional regulation: editing of mRNAs and differential exon usage**

297 Integrating genome and transcriptome data from *P. cordatum*, we examined two modes of
298 post-transcriptional regulation for generating diversity of proteins (and thus function) in
299 distinct growth conditions: (a) alteration of a single base in mRNA (i.e. differential RNA
300 editing), and (b) alternative splicing that leads to preferential usage of exons (i.e. differential
301 exon usage [DEU]). We found evidence of post-transcriptional regulation in 4,550 genes:
302 2,744 genes (implicating 4,410 sites) with differential RNA editing, 1,755 genes (implicating
303 5,837 exons) with DEU (Table S21), and 51 genes (implicating 93 sites and 72 exons) with
304 both modes of regulation (Figure 5c). The types of mRNA edits were similar under all
305 conditions (Figure 6b), but number and overlap between edited sites was condition-specific
306 (Figure 6b, Table S22).

307 The relatively few genes with both RNA editing and DEU in *P. cordatum* suggests post-
308 transcriptional regulation of a gene tends to involve only one of these two modes.
309 Accordingly, the editing of RNA was identified predominantly in genes for which no DEU
310 was identified, and vice versa (Figure 5c). The complementarity between the two modes of
311 post-transcriptional regulation may present a regulatory mechanism for generating functional
312 diversity of the encoded proteins in *P. cordatum*, and potentially more broadly in
313 dinoflagellates, under variable environmental conditions. Interestingly, the singleton genes in
314 *P. cordatum* appear to show a greater extent of DEU than the dispersed duplicates, which
315 exhibit a greater extent of RNA editing (Figure 5c). The 51 genes displaying both modes of
316 regulation encode key photosynthetic and stress-related functions, such as chloroplastic ATP
317 synthase subunits (beta and epsilon subunits), ferredoxin, photosystem I reaction center
318 subunit XI, and heat shock 70 kDa protein.

319 **Structure and expression of Hsp70 and RuBisCo gene models**

320 Interestingly, we found protein-coding sequences organized in multiple sub-regions, for
321 which each sub-region encodes a full-length copy of the functional protein; we follow [42] to
322 define such a sub-region within a gene model as a coding unit (CU).

323 The heat-shock protein 70 kDa (HSP70) is a ubiquitously formed chaperone with conserved
324 functions in the stress response. We identified 16 gene models in the genome of *P. cordatum*
325 that putatively encode HSP70s with sizes ranging from 574 to 2,575 amino acids. All of these
326 gene models contain a single CU except for s12246_g74608, which encodes four complete
327 HSP70s along with a fragmented fifth gene. This multi-CU gene model is the main
328 contributor to the heat stress response of HSP70 (Fig. 6a). Exon 1 (CU-1) of the multi-CU
329 transcript displayed a unique pattern of RNA editing (Fig. 6b), with contrasting growth phase
330 vs. temperature-specific RNA edits. We found multiple RNA-edited sites upstream of the
331 transcriptional start site (Supplementary Fig. S21), that suggest a post-transcriptionally
332 regulated heat stress response common to both growth phases. The usage of exon 3 (CU-2)
333 was biased towards a higher abundance at 20°C St phase, corresponding with multiple
334 intronic RNA-edited sites downstream of this exon only identified under this condition.
335 Interestingly, the HSP70 encoded by exon 3 (CU-2) lacks a disordered C-terminus region
336 involved in protein-protein interactions in many species.

337 The CO_2-fixing ribulose-1,5-bisphosphate carboxylase (RuBisCo) is the key enzyme of the
338 Calvin cycle. The genome of *P. cordatum* contains 32 gene models that putatively code for
339 RuBisCos of varying sizes (163 to 5,197 amino acids, Figure S10). Half of these gene models
340 encode fragments of the protein, whereas the remainder are organized in single or multiple
341 CUs, similar to what we observed with HSP70 (Fig. 6d). Previous work on *P. minimum* [43]
342 and *P. donghaiense* [42] showed a single transcript that contains three consecutive CUs. Its
343 unprocessed protein product is hypothesized to be transported into the plastid where it is split
344 into three separate, active proteins [42]. Notably, here we found four gene models that
345 harbour three CUs, one with two CUs, and eleven with a single CU (Fig. 6d).

346 We analysed the upstream regions of the gene models as well as the spacers to determine if
347 they encoded transit peptides (TP) necessary for transport into the plastid. No consensus
348 sequence for such a TP is described for dinoflagellates and TPs for plastid localization are
349 highly diverged among phototrophs [44][45]. The predicted leader sequence from *P.*

350 *donghaiense* [42] was found in eight of these gene models (Figure 5d). Comparing the
351 regions upstream of the CUs, we identified two conserved sequence motifs in (a) the leader
352 sequence, (b) the N-terminal region of proteins that lacked a leader sequence, and (c) the
353 spacer region between two CUs (Fig 6e). The two sequence motifs are moderately
354 hydrophobic and rich in basic amino acid residues (R/K) as expected for plastidic TPs [45].
355 Thus both the single-CU and multi-CU proteins can potentially be transported into the
356 plastid, based on putative TP either in their upstream region or in the spacer. Highly similar
357 sequence motifs in RuBisCo from *P. donhaiense* and other dinoflagellates (Figure S11)
358 further support the hypothesis of their involvement in transport to the plastid.

359 Proteins encoded by the different CUs share very high pairwise sequence identity (>95%)
360 making it impossible to map detected peptides to a given CU. The cumulative abundance of
361 proteins increased at 20°C in the St phase, whereas transcript abundance increased with
362 temperature (Figure 6c). Analysis of the transcriptional response showed clustering of gene
363 models according to the number of CUs. Multi-CU transcripts had a consistently low
364 expression level while single-CU derived transcripts were highly expressed and responded
365 strongly to heat stress (Fig5, Extended data Figures 7 and 8). Only seven out of 16 complete
366 RuBisCo gene models were interrupted by introns. No significant change in exon usage and
367 little mRNA editing was observed for these sequences. Analyses of the promotor regions of
368 the gene models revealed promotor motifs in eight cases, including one zinc finger motif, but
369 these were not related to the transcriptional profile. Therefore, regulation of this key enzyme
370 in dinoflagellates occurs primarily by controlling the number of CUs available for translation.

371 Discussion

372 Here we show how a HAB forming dinoflagellate responds to heat-shock based on its
373 genome, which is the largest of any dinoflagellate sequenced to date. Genome expansion is
374 mediated by the long genes and introns [46][47], as well as a huge fraction of dispersed
375 duplicates that are enriched for functions potentially relevant for bloom forming. The
376 extensive intron gain in *P. cordatum* points to active transposition and an important role of
377 introner elements, i.e. non-autonomous DNA transposons [48][49][50].

378 *P. cordatum* mounts a systems level heat-shock response already at 26°C, with few additional
379 adaptations required for it to grow and persist stably at 30°C. Multi-omics data integration

380 showed functions essential for energy production (photosynthetic electron transport chain)
381 and energy consumption (fixation of carbon dioxide, protein synthesis) to be impaired,
382 consequently resulting in reduced central metabolism and protein synthesis. Our findings
383 extend those from previous studies [40][51][52][53][54] and represent the first
384 comprehensive analysis of heat-shock response in a bloom forming dinoflagellate under
385 strictly phototrophic conditions, thus providing a basis for understanding dinoflagellate
386 ecology.

387 Heat resilience is mediated by multi-level mechanisms of regulation. Previously it was
388 thought that most genes in dinoflagellates are expressed constitutively [11][55][35][19].
389 Remarkably, we found strong evidence against this hypothesis, with ~25% of all genes being
390 DE in response to heat stress, comparable to other microalgae [56]. Rather than being a
391 structural feature of dinoflagellates, the small number of DE genes previously detected might
392 be related to the complexity of coral holobionts that obscures transcriptional regulation (e.g.
393 [57][21][58]). A large fraction of DE genes was either subject to RNA editing, *or* subject to
394 alternative exon usage, two mechanisms of posttranscriptional regulation generating
395 functional diversity. They tend to occur in different genes in *P. cordatum*, most likely
396 improving resilience towards heat stress.

397 Two important enzymes of heat-stress response, HSP70 and RuBisCo, demonstrated a novel
398 mechanism of transcriptional regulation. Some of their genes encoded several complete
399 proteins and were transcribed in a condition-specific way. Transcription of several ORFs into
400 a single (polycistronic) transcript is common in prokaryotes, but was unknown in eukaryotes
401 until recently [59]. Interestingly, HSP70 was among the first examples for a polycistronic
402 gene in eukaryotes [60]. Meanwhile many more have been discovered [61][59][62]. We
403 found them here in *P. cordatum* and show that they represent an important component of the
404 heat-shock response.

405 This first genome based multi-omics analysis of heat stress in a HAB forming dinoflagellate
406 has significantly advanced our understanding of dinoflagellate biology and evolution. It
407 provides a baseline for studies of the role of dinoflagellates in the biological pump, HABs
408 and symbiotic interactions with marine invertebrates. Beyond the many novel insights into
409 genome structure and regulation we describe, the *P. cordatum* genomic resources can be used

410 to design field experiments of HABs to explore mitigation strategies. This aspect is of
411 growing importance as the world's oceans continue to warm and HABs become more
412 frequent and widespread.

413 # Methods

414 **Strain and cultivation conditions**

415 The axenic culture of *Prorocentrum cordatum* strain CCMP1329 used in this work was
416 obtained from the Provasoli-Guillard National Center for Marine Algae and Microbiota
417 (NCMA). *P. cordatum* CCMP1329 was cultivated in L1-Si medium which is L1 medium
418 [63] with the modifications that synthetic ocean water [64] was used instead of natural
419 seawater, and $Na_2SiO_3 \cdot 9H_2O$ was omitted, since *P. cordatum* does not require silica.
420 Cultures were routinely maintained by transferring 10 ml of a 15 day old culture to 90 ml of
421 fresh L1-Si medium in 300 ml Erlenmeyer flasks. The flasks were kept in a climate chamber
422 (RUMED type 3501) at 20°C in a 12:12 h light:dark cycle with a light intensity of about 40
423 µmol photons $m^{-2}s^{-1}$ without any agitation. Lack of contaminating bacteria was routinely
424 checked by plating aliquots on LB and Difco marine agar 2216 (MB) plates. The number of
425 cells/ml after transfer fluctuated between 2.0×10^4 and 3.0×10^4.

426 **Cell counting**

427 Growth of *P. cordatum* was determined by cell counting using a BD LSR-Fortessa flow
428 cytometer (BD Biosciences, San Jose, CA, USA). *P. cordatum* was identified according to its
429 chlorophyll autofluorescence. Chlorophyll was excited with the 488 nm excitation laser line
430 and emission was detected at 695 nm. 1 ml samples were taken from three biological
431 replicates during the light period and fixed with 80 µl 25 % glutaraldehyde (final
432 concentration 2 %) for about 15 minutes at room temperature. Samples were snap frozen in
433 liquid nitrogen and stored at -70 °C until they were analyzed. Each sample was analyzed in
434 triplicate for 1 minute.

435 Four independent growth curves per temperature were obtained. For each growth curve,
436 counts were averaged from three biological replicates (with three technical replicates each).
437 Counts were plotted against time and a generalized additive model (GAM) was fitted. The

438 specific growth rate in the exponential growth phase, μ_{exp} (day^{-1}) and the doubling rate per
439 day k was calculated according to [65].

440 **Extraction of high molecular weight DNA for PacBio Sequencing**

441 DNA of *P. cordatum* cells extracted with commercially available plant kits or by applying the
442 common CTAB protocol was either too fragmented or contained too many contaminating
443 compounds to be suitable for PacBio sequencing. Therefore, we performed an initial
444 ultrasound treatment to break the cells and separate the nuclei from the debris, based on work
445 aimed at isolating nuclei for electron microscopy [66].

446 100 ml culture of *P. minimum* was transfered to two 50 ml Falcons tubes and centrifuged
447 (2,000 rpm, 5 min, RT). The supernatant was discarded, the pellets were dissolved in 15 ml
448 artificial sea water using an inoculation loop, and centrifuged again (2,000 rpm, 5 min, RT).
449 The supernatant was discarded, the pellet dissolved in 15 ml of 30% ethanol, and centrifuged
450 (2,000 rpm, 5 min, RT). This step was repeated, and the final pellets were dissolved in 3 ml
451 of 30% ethanol. Ultrasound using a Bandelin Probe Sonicator (Bandelin) was applied for 1
452 min, 5 cycles, amplitude 40% followed by centrifugation (1,000 rpm, 3 min, RT). The pellet
453 was suspended in 1.5 ml of 0.85% NaCl and centrifuged (10,000 g, 1 min, 4°C). Pellets from
454 two tubes (100 ml culture) were dissolved in 1 ml high salt buffer in an 2 ml Eppendorf tube
455 using inoculation loops. 8 µl proteinase K was added and incubated (56°C, 300 rpm, 1 h)
456 inverting the tubes every 15 min. After cooling on ice for 5 min 15 µl of RNAse A was added
457 and incubated (37°C for 30 min without shaking). The sample was centrifuged (10,000 g, 5
458 min, 4°C) and the supernatant transferred to a new tube. 200 µl 5 M NaCl were added and
459 mixed well by inverting the tube. Then, 200 µl solutions of CTAB/NaCl were added, mixed
460 well by inverting the tubes and incubated (65°C, with shaking 300 rpm, 10 min). 1 ml of
461 chloramphenicol-isoamyl alcohol was added, mixed well by inverting the tube, and after
462 phase separation the aqueous phase was collected. This treatment was repeated three times
463 until no interphase could be seen. To the collected aqueous phases an equal volume of -20°C
464 cold isopropanol was added and incubated overnight at -20°C. DNA was centrifuged (10,000
465 g, 10 min, 4°C) and the pellet was washed three times with 70% ethanol cooled to 4°C. The
466 pellet was suspended by flicking and then centrifuged at 10,000 g, 10 min, 4 °C. Following
467 air drying under the clean bench the DNA pellet was dissolved in 200 µl EB buffer and
468 stored at -70°C.

469 **RNA Isolation and sequencing**

470 For RNA extraction, the sample was thawed at room temperature and then transferred to a
471 cryotube filled with 0.3 g acid washed glass beads (100 µm). The cells were homogenized
472 using the FastPrep-24 instrument (MP Biomedicals, California, USA) at 6.0 m/s for 3 min (3
473 x 1 min + 1 min on ice). Samples were centrifuged (12,000 g, 10 min, 4°C), the supernatants
474 were transferred to fresh tubes and incubated for 5 min at room temperature. Then 100 µl of
475 1-bromo-3-chloropenol (BCP, Sigma, Germany) was added, samples were shaken vigorously
476 for 15 s and incubated for 10 min at room temperature. Samples were centrifuged (12,000 g,
477 10 min, 4°C), the aqueous phase was transferred to a new tube, 0.5 ml of isopropanol was
478 added, mixed and incubated for 10 min. The sample was then centrifuged (12,000 g, 10 min,
479 4°C). The supernatant was removed and the RNA pellet was washed with 1 ml of 75%
480 ethanol. The sample was vortexed briefly and the centrifuged (7,500 g, 5 min, 4°C). Washing
481 was repeated. The final pellet was air dried for 5 min, resuspended in 50 µl RNase-free water,
482 and incubated in a heat block at 55°C for 10 min. The RNA was stored at -80°C.

483 Removal of genomic DNA was verified via PCR using total RNA as template. The
484 concentration of the RNA was quantified using a NanoDrop spectrophotometer (Peqlab,
485 Erlangen, Germany) and the RNA integrity was assessed using a Bioanalyzer 2100 (Agilent,
486 Santa Clara, USA). The average RNA concentration in the 18 samples was 297.4 $\pm$ 109.5
487 ng/µl and the average RIN value was 5.5 $\pm$ 0.78.

488 RNA sequencing was performed at the HZI Braunschweig with 300 cycles on a NovaSeq
489 6000 using paired end 150 bp chemistry with the library kit NEBNext Ultra II directional
490 RNA. Ribosomal RNA was depleted prior to sequencing using polyA beads.

491 ***De novo* genome assembly**

492 Genome data from *P. cordatum* were generated using Illumina NovaSeq 6000 and PacBio
493 sequencing technologies, with a total data yield of 843.4 Gb (Table S23). Combining these
494 sequence reads, a hybrid genome assembly was generated using MaSuRCA v4.0.3 [67],
495 independently with CABOG (option *FLYE_ASSEMBLY=0*) and FLYE (option
496 *FLYE_ASSEMBLY=1*) as the assembly tool in the final step. Both assemblies are near
497 identical, in which 99.94% of the scaffolds in each assembly share 99.04% identity on
498 average. They exhibit the same level of data completeness (56.7%) based on recovery of

499 BUSCO single-copy orthologs in alveolata_odb10 dataset (which is known to be poor in
500 dinoflagellate data). Between these two preliminary assemblies, MaSuRCA-CABOG is more
501 contiguous (N50 length of scaffolds = 194.50 Kb) and yields better recovery of the rRNA
502 region (Table S24); this assembly was used in the subsequent refinement steps.

503 To refine the assembled genome, we first incorporated RNA-Seq data to further scaffold the
504 MaSuRCA-CABOG assembly using L_RNA_scaffolder [68]. Briefly, we mapped the *de*
505 *novo* assembled transcriptome (for transcripts ≥ 500 bp) onto the assembled genome
506 sequences using pblat v2.5 [69]. The mapping results in psl format were used as input for
507 L_RNA_scaffolder. This approach yielded a more-contiguous genome assembly (N50 length
508 of scaffolds = 346.97 Kb) with a better recovery of BUSCO genes (57.9%; Table S24).

509 Next, using BloBTool v1.1.1 [70], we assessed the assembled genome for potential outlier
510 sequences based on sequence coverage, G+C content, and shared sequence similarity to
511 known sequences in NCBI nt database (release 243; 15 April 2021). Genome scaffolds for
512 which the sequencing coverage or G+C content is external to the range of median ± 1.5 ×
513 interquartile range are considered as potential outliers. Scaffolds that have bacterial, archaeal,
514 or viral sequences as the top hits plus extreme sequencing coverage or extreme G+C content
515 are considered sequences that are putatively external to the nuclear genome of *Prorocentrum*
516 *cordatum*. In this analysis, majority (23,366; 98.2%) of the 24,295 genome scaffolds
517 (implicating 3.89 Gb) do not have hits in the BLAST searches (20,914 scaffolds; 2.95Mbp)
518 or have top hits in an undefined eukaryote sequence (2,452 scaffolds; 0.95Mbp); this
519 observation is expected given the lack of dinoflagellate data in the existing databases. Using
520 this approach, we identified 1,571 outlier sequences and removed them from the genome
521 assembly. Most of these outlier sequences do not have shared similarity to bacterial
522 sequences in the database. This is expected given the algal cultures from which the genomic
523 DNA was extracted were axenic. The final genome assembly has a total size of 4.15 Gb (N50
524 length of scaffolds = 349.2 Kb; Table S24).

525 **Transcriptome assembly**

526 RNA-Seq reads from six conditions (20_ex, 20_st, 26_ex, 26_st, 30_ex and 30_st) were
527 processed using fastp v0.21.0 [71] using parameter *–g* to remove adapter sequences and poly-
528 G artifacts known in NovaSeq 6000 data. Transcriptomes were first assembled in "de novo"
529 mode using Trinity v2.12.0 [72] independently for each condition. All *de novo* assembled

530 transcripts were combined as a single assembly, from which redundant sequences were
531 identified and removed using CD-HIT v4.8.1 [73] (98% identity; 0.9 length-difference
532 cutoff), yielding the final representative reference transcriptome.

533 To generate the genome-guided transcriptome, processed RNA-Seq reads from each
534 condition were first mapped onto the final genome assembly (above) using HISAT2 v2.2.1
535 [74]. The mapping result (i.e. describing the alignments between RNA-Seq reads and genome
536 scaffolds) was used as input for Trinity v2.12.0 in "genome-guided" mode. Using the same
537 strategy above, individual genome-guided assemblies were combined and redundant
538 sequences removed, yielding the final representative genome-guided transcriptome.

539 ***Ab initio* prediction of protein-coding genes**

540 To predict protein-coding genes, we follow the customised gene prediction workflow tailored
541 for dinoflagellate genomes following [75]. The description of this workflow is available at
542 https://github.com/TimothyStephens/Dinoflagellate_Annotation_Workflow. Briefly, this
543 approach integrates evidence of transcript and protein sequences to guide predictions using
544 multiple gene programs, after which the results were integrated to yield the final gene
545 models.

546 First, we identified novel repetitive elements in the genome assembly using RepeatModeler
547 v2.0.1 (http://www.repeatmasker.org/RepeatModeler/), combined these elements with
548 existing repeat families in RepeatMasker database release 2018/10/26 as a reference, to
549 predict and mask all repetitive sequence regions from the genome sequences using
550 RepeatMasker v4.1.0 (http://www.repeatmasker.org/); this yields the repeat-masked genome
551 assembly.

552 To predict protein-coding genes, we first mapped the representative *de novo* and genome-
553 guided transcriptome assemblies to the genome assembly using Minimap2 v2.18 [76], for
554 which the code was modified to recognise GC and GA donor splice sites. The mapping
555 information was then used to predict transcript-based genes using PASA v2.3.3 [77] for
556 which the code was modified to account for non-canonical splice sites. The proteins coded by
557 the transcript-based genes were searched (BLASTP, $E \leq 10^{-20}$, >80% query cover) against a
558 customised protein database combining RefSeq (release 98) proteins and predicted proteins
559 from available Symbiodiniaceae genomes (Table S25). The gene models were checked for

560 transposable elements using HHblits v3.3.0 [78] and TransposonPSI
561 (https://github.com/NBISweden/TransposonPSI), searching against the JAMg transposon
562 database (https://github.com/genomecuration/JAMg); those genes containing transposable
563 elements were removed from subsequent steps. Redundant sequences were removed based on
564 similarity using CD-HIT v4.8.1 [73] (*-c 0.75 -n 5*). Among the remaining transcript-based
565 gene sequences, we identified high-quality "golden genes" using the script
566 *Prepare_golden_genes_for_predictors.pl* from the JAMg pipeline
567 (https://github.com/genomecuration/JAMg), altered to recognise alternative splice sites.
568 These "golden genes" represent high-quality training set for *ab initio* gene predictors. We
569 used them as the training set for SNAP [79] and AUGUSTUS v3.3.1 [80] for gene prediction
570 from the repeat-masked genome assembly; the codes for AUGUSTUS was also modified to
571 recognise alternative splice sites (https://github.com/chancx/dinoflag-alt-splice).

572 The repeat-masked genome was also used as the input for GeneMark-ES [81]. We also
573 predicted genes using MAKER v2.31.10 [82], in which the code was modified to recognise
574 GA donor splice sites. Protein-coding genes were predicted using MAKER (*protein2genome*
575 mode) based on protein sequences from the Swiss-Prot database (retrieved 02 March 2020)
576 and other predicted *Symbiodiniaceae*. Finally, gene predictions from the five methods
577 including the *ab initio* predictors (GeneMark-ES, AUGUSTUS, SNAP), MAKER protein-
578 based predictions, and PASA transcript-based predictions were integrated using
579 EvidenceModeler v1.1.1 [83] to yield the high-confident gene models; The weighting is
580 PASA 10, MAKER protein 8, AUGUSTUS 6, SNAP 2 and GeneMark-ES 2; see [75] for
581 details. These resulted gene models were subjected three rounds of polishing during which
582 the gene models were corrected based on transcriptome re-mapping using the annotation
583 update utility in PASA [77].

584 **Functional annotation of protein-coding genes**

585 Annotation of gene function for *P. cordatum* and other representative dinoflagellate genomes
586 was conducted based on sequence similarity to known proteins in the UniProt database
587 (release 2021_03). Predicted protein sequences from the gene models were first searched
588 against the manually curated protein sequences of Swiss-Prot (UniProt release 2021_03)
589 using BLASTp v2.3.0+ ($E \leq 10^{-5}$; subject-sequence cover ≥ 70%). Sequences that have no

590 Swiss-Prot hits were then searched against TrEMBL (UniProt release 2021_03) using the
591 same parameters.

592 For predicted proteins of *P. cordatum*, we further assessed functions based on sequence-
593 similarity search against known protein sequences in EnzymeDetector, InterProScan,
594 eggNOG and Kofam.

595 **Prediction of transit peptides**

596 For each predicted protein, transit peptides were first predicted independently using TargetP
597 v2.0 (*-org pl*), Signalp v6 *(--organism eukarya --mode fast*), WoLF PSORT (*plant* mode),
598 Predotar, and ChloroP v1.1. Subcellular localization is determined based on the consensus
599 from these predictions confirmed in three or more programs.

600 **Analysis of homologous proteins**

601 To identify homologous proteins of the predicted *P. cordatum* proteins, we compiled a
602 comprehensive protein-sequence database (1,554,705 sequences from 31 dinoflagellate taxa;
603 Table S26) using available genome or transcriptome data. All data from the MMETSP
604 database [84] were downloaded at https://zenodo.org/record/1212585#.YlYczuhByUk. For
605 species where there were multiple datasets for the same isolate, the protein sequences were
606 clustered at 90% sequence identity using CD-HIT v4.8.1 [73] to reduce redundancy. Using
607 these 31 sets of protein sequences, homologous sets were then identified based on clustering
608 of protein sequences based on sequence identity using OrthoFinder v2.3.8 [85] at default
609 settings.

610 **Identification of mRNA editing sites**

611 Putative mRNA editing events were identified from single nucleotide variations observed in
612 genome-sequence reads *versus* RNA-Seq reads. An observed nucleotide variation in the
613 RNA-Seq reads but not in genome-sequence reads is considered a potential mRNA edited
614 site. Briefly, 25% of all genome-sequence reads (randomly sampled) were mapped to the
615 final genome assembly using bwa-mem v0.7.17-r1188 (https://github.com/lh3/bwa).
616 Trimmed RNA reads from each sample (6 conditions; 3 replicates) were mapped to the
617 genome assembly separately with HISAT2 v2.2.1 [74] using default parameters (--no-
618 discordant) and a HGFM index that was built using known exons and splice sites from the

619 predicted gene models. PCR duplicates were marked by *MarkDuplicates* implemented in
620 Picard v2.23.8 (https://broadinstitute.github.io/picard/). For each condition, mapping of
621 RNA-Seq reads was compared with the mapped genome-sequence reads using JACUSA v2
622 (call-2 -F 1024 -P2 RF-FIRSTSTRAND -s -a D,Y,H:condition=1). We follow the authors'
623 recommendation to assess the statistical significance of an mRNA edited site. A site is
624 considered statistically significant if it meets all the requirements: (1) a score > 1.15; (2)
625 coverage of genome reads >10; (3) coverage of RNA reads from each condition > 5; (4)
626 number of putative editing type is < 2; (5) The editing site is present in all three replicates.

627 **Analysis of horizontal gene transfer**

628 To identify putative horizontal gene transfer (HGT), *P. cordatum* proteins were searched
629 (BLASTP, $E \leq 10^{-5}$) against a customised protein-sequence database that consist of 2,773,521
630 proteins from 82 other eukaryotes (Table S27) and 688,212 proteins from 543 single-cell
631 assembled genomes (SAGs) of prokaryotes (Table S28) [86]. Excluding hits to other
632 *Prorocentrum* proteins, *P. cordatum* proteins have a bacterial top hit are considered results of
633 horizontal gene transfer involving *P. cordatum* and bacteria. To support the inferences of
634 putative HGT, we employed Orthofinder v2.5.4 [85] to infer homologous protein sets from
635 all the involved proteins (i.e. *P. cordatum* proteins, proteins from other eukaryotes and
636 SAGs). Homologous protein sets containing *P. cordatum* proteins that involving HGT were
637 used to infer multiple sequences alignments using MAFFT v7.453 [87] at --maxiterate 1000.
638 Trimmed by trimAl v1.4.1 [88] using parameter *-automated1*, the alignments were used to
639 infer phylogenetic trees using IQ-TREE2 [89] at *-B 2000 -T AUTO*.

640 **Identification of introner elements**

641 Introner elements were identified in the intronic regions using the program Pattern locator
642 [90]. The pattern we searched for were inverted repeats of 8-20 nucleotides and direct repeats
643 of 3-5 nucleotides within 30 bases of each end of the introns as described in [91].

644 **Integration/mixOmics analysis**

645 We conducted a multi-omics analysis in *P. cordatum* using the proteomic and transcriptomic
646 data to identify a systems level heat stress response across the growth phases using the
647 *mixOmics* package in R [92]. Transcriptome FPKM values were first log2 transformed prior
648 to quality filtering with normalized proteomic data, requiring features be present across 75%

649 of samples. Selected features represented the top 25% and 50% most variable transcripts
650 (15,097) and proteins (259), respectively, that were present across 75% of the samples. These
651 features were then input to mixOmics for Data Integration Analysis for Biomarker discovery
652 using a Latent cOmponents (DIABLO) [93].

653 We conducted performance testing of the initial model to identify the number of latent
654 components that contained a multi-omics signature using Mfold validation with 5 folds and
655 50 repeats. This suggested two latent components as the best fits for the model. We then
656 performed final model tuning using the max distance to select diagnostic features for both
657 component 1 (RNAseq: 170, proteins: 30) and component 2 (RNAseq: 190, proteins: 40).
658 Ordination of all features indicate the separation of the three temperature levels for both
659 transcriptome and proteome features. The transcriptome and proteome features selected for
660 component 1 discriminate a common heat stress response at both 26°C and 30°C (Extended
661 Data Figure 1, Table S12) and for component 2 a heat stress response specific to either 26°C
662 or 30°C (Extended Data Figure 2, Table S12). The variates for the pathways from both
663 components were then input to NetworkAnalyst [94] for KEGG pathway over-representation
664 analysis to identify functional categories that were enriched in each network.

665 A relevance associations network was created for each component using the *network* function
666 within *mixOmics*, where values represent a robust approximation of the Pearson correlation.
667 A heatmap displaying the features from each component was created using the *pheatmap*
668 package in R with features clustered according to their Euclidean distances and scaled within
669 rows. This revealed two main clusters within each component that were extracted using the R
670 package *dendextend* according to the corresponding dendogram. A relevance associations
671 network was then created for each subcluster as previously done for the two components.

672 **Experimental design and sampling**

673 Cultures of *P. cordatum* CCMP1329 were maintained as described above. After 15 days of
674 cultivation at 20°C, 10 ml were transferred to 90 ml of fresh L1-Si medium in 300 ml
675 Erlenmeyer flasks and placed in a climate chamber set to the desired temperature (26°C or
676 30°C) and exhibiting the same light intensity and light:dark cycle. Replicate cultures were set
677 up under identical conditions to allow sampling of 3 biological replicates each for
678 transcriptome and proteome and 5 biological replicates for metabolome. For proteome
679 measurement, up to 12 liters of culture (120 flasks with 100 ml culture volume each) were

680 cultivated in the same climate chamber to obtain ~2 g biomass (wet weight). For each growth
681 stage (exponential, stationary) and temperature (20°C, 26°C, 30°C) a complete set of cultures
682 were sacrificed. Cell counting and harvesting were started about 5 h after the onset of the
683 light period. For cell counting, random samples were chosen from the climate chamber to
684 account for small differences in light intensity.

685 Harvesting samples for transcriptome analysis. Three 100 ml cultures (biological replicates)
686 were sampled per temperature and growth phase. 100 ml culture was centrifuged in two 50
687 ml Falcon tubes at 5000 rpm (4276 g), 4 °C for 5 minutes in a Heraeus Multifuge X1R
688 (Thermo Scientific, Schwerte, Germany). The supernatant was decanted, and the pellet was
689 resuspended in the remaining medium. The two pellets were combined in a 2 ml Eppendorf
690 tube and centrifuged at 13,000 rpm, 4 °C for 3 minutes in a Heraeus PrimoR centrifuge
691 (Thermo Scientific, Schwerte, Germany). The remaining supernatant was removed by
692 pipetting, and the weight of the wet pellet was determined. The pellet was resuspended in 1
693 ml Trizol Reagent, frozen in liquid nitrogen and stored at -70 °C until further analysis.

694 Harvesting samples for proteome analysis: Tubes and buffers used in the following steps
695 were pre-chilled; all steps were carried out on ice. 400 ml culture was filled into chilled 500
696 ml centrifuge bottles, centrifuged at 5,000 rpm (4,248 g), 4°C, for 30 minutes in a Sorval
697 Lynx 4000 centrifuge (Thermo Fischer, Schwerte, Germany). The supernatant was decanted,
698 the pellet was resuspended in 100 ml buffer containing 100 mM Tris-HCl pH 7.5 + 5 mM
699 $MgCl_2$ x 6 H_2O. The re-suspended pellets were centrifuged again (30 min, 5,000 rpm, 4°C.
700 The supernatant was removed, and the pellet was re-suspended by gently pipetting in 800 µl
701 of the same buffer. The suspension was transferred into 2 ml Eppendorf tubes and centrifuged
702 at 13,000 rpm, 4°C for 5 minutes in a Heraeus PrimoR centrifuge (Thermo Scientific,
703 Schwerte, Germany). The supernatant was removed by pipetting and the weight of the wet
704 pellet was determined. Samples were frozen in liquid nitrogen and stored at -70 °C.

705 Harvesting samples for metabolome analysis: 15 ml from five 100 ml cultures (biological
706 replicates) from each temperature and growth phase were extracted immediately after cell
707 counting with a filtration unit and 0,22 µm filter (Millipore, GSWP04700). Samples were
708 filtered at 500 mbar with a vacuum controller. The cells were washed three times with 4°C
709 cold 3.5 % NaCl.

710 The filters were immediately transferred to 5 mL Eppendorf tubes containing 100 mg 0.7-100
711 µm glass beads (Kuhmichel, Germany), two 5 mm^3 and one 10 mm^3 stainless steel beads
712 (Kugel Winnie, Germany) for partially destroying the filter and to obtain cell lysis. 2 mL of -
713 20°C cold extraction fluid (EF) were immediately added. The EF for metabolite extraction
714 used in the experiments contained methanol, ethanol, chloroform and the internal standard
715 13C-ribitol (IS) [95]. The pre-cooled 5 mL tubes with filter and EF on ice were mixed for 20
716 seconds and placed back on ice until further treatment.

717 **Transcriptome analysis**

718 Reads mapping and gene counting

719 The transcriptome reads were mapped to the assembled reference genome using HISAT2
720 [74]. The reads mapped onto the exons were counted for the corresponding genes with
721 featureCounts [96]. For differential expression analysis, only uniquely mapped reads were
722 taken into account to avoid ambiguity. The fragments per kilo base of transcript per million
723 mapped fragments (FPKM) value of each gene was calculated by normalizing the fragments
724 per million with the sum of the exon length of the corresponding gene. The ternary
725 visualization of the gene expression pattern across three temperatures was produced with R
726 package ggtern [97].

727 Differential expression analysis

728 Genes with low expression were filtered out by the "filterbyexpr" function in edgeR [98]
729 with default parameters. Then, the differential expression (DE) analysis was performed with
730 edgeR using the recommended "glmQLF" test on the raw read count per gene. Genes with a
731 "Benjamini-Hochberg" corrected p-value (FDR) <=0.001 and an absolute log2 fold change
732 (log2FC) >=2 were considered as significantly differentially expressed. The Venn diagram
733 showing the DE gene overlaps between different comparisons was created with R package
734 VennDiagram [99].

735 Hierarchical clustering with the complete-linkage algorithm was used to identify gene
736 clusters based on their expression profile across temperature conditions. The input expression
737 values were centered log2 transformed FPKM values where the FPKM expression values
738 were log2(x+1) transformed, then centered by subtracting the mean. The tree was cut into
739 eight clusters to represent different expression profiles. Super cluster 4 and cluster 7 in the

740 stationary phase showed very similar expression pattern and each of them was small.
741 Therefore, we decided to merge them for the downstream functional enrichment analysis.

742 Functional enrichment and KEGG pathway analysis

743 The GO term enrichment analysis was carried out with the R package topGO (DOI:
744 10.18129/B9.bioc.topGO) for all three gene ontologies, i.e., biological process (BP), cellular
745 component (CC) and molecular function (MF). REVIGO [100] was used to summarize the
746 GO terms according to the semantic similarity for a concise visualization in Figure S9. In
747 order to perform a KEGG pathway enrichment analysis, we assigned KEGG ortholog (KO)
748 number for each gene using KOfam [101]. When a gene had multiple KO assignments with
749 an e-value <=1e-10, we chose the one with the lowest e-value. If a gene had several KO
750 assignments with an e-value of zero, we kept all those KO assignments for this gene. Using
751 this annotation, the genes with KO assigned held about 50% of mapped reads. The expression
752 profile for each KO gene was then generated by summing up the read count for all genes
753 belonging to the same KO gene. DE KO genes were then identified using edgeR similar to
754 the above DE analysis. The KEGG pathway and module enrichment analyses were performed
755 using clusterProfiler [102] on those DE gene sets. For both GO term enrichment analysis and
756 KEGG enrichment analysis, an FDR <=0.05 was considered as significantly enriched.

757 Statistical tests

758 The association between introner element containing genes, DE genes in stationary phase,
759 and the gene duplication categories was tested with Fisher exact test to identify significant
760 relations (enrichment, and depletion). An odds ratio (OR) above 1 and p-value smaller than
761 1e-5 indicates an enrichment relation, while OR below 1 and p-value below 1e-5 represents a
762 depletion relation.

763 In the central carbon metabolism pathway analysis, a paired Wilcoxon test was performed to
764 compare the expression change of the gene members belonging to the same EC number. The
765 alteration of the metabolite's concentration was analyzed using Wilcoxon test on the mean
766 concentration values of biological replicates over the three technical replicates. An FDR
767 <=0.05 was regarded as significantly changed.

768 **Metabolome analysis**

769 Metabolite extraction

770 The cell disruption was performed with an oscillating mill (Retsch, MM400,Germany) at 30
771 Hz for 2 minutes, which was repeated three times. The cooling during the treatment was
772 ensured by using -80°C pre-cooled sample containers. For metabolite extraction, the tubes
773 were centrifuged for 8 min at 12,000 rpm and 4 °C and afterwards two different volumes of
774 the supernatant (100 µl and 200 µl) were transferred to GC glass vials (Klaus Trott
775 Chromatographiezubehör, Germany) and dried under vacuum at 4 °C (Labconco, United
776 States). After drying over night the vials were capped with magnetic vial caps (Klaus Ziemer
777 GmbH, Germany) and stored at -80 °C until measurements. All samples belonging to one
778 experiment were analyzed together in one GC-MS batch.

779 Derivatization and GC-MS measurements

780 To analyze polar metabolites with gas chromatography and mass spectrometry (GC-MS) a
781 derivatization was performed to reduce polar interactions by introducing unpolar
782 trimethylsilyl (TMS) groups. This was necessary to evaporate the substances under low
783 temperatures and avoid alteration. After extraction the samples were analyzed with a gas
784 chromatograph connected to a mass spectrometer (GC-MS) (Agilent 7890A and 5975C inert
785 XL Mass Selective Detector). The sample derivatization was automatically done with a
786 multisampler (GERSTEL, Germany) and 15 µl 2 % methoxyamine hydrochlorid (MeOX) in
787 pyridine (Roth, Germany) for 60 min at 40°C followed by the addition of 15 µl N-tert-
788 Butyldimethylsilyl-N-methyl (MSTFA) and incubated for 30 min at 40°C. For measurement
789 1 µl of the sample was injected in splitless-mode into an SSL injector at 270°C. The GC was
790 equipped with a 30 m DB-35MS + 5 m Duraguard capillary column (0.25 mm inner
791 diameter, 0.25 µm film thickness). Helium was used as carrier gas at a flow rate of 1.0
792 mL/min and the GC-oven was run at following temperatures and times per sample: 6 min
793 80°C; 6 minn80 to 300°C; 10 min 300°C; 2.5 min 300 to 325°C; 4 min 325°C. Each GC-MS
794 run lasted 60 minutes. The transmission temperature from GC to MS was 280°C and the
795 temperatures of the MS and the quadrupol were 230°C and 150°C, respectively. The
796 ionisation in the mass detector was performed at 70 eV. The detector was operated in SCAN-
797 mode for full scan spectra from 70 to 800 m/z with 2 scans/s. For calibration of retention
798 indices an alcan mix ((C10-C40) (Sigma Aldrich) was used.

799 Data analysis for metabolome

800 Data analysis was performed with the software Metabolite Detector [103], RStudio and
801 LibreOffice. After calibration with the alcan mix a normalization by IS 13C-Ribitol was
802 performed to eliminate variations in sample volumes. Batch quantifications were performed
803 with Metabolite Detector and the generated files were visualized in barplots with the same
804 software or with R as heatmaps. The shown statistics were also performed with Metabolite
805 Detector.

806 Non-targeted analysis was performed with an in-house library and the following settings:
807 Peak threshold 5; minimum peak height 5; 10 bins per scan; deconvolution with 5; no
808 baseline adjustment; compound reproducibility 1.00; maximum peak discr. index 100.00; min
809 # ions 20.

810 **Proteome analysis**

811 Shotgun proteomics

812 Cells of *P. cordatum* were resuspended in solubilization buffer and disrupted by bead beating
813 (FastPrep-24 5G, MP Biomedical, Eschwege, Germany) for 10 s at 6 m/s followed by 90 s on
814 ice (three repetitions) using 0.1 mm silica spheres. Cell debris and insoluble material was
815 removed by ultracentrifugation for 1 h at 104,000 g at 17 °C and the supernatant again
816 centrifuged at 200,000 g for 1 h at 17 °C. The protein content of the preparation was
817 determined according to [104]. A total of 50 µg was subjected to reduction, alkylation and
818 tryptic digest as reported previously [105]. Generated peptides were analyzed by (i) 1D
819 reversed phase nanoLC and (ii) 2D SCX-/RP nanoLC separation coupled to MS-detection.
820 For 1D separation, one microgram of tryptic peptides were decomplexed by reversed phase
821 nanoLC separation (Ultimate 3000 nanoRSLC; Thermofisher Scientific, Dreieich, Germany)
822 using a trap column setup (2 cm length, 5 µm bead size, 75 µm inner diameter; Thermofisher
823 Scientific) with a 25 cm separation column (2 µm bead size, 75 µm inner diameter;
824 Thermofisher Scientific) and applying a 360 min linear gradient [105]. In case of 2D
825 separation, 4 µg were subjected to SCX fractionation (eluent A 5 mM H_3PO_4, 5 % (v/v)
826 acetonitrile, pH 3.0; eluent B 5 mM H_3PO_4, 5 % (v/v) acetonitrile, 1 M NaCl, pH 3.0) with a
827 linear 20 min gradient and collection of 19 fractions per sample beginning 5 min after
828 injection. Each fraction was subsequently applied for second dimension reversed phase
829 separation (see above) using a 60 min gradient. For both methods, the eluent was
830 continuously ionized (captive spray ion source; Bruker Daltonik GmbH, Bremen, Germany)

831 and ions analyzed by an ion-trap mass spectrometer (amaZon speed ETD; Bruker Daltonik
832 GmbH) as described [106].

833 Preparation of the membrane protein-enriched fraction

834 Cell pellets were gently thawed, cells resuspended in 0.5 ml membrane lysis buffer (MLB;
835 100 mM TrisHCl, 2 mM $MgCl_2$, 10 % (w/v) glycerin, 0.5 mM DTT, pH 7.5) and disrupted
836 by bead beating (see above). DNA was digested using DNase I and the obtained raw fraction
837 applied on top of a continuous sucrose gradient (30 – 80 % (w/v)) prior to centrifugation for
838 12 h at 37,800 g and 4 °C. Membrane containing fractions were collected and washed twice
839 with MLB (centrifugation 1 h at 104,000 g). Obtained membrane pellets were resuspended in
840 500 µl MLB, pooled and centrifugation repeated. The final pellet was resuspended in 300 µl
841 1.0 % (w/v) SDS, incubated at 95 °C for 5 min prior to centrifugation for 20 min at 20,000 g
842 and 20 °C. The supernatant was frozen in liquid nitrogen until usage. Protein content was
843 determined using the RCDC assay (BioRad GmbH, Munich, Germany). A total of 10 µg
844 protein per sample and fraction were separated by SDS-PAGE and gels stained with
845 Coomassie brilliant blue [107]. Each sample lane was cut into 8 slices and each slice into
846 small pieces of ~1-2 mm^2 for subsequent in-gel digest as described [106]. Generated peptide
847 solutions were analyzed by reversed phase nanoLC-MS (see above) applying a 120 min
848 gradient. Per sample, protein search results of each slice of fractions I and II were compiled.

849 Proteome data analysis

850 Protein identification was performed using Mascot (version 2.3; Matrix Science, London,
851 UK) via the ProteinScape platform (version 4.2; Bruker Daltonik GmbH) against the genomic
852 database of *P. cordatum*. A target-decoy strategy with a false discovery rate <1.0% was
853 applied as well as the following settings: enzyme trypsin; 1 missed cleavage allowed;
854 carbamidomethylation (C) as fixed, oxidation (M) as variable modification; peptide and
855 MS/MS mass tolerance 0.3 Da; monoisotopic; peptide charge 2+ and 3+; instrument type
856 ESI-TRAP; significance threshold $p < 0.05$; ion score cutoff 25.0; minimum peptide length of
857 5. Search results of individual searches per sample were compiled using the ProteinExtractor
858 function of ProteinScape.

859

860 ## Figure Legends

861 **Figure 1. Genome features of *P. cordatum*.** (a) Distribution of identified repeat types in the
862 *P. cordatum* genome. (b) Maximum likelihood tree inferred using 3,507 strictly orthologous,
863 single-copy protein sets among 31 dinoflagellate taxa, with ultrafast bootstrap support (based
864 on 2000 replicate samples) shown at each internal node; unit of branch length is number of
865 substitutions per site. The ecological niche for each taxon is shown on the right of the tree.
866 The five representative taxa and *P. cordatum* from this study are highlighted on the tree in
867 red text. Distribution of G+C content for (c) whole-genome sequences and (d) protein-coding
868 sequences relative to the other five representative genomes. (e) Genome and gene features of
869 *P. cordatum* relative to the other five taxa, showing haploid genome size estimated based on
870 sequence data, number of protein-coding genes, lengths of all introns, and separately for
871 introns that contain introner elements (IE^+), and those that do not contain introner elements
872 (IE^-), known repeat types, and types of duplicated genes.

873 **Figure 2. Gene functions encoded in *P. cordatum* genome.** (a) Gene functions encoded in
874 the genome of *P. cordatum* and the other five representative taxa based on relative abundance
875 of Gene Ontology (GO) terms per genome, shown for categories of Biological Process (BP),
876 Molecular Function (MF) and Cellular Component (CC). The ecological niche for each taxon
877 is shown at the bottom of the heatmaps. (b) GO terms that are significantly enriched ($p <$
878 0.01) among genes for each duplication type (i.e. dispersed, proximal and tandem) in *P.*
879 *cordatum*, relative to p values observed for the other taxa, and the associated number of genes
880 for each GO term.

881 **Fig. 3. Integrated analysis of transcriptome and proteome response of *P. cordatum* upon**
882 **heat stress. a**, Experimental design of the temperature-dependent growth experiments. **b**,
883 Growth of *P. cordatum* at 20°C, 26°C and 30°C. Harvesting for multi-omics analyses is
884 indicated by dashed vertical lines (Ex, exponential, St stationary growth phase). **c**, Ternary
885 plots of highly expressed gene models with an average log2CPM (log2 of count per million)
886 >5 in response to temperature and growth phase (8,593 transcripts in each plot). **d**, Clustering
887 of differentially abundant proteins (2,098 identified proteins) in response to temperature and
888 growth phase. Abundances of proteins were calculated from standardized peptide counts. **e**,
889 Heatmap of transcripts and proteins showing significant correlations for generalized heat

890 stress response (component 1, left panel) and temperature-specific response (component 2,
891 right panel). **f**, Over-represented KEGG-pathways in the networks of generalized and
892 temperature specific heat-stress response. **g**, DIABLO network of generalized heat stress
893 response (component 1) revealing positive and negative correlations (≥0.7) between
894 transcripts and proteins.

895 **Figure 4. Heat stress response of central modules of energy and carbon metabolism in**
896 ***P. cordatum*.** a, Temperature dependent dynamics of sub-transcriptomes (left) and sub-
897 proteomes (right) associated with photosynthesis (PS), central metabolism (CM) and
898 oxidative phosphorylation (OP). Colored circles represent individual transcripts and proteins,
899 respectively, with their areas proportional to the determined abundances. For proteome, the
900 heights of the CM- and OP-bands were 10-fold enlarged for better comparison with the PS-
901 band. b, Light reaction of photosynthesis. Abbreviations: CCP, carotenoid/chlorophyll-
902 binding protein; CP, chlorophyll-binding protein; FCP, fucoxanthin/chlorophyll-binding
903 protein; FNR, ferredoxin:NADP oxidoreductase; OEC, oxygen evolving complex; PS,
904 photosystem. c, Central metabolism including CO2-concentrating, Calvin cycle and central
905 C-metabolism and selected biosyntheses of amino acids. Enzyme abbreviations: PEPC,
906 phosphoenolpyruvate carboxylase; GBSS, granule-bound starch synthase (NDP-glucose-
907 starch glycosyltransferase); GLS, glutamine synthetase; PGK, phosphoglycerate kinase;
908 PRK, phosphoribulokinase; PSPH, phosphoserine phosphatase; SS, starch synthase. A
909 detailed scheme is presented in Extended data Figure 4. d, Oxidative phosphorylation.
910 Underlying data are compiled in Table S13.

911 **Figure 5: The transcriptome landscape under heat stress. a** The regulation pattern of
912 dispersed genes in ortho-groups. Up: up-regulated at 26°C St phase compared to 20°C St
913 phase, Down: down-regulated at 26°C St phase compared to 20°C St phase; NotSig: not
914 significantly changed in expression. In the right panel of **a** 2Up.1Down.0NotSig indicates the
915 ortho-groups that had 2 genes up-regulated, 1 down-regulated and 0 not significantly
916 changed. **b** Heatmap and clusters of gene expression across samples and temperatures in the
917 stationary growth phase. Hierarchical clustering with the complete-linkage algorithm was
918 used to identify gene clusters based on their expression profile across temperature conditions.
919 The input expression values were centered log2(x+1) transformed FPKM values. **c** Post-
920 transcriptional regulation in *P. cordatum*, shown for 4,550 genes in distinct duplication

921 modes (couterclockwise): dispersed, tandem, proximal, and singleton. Features shown from
922 the inner-most to the outer-most circle: differential exon usage, differential editing of mRNA
923 per gene in response to growth phase (green/brown), and differential mRNA editing per gene
924 in response to temperature (red/purple). In each circle, a bar in a light shade indicates the
925 number of corresponding features identified in a gene, a bar in a dark shade indicates the
926 number of statistically significant features in a gene. The bottom-left legend shows the
927 distinct scenarios for identifying sites for mRNA editing in all transcripts versus differentially
928 edited sites.

929 **Fig. 6. Structure, expression in response to heat stress and protein architecture of**
930 **Hsp70 (a-c) and RuBisCo (c-e) gene models. a,** Clustering of HSP70 gene models
931 according to transcription pattern across growth conditions; cumulative abundance of
932 transcripts and proteins is indicated below. **b**, Relative exon usage, CDS structure and mRNA
933 editing of gene model s12246_g74608 harbouring multiple HSP70 CUs. E, exon; kinked line,
934 intron; E01/03/05/07, complete HSP70s; E09/10, N-/C-terminal fragments of HSP70;
935 E02/04/06/08, spacers. **c**, Clustering of RuBisCo gene models according to transcription
936 pattern across growth conditions; cumulative abundance of transcripts and proteins is
937 indicated below. **d**, Scheme of proteins indicating complete CUs, N- and C-terminal
938 fragments, spacer and putative leader sequences. Detected peptides are indicated below. **e**,
939 ClustalW alignment of upstream regions of sixteen complete proteins of *P. cordatum* with or
940 without leader sequence as well as representative spacer sequences. Alignment of all 32
941 sequences and other dinoflagellate proteins is provided in Figure S11). Basic residues are
942 marked in dark blue and conserved regions are highlighted by purple background.

References

1. Anderson DM, AD Cembella, GM Hallegraeff. Progress in understanding harmful algal blooms: Paradigm shifts and new technologies for research, monitoring, and management. *Ann Rev Mar Sci* . 2012. , **4**

2. Wells ML, B Karlson, A Wulff, R Kudela, C Trick, V Asnaghi, E Berdalet, W Cochlan, K Davidson, M De Rijcke, S Dutkiewicz, G Hallegraeff, KJ Flynn, C Legrand, H Paerl, J Silke, S Suikkanen, P Thompson, VL Trainer. Future HAB science: Directions and challenges in a changing climate. *Harmful Algae* 2020; **91**.

3. Brown AR, M Lilley, J Shutler, C Lowe, Y Artioli, R Torres, E Berdalet, CR Tyler. Assessing risks and mitigating impacts of harmful algal blooms on mariculture and marine fisheries. *Rev Aquac* 2020; **12**: 1663–1688.

4. Lewis AM, LN Coates, AD Turner, L Percy, J Lewis. A review of the global distribution of Alexandrium minutum (Dinophyceae) and comments on ecology and associated paralytic shellfish toxin profiles, with a focus on Northern Europe. *J Phycol* 2018; **54**: 581–598.

5. Zhang F, M Li, PM Glibert, SH (Sophia) Ahn. A three-dimensional mechanistic model of Prorocentrum minimum blooms in eutrophic Chesapeake Bay. *Sci Total Environ* 2021; **769**.

6. Velikova V, J Larsen. The Prorocentrum cordatum / Prorocentrum minimum taxonomic problem. 2017. , **3134**

7. Hallegraeff GM. A review of harmful algal blooms and their apparent global increase. *Phycologia* 1993; **32**.

8. Gobler CJ. Climate Change and Harmful Algal Blooms: Insights and perspective. *Harmful Algae* 2020; **91**: 101731.

9. Hallegraeff GM, L Schweibold, E Jaffrezic, L Rhodes, L MacKenzie, B Hay, H Farrell. Overview of Australian and New Zealand harmful algal species occurrences and their societal impacts in the period 1985 to 2018, including a compilation of historic records. *Harmful Algae* 2021; **102**.

10. LaJeunesse TC, G Lambert, RA Andersen, MA Coffroth, DW Galbraith. Symbiodinium (Pyrrhophyta) genome sizes (DNA content) are smallest among dinoflagellates. *J Phycol* 2005; **41**: 880–886.

11. Wisecaver JH, JD Hackett. Dinoflagellate genome evolution. *Annu Rev Microbiol* 2011; **65**.

12. Lin S. Genomic understanding of dinoflagellates. *Res Microbiol* 2011; **162**: 551–569.

13. LaJeunesse TC, JE Parkinson, PW Gabrielson, HJ Jeong, JD Reimer, CR Voolstra, SR Santos. Systematic Revision of Symbiodiniaceae Highlights the Antiquity and Diversity of Coral Endosymbionts. *Curr Biol* 2018; **28**.

14. Stoecker DK, PJ Hansen, DA Caron, A Mitra. Mixotrophy in the Marine Plankton. *Ann Rev Mar Sci* 2017; **9**.

15. Johnson MD. Inducible Mixotrophy in the Dinoflagellate Prorocentrum minimum. *J*

983 *Eukaryot Microbiol* 2015; **62**.

984 16. Stephens TG, A Gabr, V Calatrava, AR Grossman, D Bhattacharya. Why is primary
985 endosymbiosis so rare? *New Phytol* 2021; **231**: 1693–1699.

986 17. Zaheri B, D Morse. Assessing nucleic acid binding activity of four dinoflagellate cold
987 shock domain proteins from Symbiodinium kawagutii and Lingulodinium polyedra.
988 *BMC Mol Cell Biol* 2021; **22**.

989 18. Bi Y, F Wang, W Zhang. Omics analysis for dinoflagellates biology research.
990 *Microorganisms* 2019; **7**: 1–25.

991 19. Savary R, DJ Barshis, CR Voolstra, A Cárdenas, NR Evensen, G Banc-Prandi, M
992 Fine, A Meibom. Fast and pervasive transcriptomic resilience and acclimation of
993 extremely heat-tolerant coral holobionts from the northern Red Sea. *Proc Natl Acad*
994 *Sci U S A* 2021; **118**.

995 20. Liew YJ, Y Li, S Baumgarten, CR Voolstra, M Aranda. Condition-specific RNA
996 editing in the coral symbiont Symbiodinium microadriaticum. *PLoS Genet* 2017; **13**.

997 21. Matthews JL, JB Raina, T Kahlke, JR Seymour, MJH van Oppen, DJ Suggett.
998 Symbiodiniaceae-bacteria interactions: rethinking metabolite exchange in reef-
999 building corals as multi-partner metabolic networks. *Environ Microbiol* 2020; **22**:
1000 1675–1687.

1001 22. Nand A, Y Zhan, OR Salazar, M Aranda, CR Voolstra, J Dekker. Genetic and spatial
1002 organization of the unusual chromosomes of the dinoflagellate Symbiodinium
1003 microadriaticum. *Nat Genet* 2021; **53**: 618–629.

1004 23. Marinov GK, AE Trevino, T Xiang, A Kundaje, AR Grossman, WJ Greenleaf.
1005 Transcription-dependent domain-scale three-dimensional genome organization in the
1006 dinoflagellate Breviolum minutum. *Nat Genet* 2021; **53**.

1007 24. Camp EF, T Kahlke, B Signal, CA Oakley, A Lutz, SK Davy, DJ Suggett, WP Leggat.
1008 Proteome metabolome and transcriptome data for three Symbiodiniaceae under
1009 ambient and heat stress conditions. *Sci Data* 2022; **9**: 1–10.

1010 25. Heil CA, PM Glibert, C Fan. Prorocentrum minimum (Pavillard) Schiller - A review
1011 of a harmful algal bloom species of growing worldwide importance. *Harmful Algae*
1012 2005; **4**: 449–470.

1013 26. Khanaychenko AN, I V. Telesh, SO Skarlato. Bloom-forming potentially toxic
1014 dinoflagellates Prorocentrum cordatum in marine plankton food webs. *Protistology*
1015 2019; **13**.

1016 27. Rodríguez I, A Alfonso, E Alonso, JA Rubiolo, M Roel, A Vlamis, P Katikou, SA
1017 Jackson, ML Menon, A Dobson, LM Botana. The association of bacterial C 9 -based
1018 TTX-like compounds with Prorocentrum minimum opens new uncertainties about
1019 shellfish seafood safety. *Sci Rep* 2017; **7**.

1020 28. Seebens H, N Schwartz, PJ Schupp, B Blasius. Predicting the spread of marine species
1021 introduced by global shipping. *Proc Natl Acad Sci U S A* 2016; **113**: 5646–5651.

1022 29. Stephens TG, RA González-Pech, Y Cheng, AR Mohamed, DW Burt, D Bhattacharya,
1023 MA Ragan, CX Chan. Genomes of the dinoflagellate Polarella glacialis encode
1024 tandemly repeated single-exon genes with adaptive functions. *BMC Biol* 2020; **18**: 1–

21.

30. González-Pech RA, TG Stephens, Y Chen, AR Mohamed, Y Cheng, S Shah, KE Dougan, MDA Fortuin, R Lagorce, DW Burt, D Bhattacharya, MA Ragan, CX Chan. Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural divergence in family Symbiodiniaceae and genus Symbiodinium. *BMC Biol* 2021; **19**.

31. Dougan KE, AJ Bellantuono, T Kahlke, RM Abbriano. Title : Whole-genome duplication in an algal symbiont serendipitously confers thermal tolerance to corals Authors : Affiliations : Abstract : 2022; 1–19.

32. Chen Y, S Shah, KE Dougan, MJH Van Oppen. Improved Cladocopium goreaui genome assembly reveals features of a facultative coral symbiont and the complex evolutionary history of dinoflagellate genes. *BioRvix* 2022; 35–36.

33. John U, Y Lu, S Wohlrab, M Groth, J Janouškovec, GS Kohli, FC Mark, U Bickmeyer, S Farhat, M Felder, S Frickenhaus, L Guillou, PJ Keeling, A Moustafa, BM Porcel, K Valentin, G Glöckner. An aerobic eukaryotic parasite with functional mitochondria that likely lacks a mitochondrial genome. *Sci Adv* 2019; **5**.

34. Stephens TG, MA Ragan, D Bhattacharya, CX Chan. Core genes in diverse dinoflagellate lineages include a wealth of conserved dark genes with unknown functions. *Sci Rep* 2018; **8**: 1–11.

35. Voolstra CR, JJ Valenzuela, S Turkarslan, A Cárdenas, BCC Hume, G Perna, C Buitrago-López, K Rowe, M V. Orellana, NS Baliga, S Paranjape, G Banc-Prandi, J Bellworthy, M Fine, S Frias-Torres, DJ Barshis. Contrasting heat stress response patterns of coral holobionts across the Red Sea suggest distinct mechanisms of thermal tolerance. *Mol Ecol* 2021; **30**: 4466–4480.

36. Ibarbalz FM, N Henry, MC Brandão, S Martini, G Busseni, H Byrne, LP Coelho, H Endo, JM Gasol, AC Gregory, F Mahé, J Rigonato, M Royo-Llonch, G Salazar, I Sanz-Sáez, E Scalco, D Soviadan, AA Zayed, A Zingone, et al. Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell* 2019.

37. Alexander MA, JD Scott, KD Friedland, KE Mills, JA Nye, AJ Pershing, AC Thomas. Projected sea surface temperatures over the 21st century: Changes in the mean, variability and extremes for large marine ecosystem regions of Northern Oceans. *Elementa* 2018; **6**.

38. Singh A, CP Shannon, B Gautier, F Rohart, M Vacher, SJ Tebbutt, KAL Cao. DIABLO: An integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* 2019; **35**.

39. Salvucci ME, SJ Crafts-Brandner. Inhibition of photosynthesis by heat stress: The activation state of Rubisco as a limiting factor in photosynthesis. *Physiol Plant* . 2004. , **120**

40. Schroda M, D Hemme, T Mühlhaus. The Chlamydomonas heat stress response. *Plant J* 2015; **82**.

41. Verma A, A Barua, R Ruvindy, H Savela, PA Ajani, SA Murray. The genetic basis of toxin biosynthesis in dinoflagellates. *Microorganisms* . 2019. MDPI AG. , **7**

42. Shi X, H Zhang, S Lin. Tandem repeats, high copy number and remarkable diel

1067 expression rhythm of form II RuBisCO in Prorocentrum donghaiense (Dinophyceae).
1068 *PLoS One* 2013; **8**.

1069 43. Zhang H, S Lin. COMPLEX GENE STRUCTURE OF THE FORM II RUBISCO IN
1070 THE DINOFLAGELLATE PROROCENTRUM MINIMUM (DINOPHYCEAE) 1. *J*
1071 *Phycol* 2003; **39**: 1160–1171.

1072 44. Bruce BD. Chloroplast transit peptides: Structure, function and evolution. *Trends Cell*
1073 *Biol* . 2000. , **10**

1074 45. Lee DW, I Hwang. Evolution and design principles of the diverse chloroplast transit
1075 peptides. *Mol Cells* . 2018. , **41**

1076 46. Csurös M, IB Rogozin, E V. Koonin. Extremely intron-rich genes in the alveolate
1077 ancestors inferred with a flexible maximum-likelihood approach. *Mol Biol Evol* 2008;
1078 **25**.

1079 47. Rogozin IB, L Carmel, M Csuros, E V. Koonin. Origin and evolution of spliceosomal
1080 introns. *Biol Direct* 2012; **7**: 1–28.

1081 48. Huff JT, D Zilberman, SW Roy. Mechanism for DNA transposons to generate introns
1082 on genomic scales. *Nature* 2016; **538**: 533–536.

1083 49. Van Der Burgt A, E Severing, PJGM De Wit, J Collemare. Birth of new spliceosomal
1084 introns in fungi by multiplication of introner-like elements. *Curr Biol* 2012; **22**.

1085 50. Roy SW, L Gozashti, BA Bowser, BN Weinstein, GE Larue. Massive Intron Gain in
1086 the Most Intron-Rich Eukaryotes is Driven by Introner-Like Transposable Elements of
1087 Unprecedented Diversity and Flexibility. *SSRN Electron J* 2020.

1088 51. Mathur S, D Agrawal, A Jajoo. Photosynthesis: Response to high temperature stress. *J*
1089 *Photochem Photobiol B Biol* 2014; **137**.

1090 52. Ivanov AG, MY Velitchkova, SI Allakhverdiev, NPA Huner. Heat stress-induced
1091 effects of photosystem I: an overview of structural and functional responses.
1092 *Photosynth Res* . 2017. , **133**

1093 53. Posch BC, BC Kariyawasam, H Bramley, O Coast, RA Richards, MP Reynolds, R
1094 Trethowan, OK Atkin. Exploring high temperature responses of photosynthesis and
1095 respiration to improve heat tolerance in wheat. *J Exp Bot* . 2019. , **70**

1096 54. Barten R, M Kleisman, G D'Ermo, H Nijveen, RH Wijffels, MJ Barbosa. Short-term
1097 physiologic response of the green microalga Picochlorum sp. (BPE23) to supra-
1098 optimal temperature. *Sci Rep* 2022; **12**.

1099 55. Barshis DJ, JT Ladner, TA Oliver, SR Palumbi. Lineage-specific transcriptional
1100 profiles of Symbiodinium spp. Unaltered by heat stress in a coral host. *Mol Biol Evol*
1101 2014; **31**: 1343–1352.

1102 56. Zhang N, EM Mattoon, W McHargue, B Venn, D Zimmer, K Pecani, J Jeong, CM
1103 Anderson, C Chen, JC Berry, M Xia, S-C Tzeng, E Becker, L Pazouki, B Evans, F
1104 Cross, J Cheng, KJ Czymmek, M Schroda, et al. Systems-wide analysis revealed
1105 shared and unique responses to moderate and acute high temperatures in the green alga
1106 Chlamydomonas reinhardtii. *Commun Biol* 2022; **5**.

1107 57. McLachlan RH, JT Price, SL Solomon, AG Grottoli. Thirty years of coral heat-stress

1108 experiments: a review of methods. *Coral Reefs* 2020; **39**: 885–902.

1109 58. Maire J, LL Blackall, MJH van Oppen. Intracellular bacterial symbionts in corals:
1110 Challenges and future directions. *Microorganisms* . 2021. , **9**

1111 59. Gallaher SD, RJ Craig, I Ganesan, SO Purvine, SR McCorkle, J Grimwood, D
1112 Strenkert, L Davidi, MS Roth, TL Jeffers, MS Lipton, KK Niyogi, J Schmutz, SM
1113 Theg, CE Blaby-Haas, SS Merchant. Widespread polycistronic gene expression in
1114 green algae. *Proc Natl Acad Sci U S A* 2021; **118**: 1–10.

1115 60. Lee MG. An RNA polymerase II promoter in the hsp70 locus of Trypanosoma brucei.
1116 *Mol Cell Biol* 1996; **16**: 1220–1230.

1117 61. Crosby MA, L Sian Gramates, G dos Santos, BB Matthews, SE St. Pierre, P Zhou, AJ
1118 Schroeder, K Falls, DB Emmert, SM Russo, WM Gelbart, W Gelbart, NH Brown, T
1119 Kaufman, M Werner-Washburne, R Cripps, K Broll, M Crosby, D Emmert, et al. Gene
1120 model annotations for Drosophila melanogaster: The rule-benders. *G3 Genes,*
1121 *Genomes, Genet* 2015; **5**: 1737–1749.

1122 62. Shimakawa G, E Shoguchi, A Burlacot, K Ifuku, Y Che, M Kumazawa, K Tanaka, S
1123 Nakanishi. Coral symbionts evolved a functional polycistronic flavodiiron gene.
1124 *Photosynth Res* 2022; **151**: 113–124.

1125 63. Guillard RRL, PE Hargraves. Stichochrysis immobilis is a diatom, not a chrysophyte.
1126 *Phycologia* 1993; **32**.

1127 64. Sunda WG, NM Price, FMM Morel. Trace Metal Ion Buffers and Their Use in Culture
1128 Studies. *Algal Culturing Techniques*. 2005.

1129 65. Michelle Wood A, RC Everroad, LM Wingard. Measuring Growth Rates in
1130 Microalgal Cultures. *Algal Culturing Techniques*. 2005.

1131 66. Levi-Setti R, KL Gavrilov, PJ Rizzo. Divalent cation distribution in dinoflagellate
1132 chromosomes imaged by high-resolution ion probe mass spectrometry. *Eur J Cell Biol*
1133 2008; **87**.

1134 67. Zimin A V., G Marçais, D Puiu, M Roberts, SL Salzberg, JA Yorke. The MaSuRCA
1135 genome assembler. *Bioinformatics* 2013; **29**.

1136 68. Xue W, JT Li, YP Zhu, GY Hou, XF Kong, YY Kuang, XW Sun. L_RNA_scaffolder:
1137 Scaffolding genomes with transcripts. *BMC Genomics* 2013; **14**.

1138 69. Wang M, L Kong. pblat: A multithread blat algorithm speeding up aligning sequences
1139 to genomes. *BMC Bioinformatics* 2019; **20**.

1140 70. Laetsch DR, ML Blaxter. BlobTools: Interrogation of genome assemblies.
1141 *F1000Research* 2017; **6**.

1142 71. Chen S, Y Zhou, Y Chen, J Gu. Fastp: An ultra-fast all-in-one FASTQ preprocessor.
1143 *Bioinformatics*. 2018.

1144 72. Grabherr MG, BJ Haas, M Yassour, JZ Levin, DA Thompson, I Amit, X Adiconis, L
1145 Fan, R Raychowdhury, Q Zeng, Z Chen, E Mauceli, N Hacohen, A Gnirke, N Rhind, F
1146 Di Palma, BW Birren, C Nusbaum, K Lindblad-Toh, et al. Full-length transcriptome
1147 assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011; **29**.

1148 73. Li W, A Godzik. Cd-hit: A fast program for clustering and comparing large sets of

1149 protein or nucleotide sequences. *Bioinformatics* 2006; **22**.

1150 74. Kim D, JM Paggi, C Park, C Bennett, SL Salzberg. Graph-based genome alignment
1151 and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019; **37**.

1152 75. Chen Y, RA González-Pech, TG Stephens, D Bhattacharya, CX Chan. Evidence That
1153 Inconsistent Gene Prediction Can Mislead Analysis of Dinoflagellate Genomes. *J*
1154 *Phycol* . 2020. , **56**

1155 76. Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;
1156 **34**.

1157 77. Haas BJ, AL Delcher, SM Mount, JR Wortman, RK Smith, LI Hannick, R Maiti, CM
1158 Ronning, DB Rusch, CD Town, SL Salzberg, O White. Improving the Arabidopsis
1159 genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res*
1160 2003; **31**.

1161 78. Remmert M, A Biegert, A Hauser, J Söding. HHblits: Lightning-fast iterative protein
1162 sequence searching by HMM-HMM alignment. *Nat Methods* 2012; **9**.

1163 79. Korf I. Gene finding in novel genomes. *BMC Bioinformatics* 2004; **5**.

1164 80. Hoff KJ, M Stanke. Predicting Genes in Single Genomes with AUGUSTUS. *Curr*
1165 *Protoc Bioinforma* 2019; **65**.

1166 81. Borodovsky M, A Lomsadze. Eukaryotic gene prediction using GeneMark.hmm-E and
1167 GeneMark-ES. *Curr Protoc Bioinforma* 2011.

1168 82. Campbell MS, C Holt, B Moore, M Yandell. Genome Annotation and Curation Using
1169 MAKER and MAKER-P. *Curr Protoc Bioinforma* 2014; **2014**.

1170 83. Haas BJ, SL Salzberg, W Zhu, M Pertea, JE Allen, J Orvis, O White, CR Robin, JR
1171 Wortman. Automated eukaryotic gene structure annotation using EVidenceModeler
1172 and the Program to Assemble Spliced Alignments. *Genome Biol* 2008; **9**.

1173 84. Johnson LK, H Alexander, CT Brown. Re-assembly, quality evaluation, and
1174 annotation of 678 microbial eukaryotic reference transcriptomes. *Gigascience* 2019; **8**.

1175 85. Emms DM, S Kelly. OrthoFinder: Phylogenetic orthology inference for comparative
1176 genomics. *Genome Biol* 2019; **20**.

1177 86. Pachiadaki MG, JM Brown, J Brown, JJ La Clair, SW Chisholm, O Bezuidt, PM
1178 Berube, SJ Biller, NJ Poulton, MD Burkart, R Stepanauskas. Charting the Complexity
1179 of the Marine Microbiome through Single-Cell Genomics. *Cell* 2019; **179**.

1180 87. Katoh K, DM Standley. MAFFT multiple sequence alignment software version 7:
1181 Improvements in performance and usability. *Mol Biol Evol* 2013; **30**.

1182 88. Capella-Gutiérrez S, JM Silla-Martínez, T Gabaldón. trimAl: A tool for automated
1183 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009; **25**.

1184 89. Minh BQ, HA Schmidt, O Chernomor, D Schrempf, MD Woodhams, A Von Haeseler,
1185 R Lanfear, E Teeling. IQ-TREE 2: New Models and Efficient Methods for
1186 Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020; **37**.

1187 90. Mŕazek J, S Xie. Pattern locator: A new tool for finding local sequence patterns in
1188 genomic DNA sequences. *Bioinformatics* 2006; **22**.

1189 91. Farhat S, P Le, E Kayal, B Noel, E Bigeard, E Corre, F Maumus, I Florent, A Alberti,
1190 JM Aury, T Barbeyron, R Cai, C Da Silva, B Istace, K Labadie, D Marie, J Mercier, T
1191 Rukwavu, J Szymczak, et al. Rapid protein evolution, organellar reductions, and
1192 invasive intronic elements in the marine aerobic parasite dinoflagellate Amoebophrya
1193 spp (BMC Biology, (2021), 19, 1, (1), 10.1186/s12915-020-00927-9). *BMC Biol* 2021;
1194 **19**: 1–21.

1195 92. Rohart F, B Gautier, A Singh, KA Lê Cao. mixOmics: An R package for 'omics
1196 feature selection and multiple data integration. *PLoS Comput Biol* 2017; **13**.

1197 93. Singh A, CP Shannon, B Gautier, F Rohart, M Vacher, SJ Tebbutt, KAL Cao.
1198 DIABLO: An integrative approach for identifying key molecular drivers from multi-
1199 omics assays. *Bioinformatics* 2019; **35**: 3055–3062.

1200 94. Zhou G, O Soufan, J Ewald, REW Hancock, N Basu, J Xia. NetworkAnalyst 3.0: A
1201 visual analytics platform for comprehensive gene expression profiling and meta-
1202 analysis. *Nucleic Acids Res* 2019; **47**.

1203 95. Paul C, MA Mausz, G Pohnert. A co-culturing/metabolomics approach to investigate
1204 chemically mediated interactions of planktonic organisms reveals influence of bacteria
1205 on diatom metabolism. *Metabolomics* 2013; **9**.

1206 96. Liao Y, GK Smyth, W Shi. FeatureCounts: An efficient general purpose program for
1207 assigning sequence reads to genomic features. *Bioinformatics* 2014; **30**.

1208 97. Hamilton NE, M Ferry. Ggtern: Ternary diagrams using ggplot2. *J Stat Softw* 2018;
1209 **87**.

1210 98. Robinson Mark D, J McCarthy Davis, K Smyth Gordon. edgeR: a Bioconductor
1211 package for differential expression analysis of digital gene expression data.
1212 *Bioinformatics* 2010; **26**.

1213 99. Chen H, PC Boutros. VennDiagram: A package for the generation of highly-
1214 customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 2011; **12**.

1215 100. Supek F, M Bošnjak, N Škunca, T Šmuc. Revigo summarizes and visualizes long lists
1216 of gene ontology terms. *PLoS One* 2011; **6**.

1217 101. Aramaki T, R Blanc-Mathieu, H Endo, K Ohkubo, M Kanehisa, S Goto, H Ogata.
1218 KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive
1219 score threshold. *Bioinformatics* 2020; **36**.

1220 102. Yu G, LG Wang, Y Han, QY He. ClusterProfiler: An R package for comparing
1221 biological themes among gene clusters. *Omi A J Integr Biol* 2012; **16**.

1222 103. Hiller K, J Hangebrauk, C Jäger, J Spura, K Schreiber, D Schomburg. Metabolite
1223 detector: Comprehensive analysis tool for targeted and nontargeted GC/MS based
1224 metabolome analysis. *Anal Chem* 2009; **81**.

1225 104. Bradford MM. A rapid and sensitive method for the quantitation of microgram
1226 quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976;
1227 **72**.

1228 105. Wöhlbrand L, R Rabus, B Blasius, C Feenders. Influence of NanoLC Column and
1229 Gradient Length as well as MS/MS Frequency and Sample Complexity on Shotgun
1230 Protein Identification of Marine Bacteria. *J Mol Microbiol Biotechnol* 2017; **27**.

1231 106. Wöhlbrand L, HS Ruppersberg, C Feenders, B Blasius, HP Braun, R Rabus. Analysis
1232 of membrane-protein complexes of the marine sulfate reducer Desulfobacula toluolica
1233 Tol2 by 1D blue native-PAGE complexome profiling and 2D blue native-/SDS-PAGE.
1234 *Proteomics* 2016; **16**.

1235 107. Neuhoff V, N Arold, D Taube, W Ehrhardt. Improved staining of proteins in
1236 polyacrylamide gels including isoelectric focusing gels with clear background at
1237 nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250.
1238 *Electrophoresis* 1988; **9**.

1239

Data availability statement

The assembled genome and transcriptome, and the predicted gene and protein sequences from *P. cordatum* are available at https://cloudstor.aarnet.edu.au/plus/s/nxC7atJTpJuyr8h. The BioProject accession for the genome and transcriptome data is PRJEB54915. Detailed information on identified proteins and peptides is deposited at fairdomhub.org under the project 293 (DOI will be available once the data are publicly available). For review, the data are available via the following links:

SOP Metabolites:
https://fairdomhub.org/sops/551?code=RrSjntdGsUnt2zRDdFKIKnfmJviLFvw3O7BzG9Lv;

SOP Proteome:
https://fairdomhub.org/sops/549?code=TKCZcudIA05oY%2FSnAFW67R2IwE91d1x%2F9eQp6dnH

Data Metabolites:
https://fairdomhub.org/data_files/6075?code=kyxQ9ljLVTxjU54Wrk5HS62bQj6Z9niiJtT30GSc

Data Shotgun proteomics:
https://fairdomhub.org/data_files/6059?code=GhhJoFC9oBKYd7fWRlE9kZ%2BpMGnGRMD2Cvtf8sLE

Data membrane fraction:
https://fairdomhub.org/data_files/6058?code=QwaeOVVnEvuCXftZqObFSFD82emv0qEqhs9oiHV1

Data 2D LC-MS/MS:
https://fairdomhub.org/data_files/6057?code=4PPrRrP6qwTuD1iekFWzxzRijAyGkbeaySaybJjI

Acknowledgements

The authors want to thank Nicole Heyer and Simone Schrader for technical assistance regarding genome sequencing. We want to thank our laboratory assistant Sinem Bektas for her enormous flexibility, unfailing reliability and good spirits even in the worst of Corona times.

1270

1271 ## Extended Data

1272 **Table 1**: Statistics of assembled genomes of *P. cordatum* and other dinoflagellates

1273 **Table 2**: Statistics of predicted genes in *P. cordatum* and other dinoflagellates

1274 **Extended Data Figure 1**: Heatmap of expression levels for the inferred multi-omics
1275 signature that is indicative of a general heat stress response. Values are scaled across rows
1276 with proteins and transcripts up-regulated during heat stress shown in (a) and those down-
1277 regulated during heat stress in (b).

1278 **Extended Data Figure 2**: Heatmap of expression levels for the inferred multi-omics
1279 signature that is indicative of a 26°C and 30°C specific heat stress response. Values are
1280 scaled across rows with proteins and transcripts up-regulated during heat stress shown in (a)
1281 and those down-regulated during heat stress in (b).

1282 **Extended Data Figure 3**: Heatmap of all metabolites ($p < 0.001$) at 20°C, 26°C and 30°C in
1283 *P. cordatum*. Five biological- and three technical replicates were run per experiment. Relative
1284 intensity values are shown and data were normalized to internal standard and cell count.
1285 Normalization for heatmap was done by using z-score and significance was calculated by
1286 two-tailed T-test.

1287 **Extended Data Figure 4**: Detailed reconstruction of the central carbon metabolism of *P.*
1288 *cordatum* and its response to heat stress on transcript (blue), protein (red) and metabolite
1289 (yellow) level.

1290 **Extended Data Figure 5**: Functional alteration of the transcriptome under different
1291 temperatures and growth phases. Molecular function (MF) GO terms enriched in up-
1292 regulated genes at 26°C compared to 20°C (a). The MF GO terms enriched in down-
1293 regulated genes (b).

1294 **Extended Data Figure 6**: Clusters of genes with similar expression patterns across different
1295 conditions. (a) Clusters based on samples from exponential phase. (b) The expression
1296 patterns across conditions of clusters for exponential phase samples. (c) The MF GO terms
1297 enriched in cluster 2, 3, 5, and 6 in the stationary phase.

1298 **Extended Data Figure 7**: Transcript exon-intron plot Hsp70

1299 **Extended Data Figure 8**: Transcript profiles of Hsp70 (a) and RuBisCo (b) gene models
1300 according to clustering in Figure 6a and 6c.

1301

1302 ## Supplementary Information

1303 ## Supplementary Figures

1304 **Figure S1**: Analysis of *P. cordatum* genome based on (a) distribution of 21-mers indicating a
1305 haploid genome, (b) genome-size estimation based on 21-mers using GenomeScope2, and (c)
1306 repeat landscape excluding unknown repeats.

1307 **Figure S2**: Maximum likelihood protein tree of a putative hydantoin racemase indicating
1308 bacterial origin of the encoding gene in *P. cordatum*.

1309 **Figure S3.** Unidirectional coding of genes in dinoflagellate genomes based on gene-
1310 orientation changes in ten-gene windows.

1311 **Figure S4.** Gene functions encoded in the genome of *P. cordatum* and the other five
1312 representative taxa based on count of Gene Ontology (GO) terms.

1313 **Figure S5.** The transcriptome landscape under different temperatures and growth phases. The
1314 PCA visualization of samples based on the FPKM expression values (a). Number of shared
1315 differentially expressed genes among different comparisons (b).

1316 **Figure S6.** The proteome landscape of *P. cordatum* heat stress response. NMDS visualisation
1317 of individual samples based on peptide count data [max error = 1.24 × 10-11] (a). Number of
1318 shared/specifc proteins across the studied conditions (b).

1319 **Figure S7.** Network visualization of multi-omics signature inferred from both proteins and
1320 transcripts that represent a general heat stress response. Colors of the lines indicate the nature
1321 of the relationships between proteins and transcripts, with positive relationships in green and
1322 negative relationships in purple.

1323 **Figure S8.** Network visualization of multi-omics signature inferred from both proteins and
1324 transcripts that represent a 26°C and 30°C specific heat stress response. Colors of the lines
1325 indicate the nature of the relationships between proteins and transcripts, with positive
1326 relationships in green and negative relationships in purple.

1327 **Figure S9**. The regulation pattern (a) and functional enrichment (b-d) of dispersed
1328 duplication genes in ortho-groups. All ortho-groups with ≥3 DE genes in stationary phase
1329 were included in panel a. For the functional enrichment analysis in panel b, only ortho-groups
1330 with ≥20 gene members were taken into account which cover 13,529 genes.

1331 **Figure S10.** Detailed information on gene models containing multiple coding units (CU).
1332 Scale model of gene models annotated as RuBisCo, that contain only fragments of the
1333 complete Rubisco CU, gene model s10_g491 as reference for a 3 CU version (a). Scale
1334 model of Rubisco gene models with respective exons (b). Scale model of gene models
1335 containing multiple CUs encoding protein functions in central metabolism (c).

1336 **Figure S11.** Comparison of RuBisCo gene models of *P. cordatum* as well as with other
1337 dinoflagellate species. Alignment of the predicted leader sequences, sequences N-terminal of
1338 the first CU and the region between two complete CUs (i.e. spacer regions) (a). Alignment of
1339 the spacer regions of chloroplast localized proteins (RuBisCo and phosphoglycerate kinase,
1340 top) and non-chloroplastic proteins (Hsp70 and glyceraldehyde 3-phosphate dehydrogenase)
1341 (mitochondrial), bottom) (b). Histogram of RuBisCo proteins encoded in the genomes of *P.*
1342 *cordatum* and *S. microadriaticum* (c).

1343 Supplementary Tables

1344 **Table S1**: The 47 *P. cordatum* proteins with top bacterial hit in GORG-Tropics.

1345 **Table S2**: *P. cordatum* genes that contain introner elements in one or more introns.

1346 **Table S3**: Introner elements in genomes of dinoflagellates.

1347 **Table S4**: Classification of gene duplicates in *P. cordatum* and the five representative
1348 dinoflagellate taxa.

1349 **Table S5**: Functional annotation of 85,849 *P. cordatum* proteins.

1350 **Table S6**: Annotated protein sequences from genome of *P. cordatum* and other
1351 dinoflagellates.

1352 **Table S7**: Tandemly duplicated gene blocks in *P. cordatum* genome.

1353 **Table S8**: RNAseq read statistics

1354 **Table S9**: FPKM values per gene model

1355 **Table S10**: Normalized peptide counts per gene model

1356 **Table S11**: Relative intensity values of metabolite data measured by GC-MS from *P.*
1357 *cordatum* grown at 20°C, 26°C and 30°C. Five biological and three technical replicates were
1358 run per experiment. Data were normalized to internal standard and cell count. Significance of
1359 differences between the conditions was calculated by two-tailed T-test from the average of
1360 the technical replicates.

1361 **Table S12**: Transcripts and proteins of components 1 and 2

1362 **Table S13**: Transcript (FPKM) and protein (normalized peptide counts) data of central
1363 carbon metabolism, oxidative phosphorylation and photosynthesis at 20°C, 26°C and 30°C in
1364 *P. cordatum* (Fig 4). Central carbon metabolism enzymes are shown by their EC-number and
1365 oxidative phosphorylation and photosynthesis were based on the annotation. Average was
1366 calculated from three biological and three technical replicates. Wilcoxon tests were applied to
1367 calculate significance (see table S14).

1368 **Table S14**: The comparison of central carbon metabolism enzymes expression levels
1369 (transcriptome) between 26°C-St and 20°C-St

1370 **Table S15**: Differential gene expression analysis (transcriptome)

1371 **Table S16**: GO term enrichment analysis

1372 **Table S17**: KO gene expression

1373 **Table S18**: KEGG enrichment analysis

1374 **Table S19:** Clustered genes during exponential growth phase

1375 **Table S20**: Clustered genes during stationary growth phase

1376 **Table S21**: Differentially edited mRNA sites identified in *P. cordatum* genes

1377 **Table S22**: Differential exon usage identified in *P. cordatum* genes.

1378 **Table S23**: Genome sequence data generated for this study.

1379 **Table S24**: Statistics of preliminary and final genome assemblies of *P. cordatum*.

1380 **Table S25**: Dinoflagellate proteins used to guide prediction of protein-coding genes from the
1381 *P. cordatum* genome

1382 **Table S26**: Protein sequences from 31 dinoflagellate taxa used for identification of
1383 homologous sequence sets.

1384 **Table S27**: The protein sequences from 82 eukaryote taxa used in this study.

1385 **Table S28**: The 688,212 protein sequences from 543 single-cell assembled genomes in
1386 GORG-Tropics used in this study.

1387

1388 **Supplementary Note**

1389 1 Horizontal gene transfer
1390 2 Introner elements & tandem repeats
1391 3 Transcriptome landscape
1392

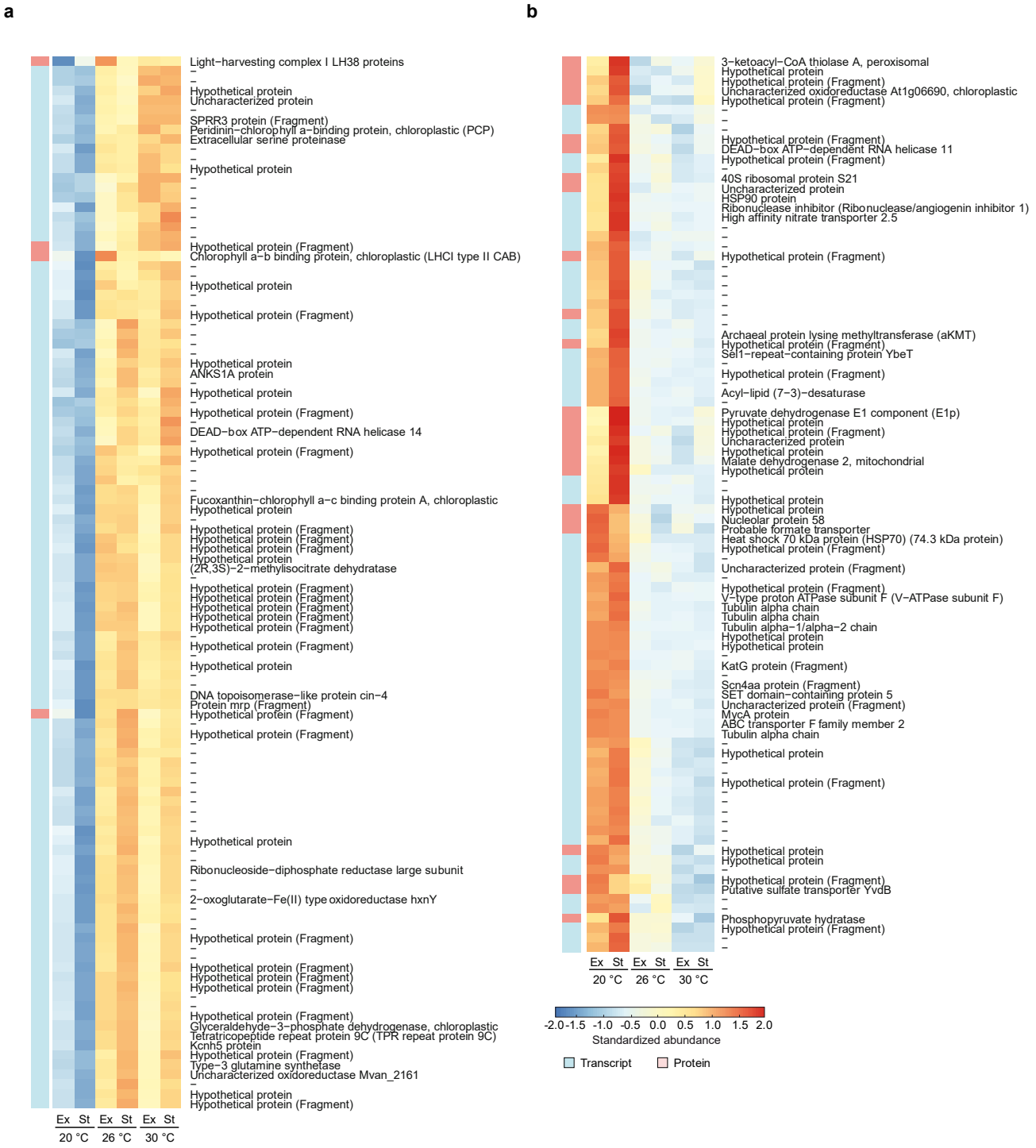
1393

Extended Data Table 1. Statistics of assembled genomes of *P. cordatum* and other dinoflagellates.

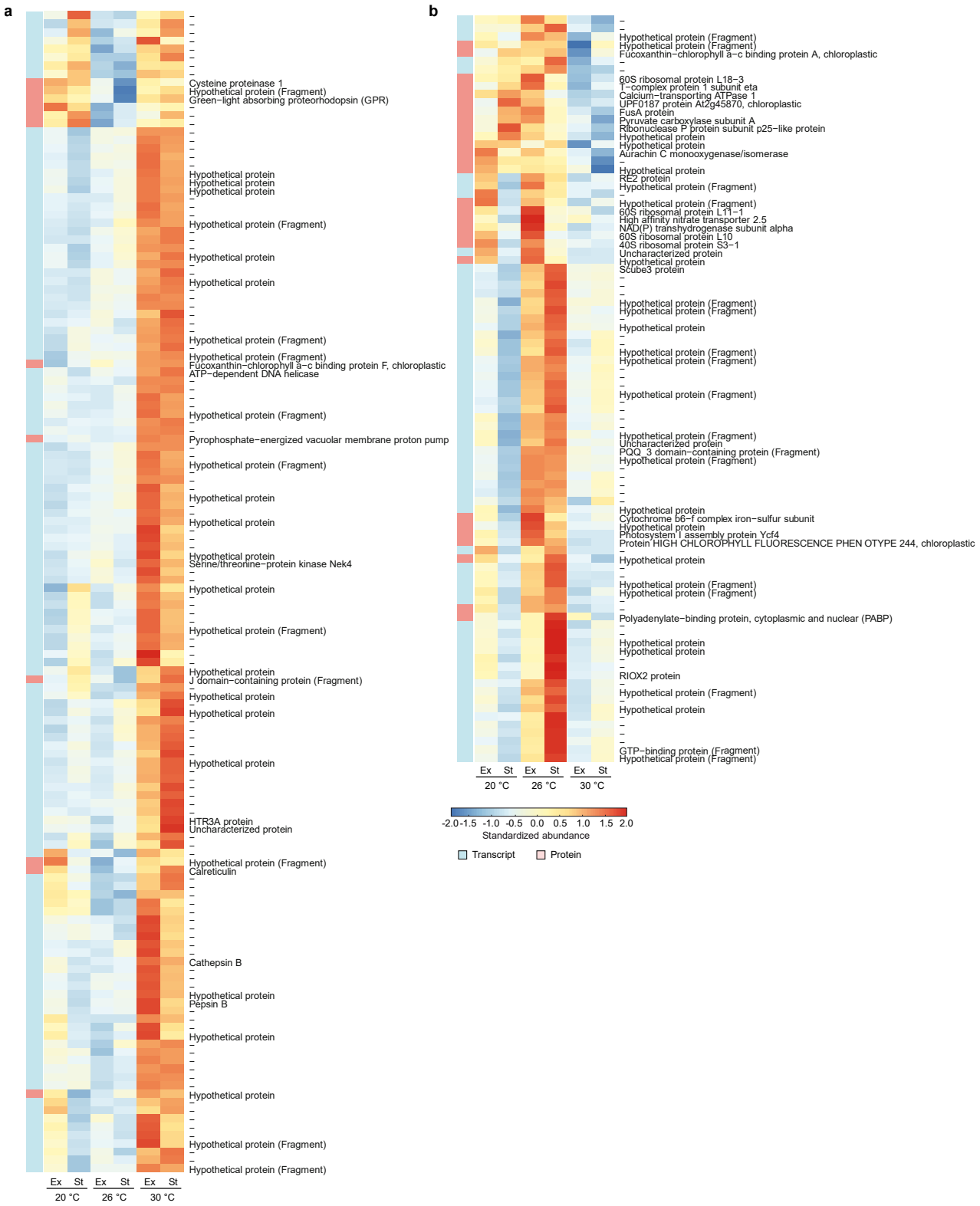
Species	*Prorocentrum cordatum*	*Polarella glacialis*	*Symbiodium natans*	*Durusdinium trenchii*	*Cladocopium goreaui*	*Amoebophrya ceratii*
Isolate	CCMP1329	CCMP1383	CCMP2548	CCMP2556	SCF055	AT5.2
Reference	This study	Stephens et al. 2020	Gonzalez-Pech et al. 2021	Dougan et al. 2022	Chen et al. 2022	John et al. 2019
Assembly size (Gbp)	4.15	2.99	0.76	1.71	1.17	0.09
Estimated genome size based on *k*-mers (Gbp)	4.75	1.48	0.74	1.05	1.30	0.12
Number of scaffolds	22,724	33,494	2,855	29,137	6,843	2,351
Genome scaffolds N50 (Kbp)	349.2	170.3	610.5	774.3	353.9	84.0
Maximum scaffold length (Mbp)	3.56	2.17	3.40	4.57	7.38	0.54
Genome GC-content (%)	59.7	45.9	51.8	49.8	44.4	55.9
percentage of BUSCO proteins recovered (genome mode; alveolata_odb10)	60.2	60.2	62.0	61.4	66.7	82.5

Extended Data Table 2. Statistics of predicted genes in *P. cordatum* and other dinoflagellates.

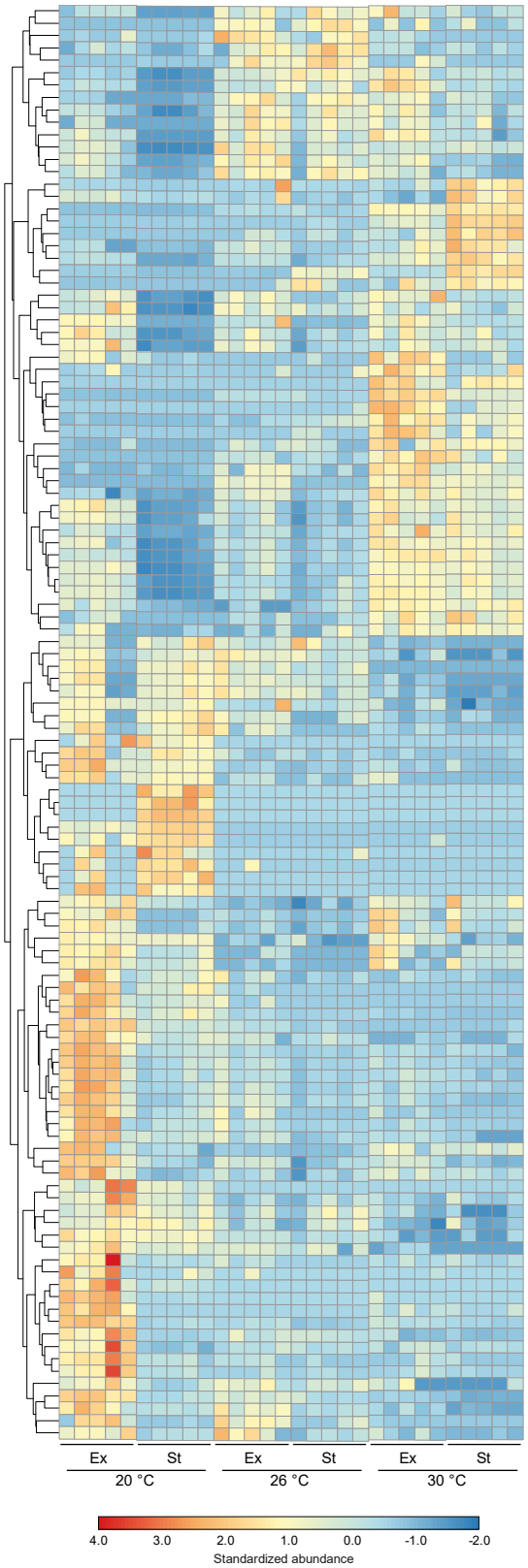
Species		*Prorocentrum cordatum*	*Polarella glacialis*	*Symbiodium natans*	*Durusdinium trenchii*	*Cladocopium goreaui*	*Amoebophrya ceratii*
Isolate		CCMP1329	CCMP1383	CCMP2548	CCMP2556	SCF055	AT5.2
Reference		This study	Stephens et al. 2020	González-Pech et al. 2021	Dougan et al. 2022	Chen et al. 2022	John et al. 2019
Number of predicted genes		85,849	58,232	35,270	55,799	45,322	19,925
% recovery of BUSCO proteins (alveolata_odb10)		61.4	70.2	74.9	69.6	82.4	86.5
Genes with transcript support (%)		84.8	94.0	83.0	75.7	82.5	24.4
Average gene length (bp)		24,462	16,206	8,780	15,334	15,745	2,772
Average CDS length (bp)		2,798	1,230	1,660	1,647	2,018	1,964
CDS GC-content (%)		65.9	57.8	58.16	55.69	54.24	60.8
Number of exons per gene		11.7	11.6	15.7	16.7	17.2	3.4
Average exon length (bp)		239.9	105.7	106	98.7	120.4	578.7
Genes with introns (%)		83.7	73.8	85.5	93.1	95.9	71.3
Number of introns per gene		9.8	10.6	14.7	15.7	16.2	2.4
Average intron length (bp)		4709	1408	486	869	839	377
Splice donor motif (%)	GT	24.7	28.8	23.6	30.3	36.6	99.9
	GC	51.4	52.7	58	52.3	43.6	<0.1
	GA	23.8	18.5	18.4	17.4	19.8	0
Splice acceptor with AGG motif (%)		79.7	96.9	97.1	96.5	96.1	56.5
Number of intergenic regions		48,574	35,271	33,042	47,452	39,720	17,856
Average length of intergenic regions (bp)		26,278	21,625	11,585	13,222	7,388	1,522



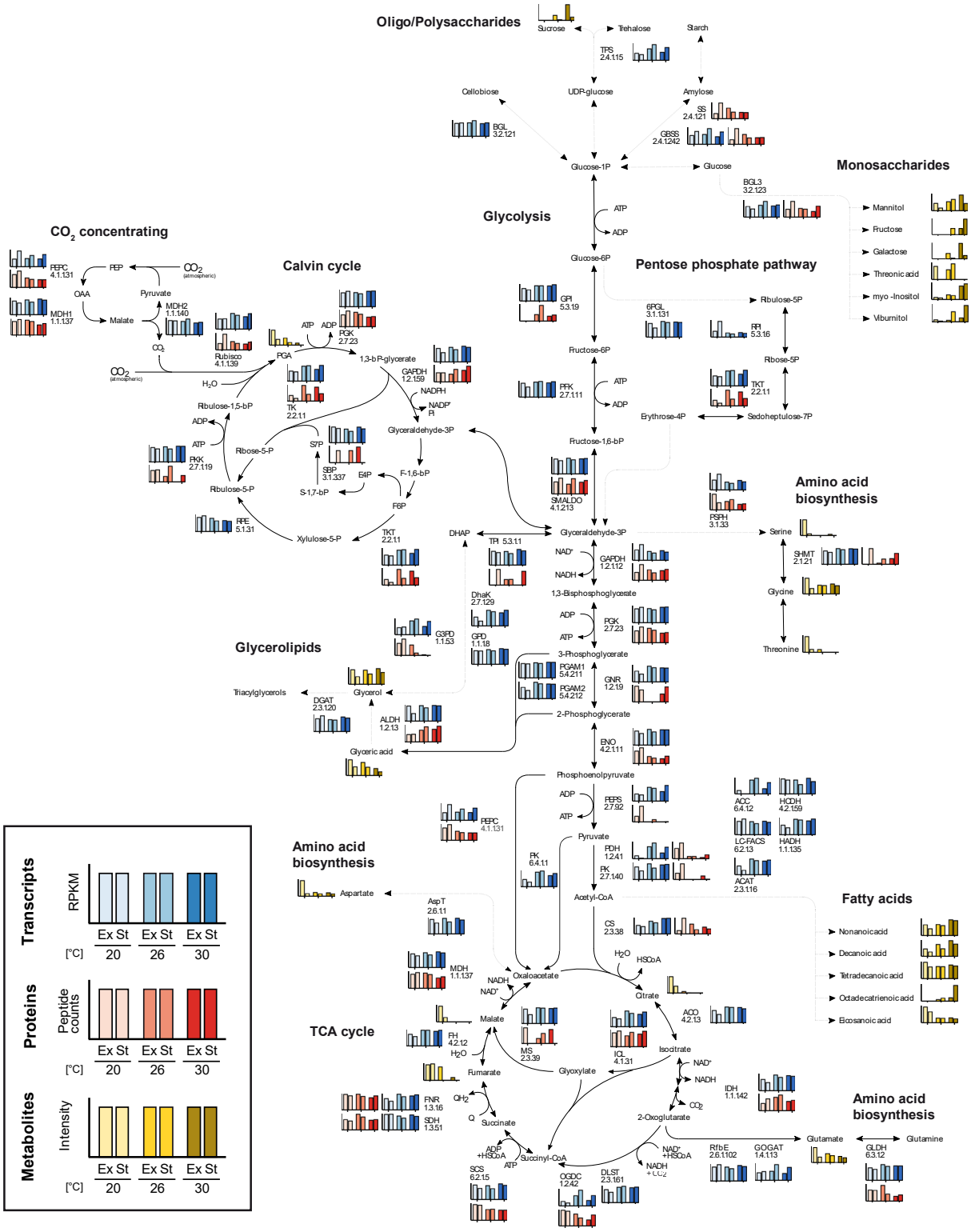
ED Figure 1: Heatmap of expression levels for the inferred multi-omics signature that is indicative of a general heat stress response. Values are scaled across rows with proteins and transcripts up-regulated during heat stress shown in (a) and those down-regulated during heat stress in (b).



ED Figure 2: Heatmap of expression levels for the inferred multi-omics signature that is indicative of a 26C- and 30C-specific heat stress response. Values are scaled across rows with proteins and transcripts up-regulated during heat stress shown in (a) and those down-regulated during heat stress in (b).

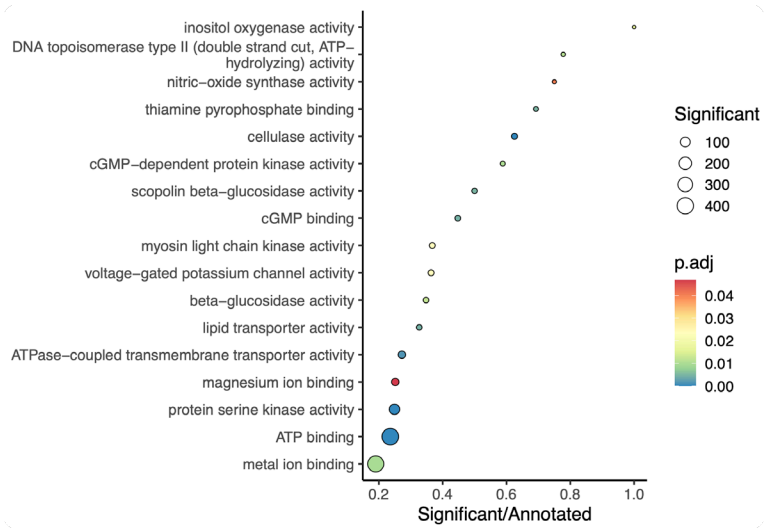


ED Figure 3: Heatmap of all metabolites ($p < 0.001$) at 20°C, 26°C and 30°C in *P. cordatum*. Five biological- and three technical replicates were run per experiment. Relative intensity values are shown and data were normalized to internal standard and cell count. Normalization for heatmap was done by using z-score and significance was calculated by two-tailed T-test.

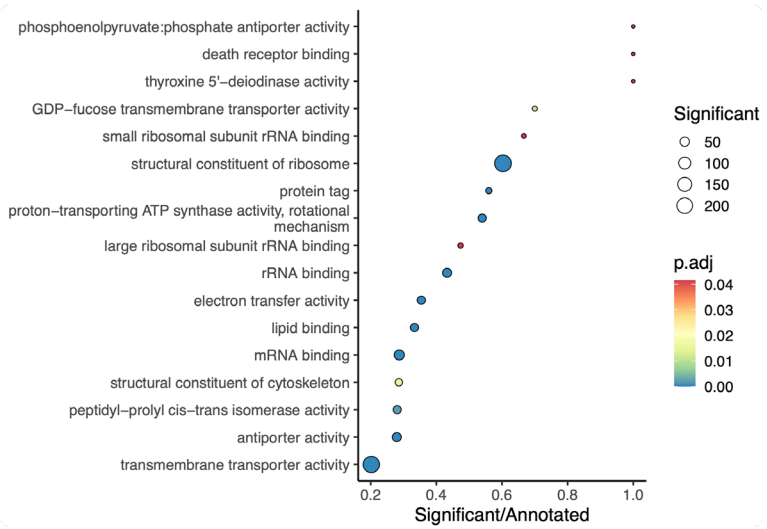


ED Figure 4: Detailed reconstruction of the central carbon metabolism of *P. cordatum* and its response to heat stress on transcript (blue), protein (red) and metabolite (yellow) level.

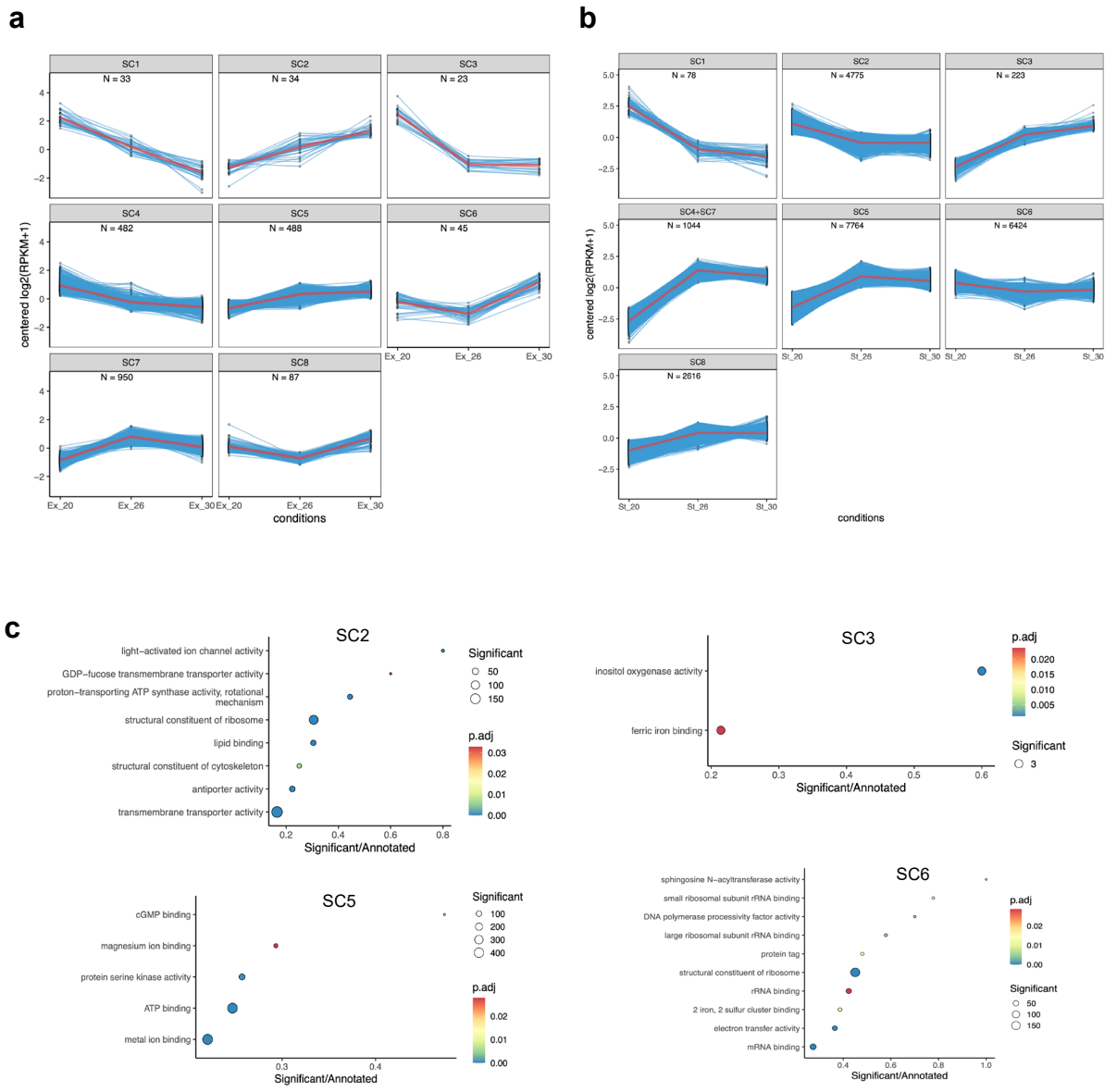
a



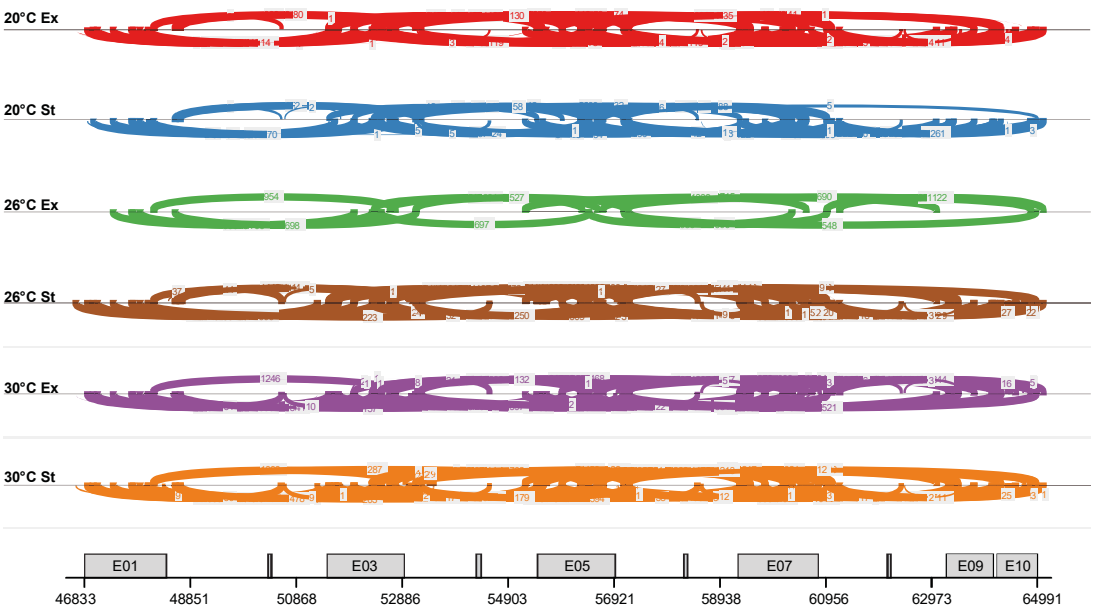
b



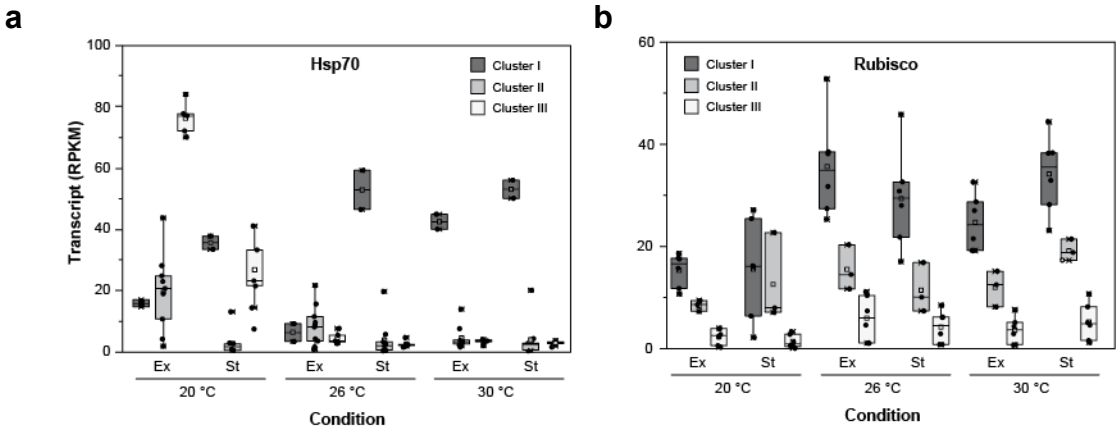
ED Figure 5: Functional alteration of the transcriptome under different temperatures and growth phases. Molecular function (MF) GO terms enriched in up-regulated genes at 26°C compared to 20°C (a). The MF GO terms enriched in down-regulated genes (b).



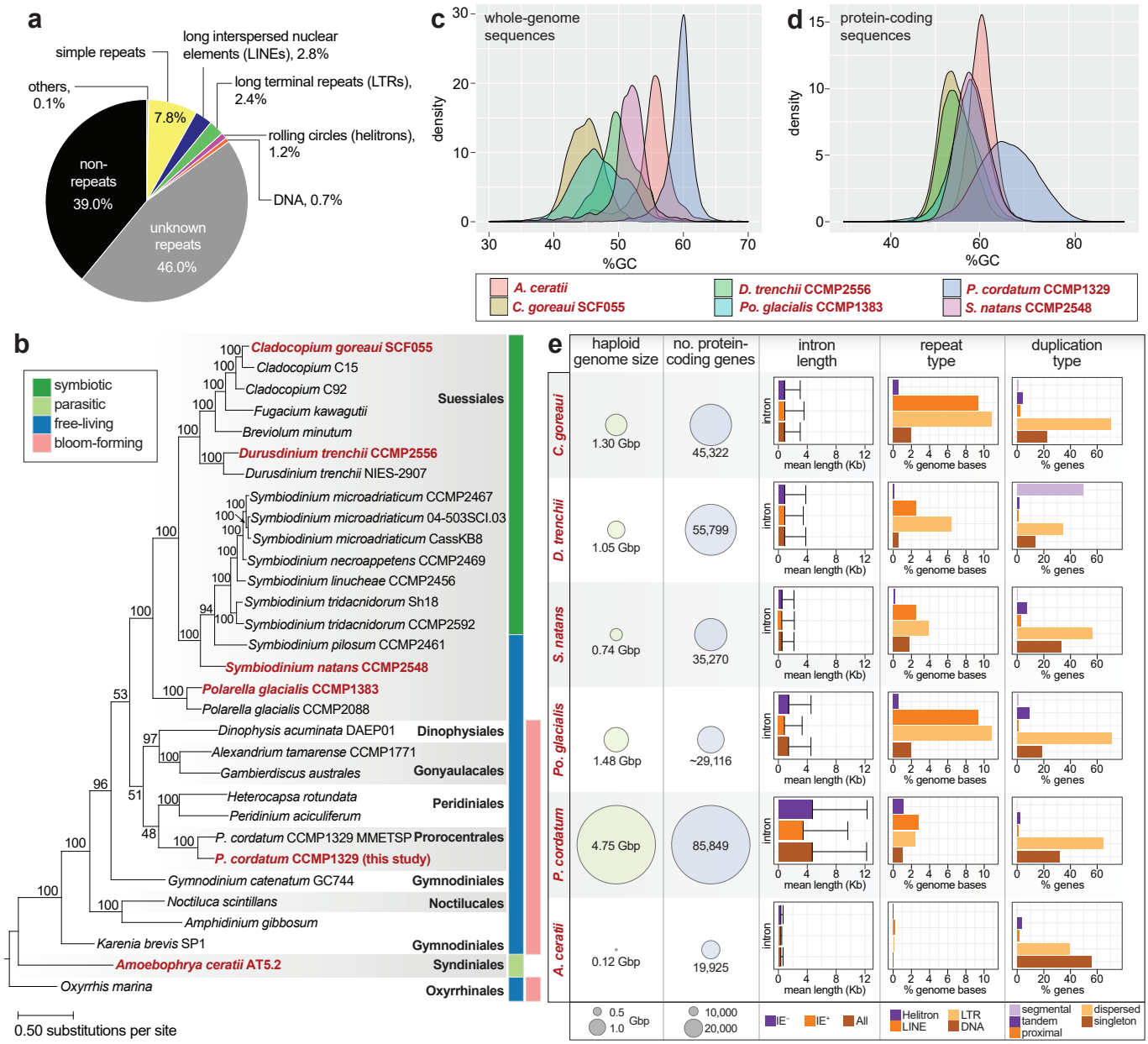
ED Figure 6: Clusters of genes with similar expression patterns across different conditions. (a) Clusters based on samples from exponential phase. (b) The expression patterns across conditions of clusters for exponential phase samples. (c) The MF GO terms enriched in cluster 2, 3, 5, and 6 in the stationary phase.

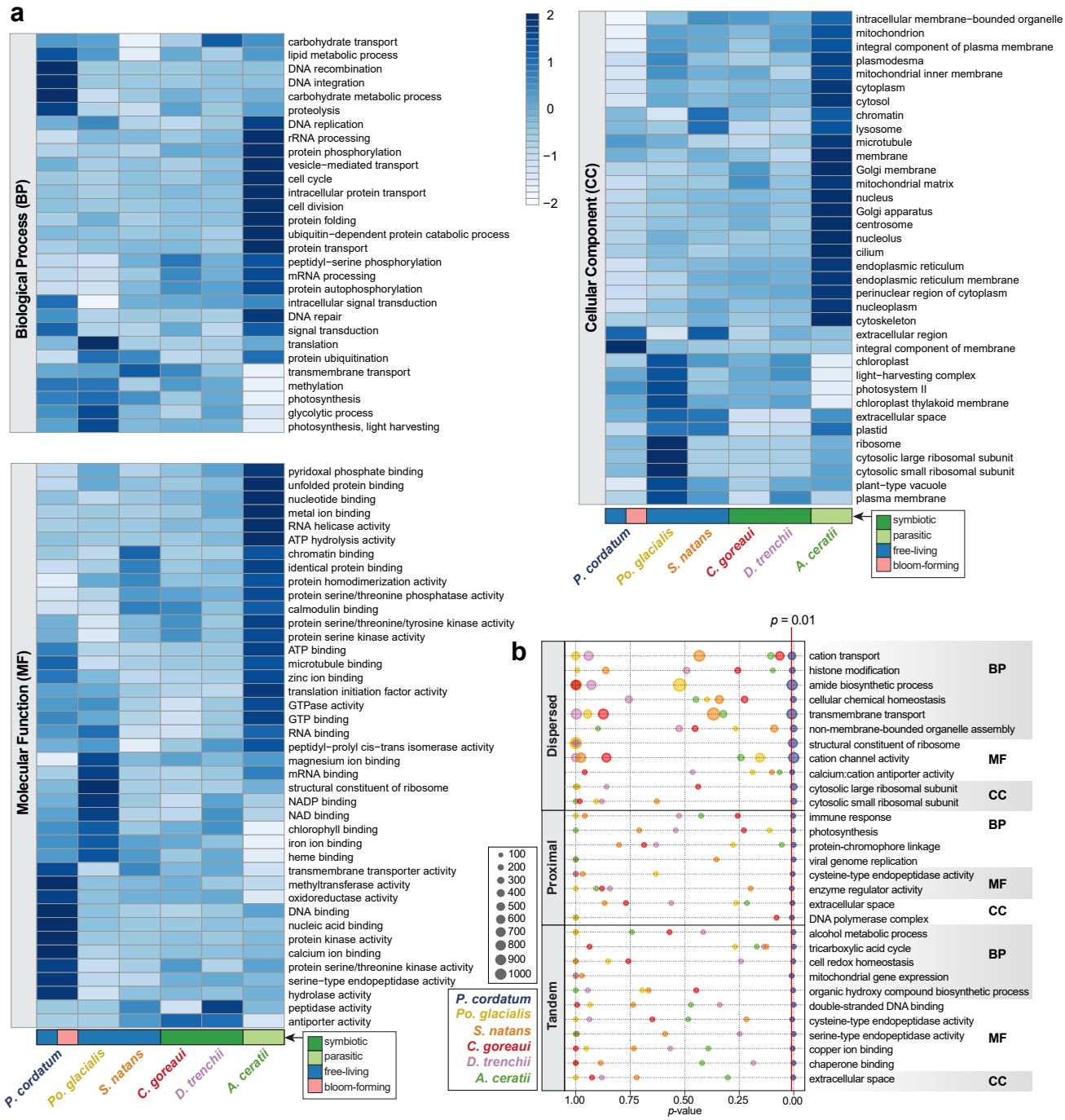


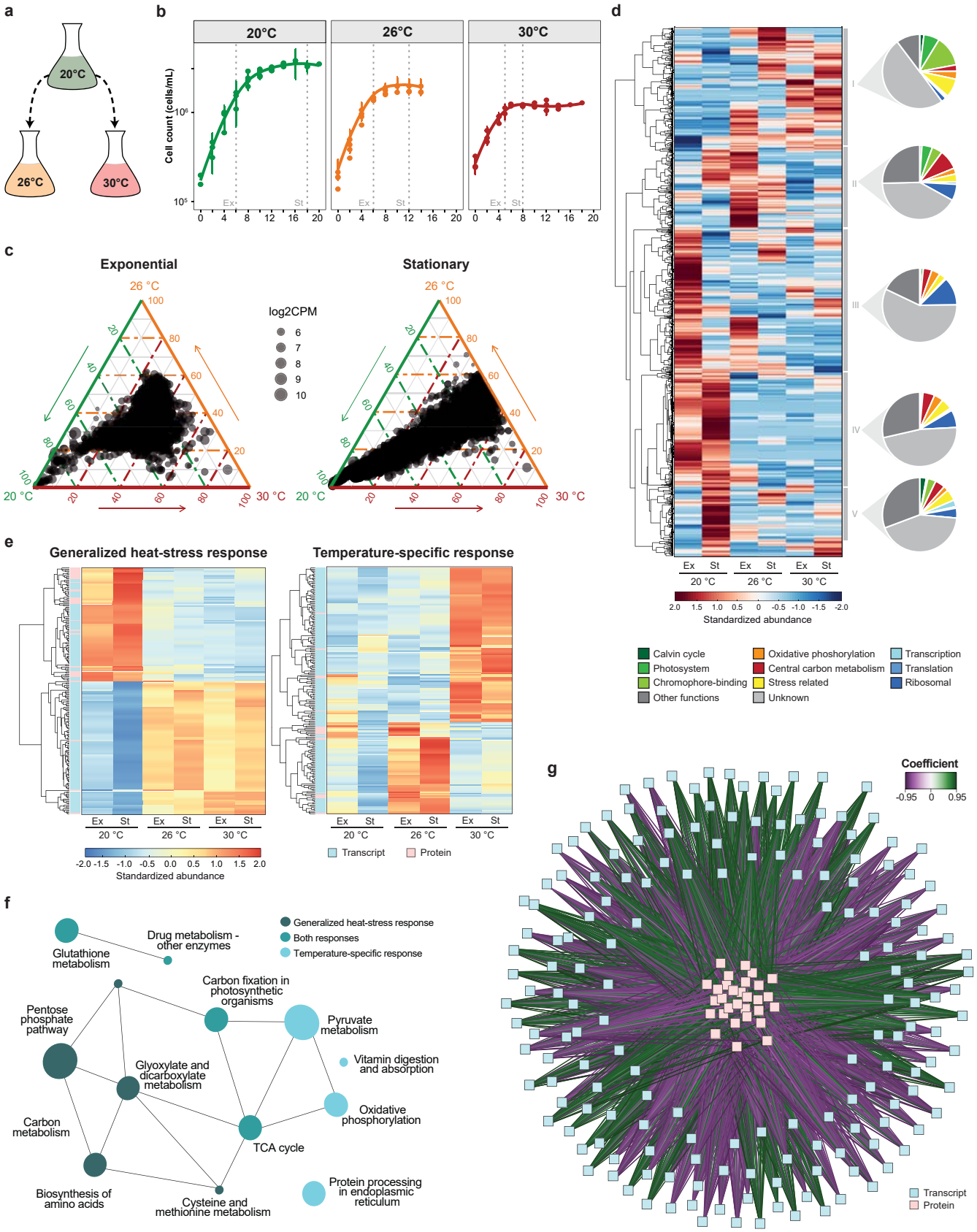
ED Figure 7: Transcript exon-intron plot Hsp70

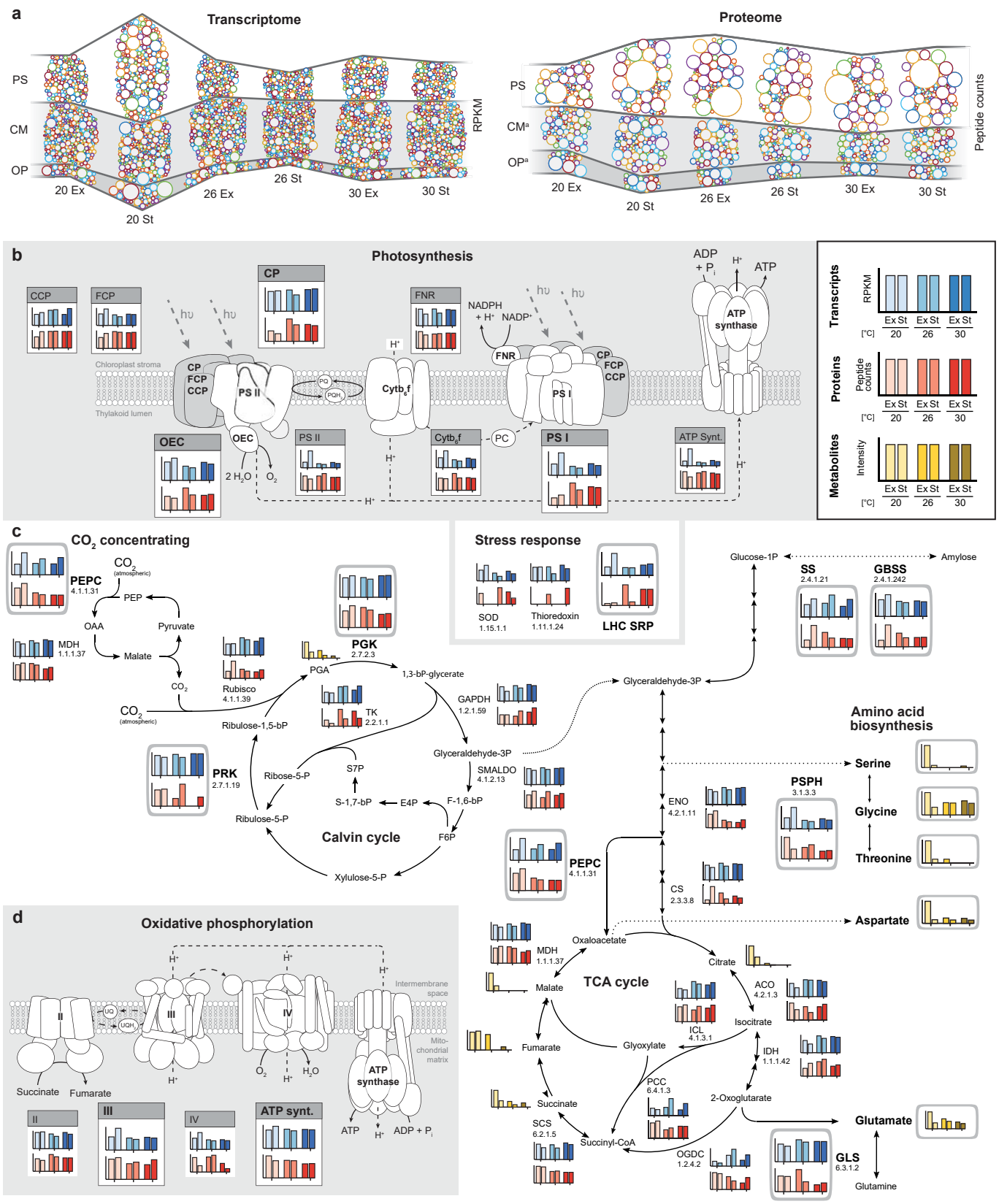


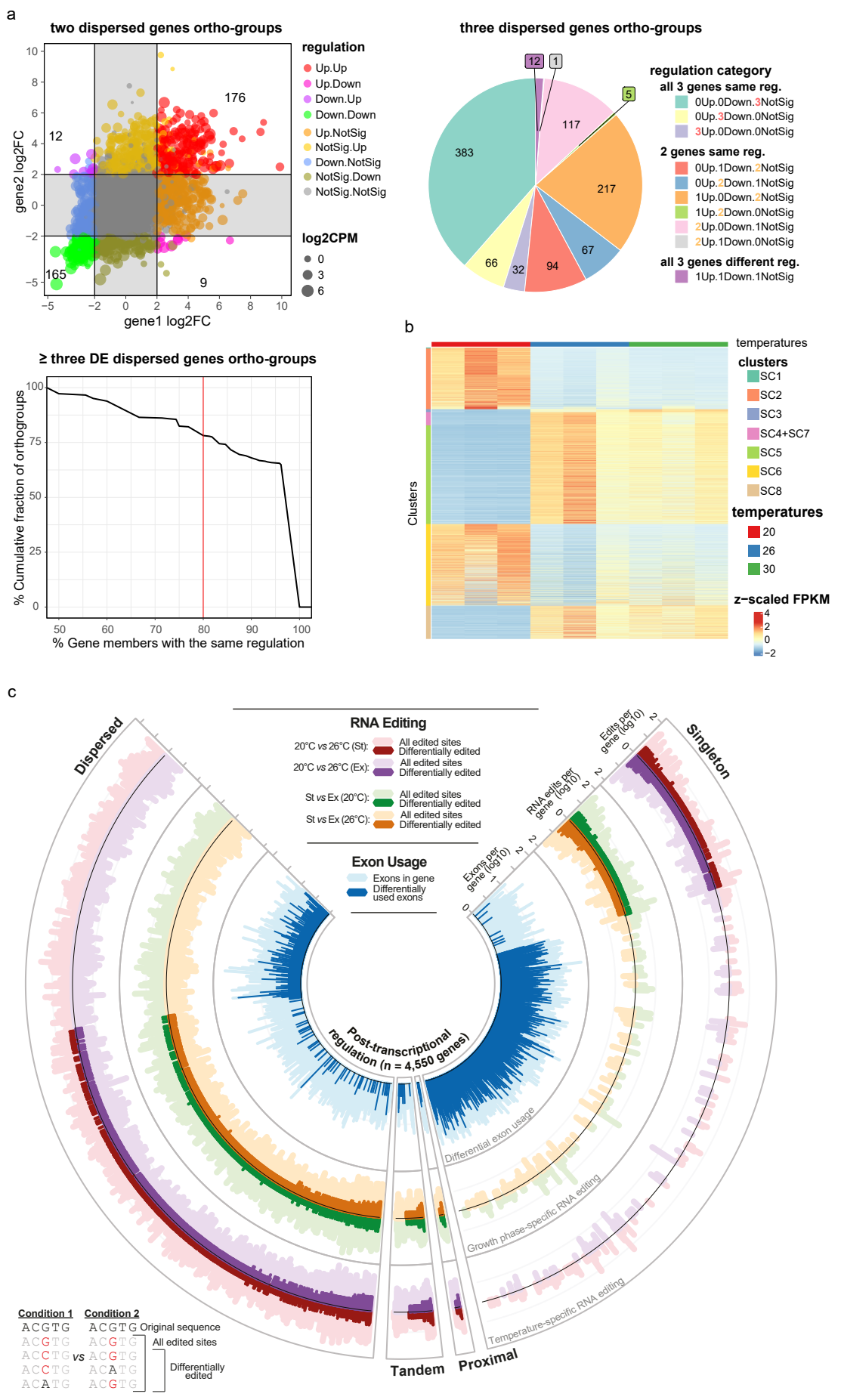
ED Figure 8: Transcript profiles of Hsp70 (a) and RuBisCo (b) gene models according to clustering in Figure 6a and 6c.

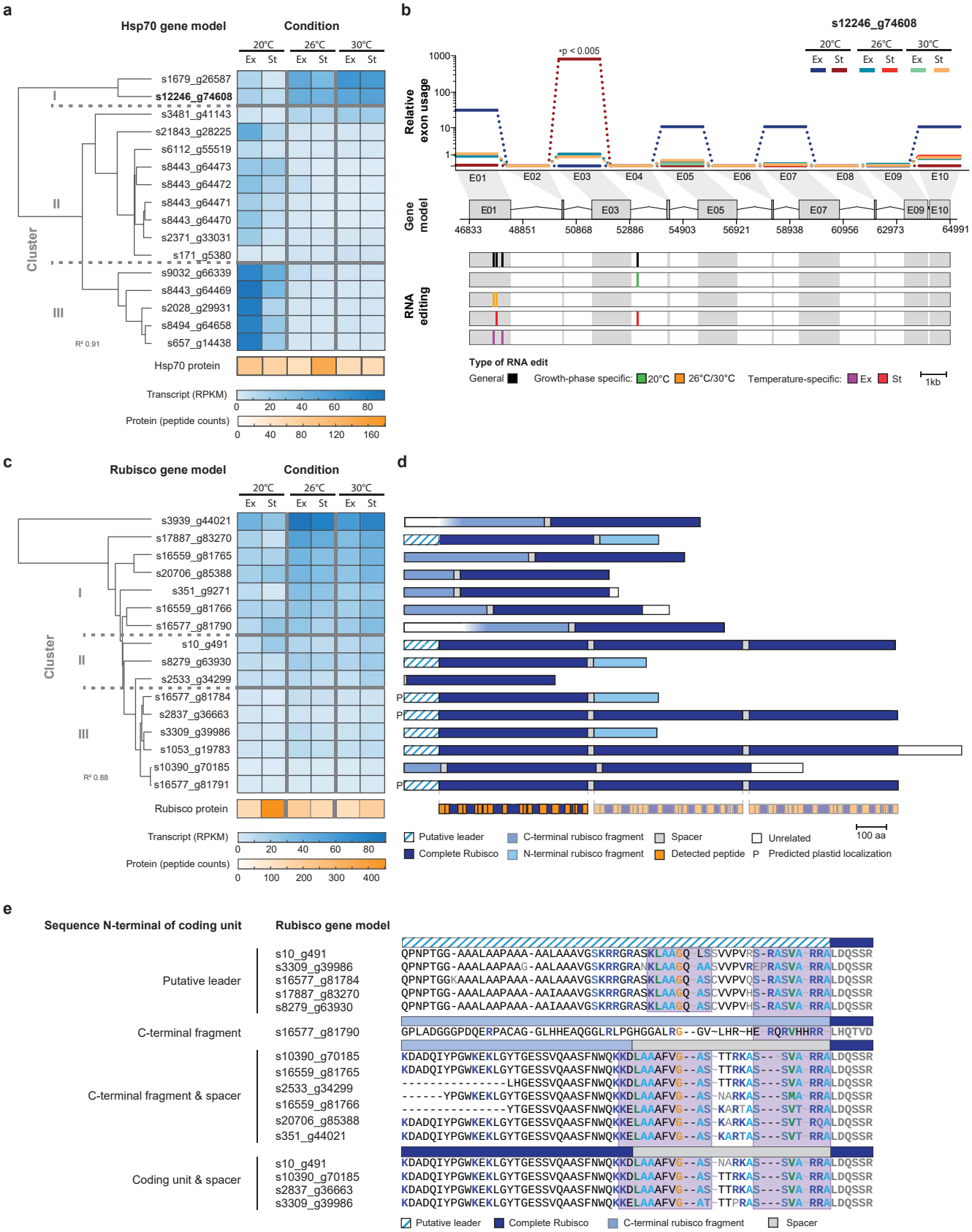

a
simple repeats
long interspersed nuclear elements (LINEs), 2.8%
others, 0.1%
long terminal repeats (LTRs), 2.4%
7.8%
rolling circles (helitrons), 1.2%
non-repeats 39.0%
DNA, 0.7%
unknown repeats 46.0%
c
whole-genome sequences
density
%GC
d
protein-coding sequences
density
%GC
A. ceratii
C. goreaui SCF055
D. trenchii CCMP2556
Po. glacialis CCMP1383
P. cordatum CCMP1329
S. natans CCMP2548
b
symbiotic
parasitic
free-living
bloom-forming
Cladocopium goreaui SCF055
Cladocopium C15
Cladocopium C92
Fugacium kawagutii
Breviolum minutum
Durusdinium trenchii CCMP2556
Durusdinium trenchii NIES-2907
Symbiodinium microadriaticum CCMP2467
Symbiodinium microadriaticum 04-503SCI.03
Symbiodinium microadriaticum CassKB8
Symbiodinium necroappetens CCMP2469
Symbiodinium linucheae CCMP2456
Symbiodinium tridacnidorum Sh18
Symbiodinium tridacnidorum CCMP2592
Symbiodinium pilosum CCMP2461
Symbiodinium natans CCMP2548
Polarella glacialis CCMP1383
Polarella glacialis CCMP2088
Suessiales
Dinophysis acuminata DAEP01
Alexandrium tamarense CCMP1771
Gambierdiscus australes
Dinophysiales
Gonyaulacales
Heterocapsa rotundata
Peridinium aciculiferum
Peridiniales
P. cordatum CCMP1329 MMETSP
P. cordatum CCMP1329 (this study)
Prorocentrales
Gymnodinium catenatum GC744
Gymnodiniales
Noctiluca scintillans
Amphidinium gibbosum
Noctilucales
Karenia brevis SP1
Gymnodiniales
Amoebophrya ceratii AT5.2
Syndiniales
Oxyrrhis marina
Oxyrrhinales
0.50 substitutions per site
e
haploid genome size
no. protein-coding genes
intron length
repeat type
duplication type
C. goreaui
D. trenchii
S. natans
Po. glacialis
P. cordatum
A. ceratii
1.30 Gbp
45,322
1.05 Gbp
55,799
0.74 Gbp
35,270
1.48 Gbp
~29,116
4.75 Gbp
85,849
0.12 Gbp
19,925
mean length (Kb)
% genome bases
% genes
IE⁻
IE⁺
All
Helitron
LINE
LTR
DNA
segmental
tandem
proximal
dispersed
singleton


a
Biological Process (BP)
carbohydrate transport
lipid metabolic process
DNA recombination
DNA integration
carbohydrate metabolic process
proteolysis
DNA replication
rRNA processing
protein phosphorylation
vesicle−mediated transport
cell cycle
intracellular protein transport
cell division
protein folding
ubiquitin−dependent protein catabolic process
protein transport
peptidyl−serine phosphorylation
mRNA processing
protein autophosphorylation
intracellular signal transduction
DNA repair
signal transduction
translation
protein ubiquitination
transmembrane transport
methylation
photosynthesis
glycolytic process
photosynthesis, light harvesting
2
1
0
−1
−2
Cellular Component (CC)
intracellular membrane−bounded organelle
mitochondrion
integral component of plasma membrane
plasmodesma
mitochondrial inner membrane
cytoplasm
cytosol
chromatin
lysosome
microtubule
membrane
Golgi membrane
mitochondrial matrix
nucleus
Golgi apparatus
centrosome
nucleolus
cilium
endoplasmic reticulum
endoplasmic reticulum membrane
perinuclear region of cytoplasm
nucleoplasm
cytoskeleton
extracellular region
integral component of membrane
chloroplast
light−harvesting complex
photosystem II
chloroplast thylakoid membrane
extracellular space
plastid
ribosome
cytosolic large ribosomal subunit
cytosolic small ribosomal subunit
plant−type vacuole
plasma membrane
symbiotic
parasitic
free-living
bloom-forming
P. cordatum
Po. glacialis
S. natans
C. goreaui
D. trenchii
A. ceratii
Molecular Function (MF)
pyridoxal phosphate binding
unfolded protein binding
nucleotide binding
metal ion binding
RNA helicase activity
ATP hydrolysis activity
chromatin binding
identical protein binding
protein homodimerization activity
protein serine/threonine phosphatase activity
calmodulin binding
protein serine/threonine/tyrosine kinase activity
protein serine kinase activity
ATP binding
microtubule binding
zinc ion binding
translation initiation factor activity
GTPase activity
GTP binding
RNA binding
peptidyl−prolyl cis−trans isomerase activity
magnesium ion binding
mRNA binding
structural constituent of ribosome
NADP binding
NAD binding
chlorophyll binding
iron ion binding
heme binding
transmembrane transporter activity
methyltransferase activity
oxidoreductase activity
DNA binding
nucleic acid binding
protein kinase activity
calcium ion binding
protein serine/threonine kinase activity
serine−type endopeptidase activity
hydrolase activity
peptidase activity
antiporter activity
symbiotic
parasitic
free-living
bloom-forming
P. cordatum
Po. glacialis
S. natans
C. goreaui
D. trenchii
A. ceratii
b
p = 0.01
Dispersed
Proximal
Tandem
cation transport
histone modification
amide biosynthetic process
cellular chemical homeostasis
transmembrane transport
non-membrane-bounded organelle assembly
BP
structural constituent of ribosome
cation channel activity
calcium:cation antiporter activity
MF
cytosolic large ribosomal subunit
cytosolic small ribosomal subunit
CC
immune response
photosynthesis
protein-chromophore linkage
viral genome replication
BP
cysteine-type endopeptidase activity
enzyme regulator activity
MF
extracellular space
DNA polymerase complex
CC
alcohol metabolic process
tricarboxylic acid cycle
cell redox homeostasis
mitochondrial gene expression
organic hydroxy compound biosynthetic process
BP
double-stranded DNA binding
cysteine-type endopeptidase activity
serine-type endopeptidase activity
copper ion binding
chaperone binding
MF
extracellular space
CC
100
200
300
400
500
600
700
800
900
1000
P. cordatum
Po. glacialis
S. natans
C. goreaui
D. trenchii
A. ceratii
1.00
0.75
0.50
0.25
0.00
p-value


a
20°C
26°C
30°C
b
20°C
26°C
30°C
Cell count (cells/mL)
Ex
St
c
Exponential
Stationary
26 °C
20 °C
30 °C
log2CPM
6
7
8
9
10
d
I
II
III
IV
V
Ex St Ex St Ex St
20 °C 26 °C 30 °C
Standardized abundance
Calvin cycle
Photosystem
Chromophore-binding
Other functions
Oxidative phoshorylation
Central carbon metabolism
Stress related
Unknown
Transcription
Translation
Ribosomal
e
Generalized heat-stress response
Temperature-specific response
Standardized abundance
Transcript
Protein
f
Glutathione metabolism
Drug metabolism - other enzymes
Generalized heat-stress response
Both responses
Temperature-specific response
Carbon fixation in photosynthetic organisms
Pentose phosphate pathway
Pyruvate metabolism
Vitamin digestion and absorption
Glyoxylate and dicarboxylate metabolism
Oxidative phosphorylation
Carbon metabolism
TCA cycle
Biosynthesis of amino acids
Cysteine and methionine metabolism
Protein processing in endoplasmic reticulum
g
Coefficient
-0.95 0 0.95
Transcript
Protein


a
Transcriptome
Proteome
PS
CM
OP
CMa
OPa
RPKM
Peptide counts
20 Ex
20 St
26 Ex
26 St
30 Ex
30 St
b
Photosynthesis
CCP
FCP
CP
FNR
Chloroplast stroma
Thylakoid lumen
PS II
Cytb6f
PS I
OEC
ATP synthase
ATP Synt.
PC
PQ
PQH2
NADPH + H+
NADP+
ADP + Pi
ATP
2 H2O
O2
Transcripts
Proteins
Metabolites
RPKM
Peptide counts
Intensity
Ex St
[°C]
c
CO2 concentrating
PEPC
4.1.1.31
CO2 (atmospheric)
PEP
OAA
Pyruvate
MDH
1.1.1.37
Malate
Rubisco
4.1.1.39
PGK
2.7.2.3
PGA
1,3-bP-glycerate
GAPDH
1.2.1.59
TK
2.2.1.1
Ribulose-1,5-bP
PRK
2.7.1.19
Ribose-5-P
S7P
Glyceraldehyde-3P
SMALDO
4.1.2.13
S-1,7-bP
E4P
F-1,6-bP
Ribulose-5-P
Calvin cycle
F6P
Xylulose-5-P
Stress response
SOD
1.15.1.1
Thioredoxin
1.11.1.24
LHC SRP
Glucose-1P
Amylose
SS
2.4.1.21
GBSS
2.4.1.242
Glyceraldehyde-3P
Amino acid biosynthesis
Serine
Glycine
Threonine
PSPH
3.1.3.3
ENO
4.2.1.11
PEPC
4.1.1.31
CS
2.3.3.8
Aspartate
Oxaloacetate
MDH
1.1.1.37
Malate
TCA cycle
Citrate
ACO
4.2.1.3
Isocitrate
ICL
4.1.3.1
Fumarate
Glyoxylate
IDH
1.1.1.42
PCC
6.4.1.3
2-Oxoglutarate
Succinate
SCS
6.2.1.5
Succinyl-CoA
OGDC
1.2.4.2
Glutamate
GLS
6.3.1.2
Glutamine
d
Oxidative phosphorylation
Intermembrane space
Mito-chondrial matrix
UQ
UQH2
Succinate
Fumarate
ATP synthase
ATP
ADP + Pi
II
III
IV
ATP synt.


a
two dispersed genes ortho-groups
gene2 log2FC
gene1 log2FC
176
12
165
9
regulation
Up.Up
Up.Down
Down.Up
Down.Down
Up.NotSig
NotSig.Up
Down.NotSig
NotSig.Down
NotSig.NotSig
log2CPM
0
3
6
three dispersed genes ortho-groups
12
1
5
117
383
217
67
66
32
94
regulation category
all 3 genes same reg.
0Up.0Down.3NotSig
0Up.3Down.0NotSig
3Up.0Down.0NotSig
2 genes same reg.
0Up.1Down.2NotSig
0Up.2Down.1NotSig
1Up.0Down.2NotSig
1Up.2Down.0NotSig
2Up.0Down.1NotSig
2Up.1Down.0NotSig
all 3 genes different reg.
1Up.1Down.1NotSig
≥ three DE dispersed genes ortho-groups
% Cumulative fraction of orthogroups
% Gene members with the same regulation
b
temperatures
clusters
SC1
SC2
SC3
SC4+SC7
SC5
SC6
SC8
Clusters
temperatures
20
26
30
z−scaled FPKM
4
2
0
−2
c
Dispersed
Singleton
RNA Editing
20°C vs 26°C (St): All edited sites Differentially edited
20°C vs 26°C (Ex): All edited sites Differentially edited
St vs Ex (20°C): All edited sites Differentially edited
St vs Ex (26°C): All edited sites Differentially edited
Exon Usage
Exons in gene
Differentially used exons
Exons per gene (log10)
RNA edits per gene (log10)
Edits per gene (log10)
Post-transcriptional regulation (n = 4,550 genes)
Differential exon usage
Growth phase-specific RNA editing
Temperature-specific RNA editing
Tandem
Proximal
Condition 1
Condition 2
ACGTG ACGTG Original sequence
ACGTG ACGTG All edited sites
ACCTG vs ACGTG
ACCTG ACATG Differentially edited
ACATG ACGTG


a
Hsp70 gene model
Condition
20°C 26°C 30°C
Ex St Ex St Ex St
s1679_g26587
s12246_g74608
s3481_g41143
s21843_g28225
s6112_g55519
s8443_g64473
s8443_g64472
s8443_g64471
s8443_g64470
s2371_g33031
s171_g5380
s9032_g66339
s8443_g64469
s2028_g29931
s8494_g64658
s657_g14438
Cluster
I
II
III
R² 0.91
Hsp70 protein
Transcript (RPKM)
0 20 40 60 80
Protein (peptide counts)
0 40 80 120 160
b
s12246_g74608
*p < 0.005
Relative exon usage
1000
100
10
1
E01 E02 E03 E04 E05 E06 E07 E08 E09 E10
Gene model
46833 48851 50868 52886 54903 56921 58938 60956 62973 64991
RNA editing
Type of RNA edit
General
Growth-phase specific: 20°C 26°C/30°C
Temperature-specific: Ex St
1kb
c
Rubisco gene model
Condition
20°C 26°C 30°C
Ex St Ex St Ex St
s3939_g44021
s17887_g83270
s16559_g81765
s20706_g85388
s351_g9271
s16559_g81766
s16577_g81790
s10_g491
s8279_g63930
s2533_g34299
s16577_g81784
s2837_g36663
s3309_g39986
s1053_g19783
s10390_g70185
s16577_g81791
Cluster
I
II
III
R² 0.88
Rubisco protein
Transcript (RPKM)
0 20 40 60 80
Protein (peptide counts)
0 100 200 300 400
d
P
Putative leader
Complete Rubisco
C-terminal rubisco fragment
N-terminal rubisco fragment
Spacer
Detected peptide
Unrelated
Predicted plastid localization
100 aa
e
Sequence N-terminal of coding unit
Rubisco gene model
Putative leader
s10_g491
s3309_g39986
s16577_g81784
s17887_g83270
s8279_g63930
QPNPTGG-AAALAAPAAA-AALAAAVGSKRRGRASKLAAGQ-LSSVVPVRS-RASVA-RRALDQSSR
QPNPTGG-AAALAAPAAG-AALAAAVGSKRRGRANKLAAGQ-AASVVPVREPRASVA-RRALDQSSR
QPNPTGGKAAALAAPAAA-AALAAAVGSKRRGRASKLAAGQ-ASCVVPVQS-RASVA-RRALDQSSR
QPNPTGG-AAALAAPAAA-AAIAAAVGSKRRGRASKLAAGQ-ASCVVPVHS-RASVA-RRALDQSSR
QPNPTGG-AAALAAPAAA-AAIAAAVGSKRRGRASKLAAGQ-ASCVVPVHS-RASVA-RRALDQSSR
C-terminal fragment
s16577_g81790
GPLADGGGPDQERPACAG-GLHHEAQGGLRLPGHGGALRG--GV~LHR~HE-RQRVHHRR-LHQTVD
C-terminal fragment & spacer
s10390_g70185
s16559_g81765
s2533_g34299
s16559_g81766
s20706_g85388
s351_g44021
KDADQIYPGWKEKLGYTGESSVQAASFNWQKKDLAAAFVG--AS~TTRKAS---SVA-RRALDQSSR
KDADQIYPGWKEKLGYTGESSVQAASFNWQKKDLAAAFVG--AS~TTRKAS---SVA-RRALDQSSR
----------LHGESSVQAASFNWQKKDLAAAFVG--AS~TTRKAS---SVA-RRALDQSSR
------YPGWKEKLGYTGESSVQAASFNWQKKDLAAAFVG--AS~NARKAS---SMA-RRALDQSSR
----------YTGESSVQAASFNWQKKELAAAFVG--AS~KARTAS---SVT-RRALDQSSR
KDADQIYPGWKEKLGYTGESSVQAASFNWQKKELAAAFVG--AS~KARKAS---SVT-RQALDQSSR
KDADQIYPGWKEKLGYTGESSVQAASFNWQKKELAAAFVG--AS~KARTAS---SVT-RRALDQSSR
Coding unit & spacer
s10_g491
s10390_g70185
s2837_g36663
s3309_g39986
KDADQIYPGWKEKLGYTGESSVQAASFNWQKKDLAAAFVG--AS~NARKAS---SVA-RRALDQSSR
KDADQIYPGWKEKLGYTGESSVQAASFNWQKKDLAAAFVG--AS~TTRKAS---SVA-RRALDQSSR
KDADQIYPGWKEKLGYTGESSVQAASFNWQKKDLAAAFVG--AS~TTRKAS---SVA-RRALDQSSR
KDADQIYPGWKEKLGYTGESSVQAASFNWQKKELAAAFVG--AT~TTPRAS---SVA-RRALDQSSR
Putative leader
Complete Rubisco
C-terminal rubisco fragment
Spacer