Diatom acclimation to elevated CO₂ via cAMP signalling and coordinated gene expression

Gwenn M. M. Hennon[1]★, Justin Ashworth[2], Ryan D. Groussman[1], Chris Berthiaume[1], Rhonda L. Morales[1], Nitin S. Baliga[2], Mónica V. Orellana[2,3] and E. V. Armbrust[1]★

Diatoms are responsible for ∼40% of marine primary productivity[1], fuelling the oceanic carbon cycle and contributing to natural carbon sequestration in the deep ocean[2]. Diatoms rely on energetically expensive carbon concentrating mechanisms (CCMs) to fix carbon efficiently at modern levels of CO₂ (refs 3–5). How diatoms may respond over the short and long term to rising atmospheric CO₂ remains an open question. Here we use nitrate-limited chemostats to show that the model diatom *Thalassiosira pseudonana* rapidly responds to increasing CO₂ by differentially expressing gene clusters that regulate transcription and chromosome folding, and subsequently reduces transcription of photosynthesis and respiration gene clusters under steady-state elevated CO₂. These results suggest that exposure to elevated CO₂ first causes a shift in regulation, and then a metabolic rearrangement. Genes in one CO₂-responsive cluster included CCM and photorespiration genes that share a putative cAMP-responsive *cis*-regulatory sequence, implying these genes are co-regulated in response to CO₂, with cAMP as an intermediate messenger. We verified cAMP-induced downregulation of CCM gene δ-CA3 in nutrient-replete diatom cultures by inhibiting the hydrolysis of cAMP. These results indicate an important role for cAMP in downregulating CCM and photorespiration genes under elevated CO₂ and provide insights into mechanisms of diatom acclimation in response to climate change.

Burning fossil fuels and land-use change have accelerated CO₂ emissions to the atmosphere by a factor ∼100 above natural levels[6]. About a third of anthropogenic emissions have been absorbed by the oceans[7,8], increasing dissolved CO₂ and reducing pH (ref. 9). Despite these changes, CO₂ concentrations in surface waters remain below half-saturation for most forms of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco)[3], the central enzyme used to fix carbon. Consequently, marine phytoplankton, including diatoms, rely on carbon concentrating mechanisms (CCMs) to ensure adequate delivery of CO₂ to the Rubisco active site, minimizing the competitive fixation of oxygen[3–5]. The required bicarbonate transporters and carbonic anhydrases of these CCMs concentrate CO₂ against a gradient, which is energetically costly[10]. Downregulation of CCMs as part of acclimation to elevated CO₂ should result in energy savings to the diatom cell and metabolic rearrangement. Here we use nitrate-limited chemostats to simulate *in situ* nutrient limitation[11] while precisely controlling cell biomass and CO₂ (ref. 12), allowing us to identify potential signalling pathways triggered either by an abrupt transition to increased CO₂, as might occur during coastal upwelling[13], or at steady-state exposure to elevated CO₂, including 800 μatm predicted for 2100 (ref. 14; Fig. 1a,b).

Metabolic and regulatory genes were differentially impacted by changes in CO₂ (Fig. 1c). The initial response to an abrupt increase in CO₂ included upregulation of genes required for transcriptional regulation and kinase activity (Fig. 1c), potentially reflecting CO₂-specific signalling mechanisms. Once steady-state conditions with elevated CO₂ were attained, genes required for energy-producing metabolic pathways, including oxidative phosphorylation, TCA cycle and photosynthesis, were preferentially downregulated, as were genes involved in transcriptional regulation, ion transport, kinase activity and protein degradation (Fig. 1c), suggesting a general reduction in metabolism under high CO₂. This corroborates our previous physiological measurements of decreased respiration and photosynthetic oxygen production under elevated CO₂ and nitrate limitation[12]. A significant upregulation of components of the ribosome under steady-state high CO₂ (Fig. 1c) may reflect either an increased capacity for protein synthesis or a high turnover of ribosomal proteins, whereas the downregulation of the same gene category in transition suggests disparate modes of regulation between transition and steady-state exposure to elevated CO₂.

To identify potential linkages between an initial response and eventual metabolic rearrangements under steady-state conditions, we identified groups of co-expressed genes based on unsupervised hierarchical clustering across a combined data set of 98 RNA sequencing and microarray data sets from a range of conditions[15–20], including our chemostat experiments. Of the 400 clusters of co-expressed genes, 160 exhibited significant correlation to CO₂ partial pressure during transition or steady-state exposure (Spearman false discovery rate (FDR) < 0.0001; Supplementary Table 1). The unsupervised clustering analysis yielded biologically meaningful insights, grouping together functionally related genes associated with energy metabolism, including light harvesting, photosynthesis, and respiration (Fig. 2a,b). These energy metabolism gene clusters were differentially expressed only after acclimation to steady-state elevated CO₂, along with other clusters containing genes implicated in cell signalling (Fig. 2c,d), suggesting a suite of gene clusters are probably involved in diatom metabolic rearrangement. Expression of a gene cluster associated with regulation, structural maintenance of chromosomes (SMC), and transcription was correlated with CO₂ level during transition and steady-state exposure (Fig. 2e,f). This cluster of regulatory genes, not previously connected to the diatom CO₂ response, could be involved in diatom cell regulation vital to sensing and acclimating to rising CO₂.

High external CO₂ is expected to reduce the need for the CCM as well as photorespiration, owing to the competitive inhibition of Rubisco's oxygenase activity by CO₂ (ref. 5). Interestingly, a single cluster contained both putative CCM and photorespiration

[1]School of Oceanography, University of Washington, Seattle, Washington 98105, USA. [2]Institute for Systems Biology, Seattle, Washington 98109, USA. [3]Polar Science Center, University of Washington, Seattle, Washington 98105, USA. *e-mail: gwennm@uw.edu; armbrust@uw.edu
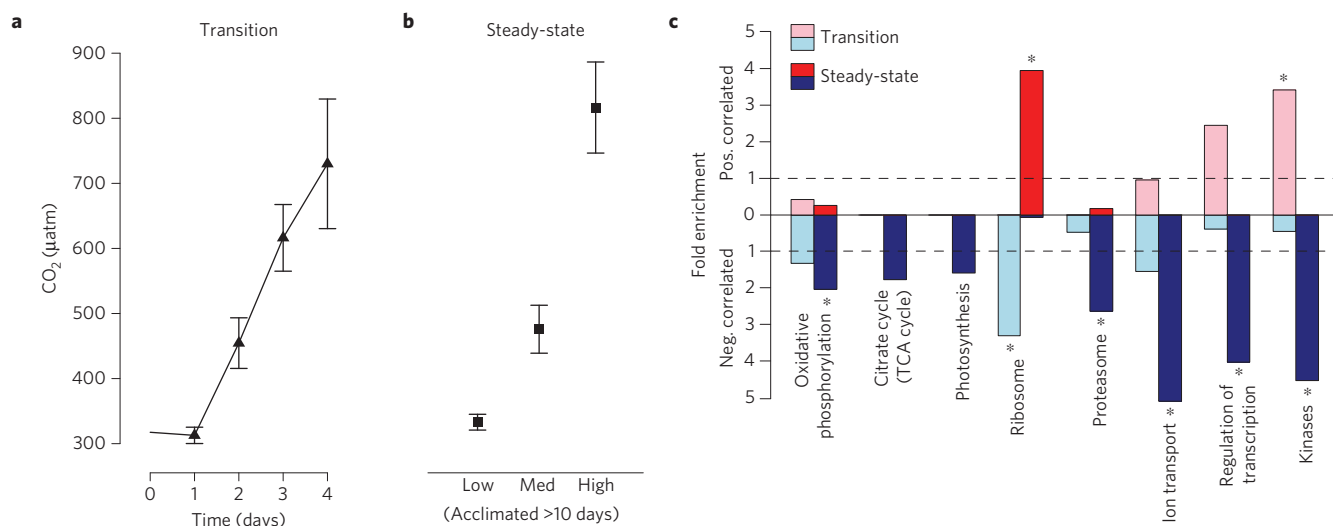
Figure 1 | Gene set enrichment in transition and steady-state nitrate-limited cultures. a,b, *Thalassiosira pseudonana* exposed to transition from low to high CO_2 over four days (**a**) or steady-state acclimation to low, medium or high CO_2 (**b**). Error bars indicate 1 s.d. from the mean, $n = 4$. **c,** Genes significantly correlated (positively or negatively) with CO_2 (Spearman rank correlation $p < 0.05$) formed gene sets categorized by KEGG pathway and GO term. The ratios of gene number to expected value (fold enrichment) in each category are plotted on the y-axis; dashed line indicates expected value and asterisk indicates significant enrichment in positively or negatively correlated gene sets (hypergeometric test $p < 0.05$).

genes and was negatively correlated to CO_2 in both transition and steady-state experiments (Supplementary Table 1). This cluster further grouped into two distinct sub-clusters that exhibited different expression patterns across the range of CO_2 concentrations examined in our experiments (Fig. 2e,f and Supplementary Fig. 1). One sub-cluster included candidates for participation in the CCM, such as the plastid-localized delta-carbonic anhydrase (δ-CA3; ref. 21) and plastid-targeted transporters. The other sub-cluster included mitochondria-localized photorespiration genes necessary to recycle by-products of the oxygenase activity of Rubisco. The zeta-carbonic anhydrase ζ-CA1 (ref. 21), also grouped with the photorespiration sub-cluster, is the only enzyme known to use cadmium as a cofactor[22] and is localized to the outer membrane space[21].

Given the strong correlation across our data set of these sub-clusters of genes from separate pathways, we hypothesized they were regulated by a common mechanism. Genes in the CCM/photorespiration cluster all share a short upstream sequence (Supplementary Fig. 1b), TGACGT, recently identified as a *cis*-regulatory sequence that downregulates expression of the CO_2-responsive β-CA in the distantly related diatom *Phaeodactylum tricornutum*[23]. Furthermore, two transcription factors, one in each sub-cluster, exhibit close homology (Supplementary Fig. 2) and are classified in the same phylogenetic group[24] as a cAMP-activated transcription factor in *P. tricornutum* shown to bind the *cis*-regulatory element and to contribute to downregulation of the β-CA gene in response to increased CO_2 (ref. 23). These lines of evidence suggest that the two transcription factors identified in the CCM/photorespiration cluster may function to downregulate transcription of their own cluster by binding to the common upstream *cis*-regulatory region. This further implies that these transcription factors repress their own transcription, creating a negative feedback on the repression of this cluster, a feature common in biological systems[25].

We tested the proposed mechanism of cAMP-regulated gene expression with two indicator genes—carbonic anhydrase δ-CA3 and a putative transporter (pID 262258)—that responded to CO_2 in both transition and steady-state experiments with a large fold change (Supplementary Table 2). We grew *T. pseudonana* in nutrient-replete batch cultures and manipulated the CO_2 and

the cellular level of cAMP by inhibiting its degradation with the phosphodiesterase inhibitor 3-isobutyl-1-methylxanthine (IBMX). The gene encoding δ-CA3 is a member of the CCM gene cluster and contains the upstream TGACGT motif. As predicted, transcript abundance for this gene was significantly ($p < 0.05$, two-way analysis of variance (ANOVA)) reduced by both high CO_2 and IBMX addition compared to low CO_2 (Fig. 3). The gene encoding the putative transporter lacks the cAMP-responsive motif. As predicted, transcript abundance for this gene was increased by elevated CO_2, but was not significantly altered by the addition of IBMX ($p > 0.05$, *t*-test) (Fig. 3). These data suggest that gene expression patterns in response to CO_2 remain consistent independent of nitrate availability and that cAMP selectively downregulates CO_2-responsive genes possessing the TGACGT motif.

Consistent with our data, mounting evidence[23,26] suggests that diatoms sense a change in external CO_2 concentrations, in part, through cAMP signalling. In other organisms[26,27] the activity of specific adenylyl cyclases are regulated by CO_2 or bicarbonate, where increased CO_2 stimulates cyclase activity and results in enhanced production of cAMP (ref. 27). This evidence suggests that increasing CO_2 concentrations promote a cyclase in *T. pseudonana* to increase production of cAMP, which in turn induces the CCM/photorespiration cluster transcription factors to bind to the TGACGT motif and repress transcription of CCM/photorespiration cluster genes, including their own transcription. The soluble cyclase (sol. CYCc, Fig. 4) and two membrane-bound cyclases (CYCc, Figs 2c,d and 4) are among the candidates for sensing CO_2 in *T. pseudonana*. This mechanism is supported by the downregulation of CCM/photorespiration cluster genes, including the transcription factors, by an increase in CO_2 and the downregulation of δ-CA3 by addition of IBMX (Supplementary Table 2 and Fig. 3).

Negative feedback mechanisms should result in a decreased repression of the CCM/photorespiration cluster genes over time unless changes in cAMP signalling or transcription factor regulation occur. We observed that cAMP metabolism genes were upregulated at steady-state elevated CO_2 (Fig. 4), including a phosphodiesterase, which degrades cAMP by hydrolysis, and two membrane-bound cyclases (CYCc, Fig. 2c,d) that enhance production of cAMP. This indicates a change in cAMP metabolism at steady-state elevated CO_2, potentially altering
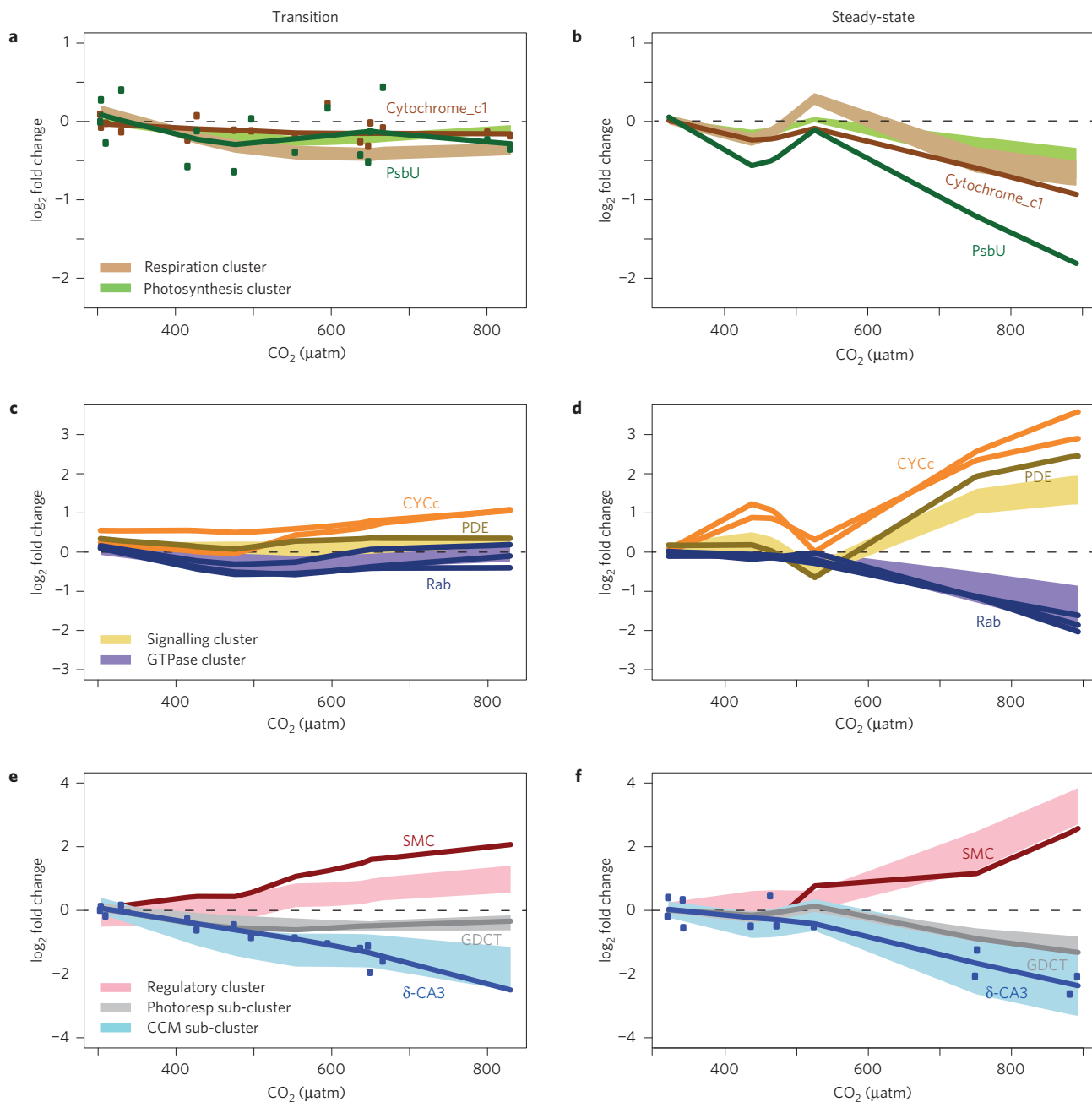
Figure 2 | Clusters of co-expressed genes versus CO_2 in transition and steady-state experiments. a–f, Fold change (mean ± 2 s.e.m.) of clusters of co-expressed genes normalized to expression under low CO_2. Gene clusters involved in energy-producing metabolic pathways (**a,b**), sensing and intracellular signal propagation (**c,d**) and a regulatory cluster and photorespiration and CCM sub-clusters (**e,f**). Expression of selected individual genes is indicated with solid lines representing smoothed mean expression and points representing single transcriptome measurements relative to low CO_2. Abbreviations for genes are: Cytochrome_c1, Cytochrome C1 oxidase; PsbU, Photosystem II subunit; CYCc, adenylate/guanylate cyclase; PDE, phosphodiesterase; Rab, Rab-like small GTPase; SMC, structural maintenance of chromosomes protein; CA, carbonic anhydrase; GDCT, glycine decarboxylase T protein.

how signals are propagated in cAMP-responsive pathways, and perhaps explaining why transcript abundance for the δ-CA3 gene is not further reduced under high-CO_2 conditions in the presence of IBMX (Fig. 3). Likewise, the CCM sub-cluster transcription factor was no longer significantly repressed at steady-state elevated CO_2 (Fig. 4), potentially allowing it to attain sufficient abundance to maintain repression of the CCM and photorespiration pathway.

Placing the CCM and photorespiration gene clusters in the context of the diatom cell, we integrate predicted protein

localization, gene transcription patterns, and metabolic pathways at steady-state elevated CO_2 (Fig. 4) to gain insight into potential mechanisms of the diatom CO_2 response. A functioning diatom CCM must transport bicarbonate into the plastid to the stroma, where it is converted to CO_2 by a carbonic anhydrase[10]. The CCM sub-cluster contains genes encoding plastid-localized membrane proteins, including bestrophin-like proteins, which are homologous to a family of gated anion-selective channels[28] permeable to bicarbonate[29] (Supplementary Table 2). These plastid-targeted proteins provide a feasible mechanism for bicarbonate transport to
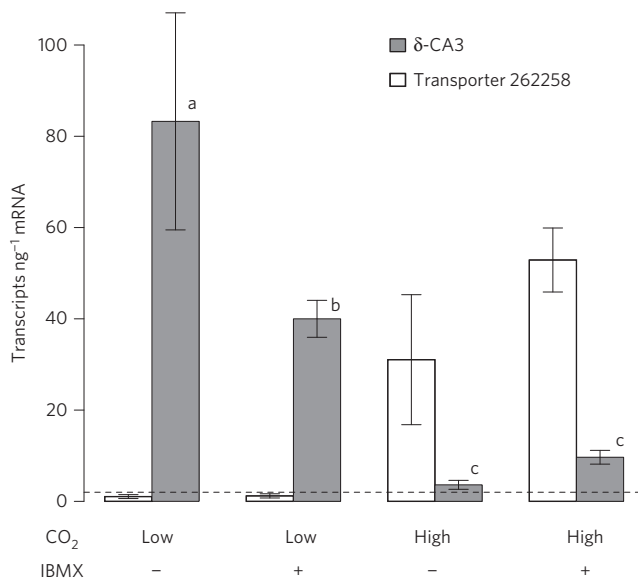
763

Figure 3 | cAMP and CO_2 dependence of gene expression in nutrient-replete cultures. Transcript abundance normalized to messenger RNA (mean ± 1.96 s.e.m.: 95% CI, $n = 3$) of genes encoding carbonic anhydrase δ-CA3 (grey) and a putative transport protein pID 262258 (white), in nutrient-replete cultures. *T. pseudonana* cells were exposed to low (240 ± 15 μatm) or high (857 ± 54 μatm) CO_2, with or without 1.0 mM of the phosphodiesterase inhibitor 3-isobutyl-1-methylxanthine (IBMX) used to elevate cellular cAMP concentrations. Dashed line indicates the quantification limit of transcript abundance by RT-qPCR and letters indicate significant differences among treatments for δ-CA3 gene expression ($p < 0.05$, two-way ANOVA and Tukey HSD).

the stroma (Fig. 4), which has not yet been elucidated in diatoms. Once inside the plastid, bicarbonate can be converted to CO_2 by the stroma-localized[21], constitutively expressed α-CA1 (Fig. 4). Interestingly, this carbonic anhydrase is not downregulated by CO_2, suggesting that, unlike *P. tricornutum*[23], expression levels of a stromal-localized CA are not a point of CCM control in *T. pseudonana*. Another essential feature of efficient CCMs is to prevent the passive diffusion of enhanced CO_2 concentrations out of the cell in general, or the plastid in particular[10]. The δ-CA3, localized to the plastid membrane space[21], and the ζ-CA1, localized to the outer membrane space[21], are well placed to serve this role by converting CO_2 into bicarbonate in the periplastid and periplasmic compartments, where it could be selectively transported back into the cytoplasm or stroma. That these genes are rapidly and sustainably downregulated under elevated CO_2 suggests that preventing diffusive loss of CO_2 is tightly regulated as a point of control for the *T. pseudonana* CCM, in strong agreement with physiology-based model predictions for an efficient diatom CCM (ref. 10).

Another feature of an efficient diatom CCM is active transport of bicarbonate from the external environment to the cell interior[10]; the SLC4-2 bicarbonate transporter is the most likely candidate, as it is distantly related to a transporter demonstrated to serve this role in the diatom *P. tricornutum*[30], and the gene encoding this transporter was significantly downregulated at steady-state high CO_2 (Fig. 4 and Supplementary Table 2). Yet the SLC4-2 gene does not group with the CCM/photorespiration genes, nor share the same upstream motif (Supplementary Table 1). Likewise, the plastid-localized phosphoglycolate phosphatase (PGP) gene, essential to the photorespiration pathway, was downregulated at steady-state (Fig. 4), but does not cluster with the mitochondrial photorespiration genes nor share the upstream region. cAMP-signal-induced regulation cannot account for all the changes
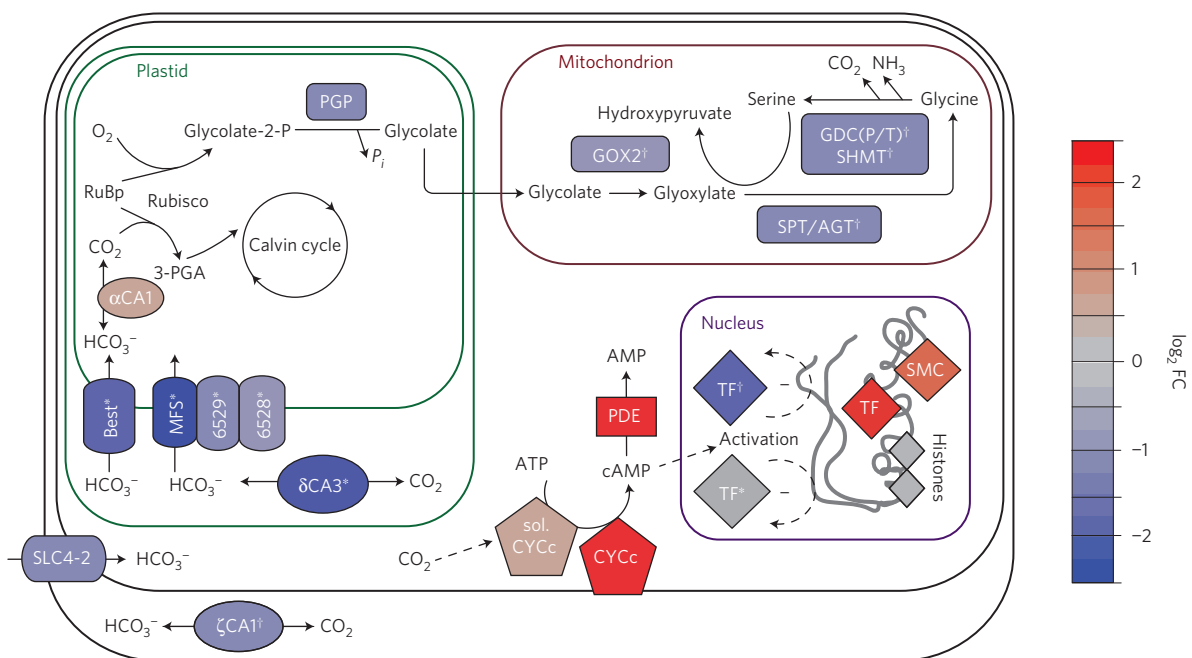

Figure 4 | Model of cell signalling and metabolite fluxes in *T. pseudonana* after acclimation to high CO_2. Gene expression changes relative to low CO_2 in high-CO_2 steady-state experiments indicated by the heat map (log$_2$ fold change) overlaid on a model of diatom CO_2-responsive signalling and metabolic pathways. Membership in the CCM sub-cluster and photorespiration cluster indicated by * and † respectively. Solid arrows indicate reactions and dashed arrows indicate regulatory relationships. Abbreviations: SPT/AGT, serine-pyruvate/aspartate aminotransferase; TF, transcription factor; SHMT, serine-glycine hydroxymethyltransferase; GDC(P/T), glycine decarboxylase (P/T) protein; GOX, glycolate oxidase; MFS, major facilitator superfamily; Best, bestrophin superfamily transporter; PGP, phosphoglycolate phosphatase; SLC4, solute carrier family 4 bicarbonate transporter; 3-PGA, 3-phosphoglyceric acid; RuBP, ribulose-1,5-bisphosphate; others as in Fig. 2.

in gene expression, because these and other known CCM and photorespiration genes were not downregulated in lock-step (Fig. 4) nor do they all contain the same upstream region. Transcription of a SMC gene, co-expressed transcription factors, and histone proteins were significantly affected by an abrupt increase in CO_2 (Supplementary Table 2). These proteins function to modify chromosome structure to alter gene expression (Fig. 4), and may explain shifts in CO_2-responsive genes not regulated by the cAMP-induced transcription factors.

The metabolic rearrangement we observed under steady-state elevated CO_2 and nitrate limitation resulted from downregulation of the energy-consuming processes of CCM and photorespiration because cells were unable to increase growth rate owing to nitrate limitation, a common limiting factor in the oceans[11]. Diatoms are also limited by other factors, such as iron or light, in vast regions of the ocean[11], and may exhibit different metabolic rearrangements or growth rate enhancement in response to elevated CO_2 under these other conditions. Our analysis discovered genes that exhibit CO_2 responses regardless of nitrogen status, and provide insights into the mechanisms of CO_2 sensing, signalling and metabolic rearrangement necessary to predict how diatoms will acclimate in future oceans. Our approach, examining the behaviour of co-expressed gene clusters in both transition and after steady-state acclimation, is a powerful tool for beginning to untangle the mechanisms of sensing and responding to rising CO_2. Many of the CO_2-responsive genes identified by our clustering analysis were not previously connected to *T. pseudonana* CCM or CCM-regulation, allowing us to investigate new genes and mechanisms involved in regulating diatom physiology and biogeochemical cycles in the face of climate change.

Methods

Methods and any associated references are available in the online version of the paper.

References

1. Nelson, D. M., Treguer, P., Brzezinski, M. A., Leynaert, A. & Queguiner, B. Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Glob. Biogeochem. Cycles* **9**, 359–372 (1995).
2. Ducklow, H. W., Steinberg, D. K. & Buesseler, K. O. Upper ocean carbon export and the biological pump. *Oceanography* **14**, 50–58 (2001).
3. Badger, M. R. *et al.* The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO_2-concentrating mechanisms in algae. *Can. J. Bot.* **76**, 1052–1071 (1998).
4. Giordano, M., Beardall, J. & Raven, J. A. CO_2 concentrating mechanisms in algae: Mechanisms, environmental modulation, and evolution. *Annu. Rev. Plant Biol.* **56**, 99–131 (2005).
5. Reinfelder, J. R. Carbon concentrating mechanisms in eukaryotic marine phytoplankton. *Annu. Rev. Mar. Sci.* **3**, 291–315 (2011).
6. Berner, R. A. The long-term carbon cycle, fossil fuels and atmospheric composition. *Nature* **426**, 323–326 (2003).
7. Sabine, C. L. *et al.* The oceanic sink for anthropogenic CO_2. *Science* **305**, 367–371 (2004).
8. Le Quéré, C. *et al.* The global carbon budget 1959–2011. *Earth Syst. Sci. Data Discuss.* **5**, 165–185 (2012).
9. Caldeira, K. & Wickett, M. Anthropogenic carbon and ocean pH. *Nature* **425**, 365 (2003).
10. Hopkinson, B. M., Dupont, C. L., Allen, A. E. & Morel, F. M. M. Efficiency of the CO_2-concentrating mechanism of diatoms. *Proc. Natl Acad. Sci. USA* **108**, 3830–3837 (2011).
11. Moore, J. K., Doney, S. C., Glover, D. M. & Fung, I. Y. Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. *Deep-Sea Res. II* **49**, 463–507 (2002).
12. Hennon, G. M. M., Quay, P., Morales, R. L., Swanson, L. M. & Armbrust, E. V. Acclimation conditions modify physiological response of the diatom *Thalassiosira pseudonana* to elevated CO_2 concentrations in a nitrate-limited chemostat. *J. Phycol.* **253**, 243–253 (2014).
13. Feely, R. A., Sabine, C. L., Hernandez-Ayon, J. M., Ianson, D. & Hales, B. Evidence for upwelling of corrosive 'acidified' water onto the continental shelf. *Science* **320**, 1490–1492 (2008).
14. Ciais, P. *et al.* in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) 465–570 (IPCC, Cambridge Univ. Press, 2013).
15. Mock, T. *et al.* Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. *Proc. Natl Acad. Sci. USA* **105**, 1579–1584 (2008).
16. Carvalho, R. N., Bopp, S. K. & Lettieri, T. Transcriptomics responses in marine diatom *Thalassiosira pseudonana* exposed to the polycyclic aromatic hydrocarbon benzo[a]pyrene. *PLoS ONE* **6**, e26985 (2011).
17. Thamatrakoln, K., Korenovska, O., Niheu, A. K. & Bidle, K. D. Whole-genome expression analysis reveals a role for death-related genes in stress acclimation of the diatom *Thalassiosira pseudonana*. *Environ. Microbiol.* **14**, 67–81 (2012).
18. Shrestha, R. P. *et al.* Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. *BMC Genomics* **13**, 499 (2012).
19. Ashworth, J. *et al.* Genome-wide diel growth state transitions in the diatom *Thalassiosira pseudonana*. *Proc. Natl Acad. Sci. USA* **110**, 7518–7523 (2013).
20. Bender, S. J., Durkin, C. A., Berthiaume, C. T., Morales, R. L. & Armbrust, E. V. Transcriptional responses of three model diatoms to nitrate limitation of growth. *Front. Mar. Sci.* **1**, 3 (2014).
21. Samukawa, M., Shen, C., Hopkinson, B. M. & Matsuda, Y. Localization of putative carbonic anhydrases in the marine diatom, *Thalassiosira pseudonana*. *Photosynth. Res.* **121**, 235–249 (2014).
22. Xu, Y., Feng, L., Jeffrey, P. D., Shi, Y. & Morel, F. M. M. Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms. *Nature* **452**, 56–61 (2008).
23. Ohno, N. *et al.* CO_2-cAMP-responsive *cis*-elements targeted by a transcription factor with CREB/ATF-like basic zipper domain in the marine diatom *Phaeodactylum tricornutum*. *Plant Physiol.* **158**, 499–513 (2012).
24. Rayko, E., Maumus, F., Maheswari, U., Jabbari, K. & Bowler, C. Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol.* **188**, 52–66 (2010).
25. Krishna, S., Andersson, A. M. C., Semsey, S. & Sneppen, K. Structure and function of negative feedback loops at the interface of genetic and metabolic networks. *Nucleic Acids Res.* **34**, 2455–2462 (2006).
26. Matsuda, Y., Nakajima, K. & Tachibana, M. Recent progresses on the genetic basis of the regulation of CO_2 acquisition systems in response to CO_2 concentration. *Photosynth. Res.* **109**, 191–203 (2011).
27. Hammer, A., Hodgson, D. R. W. & Cann, M. J. Regulation of prokaryotic adenylyl cyclases by CO_2. *Biochem. J.* **396**, 215–218 (2006).
28. Dickson, V. K., Pedi, L. & Long, S. B. Structure and insights into the function of a Ca^{2+}-activated Cl^- channel. *Nature* **516**, 213–218 (2014).
29. Qu, Z. & Hartzell, H. Bestrophin Cl^- channels are highly permeable to HCO_3^-. *Am. J. Cell Physiol.* **294**, 1371–1377 (2008).
30. Nakajima, K., Tanaka, A. & Matsuda, Y. SLC4 family transporters in a marine diatom directly pump bicarbonate from seawater. *Proc. Natl Acad. Sci. USA* **110**, 1767–1772 (2013).

Author contributions

G.M.M.H., R.L.M. and R.D.G. carried ot RNA sample preparation, sequencing and RT-qPCR. G.M.M.H., J.A. and C.B carried out bioinformatics and statistics. G.M.M.H., J.A., M.V.O., N.S.B. and E.V.A. carried out experimental design. The manuscript was prepared by G.M.M.H., J.A., R.D.G. and E.V.A. All authors contributed to discussion of results and comments on the manuscript.

Additional information

Competing financial interests

The authors declare no competing financial interests.

Methods

Chemostat cultures. For a full description of chemostat culturing methods see ref. 12. Briefly, axenic *T. pseudonana* cells in four biological replicates (duplicate chemostats × 2 experimental runs) were acclimated to nitrate limitation at 70% (1.5 day^{-1}) of maximum growth rate for more than ten days (>15 generations) under a continuous light level of 80 μmol photons m^{-2} s^{-1}. Cell biomass was maintained at ~2×10^5 cells ml^{-1} by 10 μM nitrate, carbonate chemistry stabilized to 300, 475 or 800 μatm CO_2, verified by calculating[31] f_{CO_2} from pH and dissolved inorganic carbon (DIC) measurements. After steady-state acclimation, ~1×10^8 cells were harvested on 0.2 μm polycarbonate filters by gentle vacuum filtration and then flash frozen. Transition samples and carbonate chemistry were collected daily from chemostat cultures as CO_2 levels were increased from ~300–800 μatm at a rate ≤ 0.2 μatm min^{-1} over four consecutive days (six generations) after pre-acclimation to 300 μatm CO_2 and nitrate limitation (Fig. 1a,b). During transition, ~1.5×10^7 cells were harvested daily on 0.2 μm polycarbonate filters by gentle vacuum filtration and then flash frozen.

Nutrient-replete cultures. Triplicate cultures of axenic *T. pseudonana* were grown in f/2 medium under continuous illumination (300 μmol photons m^{-2} s^{-1}) at 20 °C with constant aeration. Exponentially growing cultures were harvested with a cell density of ~4.2×10^5 cells ml^{-1} at 240 ± 19 μatm or 857 ± 54 μatm CO_2, verified by DIC and pH measurements (Supplementary Table 3). To inhibit the hydrolysis of cAMP, 1.0 mM 3-isobutyl-1-methylxanthine (IBMX) was added to the cultures, as described in previous work with the diatom *Phaeodactylum tricornutum*[32]. Before and after a 100 min exposure to 1.0 mM IBMX, ~4×10^8 cells were harvested by gentle vacuum filtration on a 0.8 μm filter and then flash frozen for RT-qPCR analysis.

SOLiD libraries. A total of 28 barcoded transcriptome libraries were prepared from chemostat culture samples. RNA was extracted from filtered cells using the ToTALLY RNA kit (Life Technologies). Messenger RNA was selectively amplified using MessageAmp II aRNA Amplification kit (Life Technologies) and used to prepare SOLiD barcoded libraries (SOLiD Total RNA-seq kit, Life Technologies). Libraries were sequenced on a SOLiD 5500XL sequencer in two runs: one containing steady-state barcoded libraries and one containing transition barcoded libraries.

***In silico* read processing.** Reads were quality controlled (using a cutoff of $p = 0.99$ and minimum length = 30), trimmed, and aligned to *T. pseudonana* gene models using the Burrows–Wheeler Alignment tool and the SEAStAR tool (https://github.com/armbrustlab/SEAStAR). The aligned reads were counted for each gene model (Joint Genome Institute: Thaps3 extended models) using the SEAStAR tool. RNA sequences and analysis products were deposited in NCBI's Gene Expression Omnibus and are accessible through GEO series accession number GSE67971.

Gene set enrichment analysis. Read counts from transition and steady-state transcriptomes were first normalized by the trimmed mean of M-values method (TMM: R package, edgeR; ref. 33), then each gene was normalized to its own average expression level at low CO_2 (<312 μatm for transition samples and <350 μatm for steady-state samples) and $\log_2$ transformed. The normalized gene expression data were tested for correlation to CO_2 using the Spearman rank correlation test (cor. test, R stats). Genes with significant ($p < 0.05$) correlation to CO_2 in transition and steady-state transcriptomes were tested for significant gene set enrichment with the hypergeometric test (phyper, R stats) in categories defined by Kyoto Encyclopedia of Genes and Genomes (KEGG 58.1, 1 June 2011), Gene Ontology (GO 2009).

Gene clustering analysis. A data set of 98 microarray and RNA-seq samples representing ten experiments[15–20] was used to identify co-expressed gene clusters. All RNA-seq data were TMM normalized, and $\log_2$ ratios versus triplicate control samples were combined with microarray $\log_2$ ratios. To estimate the co-expression of transcripts, Pearson pairwise correlation distances were computed across all samples, following normalization such that all within-sample standard deviations were equal to one. Hierarchical clustering of these distances using Ward's method (fastcluster)[34] identified a hierarchy of co-expressed genes, from which 400 co-expressed groups were selected using arbitrary cut height. Quality controls required a cluster to have at least 15 genes expressed with normalized mean square residual[35] <0.6 across the CO_2 experiments and to be significantly correlated with CO_2 in transition or steady-state experiments (FDR < 0.0001) to be considered for further analysis. Multiscale bootstrap resampling was performed to estimate the significance and reproducibility of sub-clusters among this hierarchy. Candidate *cis*-regulatory regions were identified using MEME (ref. 36) from 0 to 800 base pairs upstream of gene start sites, and TOMTOM (ref. 37) was used to assess similarity of candidate *cis*-regulatory motifs to previously characterized *cis*-regulatory elements.

Homology and alignments. Searches for homologous sequences were performed by a hidden Markov model search tool (HMMsearch; ref. 38) across the gene models. Multiple alignments were performed with MAFFT (ref. 39) and visualized in Jalview[40].

RT-qPCR. To quantify gene expression we used primers (Supplementary Table 4) to amplify genes for δ-CA3 (ref. 41) and a putative transporter (pID 262258) from poly(A) selected RNA (MicroPoly(A)Purist kit, Life Technologies) on a StepOnePlus instrument with the Power SYBR Green RNA-to-CT 1-Step kit (Life Technologies). The gene copies were quantified in technical triplicate by a standard curve generated from the gene amplicon in a linearized 2.1 TOPO vector (Life Technologies) and normalized to mRNA in each sample. A two-way ANOVA (R stats, aov) was performed on δ-CA3 gene expression with a Tukey HSD *post hoc* test (R, stats, TukeyHSD) to determine significant differences between treatments and groups respectively. A Student's *t*-test (R, stats, *t*-test) was performed on gene expression of the putative transporter at high CO_2 only, because the low-CO_2 gene expression was below the quantification limit of the standard curve.

References

31. Lewis, E. & Wallace, D. *Program Developed for CO_2 System Calculations* (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, 1998); http://cdiac.ornl.gov/oceans/co2rprt.html
32. Harada, H., Nakajima, K., Sakaue, K. & Matsuda, Y. CO_2 sensing at ocean surface mediated by cAMP in a marine diatom. *Plant Physiol.* **142,** 1318–1328 (2006).
33. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11,** R25 (2010).
34. Mullner, D. Dastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. *J. Stat. Softw.* **53,** 1–18 (2013).
35. Reiss, D. J., Baliga, N. S. & Bonneau, R. Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. *BMC Bioinform.* **7,** 280 (2006).
36. Bailey, T. & Elkan, C. *Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Biopolymers* (Univ. California, 1994); http://biofactory.org/sites/default/files/presentations/motif.pdf
37. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. *Genome Biol.* **8,** R24 (2007).
38. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7** (2011).
39. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).
40. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).
41. Crawfurd, K. J., Raven, J. A., Wheeler, G. L., Baxter, E. J. & Joint, I. The response of *Thalassiosira pseudonana* to long-term exposure to increased CO_2 and decreased pH. *PLoS ONE* **6,** 1–9 (2011).