Molecular and physiological evidence of genetic assimilation to high CO$_2$ in the marine nitrogen fixer *Trichodesmium*

Nathan G. Walworth[a], Michael D. Lee[a], Fei-Xue Fu[a], David A. Hutchins[a], and Eric A. Webb[a,1]

[a]Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089

Most investigations of biogeochemically important microbes have focused on plastic (short-term) phenotypic responses in the absence of genetic change, whereas few have investigated adaptive (long-term) responses. However, no studies to date have investigated the molecular progression underlying the transition from plasticity to adaptation under elevated CO$_2$ for a marine nitrogen-fixer. To address this gap, we cultured the globally important cyanobacterium *Trichodesmium* at both low and high CO$_2$ for 4.5 y, followed by reciprocal transplantation experiments to test for adaptation. Intriguingly, fitness actually increased in all high-CO$_2$ adapted cell lines in the ancestral environment upon reciprocal transplantation. By leveraging coordinated phenotypic and transcriptomic profiles, we identified expression changes and pathway enrichments that rapidly responded to elevated CO$_2$ and were maintained upon adaptation, providing strong evidence for genetic assimilation. These candidate genes and pathways included those involved in photosystems, transcriptional regulation, cell signaling, carbon/nitrogen storage, and energy metabolism. Conversely, significant changes in specific sigma factor expression were only observed upon adaptation. These data reveal genetic assimilation as a potentially adaptive response of *Trichodesmium* and importantly elucidate underlying metabolic pathways paralleling the fixation of the plastic phenotype upon adaptation, thereby contributing to the few available data demonstrating genetic assimilation in microbial photoautotrophs. These molecular insights are thus critical for identifying pathways under selection as drivers in plasticity and adaptation.

diazotroph | evolution | CO$_2$ | genetic assimilation | plasticity

Warming temperatures and increasing anthropogenic carbon dioxide (CO$_2$) emissions have galvanized investigations of both short- and long-term responses to global change factors in numerous biological systems. Studies assessing responses of both carbon-fixing (primary producers) and nitrogen-fixing (diazotrophs) organisms to ocean acidification have been of particular interest because of their bottom-up control of global biogeochemical cycles and food webs (1). However, attributing observed phenotypic changes to specific environmental perturbations in situ remains an ongoing challenge, particularly when delineating between phenotypic plasticity and adaptive evolution (2). Phenotypic plasticity occurs when individuals in a population of a given genotype change their phenotype as part of a rapid response to environmental change, whereas adaptive evolution occurs when the underlying genetic (allelic) composition of a population changes the phenotype as a result of natural selection (3). It is also worth noting that population-level phenotypic changes may also ultimately result from environmental stress (2).

Additionally, it has been shown that a range of phenotypic plasticity can exist within a single species (4, 5) and that phenotypic plasticity itself can evolve and aid in adaptation (3, 6, 7). As such, plasticity can potentially affect evolution in opposing ways. It may either facilitate adaptation by having natural selection fix a beneficial plastic trait (phenotype; i.e., genetic assimilation) (8), or it can shield certain genotypes from natural selection if optimal phenotypes may be produced by plasticity

alone (3). Hence, these phenomena necessitate investigations into the effects of plasticity on population-level adaptations during periods of environmental pressure. Here, we define genetic assimilation to occur when a trait that originally responded to environmental change loses environmental sensitivity (i.e., plasticity) and ultimately becomes constitutively expressed (i.e., fixed) in a population (8).

Laboratory-based experimental evolution studies enable analysis of organismal and population responses to defined experimental conditions as they transition from plastic to adaptive (7). These insights better inform environmental phenotypic observations and offer more constrained time scales of plasticity vs. adaptation. However, aside from being typically restricted to rapidly dividing microorganisms, the main experimental challenge resides in extrapolating laboratory evolutionary potential to predicting adaptive capacities in natural populations. Thus, comprehensively interpreting in situ genetic and phenotypic datasets remains challenging because of limited knowledge of fundamental biology, gene flow, population sizes, mutation, and recombination rates (3).

One promising approach is to couple molecular techniques with experimental evolution to elucidate the coordination of

Significance

The free-living cyanobacterium *Trichodesmium* is an important nitrogen-fixer in the global oceans, yet virtually nothing is known about its molecular evolution to increased CO$_2$. Here we show that *Trichodesmium* can fix a plastic, short-term response upon long-term adaptation, potentially through genetic assimilation. We provide transcriptional evidence for molecular mechanisms that parallel the fixation of the plastic phenotype, thereby demonstrating an important evolutionary capability in *Trichodesmium* CO$_2$ adaptation. Transcriptional shifts involve transposition and other regulatory mechanisms (sigma factors) that control a variety of metabolic pathways, suggesting alterations in upstream regulation to be important under genetic assimilation. Together, these data highlight potential biochemical evidence of genetic assimilation in a keystone marine N$_2$-fixer, with broad implications for microbial evolution and biogeochemistry.

underlying molecular changes as they influence both the plasticity phenotype and/or evolutionary phenotype/genotype (8, 9). For example, one recent study examining the effect of high CO_2 on gene expression changes in the eukaryotic calcifying alga *Emiliana huxleyi* found that opposing plastic and adaptive phenotypes were also reflected by their corresponding gene expression changes (10). In a preceding study, Lohbeck et al. confirmed adaptation through reciprocal transplantation and observed significant growth rate increases in high-CO_2 selected lines relative to those of low-CO_2 selected lines under elevated CO_2 conditions (11). Reduced growth and calcification in the plastic response correlated with reductions in expression of genes involved in pH regulation, photosynthesis, carbon transport, and calcification, whereas partly restored growth and calcification observed in the adaptive response were associated with the significant recovery of the expression of these genes. Hence, this experiment elucidated an opposing phenotypic relationship between plasticity and adaptation mirrored by underlying gene expression changes in *E. huxleyi*. However, no studies to date have characterized the molecular progression underlying this transition for marine nitrogen-fixing organisms.

In a prior study, we showed growth and N_2 fixation rate increases in response to short-term, elevated CO_2 (i.e., plastic response) that became fixed upon long-term adaptation in the biogeochemically important marine diazotroph *Trichodesmium* (12). High-CO_2 selected cell lines exhibited higher fitness in the ancestral (i.e., low) CO_2 condition than low-CO_2 selected and ancestral cell lines (i.e., positive correlated response), similar to Lohbeck et al. (11). This finding contrasts with other CO_2-correlated responses reported in marine phytoplankton (3, 7, 13), although other positive correlated responses have been reported in heterotrophic microbial evolution studies (14). In our study, the observed adaptive response resulted in constitutive growth and N_2 fixation rate increases under both CO_2 levels, demonstrating an apparent loss of environmental sensitivity to low CO_2.

Before our study, all other field and laboratory studies to date have only characterized *Trichodesmium* short-term responses to environmental change, which has set the stage for evolutionary investigations for factors like iron (Fe), phosphorus (P), and CO_2 (15–19). For example, multiple short-term (weeks to months) laboratory (17, 20–24) and field (1, 25) experiments have demonstrated increases in both growth and N_2 fixation in *Trichodesmium* spp. in response to elevated CO_2, revealing a plastic high-CO_2 phenotype under specific nutrient conditions. Furthermore, Hutchins et al. (26) observed divergent taxon-specific responses (i.e., reaction norms) to increased CO_2 among biogeographically distinct diazotrophic cyanobacteria including *Trichodesmium*, a result that has also been observed in eukaryotic phytoplankton (5, 27, 28). A reaction norm is defined as a short-term response of a given genotype, which describes trait values such as rate of growth as a function of two or more environments. The shape of a reaction norm will determine which genotypes can most rapidly respond under a changing environment where a greater slope reflects a greater range in plasticity that may increase the competitive advantage for that genotype (29). Hence, genus-level conservation of physiological variability across broad eukaryotic and prokaryotic genera highlights the potential importance of maintaining evolutionary plasticity in certain environmental regimes and suggests differential taxonomic CO_2 selection over geological time, which could have ultimately influenced biogeographic distributions. Additionally, Walworth et al. (30) recently validated the conservation of genome architecture and coding potential of the cyanobacterial diazotroph *Trichodesmium erythraeum* IMS101 (hereafter IMS101) with different *Trichodesmium* isolates, as well as with natural *Trichodesmium* populations sampled decades apart. These data help to environmentally contextualize the molecular results described here, which

further aid in the extrapolation of laboratory molecular adaptation to the evolutionary potential contained within natural populations.

Leveraging these phenotypic and genetic data, we investigated the global transcriptional underpinnings of long-term CO_2 selection of a single IMS101 starting population as its phenotype transitioned from plastic to adaptive. We sequenced biological triplicate transcriptomes of both long-term CO_2 treatments after 4.5 y of selection (380-selected and 750-selected), as well as both reciprocal transfers (380s-to-750 and 750s-to-380) after 2 wk in the reciprocal CO_2 concentration. One of the most striking insights separating the short- and long-term responses relative to the 380-selected phenotype (i.e., low-CO_2 phenotype) was the differential regulation of RNA polymerase sigma factors, which have been shown to induce broad shifts in metabolic pathways in response to carbon and nitrogen fluctuations in other microbial systems (31, 32). Changes in sigma factor expression have also been proposed as mechanisms for the expression of broad gene circuits to undergo canalization (i.e., fixation or loss of low-CO_2 plasticity in this case) in genetic assimilation (8), which is evidenced in our data by certain sigma factors and coexpressed genes sharing parallel expression profiles in both plastic and adaptive responses. Additionally, differential expression of transposition was detected in both short- and long-term CO_2 responses. Hence, our data suggest that differential regulation of transposition and sigma factor expression may mediate genetic assimilation to long-term CO_2 selection, potentially leading to broad, downstream changes in metabolic pathways.

Results and Discussion

One cell line was divided into two CO_2 treatments of six biological replicates each and experimentally adapted at both low [380 microatmospheres (µatm)] and high (750 µatm) concentrations for ~4.5 y (~570–850 generations, depending on CO_2 treatment) with growth rate as a proxy for reproductive fitness (9, 12). At the onset of this incubation, cell lines placed in 750 µatm CO_2 (750 ancestral) rapidly increased both growth and N_2 fixation rates, whereas cell lines in 380 µatm CO_2 (380 ancestral) sustained lower physiological rates (Fig. 1) (12). This immediate fitness increase in response to high CO_2 is consistent with the classically observed plastic response of IMS101 (high-CO_2 phenotype) as previously shown (see above). After 4.5 y of low and high CO_2 selection, no further changes in growth or N_2 fixation were observed for either the 380-selected (low-CO_2 genotype) or 750-selected (high-CO_2 genotype) lines relative to their corresponding 380- and 750-ancestral time points, respectively (Fig. 1). All six replicates in the 750-selected cell lines still maintained significantly higher growth and N_2 fixation rates relative to the 380-selected cell lines (Fig. 1; $P = 2.4 \times 10^{-5}$), but showed no further fitness increase after the initial plastic growth rate response, despite ~850 subsequent generations of selection at high CO_2 (Fig. 1, orange bars). Once subcultures of the 380-selected cell lines were placed in high (750 µatm) CO_2 for 2 wk after the 4.5-y incubation at low (380 µatm) CO_2 (380s-to -750), both growth (fitness) and N_2 fixation rapidly increased, similar to the 750-ancestral response and consistent with the aforementioned experiments (Fig. 1, green bars; $P = 1.3 \times 10^{-3}$). However, when subcultures of the 750-selected cell lines were reciprocally transplanted back to the ancestral CO_2 condition (750s-to-380; correlated response), a 44% fitness increase was observed relative to both the 380-selected and ancestral cell lines (Fig. 1, *Lower*; blue bars; $P = 3.2 \times 10^{-3}$), similar to a positive correlated response in one other study (11), but contrasting with most others reported in marine phytoplankton (3, 7, 13). This positive correlated response is corroborated by a nonsignificant selection x assay interaction from the two-way analysis of variance (ANOVA, $F = 14.99$, $P = 0.18$; Fig. 1). As such, the 750-selected cell lines after long-term high-CO_2 selection were characterized not by steady fitness increases in the selection environment, but a loss of environmental
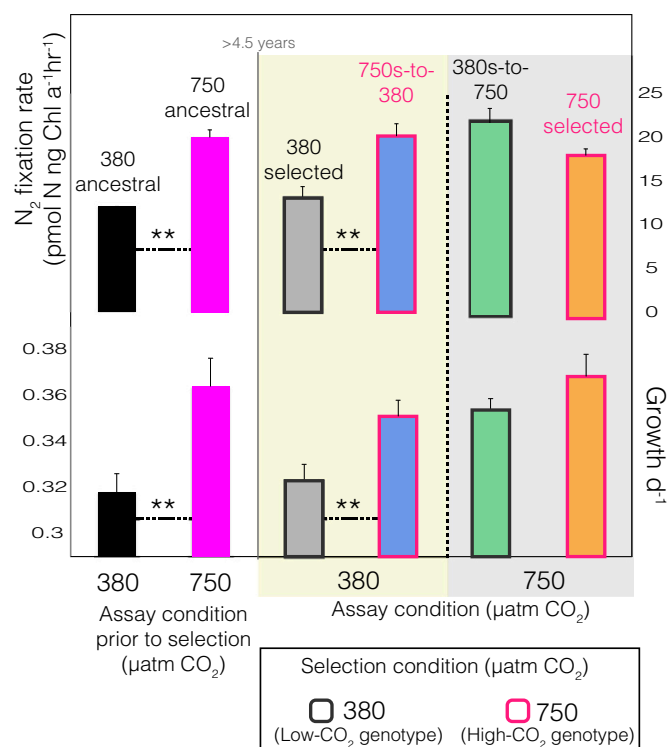
Fig. 1. Growth (*Lower*) and N_2 fixation (*Upper*) rates of the ancestral and CO_2-selected cultures before and after 4.5 y of selection, respectively. The assay condition is denoted on the x-axis, and the selection condition is denoted by the colors of the bar border. The bar colors denote the different experimental treatments that are indicated above each bar. Assays done after 4.5 y of long-term CO_2 selection are denoted after the gray solid vertical line. The yellow background denotes the 380-μatm CO_2 assay condition, and the gray background denotes that of the 750-μatm CO_2. Statistically significant differences were determined by two-way ANOVAs (selection x assay conditions) followed by Tukey's HSD post hoc tests. Asterisks denote statistical significance between two respective treatments. **$P \leq 0.01$. Error bars are SEs of six biological replicate cultures.

lines in low CO_2, but have indeed evolutionary shifted from the 750-selected cell lines under low CO_2 (Fig. S1).

Hereafter, we have chosen the terms "plastic response" and "adaptive response" to describe the physiological and transcriptional responses deriving from the low- and high-CO_2 genotypes, respectively. The 380s-to-750 is a plastic (i.e., nonadapted) response to high CO_2 deriving from the 380-selected cell lines (i.e., low-CO_2 genotype), and thus we term both the physiological and transcriptional data of the 380s-to-750 treatment a plastic response. Accordingly, because we demonstrated adaptation to have taken place in the 750-selected cell lines (12), the physiological and transcriptional data of the 750-selected cell lines are an adaptive response of the high-CO_2 genotype. Because the 750-selected, the 380s-to-750, and the 750s-to-380 all exhibited the high-CO_2 phenotype (Fig. 1), genes sharing parallel expression profiles among all three treatments represent those that both rapidly responded to increased CO_2 as part of the plastic response and subsequently maintained these profiles as part of the adaptive response, making them putative candidates for genetic assimilation (8). These changes provide evidence for genes whose expression may have been canalized (i.e., loss of low-CO_2 plasticity) (8) reflected in the 750s-to-380 condition. Because the cell lines in the 750s-to-380 treatment are the 750-selected cell lines (e.g., same high-CO_2 genotype), the transcriptional and physiological data deriving from this 750s-to-380 treatment are a mixture of the 750-selected cell lines transcriptional plasticity to low-CO_2 (e.g., the 308 down-regulated genes in the lower blue portion of the 750s-to-380 treatment circle in the Venn diagram in Fig. 2) and genes whose expressions are putatively assimilated as a product of adaptation in which expression profiles should be analogous to those of the 750-selected treatment, regardless of CO_2 concentration (i.e., the 45 down-regulated gene portion of the Venn diagram shared by the 750-selected and 750s-to-380 treatments in Fig. 2). Hence, gene expression profiles from the 750-selected and 750s-to-380 treatment derive from the same high-CO_2 genotype (i.e., adapted to high CO_2), and thus we describe these shared expression profiles as part of an adaptive response reflecting the loss of physiological CO_2 plasticity (i.e., genetic assimilation). Hence, the transcriptional

sensitivity to low CO_2 by the measured phenotypic traits (growth and N_2 fixation). Hence, because both the 380s-to-750 and 750-selected responses to increased CO_2 exhibited the same high-CO_2 phenotype in the selection environment, the plastic response appears to have been fixed upon adaptation, suggesting that the low-CO_2 genotype underwent genetic assimilation to produce the high-CO_2 genotype (8). Additionally, upon graphing the growth rate slopes of the 380-ancestral to 750 switch (380a-to-750), 380s-to-750, and the 750s-to-380 across CO_2 regimes, positive slopes are observed for both the 380a-to-750 and the 380s-to-750 going from 380 to 750 μatm CO_2 (Fig. S1). In contrast, a negative slope is observed for the 750s-to-380 cell lines, suggesting an evolutionary shift in reaction norms between the low-CO_2 genotype treatments (380-ancestral and 380-selected) and the high-CO_2 genotype (750-selected): a criterion of genetic assimilation (8). The lack of a positive slope of the 750s-to-380 going from 380 to 750 μatm CO_2 provides strong evidence that environmentally responsive traits (here growth and N_2 fixation) lose environmental sensitivity by maintaining significantly increased rates in 380 μatm CO_2 relative to the 380-selected cell lines in the same CO_2 condition (8). The similar growth rate slopes of the 380a-to-750 and 380s-to-750 corroborate that the 380-ancestral and 380-selected cell lines are genetically analogous in terms of CO_2, which demonstrates that the 380-ancestral cell lines growing in low CO_2 have not evolutionarily shifted from the 380-selected cell

A Downregulated

Carbon Metabolism
- ◆ pyruvate metabolic process
- ◆ lipoate biosynthesis
- ✳ polysaccharide transport

Transcription/Translation
- ■ sigma factor activity

Energy/General Metabolism
- ✳ phosphoribosylformylglycinamidine synthase

DNA Binding Transcription Factor
- ◆ regulation of transcription

Cell signaling/communication
- ■ phosphoric diester hydrolase

B

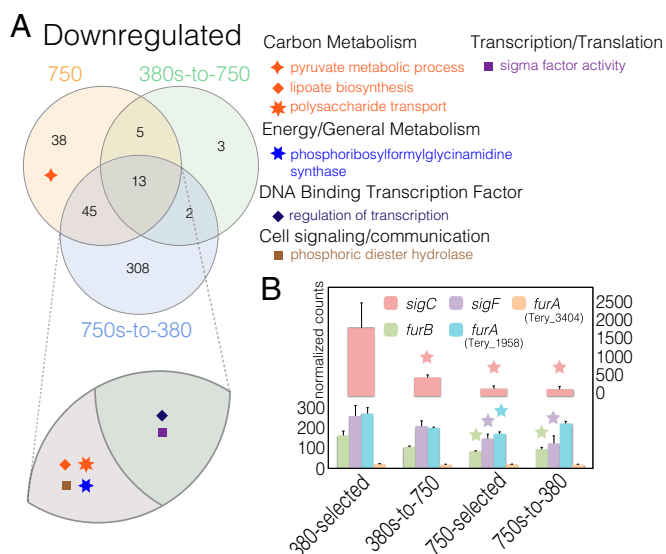**Fig. 2.** Shown are down-regulated GO-enriched pathways relative to 380-selected and transcriptional profiles of sigma factors, *sigC* and *sigF* as well as Fur proteins. (*A*) Down-regulated GO-enriched pathways for the all high-CO_2 phenotype treatments. (*B*) Differential expression of transcriptional regulators with stars representing statistical significance relative to the 380-selected and error bars being SEs.
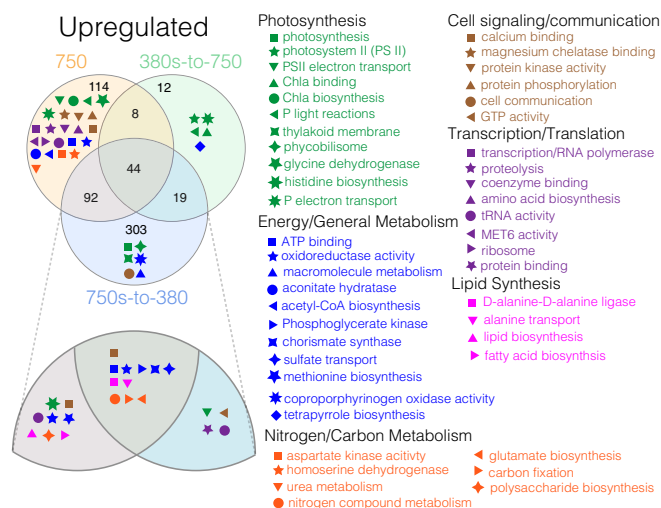
pool shared among the high-CO_2 genotype treatments (750-selected and 750s-to-380) may contain mechanisms that are potentially driving the maintenance of the high-CO_2 phenotype, even after transfer back to ancestral CO_2 levels. Therefore, we characterize gene expression profiles shared between the 750-selected and 750s-to-380 conditions that parallel this phenotypic maintenance, reflecting the loss of low-CO_2 sensitivity after long-term, high-CO_2 adaptation.

To examine transcriptional responses, we gently filtered replicate cultures of each treatment growing semicontinuously during the middle of the photoperiod (~11:00 AM) followed by flash-freezing in liquid nitrogen and storing until processing (*SI Materials and Methods*). Libraries were then constructed by using equimolar amounts of RNA per library and sequenced on the Illumina HiSeq 2000 (i.e., RNA-Seq), yielding 50-base pair, single-end reads (*SI Materials and Methods*). Reads were then quality trimmed and mapped onto the IMS101 reference genome, followed by normalization and differential expression analysis using the edgeR package (see *SI Materials and Methods* for full statistical description). Briefly, libraries were normalized using the trimmed mean of M-values method. M-values are the library size-adjusted log-ratio of counts between the control RNA-Seq library (380-selected), and the treatment of interest in which the most extreme 30% of M values are trimmed before calculating the resulting trimmed mean. This method attempts to eliminate systematic differences in the counts between the libraries (e.g., RNA pools between treatments) by assuming that most genes are not differentially expressed. Common dispersion was estimated by fitting a generalized linear model (GLM), and differentially expressed genes were determined by fitting the negative binomial GLM followed by a likelihood ratio test. Finally, genes with a Benjamini–Hochberg false discovery rate (FDR) < 0.05 were deemed differentially expressed. General sequencing and differential expression statistics can be found in Tables S1 and S2, and global expression maps can be examined in Fig. S3.

We were able to identify processes involved in the short (380s-to-750; plastic), long (750-selected; adaptive), and correlated (750s-to-380) responses, implicating them to be important for instigating and sustaining the high-CO_2 phenotype (i.e., genetic assimilation; Dataset S1). Genes exhibiting significant decreases in expression (down-regulation) in the high-CO_2 phenotype (Fig. 2A "Plastic + Adaptive") relative to low-CO_2 phenotype levels were enriched in Gene Ontology (GO) metabolisms involving broad metabolic processes—particularly of note, sigma factor activity and carbon transport [Fig. 2A, hypergeometric test with Benjamini–Hochberg correction FDR <= 0.1 (33)]. Differential expression of specific transposition types (Fig. 3 and below) was also correlated to the high-CO_2 phenotype, suggesting them to be potential targets of CO_2 selection or the result of prolonged CO_2 exposure. Similarly, genes with increased expression (up-regulation) shared across all three treatments exhibiting the high-CO_2 phenotype (380s-to-750, 750-selected, and 750s-to-380) were detected in widespread pathways (Fig. 4 and below), making them candidate genes that may have undergone genetic assimilation in the transition from plasticity to adaptation.

To test for the probability of sharing a given amount of differentially expressed genes between treatments by chance alone, pairwise hypergeometric tests were conducted for both up- and down-regulated gene sets. These results revealed very low probabilities that the number of genes exhibiting parallel expression profiles were shared by chance alone among down-regulated gene sets (Fig. 2A) between 380s-to-750 and 750s-to-380 ($P < 10^{-13}$), 750-selected and 380s-to-750 ($P < 10^{-29}$), and 750-selected to 750s-to-380 ($P < 10^{-42}$). Similarly, low probabilities were also observed for up-regulated gene sets (Fig. 4) between the 380s-to-750 and 750s-to-380 ($P < 10^{-50}$), the 750-selected and 380s-to-750 ($P < 10^{-48}$), and the 750-selected to 750s-to-380 ($P < 10^{-79}$). Because of the low probabilities of

Fig. 3. Shown are hierarchical clustering of logtwofold changes of TE centroids differentially expressed in at least one high-CO_2 phenotype treatment (A), the distribution of TE copies within different genomic elements (B), and the distribution of TE centroids and their corresponding genome copies (C). (A) Hierarchical clustering of log-twofold changes of differentially expressed TE centroids in at least one high-CO_2 phenotype treatment (*SI Materials and Methods*) resulting in two well-defined clusters with one representing TE centroids with increased average log-twofold changes in the high-CO_2 phenotype (blue) treatments relative to the 380-selected and vice versa for the "decreased" cluster (red). Bolded/asterisked labels indicate differentially expressed TE clusters in all high-CO_2 phenotype treatments vs. the 380-selected treatment. (B) Pie chart of the distribution of all detected TE copies in the genome. (C) A genome plot of TE centroids and their corresponding copies as well as all detected TE copies in the genome. Going from outside to inside: Track 1 shows TE copies in the genome on the forward strand and is colored according to genomic element. Track 2 is the same, but on the minus strand. Track 3 contains the names and symbols for differentially expressed TE centroids in A relative to the 380-selected treatment. Track 4 shows the distribution of the corresponding copies of differentially expressed TE centroids in track 3 via blue and red colored links. Underlying gray links represent paralogous copies from TE clusters showing no change in expression.

sharing these many genes or more between treatments by chance alone, it is likely that some of the downstream effects of these shared expression changes are associated with the plastic and/or adaptive responses to high CO_2.

Fig. 4. Shown are up-regulated GO-enriched pathways for the all high CO_2 phenotype treatments relative to the 380-selected replicates. Significantly up-regulated genes were classified into GO pathways and tested for significant enrichment among the treatments (*SI Materials and Methods*).

Sigma Succession Underlying Plasticity-Mediated Adaptation. Our data show that increased CO_2 correlates with lower expression of RNA polymerase sigma factors *sigC* (Tery_1956; group 2) and *sigF* (Tery_3916; group 3) (Fig. 2B; ref. 34). Differential regulation of sigma factors, "sigma switching," aids in both stress responses and adaptation via transcriptional initiation of gene sets specific to particular environmental or internal cellular changes (35). For example, *sigC* transcripts have been shown to increase under short-term nitrogen limitation in diazotrophic cyanobacteria (31), which is consistent with the simultaneous decrease in *sigC* (ortholog to *Anabaena sigC*, reciprocal best blast hit, evalue < 1e-10) expression and increase in nitrogen fixation in all high-CO_2 phenotypes (Fig. 2B, red bars).

Furthermore, homologs of the ferric uptake regulator protein, Fur, have been shown to bind to the promoter region of *sigC* in cyanobacteria, implicating *sigC* to have a connective role in both nitrogen and iron homeostasis, and potentially oxidative stress as well (36). Accordingly, the ferric uptake regulators, *furA* (Tery_1958) and *furB* (Tery_1953) genes (orthologs to *Anabaena*, reciprocal best blast hit, evalue < 1e-10), exhibited parallel decreases in expression with *sigC* under prolonged high CO_2 (Fig. 2B). In contrast, a *fur* paralog (Tery_3404) showed no changes in expression after prolonged exposure to high CO_2 in replete iron. These transcriptional reductions of *fur* homologs in high CO_2 may enhance tetrapyrrole production (Fig. 4 and below), as observed in the cyanobacterial diazotroph *Anabaena* (37). Furthermore, *furA* (Tery_1958) and *furB* slightly increased expression in the 750s-to-380 treatment compared with the 750-selected treatment, suggesting a short-term response to low CO_2 exposure relative to their decreased expression seen in the 750-selected under high CO_2. Together, these data provide some evidence of the coregulation of *sigC* and specific *fur* genes as part of both short-term plastic and long-term adaptive responses. Additionally, the maintenance of *sigC* down-regulation in the transition from the low- to high-CO_2 genotype (Figs. 1 and 2B) provides strong evidence for genetic assimilation of this sigma factor and its targets.

Conversely, *sigF* (ortholog to *Synechocystis sigF*, reciprocal best blast hit, evalue < 1e-10) transcription was only significantly decreased in 750-selected and 750s-to-380 treatments (high-CO_2 genotype), suggesting significant down-regulation of *sigF* to be primarily involved in adaptation rather than initiation of the high-CO_2 phenotype (Fig. 2B, purple bars). *sigF* is involved in a

variety of cellular processes and has been shown to target other transcriptional regulators such as *rsfA* in the Gram-positive *Bacillus subtilis* (38), as well as a phytochrome-like histidine kinase in the cyanobacterium *Synechocystis* PCC6803 (34). Interestingly, an IMS101 hypothetical protein (Tery_2530), containing an *rsfA* domain (BLASTx, default settings), as well as a PAS/PAC signal transduction histidine kinase (Tery_4221) containing several overlapping portions of conserved domains including bacteriophytochrome (COG4251), phosphate regulon sensor kinase (PhoR; TIGR02966) (39), and NtrY (COG5000) also exhibited parallel down-regulation with *sigF* (Dataset S1). Intriguingly, NtrY modulates *nifA* expression that specifically controls expression of N_2-fixing *nif* genes in the symbiotic diazotroph *Azorhizobium caulinodans* ORS571 (40). However, no IMS101 *nifA* homologs to that of ORS571 were detected. PAS-containing histidine kinases have also been shown to bind to a wide array of cofactors and are important signaling modules that monitor changes in light, redox potential, small ligands, and cellular energy (41, 42). Hence, Tery_4221 may regulate several different metabolic functions aiding in increased growth and N_2 fixation. Regardless, its significantly decreased expression in conjunction with *sigF* after long-term high CO_2 exposure implicates a role in influencing the high-CO_2 genotype/phenotype.

Together, these data suggest that the fixation of long-term, increased growth and N_2 fixation in 750-selected cell lines is associated with a short-term, plastic response reflected in *sigC* and other genes whose expression profiles were similar in both the 380s-to-750, 750-selected, and 750s-to-380 treatments. Reduced *sigC* transcription is associated with the initiation of the high-CO_2 phenotype, whereas decreased expression of both *sigC* and *sigF* ultimately contribute to its adaptive maintenance.

Transposition Regulation in Plasticity and Adaptation. In addition to sigma switching, shifts in transposable element (TE) regulation have been shown to be involved in both environmental plasticity and adaptation (43). Numerous partial TE genome sequences are marks of neutral maintenance based on TE deletion bias (43), and IMS101 is indeed enriched in TEs and TE pseudogenes (30). Accordingly, it has been shown that repetitive elements such as TEs can selectively mediate genome plasticity, whereas partial TEs can also act to inhibit transposition, which may be partly why both IMS101 and natural populations of *Trichodesmium* have retained numerous repetitive elements and TE pseudogenes (30). Hence, the long-term maintenance of TEs and their expression in situ may indicate an important role for both the plasticity and adaptation of *Trichodesmium* to environmental change in situ. However, reliably quantifying expression of repetitive DNA sequences (repeats) such as TEs (e.g., insertion sequences) remains a significant challenge for next-generation sequencing methods using short-read technology (e.g., Illumina sequencing typically 50–150 base pairs) because of difficulty in mapping repetitive sequences to a single genomic location (e.g., multireads; *SI Note 1*) (44).

To try and circumvent these challenges, we developed a method to quantify the expression of TE clusters by binning TE sequences with ≥70% identity into clusters (45) and quantifying the expression of each cluster across treatments (see *SI Note 1* for detailed methodology). These analyses resulted in several (5 of 16) of the clusters exhibiting significantly different mean expression values between the low-CO_2 (i.e., 380-selected) and high-CO_2 (i.e., 380s-to-750, 750-selected, and 750s-to-380) phenotypes (Fig. 3 *A* and *C*, bolded/asterisk labels). We also identified a cluster (TE_67) that initially responded to high CO_2 in the plastic response (380s-to-750) and maintained its expression profile in the adaptive response (750-selected and 750s-to-380), consistent with other genes' expression (e.g., *sigC*), showing evidence for genetic assimilation (Fig. S2). Other TE clusters' mean expression was also different from the 380-selected treatments in other high-CO_2

phenotype conditions, suggesting consistent responses to CO_2 across conditions (Fig. S2).

Additionally, 77% (53 of 69) of total TE clusters ($n = 69$; Dataset S2) showed no differences in expression indicating maintained cluster expression irrespective of phenotype, which suggests widespread TE activity devoid of selection (Fig. 3C, light gray links). Together, these patterns corroborate the maintenance of transposition as a result of neutral processes (43) and/or weak selection (46, 47) where stable coexistence occurs between TEs and the host genome. Furthermore, ~75% of all detected TE paralogs reside within either genic or pseudogenic bodies (Fig. 3B), suggesting pseudogenization as a mechanism for generating degenerate TE genome copies. Upon plotting the locations of all differentially expressed centroid sequences along with their corresponding paralogs (Fig. 3C), some TE clusters contained numerous copies with widespread distributions across the genome (e.g., TE_30 and TE_56), whereas others contained only one or two copies (e.g., TE_12 and TE_40). The mechanisms involved in the differing degrees of TE cluster proliferation remain unknown, but these potential TE-controlling mechanisms may also contribute to the high genome conservation observed between IMS101, other isolates, and natural populations (30). In summary, the fact that most TE clusters show no changes in expression between the low- and high-CO_2 phenotypes suggests that these are being maintained under neutral processes and/or weak selection, whereas the few that did exhibit differences between phenotypes are potential candidates under selection. Hence, these data implicate differential regulation of certain transposition types to be involved in or caused by both plastic and adaptive high-CO_2 phenotypes (see SI Note 1 for more discussion).

Functional GO-Enriched Transcription in Plasticity and Adaptation. The plastic response (i.e., 380s-to-750) exhibited significant GO enrichment of tetrapyrrole biosynthesis driven by the up-regulation of Tery_3684 (*Anabaena* sp. wa102 homolog, *hemB*, AA650_24065). These expression profiles are consistent with prior observations of *hemB* induction via decreases in Fur transcription in the cyanobacterial nitrogen-fixer *Anabaena* (37). González et al. (37) also showed dual roles of Fur regulation in some Heme proteins involved in tetrapyrrole biosynthesis, including transcriptional repression (e.g., *hemB*, *hemC*, and *ho1*) and activation (e.g., *hemK* and *hemH*). However, other *hem* genes showed either broad variability or no changes in expression and seemed to be regulated independently of Fur, leading the authors to suggest each *hem* gene to be under different regulatory mechanisms. Similar to *Anabaena*, we detected the up-regulation of *hemB*, whereas other detected *hems* either showed variable or no changes in expression. Additionally, the significant fraction of shared up-regulated genes between the 380s-to-750 and the 750s-to-380 (hypergeometric test, $P < 10^{-50}$) were enriched in GTP and cytochrome b6f complex homologs, which transfers electrons from PSII to PSI. The enrichment of PSII light-harvesting and electron transport metabolisms was also consistent with previously observed decreases in PSI:PSII ratios under short-term exposure to high CO_2 (20) (Fig. 4). Together, shifts in transcription of genes involved in electron flow deriving from PSII seem to be part of a rapid plastic response to general fluctuations in CO_2 regardless of sign. Interestingly, all high-CO_2 phenotype treatments demonstrated slightly different enrichments of photosynthesis GO subpathways, but they all shared up-regulation of proteins involving electron flow and light reactions, suggesting these transcriptional changes to be important in the genetic assimilation of the high-CO_2 genotype/phenotype.

Hutchins (2007) and Levitan et al. (2007) observed no significant changes in either photosynthetic rates or photochemical activity of PSII, respectively, between short-term, low- and high-CO_2 treatments in IMS101. These observations led them to suggest that increased growth and N_2 fixation is energized from decreased energetic demands in other cellular processes (e.g., alleviation of carbon limitation) rather than increased photosynthetic electron flow (17, 20). However, Levitan et al. (2007) also observed lower PSI:PSII ratios under high CO_2, indicating decreased investment in PSI biosynthesis generally consistent with our expression results (see above), and leading them to hypothesize that a reduction in iron-heavy PSI would free up available Fe for nitrogenase (20). Our results generally support these observations through decreases in carbon transport and increases in PSII-associated gene expression (Figs. 2 and 4). It is worth noting that the discrepancy between the increases in PSII-associated gene expression and the lack of observed changes in photosynthetic and PSII activity may be due to several possibilities, including time of sampling, posttranscriptional, posttranslational, and protein degradation regulation (48). Future studies investigating changes in gene expression and/or photosynthesis-related protein abundances involved in photosystem electron flow should include several diel sampling points with simultaneous measurement of photosynthetic rates to determine the specific roles of the photosystem expression and activity associated with increased growth and N_2 fixation.

GO-enriched groups in both plastic and adaptive responses (Fig. 4, *Center*, Plastic + Adaptive) included enhanced energy production (Fig. 4, blue symbols), carbon fixation (orange symbols; consistent with ref. 17), nitrogen storage (orange symbols), and carbon storage (magenta symbols). As such, these enriched metabolisms highlight pathways potentially influenced by CO_2 concentrations on short timescales, which appear to have been subsequently fixed upon prolonged CO_2 exposure in a stable, nutrient replete environment.

Teasing Apart the Molecular Succession Underlying Potential Genetic Assimilation. Although the physiological transition from plasticity to adaptation was phenotypically neutral in the selection environment, the same high-CO_2 phenotype shared between the low-CO_2 (380s-to-750) and high-CO_2 (750-selected and 750s-to-380) genotypes enabled identification of expression changes that were initially involved in short-term increased growth and sustained in long-term adaptive maintenance, thereby corroborating gene expression canalization via genetic assimilation. These gene expression changes may cause or be part of other phenotypic changes, so our phenotypically indistinguishable (growth and N_2 fixation) low- and high-CO_2 genotypes may in fact differ through other expressed traits unmeasured in this study. Future studies analyzing genes exhibiting diel patterns of expression should also include sampling at alternative time points to see how diel expression changes correlate with the observed physiology. However, other portions of the metabolic pool exhibited clear expression differences solely in the high-CO_2 genotype (750-selected and 750s-to-380 treatments) relative to the 380-selected, suggesting these gene expression changes to be specific to the long-term maintenance of the high-CO_2 phenotype, even in the ancestral environment. Several other lines of evidence also corroborate these molecular and physiological parallels.

First, of all differentially expressed genes, it is unlikely that the amount of genes exhibiting consistent expression between the 750-selected and 750-to-380s conditions were shared by chance alone for both the down-regulated (hypergeometric test, $P < 10^{-42}$) and up-regulated ($P < 10^{-79}$; Figs. 2 and 4, Adaptive sections) fractions, suggesting their expression to be nonrandomly associated to this genotype. Second, the strong statistical support for the differential down-regulation of specific sigma factors and other metabolic genes (see above) in the plastic vs. adaptive responses suggests differing roles in short- and long-term CO_2 phenotypes, respectively. It is worth noting that changes in expression of these sigma factors and other genes may either be part of the mechanisms producing the observed phenotype or secondary effects

after upstream metabolic/mechanistic processes responding to high CO_2. For example, differential regulation of *sigC* may either help mediate or be a product of the transition from plasticity to adaptation, whereas differences in *sigF* expression may be primarily associated with the adaptation (Fig. 2).

Additionally, the plastic response of *Trichodesmium* to high CO_2 in stable light and replete nutrients may initially shield it from adaptation on short timescales because an optimum phenotype is achieved by plasticity alone. However, upon prolonged selective CO_2 pressure, initial short-term, plastic responses appear to become fixed if held under constant conditions, which in this case seems to have led to a loss of the low-CO_2 phenotype (i.e., genetic assimilation) (Fig. 1, blue bars). Underlying this physiological trend are canalized pathways in both down-regulation (Fig. 2) and numerous up-regulated pathways (Fig. 4). Although our physiological and transcriptional data conform to prior criteria and observations set forth by other independent studies observing genetic assimilation (8, 49, 50), future studies can include various time series assays (e.g., reaction norms and functional genomics), which may further elucidate underlying mechanisms contributing to the adaptive walk of genetic assimilation as these mechanisms currently remain unclear (49). For example, in a theoretical modeling study, Kronholm and Collins (2015) suggest that one potential genetic assimilation mechanism may be that an epigenetic mutation produces an optimal (plastic) phenotype that is then later replaced by a genetic mutation to maintain it (now an adaptive phenotype) (49). This replacement thus results in a trait that is now environmentally robust to the environmental fluctuation that first triggered it (in this case, CO_2; Fig. 1).

In summary, the adaptation of IMS101 to high CO_2 in stable light and replete nutrients is mediated through an initial plastic response reflected in corresponding changes in both phenotype and gene expression. Our data suggest upstream regulatory elements (e.g., sigma factors) and differential regulation of transposition clusters to influence both short- and long-term CO_2 responses. The maintenance of the adaptive phenotype in the ancestral condition may be influenced by both plasticity-derived gene expression as well as canalized gene expression after adaptation. Additionally, increased transcription of photosystem electron flow and its mechanical components [e.g., histidine enrichment (51, 52)], in concert with the differential expression of potential iron and redox sensing regulation, possibly suggests constant light and replete iron to be synergistically acting with enhanced CO_2 to initiate and maintain increased growth and N_2 fixation. Indeed, the short-term achievement of the plastic high-CO_2 phenotype has been demonstrated in natural populations when conditions were appropriate (1, 25), but

our observed form of laboratory adaptation, defined by the apparent loss of the low-CO_2 phenotype, will likely depend on both genotype (26) and the availability of in situ compensatory environmental factors [e.g., replete phosphorus and iron (53)] to maintain increased growth and N_2 fixation. Future efforts will involve alternative DNA-sequencing technologies (i.e., long-read DNA sequencing and optical mapping) to identify potentially adaptive genomic rearrangements under high CO_2 to circumvent analysis issues produced by short-read sequencing (see above).

Because the plastic high-CO_2 phenotype seemed to have been fixed upon adaptation in IMS101, optimal plastic phenotypes that may have initially shielded adaptive genotypes can be acted on by natural selection to facilitate adaptation upon longer selection. Varying physiological results have been observed in other algal systems in which the plastic response is either maintained during evolution as in this study or ultimately reversed by adaptation (13, 54). These types of data provide environmentally relevant genetic context to physiological adaptation (see *SI Note 2* for more discussion) and future efforts examining both genetic and epigenetic effects on adaptation should provide insight into potential mechanisms driving ultimate differences in expression levels between experimental conditions (9, 49). In summary, this study supports both past observed short-term CO_2 responses and contributes evolutionary observations corroborating genetic assimilation in the globally distributed and biogeochemically important *Trichodesmium*.

Materials and Methods

Growth and N_2 fixation data were obtained from Hutchins et al. (12). Treatments were analyzed in biological triplicate with both sampling and RNA isolation being conducted as described (30). Raw fastq files were processed as described (30) and mapped onto IMS101, IMG-called genes (https://img.jgi.doe.gov/) using Bowtie2 (Version 2.2.6) (55) with default settings followed by differential expression analysis using edgeR (56).

Venn diagrams were produced by using differentially expressed gene lists per treatment (57). The "phyper" function in "R" (R Core Team 2014) was used for hypergeometric tests, and *P* values were corrected with the Benjamini–Hochberg method (58) using the "p.adjust" function (33). Transposable element sequences were downloaded from Walworth et al. (30) and clustered at 70% identity using USEARCH (59) followed by mapping of RNA-Seq reads to each cluster. Hierarchical clustering was conducted with "pvclust" (45), and Welch's T-tests assuming heteroscedasticity were conducted in Microsoft Excel. See *SI Materials and Methods* for further details.

1. Hutchins DA, Mulholland MR, Fu F-X (2009) Nutrient cycles and marine microbes in a CO_2-enriched ocean. *Oceanography (Wash DC)* 22(4):128–145.
2. Merilä J, Hendry AP (2014) Climate change, adaptation, and phenotypic plasticity: The problem and the evidence. *Evol Appl* 7(1):1–14.
3. Collins S, Rost B, Rynearson TA (2014) Evolutionary potential of marine phytoplankton under ocean acidification. *Evol Appl* 7(1):140–155.
4. Boyd PW, et al. (2013) Marine phytoplankton temperature versus growth responses from polar to tropical waters—outcome of a scientific community-wide study. *PLoS One* 8(5):e63091.
5. Schaum E, Rost B, Millar AJ, Collins S (2012) Variation in plastic responses of a globally distributed picoplankton species to ocean acidification. *Nat Clim Chang* 3(3):298–302.
6. Draghi JA, Whitlock MC (2012) Phenotypic plasticity facilitates mutational variance, genetic variance, and evolvability along the major axis of environmental variation. *Evolution* 66(9):2891–2902.
7. Schaum C-E, Rost B, Collins S (2016) Environmental stability affects phenotypic evolution in a globally distributed marine picoplankton. *ISME J* 10(1):75–84.
8. Ehrenreich IM, Pfennig DW (2016) Genetic assimilation: A review of its potential proximate causes and evolutionary consequences. *Ann Bot (Lond)* 117(5):769–779.
9. Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nat Rev Genet* 4(6):457–469.
10. Lohbeck KT, Riebesell U, Reusch TBH (2014) Gene expression changes in the coccolithophore *Emiliania huxleyi* after 500 generations of selection to ocean acidification. *Proc Biol Sci* 281(1786):20140003.

11. Lohbeck KT, Riebesell U, Reusch TBH (2012) Adaptive evolution of a key phytoplankton species to ocean acidification. *Nat Geosci* 5(5):346–351.
12. Hutchins DA, et al. (2015) Irreversibly increased nitrogen fixation in *Trichodesmium* experimentally adapted to elevated carbon dioxide. *Nat Commun* 6:8155.
13. Schaum CE, Collins S (2014) Plasticity predicts evolution in a marine alga. *Proc Biol Sci* 281(1793):20141486.
14. Travisano M, Vasi F, Lenski RE (1995) Long-term experimental evolution in *Escherichia coli*. III. Variation replicate populations in correlated responses to novel environments. *Evolution* 49(1):189–200.
15. Sañudo-Wilhelmy SA, et al. (2001) Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. *Nature* 411(6833):66–69.
16. Mills MM, Ridame C, Davey M, La Roche J, Geider RJ (2004) Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature* 429(6989):292–294.
17. Hutchins DA, et al. (2007) CO_2 control of *Trichodesmium* N_2 fixation, photosynthesis, growth rates, and elemental ratios. *Limnol Oceanogr* 52(4):1293–1304.
18. Sohm JA, Webb EA, Capone DG (2011) Emerging patterns of marine nitrogen fixation. *Nat Rev Microbiol* 9(7):499–508.
19. Chappell PD, Moffett JW, Hynes AM, Webb EA (2012) Molecular evidence of iron limitation and availability in the global diazotroph *Trichodesmium*. *ISME J* 6(9):1728–1739.
20. Levitan O, et al. (2007) Elevated CO_2 enhances nitrogen fixation and growth in the marine cyanobacterium *Trichodesmium*. *Glob Change Biol* 13(2):531–538.
21. Barcelos e Ramos J, Biswas H, Schulz KG, LaRoche J, Riebesell U (2007) Effect of rising atmospheric carbon dioxide on the marine nitrogen fixer *Trichodesmium*. *Global Biogeochem Cycles* 21:GB2028.

22. Kranz SA, et al. (2010) Combined effects of CO_2 and light on the N_2-fixing cyanobacterium *Trichodesmium* IMS101: Physiological responses. *Plant Physiol* 154(1): 334–345.

23. Levitan O, et al. (2010) Regulation of nitrogen metabolism in the marine diazotroph *Trichodesmium* IMS101 under varying temperatures and atmospheric CO_2 concentrations. *Environ Microbiol* 12(7):1899–1912.

24. Garcia NS, et al. (2011) Interactive effects of irradiance and CO_2 on CO_2 fixation and N_2 fixation in the diazotroph *Trichodesmium erythraeum* (cyanobacteria). *J Phycol* 47(6):1292–1303.

25. Shetye S, Sudhakar M, Jena B, Mohan R (2013) Occurrence of nitrogen fixing cyanobacterium *Trichodesmium* under elevated pCO_2 conditions in the western Bay of Bengal. *Int J Oceanogr* 2013: 10.1155/2013/350465.

26. Hutchins DA, Fu F-X, Webb EA, Walworth N, Tagliabue A (2013) Taxon-specific response of marine nitrogen fixers to elevated carbon dioxide concentrations. *Nat Geosci* 6(7):1–6.

27. Langer G, Nehrke G, Probert I, Ly J, Ziveri P (2009) Strain-specific responses of *Emiliania huxleyi* to changing seawater carbonate chemistry. *Biogeosciences* 6:2637–2646.

28. Kremp A, et al. (2012) Intraspecific variability in the response of bloom-forming marine microalgae to changed climate conditions. *Ecol Evol* 2(6):1195–1207.

29. Reusch TBH, Boyd PW (2013) Experimental evolution meets marine phytoplankton. *Evolution* 67(7):1849–1859.

30. Walworth N, et al. (2015) *Trichodesmium* genome maintains abundant, widespread noncoding DNA in situ, despite oligotrophic lifestyle. *Proc Natl Acad Sci USA* 112(14): 4251–4256.

31. Brahamsha B, Haselkorn R (1992) Identification of multiple RNA polymerase sigma factor homologs in the cyanobacterium *Anabaena* sp. strain PCC 7120: Cloning, expression, and inactivation of the *sigB* and *sigC* genes. *J Bacteriol* 174(22):7273–7282.

32. Caslake LF, Gruber TM, Bryant DA (1997) Expression of two alternative sigma factors of *Synechococcus* sp. strain PCC 7002 is modulated by carbon and nitrogen stress. *Microbiology* 143(Pt 12):3807–3818.

33. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.

34. Imamura S, Asayama M (2009) Sigma factors for cyanobacterial transcription. *Gene Regul Syst Bio* 3:65–87.

35. Camsund D, Lindblad P (2014) Engineered transcriptional systems for cyanobacterial biotechnology. *Front Bioeng Biotechnol* 2(Suppl.):40.

36. López-Gomollón S, et al. (2007) Cross-talk between iron and nitrogen regulatory networks in *Anabaena (Nostoc)* sp. PCC 7120: Identification of overlapping genes in FurA and NtcA regulons. *J Mol Biol* 374(1):267–281.

37. González A, Bes MT, Valladares A, Peleato ML, Fillat MF (2012) FurA is the master regulator of iron homeostasis and modulates the expression of tetrapyrrole biosynthesis genes in *Anabaena* sp. PCC 7120. *Environ Microbiol* 14(12):3175–3187.

38. Wu LJ, Errington J (2000) Identification and characterization of a new prespore-specific regulatory gene, rsfA, of *Bacillus subtilis*. *J Bacteriol* 182(2):418–424.

39. Su Z, Olman V, Xu Y (2007) Computational prediction of *Pho* regulons in cyanobacteria. *BMC Genomics* 8(1):156.

40. Pawlowski K, Klosse U, de Bruijn FJ (1991) Characterization of a novel *Azorhizobium caulinodans* ORS571 two-component regulatory system, NtrY/NtrX, involved in nitrogen fixation and metabolism. *Mol Gen Genet* 231(1):124–138.

41. Narikawa R, Okamoto S, Ikeuchi M, Ohmori M (2004) Molecular evolution of PAS domain-containing proteins of filamentous cyanobacteria through domain shuffling and domain duplication. *DNA Res* 11(2):69–81.

42. Ashby MK, Houmard J (2006) Cyanobacterial two-component proteins: Structure, diversity, distribution, and evolution. *Microbiol Mol Biol Rev* 70(2):472–509.

43. Iranzo J, Gómez MJ, López de Saro FJ, Manrubia S (2014) Large-scale genomic analysis suggests a neutral punctuated dynamics of transposable elements in bacterial genomes. *Genome Biol Evol* 3(0):959–973.

44. Treangen TJ, Salzberg SL (2012) Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nat Rev Genet* 13(1):36–46.

45. Suzuki R, Shimodaira H (2006) Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22(12):1540–1542.

46. Navarro-Quezada A, Schoen DJ (2002) Sequence evolution and copy number of Ty1-copia retrotransposons in diverse plant genomes. *Proc Natl Acad Sci USA* 99(1):268–273.

47. Delihas N (2011) Impact of small repeat sequences on bacterial genome evolution. *Genome Biol Evol* 3(0):959–973.

48. Vogel C, Marcotte EM (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 13(4):227–232.

49. Kronholm I, Collins S (2016) Epigenetic mutations can both help and hinder adaptive evolution. *Mol Ecol* 25(8):1856–1868.

50. Pigliucci M, Murren CJ, Schlichting CD (2006) Phenotypic plasticity and evolution by genetic assimilation. *J Exp Biol* 209(Pt 12):2362–2367.

51. Tang X-S, et al. (1994) Identification of histidine at the catalytic site of the photosynthetic oxygen-evolving complex. *Proc Natl Acad Sci USA* 91(2):704–708.

52. Onidas D, Stachnik JM, Brucker S, Krätzig S, Gerwert K (2010) Histidine is involved in coupling proton uptake to electron transfer in photosynthetic proteins. *Eur J Cell Biol* 89(12):983–989.

53. Walworth NG, et al. (2016) Mechanisms of increased *Trichodesmium* fitness under iron and phosphorus co-limitation in the present and future ocean. *Nat Commun* 7:12081.

54. Reusch TBH (2014) Climate change in the oceans: Evolutionary versus phenotypically plastic responses of marine animals and plants. *Evol Appl* 7(1):104–122.

55. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.

56. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1): 139–140.

57. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R (2015) InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16(1):169.

58. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B* 57(1):289–300.

59. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460–2461.

60. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.

61. Prufert-Bebout L, Paerl HW, Lassen C (1993) Growth, nitrogen fixation, and spectral attenuation in cultivated *Trichodesmium* species. *Appl Environ Microbiol* 59(5): 1367–1375.

62. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11(3):R25.

63. Nookaew I, et al. (2012) A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 40(20): 10084–10097.