Supporting Information

Allen *et al.* 10.1073/pnas.0711370105

SI Methods

Culture and Growth Conditions. For the construction of the cDNA library from which ESTs were generated, RNA was extracted from cultures acclimated to 5 nM total Fe for several generations in 1-liter polycarbonate bottles and finally grown in an 8-liter bottle to mid-exponential phase at a temperature of 18°C and a 16:8 light (150 $\mu E \cdot m^{-1} \cdot sec^{-1}$):dark regime. Additional physiological parameters were measured in parallel for 1L Fe-limited and Fe-replete cultures. These included growth rate, cell size, carbon fixation, chlorophyll, flavodoxin (1), ferric reductase activity (2), nonphotochemical quenching (3), respiration and alternative oxidase activity, and spectroscopic determination of photosystem and electron carrier concentrations. Fe′, the sum of all unchelated species, calculated according to the equations of (4) was ($Fe' = 0.026\ Fe_T$) under our conditions. Therefore, the Fe limiting condition had an Fe′ level of 13.4 $pmol \cdot liters^{-1}$ Fe′ and the Fe replete condition had an Fe′ value of 2.6 $nmol \cdot liters^{-1}$ Fe′. Cells were grown in semicontinuous batch mode and kept in steady-state with respect to Fe availability. To achieve and maintain desired Fe levels, Fe was pre-complexed with EDTA (1:1.1 mol:mol) and added to Fe-free media. Thus, the availability of free Fe ions in the culture media was effectively buffered to maintain consistent Fe′ levels over the course of the experiments. The concentration of kinetically labile inorganic species is known to be proportional to the free ion concentration (5, 6).

Chlorophyll Measurements. Ten to 50 ml of culture were filtered on 0.2 μm GF/F filters, chlorophyll was extracted in 10 ml of 90% acetone, using 1:1 2 mm/4 mm glass beads. Chlorophyll concentration was measured with a Turner fluorometer.

Variable Fluorescence. Variable fluorescence Fv/Fm was measured after 5 min of dark adaptation with a PhytoPAM (WALTZ) and analyzed with PhytoWin software (7).

Nonphotochemical Quenching (NPQ). NPQ was calculated as (Fm − Fm′)/Fm′, where Fm is the maximum fluorescence yield (i.e., in dark adapted samples) in the dark, and Fm′ is the maximum fluorescence emission in steady state light conditions. Fm and Fm′ were determined by exposure of the cell suspension to a saturating light pulse (intensity: 3,000 μE; duration: 150 ms, optical pathway of the cuvette: 1 mm). Actinic light was 1,000 μE. Measurements were performed with a laboratory-built fluorometer, as described in ref. 3.

Spectroscopy. Spectroscopic measurements were performed with a LED-based spectrophotometer having a time resolution of 10 μs (JTS 10; Biologic). PSI and PSII ratios were calculated from the extent of the ECS signal at 520–490 nm upon excitation of cells with a single turnover saturating flash. This signal is proportional to reaction center photochemistry (8). Thus, the generation of a transmembrane potential following a single turnover flash reflects the turnover of the different photosynthetic complexes. The fast phase, completed in $<1\ \mu s$ (phase a) is proportional to the charge separation performed by PSI and PSII. The slower rising signal is attributed to cytochrome b_6f turnover, whereas the slow relaxation phase is linked to H^+ flux through the ATP synthase (9). PSII contribution was deduced as the difference between the signals measured in the absence and presence of the PSII inhibitors DCMU and HA (10). The latter compound was added to destroy the manganese cluster responsible for oxygen evolution and to prevent recombination between the donor and acceptor side of PSII, which would preclude correct estimation of the PSI/PSII ratio.

The same setup was used to assess the amount of functional cytochrome c_6 and b_6f complexes. Cyt c6/P700 ratios were assessed by comparing the extent of c_6 oxidation in continuous light in the presence of the plastoquinone analogue DBMIB. This compound blocks cyt b_6f turnover, allowing full oxidation of the cytochrome c6 pool. This amplitude was normalised on the signal corresponding to oxidation of P700 under the same conditions. Cytochrome b_6f complex was estimated from the signal corresponding to the maximum oxidation of cytochrome b_6. Cytochrome c_6 redox changes were calculated as the difference between the absorption at 554 nm and a baseline drawn between 545 and 573 nm. Cytochrome b_6 signals at 564 minus the same baseline. P700 was computed as $\Delta I/I_{\ P700} = \Delta I/I$ 820 nm $-$ 0.8 $\times$ $\Delta I/I$ 870 nm (11). An extinction coefficient of 17.4 $mM \cdot cm^{-1}$ (12) was used to evaluate PSI concentration starting from P700 redox change signals.

Cell Volume. Cell volumes of fusiform cells were calculated from mean length and diameter measured under the light microscope, using a double frustum as a simplified model of the fusiform cell.

RNA Extraction for qRT-PCR Analysis. 2–4 $\times$ 10^7 cells were harvested by 2 $\times$ 10 min centrifugations at 4°C. Cell pellets were used for extraction of total RNA, using a Plant RNeasy Kit (Qiagen). RNA was eluted from the columns with 50 μl of RNase-free water and stored at −20°C. cDNA was synthesized with a Quantitec reverse transcription kit, after treatment with the g-DNA whipeout buffer to remove contaminating genomic DNA. qRTPCR was performed with an ABI 700 instrument, and ABI SyBr Green master mix in 25 μl of total reaction volume containing 5 μl of RNA sample, and 10 pmol of each of gene-specific forward and reverse primers, cycled according to the manufacturers default program of 10 min at 95°C, followed by 40 cycles between 15 seconds at 95°C and 1 min at 60°C followed by a dissociation curve measurement for each gene-specific primer pair. The efficiency of amplification for each gene was calculated from a 10-fold dilution series of template cDNA, and was >95%. RNA and cDNA samples were run in parallel assays with 18S rRNA primers to confirm that the RNA samples were free of contaminating genomic DNA. Expression levels for target genes in each culture were normalized to 18S rRNA expression.

All pairs of qPCR primers are given in Table S1.

^{14}C-Fixation. ^{14}C fixation experiments were carried out with cells from late exponential phase supplied with 1 μM or 5 nM Fe, respectively. Cell densities were 930,000 (1 μM at 5 days) and 385,000 (5 nM at 8 days) cells per ml. Culture (40 ml) was incubated for 7 h at 150 μE with 2 μCurie of $NaH^{14}CO_3$. Culture (8 ml) was filtered on cellulose acetate filters in triplicates, filters were put in vials and supplied with 5 ml of Lumagel. Vials were incubated for 1 day and activity was measured with a scintillation counter.

Dark Respiration and Oxygen Evolution. Respiration and O_2-evolution were measured with a Clark Pt/Ag-electrode (HansaTech). Cells were concentrated 50-fold to $\approx 5 \times 10^7$ cells per ml, resuspended in fresh medium and measured at 20°C. Respiration was measured with dark adapted cells, and photosyn-

thetic O_2-evolution under illumination of 150 μE. The photosynthetic quotient (moles of O_2 evolved per moles of CO_2 assimilated) for Fe replete and limited cultures was consistent with theoretical values.

Alternative Oxidase (AOX) Assay. AOX activity was estimated from respiration measurements in which the cytochrome inhibitor antimycin A (AA) (1 μM) or the AOX inhibitor salicyl hydroxamic acid (SHAM) (1 mM) were added independently to different sub-samples of the same culture. Subsequently, both inhibitors were added to the same sample to verify the same level of inhibition.

Metabolite Extraction and GC-MS Analysis. Approximately 1×10^9 cells were pelleted and used for metabolite analysis. Metabolites were extracted by the addition of 600 μL methanol and 15 μL (2 mg/ml) ribitol, shock freezing in liquid nitrogen, subsequent vortexing and shock freezing. This was followed by the addition of 250 μL dH2O and 250 μL $CHCl_3$, incubation of the samples at 70°C for 10 min, and centrifugation for 10 min at 14,000 rpm. 50 μL of the supernatant was transferred into a new reaction tube and dried for GC-MS analysis. Derivatization and GC-MS analysis were carried out as described in ref. 13. The GC-MS system comprised a CTC CombiPAL autosampler, an Agilent 6890N gas chromatograph and a LECO Pegasus III TOF-MS running in EI+ mode. Metabolites were identified compared with database entries of authentic MSRI libraries (14). Metabolites were normalized internally to the added standard Ribitol and quantified relative to total cellular protein levels (15). Recovery and recombination experiments confirmed the validity of this protocol for both qualitative and quantitative analyses.

EST Library Construction, Sequencing, Annotation, and Comparative Analysis. cDNA libraries were constructed from poly$(A)^+$ RNA, using the CloneMiner cDNA library construction Kit (Invitrogen) following the supplier's instructions. Sequencing was performed mostly from the 5′ end of the insert but for some of the libraries, an attempt was made to sequence each clone at both the 5′ and the 3′ ends. When both EST reads overlapped, the two sequences were fused into a consensus sequence, using PHRAP (www.phrap.org). Approximately 110,000 ESTs were sequenced from cells grown in 12 different growth conditions. ESTs were checked for ambiguities, vector trimmed, and quality checked and clustered by using CAP3 software. Approximately 8,000 contigs and 5,000 singletons were obtained by this method. The frequency of expressed genes under different conditions was calculated for each contig. In cases where 5′ and 3′ sequencing was performed for the same clone, the EST was only counted once for overall frequency computation. The log-likelihood ratio (*R* value) was computed to test for statistical significance of differential gene expression by comparing to randomized datasets to identify transcripts significantly over-represented in a particular library. This method was designed specifically for the purpose of statistically evaluating gene expression level across multiple cDNA libraries (16). Log-likelihood ratios calculated for the actual data and 1,000 randomized datasets indicated that a log-likelihood ratio >12 for a given contig can be considered as statistically significant at the 98% confidence level. Contigs with transcript representation in the Fe-limited library were ranked based on decreasing order of EST frequency. Contigs were catalogued as differentially expressed if they met the criteria of (*i*) having a log-likelihood ratio >12, (*ii*) were first or second most frequent in the Fe-limited library, and (*iii*) had a minimum contig size of at least 3 ESTs. 228 contigs were identified as being differentially up-regulated in the Fe-limited treatment. These contigs were mapped to the genome and the 228 contigs were assembled to 212 predicted proteins. Genome browser protein IDs for all 212 proteins can be found in Table S3. R values and EST frequency across all 12 libraries can be found by Blast-searching the PtDB EST database (www.biologie.ens.fr/diatomics/EST). The EST database can also be queried with the EST contig IDs found in Table S1. The set of 212 proteins identified through analysis of the EST data were used for the further comparative analysis against different taxonomical lineages.

Representative proteomes were compiled from 14 completed genomes and the public data repository of National Center for Biotechnology Information and used for comparative analysis. Three major taxonomical groups, heterokonts (3 genomes), plants and algae (5 genomes), and opisthokonts (6 genomes), were represented. Combined proteomes from *Thalassiosira pseudonana*, *Phytophthora sojae*, and *Phytophthora ramorum* were compiled to represent heterokonts. Complete proteomes of *Cyanidioschyzon merolae*, *Ostreococcus lucimarinus*, *Ostreococcus tauri*, *Chlamydomonas reinhardtii*, and the Chlorophyta protein sequences from National Center for Biotechnology Information were compiled to represent green and red algae. *Arabidopsis thaliana* and Streptophyta proteins from National Center for Biotechnology Information were combined to form a vascular plant dataset. Complete proteomes encoded by the genomes of *Monosiga brevicollis*, *Ciona intestinalis*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Aspergillus fumigatus*, and *Pichia stipitis* were compiled with amino acid sequences from National Center for Biotechnology Information representing opisthokonts to form an opisthikont proteome dataset. The heterokont, green and red algae, and plant proteomes were compared with the 212 proteins overexpressed in the Fe-limited condition, using the Blastp algorithm. Proteins with an E value $<10^{-5}$, coverage of >50 percent, and >30 percent overall identity were considered as homologous. *P. tricornutum* proteins that were found to be absent from other genomes were further compared against an un-published *Pseudo-nitzchia* sp. (a pennate diatom) EST database. Comparative data is found within Table S3. The 212 predicted proteins corresponding to the up-regulated transcripts were functionally annotated through a combination of BlastP (Swissprot) results, manual inspection of alignments, InterPro domain scanning results, and the occurrence of SignalP, TargetP, ChloroP, and TMHMM targeting and transmembrane signatures. Proteins determined to have signal peptides were subsequently examined for plastid targeting by removal of the signal peptide and evaluation by TargetP, ChloroP, and manual inspection (17). Occasionally, the final filtered gene model (ID provided in Table S1) was truncated and the EST extended model was more appropriate for 5′ targeting predictions. BlastP scores, hits against Swissprot, InterPro domains, targeting prediction information, EST contig IDs, and genome browser protein IDs are given in Table S3 for proteins identified through EST sequencing and microarray experiments that we were able to annotate.

Microarray Fabrication, Hybridization, and Data Processing. The array platform with all of the probe sequences and corresponding genome browser protein IDs and all of the normalized hybridization data is available in Minimum Information About a Microarray Experiment (MIAME) compliant format at the National Center for Biotechnology Information Gene Expression Omnibus (GEO) website, www.ncbi.nlm.nih.gov/projects/geo. GEO Accession GSE8675 can be viewed at www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=txohrwusmukyqpw&acc=GSE8675. Oligonucleotetide probes (701) were designed based primarily on EST sequence data from a previous *P. tricornutum* EST sequencing project of ≈17,000 ESTs (18). A very limited number of probes on the microarray were derived from early and preliminary Fe-limited EST and *P. tricornutum* genome sequence data. Spotting of the microarray slides was performed in duplicate on each slide by the Genomic Service from the Biology

Department (SGDB) at the Ecole Normale Supérieure, Paris. Technical details of the spotting procedure used by the SGDB can be found at www.transcriptome.ens.fr/sgdb.

Duplicate cultures of Fe-replete and Fe-limited *P. tricornutum* cells were grown and RNA was isolated as described above. An on-column digestion of DNA was performed (Qiagen) during an RNA cleanup step (Qiagen). Purified RNA was prepared separately for each of the culture replicates. RNA quality and quantity was inspected on a 2100 Agilent Bioanalyzer and NanoDrop (ND-1000), respectively. Using the Ambion Amino Allyl MessageAmp kit (Ambion), 500 ng of total RNA was used as a template for first strand cDNA synthesis and conversion to double-stranded DNA (dsDNA). dsDNA was purified and *in vitro* transcribed with amino allyl modified UTPs (aaUTP) to generate amplified RNA (aRNA). aRNA was purified and quantified, and 10 μg of amino allyl RNA was vacuum-dried and resuspended in 9 μL of dye coupling buffer. Dye couple reactions were performed and efficiency was evaluated spectrophotometrically on a ND-1000. Samples that contained 25–30 incorporated dye molecules per 1,000 nt were considered suitable for competitive hybridization.

Before hybridization 5 μg of labeled aRNA made from Fe-replete and Fe-limited samples were mixed together, precipitated, resuspended, and fragmented (Ambion). Each pair of RNA samples from Fe-replete and Fe-limited duplicate cultures was competitively hybridized three times for a total of six slides. One slide was a dye-flip hybridization. Microarray slides were scanned by using a GenePix4000B (Axon) two-laser scanner and image analysis was performed by using GenePix 5.0 (Axon) software. Mean intensities for both RNA populations were processed through minimum and maximum saturation and diameter filters and normalized by using global lowess followed by a block mean normalization. Image data processing and normalization was performed by using the SGDB GOULPHAR software package http://transcriptome.ens.fr/goulphar (19). To assign *P* values and evaluate statistical significance, normalized log2ratios were loaded into the Multi experiment Viewer (MeV) software package for analysis of microarray data, www.tm4.org/index.html (20). Normalized log2ratios for each of the duplicate probes on each slide were averaged. Averaged log2 ratios of the six slides were converted into two tab delimited multi sample (TDMS) text files with three slides each; each set of three corresponded to one of the duplicate culture pairs. A one class T test was performed on each group of 3 slides. P values were calculated based on *t* distributions with an overall critical value of 0.05. If the *P* value of a given gene was less than or equal to 0.05 across all of the hybridizations from each of the replicate culture pairs then the change in expression between the two treatments was considered significant. *P* values for each set of three hybridizations are provided in Table S1.

1. La Roche J, Murray H, Orellana M, Newton J (1995) Flavodoxin expression as an indicator of iron limitation in marine diatoms. *J Phycol*. 31:520–530.
2. Eckhardt U, Buckhout T (1998) Iron assimilation in Chlamydomonas reinhardtii involves ferric reduction and is similar to Strategy I higher plants. *J Exp Bot* 49:1219–1226.
3. Finazzi G, *et al.* (2004) A zeaxanthin-independent nonphotochemical quenching mechanism localized in the photosystem II core complex. *Proc Natl Acad Sci USA* 101:12375–12380.
4. Sunda W, Huntsman S (2003) Effect of pH, light, and temperature on Fe-EDTA chelation and Fe hydrolysis in seawater. *Mar Chem* 84:35–47.
5. Hudson RJM, Morel FMM (1990) Iron transport in marine phytoplankton: Kinetics of cellular and medium coordination reactions. *Limnol Oceanogr* 35:1002–1020.
6. Hudson RJM, Morel FMM (1993) Trace metal transport by marine microorganisms: Implications of metal coordination kinetics. *Deep-Sea Res Part I Oceanogr Res Pap* 40:129–151.
7. Schreiber H (1998) in *Photosynthesis: Mechanisms and Effects*, ed Garab G (Kluwer Academic, Amsterdam), Vol V pp 4253–4258.
8. Witt HT (1979) Energy conversion in the functional membrane of photosynthesis. Analysis by light pulse and electric pulse methods. The central role of the electric field. *Biochimm Biophys Acta* 505:355–427.
9. Joliot P, Delosme R (1974) Flash-induced 519 nm absorption change in green algae. *Biochim Biophys Acta* 357:267–284.
10. Bennoun P (1970) Reoxidation of the fluorescence quencher ''Q'' in the presence of 3-(3,4-dichlorophenyl)-1,1-dimethylurea. *Biochim Biophys Acta* 216:357–363.
11. Joliot P, Joliot A (2005) Quantification of cyclic and linear flows in plants. *Proc Natl Acad Sci USA* 102:4913–4918.
12. Joliot P, Joliot A (1988) The low potential electron transfer chain in the cytochrome b6f complex *Biochim Biophys Acta* 933:319–333.
13. Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatogrephy mass spectrometry-based metabolite profiling in plants. *Nat Protoc* 1:387–396.
14. Schauer N, *et al.* (2005) GC-MS libraies for the rapid indentification of metabolites in complex biological samples. *FEBS Lett*. 579:1332–1337.
15. Bradford M (1976) A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye-binding. *Anal Biochem* 72:248–254.
16. Stekel DJ, Git Y, Falciani F (2000) The comparison of gene expression from multiple cDNA libraries. *Genome Res* 10:2055–2061.
17. Kilian O, Kroth PG (2005) Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *Plant J* 41:175–183.
18. Maheswari U, *et al.* (2005) The diatom EST database. *Nucleic Acids Res* 33:344–347.
19. Lemoine S, Combes F, Servant N, Le Crom S (2006) Goulphar: Rapid access and expertise for standard two-color microarray normalization methods. *BMC Bioinformatics*. 7:467.
20. Saeed AI, *et al.* (2003) TM4: a free open-source system for microarray data management and analysis. *BioTechniques*. 34:374–378.

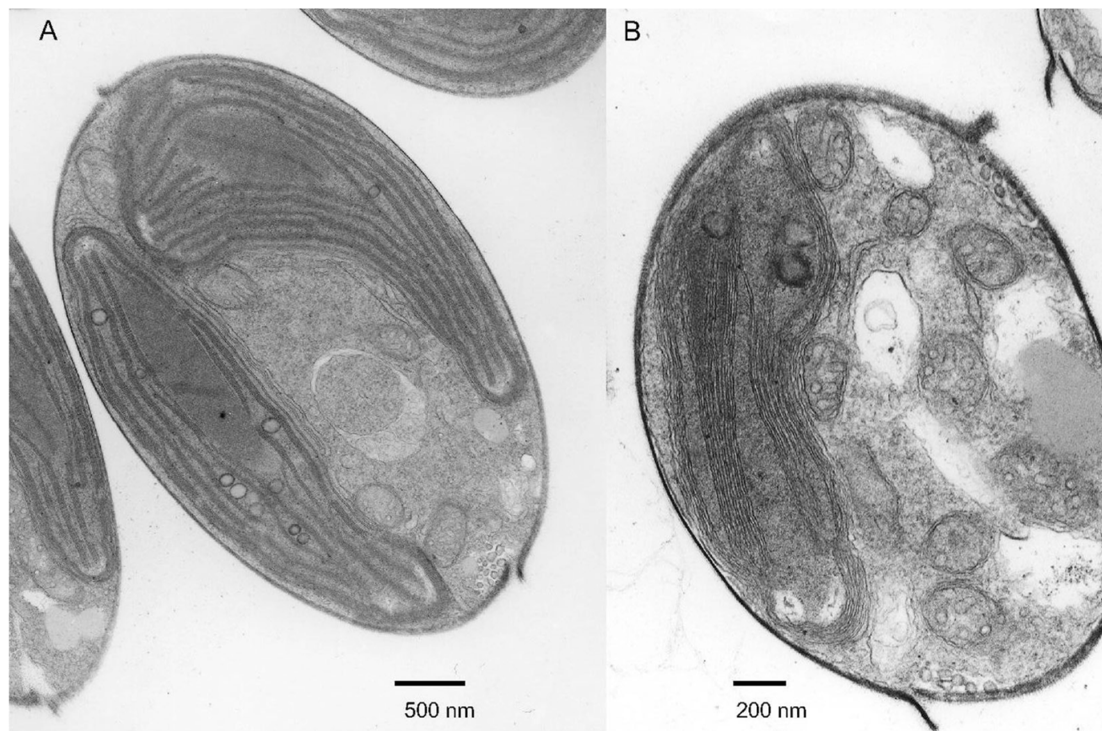

Fig. S1. Electron micrographs of Fe-replete (*A*) and Fe limited (*B*) *P. tricornutum* cells, as detailed in Table 1. (Scale bars: *A*, 500 nm; *B*, 200 nm).

Fig. S2. Pathway map of primary metabolism. This representation of metabolism reflects metabolic changes in Fe limited compared to Fe-replete cells. Data were log2 scaled and color coded. Green boxes represent an increase in metabolite content in Fe limited cells relative to the replete control, while red boxes represent a decrease in metabolite content relative to the replete control. The shading from light to dark intensities reflects the level of metabolite changes from small to large. Metabolites in gray were not detected in this analysis.

Other Supporting Information Files

Table S1
Table S2
Table S3