AGU PUBLICATIONS

[Global Biogeochemical Cycles]

Community-Level Responses to Iron Availability in Open Ocean Planktonic Ecosystems

Luigi Caputi[†1], Quentin Carradec[†2, 3, 4], Damien Eveillard[†5], Amos Kirilovsky[†6, 7], Eric Pelletier[†2, 3, 4], Juan J. Pierella Karlusich[†6], Fabio Rocha Jimenez Vieira[†6], Emilie Villar[†6, 8], Samuel Chaffron[5], Shruti Malviya[6, 9], Eleonora Scalco[1], Silvia G. Acinas[10], Adriana Alberti[2], Jean-Marc Aury[2], Anne-Sophie Benoiston[6, 11], Alexis Bertrand[2], Tristan Biard[12, 13], Lucie Bittner[6, 8, 11], Martine Boccara[6], Jennifer R. Brum[14], Christophe Brunet[1], Greta Busseni[1], Anna Carratalà[15], Hervé Claustre[13], Luis Pedro Coelho[16], Sébastien Colin[6, 8], Salvatore D'Aniello[1], Corinne Da Silva[2], Marianna Del Core[17], Hugo Doré[8], Stéphane Gasparini[13], Florian Kokoszka[1, 6, 18], Jean-Louis Jamet[19, 20], Christophe Lejeusne[1, 51], Cyrille Lepoivre[21], Magali Lescot[22], Gipsi Lima-Mendez[23, 26], Fabien Lombard[13], Julius Lukeš[27, 28], Nicolas Maillet[1, 29], Mohammed-Amin Madoui[2, 3, 4], Elodie Martinez[30], MariaGrazia Mazzocchi[1], Mario B. Néou[2, 3, 4], Javier Paz-Yepes[6], Julie Poulain[2], Simon Ramondenc[13], Jean-Baptiste Romagnan[13], Simon Roux[14], Daniela Salvagio Manta[17], Remo Sanges[1], Sabrina Speich[18], Mario Sprovieri[17], Shinichi Sunagawa[16, 31], Vincent Taillander[13], Atsuko Tanaka[6], Leila Tirichine[6], Camille Trottier[32], Julia Uitz[13], Alaguraj Veluchamy[6, 33], Jana Veselá[27], Flora Vincent[6], Sheree Yau[34], Stefanie Kandels-Lewis[16, 35], Sarah Searson[13], Céline Dimier[6, 8], Marc Picheral[13], *Tara* Oceans Coordinators, Peer Bork[35], Emmanuel Boss[36], Colomban de Vargas[8], Michael J. Follows[37], Nigel Grimsley[34], Lionel Guidi[13, 38], Pascal Hingamp[22], Eric Karsenti[6, 35], Paolo Sordino[1], Lars Stemmann[13], Matthew B. Sullivan[14], Alessandro Tagliabue[39], Adriana Zingone[1], Laurence Garczarek[8], Fabrizio d'Ortenzio[13], Pierre Testor[40], Fabrice Not[8], Maurizio Ribera d'Alcalà[*1], Patrick Wincker[*2, 3, 4], Chris Bowler[$*6], Daniele Iudicone[$*1]

Tara Oceans Coordinators: Silvia G. Acinas[10], Peer Bork[35, 41, 42, 43], Emmanuel Boss[36], Chris Bowler[$*6], Colomban de Vargas[8], Michael J. Follows[37], Gabriel Gorsky[13],

Nigel Grimsley[34], Pascal Hingamp[22], Daniele Iudicone[$*1], Olivier Jaillon[2, 3], Stefanie Kandels-Lewis[16, 35], Lee Karp-Boss[36], Eric Karsenti[6, 35], Uros Krzic[44], Fabrice Not[8], Hiroyuki Ogata[45], Stéphane Pesant[46, 47], Jeroen Raes[23], Emmanuel G. Reynaud[48], Christian Sardet[13], Mike Sieracki[49, 50], Sabrina Speich[18], Lars Stemmann[13], Matthew B. Sullivan[14], Shinichi Sunagawa[16, 31], Didier Velayoudon[51], Jean Weissenbach[2, 3, 4], Patrick Wincker[*2, 3, 4]

[1]: Stazione Zoologica Anton Dohrn, Villa Comunale, 80121, Naples, Italy. [2]: CEA - Institut François Jacob, Genoscope, 2 rue Gaston Crémieux, 91057 Evry, France. [3]: CNRS UMR 8030, Evry, France. [4]: Université d'Evry Val d'Essonne, Université Paris-Saclay, Evry, France. [5]: Laboratoire des Sciences du Numérique de Nantes (LS2N) – CNRS, Université de Nantes, École Centrale de Nantes, IMT Atlantique, 2 rue de la Houssinière, 44322 Nantes, France. [6]: Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, PSL Université Paris 75005 Paris, France. [7]: INSERM, UMRS1138, Laboratory of Integrative Cancer Immunology, Centre de Recherche des Cordeliers, 15 Rue de l'Ecole de Medecine, 75006 Paris, France. [8]: CNRS, UMR 7144, Station Biology de Roof, Place Georges Teissier, 29680 Roscoff, France. [9]: Simons Centre for the Study of Living Machines, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India. [10]: Department of Marine Biology and Oceanography, Institute of Marine Sciences (ICM)-CSIC, Pg. Marítim de la Barceloneta 37-49, Barcelona E0800, Spain. [11]: Sorbonne Universités, UPMC Univ Paris 06, Univ Antilles Guyane, Univ Nice Sophia Antipolis, CNRS, Evolution Paris Seine - Institut de Biologie Paris Seine (EPS - IBPS), 75005 Paris, France. [12]: Sorbonne Universités, UPMC Université Paris 06, CNRS, Laboratoire Adaptation et Diversitéen Milieu Marin UMR7144, Station Biologique de Roscoff, 29688 Roscoff, France. [13]: Sorbonne Universités, UPMC Université Paris 06, CNRS, Laboratoire d'Océanographie de Villefranche (LOV) UMR7093, Observatoire Océanologique, 06230 Villefranche-sur-Mer, France. [14:] Departments of Microbiology and Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus OH 43210, USA. [15]: Laboratory of Environmental Chemistry, School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. [16]: Structural and Computational

Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany. [17]: Institute for Coastal Marine Environment (IAMC) - CNR of Capo Granitola, Via del Mare 3, Torretta Granitola (TP), 91021, Italy. [18]: LMD Laboratoire de météorologiedynamique. Ecole normale supérieure de Paris, PSL Research University, 24 rue Lhomond, 75231 Paris Cedex 05, France. [19]: Université de Toulon, Laboratoire PROTEE-EBMA E. A. 3819, BP 20132, 83957 La GardeCedex, France. [20]: Université de Toulon, Aix Marseille Universités, CNRS/INSU, IRD, MIO UM 110 Mediterranean Institut of Oceanography, La Garde, France. [21]: Information Génomique et Structurale, UMR7256, CNRS, Aix-Marseille Université, Institut de Microbiologie de la Méditerranée (FR3479), ParcScientifique de Luminy, Marseille, France. [22]: Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France. [23]: Department of Microbiology and Immunology, Rega Institute, KU Leuven, Herestraat 49, 3000 Leuven, Belgium. [24]: Center for the Biology of Disease, VIB, Herestraat 49, 3000 Leuven, Belgium. [25]: Cellular and Molecular Microbiology , Faculté des Sciences, Université Libre de Bruxelles (ULB), Belgium. [26]: Interuniversity Institute for Bioinformatics in Brussels (IB2), ULB-VUB, Boulevard du Triomphe CP 263, 1050 Brussels, Belgium. [27]: Institute of Parasitology, Biology Centre CAS, Branišovská 1160/31, 370 05 ČeskéBudějovice, Czech Republic. [28]: Faculty of Science, University of South Bohemia, 370 05 Ceské Budejovice, Czech Republic. [29]: Institut Pasteur - Bioinformatics and Biostatistics Hub - C3BI, USR 3756 IP CNRS - Paris, France. [30]: IRD - BP 529 - 98713 Papeete, Tahiti. [31]: Institute of Microbiology, Department of Biology, ETH Zurich, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland. [32]: IFREMER, Physiology and Biotechnology of Algae Laboratory, rue de l'Iled'Yeu, 44311, Nantes, France. [33]: Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. [34]: Sorbonne Universités, UPMC Univ Paris 06, CNRS, Biologie Intégrative des Organismes Marins (BIOM, UMR 7232), Observatoire Océanologique, Banyuls sur Mer, France. [35]: Directors' Research European Molecular Biology Laboratory Meyerhofstr. 1, 69117 Heidelberg, Germany. [36]: School of Marine Sciences, University of Maine, Orono, Maine, USA. Dept of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [37]: Dept of Earth, Atmospheric

and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [38]: Department of Oceanography, University of Hawaii, Honolulu, Hawaii 96822, USA. [39]: Department of Earth Ocean and Ecological Sciences, School of Environmental Sciences, University of Liverpool, Liverpool, UK. [40]: Sorbonne Universités (UPMC, Univ Paris 06)-CNRS-IRD-MNHN, Laboratoire LOCEAN, Paris, France. [41]: Structural and Computational Biology, European Molecular Biology Laboratory, Meyerhofstr. 1, 69117 Heidelberg, Germany. [42]: Max Delbrück Centre for Molecular Medicine, 13125 Berlin, Germany. [43]: Department of Bioinformatics, Biocenter, University of Würzburg, 97074 Würzburg, Germany. [44]: Cell Biology and Biophysics, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. [45]: Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan. [46]: MARUM, Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany. [47]: PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, Bremen, Germany. [48]: Earth Institute, University College Dublin, Dublin, Ireland. [49] National Science Foundation, Arlington, VA 22230, USA. [50]: Bigelow Laboratory for Ocean Sciences East Boothbay, ME, USA. [51]: DVIP Consulting, Sèvres, France. [51:] Institut Méditerranéen de Biodiversité et d'Ecologie Marine et Continentale (IMBE), UMR 7263 CNRS, IRD, Aix Marseille Université, Avignon Université, Station Marine d'Endoume, Marseille, France.

* Correspondence to: pwincker@genoscope. cns. fr, cbowler@biologie. ens. fr, iudicone@szn. it,
maurizio@szn. it
† These authors contributed equally to this work
$ Equal coordinating contribution

Contents of this file

S4.c Diatom responses

S4.d Transcriptional response of metazoan communities

Figure S1: WGCNA analysis of eukaryotic plankton.

Figure S2: Abundance and expression of iron-responsive genes in the five major groups of eukaryotic phytoplankton

Figure S3: Abundance and expression of iron-responsive genes in the diatom genera *Thalassiosira* and *Pseudo-nitzschia* in surface samples from 5-20 μm size fractions

Figure S4: Oceanographic context of Marquesas archipelago sampling sites.

Figure S5. Changes in iron-associated genes in phytoplankton in Marquesas Islands sampling stations.

Figure S6 Changes in iron-associated genes in zooplankton in Marquesas Islands sampling stations.

Figure S7: Responses of diatom genera and iron-responsive genes in the Marquesas archipelago.

Table S1: Trace metals, IAA statistics and iron-related gene expression.

Table S2: Oceanographic data collected at the Marquesas Islands.

Table S3: Response of plankton community at the Marquesas Islands.

Additional Supporting Information (Files uploaded separately)

Table S1: Trace metals, I**AA statistics, and iron-related gene expression.** (a) Iron concentration estimates at Tara Oceans samplig station**s.** Table shows compilation of PISCES2 estimates, ECCO2-DARWIN estimates, and observed data. In bold: Tara Oceans sampling stations and depths used for the WGCNA analyses. Observed data: Iron concentrations measured at sampling stations which are the closest to the Tara Oceans stations. Only stations within a range of 2 degrees were considered to be sufficiently reliable to make a comparison with model-derived data. (b) Phytoplankton C to Chl a (Phyto C: Chl a) ratio at different depth extrapolated from Campbell et al., 1994. (c) Analysis of association of modeled iron concentrations with diatom *ISIP* gene abundance and expression. Correlation, nonlinear regression and rank between diatom *ISIP* abundances in the metagenome and metatranscriptome with ECCO2-DARWIN and

PISCES2-derived estimates of iron concentration are shown. (d) Characteristics of lineages associated with modules described extensively in text. Iron Associated Assemblages: Lineage, ECCO-2-DARWIN VIP scores and correlation to iron, PISCES2 VIP scores and correlation to iron, centrality. For selected groups, VIP scores and iron correlation median values are given. DarkGrey and Red modules: ECCO-2-DARWIN VIP scores and correlation to iron, PISCES2 VIP scores and correlation to iron, centrality. (e) Relative contribution of the main taxonomic groups in each module of interest. (f) Relative weight of the modules of interest in terms of OTU abundance per selected sampling station (top), and aggregated per macro-region (bottom). (g) Description of prokaryotic genes within the prokaryotic gene modules most correlated with iron concentrations. For each module associated with iron (a: SaddleBrown, b: Grey60, c: Plum1, d: Red, and e: SkyBlue), a table summarizes the putative functions of genes within it, their VIP scores, correlations to iron, and whether they were previously defined as being core or non-core in the Tara Oceans Microbial Reference Gene Catalog. (h) Analysis of *ISIP* and *SIT* gene expression. I) Distribution of ISIP genes from the eukaryote gene catalog in each of the five main groups of autotrophic eukaryotes. II) Partial correlation analysis of the effect of silicate and iron concentrations in the five clades of diatom *SIT* genes retrieved from the eukaryote gene catalog. Partial correlation coefficients (Spearman) and p-values are shown for the correlation of silicate concentration (measured in situ by *Tara* Oceans or retrieved from World Ocean Atlas 2013) and the SIT gene abundance or expression (normalized to total diatom counts), correcting for the effect of the iron concentration (PISCES2-derived estimates) (see Supplementary Text S1). The partial correlations of iron levels and the SIT gene abundance or expression are also shown, correcting for the effect of silicate levels. The analysis was repeated without considering stations TARA_84 and TARA_85 due to their extreme silicate values, but essentially the same results were obtained.

Table S2: **Oceanographic data collected at the Marquesas Islands.** (a) Summary of main oceanographic features at Stations TARA_122, TARA_123, TARA_124 and TARA_125. (b) Dissolved trace metal concentrations (nMol/L) in sampling stations TARA_122, TARA_123, TARA_124 and TARA_125. (c) PANGAEA and ENA sample IDs for (upper panel) Sample Methodological Context, (middle panel) Sample

Environmental Context and (bottom panel) Sample Number of Barcodes and Reads. (d) Summary of *Tara* Oceans sampling stations in the Marquesas archipelago. Upper panel: Sampling station dates and positions, with sampling depths for SRF, DCM, and MESO. Bottom panel: Summary of particle characteristics and carbon flux estimated in three layers (0-100 m, 100-150 m and 400 m).

Table S3: **Response of plankton community at the Marquesas Islands.** (a) Carbon content to body length ratios (in selected zooplankton lineages) used to estimate zooplankton biomass in *Tara* Oceans sampling stations. (b) Biodiversity analysis of *Tara* Oceans sampling stations in the Marquesas archipelago. I) Relative abundance of OTUs from the main taxonomic groups estimated from metabarcoding data in five size fractions. II) Exponential Shannon and OTU richness for selected lineages and size classes. III) Phytoplankton relative abundance and biomass from microscopy counts. IV) Picocyanobacterial abundance as determined by recruitment of *petB* miTags and abundance of *Prochlorococcus* and *Synechococcus* as measured by flow cytometry or recruitment of *petB* miTags from the bacterial size fraction of *Tara* Oceans metagenomes (Sunagawa et al., 2015). V) Pearson correlations between *Prochlorococcus* and *Synechococcus* cells and phages, and abundance of *Prochlorococcus* and *Synechococcus* phages based on *psbA* gene abundance in viromes. Picocyanobacterial abundance was determined by recruitment of *petB* miTags and abundance of *Prochlorococcus* and *Synechococcus* as measured by flow cytometry or recruitment of *petB* miTags (Sunagawa et al., 2015). (In the latter case, per-strain read counts were normalized to the total number of reads at each station and aggregated at the genus level). VI) Composition of the Copepoda and Urochordata assemblages based on V9 metabarcodes (relative abundances). VII) Zooplankton composition, abundance and diversity, based on net (WP2) counts. VIII) Mesozooplankton biovolumes and abundances estimated from Zooscan data. (c) Mapping of metagenomic and metatranscriptomic reads assigned to *Prochlorococcus* and *Synechococcus* genes encoding components of iron metabolism.

Introduction

In the following sections, we provide additional global ocean analyses on phytoplankton iron-responsive genes as obtained from the *Tara* Oceans meta-omics datasets (S1), and by a more specific analysis of iron-responsive genes in two diatom genera known to be common in iron-limited regions (S1). We also characterize the oceanographic dynamics in the area of the Marquesas before, during, and after the period of the *Tara* Oceans sampling in August 2011 (S2). Subsequently, we report our detailed comparative analysis of the structure of the community at the four sampling stations in the Marquesas archipelago using both molecular and morphological information (S3). Finally, we describe the response of several components, from bacteria to planktonic metazoans, at the level of gene copy number and expression, together with the genotypic and functional inferences (S4).

S1. Additional information on iron-related genes in diatoms

S1. a The diatom iron-starvation induced proteins (ISIPs)

In diatoms the iron-starvation induced protein (ISIP) family is made up of ISIP1, ISIP2a and ISIP3, proteins of uncertain function whose transcript levels are induced strongly in response to iron limitation both in cultures and in the environment (Morrissey et al., 2015; Graff van Creveld et al., 2016; Marchetti et al., 2017). The best characterized member of this family is ISIP2a, which concentrates ferric (FeIII) ions at the cell surface and is believed to facilitate its uptake non-reductively (Marchetti et al., 2017). It has further been shown to be a functional ortholog of transferrin and has been renamed phytotransferrin (McQuaid et al., 2018). In the case of ISIP1, recent studies have shown that it is involved in the non-reductive endocytotic uptake of siderophores (Kazamia, et al., 2018). The *ISIP1* gene has a palindromic motif within its promoter region that is also present in other low-iron regulated genes in diatoms, and whose role in iron-dependent transcriptional regulation has been confirmed by promoter deletion analysis in *Phaeodactylum tricornutum* (Lommer et al., 2015; Malviya et al., 2016). In addition to transcriptional regulation, analysis of diatom genomes also suggests that environmental iron concentrations correlate with the presence/absence patterns or copy numbers of *ISIP*-encoding genes (Lommer et al., 2015). While many diatom genomes encode all three

members of the ISIP family (Morrissey et al., 2015), *ISIP1* and *ISIP3* have also been found in the genomes of haptophytes, and the Fea1-encoding domain present in ISIP2a was first described in chlorophytes (*Chlamydomonas reinhardtii*, but also in the *Prasinophyceae*) (Morrissey et al., 2015). Here we identified 5,630 putative *ISIP* genes from the eukaryote gene catalogue reported in (Alberti et al., 2017; Carradec et al., 2018) and found that at least one of the three genes is present in each of the five main photosynthetic groups (Table S1h). We further report an extensive analysis of their abundance and expression profiles with respect to modeled iron concentrations in the global ocean (Figures 1, 5, 6 and S2) and in the four *Tara* Oceans sampling sites in the Marquesas archipelago (Figures 9c-d, S5a-b and S7b-d).

S1. b The diatom silicon transporters (SITs)

To examine the interaction between iron and other nutrients in diatoms we focused on silicate metabolism because iron bioavailability has been found to play a role in silicon utilization in these organisms. More than 150 gene products potentially required for silicon biomanipulation have been identified in the radial centric species *Thalassiosira pseudonana* (Mock et al., 2008). Half of these genes are upregulated under either silicon or iron starvation, suggesting that iron and silicon pathways are linked. Silicon, in the form of silicic acid, is transported into the diatom cell from seawater by silicon transporter (SIT) proteins (Durkin et al., 2016). Five distinct clades have been distinguished, denoted A through E (Mock et al., 2008), but their regulation by iron and silicon in natural environments has not been examined in detail.

Here we retrieved the unigenes assigned to silicon transporter Pfam family (PF03842) from the eukaryote gene catalog (Alberti et al., 2017; Carradec et al., 2018) and made their classification within clades A-E by phylogenetic placement as in (Durkin et al., 2016). We then examined the abundance and expression of the *SIT* clades (normalized to total diatom gene abundance and expression, respectively) as a function of the modeled iron concentrations (PISCES2) at each of the sampling sites, using partial correlation analysis for correcting the effect of environmental silicate levels, which were measured at each *Tara* Oceans station. Interestingly, our results show that the influence of iron in the

297 patterns of adaptation and acclimation of *SIT* clades is much stronger than the influence
298 of silicate (Table S1h). The only statistically significant effect of silicate was detected in
299 the gene abundance variation of Clade B (partial correlation coefficient = -0. 43, *p*-value
300 = $8x10^{-4}$), which was the only clade not affected by iron concentrations. Regarding the
301 iron effect, Clades A and D are anticorrelated with this micronutrient in both
302 metagenomic and metatranscriptomic datasets, while Clade E is anticorrelated only in the
303 metatranscriptomics sequences (partial correlation coefficients between -0. 26 and -0. 33,
304 *p*-value < 0. 05). On the contrary, the abundance of Clade C genes is positively correlated
305 with iron levels (partial correlation coefficient = 0. 27, *p*-value = 0. 06).

306 In addition to the *in situ* silicate determinations by *Tara* Oceans, we also extracted from
307 the World Ocean Atlas 2013 the monthly mean silicate concentrations from the
308 corresponding date of sampling. Both silicate values show strong correlations (data not
309 shown) and generate similar results in the partial correlation analysis of silicate, iron and
310 *SIT* gene expression and abundance (Table S1h). Moreover, the patterns are essentially
311 the same whether we take into account or not the Southern Ocean stations TARA_84 and
312 TARA_85 with their extreme silicate values (16.550 and 79.520 μmol/l, respectively,
313 while the remaining stations with sequencing data have values below ~5 μmol/l).

314 In summary, we found different iron-correlated patterns of gene abundance and
315 transcription among the five distinct clades that constitute the *SIT* multigene family.
316 These results suggest that there is a strong interaction between iron and silicate in
317 diatoms and that the diversification of SITs has led to specialized adaptations to deal with
318 it.

319

320 S1. c Two prominent diatom genera display drastically different strategies to cope with
321 iron deficiency

322 Diatom species within the *Thalassiosira* and *Pseudo-nitzschia* genera are commonly
323 found in severely iron-limited regions, and their responses to changes in iron availability
324 have been characterized rather well (Thompson et al., 2011; Lommer et al., 2015; Cohen
325 et al., 2017). *Thalassiosira* is a representaive of centric diatoms whereas *Pseudo-nitzschia*
326 are pennate diatoms and both groups have different responses to environmental iron

levels. We therefore explored changes in iron-responsive gene content and mRNA levels in these two genera in 5-20 micron size fractions at a global scale.

Expression of *ISIP* genes in *Thalassiosira* and *Pseudo-nitzschia* are consistent with the global patterns from all diatoms (Figure S3b; compare with Figure 5 and Figure S2), although the trends for *Thalassiosira* are weak, in agreement with previous expression analyses in culture and deck incubation experiments in *P. granii* and *T. oceanica* (Chappell et al., 2015; Graff van Creveld et al., 2016; Cohen et al., 2017). In relation to gene abundances, the low signals do not permit us to make strong conclusions (Figure S3a).

The flavodoxin/ferredoxin switch shows differences between both genera. The genes encoding the photosynthetic flavodoxin isoform was previously detected in four species of *Pseudo-nitzschia* and in five species of *Thalassiosira* (Groussman et al., 2015), and its absence was noted in the genome of one species of *Thalassiosira* (Whitney et al., 2011). There are many reports showing iron-driven down-regulation in expression of the photosynthetic flavodoxin in *Thalassiosira* species from culture or environmental samples (Lommer et al., 2010; Whitney et al., 2011; Marchetti et al., 2012). By contrast, the data from *P. granii* are inconsistent, as transcript abundance of *flavodoxin* was shown to increase in cultures exposed to low-iron conditions (Marchetti et al., 2017) while in on-deck bottle incubation experiments the transcripts were shown to be higher following iron enrichment (Marchetti et al., 2012). Our results show that flavodoxin gene and mRNA abundance display negative correlations with iron in *Pseudo-nitzschia*, although the signals from *Thalassiosira* are not clear (Figure S3a). In the case of ferredoxin, its gene appears to be plastid-encoded in many diatoms, although it is nuclear-encoded in some species (Lommer et al., 2010; Groussman et al., 2015). In the *Thalassiosira* genus, the gene is plastid-encoded in two species whereas it is nuclear-encoded in three others (Lommer et al., 2010; Groussman et al., 2015). In contrast, no nuclear-encoded versions have been reported in *Pseudo-nitzschia* (Marchetti et al., 2009). Consistently, we were able to detect *ferredoxin* unigenes from *Thalassiosira* (typically containing the ASAFAP motif for plastid targeting (Groussman et al., 2015) but not from *Pseudo-nitzschia* (Figure S3a), which is coherent with the fact that our protocols minimize detection of plastid-encoded genes as they are based on polyadenylated RNA (see Alberti et al., 2017). We

also found that *ferredoxin* expression levels in *Thalassiosira* did not appear to be responsive to different iron levels. Although *ferredoxin* has been reported to be one of the most abundant iron-associated proteins in *Thalassiosira weissflogii* (Claustre et al., 2008), it therefore appears difficult to use the gene as a biomarker for iron using only metatrascriptomics from nuclear encoded genes.

The plastocyanin/cytochrome c_6 switch showed similar patterns in both genera. The gene for the heme-protein cytochrome c_6 appears to be core in diatoms, even in those species with the gene for the isofunctional cuproprotein plastocyanin (Groussman et al., 2015). On the contrary, plastocyanin is absent from many species, such as one species with a sequenced genome from *Thalassiosira* and another from *Pseudo-nitzschia* (Groussman et al., 2015). In our analyses gene expression appears to be negatively correlated with iron for plastocyanin in both genera, while gene abundance was negatively correlated only in *Pseudo-nitzschia*. *Cytochrome* c_6 seems to display no detectable correlation with the predicted levels of iron in *Pseudo-nitzschia* (Figure S3a), and in *Thalassiosira* the signal was too low to take it into account. For both flavodoxin/ferredoxin and plastocyanin/cytochrome c_6 the genes encoding the iron-free proteins therefore appear to be the best markers of iron concentration.

Ferritin is present in all diatom classes, but its distribution appears to be group specific (Groussman et al., 2015). It has been found in all eight species of *Pseudo-nitzschia* with sequenced genomes or transcriptomes, but only in two *Thalassiosira* representatives, whereas it is absent in the two sequenced genomes from *Thalassiosira* (Groussman et al., 2015). This gene distribution is coherent with our results. Unigenes assigned to *Pseudo-nitzschia ferritin* were found worldwide, except in the Southern Ocean stations (Figure S3b). In contrast, we were able to find the corresponding *Thalassiosira* unigenes only in 14 stations, located in the Southern Ocean, the Atlantic Ocean and around the Marquesas islands (Figure S3b). The use of ferritin for intracellular iron management may therefore not be specific to *Pseudo-nitzschia* when compared to *Thalassiosira*, but the biogeographic pattern of gene and mRNA abundances hint at completely different functions and/or regulation. This situation may be related to the fact that many *Pseudo-nitzschia* species possess an additional ferritin isoform different from the one found in some species of *Thalassiosira* (Groussman et al., 2015). When compared within the same

station, the gene and mRNA abundances of *ferritin* were generally much higher in *Pseudo-nitzschia* than in *Thalassiosira*. In many cases the difference was in the range of orders of magnitude, as in some iron-rich stations from the Mediterranean Sea and Indian Ocean.

Analysis of the abundance and expression of *FBA* genes was not included due to their low detection levels in these genera. In the case of proteorhodopsin, we were unable to assign to *Pseudo-nitzschia* or *Thalassiosira* any of the unigenes detected as diatom *proteorhodopsins.* In relation to *SIT* genes, we found unigenes assigned to Clades A and E in *Thalassiosira*, but we were unable to detect any correlation with iron or silicate. In *Pseudo-nitzschia* we found unigenes assigned to Clades A, B and D. None of them showed correlations with iron levels, but Clade B and D gene abundances show strong negative correlations with silicate levels when corrected for iron levels (partial correlation coefficient = -0. 41 and -0. 57, respectively, p-value<0. 05).

S2 Oceanographic context of the Marquesas archipelago study

S2.a Previous studies

The Marquesas Islands are located close to the longitudinal centre of the Pacific Ocean, just south of the Equator (218°E– 222°E, 8°S – 11°S) at the boundary between the iron-limited HNLC equatorial region and the highly oligotrophic subtropical gyre (Claustre et al., 2008). The islands emerge from the deep ocean floor as part of a volcanic hotspot and, from an oceanographic point of view, they sit between the South Equatorial Current (SEC) and the northern margin of the Pacific South Tropical Gyre. Previous studies (Signorini et al., 1999; Martinez and Maamaatuaiahutapu, 2004) have shown that they are regularly characterized by phytoplankton proliferations (with Chl-a> 0. 2 mg/m^3) that extend to 500-1000 km offshore and have been attributed to an island mass effect (Dandonneau, and Charpy, 1995). These previous studies suggested three different kinds of blooms: seasonal, episodic, and La Niña-related, the latter modulating the seasonal bloom. The bloom intensity generally correlates very well with the local current intensity and thus has been associated with the vertical and lateral turbulence generated by the interaction of geostrophic currents (sometimes together with Ekman drift) with the

Marquesas archipelago. Several mechanisms have been invoked to explain these blooms, with most authors agreeing that they are the result of natural iron fertilization, although measurements of iron concentrations in the area are extremely limited. Both (Dandonneau, and Charpy, 1995) and (Signorini et al., 1999;) suggested that land drainage and hydrothermal fluxes were the sources of iron because the observed nitrate concentrations of >2 µM were too high to significantly limit growth rates in the area (see also Martinez and Maamaatuaiahutapu, 2004).

More recent sampling efforts found no strong evidence for a local positive anomaly in dissolved iron in the waters around the islands (Blain et al., 2008), supporting the idea of temporary anomalies due to land drainage at least at the close-shore Station TARA_123. Furthermore, because silicate can be limiting at times, the wide extension of the bloom suggests that while small-scale turbulence may enhance micronutrients close to the islands, there are also other mechanisms at play, possibly related to submesoscale processes within the turbulent coherent structures (eddies, filaments, mushrooms) created by the interaction of the currents and winds with the islands and then transported downstream by the SEC. Finally, the strong currents can also favor the creation of recirculating regions to the lee of the larger islands, as may occur when the current passes around an obstacle. Previous observations (in particular from the BIOSOPE cruise) showed that the plankton community is dominated by pico- and nanoplankton (Dolan et al., 2007; Ras et al., 2008), the former consisting principally of heterotrophic bacteria as well as *Prochlorococcus* and *Synechococcus* (Grob et al., 2007). The microplantkon is generally dominated by diatoms (in particular *Pseudo-nitschia delicatissima* (Gomez et al, 2007)) that potentially generate significant particle flux (Guidi et al., 2008), followed by dinoflagellates, the latter being significant mainly at the Chl-a richest sites. In addition, the high abundance of large transparent and unidentified objects, probably living organisms, was also noted around the islands (Stemmann et al., 2008). Zooplankton can reach very high abundances in enriched areas, with the peculiarity that ciliates are weakly abundant while the contrary appears to be the case for heterotrophic dinoflagellates *(*Masquelier & Vaulot, 2007*)*.

S2.b *Tara* Oceans study in the Marquesas archipelago

The oceanographic analysis and sampling program around the Marquesas archipelago before, during, and after the visit of the *Tara* Oceans expedition to the area in August 2011 was designed to comprehensively inspect an ecosystem experiencing a natural perturbation proposed to be related to iron bioavailability. In addition to the protocols used routinely during the expedition (Karsenti et al., 2011), we deployed gliders and profiling floats over four weeks, and analysed Argo float and remote sensing data over several months.

Of particular interest for a dynamical interpretation of the system is a sequence of images for the period June 30-July 4. First, on June 30 a 130 km long, N-S oriented, arc-shaped Chl anomaly was seen close to the island of Nuku Hiva. The great extension and shape suggest it to be due to the surface intensification of a solitary westward traveling wave or a front associated to a tropical instability wave (Legeckis et al., 2004), whose intensification could be due to the interaction with the bottom topography. Its extent also suggested that the fertilization was not only due to the island itself. The rapid emergence of the strong Chl anomaly further suggests that part of the satellite signal could be due to the exposure (entrainment into the mixed layer, ML) of the organic matter accumulated in the highly stratified waters that reside below the ML. Besides having a direct impact on the optical properties of the water, this entrainment could stimulate a fertilization via the injection of organic matter, possibly iron containing, that can be recycled in the trophic web.

The glider data (Figure S4a) show how the islands promote a rather complex shear pattern that is fed by the large scale westward current, with vertically averaged current velocities of over 30 cm/s. The highest shear is observed southwest of Nuku Hiva. The island acts as an obstacle for the main current and it causes a funneling of the main current on its northern and, above all, southern side. This region of high shear is also characterized by higher salinities (a result of a locally enhanced vertical mixing) and is the source region of the train of small cyclonic eddies that were observed to move westward during the experiment (Figure 4a). One of those eddies was sampled at Station TARA_124.

Relatively fresher and cooler waters were observed on the lee of the main islands (Figure S4c). A pool of lower salinity water was observed on the lee of Nuku Hiva, in correspondence to Station TARA_123, indicating a possible local impact of land drainage. The vertical CTD profiles from a transect between Station TARA_122 and the closest island, Ua Huka, are shown in Figure S4d. The CTD station #14 (Station 122c in Figure S4d) is in the high Chl patch and is on the island shelf. The analyzed parameters (CDOM, Chlorophyll and backscattering) display higher values in the first 30-40 m at this CTD station. Nuku Hiva is the first island encountered by the westward flow. In addition, the region is characterized by an intense mixing at night, reaching >100 m, while the CTD casts were performed late in the afternoon, during the period of maximal stratification. Therefore, these results suggest that the biological response is very rapid and that it is surface intensified, with possibly a dilution during the following night.

S3. Plankton biodiversity and taxonomic composition around the Marquesas archipelago

A large number of previous studies have addressed the impact of nutrients on ocean processes, yet only a few of them have examined the impact on overall plankton community structure in detail (Smetacek and Naqvi, 2008; Quéguiner, 2013), and only one, to the best of our knowledge, has used omics approaches to interpret changes in community structure *in situ* (Martin et al., 2013). Furthermore, previous work has generally been performed on manipulated communities *in situ* or in simplified on-deck bottle experiments and has tended to examine only a subcomponent of the community (Mock et al., 2008; Martin et al., 2013; Alexander et al., 2015). In the current work we integrated a combination of methods to explore the response of the entire plankton community, from viruses to zooplankton, to the natural nutrient perturbation downstream of the Marquesas Islands (Martinez and Maamaatuaiahutapu, 2004). The perturbation resulted in higher chlorophyll concentrations with respect to the surrounding HNLC waters (Figure 4a) and deeply affected different components of the plankton community in terms of abundance, diversity and taxonomic composition, discussed below.

S3.a Biodiversity assessment using metabarcoding data

The global metabarcoding dataset derived from the *Tara* Oceans expedition, based on the hypervariable V9 region of 18S rDNA and described in (de Vargas, 2015), was specifically investigated to explore the relative abundance, richness and diversity of eukaryotic plankton at the four sites sampled during the expedition around the Marquesas Islands (Figure 8d-f, Table S3b).

With respect to HNLC Station TARA_122, a major shift detected at Station TARA_123 was in the increased relative abundance of metazoan barcodes with respect to those from Rhizaria (Table S3b). This shift was detected in all of the analyzed size fractions but was most clear in the larger fractions (20 – 180 µm and 180 – 2000 µm), consistent with the relatively large sizes of these organisms. Because Rhizaria are typical of oligotrophic open ocean environments (Guidi et al., 2016) their reduced abundance at Station TARA_123 is consistent with the more productive conditions observed at the macro scale at this sampling site.

Analysis of the smaller fractions (0. 8 – 5 µm and 5 – 20 µm) from these stations allowed the identification of another major change occurring at Station TARA_123 with respect to Station TARA_122, namely the higher relative abundances of diatoms (Bacillariophyta), pelagophytes, and haptophytes (Figure 8), which collectively appear to be responsible for the increased production at this site. Station TARA_124, and to a lesser extent Station TARA_125, were also characterized by increased production. Furthermore, at Station TARA_124 the large Rhizaria became dominant again while metazoans were reduced in relative abundance (Table S3b), and the eukaryotic community at Station TARA_125 was very similar to that at Station TARA_122. At Station TARA_123, diatom and haptophyte diversity increased with respect to the HNLC Station TARA_122 (Table S3b). By contrast, the abundance, diversity and richness of small dinoflagellates (5 – 20 µm) at Station TARA_123 decreased, in agreement with light microscopy data, while large dinoflagellates displayed an opposite pattern. Concerning the zooplankton, an opposite trend was found in the two most important metazoan groups, urochordates and copepods, at Station TARA_123. Copepods showed a peak in diversity at this station, while the reduction in urochordate diversity (fraction 180

– 2000 µm) was suggestive of a quasi-monospecific urochordate assemblage at Station TARA_123 (Table S3b).

A broader characterization of the biological communities based on trophic modes was used to explore ecological patterns across the four stations and organismal size-classes in terms of the relative abundance (reads) and diversity (in terms of unique numbers of ribotypes) of different organismal groups (Table S3b). A major shift in the trophic mode was detected at Station TARA_123 with respect to the HNLC conditions at Station TARA_122, namely the decrease of the relative abundance and richness of photosymbiotic hosts (mainly rhizarians). This shift was found in all the analysed size classes, with the exception of 20-180 µm. Recent data have reported that rhizaria can host small-celled dinoflagellate symbionts (Probert et al., 2014; Yuasa et al., 2016). While our datasets (metabarcodes and light microscopy) are not able to distinguish such dinoflagellates, it is of note that the relative abundance of the genus *Scrippsiella* at Station TARA_123 decreased with respect to Station TARA_122 (from metabarcode data) and that the abundance of small free-living, likely photosynthetic, naked dinoflagellates increased at Station TARA_123 (see next section light microscopy).

S3. b Light microscopy analysis of eukaryotic phytoplankton

The Marquesas stations were generally characterized by the low abundance of nano- and microplankton cells (about 1·104 cells·L^{-1}), however a slightly higher abundance was noted at Station TARA_123 (< 4·104 cells·L^{-1}) (Table S3b). All stations were generally dominated by diatoms. Coccolithophores were observed only in the stations within the chlorophyll-rich patch, and in particular at Station TARA_123. In terms of biomass, dinoflagellates were generally dominant as compared to the other phytoplankton groups (diatoms, coccolithophores and ‘others’) at all stations, although a significant decrease in the biomass of this group was detected at Station TARA_123 (Table S3b), likely caused by the increased amount of small cells with respect to large cells, and by the increase in diatoms, which became a significant component of the biomass at this station (Figure 8b and Table S3b).

Species composition was similar among all samples. Undetermined centric and pennate diatoms, *Cylindrotheca closterium*, *Chaetoceros* and *Pseudo-nitzschia* species were the most abundant among diatoms, the only difference being their relative abundance in the individual samples. *Chaetoceros* and *Pseudo-nitzschia* cells displayed higher abundance at Stations TARA_122 and 124. Undetermined centric diatoms, which included *Thalassiosira* and *Schionodiscus* species, showed the highest abundance at Station TARA_123 in both bottle and net samples. *Planktoniella sol* (Thalassiosiraceae) was present at all stations but with low abundance. The undetermined pennate diatoms showed quite similar relative abundances in all samples, while *C. closterium* showed the highest abundance at Station TARA_123 in both bottle and net samples. *Gephyrocapsa* spp. dominated the coccolithophores at Station TARA_123. Microplankton communities, analysed in the 20-180 μm size fraction net samples, were characterized by the dominance of diatom taxa which attained a relative abundance up to 65%. Dinoflagellates showed an increase in their relative abundance from Station TARA_122 to the other stations.

S3.c Diagnostic photosynthetic pigments for taxonomy determination

Diagnostic pigment concentrations are shown in Figure S4e. Overall, the results are quite consistent with the phytoplankton abundances deduced from metabarcoding and microscopy data. In agreement with the metabarcoding data, the diatom-specific fucoxanthin pigment increased at Station TARA_123. The haptophyte specific pigment 19'-hexanoyloxyfucoxanthin also shows a very robust increase at this station with respect to Station TARA_122, while 19'-butanoyloxyfucoxanthin, which is specific to pelagophytes, showed increases at all downstream stations with respect to HNLC Station TARA_122. Alloxanthin (specific to cryptophytes) was only detected at Station TARA_123. The non-specific pigments chlorophyll *c1* and *c2* (from diatoms, haptophytes, dinophytes and other chromalveolate groups) and diadinoxanthin (a photoprotective pigment present in the chlorophyll *c*-containing algae) all show a very robust increase at Station TARA_123 and 124. Chlorophyll *b* (representative of chlorophytes) increased slightly at Station TARA_124 with respect to Station

TARA_123. Peridinin, a diagnostic pigment of photosynthetic and mixotrophic dinoflagellates, doubled in concentration at Station TARA_123 with respect to the other stations. Phaeophytin, a signal characteristic of grazing, displayed the highest levels at Station TARA_123.

Three bacterial related pigments were also analyzed: bacterio-chlorophyll *a* (from α-proteobacteria), zeaxanthin (cyanobacteria - although it also acts as a photoprotective pigment in other algae) and divinyl chlorophyll *a* (*Prochlorococcus*). Bacterio-chlorophyll *a* showed a small increase at Stations TARA_123 and 124. On the other hand, zeaxanthin and divinyl chlorophyll *a* both displayed very clear decreases at Station TARA_123, consistent with the loss of *Prochlorococcus* at this station (see below).

S3.d Bacterial abundance and diversity

For the cyanobacteria, *Prochorococcus* almost disappeared at Station TARA_123 but showed no compositional shift, suggesting that the *Prochlorococcus* clade usually dominating in warm iron replete waters (HLII) was not responsive enough to replace clades HLIII and HLIV, adapted to iron-depleted waters (Rusch et al., 2010; West et al., 2011). By contrast, *Synechococcus* increased in abundance at Stations TARA_123, 124, and 125, concomitant with a drastic shift in clade composition from a dominance of clade CRD1, known to be adapted to iron limitation (Farrant et al., 2016; Sohm et al., 2016), to a dominance of clade II, adapted to iron-replete waters (Figure 6). Thus, this shift could partially explain why *Synechococcus* was able to take over *Prochlorococcus* at Station TARA_123 (Table S3b). Of note, flow cytometry and *petB* miTags data are consistent with these community shifts (Figure 6, Table S3b).

S3.e Cyanophages

Because processes other than clade compositional shifts appeared to be involved in the dominance of *Synechococcus* over *Prochlorococcus* at Station TARA_123, we examined cyanophage abundance. Interestingly, the relational dynamics between cyanophages and cyanobacteria indicate that *Synechococcus* phages and cell abundance were not

significantly correlated, while *Prochlorococcus* phage abundance was positively correlated with *Prochlorococcus* cell counts. These results indicate that iron replete conditions may have resulted in decoupling of the dynamics between cyanophages and their hosts and suggest that *Synechococcus* but not *Prochlorococcus* is less prone to viral infections in iron-replete waters (Table S3b).

S3.f Zooplankton

The 18S-V9-based metabarcode data indicate that two metazoan lineages, namely copepods and appendicularians, were dominant in the Marquesas stations (Figure 8). No major shifts in the composition of metazoans were detected in the 20-180 μm size class, while an impressive increase of Urochordata abundance was detected at Station TARA_123, fraction 180-2000 μm. Of note, in this station metazoans represented 51% of the large-sized plankton (Table S3b). The shift was mainly due to an increase in appendicularians (Table S3b), and in particular to an unclassified *Oikopleura* species. This is of particular interest because some appendicularians in the Equatorial Pacific are believed to feed on *Synechococcus*, but not on *Prochlorococcus* or picoeukaryotes (<2 μm) (Gorsky et al., 1999). Moreover, appendicularians are thought to play a major role in detritus production and export fluxes (Berline et al., 2011).

Net counts (lacking for Station TARA_124) and Zooscan data only partially corroborate the data from metabarcodes. This is likely due to the fact that the latter were inferred from the surface, while net counts and Zooscan data were derived from water column samples. From net counts data (Table S3b) a 3-fold increase in copepods was detected at Station TARA_125 with respect to Station TARA_122, and a similar increase was found for tunicates. Of note, chaetognaths were relatively abundant at Station TARA_122 and 125, but not at 123, suggesting that these were outcompeted by other predators. Within the copepod assemblage, calanoids were the most abundant order. Richness and diversity data are mostly compatible with metabarcode data. A higher proportion of juvenile versus adult copepods was detected at Station TARA_124 (Table S3b). Zooscan data showed that the zooplankton community at Station TARA_122 was dominated by organisms well adapted to environments with scarce resources, such as Acantharia and Foraminifera

(Table S3b). The zooplankton community at Station TARA_123 showed an overall increase in biovolume compared to Station TARA_122, and was mainly composed of heterotrophic protists, appendicularians, chaetognaths and copepods. A combination of reduced stress and a shift of metabolism towards growth is likely the basis of the increased tunicate numbers at Station TARA_123. Rhizarian abundance was lower compared to Station TARA_122. At Station TARA_124 the zooplankton community was very similar to that at Station TARA_123, although appendicularian abundance diminished and rhizarian abundance increased with respect to Station TARA_123. The status of the zooplankton community at Station TARA_125 was comparable to that observed at Station TARA_122, but with larger copepods.

S3.g Dynamics of IAAs in Marquesas sampling sites

The responses of IAAs in the Marquesas Islands is complex and highly differentiated. The DarkRed module is most probably a low-iron associated IAA (see main text). It shows a negative response to iron concentration, given the prevalent role of iron anti-correlated OTUs in this module. Indeed, in this IAA the autotrophic component is the most relevant (with respect to the other IAAs): this is even more evident when only the iron-correlated OTUs are taken into account (Table S1d). It is the only IAA in which there are no ubiquitous OTUs. Moreover, richness analyses (Figure 7b) show that the OTUs which are included in this module are largely retained in Station TARA_122 (low iron), but are generally absent in downstream stations, suggesting a reduced relevance of this IAA compared with the others in high iron environments. Both absolute and relative abundances decrease in the wake of the islands (Station TARA_123) (Figure 7b) and the recovery in abundance at Station TARA_124 is mostly due to a single OTU (Prasino-clade) (Figure 7b).

The Turquoise module is a large IAA (almost 600 OTUs) made of mostly heterotrophic OTUs (Table S1e). Nonetheless, the role of autotrophic OTUs is more relevant when only the iron-responsive ones are taken into account (Table S1d). It seems to be a 'generalist' module, which includes OTUs able to respond both in low-iron and high-iron environments. One hypothesis is that this module shows patterns of resilience with

687 respect to iron because it includes both iron-correlated and iron-anti-correlated OTUs: the
688 latter are preferentially retained at the Marquesas Islands (Figure 7b, c). Its abundance
689 increases at Station TARA_123 compared to Station TARA_122 and the abundance
690 trends of iron-correlated and iron-anti-correlated OTUs is coherent (Figure 7b, c).

691 The Black IAA has the lowest proportion of autotrophs (Table S1e). This may complicate
692 the interpretation of its dynamics, given the indirect relationships between heterotrophs
693 and iron. Its richness in the Marquesas Islands is reduced, with the exception of Station
694 TARA_123, which indicates a small but consistent positive response to iron (Figure 7b).
695 However, its abundance is reduced in the wake of the islands (Figure 7b). This may be
696 due to the fact that heterotrophs in this IAA are outcompeted at Station TARA_123.
697 Moreover, 3 of the 5 iron-correlating autotrophs in this module appear only in Station
698 TARA_125, suggesting that open ocean species may have a very important role in this
699 IAA (Figure 7c).

700 The Yellow IAA includes only OTUs positively correlated to iron, which are mostly
701 heterotrophs (Table S1e). Richness in this IAA is reduced in the Marquesas Islands
702 (Figure 7b). The abundance of the Yellow IAA is decreased at Station TARA_123,
703 although iron-responsive autotrophs increase in abundance, suggesting that the
704 heterotrophic component may be affected by the increased presence of other heterotrophs
705 (Figure 7b, c).

706

707 **S4 Analysis of metagenomes and metatranscriptomes in Marquesas sampling**
708 **stations**

709 S4.a Taxonomic and functional response of the picocyanobacterial community

710 Metagenomic and flow cytometry analyses indicated that the picocyanobacterial
711 community displayed large variations of composition in the Marquesas Islands area
712 (Figure 6, Figure 8b, c, Table S3b). While *Prochlorococcus* was highly dominant over
713 *Synechococcus* at Stations TARA_122, 124 and 125, it was almost absent from Station
714 TARA_123, and all four stations were co-dominated by clades HLIII and HLIV (Figure
715 8c), known to thrive in HNLC regions. By contrast, *Synechococcus* was almost absent

716 from Station TARA_122 but was abundant in the three stations located downstream of
717 the islands (Stations TARA_123-125), concomitant with a drastic shift in clade
718 composition from a dominance of clade CRD1 (the major clade in HNLC regions) to a
719 dominance of clade II (a clade known to prevail in warm Fe-replete intertropical areas)
720 (Figure 6).

721 Metatranscriptomic analyses further revealed dramatic differences in the functional
722 response of both groups of organisms. The differential expression of genes related to iron
723 metabolism indicated that *Synechococcus* was responsive to island vicinity at Stations
724 TARA_123-125, in a way that supports an injection of iron into the ecosystem (Figure
725 9b, Table S3c). This includes the replacement of the flavoprotein flavodoxin (*isiB*,
726 downregulated 10-fold at Station TARA_123) by the Fe-containing ferredoxin (*petF;*
727 upregulated 2-fold in Stations TARA_123-125), induction of the *petJ* gene, encoding
728 cytochrome c_6, which was only detected at Stations TARA_123-125 (although its Cu-
729 containing counterpart, plastocyanin (*petE*), was also upregulated at these stations), or the
730 replacement of the gene encoding fructose bisphosphate aldolase class I (*fba*) by the
731 metal-dependent FBA class II (*fbaA*). In contrast, the iron-stress induced gene *isiA* was
732 more expressed in the HNLC Station TARA_122, which could help compensate for the
733 reduction in the number of photosystem I complexes known to occur under iron
734 limitation. Finally, *Synechococcus* also seemed to be able to efficiently scavenge and
735 store iron in island-influenced stations, as suggested by the upregulation of genes
736 encoding numerous iron transporters (*sufC*, *sufD*, *feoB*, *idiB* and an unnamed iron
737 transporter of the NRAMP family) and for the iron-storage protein ferritin (*ftn*;
738 upregulated > 10-fold) in iron-enriched stations (Table S3c).

739 By contrast, for *Prochlorococcus* several of these key genes (e. g., *petJ*, *fbaA*, *isiA*) could
740 not be detected in *Tara* Oceans meta-omics datasets (Table S3c) nor in HLIII/IV
741 metagenomes sequenced to date and, when present, only a slight differential expression
742 was observed between Station TARA_122 and the three stations downstream of the
743 islands. Consistent with the results obtained from the GOS dataset (Toulza et al., 2012)
744 no negative correlation between the presence of flavodoxin and iron bioavailability could
745 be observed for *Prochlorococcus*, and ferredoxin encoding genes (*petF*) were only
746 expressed at a low level in all four stations. Similarly, the expression levels of genes

involved in iron storage (ferritin, *ftn*) and transport (*sufC, sufD, feoB,* etc.) remained fairly constant in all stations (Table S3c), suggesting that *Prochlorococcus* was not able to increase its iron-uptake in Stations TARA_123-125. Thus, in contrast to *Synechococcus*, no clear acclimation process to iron availability was detected for *Prochlorococcus*.

Analysis of the mapping intensity of metagenomic reads (Table S3c) provides further insights into the differential adaptation of communities present at strictly HNLC regions with respect to stations influenced by the Marquesas Islands. In particular, several *Synechococcus* genes involved in iron metabolism (*fbaA, isiA, ftn*, and the transporters *sufC, sufD, feoB*) show a higher mapping intensity of metagenomic reads at Station TARA_122 than at Stations TARA_123-125, suggesting that the community present at this HNLC station (dominated by clade CRD1) has a higher copy number of these genes in their genomes compared to those found in island-influenced stations (dominated by clade II) (Table S3c). In contrast, some genes were seemingly more abundant (or only detected) at Stations TARA_123-125 (*petJ, fda* and the NRAMP-like transporter). This may indicate that the two communities display distinct genetic potentials, and reveals a possible adaptation of clade CRD1 to iron-limited (HNLC) areas, and/or an adaptation of clade II to iron-richer conditions, as suggested by recent studies.

In conclusion, we propose that the vicinity of the Marquesas Islands provokes a dramatic shift in the *Synechococcus* community, from a typical HNLC population at Station TARA_122 to a population more adapted to iron-replete conditions at Stations TARA_123-125, associated with a strong transcriptomic response to iron enrichment, while the *Prochlorococcus* community remains constant and transcriptionally unresponsive, thus unable to acclimate to the change in conditions induced by the islands. Altogether, *Synechococcus* community response through both adaptation and acclimation processes tends to support the injection of iron into the ecosystem influenced by the Marquesas islands, while the apparent inability of *Prochlorococcus* to benefit from these altered environmental conditions might explain its apparent decline in this area.

S4.b Iron-related gene expression throughout the eukaryotic phytoplankton

To explore the responses of the five principal photosynthetic protist groups in the Marquesas Islands stations we examined differential gene abundance and expression patterns. We specifically focused on Station TARA_123 because it displayed the strongest eukaryote response in terms of biomass and chlorophyll production in comparison to the HNLC Station TARA_122 (Figure 8a, b). Distribution of Pfam fold-change normalized within each taxonomic group revealed that several genes in the five main phytoplankton groups have strong differential expression (Figure S5a). Furthermore, with a flattened distribution pattern, dinoflagellates display a majority of genes stably expressed in the two stations, suggesting that dinoflagellates have a different transcriptional response in comparison to the four other groups. Interestingly, Pfam domains involved in proteolysis (Trypsin, Peptidase_S8) and ribosomal proteins are highly expressed in TARA_123 in dinoflagellates suggesting that an active heterotrophic nutrition is predominant in this station in contrast to the HNLC station where genes encoding photosynthetic components are strongly expressed (e. g., plastocyanin). By contrast, pelagophytes, haptophytes and diatoms did not show this pattern but an activation of one or several pathways reflecting stronger cellular activities: nitrogen metabolism (Glutamine synthase) for haptophytes and pelagophytes and amino acid synthesis (Ilv) for diatoms and pelagophytes. In addition, chlorophytes and pelagophytes displayed a very high increase of cadmium-containing carbonic anhydrase (Pfam CdCA1) at Station TARA_123 (Figure S5a). This enzyme is involved in dissolved inorganic carbon acquisition by catalyzing CO_2 hydration and in diatoms it has been shown to be dependent on Cd^{2+} or Zn^{2+} *(Lane et al., 2000)*. This result suggests that these two taxonomic groups may be able to activate carbon concentration mechanisms by the use of metal ions available at Station TARA_123.

We also examined genes encoding photosynthesis-related activities by examining the transcription of Light-Harvesting Chlorophyll-binding proteins (components of the light harvesting complex, LHC). Among the five photosynthetic groups the variations in relative LHC contributions show that pelagophytes, haptophytes (0. 8 – 5 filter) and diatoms (5 – 20 µm filter) are the main contributors of the total LHC gene pool in Station TARA_123 at the expense of dinophyceae (5 – 20 µm filter) and chlorophytes (0. 8 – 5 µm filter) (Figure S5c).

Together these results suggest that the strategy employed by dinoflagellates may be to use the proliferation of autotrophic organisms at Station TARA_123 as food supply rather than to activate their photosynthetic pathways. However, the taxonomic precision is insufficient to conclude if several dinoflagellates are mixotrophic and switch their mode of nutrition between the two stations or if among dinoflagellates, autotrophic species dominate the population in TARA_122 and are replaced by heterotrophic species in Station TARA_123. These dynamics during the bloom correlate with estimates of community shifts (Figure 8).

To examine the different dynamics in organismal responses to iron, several specific genes were studied. In the Marquesas stations, a large number of *ISIP* genes were expressed in different lineages (Figure S3a and S5b) and mRNA levels were increased principally at HNLC Station TARA_122 but decreased strongly in TARA_123 and recovered somewhat in TARA_124 and TARA_125. Thus, all five groups appear adapted to low-iron conditions and react to the iron perturbation by tuning down *ISIP* gene expression. Because *ISIP* gene expression is very strongly linked to iron bioavailability (Figure 5a and (Allen et al., 2008; Lommer et al., 2015; Chappell et al., 2015; Morrissey et al., 2015, Graff van Creveld et al., 2016; Marchetti et al., 2017) these observations provide very strong support for the hypothesis that increased iron bioavailability drives the increases in phytoplankton biomass downstream of the Marquesas archipelago.

Gene switches for iron-requiring enzymes are also seen to display differential mRNA levels between Station TARA_122 and the blooming stations in pelagophytes, haptophytes, and diatoms, whereas the organisms that do not use these switches efficiently (dinoflagellates and chlorophytes) are less abundant in the community (Figure S5). This suggests that activation of genes encoding metalloproteins (such as class II fructose bisphosphate aldolases and ferredoxins) may be an important factor for organism proliferation in the bloom areas, whereas *ISIP* genes may encode more general functions required for survival in low iron conditions

S4.c Taxonomic and functional responses of diatom communities

Metabarcoding, metagenomics, and microscopy indicate the presence of rich diatom communities at all sampling stations, with an increase in abundance of approximately 2-fold at Station TARA_123 with respect to Station TARA_122 (Figure 8, Table S3b). This increase is comparable with the increase in chlorophyll and biomass (Figure 8a, b), indicating that diatoms are an important contributor of photosynthesis and of phytoplankton standing stock biomass at Station TARA_123. We therefore focused on differences between this station and the HNLC Station TARA_122 to explore the diatom response in the area downstream of the islands. In terms of the microscopy and metabarcoding analyses, the major differences between Stations TARA_122 and 123 were found to be an increase in centric diatoms (Coscinodiscophyceae), notably small cells of the *Thalassiosira* genus, and a decrease in the typically open ocean genus *Planktoniella*, whereas pennate diatom abundance (Bacillariophyceae) was relatively stable, including *Pseudo-nitzschia*. Both *Thalassiosira* and *Pseudo-nitzschia* are commonly found in severely iron limited regions of the world's ocean, and their responses to changes in iron bioavailability have been characterized rather well (Mock et al., 2008; Lommer et al., 2015), so we explored changes in iron-responsive gene content and mRNA levels in these two genera at Stations TARA_122 and 123.

The overall changes in relative abundances of diatoms observed by microscopy and in the metabarcoding data are also reflected in the metagenomics data (Figure S5a and S7a); *Thalassiosira*-derived sequences increase in relative abundance in small size fractions from Station TARA_123 whereas changes in Bacillariophyceae sequences are less evident. It should be noted that *Planktoniella* is invisible in the metagenomics dataset because of the lack of sequence information from this uncultivated species.

Comparison of metatranscriptome and metagenome data indicate that *Pseudo-nitzschia*, and Bacillariophyceae in general, were more responsive transcriptionally than *Thalassiosira* and the Coscinodiscophyceae (Figure S7). Conversely, cyclin gene abundances and mRNA levels were much higher in the centric diatoms at Station TARA_123 than at Station TARA_122, which was not the case for the pennate diatoms (Figure S7c), indicating that these diatoms stimulate cell division and thus proliferate more strongly at this station with respect to the pennate diatoms.

Regarding known iron-responsive genes, *ISIP* genes decreased in abundance at Station TARA_123 (Figure S7b, c), consistent with the hypothesis that increased iron bioavailability underlies diatom proliferation at this station. The higher levels of mRNA encoding iron-containing FBAII with respect to the metal-free FBAI is further congruent with this hypothesis (Figure S7c, d). In contrast, flavodoxin mRNA levels were increased at Station TARA_123 in all diatom groups examined (Figure S7c, d). As the upregulation of flavodoxin could be part of a general up-regulation of the photosynthetic apparatus in iron replete conditions, it should be analysed in the context of the flavodoxin:ferredoxin switch. However, as ferredoxin is not always encoded in the nuclear genome of diatoms (Lommer et al., 2010; Botebol et al., 2015), our detection of the gene is biased in some species. In fact, we detected nuclear-encoded photosynthetic ferredoxins from Coscinodiscophyceae (including *Thalassiosira*) but not from Bacillariophyceae. In addition, we only detected the gene in the metagenome from Station TARA_123, and although we observed expression in both stations, it is higher at Station TARA_123 (Figure S7c, d).

On the other hand, the cytochrome c_6:plastocyanin switch (previously reported in *Thalassiosira* (Peers and Price, 2006)) is very evident in Coscinodiscophyceae in general, including in *Thalassiosira*, although according to our data this switch does not appear to be operative in the Bacillariophyceae (Figure S2, S3b and S7d). Furthermore, only plastocyanin genes were identified at Station TARA_122 in the metagenomic data from Thalassiosirales, which could indicate that these genotypes are genetically adapted to low iron levels at this station (Figure 6, 7c and S7c, d). These observations were previously noted in studies of the *T. oceanica* genome (Lommer et al., 2015). Additionally, the *Thalassiosira* genotypes at Station TARA_123 also contained ferritin genes, which was not the case at Station TARA_122 (Figure S7d). The use of ferritin for intracellular iron management may therefore not be specific to the Bacillariophyceae, as currently considered (Botebol et al., 2015; Pfaffen et al., 2015), and could also be operative in some Thalassiosirales and other centric diatoms. In the raphid pennates, including Pseudo-nitzschia, ferritin levels are nonetheless much higher than in the other diatoms, as expected (Botebol et al., 2015; Pfaffen et al., 2015) (Figure S7d). A further gene that was examined encodes proteorhodopsin (Marchetti et al., 2015). While we were unable to

detect any genes encoding proteorhodopsin in *Pseudo-nitzschia* or *Thalassiosira*, we could detect genes in Coscinodiscophyceae and Bacillariophyceae. The former were detected in the metagenomes predominantly from Station TARA_122 rather than from Station TARA_123, whereas the latter showed similar patterns in the metatranscriptomes (Figure S7c, d) which is generally consistent with our understanding of the role of this protein in eukaryotic phytoplankton.

Overall, these results indicate a dynamic response of the diatom community that is consistent with an increase in iron bioavailability at Station *TARA*_123, with a range of different strategies being employed. While the regulation of *ISIP* genes is a common underlying feature, ferritin, ferredoxin, cytochrome c_6, plastocyanin, and proteorhodopsin appear to be employed differently between centric and pennate species. Furthermore, small ferritin-containing *Thalassiosira* cells expressing cytochrome c_6 genes increase in abundance at Station TARA_123, replacing larger *Thalassiosirales* genetically adapted to low iron at Station TARA_122 by the retention of plastocyanin and loss of cytochrome c_6 genes. In addition, *Pseudo-nitzschia* cells with flavodoxin and plastocyanin genes are enriched at Station TARA_122 in comparison with TARA_123. Such dramatic community shifts further indicate that ephemeral seed populations of different genotypes must be present in order for the environment to select the best adapted genotypes in different conditions. On the other hand, *Pseudo-nitzschia* cells respond more strongly at the transcriptional level, indicative of a more flexible acclimation response as opposed to the permanent genome-level adaptations that appear to predominate in the *Thalassiosirales*. It should also be noted that the apparent responsiveness of *Pseudo-nitzschia* and *Planktoniella* to the local perturbation at Station TARA_123 with respect to iron bioavailability is underlined by the fact that these genera show very high VIP scores in the iron-responsive DarkRed module (Figure 2a, Figure S1c and Table S1d).

S4.d Transcriptional response of metazoan communities

Analysis of transcriptional activity in copepoda and appendicularia, the two most abundant and representative metazoan lineages, shows clear differences in their response to local perturbations downstream of the Marquesas archipelago. Markers of the response

928 to the biological availability of iron (e. g., encoding Ferritin and Mitochondral Carrier
929 Protein) show a strong, albeit transient (TARA_123) response in Appendicularea, while
930 Copepoda, unresponsive at TARA_123, show a more gradual and long-lasting, but
931 weaker response (Figure S6).

932 Metatranscriptomics (Catrradec et al., 2018) data show little response of the copepod
933 assemblage at Station TARA_123. mRNA levels of genes related to iron transport
934 (Hemopexin, Transferrin, HRG), iron homeostasis (Frataxin) and heme biosynthesis
935 (Ferrochelatase, URO-D), show a mild downregulation (0. 5-fold) at Station TARA_123
936 with respect to Station TARA_122, whereas the increase in mRNA levels of genes
937 encoding several iron-dependent cytochrome C oxidases (COX) at Station TARA_124
938 (up to 7-fold), is suggestive of an intense respiratory activity linked to increased iron
939 bioavailability. The mRNA levels of canonical oxidative stress biomarkers (GST, GSH)
940 were not significantly altered in copepods at any station. Nonetheless, indications of a
941 metal-induced oxidative response are detectable: phytochelatin mRNA levels were
942 increased at Station TARA_123 (11-fold), potentially reflecting exposure to metals such
943 as cadmium (Bundy and Kille, 2014) or copper (De Vos et al., 1992). Interestingly,
944 phytochelatins are known to cause the loss of glutathione (GSH) in some plants (Battaglia
945 et al., 2008). The mRNA levels of genes encoding proteins containing the seed
946 maturation protein (Battaglia et al., 2008) (SMP) motif were also higher at Station
947 TARA_123. Ubiquitin-related modifier (Urm1) mRNA levels were furthermore strongly
948 increased with respect to the HNLC Station TARA_122. The latter gene is known to
949 decrease the tolerance of arthropods to oxidative stress (Khoshnood et al., 2016). Overall,
950 these results suggest a slight responsiveness of copepods to variations in bioavailable
951 trace metals at Station TARA_123, whereas at Station TARA_124 the evidence for
952 increased respiratory activity is suggestive of a positive effect of iron enrichment on the
953 copepod diet at this station, which is in agreement with population dynamics data.

954 For the appendicularians, at Station TARA_123 the mRNA levels of genes encoding
955 thioredoxin-7 and NAD-binding proteins increased, respectively, by 8-fold and 7-fold,
956 suggesting that a transcriptional response to metal-dependent oxidative stress was
957 occurring locally. Of further note, GSHPx mRNA levels decreased at this station. Clear
958 signs of increased iron-dependent metabolism were also evident at Station TARA_123, in

that mRNA levels of genes encoding proteins with Pfam domains linked to iron transport were increased, such as hemopexin (3-fold), iron acquisition (PigA; 6-fold), and iron-dependent electron transfer (FDX-ACB; 6-fold, COQ7; 8-fold, COQ2; 6-fold, COX6B; 5-fold, p450; 3-fold). Conversely, decreases in ferritin (iron storage) and MitoNEET (iron homeostasis) mRNA levels were also detected at this station. Overall, these results are consistent with a strong response to increased iron bioavailability. This response is also strictly local, since mRNA levels of these iron-related genes was very similar at the other stations. Compared to copepods, we therefore propose that appendicularians may be able to respond more rapidly to increased iron bioavailability. Moreover, the transcriptomics data, in agreement with population dynamics data, are suggestive of a burst in energy production and metabolism that stimulate reproduction and development.

Tunicates are believed to accumulate metal ions that are potentially toxic as a strategy for predation avoidance (Crans et al., 2004) In addition, many tunicate species are well adapted to very polluted areas in which toxic concentrations of heavy metals have been documented (Arienzo et al., 2014).

Figure S1: WGCNA analysis of eukaryotic plankton. (a) Heatmap showing correlation between eukaryotic modules associated to environmental parameters, as revealed by weighted gene correlation network analysis (WGCNA). The Brown, DarkRed, Turquoise and Yellow modules (indicated in bold and with asterisks) are those showing the highest correlations with iron concentration estimates from the ECCO2-DARWIN and/or PISCES2 models. (b) Clustering of eukaryotic modules as generated by WGCNA. The dendrogram classifies all eukaryotic modules based on their relative correlations. The modules studied in this manuscript are indicated with asterisks. (c) Centrality and correlations to iron (ECCO2-DARWIN-estimates) for lineages of interest at global and local levels. Boxplots showing the centrality (right) and correlations to iron (left) of lineages of interest within Brown, DarkRed, Turquoise and Yellow IAAs. These four IAAs are compared with DarkGrey and Red modules, as well as to all other modules combined (denoted Global). (d) Correlation between ECCO2-DARWIN and PISCES2 VIP derived scores for the OTUs within the Brown, DarkRed, Turquoise and Yellow IAAs.

Figure S2: Abundance and expression of iron-responsive genes in eukaryotic phytoplankton from surface waters. Abundance (metagenome) and expression (metatranscriptome) of genes encoding ferritin, ISIP, FBAII, FBAI and proteorhodopsin. Chlorophyta, Pelagophyceae and Haptophyceae sequences were analyzed in the 0. 8 – 5 μm size fraction, whereas Bacillariophyta and Dinophyceae sequences were examined in the 5 – 20 μm size fraction. The abundance and expression values of each gene were normalized by the total abundance and expression of all unigenes assigned to the corresponding taxon, respectively. Iron levels corresponds to estimates from PISCES2 biogeochemical model. The global maps show the distribution of metagenome and metatranscriptome levels of each gene. Circle colors represent iron levels at each *Tara* Oceans sampling sites, circle areas represent the gene abundance or expression, and x's represent missing values. The scatter plots correspond to the correlations between gene abundance and iron, gene expression and iron, and between

gene abundance and expression, with values scaled to the unit interval. Pearson correlation coefficients (pcc) and p values are indicated in blue, and missing values as x's.

Figure S3: Abundance and expression of iron-responsive genes in the diatom genera *Thalassiosira* and *Pseudo-nitzschia* in surface samples from 5-20 µm size fractions. Abundance (metagenome) and expression (metatranscriptome) of genes encoding ISIP, flavodoxin, plastocyanin, cytochrome c_6 (a) ferritin and ferredoxin (b). The abundance and expression values of each gene were normalized by the total abundance and expression of all unigenes assigned to the corresponding genus, respectively. Iron levels corresponds to estimates from PISCES2 biogeochemical model. The scatter plots correspond to the correlations between gene abundance and iron, gene expression and iron, and between gene abundance and expression, with the values scaled to the unit interval. Pearson correlation coefficients (pcc) and p-values are indicated in blue, and missing values as x's. The global maps show the distribution of metagenome and metatranscriptome levels of each gene. The circle colors represent iron levels at each *Tara* Oceans sampling sites. The circle areas represent the gene abundance or expression, while missing values are represented by x's. Note that *Pseudo-nitzschia* ferredoxin is not shown because its mRNA was barely detectable as it is chloroplast encoded and our protocols are based on polyadenylated RNA.

Figure S4: **Oceanographic context of Marquesas archipelago sampling sites.** (a) Glider path during the oceanographic analysis in the Marquesas archipelago. The arrows indicate the average current velocity estimated from the deviation of the path with respect to the pre-assigned way points. Given the sawtooth path of the glider, these values refer to the average over several hundreds of metres, thus underestimating the actual current velocities at the surface. In color is depicted the average particulate backscattering on 532 nm over the upper 100 m. The fertilized patch has a strong signature in the backscatter data, that is directly related to suspended matter. (b) Vertical profiles from CTD casts. Plots show temperature, salinity, NO_2, NO_2NO_3, PO_4, and silicate. (c) Thermosalinograph

data from the *Tara* Oceans transect in the Marquesas archipelago. Surface temperature (upper panel; units in °C) and salinity (lower panel; psu). (d) Vertical CTD profiles from a transect between Station TARA_122 and the closest island, Ua Huka. Left panel: Chlorophyll (from fluorimeter). Right panel: Backscattering (b660). (Box: Surface chlorophyll concentration as determined from Aqua-MODIS on the 3rd of August 2011 (spatial resolution equals 1 km). Location of CTD casts are indicated. Black traces are from Station TARA_122. (e) Photosynthetic pigment concentrations at *Tara* Oceans sampling stations in the Marquesas archipelago. Data is from integrated (0 – 118 metres depth) diagnostic pigment concentrations ($\mu g \cdot C \cdot m^{-2}$). Chlorophyll *a* concentrations are 1/10 of actual values.

Figure S5. Changes in iron-associated genes in phytoplankton in Marquesas Islands sampling stations. Specific genes show differential abundances and expression patterns with group specific responses. (a) Distribution of fold-changes in Pfam abundance in metatranscriptomic data between the oligotrophic Station TARA_122 and the blooming Station TARA_123 in the five main eukaryotic photosynthetic groups. Dinophyceae and Bacillariophyta were analyzed in the 5 – 20 μm size fractions, whereas Pelagophyceae, Haptophyceae and Chlorophyta data are derived from the 0. 8 – 5 μm filters. Pfam abundances were normalized in percentage within each taxonomic group. Pfams with abundances lower than 0. 05% were excluded from the analysis. Pfams with interesting expression patterns are indicated (LHC = PF00504; Ferredoxin family = PF00111; Flavodoxin_1 family = PF00258; FBA II = PF01116; FBA I = PF00274; Plastocyanin = PF00127, PF13473; Ribosomal_Prot = Sum of all ribosomal Pfam domains). (b) Abundance (metagenome) and expression (metatranscriptome) of the three *ISIP* genes in five photosynthetic eukaryote lineages in the four Marquesas Islands stations. Taxonomic groups were analyzed in the 0. 8 - 5 μm size fraction, except for Bacillariophyta and Dinophyceae which were also studied in the 5- 20 μm size fraction. Horizontal scales represent the relative abundance or expression of genes as percentages in the metatranscriptome or metagenome datasets for each taxonomic group. (c) Contribution of main taxonomic groups to the expression of genes encoding Light-Harvesting Chlorophyll-binding proteins (LHC). The breakdown of the relative expression (in

percentage) of the LHC Pfam domain (PF00504) in the different taxonomic groups is indicated by a color code in each Marquesas station (TARA_122 to TARA_125). The average mRNA levels in all *Tara* Oceans stations is also shown (left column). The two smallest size fractions (0. 8 - 5 μm and 5 - 20 μm) are shown. (d-e) Abundance (metagenome and expression (metatranscriptome) of ferredoxin and flavodoxin genes (d) and Fructose bisphosphate aldolase class I and class II genes (e). Seven photosynthetic taxonomic groups are represented. The five eukaryote lineages were studied in the 0. 8 - 5 μm size fraction (Haptophyceae, Pelagophyceae and Chlorophyta) and in the 5 - 20 μm size fraction for Bacillariophyta and Dinophyceae. The two cyanobacteria genera (*Synechococcus* and *Prochlorococcus*) were analyzed in the 0.2-3 μm size fraction. Horizontal scales represent the relative abundance or expression of genes as a percentage of the metagenome or metatranscriptome for each taxonomic group.

Figure S6. Changes in iron-associated genes in zooplankton in Marquesas Islands sampling stations. Abundance (metagenome, magenta) and expression (metatranscriptome, orange) of unigenes containing Pfam domains related to iron metabolism (Ferritin, Mitochondrial carrier), development (Homeobox, PAX and wnt) and reproduction (Motile_sperm and Vitellogenin_N) in the two main metazoan taxa: Copepoda (top panel) and Appendicularia (bottom panel). The y axis represents the percentage of expression or abundance of the Pfam domain in the taxonomic group.

Figure S7: **Responses of diatom genera and iron-responsive genes in the Marquesas archipelago.** (a) Abundance of diatom V9 18S barcodes (metabarcode) and diatom-derived sequences in metagenome and metatranscriptome datasets. Diatoms were organized at genus level for *Pseudo-nitzschia*, *Planktoniella*, *Thalassiothrix* and *Thalassiosira*-derived sequences, and into Other Pennates, Other Centrics, and Unclassified Bacillariophyta for sequences from other diatoms. For a comprehensive comparison, diatoms in all size fractions are shown, and for clarity we focused on changes between Stations TARA_122 and 123. (b) Abundance (left) and expression (right) of *ISIP* genes assigned at three different levels of resolution in a diatom phylogenetic tree. The concentric disks nearest to the centre of the tree are

Bacillariophyta-assigned genes (phylum level). Next to them are the disks belonging to the four major Bacillariophyta classes. Finally, in the leaves of the tree the levels within individual genera are shown. In this analysis, the abundance and expression of *ISIP* genes in all filters and all Marquesas Islands stations are shown. (c) Abundances and expression of genes potentially responsive to iron in metagenome and metatranscriptome datasets. Values were normalized by total abundance or expression of all unigenes assigned to the corresponding taxonomic group (Bacillariophyceae, Coscinodiscophyceae, *Pseudo-nitzschia* and *Thalassiosira*). For clarity we focused only on changes in 5-20 μm size fractions from Stations TARA_122 and 123. (d) Relative ratios between pairs of genes whose presence in the genome or transcriptional activity has been reported previously to be influenced by ambient iron bioavailability. For clarity, ferritin levels have been multiplied by a factor of 10 to be comparable with ISIP levels, as have proteorhodopsin levels with respect to LHC levels, and only 5-20 μm size fractions from Stations TARA_122 and 123 are compared.