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, 8], 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 de Génomique, Genoscope, 2 rue Gaston Crémieux, 91057 Evry, France. [3]: CNRS UMR 8030, Evry, France. [4]: Université d'Evry Val d'Essonne, Evry, France. [5]: Laboratoire des Sciences du Numérique de Nantes (LS2N) – CNRS, Université de Nantes, É ole 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), É ole 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

62   Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France. 23: Department of Microbiology

63   and Immunology, Rega Institute, KU Leuven, Herestraat 49, 3000 Leuven, Belgium. 24: Center for

64   the Biology of Disease, VIB, Herestraat 49, 3000 Leuven, Belgium. 25: Cellular and Molecular

65   Microbiology , Faculté des Sciences, Université Libre de Bruxelles (ULB), Belgium. 26:

66   Interuniversity Institute for Bioinformatics in Brussels (IB2), ULB-VUB, Boulevard du Triomphe

67   CP 263, 1050 Brussels, Belgium. 27: Institute of Parasitology, Biology Centre CAS, Branišovská

68   1160/31, 370 05 ČeskéBudějovice, Czech Republic. 28: Faculty of Science, University of South

69   Bohemia, 370 05 Ceské Budejovice, Czech Republic. 29: Institut Pasteur - Bioinformatics and

70   Biostatistics Hub - C3BI, USR 3756 IP CNRS - Paris, France. 30: IRD - BP 529 - 98713 Papeete,

71   Tahiti. 31: Institute of Microbiology, Department of Biology, ETH Zurich, Vladimir-Prelog-Weg 4,

72   8093 Zurich, Switzerland. 32: IFREMER, Physiology and Biotechnology of Algae Laboratory, rue

73   de l'Iled'Yeu, 44311, Nantes, France. 33: Biological and Environmental Sciences and Engineering

74   Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi

75   Arabia. 34: Sorbonne Universités, UPMC Univ Paris 06, CNRS, Biologie Intégrative des

76   Organismes Marins (BIOM, UMR 7232), Observatoire Océanologique, Banyuls sur Mer, France.

77   35: Directors' Research European Molecular Biology Laboratory Meyerhofstr. 1, 69117

78   Heidelberg, Germany. 36: School of Marine Sciences, University of Maine, Orono, Maine, USA.

79   Dept of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology,

80   Cambridge, Massachusetts 02139, USA. 37: Dept of Earth, Atmospheric and Planetary Sciences,

81   Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. 38: Department of

82   Oceanography, University of Hawaii, Honolulu, Hawaii 96822, USA. 39: Department of Earth

83   Ocean and Ecological Sciences, School of Environmental Sciences, University of Liverpool,

84   Liverpool, UK. 40: Sorbonne Universités (UPMC, Univ Paris 06)-CNRS-IRD-MNHN, Laboratoire

85   LOCEAN, Paris, France. 41: Structural and Computational Biology, European Molecular Biology

86   Laboratory, Meyerhofstr. 1, 69117 Heidelberg, Germany. 42: Max Delbrück Centre for Molecular

87   Medicine, 13125 Berlin, Germany. 43: Department of Bioinformatics, Biocenter, University of

88   Würzburg, 97074 Würzburg, Germany. 44: Cell Biology and Biophysics, European Molecular

89   Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. 45: Institute for Chemical

90   Research, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan. 46: MARUM, Center for

91   Marine Environmental Sciences, University of Bremen, Bremen, Germany. 47: PANGAEA, Data

92   Publisher for Earth and Environmental Science, University of Bremen, Bremen, Germany. 48:

93   Earth Institute, University College Dublin, Dublin, Ireland. [49] National Science Foundation,

94   Arlington, VA 22230, USA. [50]: Bigelow Laboratory for Ocean Sciences East Boothbay, ME, USA.

95   [51]: DVIP Consulting, Sèvres, France.

96   * Correspondence to: pwincker@genoscope.cns.fr, cbowler@biologie.ens.fr, iudicone@szn.it,

97   maurizio@szn.it

98   † These authors contributed equally to this work

99    $ Equal coordinating contribution

100   **Key Points:**

101   • Coherent assemblages of taxa co-varying with iron at global level are identified in plankton

102   communities

103   • Functional responses to iron availability involve both changes in copy numbers of iron-

104   responsive genes and their transcriptional regulation

105   • Plankton responses to local variations in iron concentrations recapitulate global patterns

106

107   **Abstract**

108   Predicting responses of plankton to variations in essential nutrients is hampered by limited *in situ*

109   measurements, a poor understanding of community composition, and the lack of reference gene

110   catalogs for key taxa. Iron is a key driver of plankton dynamics and, therefore, of global

111   biogeochemical cycles and climate. To assess the impact of iron availability on plankton

112   communities we explored the comprehensive bio-oceanographic and -omics datasets from *Tara*

113   Oceans in the context of the iron products from two state-of-the-art global scale biogeochemical

114   models. We obtained novel information about adaptation and acclimation towards iron in a range

115   of phytoplankton, including picocyanobacteria and diatoms, and identified whole sub-

116   communities co-varying with iron. Many of the observed global patterns were recapitulated in the

117   Marquesas archipelago, where frequent plankton blooms are believed to be caused by natural iron

118   fertilization. This work provides a proof-of-concept that integrative analyses, spanning from genes

119   to ecosystems and viruses to zooplankton, can disentangle the complexity of plankton communities

120   and can lead to more accurate formulations of resource bioavailability in biogeochemical models,

121   thus improving our understanding of plankton resilience in a changing environment.

122   **Plain Language Summary**

123   Marine photosynthetic plankton require iron for their energetic metabolism. Indeed, according to

124   John Martin's iron hypothesis, fertilizing the ocean with iron may dramatically increase

125   photosynthetic activity, thus representing a biological means to counteract global warming.

126   However, while there is a constantly growing knowledge of how iron is distributed in the ocean

127   and about its role in cellular processes in marine photosynthetic groups such as diatoms and

128   cyanobacteria, little is known about how iron availability shapes plankton communities and how

129   they respond to it. In the present work, we exploited the recently published *Tara* Oceans datasets

130   to address these questions. We firstly defined specific subcommunities of co-occurring organisms

131   that are directly related to iron availability in the oceans. We then identified specific patterns of

132   adaptation and acclimation to iron in different groups of phytoplankton. Finally, we validated our

133   global results at local scale, specifically in the Marquesas archipelago, where recurrent iron-driven

134   phytoplankton blooms are believed to be a result of iron fertilization. By integrating global data

135   with a localized response we provide a framework for understanding the resilience of plankton

136   ecosystems in a changing environment.

1. Introduction

Marine plankton play critical roles in pelagic oceanic ecosystems. Their photosynthetic component (phytoplankton, consisting of eukaryotic phytoplankton and cyanobacteria) accounts for approximately half of Earth's net primary production, fueling marine food webs and sequestration of organic carbon to the ocean interior. Phytoplankton stocks depend on the availability of primary resources such as nutrients which are characteristically limiting in the oligotrophic ocean. For example, high nutrient low chlorophyll (HNLC) regions are often lacking the key micronutrient iron, and increased bioavailability of iron will typically trigger a phytoplankton bloom *(Boyd et al., 2007)*. Notwithstanding, the community response and its impact on food web structure and biogeochemical cycles are seldom predictable. The composition of blooms when limiting nutrients are supplied as sudden pulses with respect to the pre-existing community has been only poorly explored, and is even more elusive when comparing to situations when nutrients are in quasi-steady-state. Characterizing these responses is crucial to anticipate future changes in the ocean yet is challenged by community complexity and processes that span from genes to ecosystems. Dissecting these processes would also enhance the robustness of existing biogeochemical models and improve their predictive power (Stec et al., 2017).

In this report we explore the responsiveness of plankton communities to iron and assess the representation of iron bioavailability in biogeochemical models. Using global comprehensive metagenomics and metatranscriptomics data from *Tara* Oceans (Guidi et al., 2016; Bork et al., 2015; Alberti et al., 2017; Carradec et al., 2018), we examine abundance and expression profiles of iron-responsive genes in diatoms and other phytoplankton, together with clade composition in picocyanobacteria and the occurrence of iron-binding sites in bacteriophage structural proteins. These profiles are compared in the global ocean with the iron products from two state-of-the-art biogeochemical models. We further identify coherent sub-communities of taxa co-varying with iron in the open ocean that we denote iron-associated assemblages (IAAs). Overall, our findings are congruent with the outputs from the models and reveal a range of adaptive and acclimatory strategies to cope with iron availability within plankton communities. As a further proof-of-concept, we track community composition and gene expression changes within localized blooms downstream of the Marquesas archipelago in the equatorial Pacific Ocean, where previous observations have suggested them to be triggered by iron (Martinez and Maamaatuaiahutapu, 2004), even though the biogeochemical models currently lack the resolution to detect the

168 phenomenon. Our results indicate that iron does indeed drive the increased productivity in this

169 area, suggesting that a burst of the resource can elicit a response mimicking global steady-state

170 patterns.

171 **2. Materials and Methods**

172 2.1. Iron concentration estimates

173 Due to the sparse availability of direct observations of iron in the surface ocean, iron concentrations

174 were derived from two independent global ocean simulations. The first is the ECCO2-DARWIN

175 ocean model configured with 18 km horizontal resolution and a biogeochemical simulation which

176 resolves the cycles of nitrogen, phosphorus, iron and silicon (Menemenlis et al., 2008). The

177 simulation resolves 78 virtual phytoplankton phenotypes. The biogeochemical parameterizations,

178 including iron, are detailed in (Follows et al., 2007). In brief, iron is consumed by primary

179 producers and exported from the surface in dissolved and particulate organic form.

180 Remineralization fuels a pool of total dissolved iron which is partitioned between free iron and

181 complexed iron, with a fixed concentration and conditional stability of organic ligand. Scavenging

182 is assumed to affect only free iron but all dissolved forms are bioavailable. Atmospheric deposition

183 of iron was imposed using monthly fluxes from the model of (Mahowald et al., 2005).

184 PISCES (Aumont et al., 2015) is a more complex global ocean biogeochemical model than

185 ECCO2-DARWIN, representing two phytoplankton groups, two zooplankton grazers, two

186 particulate size classes, dissolved inorganic carbon, dissolved organic carbon, oxygen and

187 alkalinity, as well as nitrate, phosphate, silicic acid, ammonium and iron as limiting nutrients. In

188 brief, PISCES accounts for iron inputs from dust, sediments, rivers, sea ice and continental

189 margins, and flexible Michaelis-Menten-based phytoplankton uptake kinetics result in

190 dynamically varying iron stoichiometry and drives variable recycling by zooplankton and bacterial

191 activity. Iron loss accounts for scavenging onto sinking particles as a function of a prognostic iron

192 ligand model, dissolved iron levels and the concentration of particles. Iron loss from colloidal

193 coagulation is also included and accounts for both turbulent and Brownian interactions of colloids.

194 The PISCES iron cycle we use is denoted as 'PISCES2' (Tagliabue et al., 2016) performed at the

195 upper end of a recent inter-comparison of 13 global ocean models that included iron.

196

7

197   2.2. In situ data

198   To generate a limited dataset of observed dissolved iron concentrations for this analysis, we used

199   a dissolved iron database updated from (Tagliabue et al., 2012). For this we searched for the nearest

200   available observation at the same depth as the *Tara* Oceans sampling and collected data that was

201   within a horizontal radius of 2 degrees from the sampling coordinates.

202

203   2.3. Marquesas archipelago sampling

204   Four stations within the Marquesas archipelago were sampled during the *Tara* Oceans expedition

205   in August 2011 (Karsenti et al., 2011) using protocols described in (Pesant et al., 2015): they were

206   denoted TARA_122, TARA_123, TARA_124, and TARA_125. The sample details and

207   physicochemical parameters recorded during the cruise are available at PANGAEA (http://www.

208   pangaea. de), and nucleotide data are accessible at the ENA archive (http://www. ebi. ac. uk/ena/).

209   The study was initiated by releasing a glider which characterized the water column until the end

210   of the experiment. Firstly, the mapping of the water column structure via real-time analysis of

211   glider data was conducted. After this initial step, the continuous inspection of near real-time

212   satellite color chlorophyll images and altimetric data revealed a highly turbulent environment, with

213   a mixed layer up to 100 m deep and strong lateral shearing, especially downstream of the islands,

214   which generated an area of recirculation in the wake of the main island (Nuku Hiva). A series of

215   four sampling stations was then planned and executed by performing the full set of measurements

216   and sampling using the *Tara* Oceans holistic protocol (Pesant et al., 2015). Station TARA_122

217   sampled the HNLC pre-bloom waters upstream of the islands and thus served as a reference station

218   for the others. This station was located 27 km upstream of the island of Nuku Hiva.

219

220   2.4. Oceanographic observations

221   The Biogeochemical Argo (BioArgo) float deployed in the framework of the Marquesas study

222   (WMO 6900985) was a PROVBIO-1 free-drifter profiler (Xing et al., 2012). It was based on the

223   "PROVOR-CTS3" model, equipped with a standard CTD sensor (to retrieve temperature and

224   salinity parameters) together with bio-optical sensors for the estimation of chlorophyll-a

225   concentrations, Coloured Dissolved Organic Matter (CDOM) and backscatter at 700 nm. It was

226 also equipped with a radiometric sensor to estimate spectral downward irradiance at three

227 wavelengths (412, 490 and 555 nm) and with a beam transmissometer. The data processing is

228 discussed in (Xing et al., 2012). The profiling float was programmed to adopt a modified standard

229 Argo strategy (Freeland et al., 2005). After deployment, it navigated at 700 m depth, to a daily

230 maximum of 1000 m and then surfaced a first time, generally early in the morning. It then

231 submerged again to a depth up to 400 m, to again reach the surface approximately at noon. A third

232 profile to 400 m, followed by a subsequent resurfacing, was performed at the end of the day. During

233 all the ascending phases, a complete profile of all the available parameters was collected. At

234 surface, the obtained data were transmitted to land through a satellite connection and the profiler

235 descended again to 1000 m to start another cycle. The BioArgo was deployed on-site at Station

236 TARA_123 on 2nd August 2011. It performed 55 profiles in the Marquesas region, before moving

237 westward in early October (then outside the study area), and then southward. It definitively ceased

238 to function in December 2012, approximately 400 km south of the Marquesas islands and after

239 collecting more than 150 profiles.

240 An autonomous glider was also deployed in the study area. A complete description of glider

241 technology and functioning is available in (Testor et al., 2010). This glider was able to reach 1000

242 m depths. It was equipped with temperature and salinity sensors, an optode for oxygen

243 concentration measurements, two Wetlab ecopucks with two fluorometers for chlorophyll and

244 CDOM concentrations, and three backscatterometers to estimate backscatter coefficients at three

245 wavelengths (532, 700 and 880 nm). The glider was deployed on 16th July 2011 (approximately

246 one month before *TARA* arrived in the Marquesas archipelago), close to the position of Station

247 TARA_122. It was recovered on 5th August 2011 by *TARA* because a malfunction in the tail rudder

248 had been detected. It performed approximately 250 profiles, with 35 dives at 1000 m depths and

249 90 dives at 500 m depths.

250 Analysis of trace metals was performed following the methods reported in (Scelfo, 1997).

251

252 2.5. Network analysis and correlations with iron

253 A co-occurrence network analysis similar to that reported in (Guidi et al., 2016) was performed to

254 delineate feature subnetworks of prokaryotic and eukaryotic lineages, as well as viral populations,

255 based on their relative abundance. All procedures were applied on 103 sampling sites (Guidi et al.,

9

256   2016) after excluding outliers (Stations TARA_82, TARA_84 and TARA_85) on Hellinger-

257   transformed log-scaled abundances. Computations were carried out using the R package WGCNA

258   (Langfelder and Horvath, 2007). After building a co-occurrence weighted graph, a hierarchical

259   clustering was performed. This resulted in the definition of several subnetworks or modules, each

260   represented by its first principal component, called *module eigen value (ME)*. Associations

261   between the calculated subnetworks and a given trait were measured by the pairwise Pearson

262   correlation coefficients, as well as with corresponding *p* values corrected for multiple testing using

263   the Benjamini & Hochberg FDR procedure, between the considered environmental trait and their

264   respective principal components. The results are reported in the first 10 columns of the heatmap in

265   Figure S1a. The subnetworks that showed the highest correlation scores are of interest to

266   emphasize a putative community associated with a given environmental trait. In addition to the

267   multiple environmental parameters previously reported (Guidi et al., 2016) we simulated iron

268   bioavailability in many *Tara* Oceans stations based on the two different models of iron

269   concentration in the global oceans: the ECCO2-DARWIN model (Menemenlis et al., 2008) and

270   the PISCES2 model (Aumont et al., 2015). Both models performed well in the recent global iron

271   model intercomparison project (Tagliabue et al., 2016) and so we conducted an assessment of

272   model outputs at Tara sampling locations using compilations of iron observations (Tagliabue et al.,

273   2016) augmented by those from the GEOTRACES program (Mawji et al., 2014). ECCO2-

274   DARWIN-derived estimates (57 stations at surface) and PISCES2 model (83 stations at surface,

275   44 of which also at maximum chlorophyll depth) can be found in Table S1a. For further details on

276   the models and for a comparison of the two, please see Supplementary Material. We then identified

277   eukaryotic, prokaryotic and viral subnetworks that correlated most strongly with iron

278   bioavailability, denoted Iron-Associated Assemblages (IAAs). Four IAAs consisting of eukaryotic

279   metabarcodes (de Vargas et al., 2015) were significantly associated with iron. Similarly, four viral

280   IAAs could be identified by analysis of viral communities. Based on taxonomy, no prokaryotic

281   IAAs with significance could be identified, however when considering prokaryotic genes (as

282   described in (Guidi et al., 2016), five subnetworks of prokaryotic genes could be identified.

283   In addition to the network analyses, we examined whether the identified subnetworks can be used

284   as predictors of iron bioavailability. Following the protocol described in (Guidi et al., 2016), we

285   used *Partial least square (PLS) regression*, which is a dimensionality-reduction method that aims

286   to determine predictor combinations with maximum covariance with the *response variable*. The

10

287   predictors were ranked according to their *value importance in projection* (*VIP*) using the R package

288   *pls* (Mevik and Wehrens, 2007). For each eukaryotic IAA, their relative contribution to each

289   sample was estimated by computing the first eigen value.

290   2.6. Taxonomy determinations

291   Taxonomic studies were performed using various methods (photosynthetic pigments, flow

292   cytometry, optical microscopy for phyto- and zooplankton as detailed in (Villar et al., 2015);

293   phytoplankton counts using unfiltered bottles or nets as described in (Malviya et al., 2016) and

294   (Villar et al., 2015); mesozooplankton samples collected by vertical tows with a WP2 net (200 μm

295   mesh aperture) from 100 m depth to the surface during the day, followed by fixation in buffered

296   formaldehyde (2–4% final concentration) and later analyzed in the laboratory. Data from an

297   Underwater Vision Profiler (UVP) were used to determine particle concentrations and size

298   distributions >100 μm, (Campbell et al., 1994). To calculate phytoplankton biomass, the ratio of

299   phytoplankton biomass to Chlorophyll-*a* (Phyto C: Chl a) in the euphotic zone was estimated from

300   (Campbell et al., 1994). To estimate the total Phytoplankton biomass, Chl a concentration from

301   HPLC data was then used. The relative contribution of micro-, nano- and pico- plankton to the

302    $[Chla]_{tot}$ was estimated according to (Uitz et al., 2006). Large zooplankton biomass was estimated

303   using previously published conversion factors from body length to carbon content (C:L) in selected

304   zooplankton lineages. Individual body measures were estimated from literature considering similar

305   community composition, with the exception of the Copepoda prosome length (PL), which was

306   herein measured. Zooscan (Bongo net, >300 μm) derived abundance data (ind x m^{-3}) were used to

307   evaluate the total biomass along the water column.

308   2.7. Genomic analyses

309   Eukaryotes larger than 5 μm were collected directly from the ocean using nets with different mesh

310   sizes while smaller organisms and viruses were sampled by peristaltic pump followed by on-deck

311   filtration. Several filtration steps were performed using membranes with different pore sizes to

312   obtain size-fractionated samples corresponding to viruses (0-0. 1 and 0. 1-0. 2 μm), prokaryotes

313   (0. 2-3 μm) and eukaryotes (0. 8-5, 5-20, 20-180 and 180-2000 μm). In this study, we only used

314   samples collected from the surface water layer. Details about genomics methods are available in

315   (Carradec et al, 2018) and in the following publications: Virus metagenomes (Roux et al., 2016);

316   Prokaryote metagenomes (Sunagawa et al., 2015); Eukaryote metabarcoding (de Vargas et al.,

11

317 2015); Eukaryote metagenomes and metatranscriptomes (Alberti et al., 2017; Carradec et al.,

318 2018). The abundance of individual genes was assessed by normalization to the total number of

319 sequences within the same organismal group. Cyanobacterial clade absolute cell abundance was

320 assessed using the *petB* marker gene, as described in (Farrant et al., 2016), in combination with

321 flow cytometry counts using the method published by (Vandeputte et al., 2017).

322 Metatranscriptomic and metagenomic unigenes were functionally annotated using PFAM (Finn et

323 al., 2016) as the reference database and the search tool (Durbin et al., 1998). To detect the presence

324 of genes encoding silicon transporters, ferritin, proteorhodopsin, FBAI and FBAII among the

325 unigene collection, the profile hidden Markov models of the PFAMs PF03842, PF0210, PF01036,

326 PF00274 and PF01116, respectively, were used, with HMMer gathering threshold option. It is

327 important to note that flavodoxin (PF00253, PF12641 and PF12724), ferredoxin (PF00111), and

328 cytochrome c_6 (PF13442) PFAM families do not discriminate those sequences involved in

329 photosynthetic metabolism from other homologous sequences. The photosynthetic isoforms for

330 flavodoxin, ferredoxin and cytochrome c_6 were therefore determined by phylogeny, as described

331 below.

332 To discriminate the photosynthetic isoforms from other homologous sequences, we started with

333 the results from HMMer and then built libraries composed of well-known reference sequences

334 (manually and experimentally curated) from both photosynthetic and non-photosynthetic groups.

335 To enrich our libraries we used the reference sequences to find similar sequences by using BLAST

336 search tool against phyloDB reference database (Dupont et al., 2015). Next, we used (Katoh et al.,

337 2002) to build multiple sequence alignments and phyML (Felsenstein, 1993) to build the

338 corresponding phylogenetic reference trees, and then manually identified the branches containing

339 the photosynthetic versions and those with non-photosynthetic proteins. We ensured that the

340 bootstrap values of the photosynthetic and non-photosynthetic branches were higher than 0. 7 by

341 retaining only the most conserved matches in our trees. Finally, we realigned and labeled the

342 unigenes against the reference trees depending on the placement of each translated unigene on

343 them.

344 While HMMer has the highest sensitivity among the classical domain detection approaches, not

345 all the references collected by PFAM are sufficiently rich with HMMer to maintain the same

346 detection (Bernardes et al., 2016). To deal with the poor representation of ISIP genes in the PFAM

347 database and to improve their detection, we adopted a simplified version of the approach presented

348 in (Bernardes et al., 2015) to build our own pHMM to detect the different conserved regions

349 represented by ISIP1, ISIP2a and ISIP3 amino acid sequences. For this, we collected all the

350 sequences in the reference literature (Allen et al., 2008; Chappell et al., 2015; Morrissey et al.,

351 2015; Lommer et al., 2015), all 35 sequences belonging to PFAM PF07692, and the 56 most

352 conserved sequences from PF03713 (all the seeds).

353 2.8. Data and code availability

354 Sequencing data are archived at ENA under the accession number PRJEB4352 for the

355 metagenomics data and PRJEB6609 for the metatranscriptomics data (Carradec et al., 2018).

356 Environmental data are available at PANGAEA. The gene catalog, unigene functional and

357 taxonomic annotations, and unigene abundances and expression levels are accessible at

358 http://www. genoscope. cns. fr/tara/. Computer codes are available upon request to the

359 corresponding authors.

360 2.8.1 Accession numbers of metagenomics and metatranscriptomics data

361 2.8.2 Sample:

362 ERS492651, ERS492651, ERS492650, ERS492669, ERS492669, ERS492662, ERS492662,

363 ERS492658, ERS492658, ERS492650, ERS492740, ERS492742, ERS492742, ERS492751,

364 ERS492751, ERS492763, ERS492763, ERS492757, ERS492740, ERS492757, ERS492825,

365 ERS492825, ERS492824, ERS492824, ERS492829, ERS492852, ERS492852, ERS492846,

366 ERS492846, ERS492829, ERS492897, ERS492897, ERS492895, ERS492895, ERS492912,

367 ERS492912, ERS492909, ERS492909, ERS492904, ERS492904

368 2.8.3 Experiment:

369 ERX948080, ERX948010, ERX1782415, ERX1782384, ERX1782327, ERX1796912,

370 ERX1796638, ERX1796690, ERX1796805, ERX1782126, ERX1782109, ERX1782245,

371 ERX1796854, ERX1796544, ERX1782292, ERX1782172, ERX1782221, ERX1796700,

372 ERX1796855, ERX1782301, ERX1782464, ERX1782128, ERX1789668, ERX1789366,

373 ERX948029, ERX948074, ERX1796627, ERX1796773, ERX1789369, ERX1789449,

374 ERX1796931, ERX1796605, ERX1789426, ERX1789575, ERX1789524, ERX1796866,

375 ERX1796524, ERX1789649, ERX1789612, ERX1789647, ERX1796596, ERX1796836,

13

376   ERX1789655, ERX1789574, ERX1789407, ERX1782118, ERX1782283, ERX947973,

377   ERX948088, ERX1789391, ERX1789539, ERX1789587, ERX1796687, ERX1796586,

378   ERX1796703, ERX1789662, ERX1789616, ERX1789589, ERX1796662, ERX1796518,

379   ERX1796678, ERX1796698, ERX1782217, ERX1782352, ERX1796645, ERX1796858,

380   ERX1796924, ERX1789675, ERX1789597, ERX1789700, ERX1789362, ERX1782350,

381   ERX1782418, ERX947994, ERX948064, ERX1789361, ERX1789368, ERX1789532,

382   ERX1796658, ERX1796818, ERX1796632, ERX1789638, ERX1789548, ERX1789579,

383   ERX1796921, ERX1796732, ERX1796741, ERX1789714, ERX1789489, ERX1789628,

384   ERX1796689, ERX1796850, ERX1796523, ERX1782181, ERX1782370, ERX1796607,

385   ERX1796738, ERX1796714, ERX1789437, ERX1789516, ERX1789417

386   2.8.4 Run:

387   ERR868475, ERR868513, ERR1712182, ERR1712118, ERR1711869, ERR1726556,

388   ERR1726667, ERR1726938, ERR1726688, ERR1712207, ERR1711933, ERR1711897,

389   ERR1726927, ERR1726932, ERR1712069, ERR1712197, ERR1711986, ERR1726883,

390   ERR1726891, ERR1712219, ERR1711929, ERR1711951, ERR1719463, ERR1719159,

391   ERR868466, ERR868469, ERR1726762, ERR1726913, ERR1719393, ERR1719310,

392   ERR1726961, ERR1726522, ERR1719437, ERR1719413, ERR1719343, ERR1726622,

393   ERR1726721, ERR1719297, ERR1719410, ERR1719307, ERR1726770, ERR1726561,

394   ERR1719256, ERR1719298, ERR1719217, ERR1711914, ERR1711917, ERR868363,

395   ERR868489, ERR1719301, ERR1719160, ERR1719214, ERR1726564, ERR1726725,

396   ERR1726569, ERR1719448, ERR1719389, ERR1719194, ERR1726571, ERR1726533,

397   ERR1726892, ERR1726601, ERR1711949, ERR1712155, ERR1726608, ERR1726657,

398   ERR1726763, ERR1719391, ERR1719175, ERR1719381, ERR1719365, ERR1711882,

399   ERR1711999, ERR868382, ERR868352, ERR1719395, ERR1719316, ERR1719207,

400   ERR1726643, ERR1726714, ERR1726846, ERR1719404, ERR1719213, ERR1719459,

401   ERR1726822, ERR1726912, ERR1726691, ERR1719356, ERR1719145, ERR1719293,

402   ERR1726695, ERR1726666, ERR1726903, ERR1712102, ERR1711923, ERR1726745,

403   ERR1726946, ERR1726765, ERR1719295, ERR1719249, ERR1719385

14

3. Results

3.1 Modeled iron distributions are highly correlated with the expression of marker genes for iron limitation

Iron is a complex contamination-prone micronutrient whose bioavailability is difficult to assess in the ocean (Tagliabue et al., 2017). Rather than using single discrete measurements, we linked observed differences in plankton communities at sites sampled during the *Tara* Oceans expedition (Bork et al., 2015) with the range of iron conditions typical of each location. Specifically, we extracted annual mean iron concentrations and their variability from two state-of-the-art ocean models (ECCO2-DARWIN (Menemenlis et al., 2008) and PISCES2 (Aumont et al., 2015)), and analyzed their correspondence with the best available estimates based upon in situ data (a compilation of iron observations (Tagliabue et al., 2012) merged with GEOTRACES data (Tagliabue et al., 2012; Mawji et al., 2014) in a manner similar to previous studies (Toulza et al., 2012) (Figure 1).

To assess the reliability of the modeled iron distributions, we correlated the expression of diatom *ISIP* genes in metatranscriptomics datasets with the annual means of iron concentrations estimated by the DARWIN model, and with annual and monthly means by the PISCES2 model (Carradec et al., 2018) (Table S1b and Supplementary Material S1). These genes have been found in multiple previous studies to be inversely correlated with iron availability (Allen et al., 2008; Chappell et al. 2015; Morrissey et al., 2015, Graff van Creveld et al., 2016; Marchetti et al., 2017). Figure 1 presents a comparison between the estimates of dissolved iron concentrations derived from the annual mean iron field from PISCES2 and ECCO2-DARWIN (Figure 1a, b and Table S1a), with the *Tara* Oceans stations superimposed and best available estimates based upon *in situ* measurements (Figure 1c). In spite of the evident scarcity of actual iron concentration data (which illustrates the need to use models for estimating iron in the current exercise; Figure 1c), both models and *ISIP* mRNA levels describe very satisfactorily the global scale gradients, with the highest concentrations of iron observed in the Mediterranean and Arabian Seas (both highly impacted by desert dust deposition) and the lowest in the tropical Pacific and Southern Oceans. This demonstrates that the geographical coverage of the *Tara* Oceans expedition is well suited to studies of the role of iron on sunlit planktonic ecosystems. The available data (Figure 1, Table S1a) further indicates that the gradients of iron appear to be better captured by PISCES2, a more complex and recent model (Aumont et al., 2015). This is for instance the case for the North Atlantic

435 Ocean and the Mediterranean Sea, where longitudinal gradients are stronger in PISCES2 and are

436 consistent with *ISIP* gene levels, while ECCO2-DARWIN seems to overestimate iron in the

437 Eastern Atlantic Ocean and underestimate it in the Mediterranean Sea. The opposite is true in the

438 South Atlantic Ocean, where *ISIP* mRNA levels show a clear increase correlated with iron stress

439 between South America and Africa (Figure 1d). Overall, in the Atlantic ECCO2-DARWIN has

440 higher concentrations, and thus a clearer large scale Atlantic-Pacific gradient is observed.

441 The Pacific and Southern Oceans (subpolar and polar stations TARA 81-85) are both characterized

442 by low levels of iron, as mentioned above. Notably, PISCES2 has a rather flat distribution in the

443 Pacific Ocean, with very low values, while the other model shows a relatively higher level of iron

444 at the core of the subtropical gyres, i.e., close to the Hawaii Islands (Stations TARA_131 and

445 TARA_132) and offshore from South America (TARA_98 and close-by stations) that seems to be

446 in agreement with *ISIP* mRNA levels (at least for the Hawaiian sample – Figure 1d). These are

447 very oligotrophic oceanic regions, where nitrate is also a strongly limiting nutrient. Again, the *ISIP*

448 expression pattern in Figure 1d is closer to the PISCES2 model, in that it shows a clear reduction

449 of the stress resulting from iron deprivation within these gyres. Finally, while a significant increase

450 in iron at the Equator may be expected as a consequence of the upwelling in this region, both the

451 models and the *ISIP* levels (at Station TARA_128) suggest that this area is rather characterized by

452 low values of iron. Overall, our analysis indicates that both models correlated very well with *Tara*

453 Oceans transcriptomic data, with no relevant differences among monthly and yearly values and

454 with annual means from the ECCO2-DARWIN estimates showing the best reliability (Table S1b).

455 This analysis also indicates that metatranscriptomics is now mature enough to provide an

456 independent, biologically-based validation of ecosystem models.

457

3.2 Plankton response to iron availability is coordinated at sub-community level

459 The higher level organization of plankton communities, and its possible relationship with the roles

460 of individual constituents, has been highlighted previously in an analysis of the potential links

461 between community structure and carbon export using data from *Tara* Oceans (Guidi et al., 2016).

462 We here used this approach to explore plankton ecosystem responses to iron bioavailability using

463 an end-to-end approach from genes to communities and from viruses to metazoa to reveal

464 community responses at global scale (see Methods). Known as weighted gene correlation network

16

465 analysis (WGCNA; see Methods for further description, Guidi et al., 2016; Langfelder and

466 Horvath, 2007) this approach deciphers sub-communities (modules) of organisms within a global

467 co-occurrence network, and because of the high levels of co-variation of individual taxa it is

468 possible to deduce putative ecological interactions. As proxies for organism abundance we used

469 the relative abundances of eukaryotic lineages (defined as operational taxonomic units; OTUs)

470 derived from 18S-V9 rDNA metabarcoding data (de Vargas et al., 2015). WGCNA generated a

471 total of 31 modules. Each module groups a subset of eukaryotic taxa found in *Tara* Oceans samples

472 whose pairwise relative abundance was highly correlated over all the sampling sites, i.e., they have

473 a high probability of co-occurrence and to change their abundance in a coordinated way. Because

474 they react in phase, we can infer that within each sub-community these organisms have a higher

475 probability of interaction among themselves than with the organisms in other modules.

476 We found four eukaryotic subnetworks significantly associated with the ECCO2-DARWIN-

477 derived and/or with the PISCES2-derived estimates of iron concentrations in the global ocean

478 (Figure 2a, Figure S1a, b, Table S1c). The Black and Turquoise modules were associated with high

479 significance to the iron concentrations generated by both models whereas the DarkRed and Yellow

480 modules were better associated with ECCO2-DARWIN and PISCES2, respectively: Black

481 (DARWIN: R=0.37, P=6×10^{-4}; PISCES2: R=0.38, P=3×10^{-4}), Turquoise (DARWIN: R=0.46, P=

482 1×10^{-5}; PISCES2: R=0.42, P=9×10^{-5}), DarkRed (DARWIN: R=-0.43, P=5×10^{-5}; PISCES2: R=0.

483 19, P=0.08), and Yellow (DARWIN: R=0.19, P=0.09; PISCES2: R=0.56, P=5×10^{-8}), and contained

484 between 31 and 591 different OTUs (Tables S1c and S1d). These subnetworks were denoted iron-

485 associated assemblages (IAAs). For each IAA subnetwork, WGCNA computes a single

486 representative as a combination of lineages. Such a score, denoted a 'module eigengene' score

487 (hereafter termed an eigenlineage score) represents the first eigenvector of the assemblage

488 (Langfelder and Horvath, 2007). Projections of samples on such an eigenvector show the relative

489 importance of samples to the global variance of each IAA. Together with their contribution, in

490 terms of OTU abundance to the total eukaryotic abundance in each station (Table S1e), they

491 provide clues to interpret the link between modules and iron availability. The mismatch in some

492 regions between the two models (see above) is likely the reason why the significance of association

493 of the Yellow module with ECCO2-DARWIN, whose variance and representativeness is

494 particularly significant in the South Adriatic and is minimally present in the Peruvian upwelling

495 area, is much less than that with PISCES2. By contrast, the DarkRed module, which appears to be

17

496 the best indicator module for the Marquesas area (Figure 2b, upper panel) and is highly relevant

497 in the Peruvian upwelling region, displays a much less significant association and an opposite

498 variation with PISCES2 iron vs. ECCO2-DARWIN iron. The IAAs show slightly different, often

499 antagonistic, variance contributions at global scale (Figure 2b, upper panel), with each of them

500 being particularly responsive, in terms of variance, in specific sites, e. g., the Yellow module in the

501 Eastern Mediterranean Sea.

502 We examined the lineage composition of each IAA and the relevance of each taxon within them

503 by determining the relative abundance of each lineage with respect to iron concentration estimates

504 and their centrality within the module (see Methods). The results are reported in Table S1c, d. The

505 IAAs displayed significant differences in terms of numbers of lineages and compositions, with the

506 Turquoise module being the largest and dominated by consumers, predominantly metazoans, and

507 the DarkRed module being the smallest. The Black module displayed the highest proportion of

508 autotrophs, while the DarkRed IAA displayed the highest proportion of diatoms (Bacillariophyta;

509 57% of all autotrophic protists).

510 To reduce complexity further, we screened the networks in terms of the VIP score of each node

511 (i.e., the OTUs displaying the highest statistical weight in differentiating sites because of iron

512 availability; Methods, Table S1c, Figure 2a and Figure S1c). Species with high VIP scores can be

513 predicted to be particularly important in reflecting the adjustments of each module via their specific

514 interactions with other members of their sub-community. Considering the ECCO2-DARWIN-

515 derived VIP scores, lineages with the highest scores (>1) could predict as much as 61.9%, 52.6%,

516 49.1%, and 38.1% (in the Turquoise, Black, DarkRed and Yellow IAAs, respectively; leave-one-

517 out cross validated - LOOCV) of the variability of iron in the oligotrophic ocean. When the

518 PISCES2-derived VIP scores are taken into account, the predictive potential of the IAAs is even

519 higher: 73.2% (Turquoise), 61.9% (Yellow), 59.0% (Black) and 54.4% (DarkRed). More

520 importantly, the VIP scores obtained with the two models for each OTU showed an extremely good

521 covariance (Figure S1d). This confirms the biological coherence and stability of the modules and

522 their components to iron availability despite the occasional mismatch in the predictions of the two

523 models.

524 Of the photosynthetic groups, autotrophic dinoflagellate taxa were particularly relevant in the

525 Turquoise and Black modules, diatoms were relevant in the DarkRed module, and haptophytes

18

526   were significantly present in the Yellow module. Metazoans were particularly important in the

527   Black and the Turquoise modules, and MAST/MALV groups of phagotrophic and parasitic

528   heterotrophs were relevant in the Black (MALV), Turquoise and Yellow modules (MAST) (Figure

529   2a, Figure S1c, and Table S1c, d). This hints at particularly intricate, and still elusive, interactions

530   among organisms which ultimately lead to the observed collective responses.

531   To further interpret the patterns observed for the IAAs we chose two additional modules, denoted

532   DarkGrey and Red, because of the different correlations of diatoms within these modules to iron

533   concentrations with respect to the DarkRed module (Figure S1a, c). By examining the abundance

534   of the components of each module at different sampling sites (Table S1e), the results suggest that

535   the Turquoise module groups lineages relevant in all of the main oceanic biogeographic regions

536   with the exception of the Mediterranean basin, and with a prominent weight in the Southern Ocean.

537   By contrast, the Black and Yellow modules are of particular importance in the Mediterranean Sea,

538   while other IAAs have minor contributions. The DarkRed module is generally poorly represented,

539   however in the South Pacific and in particular around the Marquesas Islands, its relevance is high

540   (Figure 2b, upper panel and Table S1e).

541   Based on all of the above information, we then sketched the ecological profiles of the seven

542   modules, summarized below:

543   Black IAA: Ubiquitous, but with low abundance except in the Mediterranean basin, and composed

544   principally of heterotrophic organisms (protists and metazoans) (Table S1d, e). Dinophytes are the

545   autotrophic component of this module while diatoms are poorly represented. Around the

546   Marquesas Islands, its weight is constantly low. Lineages are positively correlated or loosely

547   anticorrelated with iron (Figure 2a, b and Table S1c). This module has an intermediate level of

548   internal connectivity and suggests top heavy (pyramidal) trophic interactions. The assemblage

549   resembles a typical pattern in a post-bloom phase, with biomass accumulated in the metazoan

550   compartment. No significant differences are seen when the ECCO2-DARWIN-derived and

551   PISCES2-derived VIPs are compared since the module is not relevant in areas where the two

552   models disagree. This pattern is consistent with the differences detected at molecular level.

553   DarkRed IAA: The module is not particularly significant at global scale in terms of abundance

554   (Figure 2b, upper panel and Table S1d). It contains a small number of lineages with a high relative

555   weight of diatoms and few metazoans but no copepods, with carbon recycling mostly in the

556   protistan compartment. This module is particularly intriguing because, with very few exceptions,

557   all the lineages including diatoms are negatively correlated with iron (Figure S1c). It is particularly

558   responsive in the Marquesas area but is also present in offshore South American upwelling areas.

559   The internal connectivity is of an intermediate level (Table S1c). These features hint at an

560   assemblage in the subtropical ocean driven by the activity of diatoms thriving in regions of low

561   iron availability (while exploiting a higher than average silicon availability), thus showing an

562   inversion of the pattern compared to high iron regions (Figure 2b, upper panel). Significantly, its

563   abundance drops at Station TARA_123 in the Marquesas archipelago (see below).

564   <u>Turquoise IAA</u>: Ubiquitous, with a general high weight in terms of abundance, and very abundant

565   in the Southern Ocean (in particular in stations TARA_85-88; Table S1e). The module includes

566   relatively few diatoms, but many dinoflagellates (both autotrophic and heterotrophic species)

567   (Tables S1c, d). Copepods are the most numerous component and show the highest VIP scores. Of

568   note, this module includes the crustacean order *Euphausiacea* (krill), which specifically emerges

569   as having high VIP scores only when the PISCES2-derived iron estimates are used. Both internal

570   connectivity and number of lineages are high (Table S1c). The module as a whole responds in the

571   Marquesas area, especially at TARA_123 (Figure 2b, upper panel and Table S1e).

572   <u>Yellow IAA</u>: This module is particularly important in South Adriatic and Eastern Mediterranean,

573   as well as in the tropical North Atlantic (Figure 2b, upper panel and Table S1e). It includes

574   relatively few metazoans and diatoms but a notable abundance of haptophytes and heterotrophic

575   protists (Table S1c, d). Displays a weak response in the Marquesas area (TARA_125) (Figure 2b,

576   upper panel) and seems to be less dependent on iron availability as compared to the other modules.

577   <u>DarkGrey</u>: Not an IAA and has a low weight in general, with a slight positive correlation to iron

578   and only low internal connectivity. Diatoms in this module are very relevant (Table S1c, d).

579   Contains a high fraction of metazoans with fewer heterotrophic protists. This module displays a

580   typical bottom heavy (pyramidal structure) with diatoms reacting positively to iron availability.

581   <u>Red module</u>: Not an IAA, but this module displays a similar response to iron than the DarkGrey

582   module, with the main differences being that it contains few metazoans and the protist

583   compartment is dominated by Dinophyceae. Diatoms are also dominant as autotrophic protists. It

584   is the module that correlates the most with chlorophyll and primary productivity (Figure S1a) and

585   seems to be associated with highly productive areas. It is thus not very relevant globally, with the

586 exception of the South Atlantic Ocean, where it dominates the Benguela upwelling (Station

587 TARA_67), a very rich region that is not iron limited. It is apparently driven by bottom-up flexible

588 responses to iron availability, most probably by macronutrient availability (Table S1c, d). It

589 displays variable correlations of its members to iron, and has also a bottom heavy pyramidal

590 trophic structure.

591 Overall, our analysis strongly suggests that different sub-assemblages of co-occurring lineages can

592 be pinpointed within communities that respond differently to resource limitation, mostly without

593 marked geographical preferences albeit with high plasticity to iron availability. Particularly

594 remarkable is the contrasting role shown by diatoms, with different lineages covering the full range

595 of correlations with iron (Figure S1c), possibly linked to their different strategies for responding

596 to the lack of a crucial resource. In some cases their communities share a similar response while

597 in others the structure of the assemblage is modified. The further observation that co-occurrence

598 of IAAs can show biogeographical patterns (Figure 2b, upper panel) that are not clearly

599 emphasized by analysis of single eukaryotic groups (Figure S1c) is suggestive of a

600 compartmentalization of communities in sub-communities or modules. Our analysis also infers

601 that it is the module as a whole that responds to perturbation, reinforcing the need to dissect

602 plankton responses to iron bioavailability at community scale, while investigating the

603 physiological responses of key species.

604 Interpreting why high VIP taxa are related to iron bioavailability is severely restricted by our

605 knowledge of plankton functional ecology and inter-organism interactions. Nonetheless, the

606 identification of an IAA in which several diatoms have the highest VIPs (DarkRed module, 8

607 subnetwork members, -0.337 correlation with iron), commonly found in the most severely iron-

608 limited regions of the world's ocean and often the most responsive groups in mesoscale iron

609 fertilization experiments (Boyd et al., 2007; Marchetti et al., 2006), suggests a strong physiological

610 plasticity of these groups. The fact that *Pseudo-nitzschia* is among the highest scoring VIP genera

611 in the DarkRed module further suggests that this genus tracks regions with low iron bioavailability,

612 being able to profit from it when it becomes available. Concerning metazoans, copepods from the

613 genus *Temora* (high subnetwork centrality, strong correlation with iron) are known to be iron-

614 limited (Chen et al., 2011), and the two cnidarian lineages - the class *Hydrozoa* and the genus

21

615   *Pelagia* (both of which display relatively strong subnetwork centrality and strong correlations with

616   iron) - suggest strong predator-prey links.

617   In addition to eukaryotes, WGCNA analysis was also performed on prokaryotic communities, as

618   well as on prokaryotic genes from the Ocean Microbial Reference Gene Catalog (Sunagawa et al.,

619   2015; Alberti et al., 2017). Using relative abundances of prokaryotic 16S rDNA miTags, no

620   subnetwork could be associated significantly with iron (maximum r=0.19, $P<10^{-2}$). However,

621   following the same procedure but using the relative abundances of prokaryotic genes rather than

622   taxa, five subnetworks were significantly associated with iron (ECCO2-DARWIN iron data;

623   Figure 2b, lower panel and Table S1f) ($P<10^{-5}$): Grey60 (r= 0.38, P=6. 10^{-5}), Plum1 (r= 0.54, P=3.

624   10^{-9}), Red (r=-0.42, P=10^{-5}), SkyBlue (r=-0.44, P=2. 10^{-6}), and SaddleBrown (r=-0.47, P=6. 10^{-7}).

625   VIPs obtained from each of the two models displayed high correlations (Grey60= 0.99,

626   Plum1=0.94, Red=0.99, SkyBlue=0.96, SaddleBrown=0.98). The VIP genes of the SaddleBrown

627   subnetwork represent 25% (N=41) of the total number of genes, and several genes that could be

628   functionally identified encode proteins associated with iron transport, saccharopine

629   dehydrogenase, aminopeptidase N, and ABC-type transporters (Table S1f). The Plum1 subnetwork

630   is a small subnetwork of around 100 genes that is solely associated with iron concentration

631   variability, and 30% of its VIP genes encode principally specialized functions defined as non-core

632   functions in a previous study of the *Tara* Oceans Global Ocean Microbiome (Sunagawa et al.,

633   2015) (Table S1f). Not surprisingly, 75% of the genes within this subnetwork encode proteins with

634   unknown functions, although some known functions are linked to iron, such as ferredoxin and

635   regulation of citrate/malate metabolism. The contribution to the global variance by stations located

636   within the Red Sea (Stations TARA_31-34) is particularly high (Figure 2b, lower panel). The Red

637   subnetwork is very large, composed of 3,059 genes. However, only 9% represent high scoring

638   VIPs, among which functions related to iron are evident (e.g., ABC-type Fe^{3+} siderophore transport

639   system, putative heme iron utilization protein, metalloendopeptidase - Table S1f). Finally, the

640   SkyBlue subnetwork is a small subnetwork (172 genes) containing 33% of VIPs whose functions

641   are generally unknown (Table S1f). The global variance of this gene subnetwork can be correlated

642   principally with several oligotrophic regions of the Pacific Ocean (e. g., Stations TARA_93, 100,

643   112, 128).

644   In summary, association of prokaryotes with iron is detectable at the functional level (gene

645   abundance) but not at the taxonomic level, which would suggest a low level of specialization, at

22

646 least with the resolution allowed by the 16S marker. To further analyze this aspect we focused on

647 *Prochlorococcus* and *Synechococcus*, the two most abundant and widespread

648 bacteriophytoplankton in the global ocean, and for which a higher resolution genetic marker is

649 available. Combining the information from the taxonomic marker *petB*, which encodes

650 cytochrome b_6 (Farrant et al., 2016), with flow cytometry cell counts we estimated the absolute

651 cell abundance of the picocyanobacterial clades and found that many of them have a strong

652 correlation with predicted iron levels from PISCES2 (Figure 3a) and ECCO2-DARWIN models

653 (not shown). *Prochlorococcus* HLIII and IV ecotypes showed the highest anti-correlation with

654 iron, in agreement with previous descriptions that they are the dominant populations in HNLC

655 areas (Rusch et al., 2010; West et al., 2011). Prochlorococcus LLI, a minor component in surface

656 waters, also showed anticorrelation with iron. In the case of *Synechococcus*, the strongest positive

657 correlation was found for clade III, whereas a weaker pattern is displayed by clade II. On the

658 contrary, CRD1 showed the highest negative correlation with iron, consistent with it being reported

659 as the major *Synechococcus* clade in HNLC regions (Farrant et al., 2016; Sohm et al., 2016). In

660 addition, clade EnvB also displayed a negative correlation with estimated iron concentrations.

661 These results demonstrate that iron affects picocyanobacterial community composition and raise

662 the question of whether the lack of correlation with taxonomic networks depends on a poor

663 taxonomic resolution or to being more pronounced for autotrophs with respect to heterotrophs.

664 Finally, and in view of the present debate on the impact of viruses in iron biogeochemistry (Brum

665 et al., 2015; Bonnain et al., 2016), we used relative abundance of viral populations (Brum et al.,

666 2015) to apply WGCNA and explore whether the viral module subnetworks display any kind of

667 association to the same suite of environmental factors used above for prokaryotes and eukaryotes

668 (Figure S2a). We found four viral IAAs significantly associated with iron using the ECCO2-

669 DARWIN iron estimates: DarkGreen (r=0.59, P=10^{-7}), SteelBlue (r=0.48, P=$10^{-5)}$, LightCyan1

670 (r=0.63, P=10^{-8}), and Tan (r=0.47, P=10^{-4}) (Figure S2b). The association of these modules with

671 iron was strongest with PISCES2. Projection of samples on eigenlineages shows the relative

672 contribution of samples to the variance in the global oligotrophic ocean (Figure S2b). Interestingly,

673 the IAAs again show antagonistic contributions at global scale, with DarkGreen and LightCyan1

674 subnetworks responding most strongly in the Red Sea (like the Yellow eukaryotic subnetwork).

675 Within the viral IAAs, the highest VIP populations were detected in multiple oceanic basins,

676 particularly in the Red Sea and Indian Ocean (Roux et al., 2016), consistent with the network

23

677   analysis results. The highest VIP populations were not associated with specific abundance patterns

678   or hosts: these populations included both rare and abundant viruses, and had predicted hosts in the

679   alphaproteobacteria (SAR11 and SAR116), Gammaproteobacteria, Betaproteobacteria,

680   Deltaproteobacteria and cyanobacteria (*Synechococcus* and *Prochlorococcus*) groups (Figure

681   S2b). Besides these examples, the vast majority of viruses in the IAAs have unknown host ranges.

682   Since the WGCNA results suggest that eukaryotes are the principal drivers of the iron response,

683   this finding may be indicative that they are predominantly eukaryotic viruses, for which very little

684   is known (Chow and Suttle, 2015; Bonnain et al., 2016).

685   Overall, our global analysis confirms that geochemistry affects the structure of plankton

686   assemblages down to the level of viruses. Although unlikely to be influenced directly by iron,

687   viruses appear to be an integral component of the dynamics of plankton consortia, indicating that

688   the response to resource availability occurs at the level of sub-communities which respond in a

689   coordinated way and impact the organization of the whole community across kingdoms.

690   In addition to the potential impact of viruses through the release of iron during host lysis, there is

691   a current discussion about their potential role in complexing iron (Bonnain et al., 2016). To explore

692   this latter point, we surveyed the *Tara* Oceans metagenomes for genes encoding viral structural

693   proteins with putative iron-binding sites. Specifically, we searched for paired histidine residues

694   (HxH motifs) in tail proteins (Bartual et al., 2010) and baseplate assembly proteins (Browning et

695   al., 2012), which has been experimentally implicated in the octahedral coordination of iron. We

696   also analysed the presence of four conserved cysteine residues involved in the coordination of a

697   4Fe-4S cluster in tail tip proteins (Tam et al., 2013). Remarkably, these potential iron-binding

698   motifs are present in 87% unigenes encoding viral tail proteins, 47% of baseplate assembly

699   proteins and 12% in those coding for tip proteins (Figure S2c-e). The corresponding viral contigs

700   are distributed ubiquitously and with high abundance (Figure S2 c-e), suggesting that a significant

701   fraction of colloidal iron may be associated with viruses in the ocean, a factor that is not currently

702   considered in the modeling of ocean biogeochemistry. The question is then how substantial this

703   contribution could be. (Bonnain et al., 2016) made a broad estimation based on the number of iron

704   ions experimentally determined in tails of non-marine phages, and the amount of tailed viruses

705   typically found in marine surface waters. They thereby suggested that between 6 and 70% of the

706   colloidal iron fraction from surface waters could be bound to tail fibers of phages. In this context,

707   the recent "Ferrojan Horse Hypothesis" posits that iron ions present in phage tails enable phages

708 to exploit their bacterial host's iron-uptake mechanism, where the apparent gift of iron leads to cell

709 lysis (Bonnain et al., 2016). Although our analysis does not allow to confirm this hypothesis

710 (Bonnain et al., 2016), it provides a useful context to explore it further.

711

712 3.3 Functional responses are mediated either by changes in gene copy number or by

713 expression regulation

714 Given the clear patterns in the community responses to iron availability, we next wondered which

715 molecular patterns were associated with them. We first examined the prevalence of the diatom *ISIP*

716 genes in more detail using both metagenomics and metatranscriptomics data to detect changes in

717 gene abundance and expression, respectively. We found that both the abundance and expression of

718 this gene family displayed a strong negative correlation with iron (Figure 3b and Figure S3a).

719 Figure 3b shows a strong hyperbolic profile of *ISIP* gene abundance and mRNA levels with respect

720 to iron concentrations (non-linear regression fitness of 97.01 and 98.14, respectively; Table S1b).

721 Furthermore, density clustering algorithms detected two types of responses – stations in which

722 *ISIP* was only increased in metagenomics data (denoted Group 0) and others in which both

723 metagenomic and metatranscriptomic data showed increases in *ISIP* levels (denoted Group 1;

724 Figure 3b). The former likely correspond to locales where *ISIP* copy numbers vary in diatom

725 genomes as a function of iron, implying that the diatoms at these stations display permanent genetic

726 adaptations to the ambient iron concentrations, whereas the latter display transcriptional variation,

727 indicative of more flexible short-term acclimatory rather than permanent adaptive evolutionary

728 processes. Taxonomic analyses revealed that diatoms from the *Thalassiosira* genus were typical

729 of Group 0, whereas *Pseudo-nitzschia* was found largely in Group 1 (Figure 3c). Representatives

730 from both these genera are well known to respond to fluctuations in iron (Marchetti et al., 2017;

731 see Supplementary Material S1), so these different iron-response strategies may underlie why they

732 are present in different IAAs; *Thalassiosira* is present in the Black and Turquoise IAAs whereas

733 *Pseudo-nitzschia* is only present in the DarkRed module, where it is negatively correlated to iron

734 (Figure 2a, Table S1c).

735 It is interesting to note that sampling sites can be grouped in a similar way according to either their

736 picocyanobacterial community or diatom *ISIP* patterns in relation to iron levels (Figure 3d). HLIV

737 and HLIII codominate the *Prochlorococcus* community in Group 1 stations, and these sites are

738 also characterized by the presence of LLI, as well as the *Synechococcus* clades CRD1 and EnvB.

25

739 Based on picocyanobacteria community composition, these stations tend to cluster together in a

740 group of low-iron stations from Indian and Pacific Oceans (TARA_52, 100, 102, 110, 111, 122,

741 124, 125, 128, 137). On the contrary, Group 0 ISIP stations were dominated by either

742 *Prochlorococcus* HLI or HLII and by *Synechococcus* clades II or III. Among these stations, those

743 from the high-iron Mediterranean Sea (TARA_7, 9, 18, 23, 25, 30) clustered together based on

744 picocyanobacteria community composition (Figure 3d).

745 Besides diatoms, we carried out a detailed analysis of *ISIP* distributions among other

746 phytoplankton taxa. We found that in chlorophytes Fea1-domain-encoding genes (related to

747 *ISPI2a*; Marchetti et al., 2017) vary in copy number as a function of predicted iron levels, and that

748 *ISIP* expression also varies in haptophytes and pelagophytes (Figure S3a). Dinoflagellates display

749 the lowest correlations of *ISIP* gene abundance and expression with respect to iron. This may

750 indicate that dinoflagellates respond differently to iron concentrations or with different genes.

751 We continued the analysis of additional iron response genes within the five principal groups of

752 photosynthetic eukaryotes (*Bacillariophyta*, *Chlorophyta*, *Dinophyceae*, *Haptophyceae*, and

753 *Pelagophyceae*) (Figure S3). A significant component of the cellular quota of iron in a

754 photosynthetic cell is the photosynthetic electron transport chain. The soluble electron shuttles of

755 photosynthesis can vary according to metal bioavailability in the environment. For example, while

756 the copper-containing plastocyanin is able to functionally replace the hemoprotein cytochrome c_6

757 (Peers and Price, 2006), the iron-free flavodoxin can substitute the Fe-S containing ferredoxin

758 (Pierella Karlusich et al., 2015). As these proteins also include isoforms involved in other

759 metabolisms, or constitute functional domains of complex multidomain redox proteins, the

760 unigenes corresponding to the photosynthetic isoforms were determined by their phylogenetic

761 placement on manually curated reference trees (see Methods). We were unable to detect any strong

762 correlation between *ferredoxin* expression and iron levels in any of the photosynthetic groups

763 studied (Figure S3a). Regarding its gene abundance, we also found no clear pattern either, although

764 this is expected because analysis of sequenced genomes shows that *ferredoxin* is a core gene in

765 photosynthetic eukaryotes and that there are no large differences in the copy number of the

766 canonical photosynthetic isoforms among related species. On the other hand, our analysis of all

767 five photosynthetic groups indicated that the *flavodoxin* gene is highly iron-responsive (Figure

768 S3a).

26

769 In diatoms, the genes encoding the photosynthetic flavodoxin and ISIP1 contain a palindromic

770 motif involved in iron-dependent gene regulation (Yoshinaga et al., 2014; Lommer et al., 2015).

771 Conversely, it has been shown in diatoms that *flavodoxin* and *ISIP3* genes are highly induced under

772 low iron conditions while *ferredoxin* expression was not affected (*Marchetti* et al., *2017*), and in

773 conditions of chronic iron deficiency induction of the *flavodoxin* gene was at least an order of

774 magnitude higher than the repression of the *ferredoxin* gene (Lommer et al., 2010; Whitney et al.,

775 2011; Graff van Creveld et al., 2016). The analysis of sequenced genomes and transcriptomes

776 further revealed that patterns of flavodoxin gene absence/presence is strongly negatively correlated

777 with environmental iron levels in photosynthetic organisms (Pierella Karlusich et al., 2015).

778 Therefore, our results from the gene copy number variation in the five phytoplankton groups are

779 consistent with previous studies, although the trends for diatoms are weak (Figure S3a).

780 In relation to the plastocyanin/cytochrome c_6 switch, it has been proposed that coastal diatoms in

781 iron-rich environments use cytochrome c_6 as electron carrier, which is replaced by the copper-

782 containing plastocyanin in open ocean environments. In our analysis diatom *cytochrome c6*

783 abundance/expression showed no clear tendency with respect to the simulated iron concentrations,

784 although the *plastocyanin* gene showed a strong negative correlation (Figure S3b). Additionally,

785 our results indicate that pelagophytes use only cytochrome c_6 while chlorophytes tend to use only

786 plastocyanin, as only nine distinct *cytochrome c6* unigenes could be assigned to this group. In

787 haptophytes and dinoflagellates, genes encoding cytochrome c_6 showed no clear changes in

788 abundance with respect to iron, while the *plastocyanin* gene showed a negative correlation in

789 abundance in dinoflagellates and in mRNA levels in haptophytes (Figure S3b).

790 A similar analysis was performed to examine the abundance and expression of type I (metal-free)

791 and type II (iron-containing) FBAs (Allen et al., 2012). We found that the *FBAII* gene showed a

792 clear up-regulation at high-iron stations in all groups, while diatoms showed a concomitant

793 reduction in *FBAI* gene abundance and mRNA levels, pelagophytes and dinoflagellates displayed

794 decreased gene abundance, while haptophytes displayed a response at the mRNA level (Figure

795 S3b). The chlorophytes displayed no consistent trends.

796 In summary then, while the FBAI/FBAII switch appears to display a broad distribution throughout

797 all the phytoplankton groups examined (with the exception of chlorophytes), the substitutive

798 responses   affecting   photosynthetic   electron   shuttles   (flavodoxin/ferredoxin,

27

799 plastocyanin/cytochrome c_6) tend to display iron-dependent responses of the genes encoding the

800 iron-free proteins but not the iron-dependent proteins (with the exception of *plastocyanin* in

801 pelagophytes, which appears to be absent). Analyses of other putative iron-responsive components

802 such as ferritin and proteorhodopsin were also carried out, as well as silicate responsive Si

803 transporters (see Supplementary Material S1). Collectively, our results indicate that individual

804 genes implicated in iron metabolism in specific organismal groups do not provide an unequivocal

805 evaluation of iron availability in the environment, and are thus of only limited use as sentinel genes

806 of iron bioavailability. Instead, the integration of all these iron-driven patterns, spanning from

807 genes to ecosystems, is a promising strategy for designing omics-enabled tools that can improve

808 the representation of key nutrients in biogeochemical models. In this sense, the co-variation of

809 picocyanobacterial communities with the transcriptional regulation and altered copy numbers of

810 diatom *ISIP* genes can potentially be exploited to predict actual iron bioavailability in the

811 ecosystem (Figure 3d).

812 **3.4 Plankton respond to a resource burst in the Marquesas archipelago by reorganization of**

813 **IAAs**

814 The global analyses of IAAs and iron responsive genes in the context of the ranges of geographic

815 iron availability provide a first order approximation of plankton community structure organization

816 and responses for large scale, iron-linked biogeochemical regions. In other words, they possibly

817 reflect the integrated, albeit diversified, response to average conditions and in a stationary or quasi-

818 stationary phase. They further provide support for the iron products of the two biogeochemical

819 models. We reasoned that they might also be able to indicate increases in iron in regions where

820 biogeochemical models do not have sufficient resolution and to highlight mechanisms in action

821 when the resource is provided in bursts that drive the community out of a previous steady state,

822 e.g., leading to blooms. One such case is the Marquesas archipelago in the sub-tropical Pacific

823 Ocean, where previous studies (Martinez and Maamaatuaiahutapu, 2004) have highlighted a

824 dynamic natural perturbation resulting in perennial plankton blooms that are visible from space.

825 Although iron concentrations have not been measured extensively in the region, these and similar

826 blooms (Gong et al., 2016) are triggered by different processes due to the presence of the islands

827 (vertical mixing, horizontal stirring, local precipitation and runoff) which are typically coupled to

828 iron injection (Martinez and Maamaatuaiahutapu, 2004), a phenomenon that has been termed

829 Island Mass Effect (Gove et al., 2016). We therefore focused on this region to examine the

28

830   relationship between the global patterns in plankton sub-communities and iron-responsive gene

831   abundance/expression in a more localized dynamic setting (Supplementary Material S2-S4).

832   Satellite chlorophyll estimates showed that in the days preceding the visit of the *Tara* Oceans

833   expedition to the archipelago in August 2011 the area was characterized by intense variability. Our

834   analyses also revealed a highly turbulent environment, with mixing up to 100 m depth and strong

835   lateral shearing downstream of the islands, which generated an area of recirculation in the wake of

836   the main island and the formation of small eddies where the blooms were occurring (Figs. 4a and

837   S4a). Station TARA_122 sampled the HNLC pre-bloom waters upstream of the islands (Figure

838   4a). Waters of Station TARA_122 were characterized by low chlorophyll concentrations in the

839   water column ([Chl-*a*]$_{int:}$ 16.6 mg·m^{-2}) but high concentrations of nutrients (NO$_2$: 0.12 mmol. m^{-3},

840   PO$_4$: 0.57 mmol. m^{-3}, NO$_2$NO$_3$: 5.5 mmol. m^{-3}, Si: 2. 2 mmol. m^{-3}; Figure S4b and Table S2a),

841   characteristic of an HNLC region (Smetacek and Naqvi, 2008; Quéguiner 2013). Station

842   TARA_123 is coastal, 8 km downstream of Nuku Hiva island and with a seabed depth of 1,903 m

843   and higher chlorophyll levels ([Chl-*a*]$_{int}$ 33.6 mg·m^{-2}), indicative of a bloom. Nutrients were as

844   elevated as in the pre-bloom HNLC area, and with a particular increase of NO$_2$ around 150 m

845   depth (1.47 mmol·m^{-3}). Station TARA_124 is away from the coast, 43 km from Nuku Hiva, in

846   even deeper water (2,414 m bottom depth), and in an eddy also characterized by high chlorophyll

847   content with respect to Station TARA_122 ([Chl-*a*]$_{int}$ 28.5 mg·m^{-2}). The chlorophyll patch was

848   possibly seeded near the islands and transported by currents far from the coast, but sustained by

849   the eddy dynamics and its interaction with underlying water. Station TARA_125 is located 300 km

850   downstream of the islands. The chlorophyll patch was still clearly evident ([Chl-*a*]$_{int}$ 27.6 mg·m^{-}

851   ^{2}). Of note, the large NO$_2$ reservoir at the base of the mixed layer (120-180 m) indicated significant

852   bacterial activity (Figure S4b).

853   We estimated carbon fluxes potentially due to the natural iron fertilization phenomenon (Figure

854   4b, c and Table S2a). Mean POC fluxes within the top 100 m of the water column varied between

855   37.2 ± 14.6, 78.6 ± 31.2, 47.6 ± 13.3 and 41.4 ± 15.5 mg. m^{-2}. d^{-1} at Stations TARA_122, 123, 124

856   and 125, respectively. POC fluxes at the coastal station (TARA_123) were approximately two

857   times higher than the fluxes at the other stations. POC exported (mean POC flux in 100 to 150 m

858   layer) varied between 24.2 ± 9.4, 81.9 ± 46.2, 11.2 ± 3.6 and 14.2 ± 4.8 mg. m^{-2}. d^{-1} for Stations

859   TARA_122, 123, 124 and 125, respectively, therefore showing a five-fold increase at Station

860   TARA_123 compared to the others. Finally, in the deep layer (mean POC flux between 380 to 420

29

861　m), POC flux varied from 7.0 ± 5.7, 21.3 ± 1.9, 6.4 ± 4.0 and 6.9 ± 2.2 mg. m^{-2}. d^{-1} at Stations

862　TARA_122, 123, 124 and 125, respectively. These data thus highlighted a much stronger flux at

863　Station TARA_123 compared to the other stations.

864　The concentration of measured biologically relevant metals was generally reduced in Stations

865　TARA_123 to TARA_125 with respect to HNLC station TARA_122 (Table S2b). The reduction

866　of dissolved ions was particularly significant in the case of cobalt, nickel, copper and cadmium.

867　These data indicate that the planktonic uptake of biologically available ions was strongly enhanced

868　in the leeward stations. Since these metals were not limiting in the HNLC conditions, it is possible

869　that the removal of iron limitation affected the biological pathways related to metal ion uptake in

870　general. For more information on the oceanographic context of the Marquesas Island at the time

871　of sampling, see Supplementary Material S2.

872　At the four Marquesas sampling sites the IAAs displayed dynamic patterns (Figure 4d, e; Table

873　S1e; Supplementary Material S3). The low-iron adapted DarkRed IAA showed a progressive

874　decrease in its prominence leeward of the islands, consistent with its negative correlations to iron

875　at global level, while the Turquoise IAA showed increases in abundance. The Turquoise IAA is the

876　only module containing autotrophs both positively and negatively correlated with iron, and while

877　the latter were prominent at Station TARA_122 the former were prevalent at stations TARA_123-

878　125 (Figure 4e). The observed changes in IAA prevalence in the Marquesas stations therefore

879　supports a role for iron in the modulation of plankton communities in the region. Prokaryote IAAs,

880　although not taxonomy-based, are dynamically responsive at the Marquesas Islands (Figure 2b,

881　lower panel). Two types of response can be detected: a) the prokaryote IAAs Grey60 and Plum1

882　show a shift from negative to positive eigenlineage scores from TARA_122 to TARA_123, and b)

883　the SaddleBrown, Red and SkyBlue IAAs show eigenvalue peaks in Station TARA_123. Viral

884　IAAs are not responsive at the Marquesas Islands stations, with the exception of the Blue module

885　(Figure S2a). This IAA anticorrelates significantly with the ECCO2-DARWIN iron estimates (r=-

886　0.43, p=4e-04) but not with the PISCES2 estimates.

887　Further analysis of the plankton communities at the Marquesas stations showed that the biomass

888　of primary producers was around 50% higher at the leeward stations (Stations TARA_123, 124

889　and 125) than at HNLC Station TARA_122, with increases in diatoms, haptophytes, pelagophytes,

890　and *Synechococcus* (Figure 5a-d; Supplementary Material S3). The higher productivity likely

891　fueled increases in zooplankton standing stock at these three stations, in particular copepods,

30

892 chaetognaths, and appendicularians (Figure 5b, e), although carbon export to depth was only

893 increased substantially at Station TARA_123 (Figure 4b, c), stressing the impact of different

894 community compositions on the vertical carbon export (see later).

895 Eukaryotic phytoplankton diversity increased at TARA_123 and TARA_124 (Figure 5f,

896 Supplementary Material S3), likely favored by the intense physical dynamics (Barton et al., 2010;

897 Biard et al., 2016). At these stations the increased number of diatoms was due principally to

898 *Thalassiosira* and *Minutocellus* (Supplementary Material S3). Increases in haptophyte and

899 pelagophyte abundance were due to *Phaeocystis* and *Pelagomonas*, respectively. By contrast, the

900 community at Station TARA_122 was more characteristic of an extremely oligotrophic

901 environment, with an abundance of Rhizaria (Biard et al., 2016), *Planktoniella* diatoms (Malviya

902 et al., 2016), *Chrysochromulina* haptophytes (Stibor et al., 2003), and *Pelagococcus* pelagophytes

903 (Guillou et al., 1999), as well as *Prochlorococcus* (Rusch et al., 2010).

904 Analysis of picocyanobacteria also revealed alterations consistent with increased iron

905 bioavailability in the wake of the islands with respect to TARA_122 (Figs. 3d and 5a-c). For

906 example, we observed an almost complete shift of *Synechococcus* community composition from

907 clade CRD1 at TARA_122 to clade II at TARA_123 and TARA_124, while absolute abundances

908 of *Prochlorococcus* HLIII and IV, previously shown to dominate in iron-depleted waters (Rusch

909 et al., 2010; West at al., 2011), were significantly reduced (Figure 3a, d; Supplementary Material

910 S3 and S4).

911 Using transcriptomes from MMETSP together with metatranscriptomes from *Tara* Oceans

912 (Sunagawa et al., 2015; Louca et al., 2016; Alberti et al., 2017; Carradec et al., 2018) we could

913 further compare the qualitative shifts in genotypes highlighted above with changes in

914 transcriptional outputs in cyanobacteria (Figure 6a, b), eukaryotic phytoplankton (Figure 6c, d,

915 Figure S5a-e, S6) and metazoans (Figure S5f and Supplementary Material S4). Importantly, *ISIP*

916 levels were decreased in the leeward stations (Figure 6c, d), and study of gene switches proposed

917 to be responsive to ambient iron concentrations such as ferredoxin/flavodoxin,

918 plastocyanin/cytochrome c_6, and FBAI/FBAII (Peers and Price, 2006; Thompson et al., 2011;

919 Allen et al., 2012; Marchetti et al., 2012; Mackey et al., 2015; Pierella Karlusich et al., 2015)

920 revealed patterns generally consistent with increased bioavailability at Stations TARA_123-125

921 with respect to HNLC Station TARA_122 both in *Synechococcus* (Figure 6b) and in the major

922 groups of eukaryotic phytoplankton (Figs. 6c, d and S5a-e). The expression patterns of

31

923   proteorhodopsin and ferritin genes displayed the same trends. These expression patterns of known

924   iron-responsive genes provide strong support that iron bioavailability is an important driver of the

925   phytoplankton blooms in the Marquesas Islands (Supplementary Material S4).

926   Furthermore, and consistently with the global analyses, *Thalassiosira* and *Pseudo-nitzschia* appear

927   to employ different mechanisms to respond to iron in the Marquesas stations. Specifically, small

928   ferritin-containing *Thalassiosira* cells expressing cytochrome c_6 genes increase in abundance at

929   Station TARA_123, replacing larger Thalassiosirales genetically adapted to low iron at Station

930   TARA_122 by their almost exclusive expression of plastocyanin with respect to cytochrome c_6

931   (Figure 6c, d, Figure S6; Supplementary Material S4). On the other hand, *Pseudo-nitzschia* cells

932   with flavodoxin and plastocyanin genes are enriched in TARA_122 in comparison with

933   TARA_123. For these two diatom genera, the investigation of the local response around the

934   Marquesas Islands therefore corroborates their behavior within IAAs at the global level, and their

935   compartmentalization into different groups based on *ISIP* gene abundance and expression supports

936   the hypothesis that they have evolved fundamentally different mechanisms to respond to iron

937   resource availability.

938   The outcome of the taxon-specific responses summarized above and discussed more

939   comprehensively in the Supplementary Material S4 are shifts in abundance and occurrence of taxa

940   within IAAs that change the overall structure of the food web and correlate with alterations in

941   carbon export. Our observations also reveal novel information about the genetic strategies and

942   specialized mechanisms employed by each taxon to cope with iron availability (Supplementary

943   Material S4) and illustrate that these responses may ensure resilience of each IAA in a subset of

944   conditions within a highly variable environment. Collectively, our results therefore demonstrate

945   that the delineation of co-responsive sub-communities at global scale can provide a valuable

946   framework for identifying key lineages whose adaptive capacities can be compared and contrasted

947   in specific dynamic contexts. Our in-depth analysis of community structure and gene expression

948   around the Marquesas Islands further illustrates how biological data can be used to inform

949   biogeochemical models, because neither of the models used here was able to predict increased iron

950   availability in the wake of the islands.

951

4. Discussion

953 In this study we have shown how the turnover of organisms coping with ocean variability involves

954 a combination of ontogenetic responses driven essentially by modulation of gene expression

955 patterns, i.e., acclimation, together with phylogenetic responses driven by changes in plankton

956 community structure as well as different genotypes adapted to local conditions by altered copy

957 numbers of iron responsive genes. Different organismal groups appear to use different strategies,

958 meaning that they will not all respond over the same evolutionary timescales. The island mass

959 effect in the wake of the Marquesas Islands leads to the selection of preferred genotypes at the

960 community level and triggers acclimatory responses to fine-tune metabolic functioning via

961 transcriptional responses. These local observations of the most affected organisms are consistent

962 with IAAs identified in the global ocean, suggesting that large scale equilibria are in fact dynamic

963 and responsive to smaller scale perturbations.

964 Previous studies at global scale of the effects of iron on marine plankton were focused on a specific

965 subset of bacterial genes involved in iron metabolism using metagenomics samples from North

966 West Atlantic, Equatorial Pacific and Indian Oceans (Toulza et al., 2012). Our current study

967 extends this analysis because of its broader geographical coverage and the vastly expanded

968 sequencing dataset, which has permitted us to explore both community-level and gene-level

969 responses throughout the entire plankton community, from viruses to zooplankton. Our work thus

970 provides an extensive global scale analysis of the different levels at which plankton biodiversity

971 may be impacted by iron availability, although it should not be assumed that all the responses we

972 highlighted depend solely on iron because one single resource is very unlikely to drive the

973 physiological and structural dynamics of a community. Nonetheless, our extensive statistical

974 analyses suggest that the responses we define do certainly involve iron bioavailability and that the

975 responses occur at molecular, physiological and compositional levels. Of note is the evidence of

976 modularity in the community structure with modules of co-occurring taxa being sensitive to the

977 resource yet displaying often contrasting strategies. This extends the results obtained by (Guidi et

978 al., 2016) who focused on a specific process, indicating that modularity is a general feature of

979 plankton communities, which might be related to their continuous turnover. To the extent allowed

980 by available gene catalogs and taxonomic resolution we were able to link the sub-community

33

981   responses to the molecular toolkits of the organisms but in many cases we emphasize that the

982   response is not unequivocal but rather maps to a suite of strategies that had already been recorded

983   previously in localized or laboratory experiments.

984   The complexity of the plankton ecosystem that emerges from the analysis of each IAA and their

985   VIPs, whose dynamics have a certain degree of freedom with respect to the response of the others,

986   indicates that there is some flexibility between the composition of primary producers and their

987   consumers, even though the former are the organisms most directly impacted by nutrient

988   availability. In particular, heterotrophic grazers appear to be central for responses to such bottom-

989   up processes as nutrient acquisition. We interpret the VIP values vs. correlation to iron and

990   community centrality as follows: that communities are assemblages of several organisms with

991   multiple interactions among them that cannot be reduced to just a handful of opportunistic

992   autotrophic species able to benefit from nutrient injection and that supply organic carbon to higher

993   trophic levels. Rather, organisms respond to resource availability according to their functional

994   traits but also modulate interactions within their communities, thus affecting their structure. These

995   changes will nonetheless depend on the resident community, immigration from beyond, and

996   changes in the ambient conditions. Some organisms may thrive in different contexts and therefore

997   not be strongly dependent on iron, but rather be good exploiters of primary production stimulated

998   by increased nutrient bioavailability; most of the VIPs are indeed consumers. Furthermore, the

999   relatively low subnetwork centrality of these consumers may suggest that they co-occur with only

1000   a subcomponent of the other species. Finally, the nature of the modules composed of parasitic and

1001   mixotrophic organisms further suggests that recycling of matter, e.g., through remineralization,

1002   parasitism and pathogenesis, are additional strategies within plankton communities to overcome

1003   resource limitation. Such strategies would be expected to confer further elasticity, and lead to an

1004   improved capacity to respond to sporadic bursts of favorable conditions.

1005   Taxonomy-based network analysis for the prokaryotes did not reveal significant associations with

1006   iron bioavailability, whereas their gene subnetworks did. In accordance with a recent study based

1007   largely on *Tara* Oceans data (Louca et al., 2016), this result advocates for the use of prokaryotic

1008   functional signals rather than standard taxonomic criteria to study functional responses of

1009   prokaryotes in the global ocean, at least at the level of present taxonomic resolution. In fact,

1010   cyanobacteria displayed a remarkable strain-dependent sensitivity to iron availability. The

1011   observations further indicate the need for a better assignation of functional taxonomy, and more

1012 studies to better characterize prokaryotic genes of relevance for interpreting the mechanistic

1013 changes in prokaryotes following perturbations in iron bioavailability. Furthermore, while standard

1014 steady-state analyses of ocean systems do not consider biological responses to perturbation per se,

1015 our approach of identifying steady-state global IAA subnetworks and then investigating their

1016 responses to local, short-term perturbation represents a promising new approach.

1017 Comparison of the local response to the inferred iron injection in the Marquesas archipelago with

1018 the global patterns indicates that the community response to iron availability cannot be

1019 characterized by an even increase in biomass among existing components but involves a change

1020 in their relative weights reflecting their different adaptive solutions and the concurrent

1021 reorganization of the sub-communities. In other words, our results infer that the rate of supply of

1022 a resource is a factor that modulates the response of organisms and their communities.

1023 Our analysis is based on iron distribution derived from two advanced biogeochemical models

1024 rather than from discrete measurements. This is because we considered them to be more

1025 representative than the instantaneous in situ measurements whose coverage is also scarce and could

1026 not be improved by our expedition, since Tara was not equipped to accurately perform iron

1027 concentration assessments. While this may be viewed as a limitation of our work, we provide

1028 evidence from independent data of the reliability of these estimates, thus providing a valuable

1029 demonstration of the utility of omics data as a tool to validate (and consequently improve) current

1030 models of earth system dynamics. The good correspondence between the molecular response and

1031 the model simulations demonstrates that metatranscriptomics is now mature enough to provide an

1032 independent, biologically-based validation of ecosystem models especially when the data are

1033 scarce or hard to obtain in a reliable way. The quality and number of iron measurements are

1034 continuously improving but metatranscriptomics may anticipate and suggest the presence of

1035 biogeochemical constraints that are still undetectable with analytical methods. In addition, it could

1036 significantly integrate the formulation of processes in current ecological models because, on the

1037 long term, it can complement the missing information about organism interactions (see above)

1038 which cannot be derived from the availability of resources (e. g., Stec et al., 2017).

1039 In conclusion, our study reinforces the results obtained in smaller-scale studies and significantly

1040 expands the suite of indicators that can be monitored to detect responses to changes in

1041 environmental conditions, from target genes to higher levels of biological organization. Our work

1042   paves the way to a suite of possible developments in experimental design and in model

1043   formulations that prompt for the improvement of statistical tools to better characterize responses

1044   at system level. Numerical simulations of ocean processes aimed at capturing the fluxes of key

1045   elements are currently based on just a handful of plankton functional types (Le Quere et al., 2005)

1046   or functional genes (Coles et al., 2017). Our results highlight the need to incorporate the response

1047   of entire plankton assemblages to more accurately determine responses at different levels, such as

1048   gene expression, gene copy numbers, or community composition. To determine the relevance of

1049   such processes omics should become a routine component of ocean observation, and we further

1050   demonstrate here that it can contribute to assessing the validity of ecosystem models by

1051   complementing biogeochemical measurements in the field and adding critical information about

1052   the actual bioavailability of nutrients, which is currently difficult to measure. Finally, the IAAs

1053   and other modules described herein provide a framework that is independent of taxonomic or

1054   functional groupings to tackle the complexity of natural communities, thus assisting our capacity

1055   to predict the responses and resilience of planktonic ecosystems to natural and human-induced

1056   perturbations.

1057   **Acknowledgements**

1058

1077

Competing interests
1078

1079

1080   The authors declare that no competing interests exist.

1081

Author Contribution
1082

1083   DI, FN, Fd'O and PT designed the study with input from *Tara* Oceans coordinators. DI directed

1084   the project. FN directed the field work in the Marquesas archipelago. CB wrote the paper with

1085   substantial input from MRd'A, DI, LC, and other first authors. Fd'O, PT, FN, EM, DI, HC, LG,

1086   SS, FK performed oceanographic analyses, AT and MJF provided iron concentration data from

1087   biogeochemical models, FRJV, GB and AT compared iron products from different biogeochemical

1088   models, ES, AZ, SM, JV, JL, SC, FV, AtT, CB performed analysis of eukaryotic phytoplankton,

1089   MGM, J-LJ, J-BR, SG, LC, LS, FL, TB performed analysis of metazoans and other zooplankton,

1090   JJPK, SC, HD and LG performed analysis of cyanobacteria, QC, EP, FRJV, JJPK, EV, SGA, AA,

1091   SSu, PB, PW, AV, RS, JP, GL-M, ML performed global omics analyses, FRJV, JJPK, AK, JP-Y,

1092   LT performed analysis of omics data from eukaryotic phytoplankton, AK, EP, LC, PS, SdA

1093   performed analysis of omics data from metazoans and other zooplankton, JJPK, HD and LG

1094   performed gene expression analysis of cyanobacteria, JJPK, JRB, SR, MBS, and MB performed

1095   analysis of viruses, LB, SC A-SB and DE performed WGCNA analyses, SR, FN, CD, MP, SKL,

1096   SSe and SP collected and managed *Tara* Oceans samples, LC, JJPK, MRd'A and CB assembled

1097   the manuscript. *Tara* Oceans coordinators provided a creative environment and constructive

1098   criticism throughout the study. All authors discussed the results and commented on the manuscript.

References
1099

1100   Alberti, A., Poulain, J., Engelen, S., Labadie, K., Romac, S., Ferrera, I., et al. (2017). Viral to

1101   metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Scientific*

1102   *Data*. https://doi.org/10.1038/sdata.2017.93

1103 Alexander, H., Rouco, M., Haley, S., Wilson, S., Karl, D., & Dyhrman, S. (2015). Functional

1104   group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proceedings*

1105   *of the National Academy of Sciences*, E5972–E5979.

1106   https://doi.org/10.1073/pnas.1518165112

1107 Allen, A. E., Moustafa, A., Montsant, A., Eckert, A., Kroth, P. G., & Bowler, C. (2012).

1108   Evolution and functional diversification of fructose bisphosphate aldolase genes in

1109   photosynthetic marine diatoms. *Molecular Biology and Evolution*, *29*(1), 367–379.

1110   https://doi.org/10.1093/molbev/msr223

1111 Allen, A. E., Laroche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., et al. (2008).

1112   Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation.

1113   *Proceedings of the National Academy of Sciences of the United States of America*, *105*(30),

1114   10438–10443. https://doi.org/10.1073/pnas.0711370105

1115 Arienzo, M., Toscano, F., Di Fraia, M., Caputi, L., Sordino, P., Guida, M., et al. (2014). An

1116   assessment of contamination of the Fusaro Lagoon (Campania Province, southern Italy) by

1117   trace metals. *Environmental Monitoring and Assessment*, *186*(9), 5731–5747.

1118   https://doi.org/10.1007/s10661-014-3816-4

1119 Aumont, O., Tagliabue, A., Bopp, L., & Gehlen, M. (2015). PISCES-v2 : an ocean

1120   biogeochemical model for carbon and. *Geoscientific Model Development*, *8*(1990), 2465–

1121   2513. https://doi.org/10.5194/gmd-8-2465-2015

1122 Barton, A. D., Dutkiewicz, S., Flierl, G., Bragg, J., & Follows, M. J. (2010). Patterns of diversity

1123   in marine phytoplankton. *Science (New York, N.Y.)*.

1124   https://doi.org/10.1126/science.1184961

1125 Bartual, S. G., Otero, J. M., Garcia-Doval, C., Llamas-Saiz, A. L., Kahn, R., Fox, G. C., & van

1126   Raaij, M. J. (2010). Structure of the bacteriophage T4 long tail fiber receptor-binding tip.

1127   *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1011218107

1128 Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F., & Covarrubias, A. A. (2008).
1129   The enigmatic LEA proteins and other hydrophilins. *Plant Physiology*, *148*(1), 6–24.
1130   https://doi.org/10.1104/pp.108.120725

1131 Berline, L., Stemmann, L., Vichi, M., Lombard, F., & Gorsky, G. (2011). Impact of
1132   appendicularians on detritus and export fluxes: A model approach at DyFAMed site.
1133   *Journal of Plankton Research*, *33*(6), 855–872. https://doi.org/10.1093/plankt/fbq163

1134 Bernardes, J. S., Vieira, F. R. J., Zaverucha, G., & Carbone, A. (2015). A multi-objective
1135   optimization approach accurately resolves protein domain architectures. *Bioinformatics*,
1136   *32*(3), 345–353. https://doi.org/10.1093/bioinformatics/btv582

1137 Bernardes, J., Zaverucha, G., Vaquero, C., & Carbone, A. (2016). Improvement in Protein
1138   Domain Identification Is Reached by Breaking Consensus, with the Agreement of Many
1139   Profiles and Domain Co-occurrence. *PLoS Computational Biology*, *12*(7), 1–39.
1140   https://doi.org/10.1371/journal.pcbi.1005038

1141 Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., et al. (2016). In
1142   situ imaging reveals the biomass of giant protists in the global ocean. *Nature*, *532*(7600),
1143   504–507. https://doi.org/10.1038/nature17652

1144 Blain, S., Bonnet, S., & Guieu, C. (2008). Dissolved iron distribution in the tropical and sub
1145   tropical South Eastern Pacific. *Biogeosciences*, *5*(1), 269–280. https://doi.org/10.5194/bgd-
1146   4-2845-2007

1147 Bonnain, C., Breitbart, M., & Buck, K. N. (2016). The Ferrojan Horse Hypothesis: iron-virus
1148   interactions in the ocean. *Frontiers in Marine Science*, *3*(June), 82.
1149   https://doi.org/10.3389/fmars.2016.00082

1150 Bork, P., Bowler, C., Vargas, C. de, Gorsky, G., Karsenti, E., & Wincker, P. (2015). Tara Oceans
1151   studies plankton at planetary scale. *Science*, *348*(6237), 873.
1152   https://doi.org/10.1126/science.aac5605

1153   Botebol, H., Lesuisse, E., Šuták, R., Six, C., Lozano, J.-C., Schatt, P., et al. (2015). Central role

1154   for ferritin in the day/night regulation of iron homeostasis in marine phytoplankton.

1155   *Proceedings of the National Academy of Sciences*, *112*(47), 14652–14657.

1156   https://doi.org/10.1073/pnas.1506074112

1157   Boyd, P. W., Jickells, T., Law, C. S., Blain, S., Boyle, E. a, Buesseler, K. O., et al. (2007).

1158   Mesoscale iron enrichment experiments 1993-2005: synthesis and future directions. *Science*

1159   *(New York, N.Y.)*, *315*(5812), 612–617. https://doi.org/10.1126/science.1131669

1160   Browning, C., Shneider, M. M., Bowman, V. D., Schwarzer, D., & Leiman, P. G. (2012). Phage

1161   pierces the host cell membrane with the iron-loaded spike. *Structure*.

1162   https://doi.org/10.1016/j.str.2011.12.009

1163   Brum, J. R., Ignacio-espinoza, J. C., Roux, S., Doulcier, G., Acinas, S. G., Alberti, A., &

1164   Chaffron, S. (2015). Patterns and ecological drivers of ocean viral communities. *Science*,

1165   *348*(6237), 1261498-1–11. https://doi.org/10.1126/science.1261498

1166   Bundy, J. G., & Kille, P. (2014). Metabolites and metals in Metazoa--what role do

1167   phytochelatins play in animals? *Metallomics : Integrated Biometal Science*, *6*(9), 1576–82.

1168   https://doi.org/10.1039/c4mt00078a

1169   Campbell, L., Nolla, H. A., & Vaulot, D. (1994). The importance of Prochlorococcus to

1170   community structure in the central North Pacific Ocean. *Limnol. Oceanogr.*, *39*(4), 954–

1171   961.

1172   Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., et al.

1173   (2018). A global ocean atlas of eukaryotic genes. *Nature Communications*.

1174   https://doi.org/10.1038/s41467-017-02342-1

1175   Chappell, P. D., Whitney, L. P., Wallace, J. R., Darer, A. I., Jean-Charles, S., & Jenkins, B. D.

1176   (2015). Genetic indicators of iron limitation in wild populations of Thalassiosira oceanica

1177   from the northeast Pacific Ocean. *Isme J*, *9*(3), 592–602.

1178   https://doi.org/10.1038/ismej.2014.171

1179   Chen, X., Wakeham, S. G., & Fisher, N. S. (2011). Influence of iron on fatty acid and sterol

1180   composition of marine phytoplankton and copepod consumers. *Limnology and*

1181   *Oceanography*, *56*(2), 716–724. https://doi.org/10.4319/lo.2011.56.2.0716

1182   Chow, C.-E. T., & Suttle, C. A. (2015). Biogeography of Viruses in the Sea. *Annual Review of*

1183   *Virology*, *2*(1), 41–66. https://doi.org/10.1146/annurev-virology-031413-085540

1184   Claustre, H., Sciandra, A., & Vaulot, D. (2008). Introduction to the special section bio-optical

1185   and biogeochemical conditions in the South East Pacific in late 2004: The BIOSOPE

1186   program. *Biogeosciences*. https://doi.org/10.5194/bg-5-679-2008

1187   Coale, K. H., Fitzwater, S. E., Gordon, R. M., Johnson, K. S., & Barber, R. T. (1996). Control of

1188   community growth and export production by upwelled iron in the equatorial Pacific Ocean

1189   Kenneth. *Nature*. https://doi.org/10.1038/379621a0

1190   Cohen, N. R., Brzezinski, M. A., Twining, K. T. B. S., Ellis, K. A., Lampe, R. H., McNair, H., et

1191   al. (2017). Diatom transcriptional and physiological responses to changes in iron

1192   bioavailability across ocean provinces. *Environmental Microbiology*, *4*(November), 1–20.

1193   https://doi.org/10.3389/fmars.2017.00360

1194   Coles, V. J., Stukel, M. R., Brooks, M. T., Burd, A., Crump, B. C., Moran, M. A., et al. (2017).

1195   Ocean biogeochemistry modeled with emergent trait-based genomics. *Science*,

1196   *1154*(December), 1–26. https://doi.org/10.1126/science.aan5712

1197   Crans, D. C., Smee, J. J., Gaidamauskas, E., & Yang, L. (2004). The Chemistry and

1198   Biochemistry of Vanadium and the Biological Activities Exerted by Vanadium Compounds.

1199   *Chemical Reviews*, *104*(2), 849–902. https://doi.org/10.1021/cr020607t

1200   Dandonneau, Y., & Charpy, L. (1985). An empirical approach to the island mass effect in the

1201   south tropical Pacific based on sea surface chlorophyll concentrations. *Deep Sea Research*

1202   *Part A, Oceanographic Research Papers*, *32*(6). https://doi.org/10.1016/0198-

1203   0149(85)90074-3

41

1204   de Vargas, C., Audic, S., Henry, N., Decelle, J., & Mahé, F. (2015). Eukaryotic plankton
1205   diversity in the sunlit ocean. *Science*, *348*(MAY), 1–12. https://doi.org/10.1007/s13398-
1206   014-0173-7.2

1207   De Vos, C. H., Vonk, M. J., Vooijs, R., & Schat, H. (1992). Glutathione Depletion Due to
1208   Copper-Induced Phytochelatin Synthesis Causes Oxidative Stress in Silene cucubalus. *Plant*
1209   *Physiology*, *98*(3), 853–858. https://doi.org/10.1104/pp.98.3.853

1210   Dolan, J. R., Ritchie, M. E., & Ras, J. (2007). The "neutral" community structure of planktonic
1211   herbivores, tintinnid ciliates of the microzooplankton, across the SE Tropical Pacific Ocean.
1212   *Biogeosciences Discussions*, *4*, 561–593. https://doi.org/10.5194/bgd-4-561-2007

1213   Dupont, C. L., Mccrow, J. P., Valas, R., Moustafa, A., Walworth, N., Goodenough, U., et al.
1214   (2015). Genomes and gene expression across light and productivity gradients in eastern
1215   subtropical Pacific microbial communities. *ISME Journal*, *9*(5), 1076–1092.
1216   https://doi.org/10.1038/ismej.2014.198

1217   Durkin, C. A., Koester, J. A., Bender, S. J., Armbrust, E. V., & Kroth, P. (2016). The evolution
1218   of silicon transporters in diatoms. *Journal of Phycology*, *52*(5), 716–731.
1219   https://doi.org/10.1111/jpy.12441

1220   Farrant, G. K., Doré, H., Cornejo-Castillo, F. M., Partensky, F., Ratin, M., Ostrowski, M., et al.
1221   (2016). Delineating ecologically significant taxonomic units from global patterns of marine
1222   picocyanobacteria. *Proceedings of the National Academy of Sciences*, *113*(24), E3365–
1223   E3374. https://doi.org/10.1073/pnas.1524865113

1224   Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016).
1225   The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids*
1226   *Research*, *44*(D1), D279–D285. https://doi.org/10.1093/nar/gkv1344

1227   Follows, M. J., Dutkiewicz, S., Grant, S., & Chisholm, S. W. (2007). Emergent Biogeography of
1228   Microbial. *Science*, (C), 1843–1847. https://doi.org/10.1126/science.1138544

1229   Freeland, H. J., & Cummins, P. F. (2005). Argo: A new tool for environmental monitoring and

1230   assessment of the world's oceans, an example from the N.E. Pacific. *Progress in*

1231   *Oceanography*, *64*(1), 31–44. https://doi.org/10.1016/j.pocean.2004.11.002

1232   Gómez, F., Claustre, H., Raimbault, P., & Souissi, S. (2007). Two High-Nutrient Low-

1233   Chlorophyll phytoplankton assemblages: the tropical central Pacific and the offshore

1234   Per&uacute;-Chile Current. *Biogeosciences Discussions*, *4*(3), 1535–1554.

1235   https://doi.org/10.5194/bgd-4-1535-2007

1236   Gong, W., Browne, J., Hall, N., Schruth, D., Paerl, H., & Marchetti, A. (2016). Molecular

1237   insights into a dinoflagellate bloom. *The ISME Journal*, 1–14.

1238   https://doi.org/10.1038/ismej.2016.129

1239   Gorsky, G., Chrétiennot-Dinet, M. J., Blanchot, J., & Palazzoli, I. (1999). Picoplankton and

1240   nanoplankton aggregation by appendicularians: Fecal pellet contents of Megalocercus

1241   huxleyi in the equatorial Pacific. *Journal of Geophysical Research*, *104*(C2), 3381–3390.

1242   https://doi.org/10.1029/98jc01850

1243   Gove, J. M., McManus, M. A., Neuheimer, A. B., Polovina, J. J., Drazen, J. C., Smith, C. R., et

1244   al. (2015). Ocean Oases: near-island biological hotspots in barren ocean basins. *In Review*,

1245   *7*, 1–34. https://doi.org/10.1038/ncomms10581

1246   Graff van Creveld, S., Rosenwasser, S., Levin, Y., & Vardi, A. (2016). Chronic iron limitation

1247   confers transient resistance to oxidative stress in marine diatoms. *Plant Physiology*.

1248   https://doi.org/10.1104/pp.16.00840

1249   Grob, C., Ulloa, O., Claustre, H., Huot, Y., Alarcón, G., & Marie, D. (2007). Contribution of

1250   picoplankton to the total particulate organic carbon concentration in the eastern South

1251   Pacific. *Biogeosciences*, *4*(5), 837–852. https://doi.org/10.5194/bg-4-837-2007

1252   Groussman, R. D., Parker, M. S., & Armbrust, E. V. (2015). Diversity and evolutionary history

1253   of iron metabolism genes in diatoms. *PLoS ONE*, *10*(6), 1–25.

1254   https://doi.org/10.1371/journal.pone.0129081

1255 Guidi, L., Gorsky, G., Claustre, H., Miquel, J. C., Picheral, M., & Stemmann, L. (2008).

1256 Distribution and fluxes of aggregates >100 μm in the upper kilometer of the South-

1257 Eastern Pacific. *Biogeosciences*, *5*(5), 1361–1372. https://doi.org/10.5194/bg-5-1361-2008

1258 Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., et al. (2016). Plankton

1259 networks driving carbon export in the oligotrophic ocean. *Nature*, *532*, 465–470.

1260 https://doi.org/10.1038/nature16942

1261 Guillou, L., Moon-Van Der Staay, S. Y., Claustre, H., Partensky, F., & Vaulot, D. (1999).

1262 Diversity and abundance of Bolidophiceae (Heterokonta) in two oceanic regions. *Appl.*

1263 *Environ. Microbiol.*, *65*(10), 4528–4536.

1264 Karsenti, E., Acinas, S. G., Bork, P., Bowler, C., de Vargas, C., Raes, J., et al. (2011). A holistic

1265 approach to marine Eco-systems biology. *PLoS Biology*, *9*(10), 7–11.

1266 https://doi.org/10.1371/journal.pbio.1001177

1267 Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid

1268 multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*,

1269 *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

1270 Kazamia, E., Sutak, R., Paz-Yepes, J., Dorrell, R. G., Vieira, F. R. J., Mach, J., et al. (2018).

1271 Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms.

1272 *Science Advances*, *4*(5), eaar4536. https://doi.org/10.1126/sciadv.aar4536

1273 Khoshnood, B., Dacklin, I., & Grabbe, C. (2016). Urm1: An essential regulator of JNK signaling

1274 and oxidative stress in Drosophila melanogaster. *Cellular and Molecular Life Sciences*,

1275 *73*(9), 1939–1954. https://doi.org/10.1007/s00018-015-2121-x

1276 Lane, T. W., & Morel, F. M. M. (2000). A biological function for cadmium in marine diatoms.

1277 *Proceedings of the National Academy of Sciences*, *97*(9), 4627–4631.

1278 https://doi.org/10.1073/pnas.090091397

1279 Langfelder, P., & Horvath, S. (2007). Eigengene networks for studying the relationships between
1280   co-expression modules. *BMC Systems Biology*, *1*(1), 54. https://doi.org/10.1186/1752-0509-
1281   1-54

1282 Le Quere, C., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., et al.
1283   (2005). Ecosystem dynamics based on plankton functional types for global ocean
1284   biogeochemistry models\rdoi:10.1111/j.1365-2486.2005.1004.x. *Global Change Biology*,
1285   *11*(11), 2016–2040. https://doi.org/10.1111/j.1365-2486.2005.01004.x

1286 Legeckis, R., Brown, C. W., Bonjean, F., & Johnson, E. S. (2004). The influence of tropical
1287   instability waves on phytoplankton blooms in the wake of the Marquesas Islands during
1288   1998 and on the currents observed during the drift of the Kon-Tiki in 1947. *Geophysical*
1289   *Research Letters*, *31*(23), 1–4. https://doi.org/10.1029/2004GL021637

1290 Lommer, M., Roy, A.-S., Schilhabel, M., Schreiber, S., Rosenstiel, P., & LaRoche, J. (2010).
1291   Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an
1292   oceanic diatom adapted to chronic iron limitation. *BMC Genomics*, *11*(1), 718.
1293   https://doi.org/10.1186/1471-2164-11-718

1294 Louca, S., Parfrey, L. W., & Doebeli, M. (2016). Decoupling function and taxonomy in the
1295   global ocean microbiome. *Science*, *353*(6305), 1272–1277.

1296 Mackey, K. R. M., Post, A. F., McIlvin, M. R., Cutter, G. A., John, S. G., & Saito, M. A. (2015).
1297   Divergent responses of Atlantic coastal and oceanic Synechococcus to iron limitation.
1298   *Proceedings of the National Academy of Sciences of the United States of America*, *112*(32),
1299   9944–9. https://doi.org/10.1073/pnas.1509448112

1300 Mahowald, N. M., Baker, A. R., Bergametti, G., Brooks, N., Duce, R. A., Jickells, T. D., et al.
1301   (2005). Atmospheric global dust cycle and iron inputs to the ocean. *Global Biogeochemical*
1302   *Cycles*, *19*(4). https://doi.org/10.1029/2004GB002402

1303 Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., et al. (2016). Insights
1304   into global diatom distribution and diversity in the world's ocean. *Proceedings of the*

1305 *National Academy of Sciences*, *113*(11), E1516–E1525.

1306 https://doi.org/10.1073/pnas.1509523113

1307 Marchetti, a., Schruth, D. M., Durkin, C. a., Parker, M. S., Kodner, R. B., Berthiaume, C. T., et

1308 al. (2012). Comparative metatranscriptomics identifies molecular bases for the

1309 physiological responses of phytoplankton to varying iron availability. *Proceedings of the*

1310 *National Academy of Sciences*, *109*(6), E317–E325.

1311 https://doi.org/10.1073/pnas.1118408109

1312 Marchetti, A., Catlett, D., Hopkinson, B. M., Ellis, K., & Cassar, N. (2015). Marine diatom

1313 proteorhodopsins and their potential role in coping with low iron availability. *ISME*

1314 *Journal*, *9*(12), 2745–2748. https://doi.org/10.1038/ismej.2015.74

1315 Marchetti, A., Juneau, P., Whitney, F. A., Wong, C. S., & Harrison, P. J. (2006). Phytoplankton

1316 processes during a mesoscale iron enrichment in the NE subarctic Pacific: Part II-Nutrient

1317 utilization. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *53*(20–22),

1318 2114–2130. https://doi.org/10.1016/j.dsr2.2006.05.031

1319 Marchetti, A., Moreno, C. M., Cohen, N. R., Oleinikov, I., Twining, B. S., Armbrust, E. V., &

1320 Lampe, R. H. (2017). Development of a molecular-based index for assessing iron status in

1321 bloom-forming pennate diatoms. *Journal of Phycology*, 1–13.

1322 https://doi.org/10.1111/jpy.12539

1323 Marchetti, A., Parker, M. S., Moccia, L. P., Lin, E. O., Arrieta, A. L., Ribalet, F., et al. (2009).

1324 Ferritin is used for iron storage in bloom-forming marine pennate diatoms. *Nature*,

1325 *457*(7228), 467–470. https://doi.org/10.1038/nature07539

1326 Martin, P., Van Der Loeff, M. R., Cassar, N., Vandromme, P., D'Ovidio, F., Stemmann, L., et al.

1327 (2013). Iron fertilization enhanced net community production but not downward particle

1328 flux during the Southern Ocean iron fertilization experiment LOHAFEX. *Global*

1329 *Biogeochemical Cycles*, *27*(3), 871–881. https://doi.org/10.1002/gbc.20077

1330   Martinez, E., & Maamaatuaiahutapu, K. (2004). Island mass effect in the Marquesas Islands:

1331   Time variation. *Geophysical Research Letters*, *31*(18), 1–4.

1332   https://doi.org/10.1029/2004GL020682

1333   Masquelier, S., & Vaulot, D. (2007). Distribution of micro-organisms along a transect in the

1334   South-East Pacific Ocean (BIOSOPE cruise) from epifluorescence microscopy.

1335   *Biogeosciences Discussions*, *4*(4), 2667–2697. https://doi.org/10.5194/bgd-4-2667-2007

1336   Mawji, E., Schlitzer, R., Dodas, E. M., Abadie, C., Abouchami, W., Anderson, R. F., et al.

1337   (2014). The GEOTRACES Intermediate Data Product 2014. *Marine Chemistry*, *177*, 1–8.

1338   https://doi.org/10.1016/j.marchem.2015.04.005

1339   McQuaid, J. B., Kustka, A. B., Oborník, M., Horák, A., McCrow, J. P., Karas, B. J., et al. (2018).

1340   Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. *Nature*.

1341   https://doi.org/10.1038/nature25982

1342   Menemenlis, B. D., Campin, J., Heimbach, P., Hill, C., & Lee, T. (2008). ECCO2 : High

1343   Resolution Global Ocean and Sea Ice Data Synthesis. *Mercator Ocean Quarterly*

1344   *Newsletter*, *31*(October), 13–21.

1345   Mevik, B.-H., & Wehrens, R. (2007). The pls Package: Principle Component and Partial Least

1346   Squares Regression in R. *Journal of Statistical Software*, *18*(2), 1–24.

1347   https://doi.org/10.1002/wics.10

1348   Mock, T., Samanta, M. P., Iverson, V., Berthiaume, C., Robison, M., Holtermann, K., et al.

1349   (2008). Whole-genome expression profiling of the marine diatom Thalassiosira pseudonana

1350   identifies genes involved in silicon bioprocesses. *Proceedings of the National Academy of*

1351   *Sciences*, *105*(5), 1579–1584. https://doi.org/10.1073/pnas.0707946105

1352   Morrissey, J., Sutak, R., Paz-Yepes, J., Tanaka, A., Moustafa, A., Veluchamy, A., et al. (2015).

1353   A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface

1354   and facilitates uptake. *Current Biology*, *25*(3), 364–371.

1355   https://doi.org/10.1016/j.cub.2014.12.004

Peers, G., & Price, N. M. N. (2006). Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature*, *441*(7091), 341–4. https://doi.org/10.1038/nature04630

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., et al. (2015). Open science resources for the discovery and analysis of Tara Oceans data. *Scientific Data*, *2*(Lmd), 150023. https://doi.org/10.1038/sdata.2015.23

Pfaffen, S., Bradley, J. M., Abdulqadir, R., Firme, M. R., Moore, G. R., Brun, N. E. L., & Murphy, M. E. P. (2015). A diatom ferritin optimized for iron oxidation but not iron storage. *Journal of Biological Chemistry*, *290*(47), 28416–28427. https://doi.org/10.1074/jbc.M115.669713

Picheral, M., Guidi, L., Stemmann, L., Karl, D. M., Iddaoud, G., & Gorsky, G. (2010). The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnology and Oceanography: Methods*, *8*(9), 462–473. https://doi.org/10.4319/lom.2010.8.462

Pierella Karlusich, J. J., Ceccoli, R. D., Graña, M., Romero, H., & Carrillo, N. (2015). Environmental selection pressures related to iron utilization are involved in the loss of the flavodoxin gene from the plant genome. *Genome Biology and Evolution*, *7*(3), 750–767. https://doi.org/10.1093/gbe/evv031

Probert, I., Siano, R., Poirier, C., Decelle, J., Biard, T., Tuji, A., et al. (2014). Brandtodinium gen. nov. and B.nutricula comb. Nov. (Dinophyceae), a dinoflagellate commonly found in symbiosis with polycystine radiolarians. *Journal of Phycology*, *50*(2), 388–399. https://doi.org/10.1111/jpy.12174

Quéguiner, B. (2013). Iron fertilization and the structure of planktonic communities in high nutrient regions of the Southern Ocean. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *90*, 43–54. https://doi.org/10.1016/j.dsr2.2012.07.024

1380   Ras, J., Claustre, H., & Uitz, J. (2008). Spatial variability of phytoplankton pigment distributions

1381   in the Subtropical South Pacific Ocean: comparison between in situ and predicted data.

1382   *Biogeosciences*, *5*, 353–369. https://doi.org/10.5194/bgd-4-3409-2007

1383   Roux, S., Brum, J. R., Dutilh, B. E., Sunagawa, S., Duhaime, M. B., Loy, A., et al. (2016).

1384   Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses.

1385   *Nature*, *537*(7622), 689–693. https://doi.org/10.1038/nature19366

1386   Rusch, D. B., Martiny, A. C., Dupont, C. L., Halpern, A. L., & Venter, J. C. (2010).

1387   Characterization of Prochlorococcus clades from iron-depleted oceanic regions.

1388   *Proceedings of the National Academy of Sciences of the United States of America*, *107*(37),

1389   16184–16189. https://doi.org/10.1073/pnas.1009513107

1390   Signorini, S. R., McClain, C. R., & Dandonneau, Y. (1999). Mixing and phytoplankton bloom in

1391   the wake of the Marquesas Islands. *Geophysical Research Letters*, *26*(20), 3121–3124.

1392   https://doi.org/10.1029/1999GL010470

1393   Smetacek, V., & Naqvi, S. W. a. (2008). The next generation of iron fertilization experiments in

1394   the Southern Ocean. *Philosophical Transactions of the Royal Society A*, *366*(August),

1395   3947–3967. https://doi.org/10.1098/rsta.2008.0144

1396   Sohm, J. A., Ahlgren, N. A., Thomson, Z. J., Williams, C., Moffett, J. W., Saito, M. A., et al.

1397   (2016). Co-occurring Synechococcus ecotypes occupy four major oceanic regimes defined

1398   by temperature, macronutrients and iron. *ISME Journal*, *10*(2), 333–345.

1399   https://doi.org/10.1038/ismej.2015.115

1400   Stec, K. F., Caputi, L., Buttigieg, P. L., D'Alelio, D., Ibarbalz, F. M., Sullivan, M. B., et al.

1401   (2017). Modelling plankton ecosystems in the meta-omics era. Are we ready? *Marine*

1402   *Genomics*, *32*, 1–17. https://doi.org/10.1016/j.margen.2017.02.006

1403   Stemmann, L., Eloire, D., Sciandra, A., Jackson, G., Guidi, L., Picheral, M., & Gorsky, G.

1404   (2008). Volume distribution for particles between 3.5 to 2000 μm in the upper 200 m region

1405   of the South Pacific Gyre. *Biogeosciences Discussions*, *5*, 299–310.

1406   https://doi.org/10.5194/bgd-4-3377-2007

1407   Stibor, H., & Sommer, U. (2003). Mixotrophy of a photosynthetic flagellate viewed from an

1408   optimal foraging perspective. *Protist*, *154*(1), 91–98.

1409   https://doi.org/10.1078/143446103764928512

1410   Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015).

1411   Structure and function of the global ocean microbiome. *Science*, *348*(6237), 1261359.

1412   https://doi.org/10.1126/science.1261359

1413   Tagliabue, A., Aumont, O., DeAth, R., Dunne, J. P., Dutkiewicz, S., Galbraith, E., et al. (2016).

1414   How well do global ocean biogeochemistry models simulate dissolved iron distributions?

1415   *Global Biogeochemical Cycles*, *30*, 149–174.

1416   https://doi.org/10.1002/2015GB005289.Received

1417   Tagliabue, A., Mtshali, T., Aumont, O., Bowie, A. R., Klunder, M. B., Roychoudhury, A. N., &

1418   Swart, S. (2012). A global compilation of dissolved iron measurements: Focus on

1419   distributions and processes in the Southern Ocean. *Biogeosciences*, *9*(6), 2333–2349.

1420   https://doi.org/10.5194/bg-9-2333-2012

1421   Tagliabue, A., Bowie, A. R., Philip, W., Buck, K. N., Johnson, K. S., & Saito, M. A. (2017). The

1422   integral role of iron in ocean biogeochemistry. *Nature*, *543*(7643), 51–59.

1423   https://doi.org/10.1038/nature21058

1424   Tam, W., Pell, L. G., Bona, D., Tsai, A., Dai, X. X., Edwards, A. M., et al. (2013). Tail tip

1425   proteins related to bacteriophage λ gpL coordinate an iron-sulfur cluster. *Journal of*

1426   *Molecular Biology*. https://doi.org/10.1016/j.jmb.2013.03.032

1427   Thompson, A. W., Huang, K., Saito, M. A., & Chisholm, S. W. (2011). Transcriptome response

1428   of high- and low-light-adapted Prochlorococcus strains to changing iron availability. *The*

1429   *ISME Journal*, *5*(10), 1580–94. https://doi.org/10.1038/ismej.2011.49

1430 Toulza, E., Tagliabue, A., Blain, S., & Piganeau, G. (2012). Analysis of the global ocean

1431   sampling (GOS) project for trends in iron uptake by surface ocean microbes. *PLoS ONE*,

1432   *7*(2). https://doi.org/10.1371/journal.pone.0030931

1433 Uitz, J., Claustre, H., Morel, A., & Hooker, S. B. (2006). Vertical distribution of phytoplankton

1434   communities in open ocean: An assessment based on surface chlorophyll. *Journal of*

1435   *Geophysical Research: Oceans*, *111*(8). https://doi.org/10.1029/2005JC003207

1436 Vandeputte, D., Kathagen, G., D'Hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., et al.

1437   (2017). Quantitative microbiome profiling links gut community variation to microbial load.

1438   *Nature*. https://doi.org/10.1038/nature24460

1439 Villar, E., Farrant, G. K., Follows, M., Garczarek, L., Speich, S., Audic, S., et al. (2015).

1440   Environmental characteristics of Agulhas rings affect interocean plankton transport.

1441   *Science*, *348*(6237), 1–12. https://doi.org/10.1126/science.1261447

1442 West, N. J., Lebaron, P., Strutton, P. G., & Suzuki, M. T. (2011). A novel clade of

1443   Prochlorococcus found in high nutrient low chlorophyll waters in the South and Equatorial

1444   Pacific Ocean. *The ISME Journal*, *5*(6), 933–944. https://doi.org/10.1038/ismej.2010.186

1445 Whitney, L. A. P., Lins, J. J., Hughes, M. P., Wells, M. L., Dreux Chappell, P., & Jenkins, B. D.

1446   (2011). Characterization of putative iron responsive genes as species-specific indicators of

1447   iron stress in thalassiosiroid diatoms. *Frontiers in Microbiology*, *2*(NOV), 1–14.

1448   https://doi.org/10.3389/fmicb.2011.00234

1449 Xing, X., Morel, A., Claustre, H., D'Ortenzio, F., & Poteau, A. (2012). Combined processing

1450   and mutual interpretation of radiometry and fluorometry from autonomous profiling Bio-

1451   Argo floats: 2. Colored dissolved organic matter absorption retrieval. *Journal of*

1452   *Geophysical Research: Oceans*, *117*(4), 1–14. https://doi.org/10.1029/2011JC007632

1453 Yoshinaga, R., Niwa-Kubota, M., Matsui, H., & Matsuda, Y. (2014). Characterization of iron-

1454   responsive promoters in the marine diatom Phaeodactylum tricornutum. *Marine Genomics*,

1455   *16*(1), 55–62. https://doi.org/10.1016/j.margen.2014.01.005

1456 Yuasa, T., Horiguchi, T., Mayama, S., & Takahashi, O. (2016). Gymnoxanthella radiolariae gen.

1457   et sp. nov. (Dinophyceae), a dinoflagellate symbiont from solitary polycystine radiolarians.
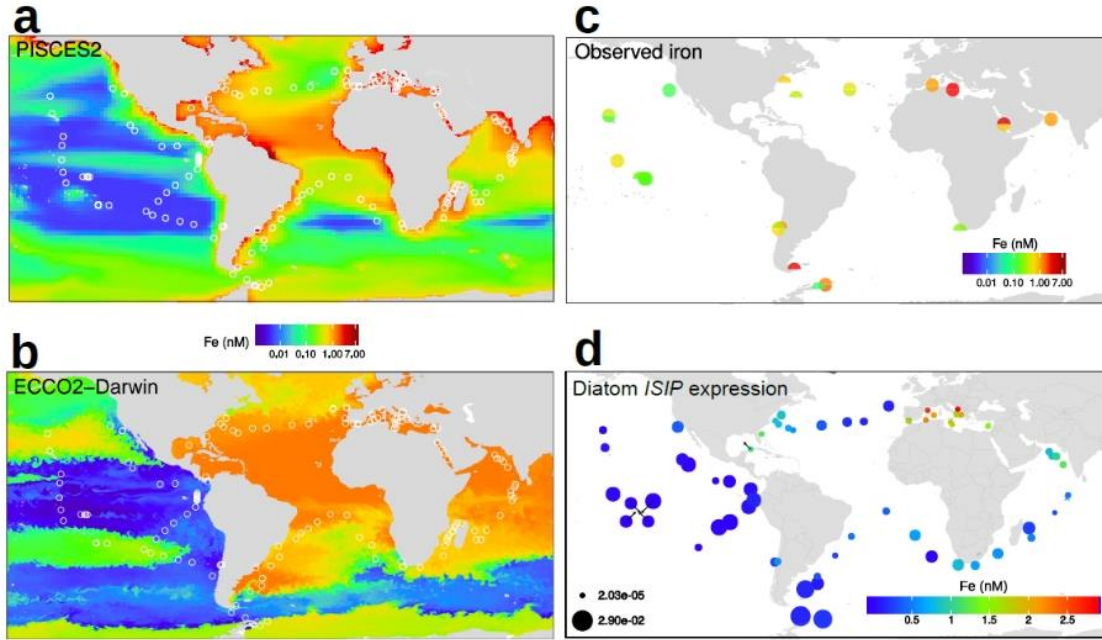
1458   *Journal of Phycology*, *52*(1), 89–104. https://doi.org/10.1111/jpy.12371

1459

1460

1461

1462

1463

1464

1465

1466

1467

1468

1469

1470

1471

1472

1473

1474
1475

Figures and Tables

Figure 1: Comparison of ECCO2-DARWIN, PISCES2 iron estimates with observed data and expression of diatom *ISIP* genes at *Tara* Oceans stations. Maps of (a) annual average iron concentrations from the ECCO2-DARWIN model (57 stations at surface), (b) from the PISCES2 model (83 stations at surface, 44 of which also at deep chlorophyll maximum depth), and (c) from the observed data where it was available at less than 2 degrees radius distance from locations of the *Tara* Oceans sampling sites (20 stations at surface, 16 of which also at deep chlorophyll maximum depth). Each circle corresponds to a sampling site, where the upper semicircle is filled according to the surface iron concentration while the lower semicircle is filled according to the deep chlorophyll maximum depth where available. Color scale indicates dissolved iron concentrations expressed in nM. (d) Biogeographical pattern of diatom *ISIP* gene expression. The circle colors represent iron concentration estimates at each *Tara* Oceans sampling site according to PISCES2 model (Table S1a). The abundance of *ISIP* transcripts was normalized by the total abundance of all diatom unigenes at each station, and the corresponding values are represented by the circle area.
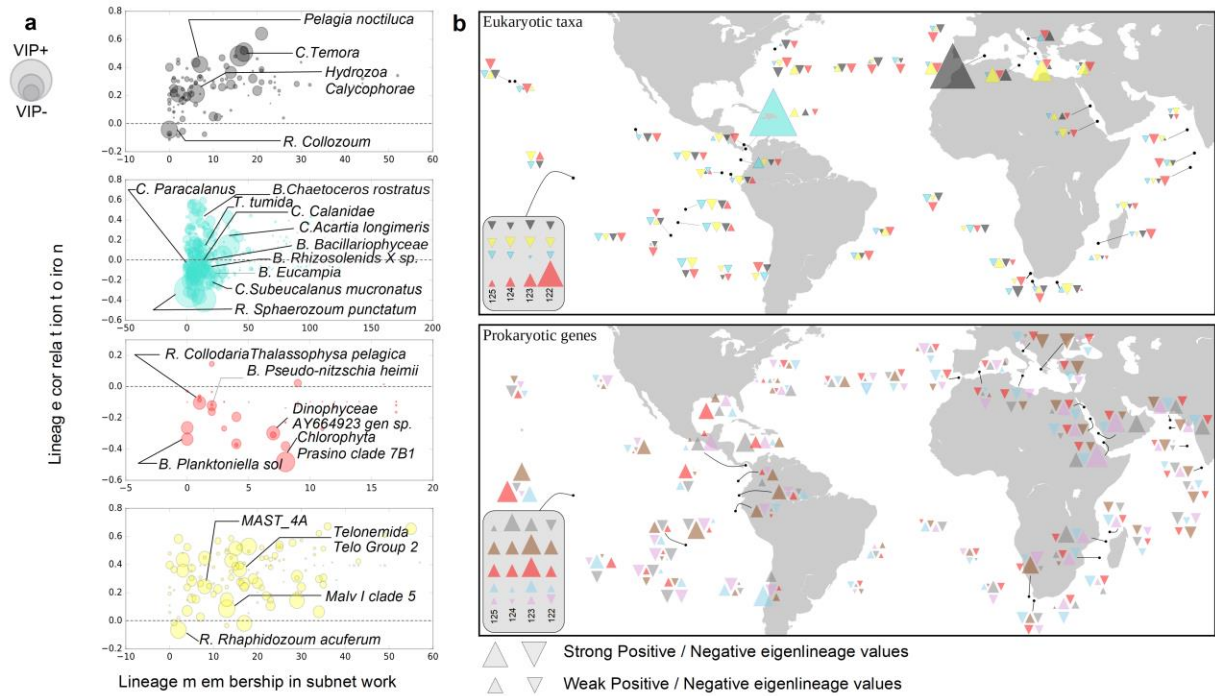
53

1510

1511

1512

Figure 2: **Planktonic Iron-Associated Assemblages (IAAs) in the global ocean and in the Marquesas Islands stations.** (a) Description of eukaryotic modules associated with iron. Relative abundances and co-occurrences of eukaryotic lineages were used to decipher modules. Four modules can predict iron with high accuracy: Black, DarkRed, Turquoise, and Yellow. For each IAA, lineages are associated to their score of centrality (x axis), to their correlation with iron concentrations (y axis), and their VIP score (circle area). Representative lineages within each module are emphasized by circles and named (C: Copepoda, B: Bacillariophyta, R: Rhizaria). (b) Upper panel: contribution of *Tara* Oceans stations to the global variance of IAAs of eukaryotic lineages. For each IAA, we represent the projection of stations on the first principal component (upper panel). Lower panel: projection of the relative contribution of the *Tara* Oceans stations to the global variance of iron-associated prokaryotic gene assemblages, as revealed by WGCNA. For each prokaryotic gene module associated with iron (from top to bottom: Grey60, Plum1, Red, SkyBlue, and SaddleBrown), we represent the projection of stations on the first principal component, proportional to triangle sizes for each module. The behavior of each IAA in the Marquesas archipelago stations is shown in the inset.
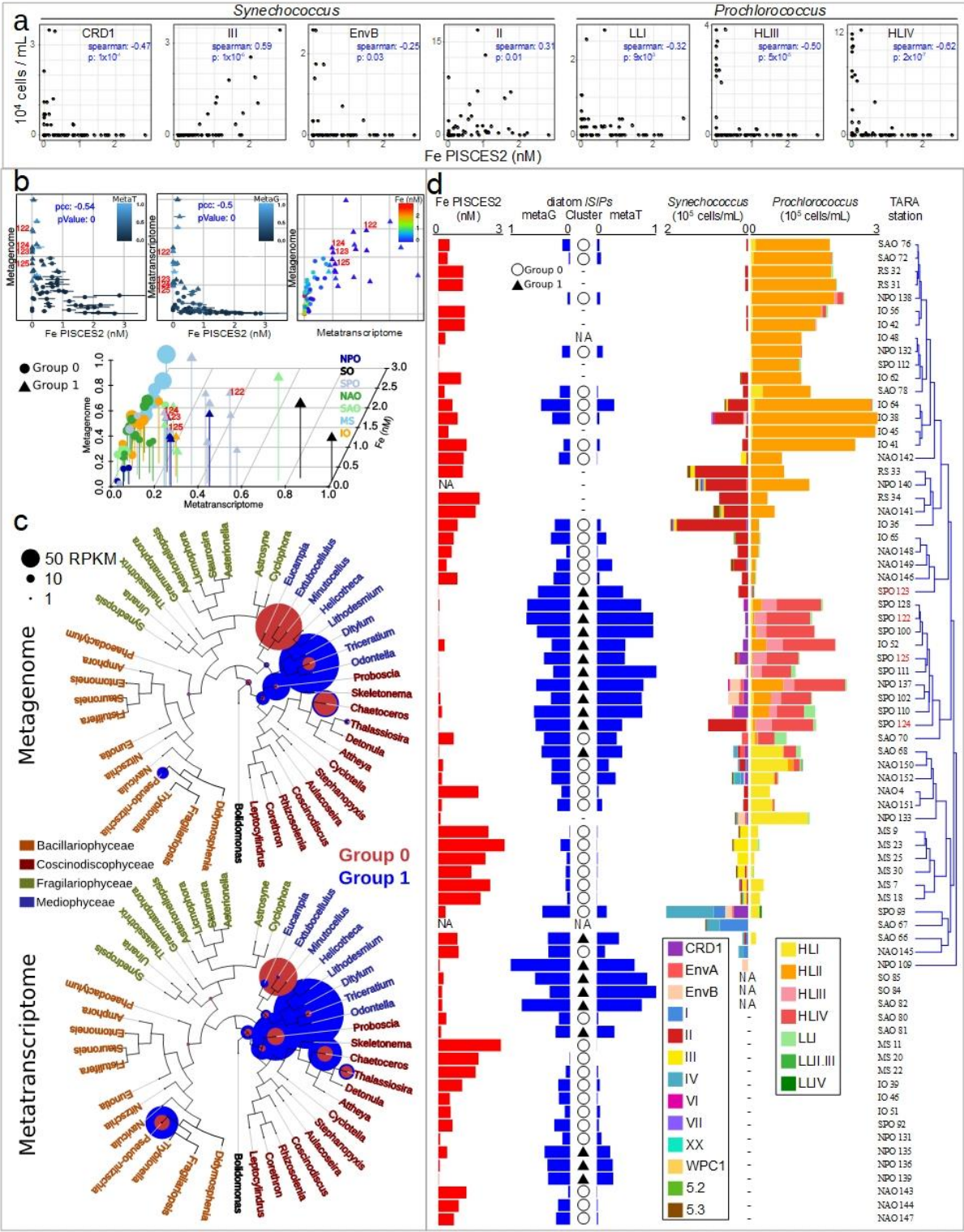
1528

54

1530 **Figure 3**: **Impact of predicted iron concentrations on marine picocyanobacterial community**

1531 **and diatom *ISIP* gene abundance and expression**. (a) Correlation analysis between absolute cell

1532 abundance of marine picocyanobacterial clades and iron concentration estimates from PISCES2

1533 model in surface waters. Only statistically significant correlations are displayed (p-value<0. 05).

1534 Spearman correlation coefficients and p-values are indicated. (b) Abundance and expression of

1535 diatom *ISIP* genes with respect to iron concentration estimates. 2D scatter plots correspond to the

1536 correlation of *ISIP* relative gene abundance and iron (left), *ISIP* relative gene expression and iron

1537 (middle), and relative abundance and expression of *ISIP* genes (right). Pearson correlation

1538 coefficients (pcc) and p-values are indicated in blue. Iron concentrations were estimated using

1539 PISCES2 model (Table S1a). Horizontal error bars correspond to the standard deviation of

1540 modeled iron concentrations, which is related to the modeled annual average fluctuation of their

1541 levels. In all cases, the abundance and expression of *ISIP* genes were normalized by the total

1542 diatom unigene abundance and expression, respectively, and were then scaled to the unit interval.

1543 The 3D plot shown below is derived from the three 2D scatter plots above, with the color gradient

1544 representing the third dimension. The data were clustered using density clustering algorithms,

1545 resulting in a group of *Tara* Oceans sampling sites in which *ISIP* was only increased in

1546 metagenomics data (denoted Group 0 stations (40 stations; circles)) and others in which both

1547 metagenomic and metatranscriptomic data showed increases in *ISIP* levels (denoted Group 1

1548 stations; 21 stations; triangles). The values corresponding to *Tara* Oceans stations in the Marquesas

1549 archipelago are labeled (122 - 125). *Tara* Oceans sampling sites are colored according to the ocean

1550 region in the 3D plot: NPO, North Pacific Ocean; SO, Southern Ocean; SPO, South Pacific Ocean;

1551 NAO, North Atlantic Ocean; SAO, South Atlantic Ocean, MS, Mediterranean Sea; IO, Indian

1552 Ocean. An interactive version of the 3D plot can be found at https://figshare.

1553 com/s/0e60410ce0b752087d21. (c) Relative abundance (above) and expression (below) of *ISIP*

1554 genes assigned at different levels of resolution in a diatom phylogenetic tree. The color code

1555 corresponds to the two clusters of stations defined in panel B based on *ISIP* patterns (red for Group

1556 0 with variations only at metagenome levels; blue for Group 1 with variations in both metagenome

1557 and metatranscriptome levels). (d) Comparisson of iron-driven changes in diatom *ISIP* gene

1558 abundance and expression and in the picocyanobacterial community from surface waters.

1559 Histograms of cell abundance of *Synechococcus* and *Prochlororococcus* clades at each *Tara*

1560 Oceans station are displayed, with stations sorted by hierarchical clustering of a Bray-Curtis

56

1561   distance matrix. The left panels indicate iron concentration estimates from PISCES2 model, and

1562   metagenome and metatranscriptome levels of diatom *ISIP* genes, including the resulting cluster

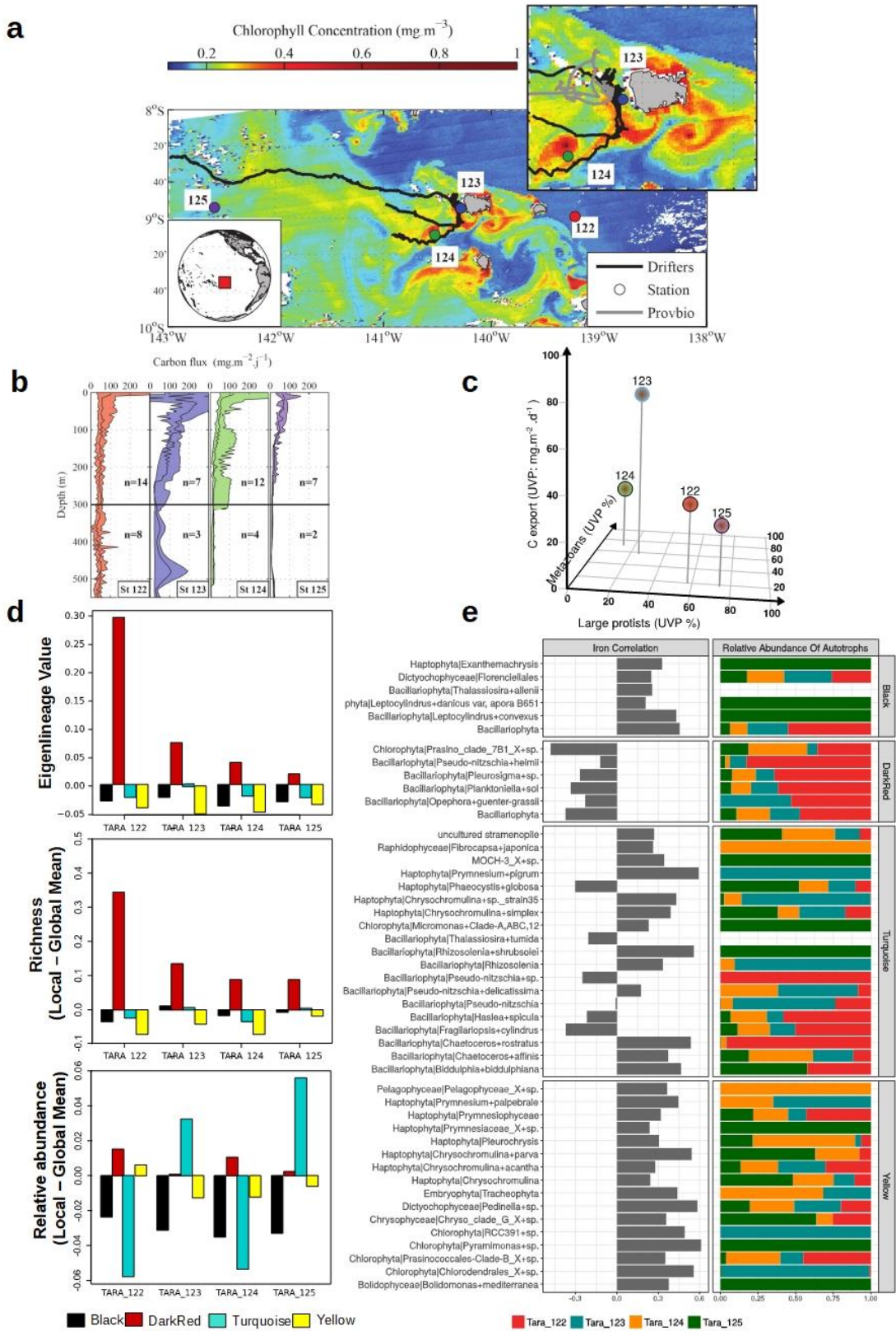1563   type (circles and triangles as described in panel b).

1564

1565

1566

1567

1568

1569

1570

1571

1572

1573

1574

1575

1576

1577

1578

1579

1580

1581

1582

1583

1584

1585

1586

1587

1588

1589

1590

1591

1592

1593

1594

1595

1596

1597

1599 **Figure 4: The Marquesas study site, showing sampling sites, surface chlorophyll**
1600 **concentrations, and the local dynamics of the four eukaryotic IAAs.** (a) Map of surface
1601 chlorophyll in the Marquesas area. Drifter and Provbio trajectories are indicated as well as the *Tara*
1602 Oceans sampling stations, with a zoom on Stations TARA_123 and TARA_124. For further details
1603 see main text and Supplementary Material S2. (b) Profile of carbon flux in mg $\cdot$ m$^{-2.}$ d^{-1} along the
1604 water column at Stations TARA_122-125 estimated from UVP measurements of particle size
1605 distribution and abundance. A significant change in particle size spectra was observed at
1606 TARA_123, indicating strong vertical export of organic matter, whereas the carbon flux to depth
1607 at TARA_124 was similar to what was found at TARA_122 (30 mg $\cdot$ m$^{-2.}$ d^{-1}). c) Relative
1608 contribution (%) of Metazoa (x axis) and of large protists (y axis) to the total carbon export (z axis
1609 - mg $\cdot$ m$^{-2.}$ d^{-1}). The data indicate that metazoan lineages played a major role in the significant
1610 increase in carbon export detected at Station TARA_123. Colors used for each TARA station are
1611 the same in A, B and C. (d) Analysis of the dynamics of the four IAAs at the Marquesas archipelago
1612 stations in relation to their eigenlineage values (upper), richness (middle) and relative abundance
1613 (lower). All modules show negative eigenlineage values, with the exception of the DarkRed IAA.
1614 The DarkRed module positive eigenlineage scores significantly decrease within the bloom
1615 stations. The mean IAA relative abundance calculated over the global *Tara* Oceans dataset was
1616 subtracted from IAA relative abundance calculated at the Marquesas Islands. The increase in
1617 DarkRed relative abundance in station TARA_124 was due to a single *Prasinophycae* OTU. The
1618 mean IAA richness calculated over the global *Tara* Oceans dataset was subtracted from IAA
1619 richness calculated at the Marquesas Islands. Data indicates that the DarkRed IAA retains ca. 60%
1620 of its OTUs in low iron conditions, a percentage that decreases in the bloom stations. (e) Relative
1621 abundance changes at the Marquesas Islands stations for IAA photosynthetic lineages with high
1622 iron correlation. The graph shows the list of IAA autotroph lineages with the highest statistically
1623 significant correlations against PISCES2 iron estimates ($p<0.05$) at a global scale, with the
1624 corresponding Pearson correlation coefficient, and their relative abundance at the Marquesas
1625 Islands sampling sites.
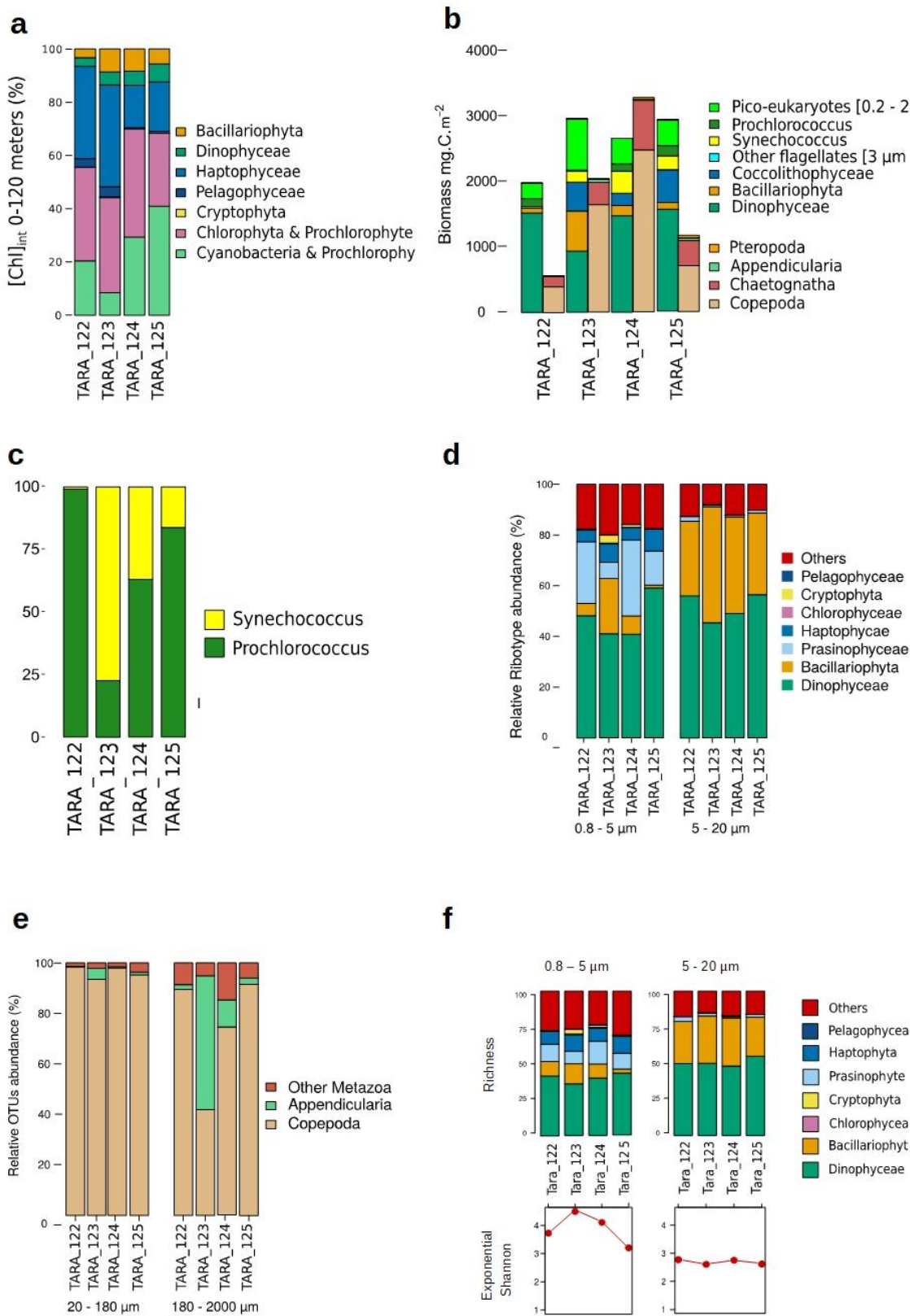
1626
1627
1628
1629
1630

59

Figure 5: Variations in plankton community composition at Marquesas sampling sites. (a) Relative contribution of different autotrophic lineages to the total chlorophyll concentration in the euphotic zone (0 - 120 m), derived from photosynthetic pigment analysis and expressed as percent of the total measured chlorophyll. (b) Depth-integrated biomass (mg C m^{-2}) of autotrophs and mesozooplankton (> 300 μm) in the euphotic zone (0 – 120 m). (c) Relative abundance of *Prochlorococcus* and *Synechococcus* picocyanobacteria expressed as percent of the total *Prochlorococcus* plus *Synechococcus* abundance estimated from flow cytometry data. Genetic markers (*petB*) showed exactly the same trends (Supplementary Material S3). (d) Relative abundance (%) of ribotypes (18S-V9 tags) assigned to autotrophic eukaryote lineages at the surface (5 meters depth). Abundances were computed for the two size fractions containing the majority of autotrophic lineages, namely 0. 8 - 5 μm and 5 – 20 μm size fractions. (e) Relative abundance (%) of ribotypes (18S-V9 Tags) assigned to metazoan lineages at the surface (5 m depth). Abundances were computed for the two size fractions containing the majority of metazoans, namely the 20 – 180 μm and 180 – 2000 μm size fractions. (f) Richness and diversity (exponential Shannon index) of eukaryotic autotrophs in two different size fractions estimated from the metabarcode data.
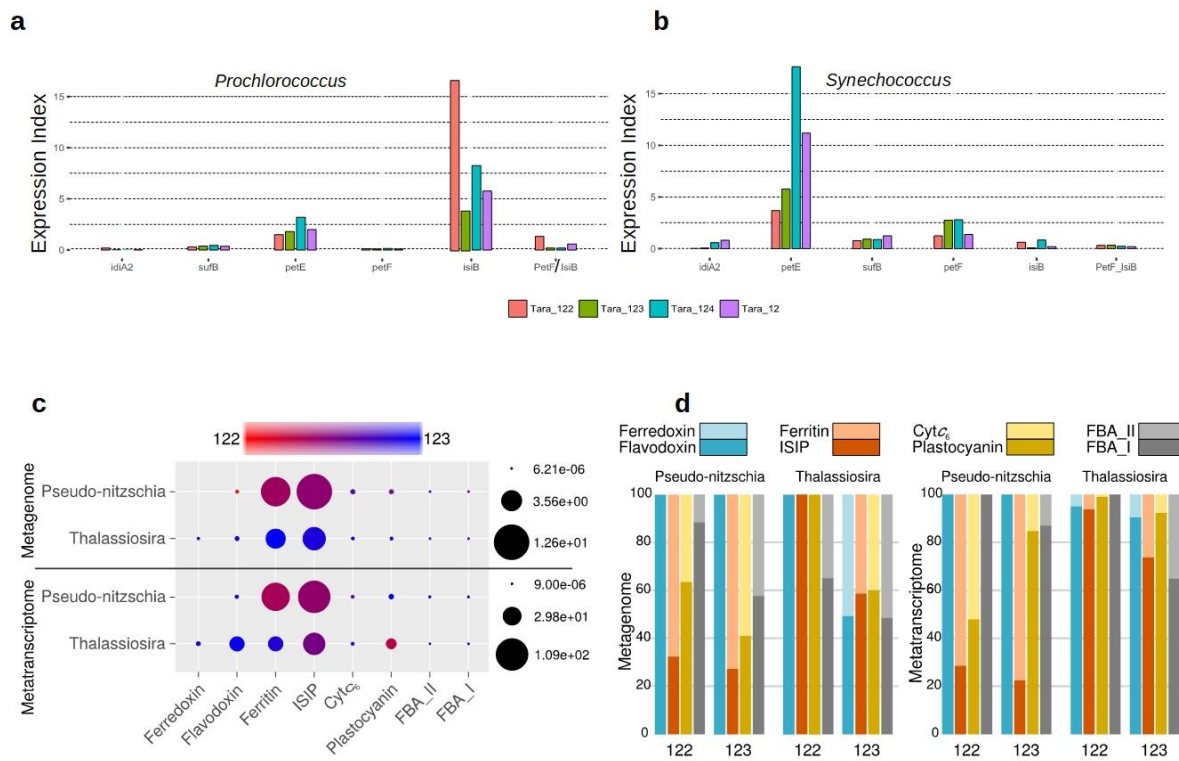
Figure 6: Variations in gene abundance and expression in cyanobacteria and diatoms at Marquesas sampling sites. (a-b) Differential expression patterns of iron-related genes from cyanobacteria *Prochlorococcus* (a) and *Synechococcus* (b) at stations TARA_122-125. Transcription values were normalized over genomic occurrence and are expressed relative to the levels observed at Station TARA_122 (index 100). The flavodoxin/ferredoxin ratio is also plotted (PetF/IsiB). (c) Relative abundances and mRNA levels of diatom genes potentially responsive to iron in metagenome and metatranscriptome datasets from Stations TARA_122 and 123. Values were normalized by total abundance or expression of all unigenes assigned to the corresponding taxonomic group (*Pseudo-nitzschia* and *Thalassiosira*). For clarity we focused only on changes in 5-20 μm size fractions. Colors indicate the contribution of each station to the total levels. (d) Relative ratios between pairs of genes whose presence in the genome or transcriptional activity has been reported previously to be potentially responsive to iron bioavailability. For clarity, ferritin

1700 levels have been multiplied by a factor of 10 to be comparable with ISIP levels, and only 5-20 μm

1701 size fractions from Stations TARA_122 and 123 are compared.