



HAL
open science

Restructuring of genomic provinces of surface ocean plankton under climate change

Paul Frémont, Marion Gehlen, Mathieu Vrac, Jade Leconte, Patrick Wincker,
Daniele Iudicone, Olivier Jaillon

► **To cite this version:**

Paul Frémont, Marion Gehlen, Mathieu Vrac, Jade Leconte, Patrick Wincker, et al.. Restructuring of genomic provinces of surface ocean plankton under climate change. 2021. cea-03166589v1

HAL Id: cea-03166589

<https://cea.hal.science/cea-03166589v1>

Preprint submitted on 11 Mar 2021 (v1), last revised 27 Aug 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

1 **Restructuring of genomic provinces of surface ocean** 2 **plankton under climate change**

3 Paul Frémont^{1,2*}, Marion Gehlen^{3*}, Mathieu Vrac³, Jade Leconte^{1,2}, Patrick Wincker^{1,2},
4 Daniele Iudicone⁴, Olivier Jaillon^{1,2*}

5 ¹Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Université d'Evry, Université Paris-
6 Saclay, 91057 Evry, France. ²Research Federation for the study of Global Ocean Systems Ecology and
7 Evolution, FR2022/Tara Oceans, Paris, France. ³LSCE-IPSL, CEA/CNRS/Université Paris-Saclay, Gif-sur-
8 Yvette, France. ⁴Stazione Zoologica Anton Dorn. Villa Comunale, 80121, Naples, Italy.

9 *Corresponding authors: pfremont@genoscope.cns.fr; marion.gehlen@lsce.ipsl.fr; ojaillon@genoscope.cns.fr
10

11 **Abstract**

12 **The impact of climate change on diversity, functioning and biogeography of marine**
13 **plankton is a major unresolved scientific issue. Here, niche theory is applied on**
14 **plankton metagenomes sampled during the *Tara Oceans* expedition to derive pan-**
15 **ocean geographical structuring in *climato-genomic* provinces characterized by**
16 **signature genomes for 6 size fractions, from viruses to meso-zooplankton. Assuming**
17 **a high warming scenario (RCP8.5), the identified tropical provinces would expand**
18 **and temperate provinces would shrink. Poleward shifts are projected for 96% of**
19 **provinces in five major basins leading to their reorganization over ~50% of the**
20 **surface ocean south of 60°N, of which 3% correspond to novel assemblages of**
21 **provinces. Sea surface temperature is identified as the main driver and accounts only**
22 **for ~51 % of the changes followed by phosphate (11%) and salinity (10.3%). These**
23 **results demonstrate the potential of integration of genomics with physico-chemical**
24 **data for higher scale modeling and understanding of ocean ecosystems.**

25 Planktonic communities are composed of complex and heterogeneous assemblages of small
26 animals, small single-celled eukaryotes (protists), prokaryotes and viruses - that drift with
27 currents. They contribute to the regulation of the Earth system notably through primary
28 production via photosynthesis¹, carbon export to the deep oceans^{2,3} and form the base of
29 the food webs that sustain the whole trophic chain in the oceans and beyond⁴.

30 The composition of communities is known to vary over time at a given site with daily⁵ to
31 seasonal fluctuations⁶ following environmental variability^{7,8}. Overlying these relatively

32 short scale spatio-temporal variations, a more macroscale partitioning of the ocean was
33 evidenced by different combinations of biological and physico-chemical data⁹⁻¹¹, and
34 recently at the resolution of community genomics¹². The basin scale biogeographical
35 structure was proposed to result from a combination of multiple bio-physico-chemical
36 processes named the seascape^{7,8}, including both abiotic and biotic interactions¹³, neutral
37 genetic drift¹⁴, natural selection¹⁵⁻¹⁷, temperature variations, nutrient supply but also
38 advection and mixing along currents^{12,14}.

39 Today, knowledge of global scale plankton biogeography at the DNA level is in its infancy.
40 We lack understanding and theoretical explanations for the emergence and maintenance of
41 biogeographical patterns at genomic resolution. Omics data (*i.e.* the DNA/RNA sequences
42 representative of the variety of coding and non-coding sequences of organisms) provide
43 the appropriate resolution to track and record global biogeographical features¹²,
44 modulation of the repertoire of expressed genes in a community in response to
45 environmental conditions^{2,18,19} as well as eco-evolutionary processes^{14,16,17}. Importantly,
46 metagenomic sequencing can be consistently analyzed across plankton organisms as
47 recently demonstrated by global expeditions²⁰⁻²³. Furthermore, the strong links between
48 plankton and environmental conditions suggest potentially major consequences of climate
49 change on community composition and biogeography^{24,25}. Time series observations have
50 highlighted recent changes in the planktonic ecosystem attributed to this anthropogenic
51 pressure, such as changes in community composition²⁶⁻²⁸ or poleward shifts of some
52 species^{29,30}. These changes are expected to intensify with ongoing climate warming and
53 could lead to major reorganizations in plankton community composition^{24,25}, with a
54 potential decline in diversity³¹⁻³³. Another major consequence of a global reorganization of
55 the seascape on biological systems (e.g. growth, grazing) would be a decrease of primary
56 production at mid-latitudes and an increase at higher latitudes³⁴.

57 Here we report a global structure of plankton biogeography based on metagenomic data
58 using niche models and its putative modifications under climate change. First, we show
59 that environmental niches³⁵, *i.e.* the envelope of environmental parameters suitable for an
60 organism or a population, can be defined at the scale of genomic provinces across 6
61 organism size fractions representing major plankton groups from nano- (viruses) to meso-
62 zooplankton (small metazoans). Then, we spatially extrapolate their niches into *climato-*

63 *genomic* provinces to depict the structure of plankton biogeography of all but arctic regions
64 for each size fraction and for all combined. Then, considering the same niches, we assess
65 putative spatial reorganization of the same provinces and their associated environmental
66 drivers under climate change at the end of the century.

67 **Niche models and signature genomes from genomic provinces**

68 We use 38 previously defined genomic provinces¹² containing at least 4 sampling sites;
69 they correspond to 595 metagenomes for 6 size fractions (ranging from 0 to 2000 μm) and
70 sampled in 95 sites from all oceans except the Arctic (Supplementary Figs. 1-2).

71 To compute and test the validity of realized environmental niches, we train four machine
72 learning techniques to probabilistically associate genomic provinces with environmental
73 data: sea surface temperature, salinity, three macronutrients (dissolved silica, nitrate and
74 phosphate), one micronutrient (dissolved iron) plus a seasonality index of nitrate. A valid
75 environmental niche is obtained for 27 out of 38 initial provinces (71%) comforting their
76 definition and covering 529 samples out of 595 (89%, Supplementary Fig. 2). Rejected
77 provinces contain relatively few stations (mean of 6 ± 2.6 versus 19 ± 15.3 for valid
78 provinces, $p\text{-value} < 10^{-3}$ Wilcoxon test). For spatial and temporal extrapolations of the
79 provinces presented below, we use the ensemble model approach³⁶ that considers mean
80 predictions of machine learning techniques.

81 The signal of ocean partitioning is likely due to abundant and compact genomes whose
82 geographical distributions closely match provinces. Within a collection of 523 prokaryotic
83 and 713 eukaryotic genomes^{37,38} from *Tara* Oceans samples, we find signature genomes for
84 all but 4 provinces. In total, they correspond to respectively 96 and 52 of the genomes and
85 their taxonomies are coherent with the size fractions (Fig. 1 for eukaryotes and
86 Supplementary Fig. 3 for prokaryotes). Some of them correspond to unexplored lineages
87 highlighting the gap of knowledge for organisms that structure plankton biogeography and
88 the strength of a rationale devoid of any *a priori* on reference genomes or species.

89 **Structure of present day biogeography of plankton**

90 To extrapolate to a global ocean biogeography for each size fraction, we define the most
91 probable provinces, named hereafter as *dominant* and assigned to a climatic annotation
92 (Supplementary Table 1), on each $1^\circ \times 1^\circ$ resolution grid point using 2006-13 WOA13
93 climatology³⁹ (Supplementary Fig. 4 and Fig. 3).

94 Overall, in agreement with previous observations¹², provinces of large size fractions (>20
95 μm) are wider and partially decoupled from those of smaller size fractions, probably due to
96 differential responses to oceanic circulation and environmental variations, different life
97 cycle constraints, lifestyles^{7,8,12} and trophic networks positions⁴⁰. Biogeographies of small
98 metazoans that enrich the largest size fractions (180-2000 μm and 20-180 μm) are broadly
99 aligned with latitudinal bands (tropico-equatorial, temperate and (sub)-polar) dominated
100 by a single province (Fig. 2a,b). A more complex oceanic structuring emerges for the
101 smaller size fractions (<20 μm) (Fig. 2c-f) with several provinces per large geographical
102 region. Taking size fraction 0.8-5 μm enriched in small protists (Fig. 2d) as an example,
103 distinct provinces are identified for oligotrophic gyres in the Atlantic and Pacific Oceans,
104 and one for the nutrient-rich equatorial upwelling region. A complex pattern of provinces,
105 mostly latitudinal, is also found for the bacteria enriched size class (Fig. 2e, 0.22-3 μm) and
106 the virus enriched size class (Fig. 2f, 0-0.2 μm) though less clearly linked to large-scale
107 oceanographic regions. A single province extending from temperate to polar regions
108 emerges from the size fraction 5-20 μm enriched in protists (Fig. 2c), for which a smaller
109 number of samples is available (Supplementary Fig. 2b-c), which probably biases this
110 result. Finally, we use PHATE⁴¹ dimension reduction algorithm to combine all provinces for
111 all size classes into a single consensus biogeography revealing 4 or 7 robust clusters (Fig.
112 2g,h). The 4 cluster consensus biogeography is mainly latitudinally organized
113 distinguishing polar, subpolar, temperate and tropico-equatorial regions. The 7 cluster
114 consensus biogeography distinguishes the equatorial pacific upwelling biome and three
115 subpolar biomes that most likely reflect the chemico-physical structuring of the Southern
116 Ocean and known polar fronts (red lines Fig. 2h). However, learning data are scarcer south
117 of 60°S so these extrapolations need to be taken with caution.

118 Previous ocean partitioning either in biomes⁹⁻¹¹ or biogeochemical provinces (BGCPs)^{9,11}
119 are based on physico-biogeochemical characteristics including SST⁹⁻¹¹, chlorophyll *a*⁹⁻¹¹,
120 salinity⁹⁻¹¹, bathymetry^{9,11}, mixed layer depth¹⁰ or ice fraction¹⁰. Considering three of these
121 partitions as examples we notice differences with our partitions (Supplementary Fig. 7-8)
122 for example in terms of number of regions in the considered oceans (56 for 2013
123 Reygondeau et al. BGCPs¹¹, 17 for Fay and McKingley¹⁰) and structure (the coastal biome
124 for 2013 Reygondeau et al. biomes¹¹). Numerical comparison of our partitions with others

125 (*Methods*), reveals low similarity between them, the highest being with Reygondeau biomes
126 (Supplementary Figs. 7-9). However, biomes and BGCP frontiers closely match our
127 province frontiers in many cases. Near the frontiers, *dominant* provinces have smaller
128 probabilities in agreement with smooth transitions instead of sharp boundaries as already
129 proposed⁹ and with a seasonal variability of the frontiers¹¹ (Supplementary Fig 7). Some of
130 these transitions are very large and match entire BGCPs, for example in subtropical North
131 Atlantic and subpolar areas where high annual variations are well known¹¹.

132 **Future changes in plankton biogeography structure**

133 We assess the impacts of climate change on plankton biogeography at the end of the 21st
134 century following the Representative Concentration Pathway 8.5 (RCP8.5)⁴² greenhouse
135 gas emission scenario. To consistently compare projections of present and future
136 biogeographies with coherent spatial structures, we use a bias-adjusted mean of 6 Earth
137 System Model (ESM) climatologies (Supplementary Table 2, Supplementary Fig. 10,
138 *Materials and Methods*). The highest warming (7.2°C) is located off the east coast of Canada
139 in the North Atlantic while complex patterns of salinity and nutrient variations are
140 projected in all oceans (Supplementary Figure 11). According to this scenario, future
141 temperature at most sampling sites will be higher than the mean and maximum
142 contemporary temperature within their current province (Supplementary Fig. 12).

143 Our projections indicate multiple large-scale changes in biogeographical structure
144 including plankton organism size-dependent province expansions or shrinkages and shifts
145 (Supplementary Fig. 13-15). A change in the *dominant* province in at least one size fraction
146 would occur over 60.1 % of the ocean surface, ranging from 12 % (20-180 µm) to 31%
147 (0.8-5 µm) (Table 1).

148 Centroids of provinces with *dominance* areas larger than 10⁶ km² within a basin would be
149 moved at least 200 km away for 77% of them, 96 % of which move poleward
150 (Supplementary Figs. 14 and 14). Most longitudinal shift distances are smaller (50% <190
151 km) but a few are larger than 1000 km while the distribution of latitudinal shifts is more
152 concentrated around the mean (290 km) with no shifts superior to 1000 km
153 (Supplementary Fig. 15b). These important longitudinal shifts corroborate existing
154 projections^{24,43,44} and differ from trivial poleward shifts due to temperature increase,
155 reflecting more complex spatial rearrangements of the other environmental drivers

156 (Supplementary Fig. 11). We find an average displacement speed of the provinces'
157 centroids of $76 \pm 79 \text{ km.dec}^{-1}$ (latitudinally mean of $34 \pm 82 \text{ km.dec}^{-1}$ and $59 \pm 82 \text{ km.dec}^{-1}$
158 longitudinally; median of 47 km.dec^{-1} , 23 km.dec^{-1} latitudinally and 27 km.dec^{-1}
159 longitudinally). We project phytoplankton provinces displacements similar to previously
160 published shifts for the North Pacific (poleward shift of 118 km.dec^{-1} for province C4
161 versus 100 km.dec^{-1} for the subtropical biome of Polovina et al.⁴⁴ and eastward shift of 195
162 km.dec^{-1} for province C9 versus 200 km.dec^{-1} for the equatorial biome of Polovina et al.⁴⁴).
163 For all size fractions, climate change would lead to a poleward expansion of tropical and
164 equatorial provinces at the expense of temperate provinces (Table 1 and Supplementary
165 Fig. 13). This is illustrated for the size fraction 180-2000 by the example of the temperate
166 province F5 (Supplementary Fig. 16) for which significant shrinkage is projected in the five
167 major ocean basins. In the North Atlantic, its centroid would move approximately 800 km
168 to the northeast (Supplementary Fig. 16c).

169 To simplify comparisons between future and present biogeographies, we combine all
170 projections into two comparable consensus maps (Fig. 2e,f). Some particularly visible
171 patterns of geographical reorganization are common to several or even all size fractions
172 and visible on the consensus maps (Fig. 2e,f compared to Fig. 2a-d and Supplementary Fig.
173 13). For example, the tropico-equatorial and tropical provinces expand in all size fractions
174 and the provinces including the pacific equatorial upwelling shrink for size fractions
175 smaller than $20 \mu\text{m}$.

176 To further quantify patterns of expansion or shrinkage, we calculate the surface covered by
177 the *dominant* provinces weighted by probabilities of presence (Supplementary Fig. 17,
178 Supplementary Table 1). In this way, *dominant* provinces are defined on 100% of the
179 surface ocean (327 millions of km^2) but their presence probabilities correspond to the
180 equivalent of 45 to 74% (due to sampling variability and niche overlaps) of the surface
181 ocean depending on plankton size fraction (Table 1). Overall, our results indicate
182 expansions of the surface of tropical and tropico-equatorial provinces but in very different
183 ways depending on the size fractions of organisms. The surface area of temperate
184 provinces is ~ 22 million km^2 on average (from 10 Mkm^2 for $0\text{-}0.2 \mu\text{m}$ to 49 Mkm^2 for 20-
185 $180 \mu\text{m}$) and should decrease by 36% on average (from -20 % for $5\text{-}20 \mu\text{m}$ up to -54% for
186 $0.8\text{-}5 \mu\text{m}$, -12 million km^2 on average, +6 % for $0.22\text{-}3 \mu\text{m}$). Tropical provinces cover ~ 118

187 million km² on average (from 86 Mkm² for 0.8-5 μm up to 169 Mkm² for 180-2000 μm) and
188 their coverage should increase by 32% on average (from +13% for 0-0.2 μm up to +75%
189 for 0.8-5 μm, +25 million km² on average) (Supplementary Fig. 17 and Supplementary
190 Table 1).

191 We calculate at each grid point a single dissimilarity index (*Materials and Methods*)
192 between probabilities of future and present *dominant* provinces for all size fractions
193 combined (Fig. 4a). Areas located between future and present borders of expanding and
194 shrinking provinces would be the most subject to replacements by other contemporary
195 provinces, as exemplified by the poleward retraction of the southern/northern edges of the
196 temperate provinces (red arrows on Fig. 4a, Table 1). High dissimilarities are obtained
197 over northern (25° to 60°) and symmetrically southern (-25 to -60°) temperate regions
198 (mean of 0.29 and 0.24 respectively). Despite important environmental changes in austral
199 and equatorial regions (Supplementary Fig. 11) and projected change in diversity³¹⁻³³ and
200 biomass⁴⁵, the contemporary provinces remain the most probable at the end of the century
201 using our statistical models (mean dissimilarities of 0.18 and 0.02 respectively) as no
202 known contemporary provinces could replace them.

203 To further study the decoupling between provinces of different size fractions in the future
204 we considered the assemblages of *dominant* provinces of each size fraction. By using two
205 differently stringent criteria, from 45.3 to 57.1% of ocean surface, mainly located in
206 temperate regions, would be inhabited in 2090-99 by assemblages that exist elsewhere in
207 2006-15 (Fig. 4b versus Fig. 4c). Contemporary assemblages would disappear on 3.5 to
208 3.8% of the surface, and, conversely, novel assemblages, not encountered today, would
209 cover 2.9 to 3.0% of the surface. These areas appear relatively small but they include some
210 important economic zones (Fig 4b, Supplementary Fig. 18). On 41.8% to 51.8% of the
211 surface of the main fisheries and 41.2% to 54.2% of Exclusive Economic Zones (*Materials*
212 *and Methods*), future assemblages would differ from those present today (Supplementary
213 Fig. 18).

214 **Drivers of plankton biogeography reorganization**

215 We quantify the relative importance of considered environmental parameters
216 (temperature, salinity, dissolved silica, macronutrients and seasonality of nitrate) into
217 niche definition and in driving future changes of the structure of plankton biogeography.

218 Among environmental properties that define the niches, temperature is the first influential
219 parameter (for 19 niches out of 27) but only at 22.6% on average. Particularly, the
220 distribution of relative influences of temperature is spread over a much wider range than
221 that of other parameters (Supplementary Fig. 18a).

222 The relative impact of each environmental parameter is calculated²⁴ for each site
223 presenting a significant dissimilarity between 2006-15 and 2090-99 (Fig 5a). Overall, SST
224 would be responsible for reorganization of half of the provinces followed by Phosphate (11
225 %) and Salinity (10.3%) (Supplementary Fig. 20). SST is the primary driver over the
226 majority of the ocean (Fig. 5a). In some regions, salinity (e.g. eastern North Atlantic) and
227 Phosphate (e.g. equatorial region) dominate (Fig. 5a) and excluding the effect of SST, they
228 are the primary drivers of global reorganization of the provinces (Fig. 5 b). The impact of
229 SST varies across size classes with a significantly higher contribution in large size classes
230 (>20 μm) compared to the small ones (mean of $\sim 73\%$ versus $\sim 49\%$; Fig. 5c). Even though
231 the contribution of combined nutrients to niche definition is similar for small and large size
232 classes under present day conditions (mean of $\sim 56\%$ versus $\sim 61\%$, Supplementary Fig. 19,
233 Supplementary Table 3), their future projected variations have a higher relative impact on
234 the reorganization of small organisms' biogeographies (mean of $\sim 39\%$ versus $\sim 20\%$, t-test
235 p-value < 0.05, Supplementary Fig. 19, Supplementary Table 3). For instance, in the tropical
236 zone, the shrinkage of the equatorial province (province C9, size fraction 0.8 – 5 μm , Fig.
237 2b,d, Supplementary Fig. 16e) is driven at 24 % by reduction of dissolved phosphate
238 concentrations and at 25% by SST increase. In contrast, SST and Salinity would drive
239 respectively at 56% and 16% the shrinkage of the temperate province F5 of size fraction
240 180-2000 μm (Supplementary Fig. 16d versus e). Non-poleward shifts are found only
241 within small size fractions (<20 μm) (Supplementary Fig. 14, 15) highlighting differential
242 responses to nutrients and SST changes between large and small size classes, the latter
243 being enriched in phytoplankton that directly rely on nutrient supplies.

244 **Discussion**

245 We propose a novel partitioning of the ocean in plankton size dependent *climato-genomic*
246 provinces, complementing pioneer and recent efforts based on other bio-physico-chemical
247 data⁹⁻¹¹. Though they are initially built at genomic scale, our biogeographies paradoxically
248 reveal basin scale provinces that are larger than BGCPs¹¹ and biomes¹⁰, and probably

249 relatively stable across seasons evoking limited effects of seasonality on frontiers positions
250 of BGCPs provinces¹¹. We propose that this apparent paradox emerges from the
251 combination of the scale, nature and resolution of sampling. First, two proximal samples
252 from the *Tara* Oceans expedition are separated by ~300 km on average sampled over three
253 years; this relatively large spatio-temporal scale overlies shorter scale compositional
254 variations previously observed^{5,6}. Second, our estimates of plankton community
255 dissimilarities are highly resolute; they are computed at genomic scale with billions of
256 small DNA fragments^{12,20} thus smoothing the more discrete species level signal. Together,
257 from these combinations of processes and patterns occurring at multiple scales emerge
258 basin scale provinces associated with coherent environmental niches and signature
259 genomes.

260 These *climato-genomic* provinces are structured in broad latitudinal bands with smaller
261 organisms (<20 μ m) displaying more complex patterns and partially decoupled from larger
262 organisms. This decoupling is the result of distinct statistical links between provinces
263 based on organism size fractions and environmental parameters and could reflect their
264 respective trophic modes⁴⁰.

265 Complex changes of the parameters defining the niches are projected under climate change
266 leading to size-dependent modifications of biogeographical patterns, as for example
267 smaller organisms being more sensitive to nutrient changes. Assuming the maintenance of
268 environmental characteristics that define the *climato-genomic* provinces, climate change is
269 projected to restructure them over approximately 50% of surface oceans south of 60°N by
270 the end of the century (Fig. 4). The largest reorganization is detected in subtropical and
271 temperate regions in agreement with other studies^{32,44} and is accompanied by appearance
272 and disappearance of size-fractionated provinces' assemblages. For tropico-equatorial and
273 austral regions, out of contemporary range and novel environmental conditions are
274 projected. While some studies extrapolate important diversity and biomass changes in
275 these zones^{31-33,45}, here we project shifts of their boundaries and maintain their climatic
276 label. However, the present approach does not account for putative changes in community
277 composition nor the emergence of novel niches over these regions for which novel
278 environmental selection pressure is expected.

279 Overall, our projections for the end of the century do not take into account possible future
280 changes of major bio-physico-chemical factors such as dynamics of community mixing,
281 trophic interactions through transport⁴⁶, the potential dynamics of the genomes^{14,16,17}
282 (adaptation or acclimation) and biomass variations⁴⁵. New sampling in current and future
283 expeditions⁴⁷, as well as ongoing technological improvements in bio-physico-chemical
284 characterization of seawater samples^{38,47,48} will soon refine functional^{18,49}, environmental
285 (micronutrients⁵⁰) and phylogenetic^{16,17} characterization of plankton ecosystems for
286 various biological entities (genotypes, species or communities) and spatio-temporal
287 scales⁴⁷. Ultimately, crossing this varied information will allow a better understanding of
288 the conditions of emergence of ecological niches in the seascape and their response to a
289 changing ocean.

290 **References**

- 291 1. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the
292 biosphere: Integrating terrestrial and oceanic components. *Science* (80-.). (1998).
293 doi:10.1126/science.281.5374.237
- 294 2. Guidi, L. *et al.* Plankton networks driving carbon export in the oligotrophic ocean. *Nature*
295 (2016). doi:10.1038/nature16942
- 296 3. Henson, S. A., Sanders, R. & Madsen, E. Global patterns in efficiency of particulate organic
297 carbon export and transfer to the deep ocean. *Global Biogeochem. Cycles* (2012).
298 doi:10.1029/2011GB004099
- 299 4. Azam, F. *et al.* The Ecological Role of Water-Column Microbes in the Sea. *Mar. Ecol. Prog. Ser.*
300 (1983). doi:10.3354/meps010257
- 301 5. Saab, M. A. abi. Day-to-day variation in phytoplankton assemblages during spring blooming
302 in a fixed station along the Lebanese coastline. *J. Plankton Res.* (1992).
303 doi:10.1093/plankt/14.8.1099
- 304 6. Djurhuus, A. *et al.* Environmental DNA reveals seasonal shifts and potential interactions in a
305 marine community. *Nat. Commun.* (2020). doi:10.1038/s41467-019-14105-1
- 306 7. Kavanaugh, M. T. *et al.* Seascapes as a new vernacular for pelagic ocean monitoring,
307 management and conservation. *ICES J. Mar. Sci.* (2016). doi:10.1093/icesjms/fsw086
- 308 8. Steele, J. H. The ocean 'landscape'. *Landsc. Ecol.* (1989). doi:10.1007/BF00131537
- 309 9. Longhurst, A. R. *Ecological Geography of the Sea. Ecological Geography of the Sea* (2007).
310 doi:10.1016/B978-0-12-455521-1.X5000-1

- 311 10. Fay, A. R. & McKinley, G. A. Global open-ocean biomes: Mean and temporal variability. *Earth*
312 *Syst. Sci. Data* (2014). doi:10.5194/essd-6-273-2014
- 313 11. Reygondeau, G. *et al.* Dynamic biogeochemical provinces in the global ocean. *Global*
314 *Biogeochem. Cycles* (2013). doi:10.1002/gbc.20089
- 315 12. Richter, D. J. *et al.* Genomic evidence for global ocean plankton biogeography shaped by
316 large-scale current systems. *bioRxiv* 867739 (2019). doi:10.1101/867739
- 317 13. Dutkiewicz, S. *et al.* Dimensions of marine phytoplankton diversity. *Biogeosciences* (2020).
318 doi:10.5194/bg-17-609-2020
- 319 14. Hellweger, F. L., Van Sebille, E. & Fredrick, N. D. Biogeographic patterns in ocean microbes
320 emerge in a neutral agent-based model. *Science* (80-.). (2014). doi:10.1126/science.1254421
- 321 15. Sauterey, B., Ward, B., Rault, J., Bowler, C. & Claessen, D. The Implications of Eco-Evolutionary
322 Processes for the Emergence of Marine Plankton Community Biogeography. *Am. Nat.* (2017).
323 doi:10.1086/692067
- 324 16. Laso-Jadart, R. *et al.* Investigating population-scale allelic differential expression in wild
325 populations of *Oithona similis* (Cyclopoida, Claus, 1866). *Ecol. Evol.* (2020).
326 doi:10.1002/ece3.6588
- 327 17. Delmont, T. O. *et al.* Single-amino acid variants reveal evolutionary processes that shape the
328 biogeography of a global SAR11 subclade. *Elife* (2019). doi:10.7554/eLife.46497
- 329 18. Carradec, Q. *et al.* A global ocean atlas of eukaryotic genes. *Nat. Commun.* (2018).
330 doi:10.1038/s41467-017-02342-1
- 331 19. Salazar, G. *et al.* Gene Expression Changes and Community Turnover Differentially Shape the
332 Global Ocean Metatranscriptome. *Cell* (2019). doi:10.1016/j.cell.2019.10.014
- 333 20. Alberti, A. *et al.* Viral to metazoan marine plankton nucleotide sequences from the Tara
334 Oceans expedition. *Sci. Data* (2017). doi:10.1038/sdata.2017.93
- 335 21. Pesant, S. *et al.* Open science resources for the discovery and analysis of Tara Oceans data.
336 *Sci. Data* (2015). doi:10.1038/sdata.2015.23
- 337 22. Karsenti, E. *et al.* A holistic approach to marine Eco-systems biology. *PLoS Biol.* (2011).
338 doi:10.1371/journal.pbio.1001177
- 339 23. Duarte, C. M. Seafaring in the 21st century: The Malaspina 2010 circumnavigation
340 expedition. *Limnology and Oceanography Bulletin* (2015). doi:10.1002/lob.10008
- 341 24. Barton, A. D., Irwin, A. J., Finkel, Z. V. & Stock, C. A. Anthropogenic climate change drives shift
342 and shuffle in North Atlantic phytoplankton communities. *Proc. Natl. Acad. Sci.* (2016).
343 doi:10.1073/pnas.1519080113

- 344 25. Benedetti, F., Guilhaumon, F., Adloff, F. & Ayata, S. D. Investigating uncertainties in
345 zooplankton composition shifts under climate change scenarios in the Mediterranean Sea.
346 *Ecography (Cop.)*. (2018). doi:10.1111/ecog.02434
- 347 26. Beaugrand, G. *et al.* Prediction of unprecedented biological shifts in the global ocean. *Nature*
348 *Climate Change* (2019). doi:10.1038/s41558-019-0420-1
- 349 27. McMahon, K. W., McCarthy, M. D., Sherwood, O. A., Larsen, T. & Guilderson, T. P. Millennial-
350 scale plankton regime shifts in the subtropical North Pacific Ocean. *Science (80-.)*. (2015).
351 doi:10.1126/science.aaa9942
- 352 28. Rivero-Calle, S., Gnanadesikan, A., Del Castillo, C. E., Balch, W. M. & Guikema, S. D.
353 Multidecadal increase in North Atlantic coccolithophores and the potential role of rising CO₂.
354 *Science (80-.)*. (2015). doi:10.1126/science.aaa8026
- 355 29. Beaugrand, G. Decadal changes in climate and ecosystems in the North Atlantic Ocean and
356 adjacent seas. *Deep. Res. Part II Top. Stud. Oceanogr.* (2009). doi:10.1016/j.dsr2.2008.12.022
- 357 30. Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L. & Levin, S. A. Marine taxa track local
358 climate velocities. *Science (80-.)*. (2013). doi:10.1126/science.1239352
- 359 31. Thomas, M. K., Kremer, C. T., Klausmeier, C. A. & Litchman, E. A global pattern of thermal
360 adaptation in marine phytoplankton. *Science (80-.)*. (2012). doi:10.1126/science.1224836
- 361 32. Busseni, G. *et al.* Large scale patterns of marine diatom richness: Drivers and trends in a
362 changing ocean. *Glob. Ecol. Biogeogr.* (2020). doi:10.1111/geb.13161
- 363 33. Ibarbalz, F. M. *et al.* Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell*
364 (2019). doi:10.1016/j.cell.2019.10.008
- 365 34. Bopp, L. *et al.* Multiple stressors of ocean ecosystems in the 21st century: Projections with
366 CMIP5 models. *Biogeosciences* (2013). doi:10.5194/bg-10-6225-2013
- 367 35. Hutchinson, G. E. Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.* (1957).
- 368 36. Jones, M. C. & Cheung, W. W. L. Multi-model ensemble projections of climate change effects
369 on global marine biodiversity. *ICES J. Mar. Sci.* (2015). doi:10.1093/icesjms/fsu172
- 370 37. Delmont, T. O. *et al.* Nitrogen-fixing populations of Planctomycetes and Proteobacteria are
371 abundant in surface ocean metagenomes. *Nat. Microbiol.* (2018). doi:10.1038/s41564-018-
372 0176-9
- 373 38. Delmont, T. O. *et al.* Functional repertoire convergence of distantly related eukaryotic
374 plankton lineages revealed by genome-resolved metagenomics. *bioRxiv* 2020.10.15.341214
375 (2020). doi:10.1101/2020.10.15.341214
- 376 39. Boyer, T. P. *et al.* WORLD OCEAN DATABASE 2013, NOAA Atlas NESDIS 72. *Sydney Levitus*,

- 377 *Ed.; Alexey Mishonoc, Tech. Ed. (2013). doi:10.7289/V5NZ85MT*
- 378 40. Sunagawa, S. *et al.* Tara Oceans: towards global ocean ecosystems biology. *Nature Reviews*
379 *Microbiology* (2020). doi:10.1038/s41579-020-0364-5
- 380 41. Moon, K. R. *et al.* Visualizing structure and transitions in high-dimensional biological data.
381 *Nat. Biotechnol.* (2019). doi:10.1038/s41587-019-0336-3
- 382 42. van Vuuren, D. P. *et al.* The representative concentration pathways: An overview. *Clim.*
383 *Change* (2011). doi:10.1007/s10584-011-0148-z
- 384 43. Marinov, I. *et al.* North-South asymmetry in the modeled phytoplankton community
385 response to climate change over the 21st century. *Global Biogeochem. Cycles* (2013).
386 doi:10.1002/2013GB004599
- 387 44. Polovina, J. J., Dunne, J. P., Woodworth, P. A. & Howell, E. A. Projected expansion of the
388 subtropical biome and contraction of the temperate and equatorial upwelling biomes in the
389 North Pacific under global warming. *ICES J. Mar. Sci.* (2011). doi:10.1093/icesjms/fsq198
- 390 45. Flombaum, P., Wang, W. L., Primeau, F. W. & Martiny, A. C. Global picophytoplankton niche
391 partitioning predicts overall positive response to ocean warming. *Nat. Geosci.* (2020).
392 doi:10.1038/s41561-019-0524-2
- 393 46. Iudicone, D. Some may like it hot. *Nature Geoscience* (2020). doi:10.1038/s41561-020-0535-
394 z
- 395 47. Gorsky, G. *et al.* Expanding Tara Oceans Protocols for Underway, Ecosystemic Sampling of
396 the Ocean-Atmosphere Interface During Tara Pacific Expedition (2016–2018). *Front. Mar.*
397 *Sci.* (2019). doi:10.3389/fmars.2019.00750
- 398 48. Istace, B. *et al.* de novo assembly and population genomic survey of natural yeast isolates
399 with the Oxford Nanopore MinION sequencer. *Gigascience* (2017).
400 doi:10.1093/gigascience/giw018
- 401 49. Busseni, G. *et al.* Meta-Omics Reveals Genetic Flexibility of Diatom Nitrogen Transporters in
402 Response to Environmental Changes. *Mol. Biol. Evol.* (2019). doi:10.1093/molbev/msz157
- 403 50. Grand, M. M. *et al.* Developing autonomous observing systems for micronutrient trace
404 metals. *Frontiers in Marine Science* (2019). doi:10.3389/fmars.2019.00035

405 **Acknowledgments**

406 PF was supported by a CFR doctoral fellowship and the *NEOGEN impulsion* grant from the
407 Direction de la recherche fondamentale (DRF) of the CEA. We thank the LSCE (Laboratoire
408 des sciences du climat et de l'environnement, CEA) for providing Earth System Models
409 outputs, Tilla Roy for preparation of the data, LAGE (Laboratoire d'Analyses Génomiques

410 des Eucaryotes, CEA) members for stimulating discussions on this project, Mahendra
411 Mariadassou, Sakina Dorothee Ayata and Bruno Hay Mele for discussions on statistics and
412 climate envelop models, Laurent Bopp for initial discussions on this project and on climate
413 models, Tom Delmont for providing Metagenome-Assembled Genomes data and Noan Le
414 Bescot (TernogDesign) for artwork on Figures. We thank all members of the *Tara Oceans*
415 consortium for maintaining a creative environment and for their constructive criticism.
416 *Tara Oceans* would not exist without the *Tara Ocean Foundation* and the continuous
417 support of 23 institutes (<https://oceans.taraexpeditions.org/>).

418 This article is contribution number XX of *Tara Oceans*.

419 **Competing interests**

420 The authors declare no competing interests.

421 **Author contributions**

422 PF, OJ and MG conceived the study. MV wrote the bias correction algorithm. PF computed
423 the results, compiled and analyzed the data. PF wrote the initial draft of the paper. JL, OJ
424 and MG conducted a preliminary study. PF, OJ, MG, MV, DI, and PW discussed the results
425 and contributed to write the paper.

426 **Online content**

427 Supplementary information, additional references, source data and codes are available at
428 www.doi.xx.com/

429 **Materials and methods**

430 **Genomic provinces of plankton**

431 Environmental niches are computed for trans-kingdom plankton genomic provinces from
432 Richter et al.¹². They consist of the clustering of metagenomic dissimilarity matrices from 6
433 available size fractions with sufficient metagenomic data from the *Tara Oceans* dataset. The
434 six size fractions (0-0.2, 0.22-3, 0.8-5, 5-20, 20-180 and 180-2000 μm) represent major
435 plankton groups. Two large size classes (180-2000 μm and 20-180 μm) are enriched in
436 zooplankton dominated by arthropods (mainly copepods) and cnidarians. They are
437 expected to directly depend on smaller eukaryotes as they feed on them. Size classes 5-20
438 μm and 0.8-5 μm are enriched in smaller eukaryotic algae, such as dinophytes (5-20 μm),

439 pelagophytes and haptophytes (0.8-5 μm). The distribution of these photoautotrophs
440 presumably depends on nutrient availability. Finally, size classes 0.22-3 μm and 0-0.2 μm
441 are respectively enriched in bacteria and viruses. Bacteria are characterized by a wide
442 range of trophisms including autotrophy (cyanobacteria), mixotrophy and heterotrophy,
443 while viruses are mainly parasites. Within each size fraction (from large to small), there are
444 respectively 8, 8, 11, 6, 6 and 8 (48 in total) provinces defined in Richter et al.¹² formed by
445 *Tara* Oceans metagenomes (644 metagenomes sampled either at the surface (SUR) or at
446 the Deep Chlorophyll Maximum (DCM) across 102 sites). The clustering of individual size
447 fractions is independent.

448 **Genome signature of the provinces**

449 We analyzed the distribution of 713 eukaryotic and 523 prokaryotic genomes^{37,38} within
450 the genomic provinces. These genomes are Metagenome-Assembled Genomes (MAGs)
451 obtained from *Tara* Oceans metagenomes. For each size class, we select MAGs that are
452 present (according to a criteria defined in Delmont et al.³⁸) in at least 5 samples. We
453 computed an index of presence enrichment of MAGs within provinces as the Jaccard
454 index⁵¹, defined as follows:

$$J = \frac{M_{11}}{M_{11} + M_{01} + M_{10}}$$

455
456 M_{11} is the number of samples where the MAG is present and match a sample of the
457 province. M_{01} and M_{10} are respectively the number of samples where the MAG is not
458 present in a sample of the province and inversely. A MAG is considered to be signature of a
459 province if the Jaccard index is superior to 0.5 with this province and inferior to 0.1 for all
460 other provinces of the given size class (Fig. 1 and Supplementary Fig. 3).

461 **World Ocean Atlas data**

462 Physicochemical parameters proposed to have an impact on plankton genomic provinces¹²
463 are used to define environmental niches: sea surface temperature (SST), salinity (Sal);
464 dissolved silica (Si), nitrate (NO_3), phosphate (PO_4), iron (Fe), and a seasonality index of
465 nitrate (SI NO_3). With the exception of Fe and SI NO_3 , these parameters are extracted from
466 the gridded World Ocean Atlas 2013 (WOA13)³⁹. Climatological Fe fields are provided by
467 the biogeochemical model PISCES-v2⁵². The seasonality index of nitrate is defined as the

468 range of nitrate concentration in one grid cell divided by the maximum range encountered
469 in WOA13 at the *Tara* Oceans sampling stations. All parameters are co-located with the
470 corresponding stations and extracted at the month corresponding to the *Tara* Oceans
471 sampling. To compensate for missing physicochemical samples in the *Tara* Oceans *in situ*
472 data set, climatological data (WOA) are preferred. The correlation between *in situ* samples
473 and corresponding values extracted from WOA are high (r^2 : SST: 0.96, Sal: 0.83, Si: 0.97,
474 NO_3 : 0.83, PO_4 : 0.89). In the absence of corresponding WOA data, a search is done within 2°
475 around the sampling location and values found within this square are averaged.
476 Nutrients, such as NO_3 and PO_4 , display a strong collinearity when averaged over the global
477 ocean (correlation of 0.95 in WOA13) which could complicate disentangling their
478 respective contributions to niche definition. However, observations and experimental data
479 allow identification of limiting nutrients at regional scale characterized by specific plankton
480 communities⁵³. The projection of niches into future climate would yield spurious results
481 when the present-day collinearity is not maintained⁵⁴ but there is up to now no evidence
482 for large scale changes in global nutrient stoichiometry⁵⁵.

483 **Earth System Models and bias correction**

484 Outputs from 6 Earth System Models (ESM) (Supplementary Table 2) are used to project
485 environmental niches under greenhouse gas emission scenario RCP8.5⁴². Environmental
486 drivers are extracted for present day (2006-2015) and end of century (2090-2099)
487 conditions for each model and the multi-model mean is computed. A bias correction
488 method, the Cumulative Distribution Function transform, CDFt⁵⁶, is applied to adjust the
489 distributions of SST, Sal, Si, NO_3 and PO_4 of the multi-model mean to the WOA database.
490 CDFt is based on a quantile mapping (QM) approach to reduce the bias between modeled
491 and observed data, while accounting for climate change. Therefore, CDFt does not rely on
492 the stationarity hypothesis and present and future distributions can be different. CDFt is
493 applied on the global fields of the mean model simulations. By construction, CDFt preserves
494 the ranks of the simulations to be corrected. Thus, the spatial structures of the model fields
495 are preserved.

496 **Environmental niche models: training, validation and projections**

497 Provinces with similar metagenomic content are retrieved from Richter et al.¹². From a
498 total of 48 initial provinces, 10 provinces are removed either because they are represented
499 by too few samples (7 out of 10) or they are found in environments not resolved by ESMs
500 (e.g. lagoons of Pacific Ocean islands, 3 out of 10). Four machine learning methods are
501 applied to compute environmental niches for each of the 38 provinces: Gradient Boosting
502 Machine (gbm)⁵⁷, Random Forest (rf)⁵⁸, fully connected Neural Networks (nn)⁵⁹ and
503 Generalized Additive Models (gam)⁶⁰. Hyper parameters of each technique (except gam)
504 are optimized. These are (1) for gbm, the interaction depth (1, 3 and 5), learning rate (0.01,
505 0.001) and the minimum number of observations in a tree node (1 to 10); (2) for rf, the
506 number of trees (100 to 900 with step 200 and 1000 to 9000 with step 2000) and the
507 number of parameters used for each tree (1 to 7); (3) for nn, the number of layers of the
508 network (1 to 10) and the decay ($1 \cdot 10^{-4}$ to $9 \cdot 10^{-4}$ and $1 \cdot 10^{-5}$ to $9 \cdot 10^{-5}$). For gam the number
509 of splines is set to 3, respectively 2 only when not enough points are available (for fraction
510 0-0.2, 65 points). R packages gbm (2.1.3), randomForest (4.6.14), mgcv (1.8.16) and nnet
511 (7.3.12) are used for gbm, rf, nn and gam models.

512 To define the best combination of hyper parameters for each model, we perform 30
513 random cross-validations by training the model on 75% (85 % for gbm and gam) of the
514 dataset randomly sampled and by calculating the Area Under the Curve⁶¹ (AUC) on the
515 25% (15 % for gbm and gam) remaining points of the dataset. The best combination of
516 hyper parameters is the one for which the mean AUC over the 30 cross-validation is the
517 highest. A model is considered valid if at least 3 out of the 4 techniques have a mean AUC
518 superior to 0.65, which is the case for 27 out of the 38 provinces (Supplementary Fig. 2a). A
519 climatic annotation is given to the 27 validated niches (Supplementary Table 1). Final
520 models are trained on the full dataset and only the techniques that have a mean AUC higher
521 than 0.65 are considered to make the projections. The majority (23) of the 27 validated
522 niches is validated by all four models and 4 by only 3 models. Relative influences of each
523 parameter in defining environmental niches are calculated using the feature_importance
524 function from the DALEX R package⁶² for all four statistical methods (Supplementary Fig.
525 19a). To evaluate the consistency and coherence of environmental niche models, we first
526 make global projections on the 2006-13 WOA2013 climatology. Projections are consistent
527 with sampling regions for provinces encompassing vast oceanic areas. For example, the

528 genomic province sampled in temperate Atlantic regions of size fraction 180-2000 μm is
529 projected to be present in the north and south temperate Atlantic but also other temperate
530 regions (Supplementary Fig. 4). For model training and projections, physicochemical
531 variables are scaled to have a mean of 0 and a variance of 1. For this scaling, the mean and
532 standard deviation of each WOA13 variable (+ PISCES-v2 Fe) co-localized with *Tara* Oceans
533 stations with a value available are used. This standardization procedure allows for better
534 performance of nn models. Finally, as statistical models often disagree on projection sets
535 whereas they give similar predictions on the training set (Supplementary Fig. 5, 6), we use
536 the ensemble model approach for global-scale projections of provinces³⁶ *i.e.* the mean
537 projections of the validated machine learning techniques.

538 **Combined size classes provinces and ocean partitioning comparisons**

539 To combine all size classes' provinces, we use the PHATE algorithm^{41,63} from R package
540 phateR. This algorithm allows visualization of high dimensional data in the requested
541 number of dimensions while best preserving the global data structure⁶³. We choose to train
542 PHATE separately on WOA13 projections and present day and end of century projections
543 including presence probabilities of *non dominant* provinces. We use 3 dimensions and set
544 hyper parameter k-nearest neighbors (knn) and decay respectively to 1000 for WOA13 and
545 2000 for model data as in this case there are twice as many points. The hyper parameter
546 knn reflects the degree to which the mapping of PHATE from high to low dimensionality
547 should respect the global features of the data. We argue that 1000 and 2000 are good
548 choices as it will be sufficient to have a highly connected graph, conserve global structure,
549 allow visualization of structures of the size of the provinces (mean number of points in a
550 province: 4867) and have a reasonable computational time. Decay is set to 20 in both cases.
551 Then we cluster the resulting distance matrix using the k-medoids algorithm⁶⁴ and the
552 silhouette average width criteria⁶⁵ is used as an indicator of good fit. The silhouette
553 criterion is maximal for 2, 3 and 4 clusters and 2 peaks are found at 7 and 14 clusters (the
554 peak at 7 is slightly less high than the one at 14, data not shown). We choose to present the
555 4 and 7 clusters geographical patterns as they seem more relevant with respect to the
556 resolutions of our environmental datasets (WOA13 and climate models). We compare the
557 three polar clusters of the 7 cluster geographical patterns with Antarctic Circumpolar
558 Currents fronts⁶⁶ by overlying them on the map (red lines Fig. 2h).

559 To visualize the global biogeography structure, the resulting 3 vectors of PHATE are
560 plotted using an RGB color code. Each coordinate of each vector is respectively assigned to
561 a given degree of color component between 0 and 255 (8 bits red, green or blue) using the
562 following formula (Fig. 1 g,h; Fig 2 e,f; Supplementary Fig. 7):

$$C_{col}(i) = \frac{C(i) - \min(C(1), C(2), C(3))}{\max(C(1), C(2), C(3)) - \min(C(1), C(2), C(3))} * 255$$

563
564 $C(i)$ is the i th component of the PHATE axes. Respectively, components 1, 2 and 3 are
565 assigned to red, green and blue.

566 To compare the six size fraction provinces, the combined size class with existing
567 biogeochemical partitions of the oceans^{10,11} and with each other, we use the adjusted rand
568 index⁶⁷ (Supplementary Fig. 7-9) and overlay their masks above our partitions. In this case,
569 presence probabilities of *dominant* provinces are not used anymore. Instead, each ocean
570 grid point is assigned to the *dominant* provinces or the phate clusters.

571 **Bray-Curtis dissimilarity index**

572 Climate change impact on global projections is calculated at each grid point as the Bray-
573 Curtis dissimilarity index^{68,69} defined as follows:

$$BC = \frac{\sum_n |P_n^{future} - P_n^{present}|}{\sum_n |P_n^{future} + P_n^{present}|}$$

574
575 Where ($P_n^{present}$ and P_n^{future}) are respectively the probability of presence of the province n
576 in present day and at the end of the century. Only the probabilities of *dominant* provinces
577 are non-null and all others are set to zero. The mask of main fisheries⁷⁰ (chosen as the first
578 4 deciles) and Exclusive Economical Zones⁷¹ is overlaid on the Bray-Curtis map.

579 **Change in province assemblages**

580 A province assemblage is defined as the assemblage of *dominant* provinces of each size
581 fraction at a given grid point of the considered ocean. We consider two criteria of change in
582 province assemblage between present day and end of the century conditions. The first one,
583 more straightforward and less stringent, considers that a province assemblage occurs
584 when a change of *dominant* province is found in at least one size fraction. In a more
585 stringent way, a change of assemblage is considered significant for $BC > \frac{1}{6}$ (previous
586 section). This threshold corresponds to an idealized case where each *dominant* province

587 has a probability of one and a change of *dominant* province is found in only one size
588 fraction. For example, the *dominant* province assemblage goes from vector
589 (F5,E6,D3,C8,B7,A7) (with the size fractions in decreasing order) corresponding to all
590 temperate provinces to vector (F8,E6,D3,C8,B7,A7). This example corresponds to the
591 replacement of the temperate province of size fraction 180-2000 μm (F5) by the tropico-
592 equatorial province (F8). This criterion allows us to discard assemblage changes for which
593 the changes in probability of presence of *dominant* provinces are very low. With this
594 criterion, only on a small oceanic area is found to have no changes of assemblage (Fig. 4c
595 light blue zones).

596 **Centroids and migration shifts**

597 The centroid of each province is defined as the average latitude and longitude for which the
598 probability of presence is superior to 0.5 and weighted by both the probability of presence
599 at each grid point and the grid cell area. It is calculated for both present day conditions and
600 end of the century conditions. The migration shift is calculated as the distance between the
601 present day and the end of the century centroids considering the earth as a perfect sphere
602 of radius 6371 km. For consistency (*i.e.* avoid long distance aberrant shifts), it is only
603 calculated for provinces with an area of dominance larger than 10^6 km^2 in the given basin.

604 **Driver analysis**

605 To assess the relative importance of each driver in province changes, the methodology
606 from Barton et al.²⁴ is adopted. For a set N of n provinces (individual provinces or all
607 provinces together), the probability of presence of each province is recalculated for present
608 day conditions except for driver d (from the set of drivers D) for which end of the century
609 condition is used ($P_n^{\text{future for } d\text{th driver only}}$). The set of driver D can be either all drivers
610 (Fig. 5a,c) or all drivers except SST (Fig. 5b). The relative importance of driver d at a given
611 grid point for the set N of provinces is computed as follow:

$$RI(d) = \frac{\sum_{n \in N} |P_n^{\text{future for } d\text{th driver only}} - P_n^{\text{present}}|}{\sum_{d \in D} \sum_{n \in N} |P_n^{\text{future for } d\text{th driver only}} - P_n^{\text{present}}|}$$

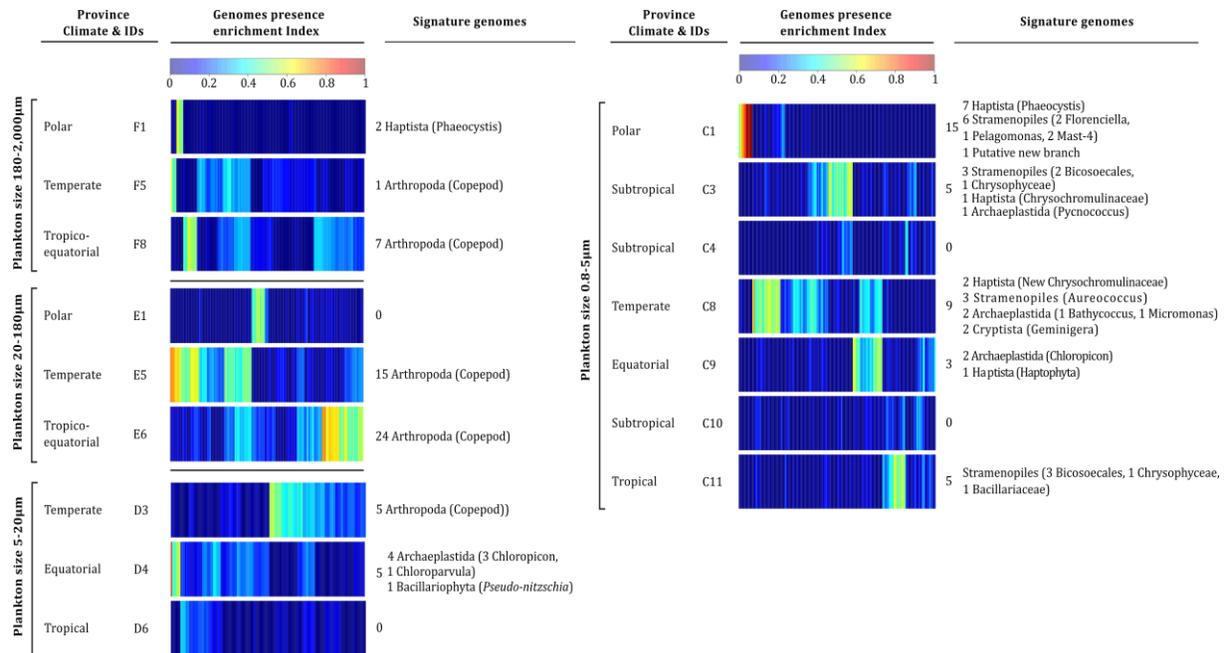
612
613 $RI(d)$ is computed at grid cells where $BC > \frac{1}{6}$, BC being calculated with either the set of all
614 drivers (Fig. 5a,c) or all drivers except SST (Fig. 5b). When $RI(d)$ is calculated for individual
615 provinces (Fig. 5c and Supplementary Fig. 16d,e), it is computed only at grid cells where

616 $BC > \frac{1}{6}$ and the concerned province is either *dominant* in present day and/or end of
617 century conditions.

618 **References**

- 619 51. Jaccard, P. Distribution comparée de la flore alpine dans quelques régions des Alpes
620 occidentales et orientales. *Bull. la Murithienne* (1902).
- 621 52. Aumont, O., Ethé, C., Tagliabue, A., Bopp, L. & Gehlen, M. PISCES-v2: An ocean biogeochemical
622 model for carbon and ecosystem studies. *Geosci. Model Dev.* (2015). doi:10.5194/gmd-8-
623 2465-2015
- 624 53. Moore, C. M. *et al.* Processes and patterns of oceanic nutrient limitation. *Nature Geoscience*
625 (2013). doi:10.1038/ngeo1765
- 626 54. Brun, P., Kiørboe, T., Licandro, P. & Payne, M. R. The predictive skill of species distribution
627 models for plankton in a changing climate. *Glob. Chang. Biol.* (2016). doi:10.1111/gcb.13274
- 628 55. Redfield, A. C. On the Proportions of Organic Derivatives in Sea Water and Their Relation to
629 the Composition of Plankton. in *James Johnstone Memorial Volume 1767–192* (Liverpool
630 Univ. Press, Liverpool, U.K., 1934).
- 631 56. Michelangeli, P. A., Vrac, M. & Loukos, H. Probabilistic downscaling approaches: Application
632 to wind cumulative distribution functions. *Geophys. Res. Lett.* (2009).
633 doi:10.1029/2009GL038401
- 634 57. Ridgeway, G. gbm: Generalized Boosted Regression Models. *R Packag. version 1.6-3.1* (2010).
- 635 58. Breiman, L. & Cutler, A. Breiman and Cutler's random forests for classification and
636 regression. *Packag. 'randomForest'* (2012). doi:10.5244/C.22.54
- 637 59. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S Fourth edition by. World*
638 (2002). doi:10.2307/2685660
- 639 60. Wood, S. N. Stable and efficient multiple smoothing parameter estimation for generalized
640 additive models. *J. Am. Stat. Assoc.* (2004). doi:10.1198/016214504000000980
- 641 61. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* (2006).
642 doi:10.1016/j.patrec.2005.10.010
- 643 62. Biecek, P. DALEX: explainers for complex predictive models. *J. Mach. Learn. Res.* **19**, 1–5
644 (2018).
- 645 63. Vallejos, C. A. Exploring a world of a thousand dimensions. *Nature Biotechnology* (2019).
646 doi:10.1038/s41587-019-0330-9
- 647 64. L., K. & P., R. Clustering by means of Medoids. in *Statistical Data Analysis Based on the L1*
648 *Norm and Related Methods* (1987).

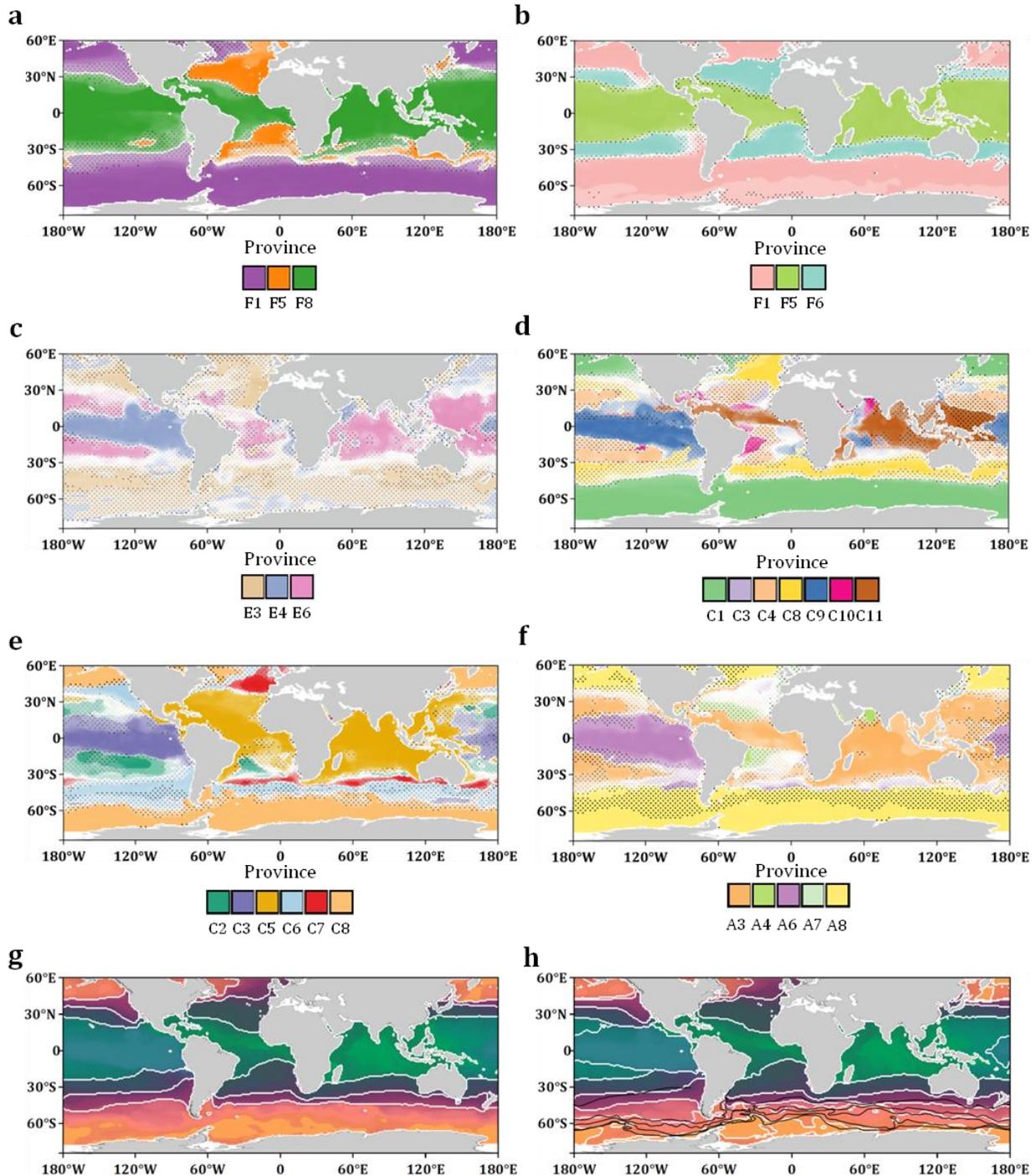
- 649 65. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster
650 analysis. *J. Comput. Appl. Math.* (1987). doi:10.1016/0377-0427(87)90125-7
- 651 66. Orsi, A. H., Whitworth, T. & Nowlin, W. D. On the meridional extent and fronts of the Antarctic
652 Circumpolar Current. *Deep. Res. Part I* (1995). doi:10.1016/0967-0637(95)00021-W
- 653 67. Hubert, L. & Arabie, P. Comparing partitions. *J. Classif.* (1985). doi:10.1007/BF01908075
- 654 68. Somerfield, P. J. Identification of the Bray-Curtis similarity index: Comment on Yoshioka
655 (2008). *Marine Ecology Progress Series* (2008). doi:10.3354/meps07841
- 656 69. Bloom, S. Similarity Indices in Community Studies: Potential Pitfalls. *Mar. Ecol. Prog. Ser.*
657 (1981). doi:10.3354/meps005125
- 658 70. Watson, R. A. A database of global marine commercial, small-scale, illegal and unreported
659 fisheries catch 1950-2014. *Sci. Data* (2017). doi:10.1038/sdata.2017.39
- 660 71. Flanders Marine Institute (2018). Maritime Boundaries Geodatabase: Maritime Boundaries
661 and Exclusive Economic Zones (200NM), version 10. (2018).
662 doi:<https://doi.org/10.14284/313>.
- 663



664

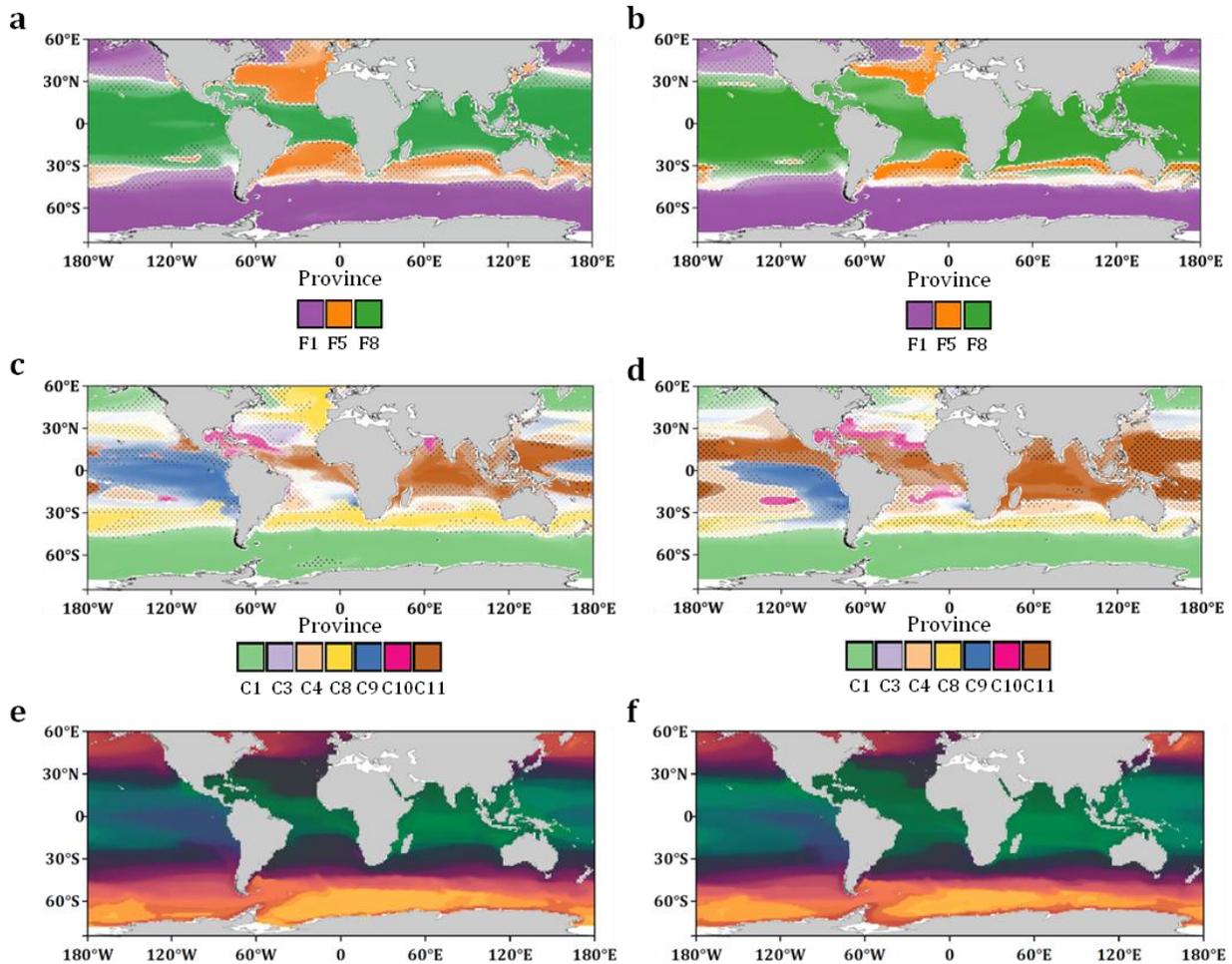
665 **Fig. 1 | Signature genomes of provinces of eukaryotes enriched size classes.** For each plankton
 666 size class, indexes of presence enrichment for 713 genomes of eukaryotic plankton³⁸ in
 667 corresponding provinces are clustered and represented in a color scale. Signature genomes
 668 (see *Methods*) are found for almost all provinces, their number and taxonomies are summarized
 669 (detailed list in Supplementary Table 6).

670



671 **Fig. 2 | Global biogeographies of size-structured plankton provinces projected on WOA2013**
 672 **dataset.** (a) Metazoans enriched (180-2000 μm) (b) Small metazoans enriched (20-180 μm) (c)
 673 Small eukaryotes enriched (5-20 μm) (d) Small eukaryotes enriched (0.8-5 μm) (e) Bacteria
 674 enriched (0.22-3 μm) (f) Viruses enriched (0-0.2 μm). (a-f) Dotted areas represent uncertainty
 675 areas where the delta of presence probability of the dominant province and an other (from the
 676 same size fraction) is inferior to 0.5. (g) Combined size class biogeography using PHATE algorithm

677 partitioned in 4 clusters. Areas of uncertainty are highlighted with dotted lines. **(h)** Combined size
678 class biogeography using PHATE algorithm partitioned in 7 clusters overlaid with Antarctic
679 Circumpolar Current boundaries (red). Simple biogeographies are observed in large size fractions
680 ($>20 \mu\text{m}$) with a partitioning in three major oceanic areas: tropico-equatorial, temperate and polar.
681 More complex geographic patterns and patchiness are observed in smaller size fractions with the
682 distinction of pacific equatorial provinces and provinces associated to oligotrophic tropical gyres.
683



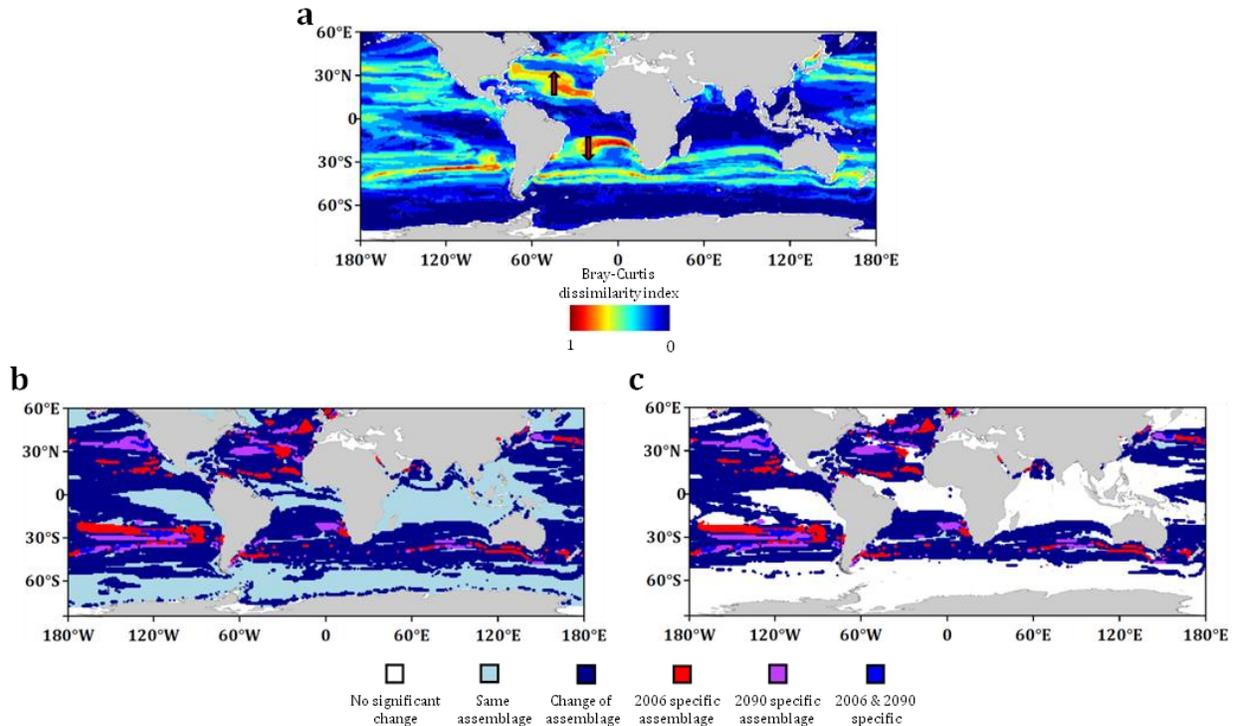
684 **Fig. 3 | Global biogeography of small metazoans enriched size fraction (180-2000 μm),**
685 **protists enriched (0.8-5 μm) and combined size classes in modeled present day (a, c, e) and**
686 **end of the century (b, d, f). (a-d)The dominant niche *i.e.* the one predicted to have the highest**
687 **probability of presence is represented at each grid point of the map. The color transparency is the**
688 **probability of presence of the dominant niche. A simple biogeography is observed for size fraction**
689 **180-2000 μm (a, b) with a polar niche, a temperate and a tropico-equatorial niche. Biogeography of**
690 **size fraction 0.8-5 μm size fraction is more complex and patchy with several temperate and tropical**
691 **niches (c, d). Biogeography of large size plankton and small size plankton are therefore decoupled.**
692 **Climate change impacts tropical niches which are expanding towards the poles in both size**
693 **fractions with a more complex behavior in size fraction 0.8-5 μm (b and d). (e, f) Present and end of**
694 **century combined size classes using PHATE algorithm.**
695

696

Size fraction (μm)	Present day covered area (%)	End of century covered area (%)	% area with a change of dominant province	Most frequent transition	%	2 nd most frequent transition	%
180-2000	74	74	13	temperate->tropico-equatorial (5->8)	67	polar->temperate (1->5)	14
20-180	78	77	12	temperate->tropico-equatorial (6->5)	67	polar->temperate (1->6)	29
5-20	45	49	22	temperate -> equatorial (3->4)	47	equatorial -> tropical (4->6)	25
0.8-5	56	59	31	equatorial -> tropical (9->11)	22	temperate -> equatorial (8->9)	15
0.22-3	60	61	15	temperate-> tropical (7->5)	22	polar -> temperate (8->6)	16
0-0.2	64	66	16	equatorial -> tropical (6->3)	32	temperate -> equatorial (8->6)	31
Total			60				

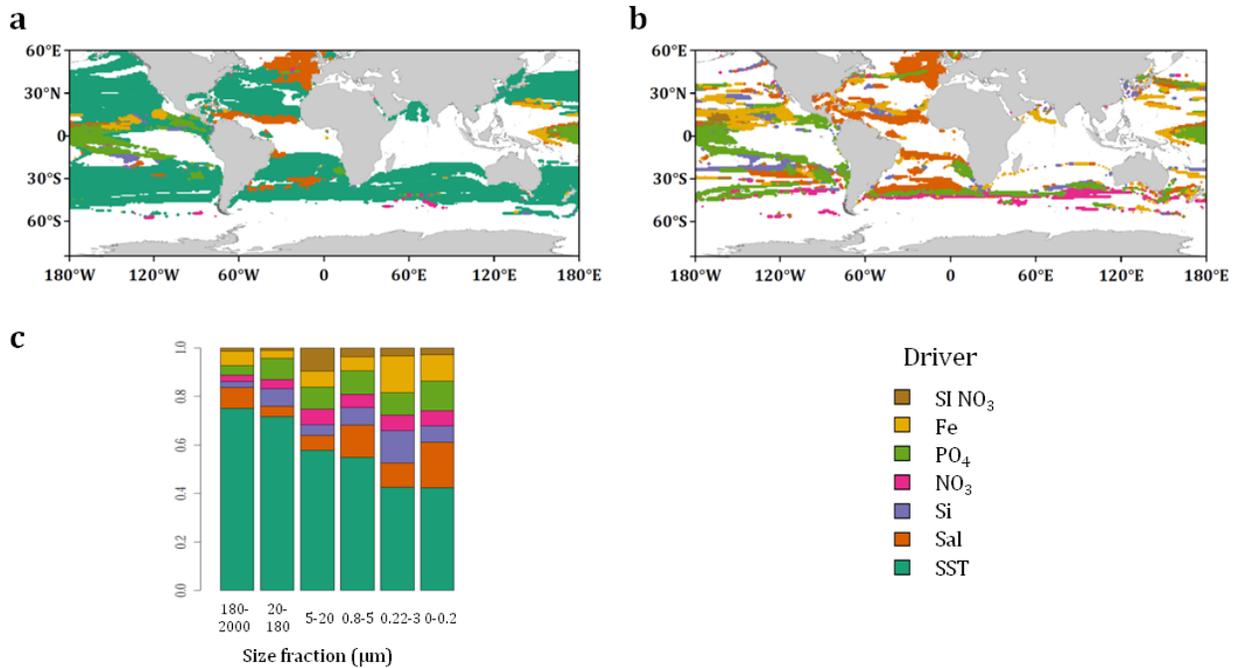
697 **Table 1 | Global statistics of covered areas and provinces' changes and transitions.** From 12 %
698 to 31% of the total covered area is estimated to be replaced by a different province at the end of the
699 century compared to present day depending on the size fraction. In total, considering all size
700 fractions this represents 60 % of the total covered area with at least one predicted change of
701 dominant province across the six size fractions.

702

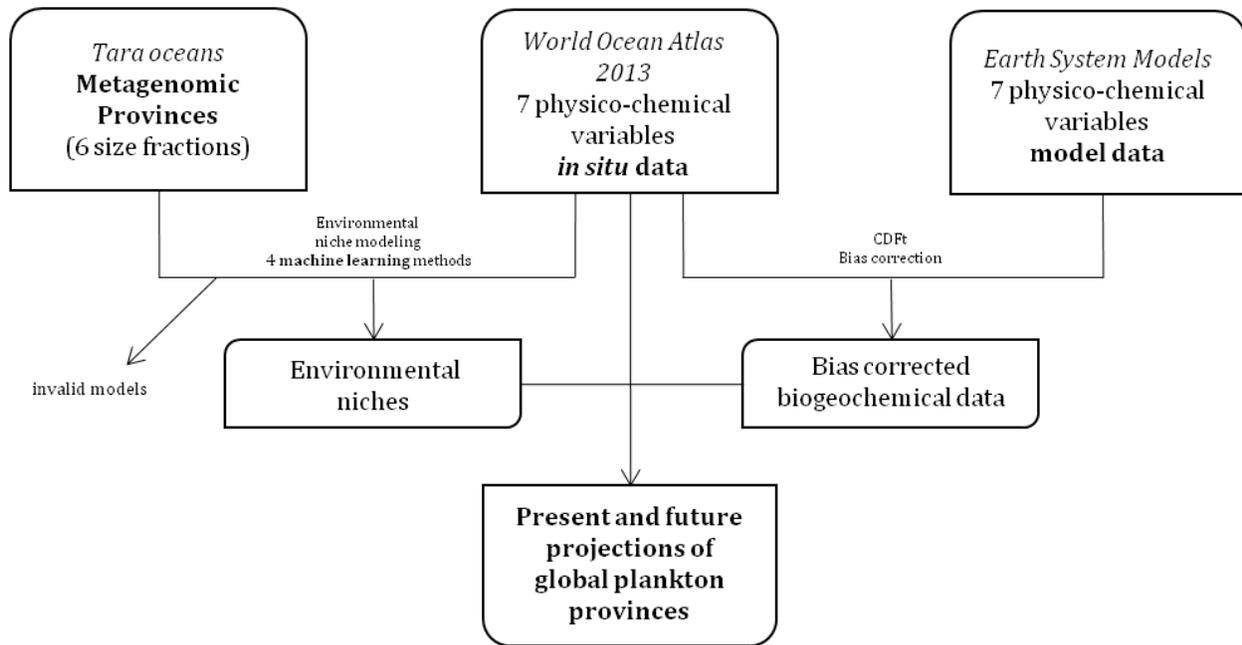


703 **Fig. 4 | (a) Bray-Curtis dissimilarity index map comparing present day with end of the**
704 **century projections of dominant provinces. Maps of trans-kingdom assemblages**
705 **reorganization of dominant provinces (b) and with a criterion of significance (c). (a) Bray-**
706 **Curtis dissimilarity index is calculated by integrating all the dominant provinces presence**
707 **probabilities over the six size fraction. Most important changes appear in subtropical, temperate**
708 **and subpolar regions. These changes are due to the displacement of tropical and temperate**
709 **provinces towards the pole but also the geographical decoupling between large and small size**
710 **plankton. The mean change in niche dissimilarity index is 0.25. (b) An assemblage is the combined**
711 **projected presence of the dominant province of each size class. Assemblage reorganization (present**
712 **day versus end of the century) is either mapped on all considered oceans or with a criterion on the**
713 **Bray-Curtis dissimilarity index ($BC > 1/6$, see *Materials and Methods*) (c). Depending on the**
714 **criterion from 60.1 % (b, dark blue) to 48.7 % (c, dark blue) of the oceanic area is projected to**
715 **change of assemblage. New assemblage are expected to appear in 2090 (purple+blue) whereas**
716 **some 2006 specific assemblages are projected to disappear (red+blue). New assemblages as well as**
717 **lost assemblages are mostly found in temperate, subtropical and tropical regions where most of the**
718 **rearrangements are projected.**

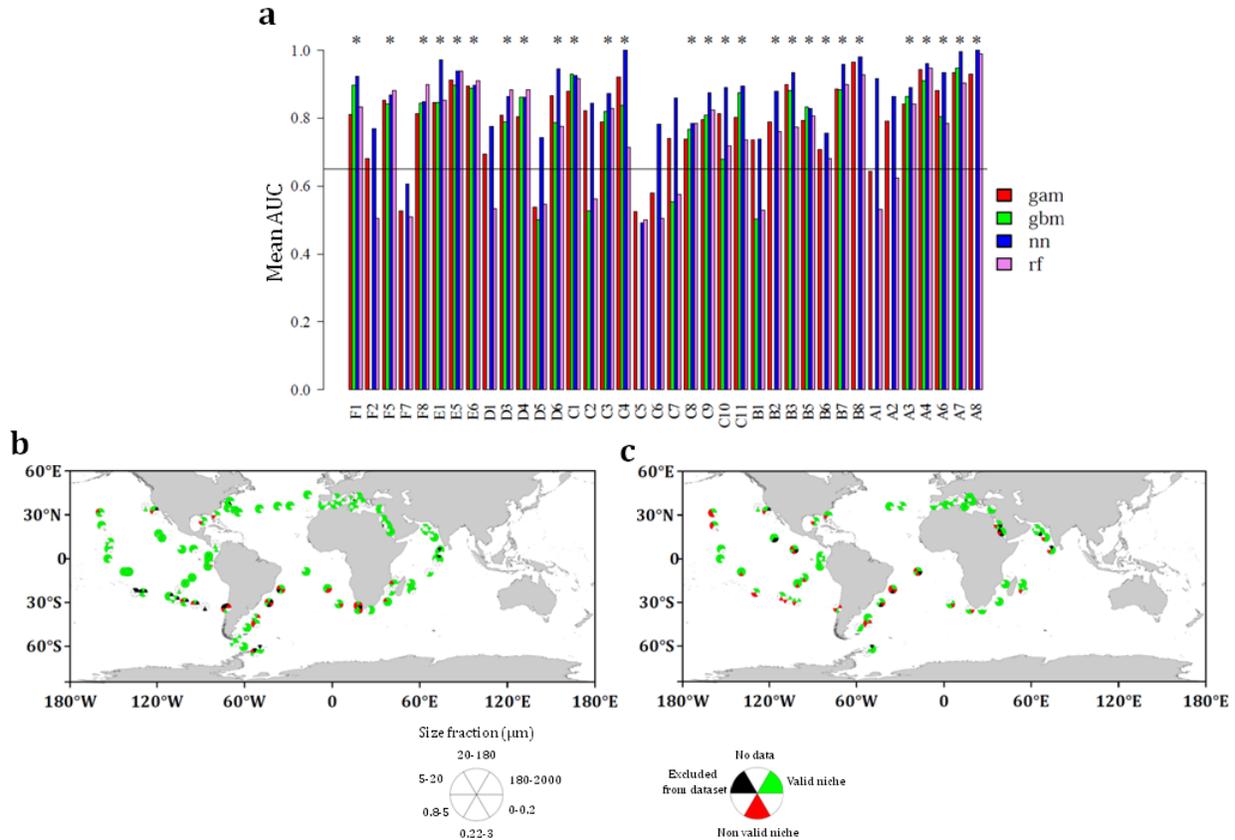
719



720 **Fig. 5 | Map of most impacting drivers on dominant province changes (a), most impacting**
721 **driver without considering temperature change (b) and relative importance of the drivers in**
722 **the different size fractions (c).** (a) Temperature appears as the top impacting driver on the
723 majority of the projected ocean with a significant change of province (Fig. 4). (b) Salinity and
724 dissolved phosphate are found to be the second and third driver of province reorganization notably
725 at tropical and subpolar latitudes. Note the importance of nitrate at temperate southern latitudes.
726 (c) Temperature is found to be the most important driver for all size classes but has a more
727 important impact in large size classes (>20 μm). Nutrients have on average a relative more
728 important impact in small size classes in driving provinces reorganization.

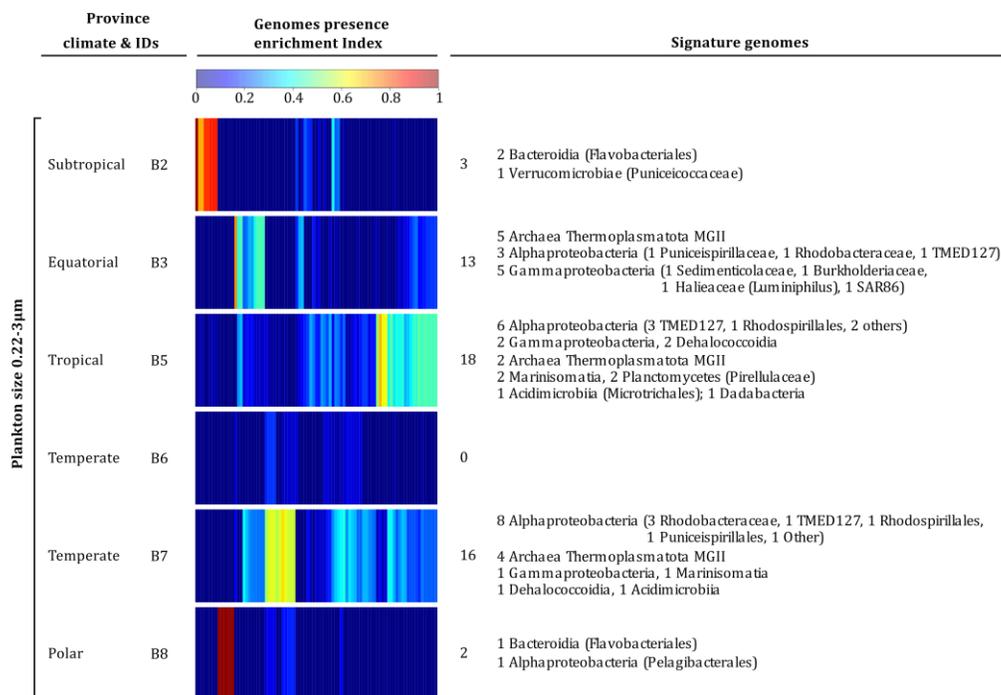


729
730 **Supplementary Fig. 1 | Study pipeline.** Metagenomic data from the 2009-2013 *Tara Oceans*
731 expedition and *in situ* measurements of physicochemical variables (*World Ocean Atlas 2013*,
732 WOA13)¹ are combined to define environmental niches at the plankton community level across 6
733 size fractions of the plankton realm. Bias corrected outputs from a mean model of 6 Earth System
734 Models²⁻⁷ and WOA13 data are then used to project global plankton provinces for present day
735 conditions and end of the century conditions under a high emission climate change scenario
736 (RCP8.5)⁸. Physico-chemical variables are Sea Surface Temperature (SST), Salinity (Sal), Dissolved
737 silica (Si), Nitrate (NO₃), Phosphate (PO₄), Iron (Fe) and a seasonality index of nitrate (SI NO₃).



738 **Supplementary Fig. 2 | (a) Barplot of mean AUC over 30-fold cross validation process of the**
 739 **38 initial metagenomic provinces. (b,c) Map of validated and non validated stations across**
 740 **the six size fractions in surface samples (b) and DCM samples (c). (a) Mean AUC (Area Under**
 741 **the receiver operating Characteristic)⁹ is plotted for the best hyperparameter combination of the 4**
 742 **machine learning techniques used for each of the 38 metagenomic provinces. General Additive**
 743 **Models¹⁰ (gam) are shown in red (no optimization), Gradient Boosting Machines¹¹ (gbm) in green,**
 744 **fully connected Neural Networks¹² (nn) in blue and Random Forest¹³ models in purple (rf). A star**
 745 **for valid models is drawn at the top of each considered valid model. A model is validated when at**
 746 **least 3 out of the four models have a mean AUC superior to 0.65. A valid model is found for 27 out of**
 747 **38 initial provinces. Out of the 27 validated models 4 are not valid for the gbm method. (b-c) For**
 748 **each Tara sample present in the dataset at surface (b) or Deep Chlorophyll Maximum (DCM) (c) and**
 749 **for each size fraction, the filter (one size fraction at one location and one depth) belongs either to a**
 750 **validated niche (green), a non validated niche (red), an excluded filter (see Materials & Methods)**
 751 **(black) or no data is available (white). Non validated niches represent only 11 % of filters present**
 752 **in the dataset (66 out of 595).**

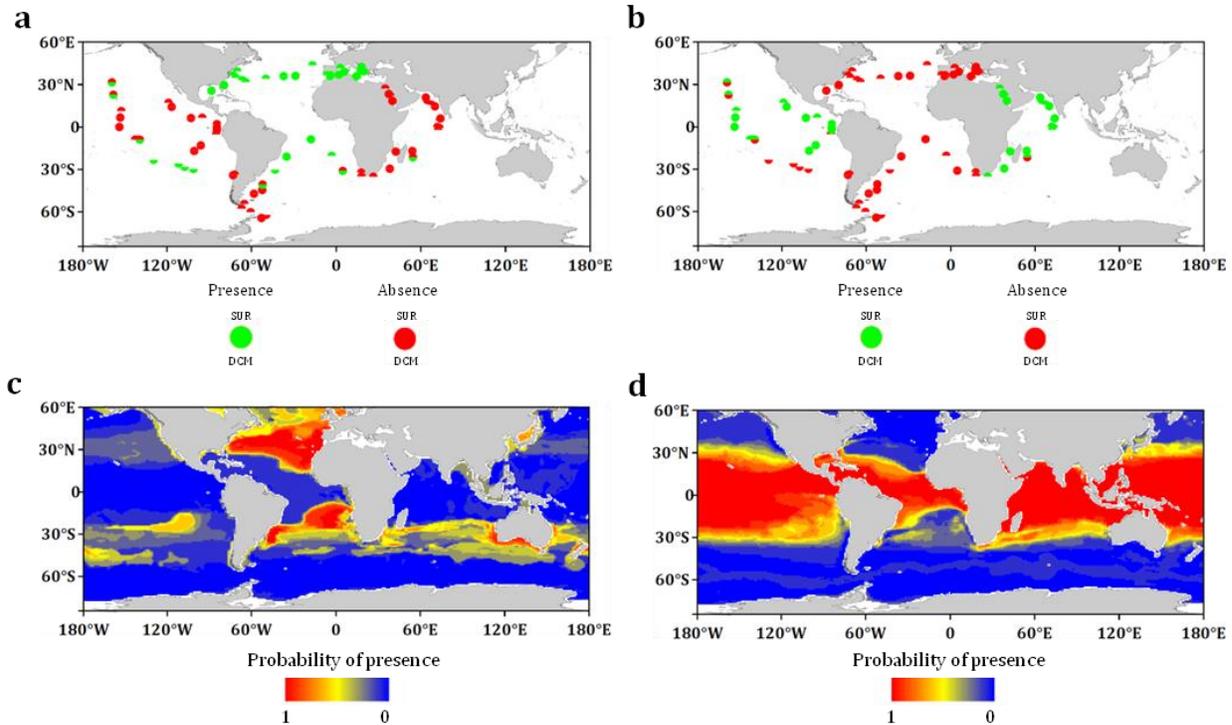
753



754 **Supplementary Figure 3 | Signature genomes of provinces of the prokaryotes enriched size.**

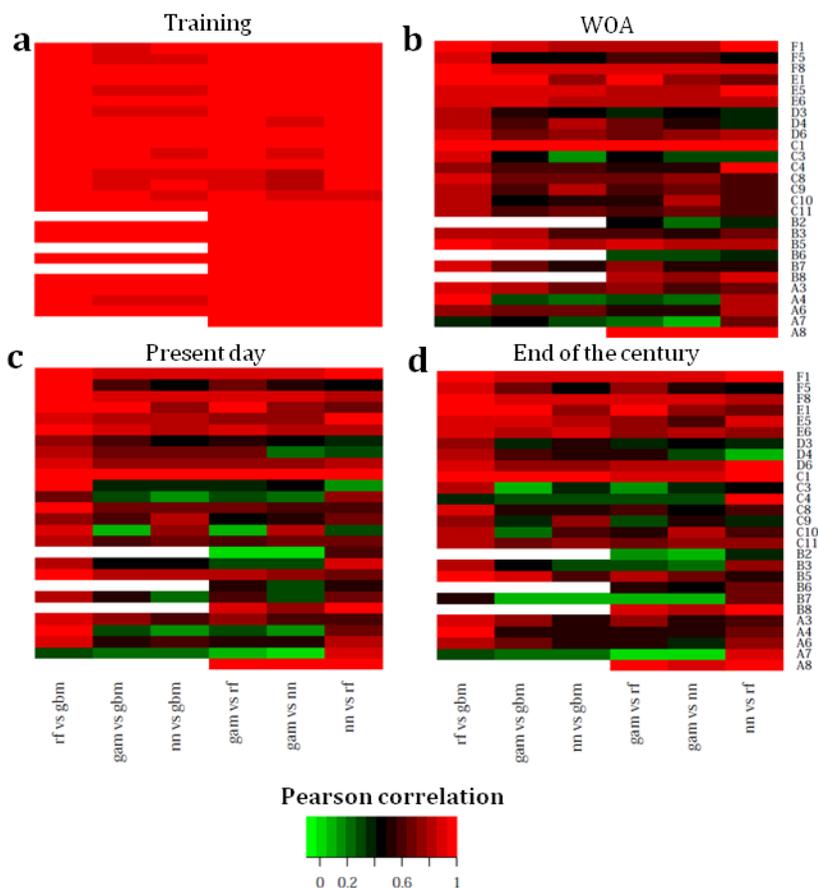
755 Indexes of presence enrichment¹⁴ for 523 genomes of prokaryotic plankton¹⁵ in corresponding
 756 provinces are clustered and represented in a color scale. Signature genomes (see *Methods*) are
 757 found for almost all provinces, their number and taxonomies are summarized (detailed list in
 758 Supplementary Table 6).

759



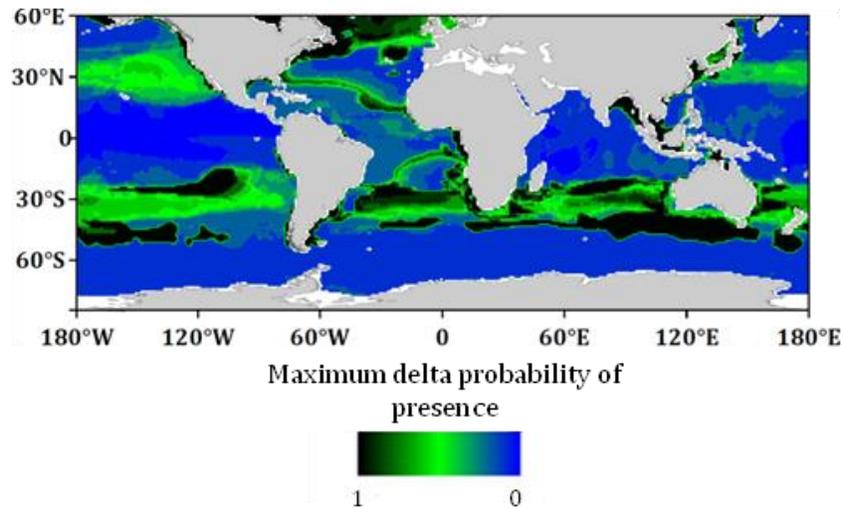
760 **Supplementary Fig. 4 | Sampling and projection maps on WOA13 climatological data of two**
761 **example provinces from size fraction 180-2000 μm .** (a) Sampling map of province F5. (b)
762 Province F8. (c) Projection map of province F5 on WOA13¹. (d) Province F8. Qualitatively,
763 projection maps are coherent with sampling maps of the two provinces with the highest probability
764 of presence projected in the sampling regions. Other presence zones are also predicted by the
765 projection. Sampling of these zones might be interesting to confirm our approach and projections
766 such as South of Australia where a high probability of presence for province F5 is predicted.

767



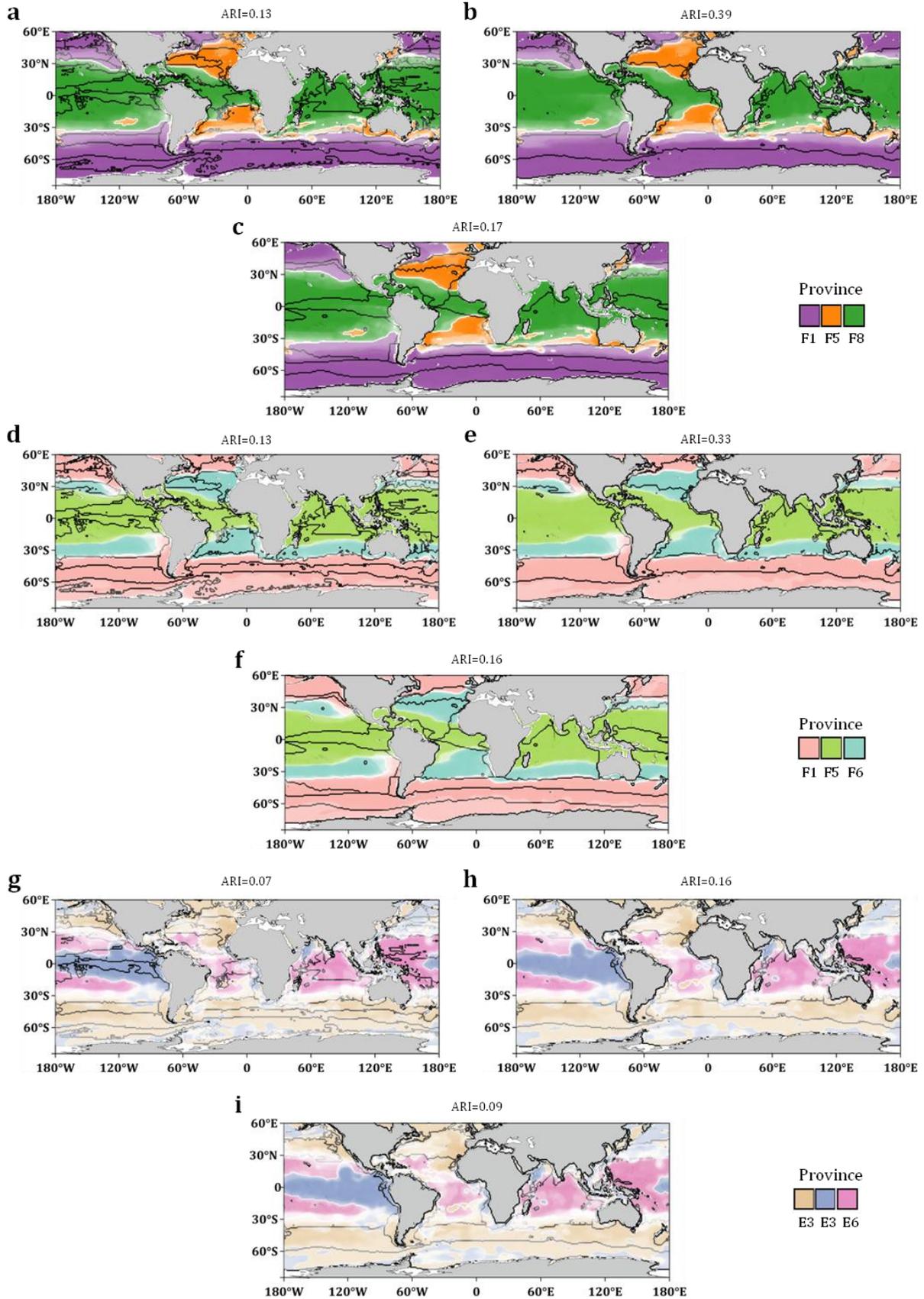
768 **Supplementary Fig. 5 | Pairwise Pearson correlation coefficient heat maps between outputs**
 769 **of the 4 machine learning models.** rf: Random Forest, gbm: gradient boosting machines, gam:
 770 general additive models, nn: neural networks. In rows are the provinces of the different size
 771 fractions. In columns are the pairwise comparisons of each machine learning technique. (a)
 772 Training set outputs. (b) WOA13 average data (except for Iron, PISCES-v2¹⁶ is used) projection
 773 outputs. (c) Present day data projection outputs (bias corrected mean model of 6 Earth System
 774 Models). (d) End of the century projection outputs. On the training set, models are in agreement
 775 with most of the correlation coefficients superior to 0.9. A drop of correlation is observed for
 776 modeled data (c, d) especially in small size fractions. This shows modeled data are more distant
 777 from the training set than WOA data. Random Forest and Gradient Boosting Machine are in very
 778 good agreement (first columns) which could be expected as they are both based on multiple
 779 decision trees.

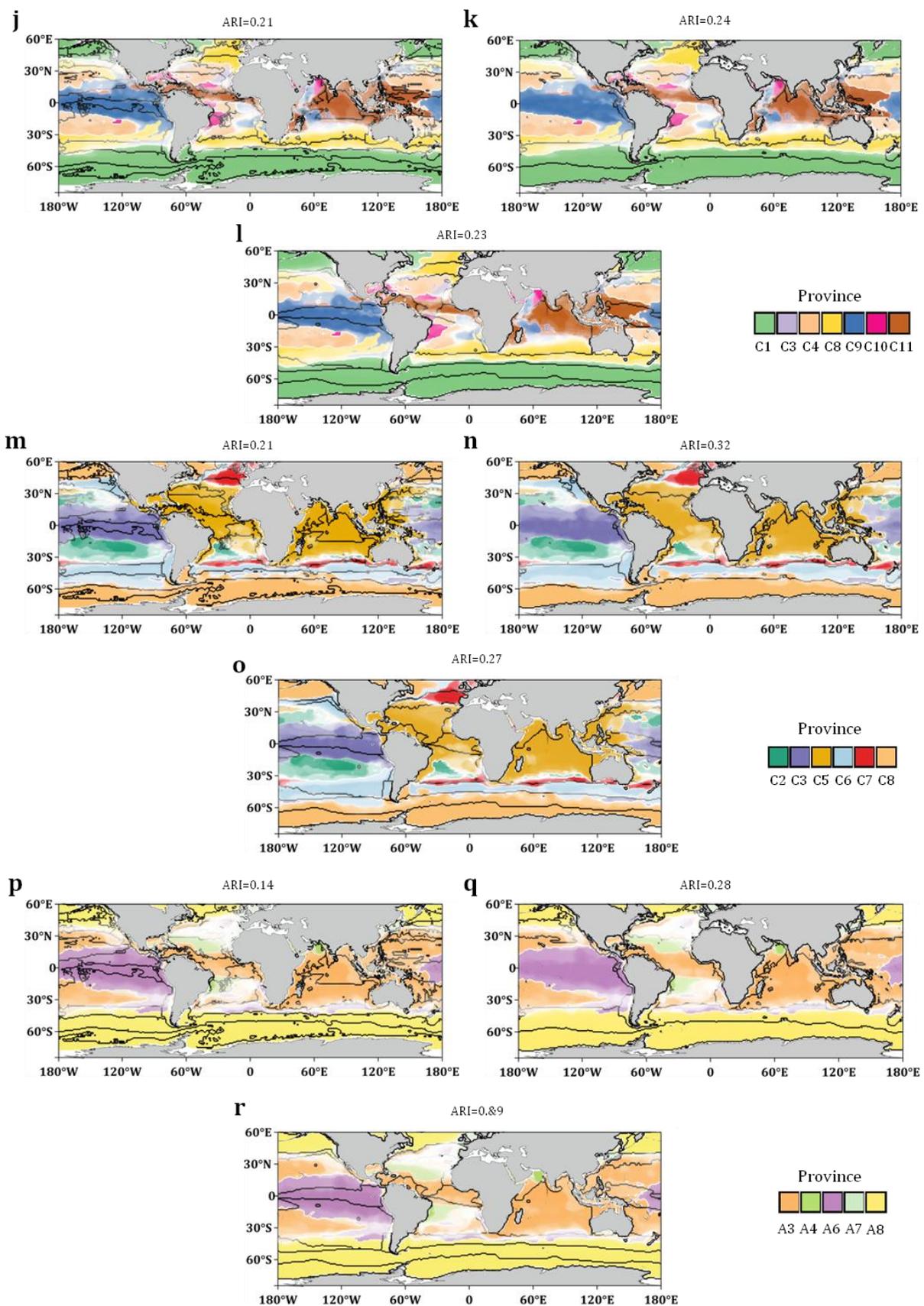
780



781 **Supplementary Fig. 6 | Probability range map (WOA data) of province F5 of size fraction 180-**
782 **2000 μm .** At each point of the grid, the maximum delta probability of presence between the 4
783 machine learning projections is calculated. In black are the zones where two models completely
784 disagree: one model predicts presence with certainty whereas the other predicts absence with
785 certainty. Disagreement appears mostly in uncertain presence areas (Supplementary Fig. 3c)
786 whereas models are generally in good agreement in absence areas (blue zones).

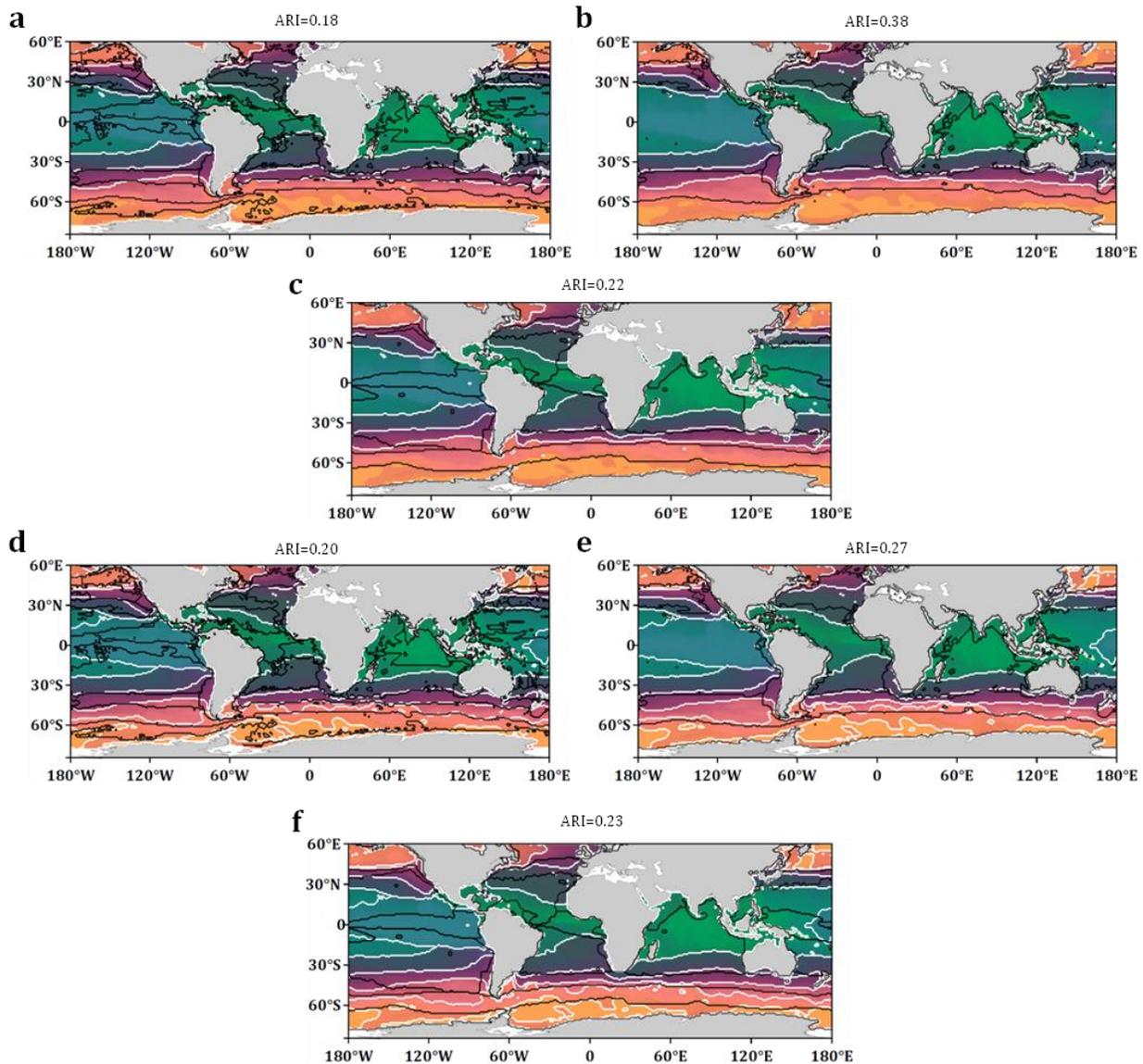
787



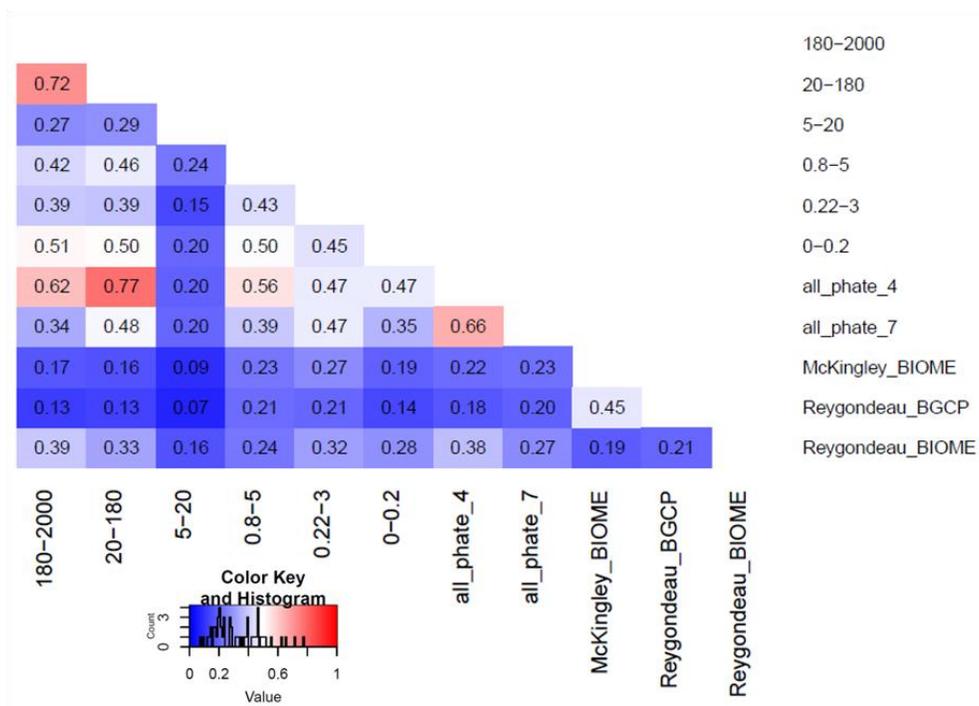


789 Supplementary Fig. 7 | Three existing oceanic partitions overlaid on top of plankton

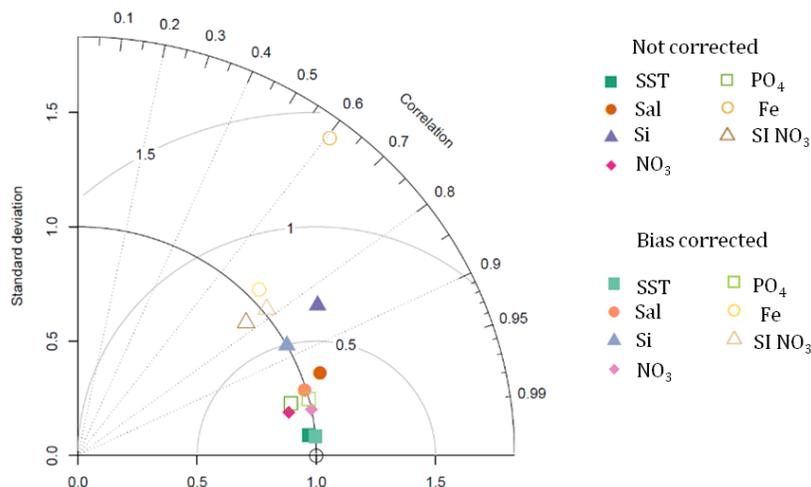
790 **provinces for the six fractions and combined size fractions.** The three oceanic partitions are
791 Reygondeau et al. Biogeochemical Provinces (BGCP)¹⁷; Reygondeau et al. Biomes¹⁷; Fay and
792 McKingley Biomes¹⁸. Each partitioning mask is overlaid in the above order on top of plankton
793 provinces for the six size fractions. **(a-c)** 180-2000 μm **(d-f)** 20-180 μm **(g-i)** 5-20 μm **(j-l)** 0.8-5
794 μm **(m-o)** 0.22-3 μm and **(p-r)** 0-0.2 μm . Above all maps, the adjusted rand index (ARI), an index
795 used to compare partitions, comparing mapped biogeographies with the black lines mask is shown.
796



797 **Supplementary Fig. 8 | Three existing oceanic partitions overlaid on top of plankton**
798 **provinces for the combined size fractions.** The three oceanic partitions are Reygondeau et al.¹⁷
799 Biogeochemical Provinces (BGCP); Reygondeau et al. Biomes¹⁷; Fay and McKingley Biomes¹⁸. Each
800 partitioning mask is overlaid in the above order on top of size fraction plankton provinces
801 combined by the PHATE algorithm in (a-c) 4 clusters and (d-f) 7 clusters.

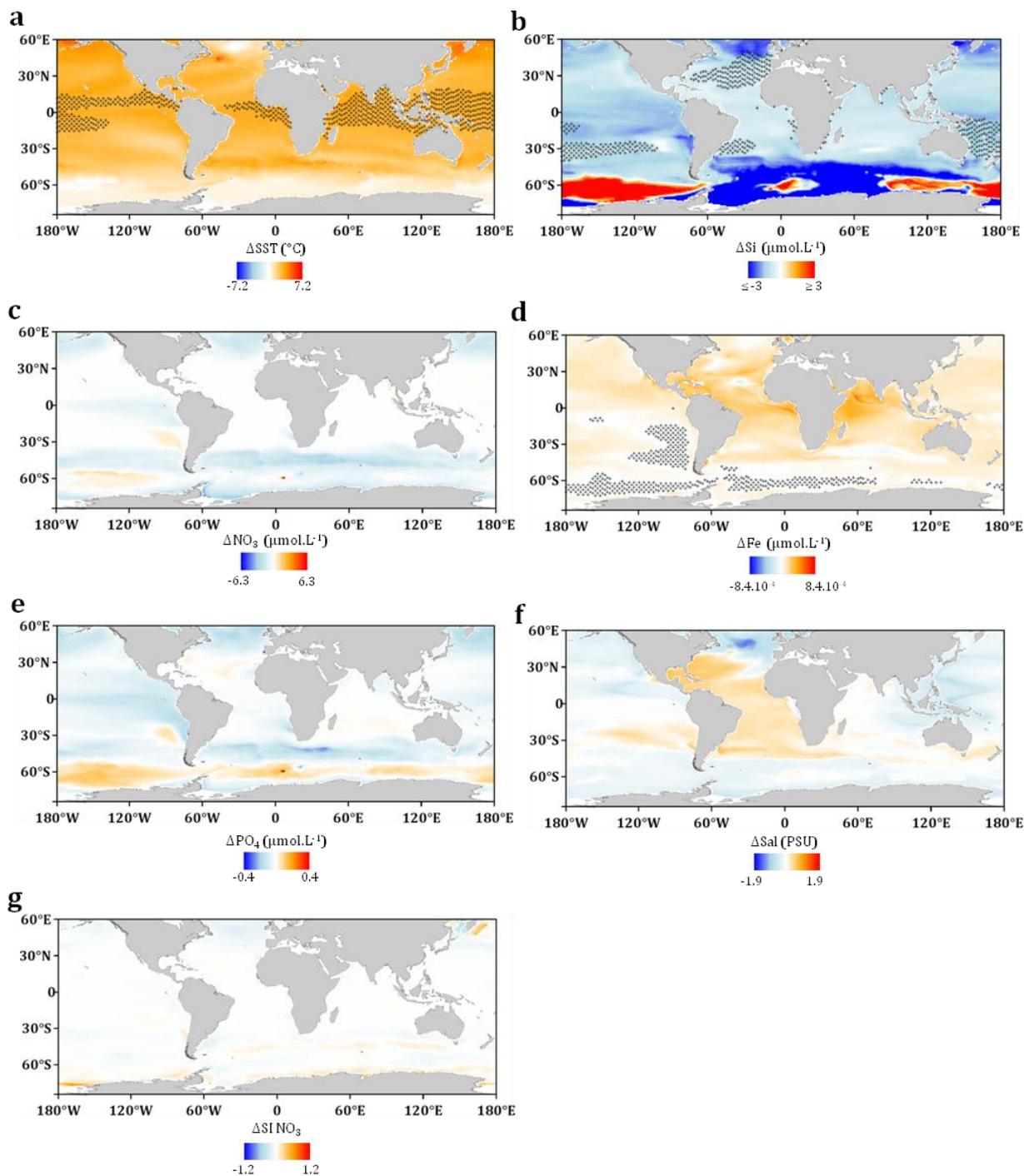


802 **Supplementary Fig. 9 | Pairwise comparisons of ocean partitions based on plankton**
 803 **provinces and existing biogeochemical partitions.** The three oceanic partitions are Reygondeau
 804 et al.¹⁷ Biogeochemical Provinces (BGCP); Reygondeau et al. Biomes¹⁷; Fay and McKingley Biomes¹⁸.
 805



806 **Supplementary Fig. 10 | Taylor diagram exhibiting statistical comparison of WOA 2013**
807 **observations and present day ESM model mean of the different drivers.** Each variable is
808 centered and scaled according to the mean and standard deviation of the observed variable (black
809 circle point at standard deviation 1 on the x-axis). Dark color points are the ESM model mean
810 without Cumulative Distribution Function transform (CDFt¹⁹) bias correction. They are to be
811 compared with light color points for which the bias correction is performed. Overall, good spatial
812 correlations are found between the model mean and the observations. CDFt bias correction
813 performs well by bringing standard deviations of the model to the observed standard deviations
814 without decreasing correlations.

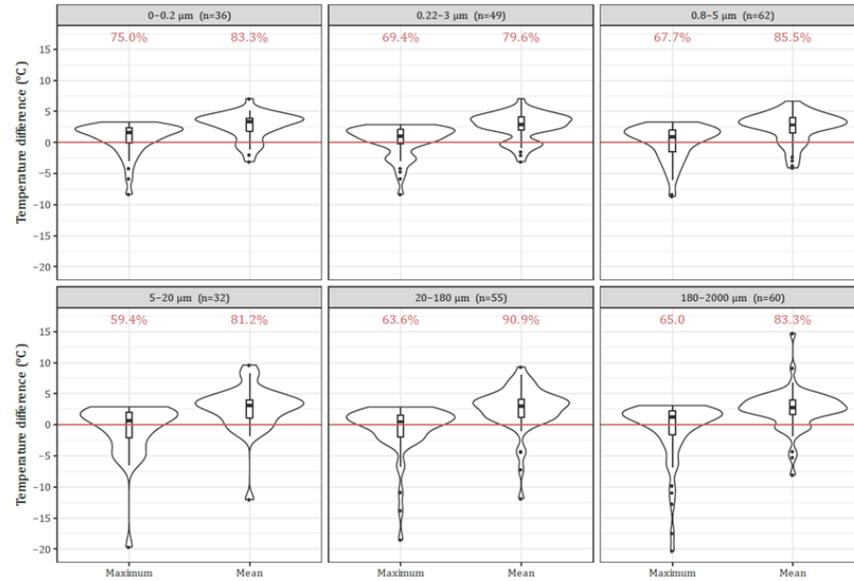
815



816 **Supplementary Fig. 11 | Differences in drivers' intensity (2090/99-2006/15) in the bias**
817 **corrected ESM model-mean under RCP8.5. (a) Sea Surface Temperature (SST). (b) Dissolved**
818 **silica. (c) Nitrate. (d) Iron. (e) Phosphate. (f) Salinity (Sal). (g) Seasonality Index of Nitrate (SI NO₃).**
819 Note that the scale for dissolved silica variations is restricted to visualize small variations. Regions
820 for which out of range values (*i.e.* inferior to the minimum or superior to the maximum found in

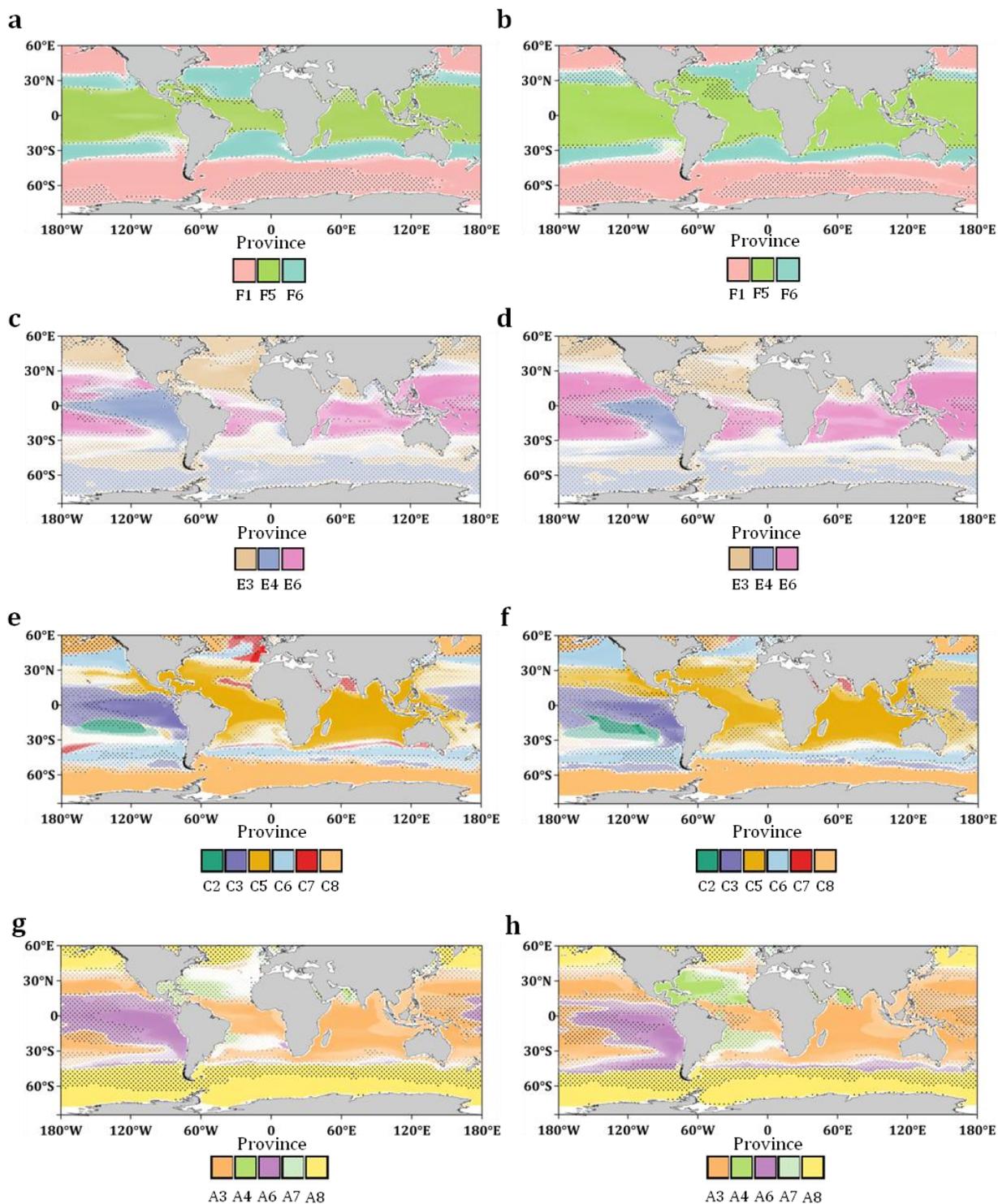
821 2006-15) are reached at the end of the century are highlighted with small stars reflecting high
822 uncertainty zones for machine learning approaches.

823



824 **Supplementary Fig. 12 | Distribution of deltas between future temperature at each sampling**
825 **site minus either the mean or maximum temperature within their contemporary genomic**
826 **province.** For most of the sites and across size fractions the future temperature projected by the
827 bias adjusted ESM ensemble model is higher than both the maximum and mean contemporary
828 temperature of their genomic province.

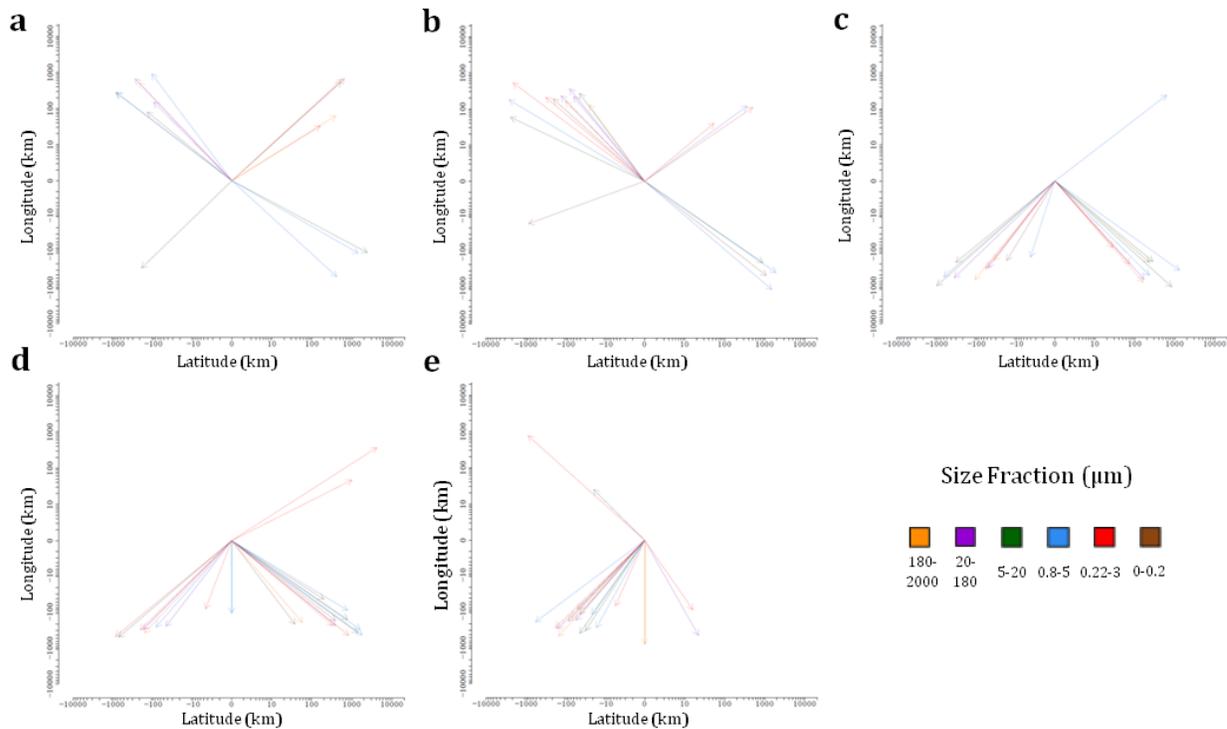
829



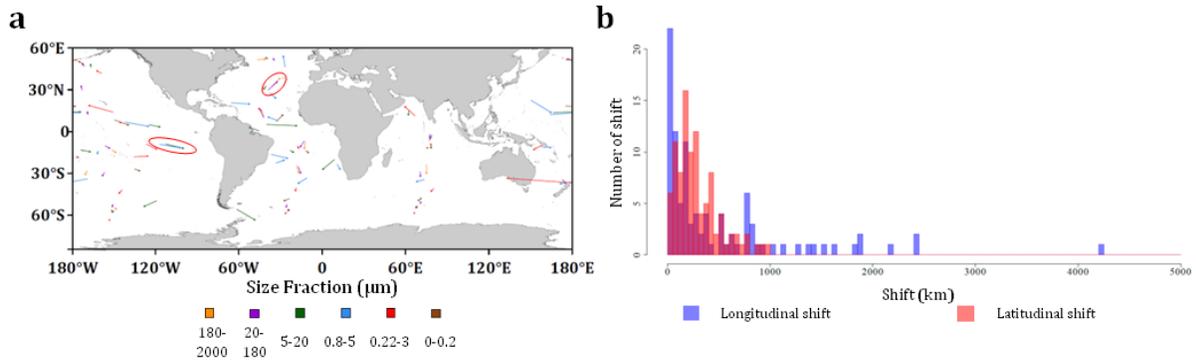
830 **Supplementary Fig. 13 | Global geographical patterns for 20-180, 5-20, 0.22-3, 0-0.2 μm**
 831 **plankton size fractions in present day (a, c, e, g) and at the end of the century (b, d, f, h). The**
 832 **dominant province *i.e.* the one predicted to have the highest probability of presence is represented**
 833 **at each grid point of the map. The color transparency is the probability of presence of the dominant**

834 province. Expansion of tropical provinces and shrinkage of temperate communities is consistently
835 projected in all size fractions.

836

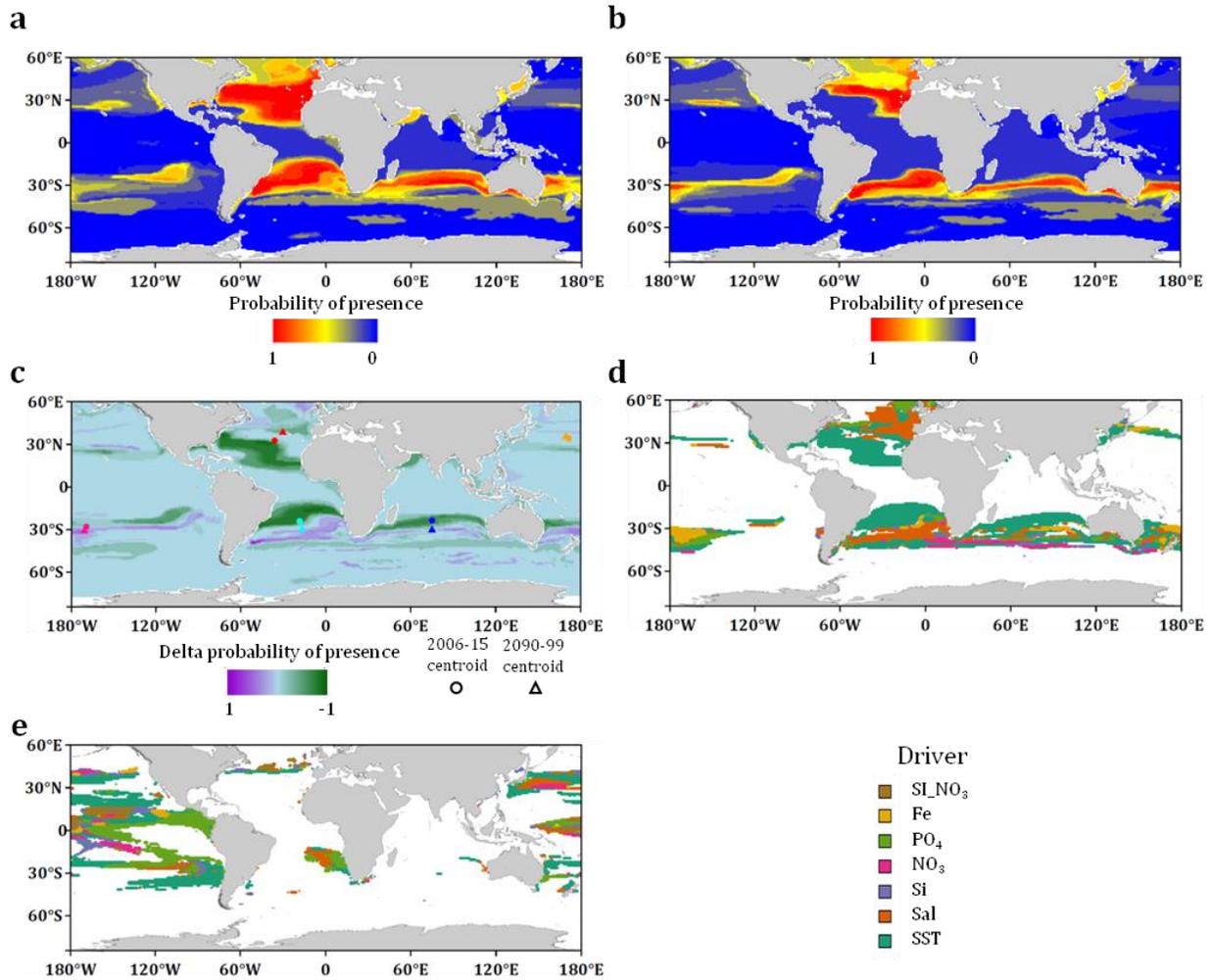


837 **Supplementary Fig. 14 | Projected migration shifts of the 27 provinces between present day**
838 **and end of the century.** Predicted migration shifts are presented in 5 major ocean basins: (a)
839 North Atlantic (b) North Pacific (c) South Atlantic (d) South Pacific and (e) Indian Ocean. 96 % of
840 migration shifts (larger than 200 km) are oriented towards the pole. Mean shift is 641 km (76 ± 79
841 km.dec⁻¹) and median shift is 394 km (47 km.dec⁻¹). Few provinces are projected to shift more than
842 thousands of kilometers toward suitable environmental conditions with a maximum shift of 4325
843 km.



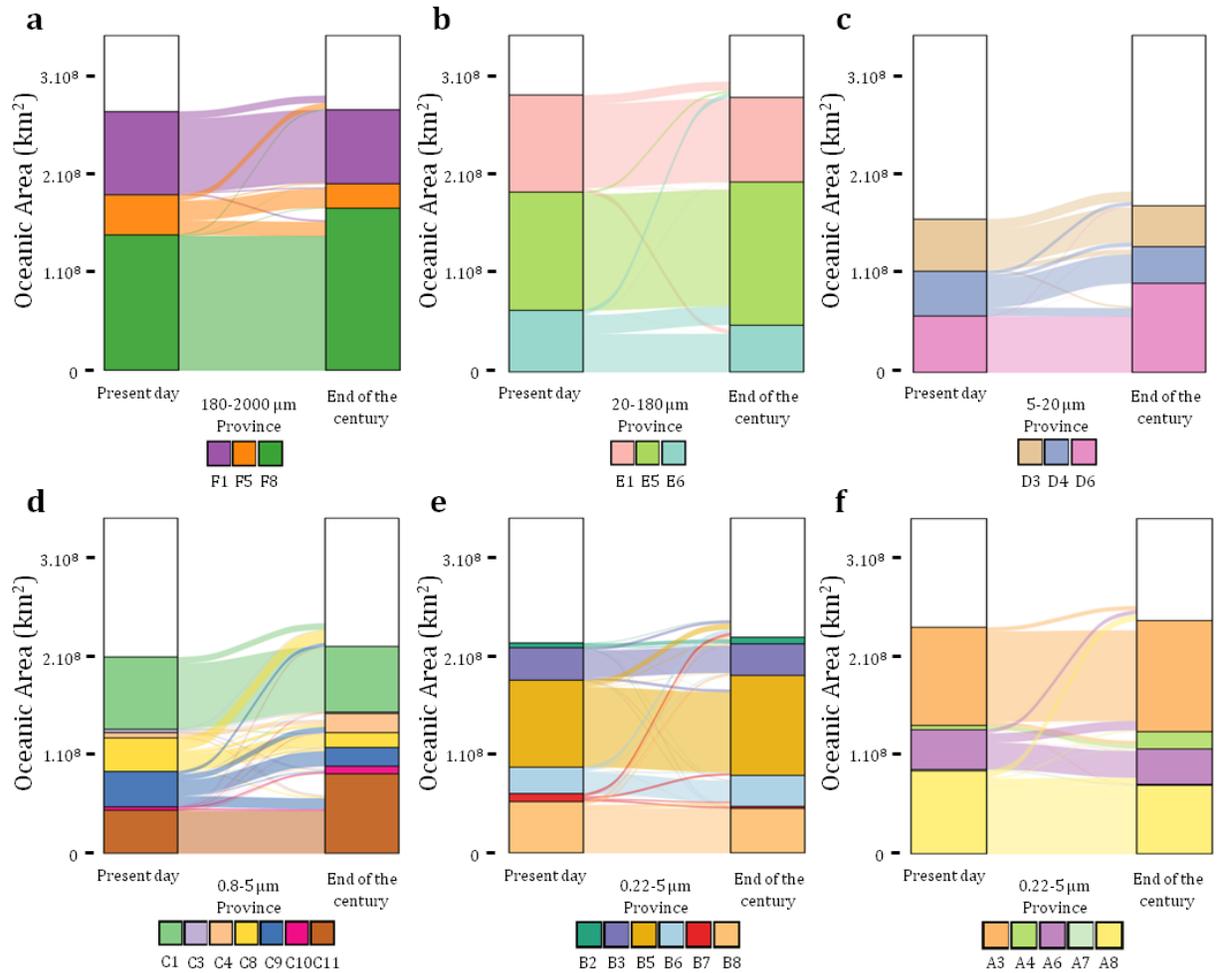
844 **Supplementary Fig. 15 | (a) Projected migration shifts of provinces on the world map. (b)**
845 **Latitudinal shift distribution (red bars) and longitudinal shift distribution (blue bars). (a)**
846 Migration shifts are represented as arrows pointing at the end of century centroid. Arrows are
847 colored according to the size fraction. Some shifts seem to correlate with each other (exemplified
848 with circled arrows). For instance, parallel shifts are projected in the southern pacific equatorial
849 communities of size fractions 0.8-5 μm and 5-20 μm (blue and green circled arrows). All non-
850 poleward arrows belong to small size classes (<20μm) showing differential responses to climate
851 change depending on the size class. (b) Some longitudinal shifts are more important than
852 latitudinal shifts with 14 longitudinal shifts superior to 1000 kms. Mean longitudinal shift (around
853 500 kms) is significantly higher (Student t-test p-value<0.01) than mean latitudinal shift (around
854 290 kms) while medians (longitudinal 190 kms vs latitudinal 230 kms) are not significantly
855 different (Wilcoxon test).

856



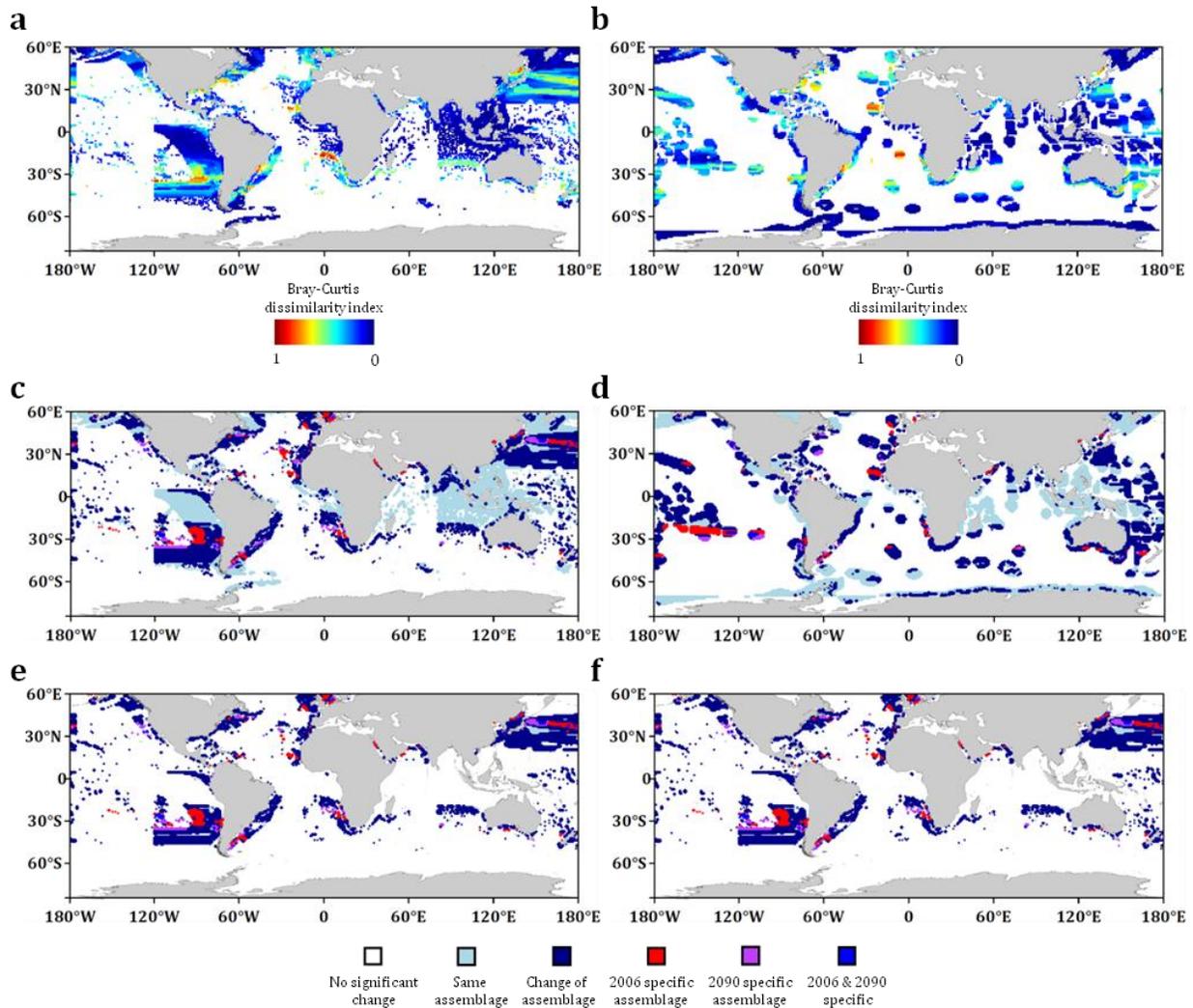
857 **Supplementary Fig. 16 | Projection maps of province F5 of size fraction 180-2000 μm in**
 858 **present day (a) and at the end of the century (b).** At each grid point, the probability of presence
 859 of the province is computed as the average of the predicted probability of each of the four machine
 860 learning techniques (gbm, nn, rf and gam). Red color indicates a high probability of presence.
 861 (a) The projected province corresponds to the sampled province (North Atlantic and South Atlantic)
 862 but several other places have high probabilities of presence such as South Australia where no
 863 sampling is available. (b) At the end of the century, the province is projected to reduce significantly
 864 in size. (c) Delta probability of presence map (2006/15 – 2090/99) and core range shift in the 5
 865 major oceanic basins of province F5 of size fraction 180-2000 μm . In all the basins, the centroid of
 866 the province is projected to migrate poleward. (d) Main drivers associated with the projected
 867 changes. Changes are mainly driven by sea surface temperature (55%) followed by salinity (14%).
 868 (e) Main drivers associated with the shrinkage of the equatorial cluster C9 of size fraction 0.8-5.

869 Considering only latitudes between the two tropics, changes are mainly driven by decreases in PO₄
870 (24 %) in addition to SST (27 %) (overall 34 % STT and 20 % PO₄).
871

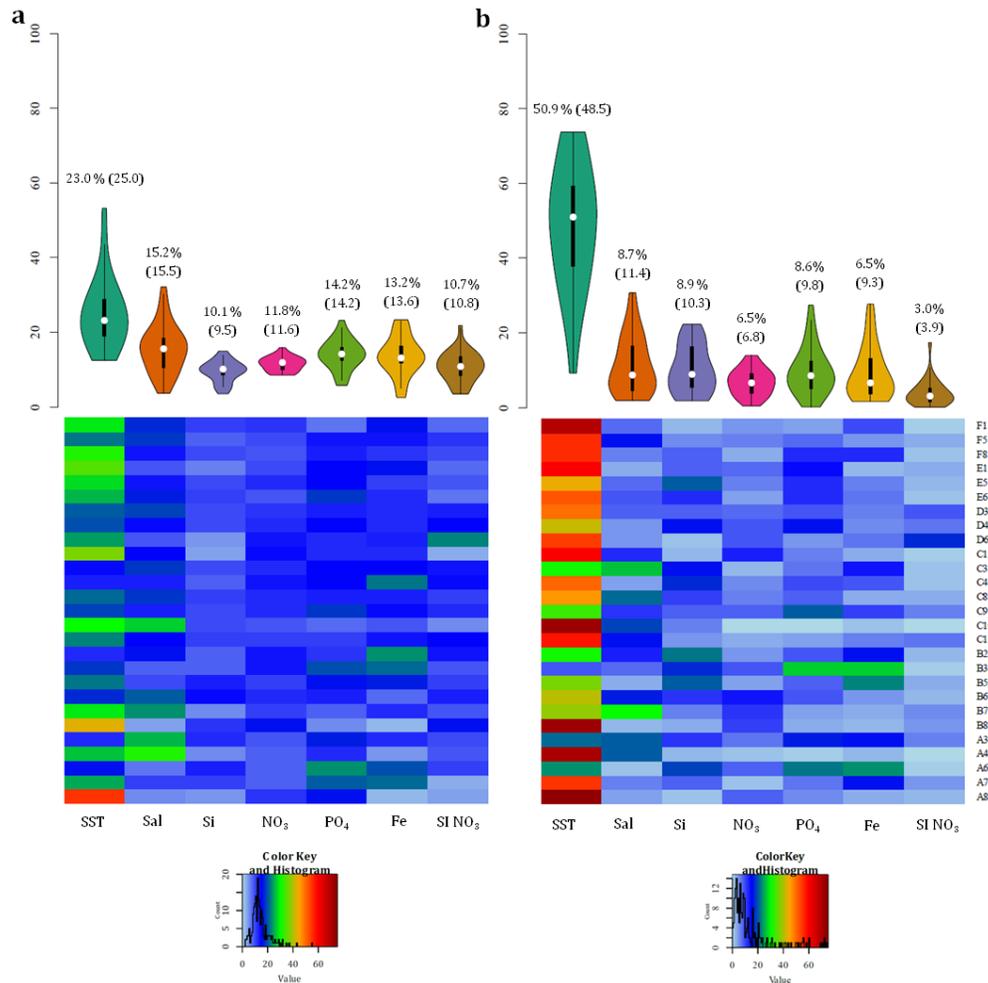


872 **Supplementary Fig. 17 | Probabilistic covered areas of the provinces projected in present**
873 **day and at the end of the century. (a) 180-2000 μm (b) 20-180 μm (c) 5-20 μm (d) 5-20 μm (e)**
874 **0.8-5 μm (f) 0-0.2 μm.** The covered area by a province is defined as the area in which this province
875 is dominant and weighted by its probability of presence at each point and grid cell area. Areas not
876 covered by the provinces are represented in white.

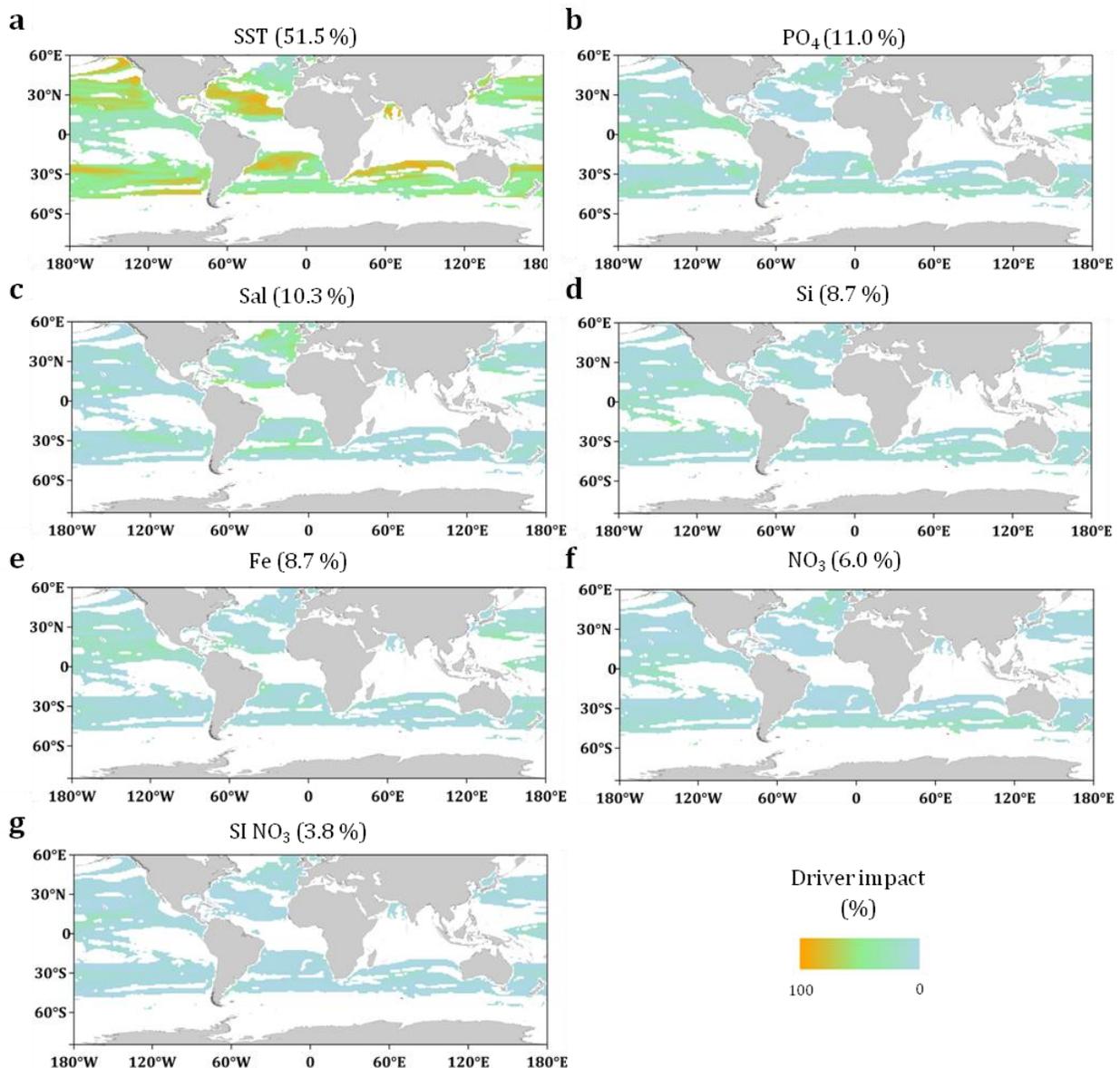
877



878 **Supplementary Fig. 18 | Bray-Curtis dissimilarity index and assemblage changes maps**
879 **comparing present day with end of the century projections of *dominant* provinces in (a)**
880 **principal fisheries (4 last deciles²⁰) and (b) Exclusive Economic Zones²¹. Assemblage changes in (c)**
881 **Principal fisheries (d) Exclusive Economic Zones. Assemblage changes in (e) Principal fisheries (f)**
882 **Exclusive Economic Zones with a Bray-Curtis dissimilarity index superior to 1/6.**
883



884 **Supplementary Fig. 19 | Distributions and heat maps of the relative influences of the**
 885 **different drivers in (a) defining single niches associated to the provinces from DALEX R**
 886 **package²² (b) driving climate change associated reorganizations of single provinces.** Median
 887 relative influence of Temperature is significantly higher than for all other environmental
 888 parameters (Pairwise Wilcoxon test $p < 0.01$ for all parameters) in (a) defining the niches and (b)
 889 driving province reorganization. It is also the case within individual size fractions for (b) but not for
 890 (a). Respectively, Salinity (Sal) and Phosphate (PO₄), have second and third highest median relative
 891 influences far behind Temperature whereas dissolved Silica and seasonality index of Nitrate (SI
 892 NO₃) have the lowest median relative importance in (a). Numbers above violins are median (and
 893 mean) relative influences.
 894



895 **Supplementary Fig. 20 | Maps of environmental drivers' relative influences in driving**
896 **province reorganization.** Each environmental parameter relative influence is quantified by
897 considering only the variation of each parameter individually between present day and end of the
898 century as defined in Barton et al.²³ (*Materials and Methods*) and where a significant change is
899 projected. Importantly the mean impact of SST (51.5 %) is largely the highest. PO₄ is the second
900 most impacting drivers (11.0 %). Contrary to supplementary Fig. 16, relative influence is calculated
901 here by combining all provinces together. Therefore, mean relative influences slightly differ
902 especially for dissolved silica (Si).
903

904

Model	Reference
CESM1-BGC	Gent et al., 2011
GFDL-ESM2G	Dunne et al., 2013
GFDL-ESM2M	Dunne et al., 2013
HadGEM2-ES	Collins et al., 2011
IPSL-CM5A-LR	Dufresne et al., 2013
IPSL-CM5A-MR	Dufresne et al., 2013
MPI-ESM-LR	Giorgetta et al., 2013
MPI-ESM-MR	Giorgetta et al., 2013
NorESM1-ME	Bentsen et al., 2013

905

Supplementary Table 1 | Earth System models used to compute the mean model.

906

907

Fraction	Province	Climatic annotation	Area 2006-15 (MKm ²)	Area 2090-99 (MKm ²)	Delta area (MKm ² (%))
180-2000	F1	polar	82	73	-9 (-11%)
180-2000	F5	temperate	44	26	-18 (-41%)
180-2000	F8	tropico-equatorial	140	169	29 (+21%)
20-180	E1	polar	97	84	-13 (-13%)
20-180	E5	tropico-equatorial	119	145	+26 (+22%)
20-180	E6	temperate	65	49	15 (-23%)
5-20	D3	temperate	51	41	-10 (-20%)
5-20	D4	equatorial	46	37	-8 (-18%)
5-20	D6	tropical	65	97	32 (+48%)
0.8-5	C1	polar	71	64	-6 (-8%)
0.8-5	C3	subtropical	3,6	1,4	-2,2 (-61%)
0.8-5	C4	subtropical	6	19	+13 (+217%)
0.8-5	C8	temperate	33	15	-18 (-54%)
0.8-5	C9	equatorial	35	19	-16 (-45%)
0.8-5	C10	subtropical	3.8	7.4	+3,6 (+95%)
0.8-5	C11	tropical	49	86	37 (+75%)
0.22-3	B2	subtropical	5,1	6,7	+1,6 (+31 %)
0.22-3	B3	equatorial	33	32	-1 (-0.3%)
0.22-3	B5	tropical	88	102	+13 (+15%)
0.22-3	B6	temperate	30	35	+5 (+17%)
0.22-3	B7	temperate	8	2	-6 (-75%)
0.22-3	B8	polar	51	44	-7 (-14%)
0-0.2	A3	tropical	101	114	+13 (+13 %)
0-0.2	A4	subtropical	6	19	+13 (+216 %)
0-0.2	A6	equatorial	40	36	-4 (-10%)
0-0.2	A7	temperate	1,7	1	-0,7 (-41%)
0-0.2	A8	polar	81	67	-14 (-17%)

908 **Supplementary Table 2 | Genomic provinces climatic annotations and oceanic surfaces they**
 909 **cover in present day (2006-15) and at the end of the century (2090-990).**

910

911

	Fraction (μm)	SST	Sal	Si	NO3	PO4	Fe	SI NO3
Niche definition	180-2000	27,7	16,8	10,1	10,3	11,4	14	9,7
	20-180	29,7	13,3	8,9	10,7	16,3	13,2	7,9
	5-20	21,5	14,2	8,4	12,7	13,7	12,6	17
	0,8-5	22,6	17,3	9,2	12,3	13,4	14,5	10,8
	0,22-3	23,9	13,7	10,5	13	12,5	14,2	12,2
	0-0,2	27,1	16,4	9,3	9,5	17,9	12,1	7,6
	all	25	15,5	9,5	11,6	14,2	13,6	10,8
Climate change	180-2000	75,2	8,4	2,5	2,6	4,1	5,8	1,4
	20-180	71,7	4,2	7,3	3,9	8,7	3,3	1,0
	5-20	57,9	6,1	4,4	6,6	9,0	6,5	1,0
	0,8-5	54,9	13,4	7,2	5,5	9,7	5,9	3,6
	0,22-3	42,6	9,9	13,5	6,4	9,1	15,2	3,3
	0-0,2	42,3	18,9	6,6	6,4	12,1	10,8	2,9
	all	51,5	10,3	8,7	6	11	8,7	3,8

912 **Supplementary Table 3 | Summary table of relative importance of each environmental**
 913 **driver in niche definition and in driving geographical reorganization in response to climate**
 914 **change.** Note that in both cases (niche definition and climate change), the row 'all' is not the mean
 915 over the size fractions. In the case of niche definition, this is due to a different number of niches in
 916 each size class. In the case of climate change, relative influence is either calculated for single
 917 provinces at a given grid point then recalculated for individual size class or calculated for all
 918 provinces together (row 'all').

919

920 **References**

- 921 1. Boyer, T. P. et al. WORLD OCEAN DATABASE 2013, NOAA Atlas NESDIS 72. Sydney Levitus,
922 Ed.; Alexey Mishonoc, Tech. Ed. (2013). doi:10.7289/V5NZ85MT
- 923 2. Gent, P. R. et al. The community climate system model version 4. *J. Clim.* (2011).
924 doi:10.1175/2011JCLI4083.1
- 925 3. Dunne, J. P. et al. GFDL's ESM2 global coupled climate-carbon earth system models. Part II:
926 Carbon system formulation and baseline simulation characteristics. *J. Clim.* (2013).
927 doi:10.1175/JCLI-D-12-00150.1
- 928 4. Collins, W. J. et al. Development and evaluation of an Earth-System model - HadGEM2.
929 *Geosci. Model Dev.* (2011). doi:10.5194/gmd-4-1051-2011
- 930 5. Dufresne, J. L. et al. Climate change projections using the IPSL-CM5 Earth System Model:
931 From CMIP3 to CMIP5. *Clim. Dyn.* (2013). doi:10.1007/s00382-012-1636-1
- 932 6. Giorgetta, M. A. et al. Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM
933 simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.*
934 (2013). doi:10.1002/jame.20038
- 935 7. Bentsen, M. et al. The Norwegian Earth System Model, NorESM1-M – Part 1: Description and
936 basic evaluation of the physical climate. *Geosci. Model Dev.* (2013). doi:10.5194/gmd-6-687-2013
- 937 8. van Vuuren, D. P. et al. The representative concentration pathways: An overview. *Clim.*
938 *Change* (2011). doi:10.1007/s10584-011-0148-z
- 939 9. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* (2006).
940 doi:10.1016/j.patrec.2005.10.010
- 941 10. Wood, S. N. Stable and efficient multiple smoothing parameter estimation for generalized
942 additive models. *J. Am. Stat. Assoc.* (2004). doi:10.1198/016214504000000980
- 943 11. Ridgeway, G. gbm: Generalized Boosted Regression Models. R Packag. version 1.6-3.1
944 (2010).
- 945 12. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* Fourth edition by. World
946 (2002). doi:10.2307/2685660
- 947 13. Breiman, L. & Cutler, A. Breiman and Cutler's random forests for classification and
948 regression. Packag. 'randomForest' (2012). doi:10.5244/C.22.54

- 949 14. Jaccard, P. Distribution comparée de la flore alpine dans quelques régions des Alpes
950 occidentales et orientales. Bull. la Murithienne (1902).
- 951 15. Delmont, T. O. et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are
952 abundant in surface ocean metagenomes. Nat. Microbiol. (2018). doi:10.1038/s41564-018-0176-9
- 953 16. Aumont, O., Ethé, C., Tagliabue, A., Bopp, L. & Gehlen, M. PISCES-v2: An ocean
954 biogeochemical model for carbon and ecosystem studies. Geosci. Model Dev. (2015).
955 doi:10.5194/gmd-8-2465-2015
- 956 17. Reygondeau, G. et al. Dynamic biogeochemical provinces in the global ocean. Global
957 Biogeochem. Cycles (2013). doi:10.1002/gbc.20089
- 958 18. Fay, A. R. & McKinley, G. A. Global open-ocean biomes: Mean and temporal variability. Earth
959 Syst. Sci. Data (2014). doi:10.5194/essd-6-273-2014
- 960 19. Michelangeli, P. A., Vrac, M. & Loukos, H. Probabilistic downscaling approaches: Application
961 to wind cumulative distribution functions. Geophys. Res. Lett. (2009). doi:10.1029/2009GL038401
- 962 20. Watson, R. A. A database of global marine commercial, small-scale, illegal and unreported
963 fisheries catch 1950-2014. Sci. Data (2017). doi:10.1038/sdata.2017.39
- 964 21. Flanders Marine Institute (2018). Maritime Boundaries Geodatabase: Maritime Boundaries
965 and Exclusive Economic Zones (200NM), version 10. (2018). doi:<https://doi.org/10.14284/313>.
- 966 22. Biecek, P. DALEX: explainers for complex predictive models. J. Mach. Learn. Res. 19, 1–5
967 (2018).
- 968 23. Barton, A. D., Irwin, A. J., Finkel, Z. V. & Stock, C. A. Anthropogenic climate change drives shift
969 and shuffle in North Atlantic phytoplankton communities. Proc. Natl. Acad. Sci. (2016).
970 doi:10.1073/pnas.1519080113