Restructuring of genomic provinces of surface ocean plankton under climate change

Paul Frémont[1,2*], Marion Gehlen[3*], Mathieu Vrac[3], Jade Leconte[1,2], Patrick Wincker[1,2], Daniele Iudicone[4], Olivier Jaillon[1,2*]

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

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

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

Planktonic communities are composed of complex and heterogeneous assemblages of small animals, small single-celled eukaryotes (protists), prokaryotes and viruses - that drift with currents. They contribute to the regulation of the Earth system notably through primary

27 production via photosynthesis[1], carbon export to the deep oceans[2,3] and form the base of
28 the food webs that sustain the whole trophic chain in the oceans and beyond[4].
29 The composition of communities is known to vary over time at a given site with daily[5] to
30 seasonal fluctuations[6] following environmental variability[7,8]. Overlying these relatively
31 short scale spatio-temporal variations, a more macroscale partitioning of the ocean was
32 evidenced by different combinations of biological and physico-chemical data[9–11], and
33 recently at the resolution of community genomics[12]. The basin scale biogeographical
34 structure was proposed to result from a combination of multiple bio-physico-chemical
35 processes named the seascape[7,8], including both abiotic and biotic interactions[13], neutral
36 genetic drift[14], natural selection[15–17], temperature variations, nutrient supply but also
37 advection and mixing along currents[12,14].
38 Today, knowledge of global scale plankton biogeography at the DNA level is in its infancy.
39 We lack understanding and theoretical explanations for the emergence and maintenance of
40 biogeographical patterns at genomic resolution. Omics data (*i.e.* the DNA/RNA sequences
41 representative of the variety of coding and non-coding sequences of organisms) provide
42 the appropriate resolution to track and record global biogeographical features[12],
43 modulation of the repertoire of expressed genes in a community in response to
44 environmental conditions[2,18,19] as well as eco-evolutionary processes[14,16,17]. Importantly,
45 metagenomic sequencing can be consistently analyzed across plankton organisms as
46 recently demonstrated by global expeditions[20–23]. Furthermore, the strong links between
47 plankton and environmental conditions suggest potentially major consequences of climate
48 change on community composition and biogeography[24,25]. Time series observations have
49 highlighted recent changes in the planktonic ecosystem attributed to this anthropogenic

50 pressure, such as changes in community composition[26–28] or poleward shifts of some
51 species[29,30]. These changes are expected to intensify with ongoing climate warming and
52 could lead to major reorganizations in plankton community composition[24,25], with a
53 potential decline in diversity[31–33]. Another major consequence of a global reorganization of
54 the seascape on biological systems (e.g. growth, grazing) would be a decrease of primary
55 production at mid-latitudes and an increase at higher latitudes[34].

56 Here we report a global structure of plankton biogeography based on metagenomic data
57 using niche models and its putative modifications under climate change. First, we show
58 that environmental niches[35], *i.e.* the envelope of environmental parameters suitable for an
59 organism or a population, can be defined at the scale of genomic provinces across 6
60 organism size fractions representing major plankton groups from nano- (viruses) to meso-
61 zooplankton (small metazoans). Then, we spatially extrapolate their niches into climatic
62 provinces to depict the structure of plankton biogeography of all but arctic regions for each
63 size fraction and for all combined. Then, considering the same niches, we assess putative
64 spatial reorganization of the same provinces and their associated environmental drivers
65 under climate change at the end of the century.

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

67 We use 38 previously defined genomic provinces[12] containing at least 4 sampling sites;
68 they correspond to 595 metagenomes for 6 size fractions (ranging from 0 to 2000 μm) and
69 sampled in 95 sites from all oceans except the Arctic (Supplementary Figs. 1-2).

70 To compute and test the validity of realized environmental niches, we train four machine
71 learning techniques to probabilistically associate genomic provinces with environmental
72 data: sea surface temperature, salinity, three macronutrients (dissolved silica, nitrate and

73 phosphate), one micronutrient (dissolved iron) plus a seasonality index of nitrate. A valid
74 environmental niche is obtained for 27 out of 38 initial provinces (71%) comforting their
75 definition and covering 529 samples out of 595 (89%, Supplementary Fig. 2). Rejected
76 provinces contain relatively few stations (mean of 6 ± 2.6 versus 19 ± 15.3 for valid
77 provinces, p-value<10^{-3} Wilcoxon test). For spatial and temporal extrapolations of the
78 provinces presented below, we use the ensemble model approach[36] that considers mean
79 predictions of machine learning techniques.
80 The signal of ocean partitioning is likely due to abundant and compact genomes whose
81 geographical distributions closely match provinces. Within a collection of 523 prokaryotic
82 and 713 eukaryotic genomes[37,38] from *Tara* Oceans samples, we find signature genomes for
83 all but 4 provinces. In total, they correspond to respectively 96 and 52 of the genomes and
84 their taxonomies are coherent with the size fractions (Fig. 1 for eukaryotes and
85 Supplementary Fig. 3 for prokaryotes). Some of them correspond to unexplored lineages
86 highlighting the gap of knowledge for organisms that structure plankton biogeography and
87 the strength of a rationale devoid of any *a priori* on reference genomes or species.
88 **Structure of present day biogeography of plankton**
89 To extrapolate to a global ocean biogeography for each size fraction, we define the most
90 probable provinces, named hereafter as *dominant* and assigned to a climatic annotation
91 (Supplementary Table 1), on each 1°x1° resolution grid point using 2006-13 WOA13
92 climatology[39] (Supplementary Fig. 4 and Fig. 3).
93 Overall, in agreement with previous observations[12], provinces of large size fractions (>20
94 µm) are wider and partially decoupled from those of smaller size fractions, probably due to
95 differential responses to oceanic circulation and environmental variations, different life

96 cycle constraints, lifestyles[7,8,12] and trophic networks positions[40]. Biogeographies of small
97 metazoans that enrich the largest size fractions (180-2000 µm and 20-180 µm) are broadly
98 aligned with latitudinal bands (tropico-equatorial, temperate and (sub)-polar) dominated
99 by a single province (Fig. 2a,b). A more complex oceanic structuring emerges for the
100 smaller size fractions (<20 µm) (Fig. 2c-f) with several provinces per large geographical
101 region. Taking size fraction 0.8-5 µm enriched in small protists (Fig. 2d) as an example,
102 distinct provinces are identified for oligotrophic gyres in the Atlantic and Pacific Oceans,
103 and one for the nutrient-rich equatorial upwelling region. A complex pattern of provinces,
104 mostly latitudinal, is also found for the bacteria enriched size class (Fig. 2e, 0.22-3 µm) and
105 the virus enriched size class (Fig. 2f, 0-0.2 µm) though less clearly linked to large-scale
106 oceanographic regions. A single province extending from temperate to polar regions
107 emerges from the size fraction 5-20 µm enriched in protists (Fig. 2c), for which a smaller
108 number of samples is available (Supplementary Fig. 2b-c), which probably biases this
109 result. Finally, we use PHATE[41] dimension reduction algorithm to combine all provinces for
110 all size classes into a single consensus biogeography revealing 4 or 7 robust clusters (Fig.
111 2g,h). The 4 cluster consensus biogeography is mainly latitudinally organized
112 distinguishing polar, subpolar, temperate and tropico-equatorial regions. The 7 cluster
113 consensus biogeography distinguishes the equatorial pacific upwelling biome and three
114 subpolar biomes that most likely reflect the chemico-physical structuring of the Southern
115 Ocean and known polar fronts (red lines Fig. 2h). However, learning data are scarcer south
116 of 60°S so these extrapolations need to be taken with caution.
117 Previous ocean partitioning either in biomes[9–11] or biogeochemical provinces (BGCPs)[9,11]
118 are based on physico-biogeochemical characteristics including SST[9–11], chlorophyll *a*[9–11],

salinity[9–11], bathymetry[9,11], mixed layer depth[10] or ice fraction[10]. Considering three of these partitions as examples we notice differences with our partitions (Supplementary Fig. 7-8) for example in terms of number of regions in the considered oceans (56 for 2013 Reygondeau et al. BGCPs[11], 17 for Fay and McKingley[10]) and structure (the coastal biome for 2013 Reygondeau et al. biomes[11]). Numerical comparison of our partitions with others (*Methods*), reveals low similarity between them, the highest being with Reygondeau biomes (Supplementary Figs. 7-9). However, biomes and BGCP frontiers closely match our province frontiers in many cases. Near the frontiers, *dominant* provinces have smaller probabilities in agreement with smooth transitions instead of sharp boundaries as already proposed[9] and with a seasonal variability of the frontiers[11] (Supplementary Fig 7). Some of these transitions are very large and match entire BGCPs, for example in subtropical North Atlantic and subpolar areas where high annual variations are well known[11].

Future changes in plankton biogeography structure

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

142 Our projections indicate multiple large-scale changes in biogeographical structure
143 including plankton organism size-dependent province expansions or shrinkages and shifts
144 (Supplementary Fig. 13-15). A change in the *dominant* province in at least one size fraction
145 would occur over 60.1 % of the ocean surface, ranging from 12 % (20-180 µm) to 31%
146 (0.8-5 µm) (Table 1).
147 Centroids of provinces with *dominance* areas larger than 10^6 km^2 within a basin would be
148 moved at least 200 km away for 77% of them, 96 % of which move poleward
149 (Supplementary Figs. 14 and 14). Most longitudinal shift distances are smaller (50% <190
150 km) but a few are larger than 1000 km while the distribution of latitudinal shifts is more
151 concentrated around the mean (290 km) with no shifts superior to 1000 km
152 (Supplementary Fig. 15b). These important longitudinal shifts corroborate existing
153 projections[24,43,44] and differ from trivial poleward shifts due to temperature increase,
154 reflecting more complex spatial rearrangements of the other environmental drivers
155 (Supplementary Fig. 11). We find an average displacement speed of the provinces'
156 centroids of 76 ± 79 $km.dec^{-1}$ (latitudinally mean of 34 ± 82 $km.dec^{-1}$ and 59 ± 82 $km.dec^{-1}$
157 longitudinally; median of 47 $km.dec^{-1}$, 23 $km.dec^{-1}$ latitudinally and 27 $km.dec^{-1}$
158 longitudinally). We project phytoplankton provinces displacements similar to previously
159 published shifts for the North Pacific (poleward shift of 118 $km.dec^{-1}$ for province C4
160 versus 100 $km.dec^{-1}$ for the subtropical biome of Polovina et al.[44] and eastward shift of 195
161 $km.dec^{-1}$ for province C9 versus 200 $km.dec^{-1}$ for the equatorial biome of Polovina et al.[44]).
162 For all size fractions, climate change would lead to a poleward expansion of tropical and
163 equatorial provinces at the expense of temperate provinces (Table 1 and Supplementary
164 Fig. 13). This is illustrated for the size fraction 180-2000 by the example of the temperate

province F5 (Supplementary Fig. 16) for which significant shrinkage is projected in the five major ocean basins. In the North Atlantic, its centroid would move approximately 800 km to the northeast (Supplementary Fig. 16c).

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

To further quantify patterns of expansion or shrinkage, we calculate the surface covered by the *dominant* provinces weighted by probabilities of presence (Supplementary Fig. 17, Supplementary Table 1). In this way, *dominant* provinces are defined on 100% of the surface ocean (327 millions of km^2) but their presence probabilities correspond to the equivalent of 45 to 74% (due to sampling variability and niche overlaps) of the surface ocean depending on plankton size fraction (Table 1). Overall, our results indicate expansions of the surface of tropical and tropico-equatorial provinces but in very different ways depending on the size fractions of organisms. The surface area of temperate provinces is ~22 million km^2 on average (from 10 Mkm^2 for 0-0.2 µm to 49 Mkm^2 for 20-180 µm) and should decrease by 36% on average (from -20 % for 5-20 µm up to -54% for 0.8-5 µm, -12 million km^2 on average, +6 % for 0.22-3 µm). Tropical provinces cover ~118 million km^2 on average (from 86 Mkm^2 for 0.8-5 µm up to 169 Mkm^2 for 180-2000 µm) and their coverage should increase by 32% on average (from +13% for 0-0.2 µm up to +75%

188 for 0.8-5 µm, +25 million km^2 on average) (Supplementary Fig. 17 and Supplementary
189 Table 1).
190 We calculate at each grid point a single dissimilarity index (*Materials and Methods*)
191 between probabilities of future and present *dominant* provinces for all size fractions
192 combined (Fig. 4a). Areas located between future and present borders of expanding and
193 shrinking provinces would be the most subject to replacements by other contemporary
194 provinces, as exemplified by the poleward retraction of the southern/northern edges of the
195 temperate provinces (red arrows on Fig. 4a, Table 1). High dissimilarities are obtained
196 over northern (25° to 60°) and symmetrically southern (-25 to -60°) temperate regions
197 (mean of 0.29 and 0.24 respectively). Despite important environmental changes in austral
198 and equatorial regions (Supplementary Fig. 11) and projected change in diversity[31–33] and
199 biomass[45], the contemporary provinces remain the most probable at the end of the century
200 using our statistical models (mean dissimilarities of 0.18 and 0.02 respectively) as no
201 known contemporary provinces could replace them.
202 To further study the decoupling between provinces of different size fractions in the future
203 we considered the assemblages of *dominant* provinces of each size fraction. By using two
204 differently stringent criteria, from 45.3 to 57.1% of ocean surface, mainly located in
205 temperate regions, would be inhabited in 2090-99 by assemblages that exist elsewhere in
206 2006-15 (Fig. 4b versus Fig. 4c). Contemporary assemblages would disappear on 3.5 to
207 3.8% of the surface, and, conversely, novel assemblages, not encountered today, would
208 cover 2.9 to 3.0% of the surface. These areas appear relatively small but they include some
209 important economic zones (Fig 4b, Supplementary Fig. 18). On 41.8% to 51.8% of the
210 surface of the main fisheries and 41.2% to 54.2% of Exclusive Economic Zones (*Materials*

211 *and Methods*), future assemblages would differ from those present today (Supplementary
212 Fig. 18).

213 **Drivers of plankton biogeography reorganization**

214 We quantify the relative importance of considered environmental parameters
215 (temperature, salinity, dissolved silica, macronutrients and seasonality of nitrate) into
216 niche definition and in driving future changes of the structure of plankton biogeography.
217 Among environmental properties that define the niches, temperature is the first influential
218 parameter (for 19 niches out of 27) but only at 22.6% on average. Particularly, the
219 distribution of relative influences of temperature is spread over a much wider range than
220 that of other parameters (Supplementary Fig. 18a).

221 The relative impact of each environmental parameter is calculated[24] for each site
222 presenting a significant dissimilarity between 2006-15 and 2090-99 (Fig 5a). Overall, SST
223 would be responsible for reorganization of half of the provinces followed by Phosphate (11
224 %) and Salinity (10.3%) (Supplementary Fig. 20). SST is the primary driver over the
225 majority of the ocean (Fig. 5a). In some regions, salinity (e.g. eastern North Atlantic) and
226 Phosphate (e.g. equatorial region) dominate (Fig. 5a) and excluding the effect of SST, they
227 are the primary drivers of global reorganization of the provinces (Fig. 5 b). The impact of
228 SST varies across size classes with a significantly higher contribution in large size classes
229 (>20 μm) compared to the small ones (mean of ~73% versus ~49%; Fig. 5c). Even though
230 the contribution of combined nutrients to niche definition is similar for small and large size
231 classes under present day conditions (mean of ~56% versus ~61%, Supplementary Fig. 19,
232 Supplementary Table 3), their future projected variations have a higher relative impact on
233 the reorganization of small organisms' biogeographies (mean of ~39% versus ~20 %, t-test

234 p-value < 0.05, Supplementary Fig. 19, Supplementary Table 3). For instance, in the tropical
235 zone, the shrinkage of the equatorial province (province C9, size fraction 0.8 – 5 μm, Fig.
236 2b,d, Supplementary Fig. 16e) is driven at 24 % by reduction of dissolved phosphate
237 concentrations and at 25% by SST increase. In contrast, SST and Salinity would drive
238 respectively at 56% and 16% the shrinkage of the temperate province F5 of size fraction
239 180-2000 μm (Supplementary Fig. 16d versus e). Non-poleward shifts are found only
240 within small size fractions (<20 μm) (Supplementary Fig. 14, 15) highlighting differential
241 responses to nutrients and SST changes between large and small size classes, the latter
242 being enriched in phytoplankton that directly rely on nutrient supplies.

243 **Discussion**

244 We propose a novel partitioning of the ocean based on metagenomic sampling,
245 complementing pioneer and recent efforts based on other bio-physico-chemical data[9–11].
246 Though they are initially built at genomic scale, our biogeographies paradoxically reveal
247 basin scale provinces that are larger than BGCPs[11] and biomes[10], and probably relatively
248 stable across seasons evoking limited effects of seasonality on frontiers positions of BGCPs
249 provinces[11]. We propose that this apparent paradox emerges from the combination of the
250 scale, nature and resolution of sampling. First, two proximal samples from the *Tara* Oceans
251 expedition are separated by ~300 km on average sampled over three years; this relatively
252 large spatio-temporal scale overlies shorter scale compositional variations previously
253 observed[5,6]. Second, our estimates of plankton community dissimilarities are highly
254 resolutive; they are computed at genomic scale with billions of small DNA fragments[12,20]
255 thus smoothing the more discrete species level signal. Together, from these combinations

256 of processes and patterns occurring at multiple scales emerge basin scale provinces
257 associated with coherent environmental niches and signature genomes.
258 Our provinces are structured in broad latitudinal bands with smaller organisms (<20 μm)
259 displaying more complex patterns and partially decoupled from larger organisms. This
260 decoupling is the result of distinct statistical links between provinces based on organism
261 size fractions and environmental parameters and could reflect their respective trophic
262 modes[40].
263 Complex changes of the parameters defining the niches are projected under climate change
264 leading to size-dependent modifications of biogeographical patterns, as for example
265 smaller organisms being more sensitive to nutrient changes. Assuming the maintenance of
266 environmental characteristics that define provinces, climate change is projected to
267 restructure plankton provinces over approximately 50% of surface oceans south of 60°N
268 by the end of the century (Fig. 4). The largest reorganization is detected in subtropical and
269 temperate regions in agreement with other studies[32,44] and is accompanied by appearance
270 and disappearance of size-fractionated provinces' assemblages. For tropico-equatorial and
271 austral regions, out of contemporary range and novel environmental conditions are
272 projected. While some studies extrapolate important diversity and biomass changes in
273 these zones[31–33,45], here we project shifts of their boundaries and maintain their climatic
274 label. However, the present approach does not account for putative changes in community
275 composition nor the emergence of novel niches over these regions for which novel
276 environmental selection pressure is expected.
277 Overall, our projections for the end of the century do not take into account possible future
278 changes of major bio-physico-chemical factors such as dynamics of community mixing,

279 trophic interactions through transport[46], the potential dynamics of the genomes[14,16,17]
280 (adaptation or acclimation) and biomass variations[45]. New sampling in current and future
281 expeditions[47], as well as ongoing technological improvements in bio-physico-chemical
282 characterization of seawater samples[38,47,48] will soon refine functional[18,49], environmental
283 (micronutrients[50]) and phylogenetic[16,17] characterization of plankton ecosystems for
284 various biological entities (genotypes, species or communities) and spatio-temporal
285 scales[47]. Ultimately, crossing this varied information will allow a better understanding of
286 the conditions of emergence of ecological niches in the seascape and their response to a
287 changing ocean.

288 **References**

289 1. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of
290 the biosphere: Integrating terrestrial and oceanic components. *Science (80-.).* (1998).
291 doi:10.1126/science.281.5374.237

292 2. Guidi, L. *et al.* Plankton networks driving carbon export in the oligotrophic ocean.
293 *Nature* (2016). doi:10.1038/nature16942

294 3. Henson, S. A., Sanders, R. & Madsen, E. Global patterns in efficiency of particulate
295 organic carbon export and transfer to the deep ocean. *Global Biogeochem. Cycles*
296 (2012). doi:10.1029/2011GB004099

297 4. Azam, F. *et al.* The Ecological Role of Water-Column Microbes in the Sea. *Mar. Ecol.*
298 *Prog. Ser.* (1983). doi:10.3354/meps010257

299 5. Saab, M. A. abi. Day-to-day variation in phytoplankton assemblages during spring
300 blooming in a fixed station along the Lebanese coastline. *J. Plankton Res.* (1992).
301 doi:10.1093/plankt/14.8.1099

302 6. Djurhuus, A. *et al.* Environmental DNA reveals seasonal shifts and potential
303 interactions in a marine community. *Nat. Commun.* (2020). doi:10.1038/s41467-
304 019-14105-1
305 7. Kavanaugh, M. T. *et al.* Seascapes as a new vernacular for pelagic ocean monitoring,
306 management and conservation. *ICES J. Mar. Sci.* (2016). doi:10.1093/icesjms/fsw086
307 8. Steele, J. H. The ocean 'landscape'. *Landsc. Ecol.* (1989). doi:10.1007/BF00131537
308 9. Longhurst, A. R. *Ecological Geography of the Sea*. *Ecological Geography of the Sea*
309 (2007). doi:10.1016/B978-0-12-455521-1.X5000-1
310 10. Fay, A. R. & McKinley, G. A. Global open-ocean biomes: Mean and temporal variability.
311 *Earth Syst. Sci. Data* (2014). doi:10.5194/essd-6-273-2014
312 11. Reygondeau, G. *et al.* Dynamic biogeochemical provinces in the global ocean. *Global*
313 *Biogeochem. Cycles* (2013). doi:10.1002/gbc.20089
314 12. Richter, D. J. *et al.* Genomic evidence for global ocean plankton biogeography shaped
315 by large-scale current systems. *bioRxiv* 867739 (2019). doi:10.1101/867739
316 13. Dutkiewicz, S. *et al.* Dimensions of marine phytoplankton diversity. *Biogeosciences*
317 (2020). doi:10.5194/bg-17-609-2020
318 14. Hellweger, F. L., Van Sebille, E. & Fredrick, N. D. Biogeographic patterns in ocean
319 microbes emerge in a neutral agent-based model. *Science (80-.).* (2014).
320 doi:10.1126/science.1254421
321 15. Sauterey, B., Ward, B., Rault, J., Bowler, C. & Claessen, D. The Implications of Eco-
322 Evolutionary Processes for the Emergence of Marine Plankton Community
323 Biogeography. *Am. Nat.* (2017). doi:10.1086/692067
324 16. Laso-Jadart, R. *et al.* Investigating population-scale allelic differential expression in

325 wild 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
328 shape the biogeography of a global SAR11 subclade. *Elife* (2019).
329 doi:10.7554/eLife.46497
330 18. Carradec, Q. *et al.* A global ocean atlas of eukaryotic genes. *Nat. Commun.* (2018).
331 doi:10.1038/s41467-017-02342-1
332 19. Salazar, G. *et al.* Gene Expression Changes and Community Turnover Differentially
333 Shape the Global Ocean Metatranscriptome. *Cell* (2019).
334 doi:10.1016/j.cell.2019.10.014
335 20. Alberti, A. *et al.* Viral to metazoan marine plankton nucleotide sequences from the
336 Tara Oceans expedition. *Sci. Data* (2017). doi:10.1038/sdata.2017.93
337 21. Pesant, S. *et al.* Open science resources for the discovery and analysis of Tara Oceans
338 data. *Sci. Data* (2015). doi:10.1038/sdata.2015.23
339 22. Karsenti, E. *et al.* A holistic approach to marine Eco-systems biology. *PLoS Biol.*
340 (2011). doi:10.1371/journal.pbio.1001177
341 23. Duarte, C. M. Seafaring in the 21st century: The Malaspina 2010 circumnavigation
342 expedition. *Limnology and Oceanography Bulletin* (2015). doi:10.1002/lob.10008
343 24. Barton, A. D., Irwin, A. J., Finkel, Z. V. & Stock, C. A. Anthropogenic climate change
344 drives shift and shuffle in North Atlantic phytoplankton communities. *Proc. Natl.*
345 *Acad. Sci.* (2016). doi:10.1073/pnas.1519080113
346 25. Benedetti, F., Guilhaumon, F., Adloff, F. & Ayata, S. D. Investigating uncertainties in
347 zooplankton composition shifts under climate change scenarios in the Mediterranean

Sea. *Ecography (Cop.).* (2018). doi:10.1111/ecog.02434

26. Beaugrand, G. *et al.* Prediction of unprecedented biological shifts in the global ocean. *Nature Climate Change* (2019). doi:10.1038/s41558-019-0420-1

27. McMahon, K. W., McCarthy, M. D., Sherwood, O. A., Larsen, T. & Guilderson, T. P. Millennial-scale plankton regime shifts in the subtropical North Pacific Ocean. *Science (80-.).* (2015). doi:10.1126/science.aaa9942

28. Rivero-Calle, S., Gnanadesikan, A., Del Castillo, C. E., Balch, W. M. & Guikema, S. D. Multidecadal increase in North Atlantic coccolithophores and the potential role of rising CO2. *Science (80-.).* (2015). doi:10.1126/science.aaa8026

29. Beaugrand, G. Decadal changes in climate and ecosystems in the North Atlantic Ocean and adjacent seas. *Deep. Res. Part II Top. Stud. Oceanogr.* (2009). doi:10.1016/j.dsr2.2008.12.022

30. Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L. & Levin, S. A. Marine taxa track local climate velocities. *Science (80-.).* (2013). doi:10.1126/science.1239352

31. Thomas, M. K., Kremer, C. T., Klausmeier, C. A. & Litchman, E. A global pattern of thermal adaptation in marine phytoplankton. *Science (80-.).* (2012). doi:10.1126/science.1224836

32. Busseni, G. *et al.* Large scale patterns of marine diatom richness: Drivers and trends in a changing ocean. *Glob. Ecol. Biogeogr.* (2020). doi:10.1111/geb.13161

33. Ibarbalz, F. M. *et al.* Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell* (2019). doi:10.1016/j.cell.2019.10.008

34. Bopp, L. *et al.* Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models. *Biogeosciences* (2013). doi:10.5194/bg-10-6225-2013

35. Hutchinson, G. E. Concludig remarks. *Cold Spring Harb. Symp. Quant. Biol.* (1957).

36. Jones, M. C. & Cheung, W. W. L. Multi-model ensemble projections of climate change effects on global marine biodiversity. *ICES J. Mar. Sci.* (2015). doi:10.1093/icesjms/fsu172

37. Delmont, T. O. *et al.* Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* (2018). doi:10.1038/s41564-018-0176-9

38. Delmont, T. O. *et al.* Functional repertoire convergence of distantly related eukaryotic plankton lineages revealed by genome-resolved metagenomics. *bioRxiv* 2020.10.15.341214 (2020). doi:10.1101/2020.10.15.341214

39. Boyer, T. P. *et al.* WORLD OCEAN DATABASE 2013, NOAA Atlas NESDIS 72. *Sydney Levitus, Ed.; Alexey Mishonoc, Tech. Ed.* (2013). doi:10.7289/V5NZ85MT

40. Sunagawa, S. *et al.* Tara Oceans: towards global ocean ecosystems biology. *Nature Reviews Microbiology* (2020). doi:10.1038/s41579-020-0364-5

41. Moon, K. R. *et al.* Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* (2019). doi:10.1038/s41587-019-0336-3

42. van Vuuren, D. P. *et al.* The representative concentration pathways: An overview. *Clim. Change* (2011). doi:10.1007/s10584-011-0148-z

43. Marinov, I. *et al.* North-South asymmetry in the modeled phytoplankton community response to climate change over the 21st century. *Global Biogeochem. Cycles* (2013). doi:10.1002/2013GB004599

44. Polovina, J. J., Dunne, J. P., Woodworth, P. A. & Howell, E. A. Projected expansion of the subtropical biome and contraction of the temperate and equatorial upwelling

394 biomes in the North Pacific under global warming. *ICES J. Mar. Sci.* (2011).
395 doi:10.1093/icesjms/fsq198
396 45. Flombaum, P., Wang, W. L., Primeau, F. W. & Martiny, A. C. Global picophytoplankton
397 niche partitioning predicts overall positive response to ocean warming. *Nat. Geosci.*
398 (2020). doi:10.1038/s41561-019-0524-2
399 46. Iudicone, D. Some may like it hot. *Nature Geoscience* (2020). doi:10.1038/s41561-
400 020-0535-z
401 47. Gorsky, G. *et al.* Expanding Tara Oceans Protocols for Underway, Ecosystemic
402 Sampling of the Ocean-Atmosphere Interface During Tara Pacific Expedition (2016–
403 2018). *Front. Mar. Sci.* (2019). doi:10.3389/fmars.2019.00750
404 48. Istace, B. *et al.* de novo assembly and population genomic survey of natural yeast
405 isolates with the Oxford Nanopore MinION sequencer. *Gigascience* (2017).
406 doi:10.1093/gigascience/giw018
407 49. Busseni, G. *et al.* Meta-Omics Reveals Genetic Flexibility of Diatom Nitrogen
408 Transporters in Response to Environmental Changes. *Mol. Biol. Evol.* (2019).
409 doi:10.1093/molbev/msz157
410 50. Grand, M. M. *et al.* Developing autonomous observing systems for micronutrient
411 trace metals. *Frontiers in Marine Science* (2019). doi:10.3389/fmars.2019.00035

412 **Acknowledgments**

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

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

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

Competing interests

The authors declare no competing interests.

Author contributions

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

Online content

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

439 # Materials and methods

440 ## Genomic provinces of plankton

441 Environmental niches are computed for trans-kingdom plankton genomic provinces from
442 Richter et al.[12]. They consist of the clustering of metagenomic dissimilarity matrices from 6
443 available size fractions with sufficient metagenomic data from the *Tara* Oceans dataset. The
444 six size fractions (0-0.2, 0.22-3, 0.8-5, 5-20, 20-180 and 180-2000 µm) represent major
445 plankton groups. Two large size classes (180-2000 µm and 20-180 µm) are enriched in
446 zooplankton dominated by arthropods (mainly copepods) and cnidarians. They are
447 expected to directly depend on smaller eukaryotes as they feed on them. Size classes 5-20
448 µm and 0.8-5 µm are enriched in smaller eukaryotic algae, such as dynophytes (5-20 µm),
449 pelagophytes and haptophytes (0.8-5 µm). The distribution of these photoautotrophs
450 presumably depends on nutrient availability. Finally, size classes 0.22-3 µm and 0-0.2 µm
451 are respectively enriched in bacteria and viruses. Bacteria are characterized by a wide
452 range of trophisms including autotrophy (cyanobacteria), mixotrophy and heterotrophy,
453 while viruses are mainly parasites. Within each size fraction (from large to small), there are
454 respectively 8, 8, 11, 6, 6 and 8 (48 in total) provinces defined in Richter et al.[12] formed by
455 *Tara* Oceans metagenomes (644 metagenomes sampled either at the surface (SUR) or at
456 the Deep Chlorophyll Maximum (DCM) across 102 sites). The clustering of individual size
457 fractions is independent.

458 ## Genome signature of the provinces

459 We analyzed the distribution of 713 eukaryotic and 523 prokaryotic genomes[37,38] within
460 the genomic provinces. These genomes are Metagenome-Assembled Genomes (MAGs)
461 obtained from *Tara* Oceans metagenomes. For each size class, we select MAGs that are

462 present (according to a criteria defined in Delmont et al.[38]) in at least 5 samples. We
463 computed an index of presence enrichment of MAGs within provinces as the Jaccard
464 index[51], defined as follows:

465 $$J = \frac{M_{11}}{M_{11} + M_{01} + M_{10}}$$

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

471 **World Ocean Atlas data**

472 Physicochemical parameters proposed to have an impact on plankton genomic provinces[12]
473 are used to define environmental niches: sea surface temperature (SST), salinity (Sal);
474 dissolved silica (Si), nitrate (NO_3), phosphate (PO_4), iron (Fe), and a seasonality index of
475 nitrate (SI NO_3). With the exception of Fe and SI NO_3, these parameters are extracted from
476 the gridded World Ocean Atlas 2013 (WOA13)[39]. Climatological Fe fields are provided by
477 the biogeochemical model PISCES-v2[52]. The seasonality index of nitrate is defined as the
478 range of nitrate concentration in one grid cell divided by the maximum range encountered
479 in WOA13 at the *Tara* Oceans sampling stations. All parameters are co-located with the
480 corresponding stations and extracted at the month corresponding to the *Tara* Oceans
481 sampling. To compensate for missing physicochemical samples in the *Tara* Oceans *in situ*
482 data set, climatological data (WOA) are preferred. The correlation between in situ samples
483 and corresponding values extracted from WOA are high (r^2: SST: 0.96, Sal: 0.83, Si: 0.97,

484 NO_3: 0.83, PO_4: 0.89). In the absence of corresponding WOA data, a search is done within 2°
485 around the sampling location and values found within this square are averaged.
486 Nutrients, such as NO_3 and PO_4, display a strong collinearity when averaged over the global
487 ocean (correlation of 0.95 in WOA13) which could complicate disentangling their
488 respective contributions to niche definition. However, observations and experimental data
489 allow identification of limiting nutrients at regional scale characterized by specific plankton
490 communities[53]. The projection of niches into future climate would yield spurious results
491 when the present-day collinearity is not maintained[54] but there is up to now no evidence
492 for large scale changes in global nutrient stoichiometry[55].

493 **Earth System Models and bias correction**

494 Outputs from 6 Earth System Models (ESM) (Supplementary Table 2) are used to project
495 environmental niches under greenhouse gas emission scenario RCP8.5[42]. Environmental
496 drivers are extracted for present day (2006-2015) and end of century (2090-2099)
497 conditions for each model and the multi-model mean is computed. A bias correction
498 method, the Cumulative Distribution Function transform, CDFt[56], is applied to adjust the
499 distributions of SST, Sal, Si, NO_3 and PO_4 of the multi-model mean to the WOA database.
500 CDFt is based on a quantile mapping (QM) approach to reduce the bias between modeled
501 and observed data, while accounting for climate change. Therefore, CDFt does not rely on
502 the stationarity hypothesis and present and future distributions can be different. CDFt is
503 applied on the global fields of the mean model simulations. By construction, CDFt preserves
504 the ranks of the simulations to be corrected. Thus, the spatial structures of the model fields
505 are preserved.

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

507 Provinces with similar metagenomic content are retrieved from Richter et al.[12]. From a
508 total of 48 initial provinces, 10 provinces are removed either because they are represented
509 by too few samples (7 out of 10) or they are found in environments not resolved by ESMs
510 (e.g. lagoons of Pacific Ocean islands, 3 out of 10). Four machine learning methods are
511 applied to compute environmental niches for each of the 38 provinces: Gradient Boosting
512 Machine (gbm)[57], Random Forest (rf)[58], fully connected Neural Networks (nn)[59] and
513 Generalized Additive Models (gam)[60]. Hyper parameters of each technique (except gam)
514 are optimized. These are (1) for gbm, the interaction depth (1, 3 and 5), learning rate (0.01,
515 0.001) and the minimum number of observations in a tree node (1 to 10); (2) for rf, the
516 number of trees (100 to 900 with step 200 and 1000 to 9000 with step 2000) and the
517 number of parameters used for each tree (1 to 7); (3) for nn, the number of layers of the
518 network (1 to 10) and the decay (1.10^{-4} to 9.10^{-4} and 1.10^{-5} to 9.10^{-5}). For gam the number
519 of splines is set to 3, respectively 2 only when not enough points are available (for fraction
520 0-0.2, 65 points). R packages gbm (2.1.3), randomForest (4.6.14), mgcv (1.8.16) and nnet
521 (7.3.12) are used for gbm, rf, nn and gam models.
522 To define the best combination of hyper parameters for each model, we perform 30
523 random cross-validations by training the model on 75% (85 % for gbm and gam) of the
524 dataset randomly sampled and by calculating the Area Under the Curve[61] (AUC) on the
525 25% (15 % for gbm and gam) remaining points of the dataset. The best combination of
526 hyper parameters is the one for which the mean AUC over the 30 cross-validation is the
527 highest. A model is considered valid if at least 3 out of the 4 techniques have a mean AUC
528 superior to 0.65, which is the case for 27 out of the 38 provinces (Supplementary Fig. 2a). A
529 climatic annotation is given to the 27 validated niches (Supplementary Table 1). Final

530 models are trained on the full dataset and only the techniques that have a mean AUC higher
531 than 0.65 are considered to make the projections. The majority (23) of the 27 validated
532 niches is validated by all four models and 4 by only 3 models. Relative influences of each
533 parameter in defining environmental niches are calculated using the feature_importance
534 function from the DALEX R package[62] for all four statistical methods (Supplementary Fig.
535 19a). To evaluate the consistency and coherence of environmental niche models, we first
536 make global projections on the 2006-13 WOA2013 climatology. Projections are consistent
537 with sampling regions for provinces encompassing vast oceanic areas. For example, the
538 genomic province sampled in temperate Atlantic regions of size fraction 180-2000 µm is
539 projected to be present in the north and south temperate Atlantic but also other temperate
540 regions (Supplementary Fig. 4). For model training and projections, physicochemical
541 variables are scaled to have a mean of 0 and a variance of 1. For this scaling, the mean and
542 standard deviation of each WOA13 variable (+ PISCES-v2 Fe) co-localized with *Tara* Oceans
543 stations with a value available are used. This standardization procedure allows for better
544 performance of nn models. Finally, as statistical models often disagree on projection sets
545 whereas they give similar predictions on the training set (Supplementary Fig. 5, 6), we use
546 the ensemble model approach for global-scale projections of provinces[36] *i.e.* the mean
547 projections of the validated machine learning techniques.

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

549 To combine all size classes' provinces, we use the PHATE algorithm[41,63] from R package
550 phateR. This algorithm allows visualization of high dimensional data in the requested
551 number of dimensions while best preserving the global data structure[63]. We choose to train
552 PHATE separately on WOA13 projections and present day and end of century projections

including presence probabilities of *non dominant* provinces. We use 3 dimensions and set hyper parameter k-nearest neighbors (knn) and decay respectively to 1000 for WOA13 and 2000 for model data as in this case there are twice as many points. The hyper parameter knn reflects the degree to which the mapping of PHATE from high to low dimensionality should respect the global features of the data. We argue that 1000 and 2000 are good choices as it will be sufficient to have a highly connected graph, conserve global structure, allow visualization of structures of the size of the provinces (mean number of points in a province: 4867) and have a reasonable computational time. Decay is set to 20 in both cases. Then we cluster the resulting distance matrix using the k-medoïds algorithm[64] and the silhouette average width criteria[65] is used as an indicator of good fit. The silhouette criterion is maximal for 2, 3 and 4 clusters and 2 peaks are found at 7 and 14 clusters (the peak at 7 is slightly less high than the one at 14, data not shown). We choose to present the 4 and 7 clusters geographical patterns as they seem more relevant with respect to the resolutions of our environmental datasets (WOA13 and climate models). We compare the three polar clusters of the 7 cluster geographical patterns with Antarctic Circumpolar Currents fronts[66] by overlying them on the map (red lines Fig. 2h).

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

$$C_col(i) = \frac{C(i) - \min\left(C(1), C(2), C(3)\right)}{\max\left(C(1), C(2), C(3)\right) - \min\left(C(1), C(2), C(3)\right)} * 255$$

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

To compare the six size fraction provinces, the combined size class with existing biogeochemical partitions of the oceans[10,11] and with each other, we use the adjusted rand index[67] (Supplementary Fig. 7-9) and overlay their masks above our partitions. In this case, presence probabilities of *dominant* provinces are not used anymore. Instead, each ocean grid point is assigned to the *dominant* provinces or the phate clusters.

Bray-Curtis dissimilarity index

Climate change impact on global projections is calculated at each grid point as the Bray-Curtis dissimilarity index[68,69] defined as follows:

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

Where ($P_n^{present}$ and P_n^{future}) are respectively the probability of presence of the province n in present day and at the end of the century. Only the probabilities of *dominant* provinces are non-null and all others are set to zero. The mask of main fisheries[70] (chosen as the first 4 deciles) and Exclusive Economical Zones[71] is overlaid on the Bray-Curtis map.

Change in province assemblages

A province assemblage is defined as the assemblage of *dominant* provinces of each size fraction at a given grid point of the considered ocean. We consider two criteria of change in province assemblage between present day and end of the century conditions. The first one, more straightforward and less stringent, considers that a province assemblage occurs when a change of *dominant* province is found in at least one size fraction. In a more stringent way, a change of assemblage is considered significant for $BC > \frac{1}{6}$ (previous

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

Centroids and migration shifts

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

Driver analysis

To assess the relative importance of each driver in province changes, the methodology from Barton et al.[24] is adopted. For a set N of n provinces (individual provinces or all provinces together), the probability of presence of each province is recalculated for present day conditions except for driver d (from the set of drivers D) for which end of the century

condition is used ($P_n^{future\ for\ dth\ driver\ only}$). The set of driver D can be either all drivers (Fig. 5a,c) or all drivers except SST (Fig. 5b). The relative importance of driver d at a given grid point for the set N of provinces is computed as follow:

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

$RI(d)$ is computed at grid cells where $BC > \frac{1}{6}$, BC being calculated with either the set of all drivers (Fig. 5a,c) or all drivers except SST (Fig. 5b). When RI(d) is calculated for individual provinces (Fig. 5c and Supplementary Fig. 16d,e), it is computed only at grid cells where $BC > \frac{1}{6}$ and the concerned province is either *dominant* in present day and/or end of century conditions.

References

51. Jaccard, P. Distribution comparée de la flore alpine dans quelques régions des Alpes occidentales et orientales. *Bull. la Murithienne* (1902).
52. Aumont, O., Ethé, C., Tagliabue, A., Bopp, L. & Gehlen, M. PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geosci. Model Dev.* (2015). doi:10.5194/gmd-8-2465-2015
53. Moore, C. M. *et al.* Processes and patterns of oceanic nutrient limitation. *Nature Geoscience* (2013). doi:10.1038/ngeo1765
54. Brun, P., Kiørboe, T., Licandro, P. & Payne, M. R. The predictive skill of species distribution models for plankton in a changing climate. *Glob. Chang. Biol.* (2016). doi:10.1111/gcb.13274
55. Redfield, A. C. On the Proportions of Organic Derivatives in Sea Water and Their

Relation to the Composition of Plankton. in *James Johnstone Memorial Volume* 1767–192 (Liverpool Univ. Press, Liverpool, U.K., 1934).

56. Michelangeli, P. A., Vrac, M. & Loukos, H. Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophys. Res. Lett.* (2009). doi:10.1029/2009GL038401

57. Ridgeway, G. gbm: Generalized Boosted Regression Models. *R Packag. version 1.6-3.1* (2010).

58. Breiman, L. & Cutler, A. Breiman and Cutler's random forests for classification and regression. *Packag. 'randomForest'* (2012). doi:10.5244/C.22.54

59. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S Fourth edition by*. *World* (2002). doi:10.2307/2685660

60. Wood, S. N. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Am. Stat. Assoc.* (2004). doi:10.1198/016214504000000980

61. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* (2006). doi:10.1016/j.patrec.2005.10.010

62. Biecek, P. DALEX: explainers for complex predictive models. *J. Mach. Learn. Res.* **19**, 1–5 (2018).

63. Vallejos, C. A. Exploring a world of a thousand dimensions. *Nature Biotechnology* (2019). doi:10.1038/s41587-019-0330-9

64. L., K. & P., R. Clustering by means of Medoids. in *Statistical Data Analysis Based on the L1 Norm and Related Methods* (1987).

65. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of

cluster analysis. *J. Comput. Appl. Math.* (1987). doi:10.1016/0377-0427(87)90125-7

66. Orsi, A. H., Whitworth, T. & Nowlin, W. D. On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep. Res. Part I* (1995). doi:10.1016/0967-0637(95)00021-W

67. Hubert, L. & Arabie, P. Comparing partitions. *J. Classif.* (1985). doi:10.1007/BF01908075

68. Somerfield, P. J. Identification of the Bray-Curtis similarity index: Comment on Yoshioka (2008). *Marine Ecology Progress Series* (2008). doi:10.3354/meps07841

69. Bloom, S. Similarity Indices in Community Studies: Potential Pitfalls. *Mar. Ecol. Prog. Ser.* (1981). doi:10.3354/meps005125

70. Watson, R. A. A database of global marine commercial, small-scale, illegal and unreported fisheries catch 1950-2014. *Sci. Data* (2017). doi:10.1038/sdata.2017.39

71. Flanders Marine Institute (2018). Maritime Boundaries Geodatabase: Maritime Boundaries and Exclusive Economic Zones (200NM), version 10. (2018). doi:https://doi.org/10.14284/313.

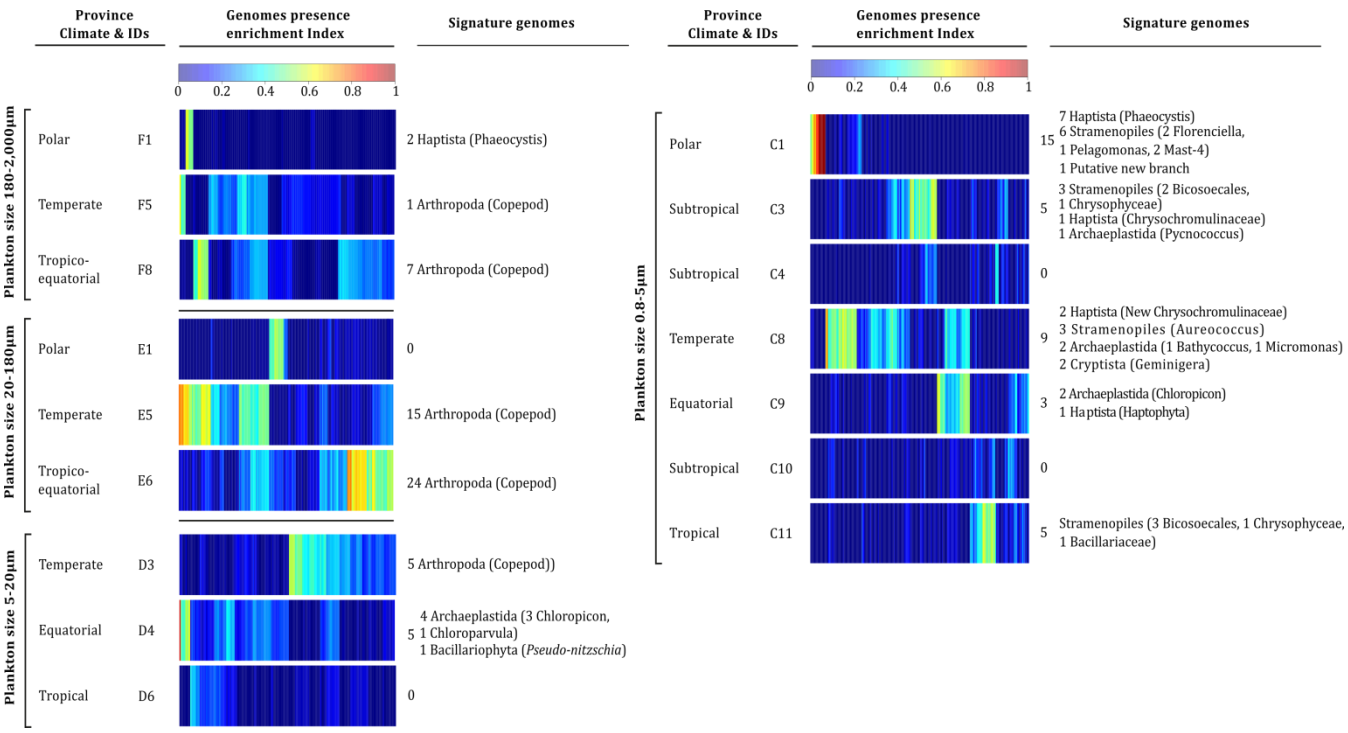


Fig. 1 | Signature genomes of provinces of eukaryotes enriched size classes. For each plankton size class, indexes of presence enrichment for 713 genomes of eukaryotic plankton[38] in corresponding provinces are clustered and represented in a color scale. Signature genomes (see *Methods*) are found for almost all provinces, their number and taxonomies are summarized (detailed list in Supplementary Table 6).

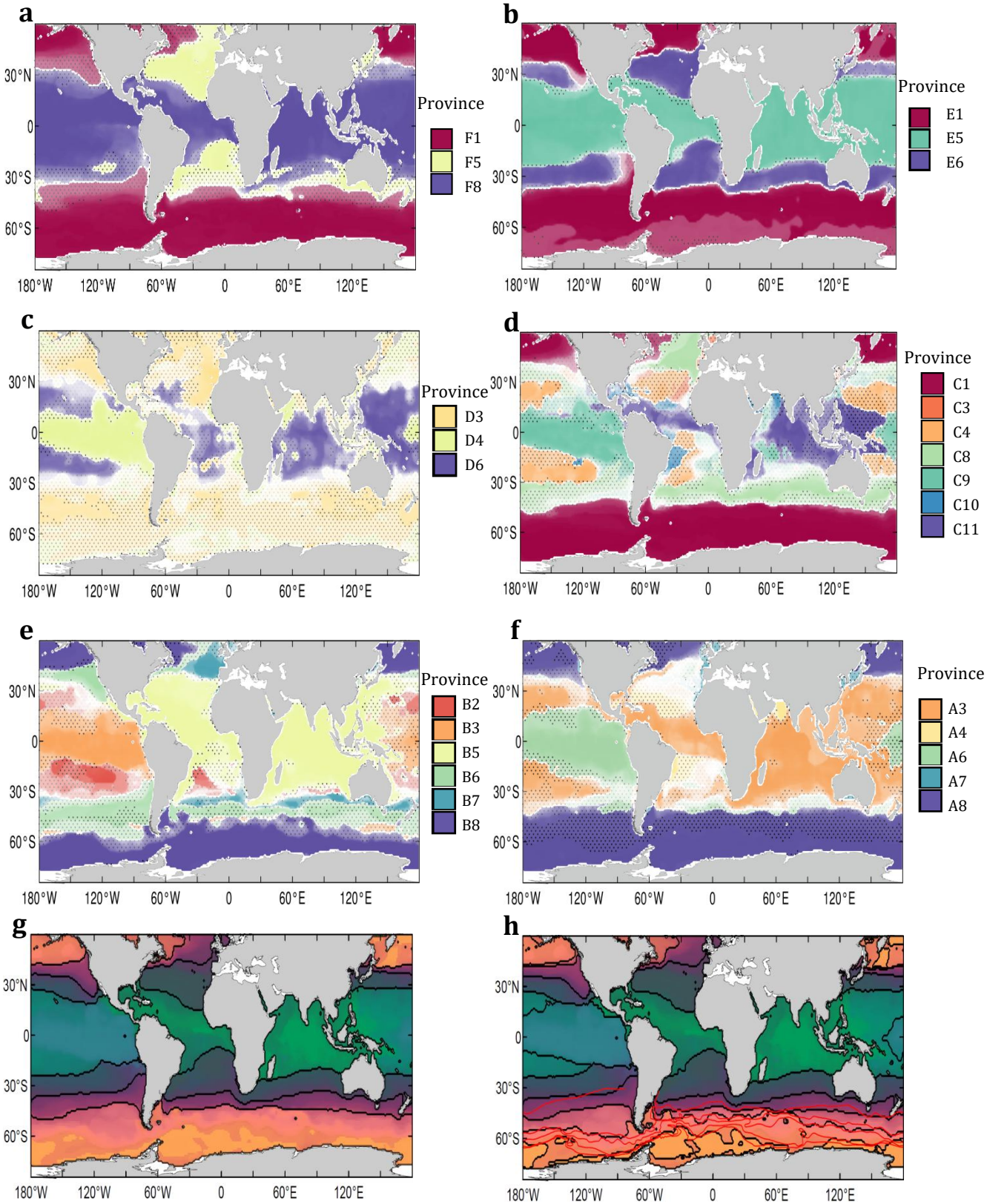


Fig. 2 | Global biogeographies of size-structured plankton provinces projected on WOA2013 dataset. (**a**) Metazoans enriched (180-2000 µm) (**b**) Small metazoans enriched (20-180 µm) (**c**) Small eukaryotes enriched (5-20 µm) (**d**) Small eukaryotes enriched (0.8-5 µm) (**e**) Bacteria enriched (0.22-3 µm) (**f**) Viruses enriched (0-0.2 µm). (**a-f**) Dotted areas represent uncertainty areas where the delta of presence probability of the dominant province and an other (from the same size fraction) is inferior to 0.5. (**g**) Combined size class biogeography using PHATE algorithm partitioned in 4 clusters. Areas of uncertainty are highlighted with dotted lines. (**h**) Combined size class biogeography using PHATE algorithm partitioned in 7 clusters overlaid with Antarctic Circumpolar Current boundaries (red). Simple biogeographies are observed in large size fractions (>20 µm) with a partitioning in three major oceanic areas: tropico-equatorial, temperate and polar. More complex geographic patterns and patchiness are observed in smaller size fractions with the distinction of pacific equatorial provinces and provinces associated to oligotrophic tropical gyres.

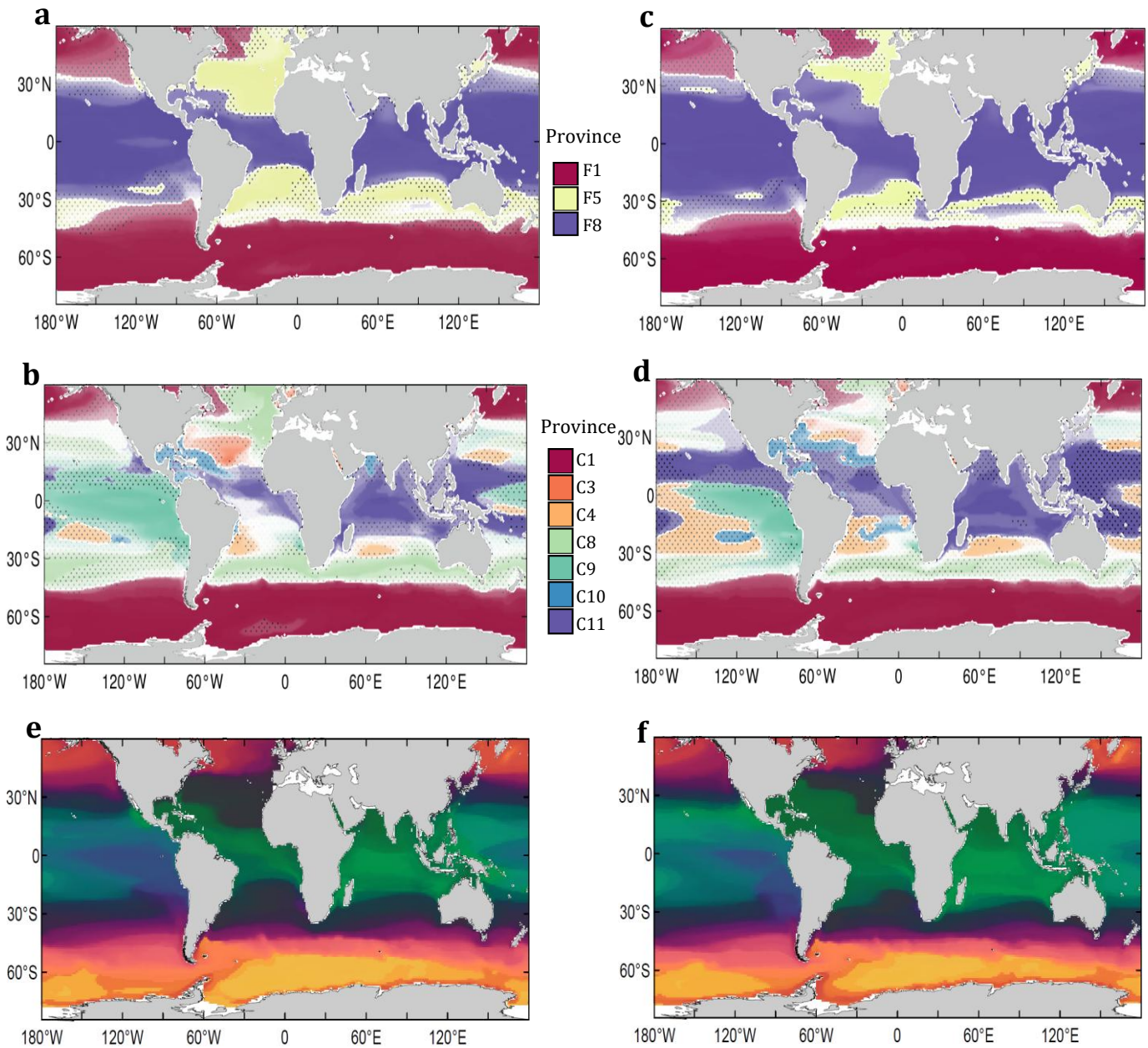


Fig. 3 | Global biogeography of small metazoans enriched size fraction (180-2000 µm), protists enriched (0.8-5 µm) and combined size classes in modeled present day (a, c, e) and end of the century (b, d, f). (**a-d**)The dominant niche *i.e.* the one predicted to have the highest probability of presence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

Size fraction (µm)	Present day covered area (%)	End of century covered area (%)	% area with a change of dominant province	Most frequent transition	%	2nd 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**				

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

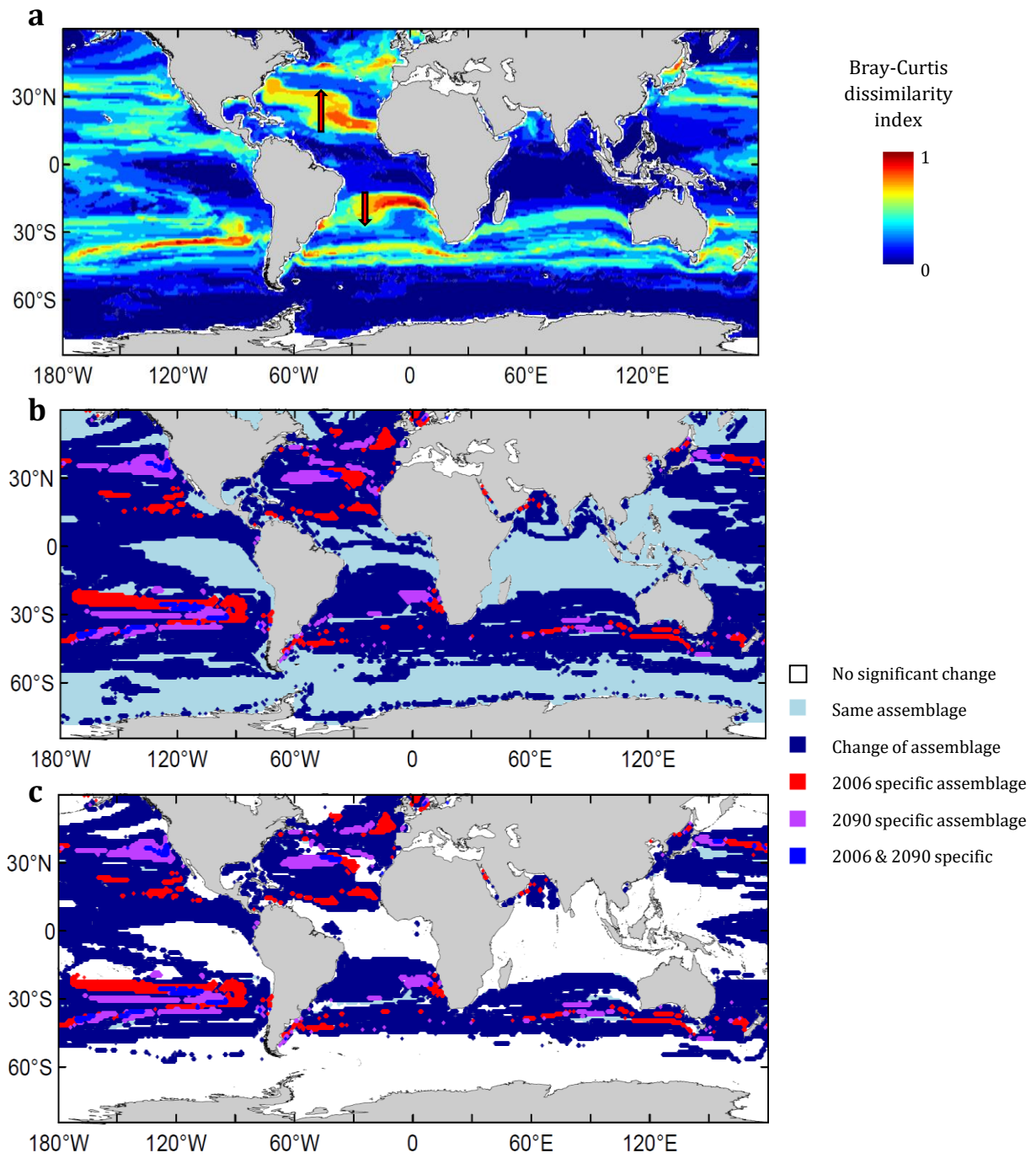


Fig. 4 | (a) Bray-Curtis dissimilarity index map comparing present day with end of the century projections of dominant provinces. Maps of trans-kingdom assemblages reorganization of dominant provinces (b) and with a criterion of significance (c). (**a**) Bray-Curtis dissimilarity index is calculated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

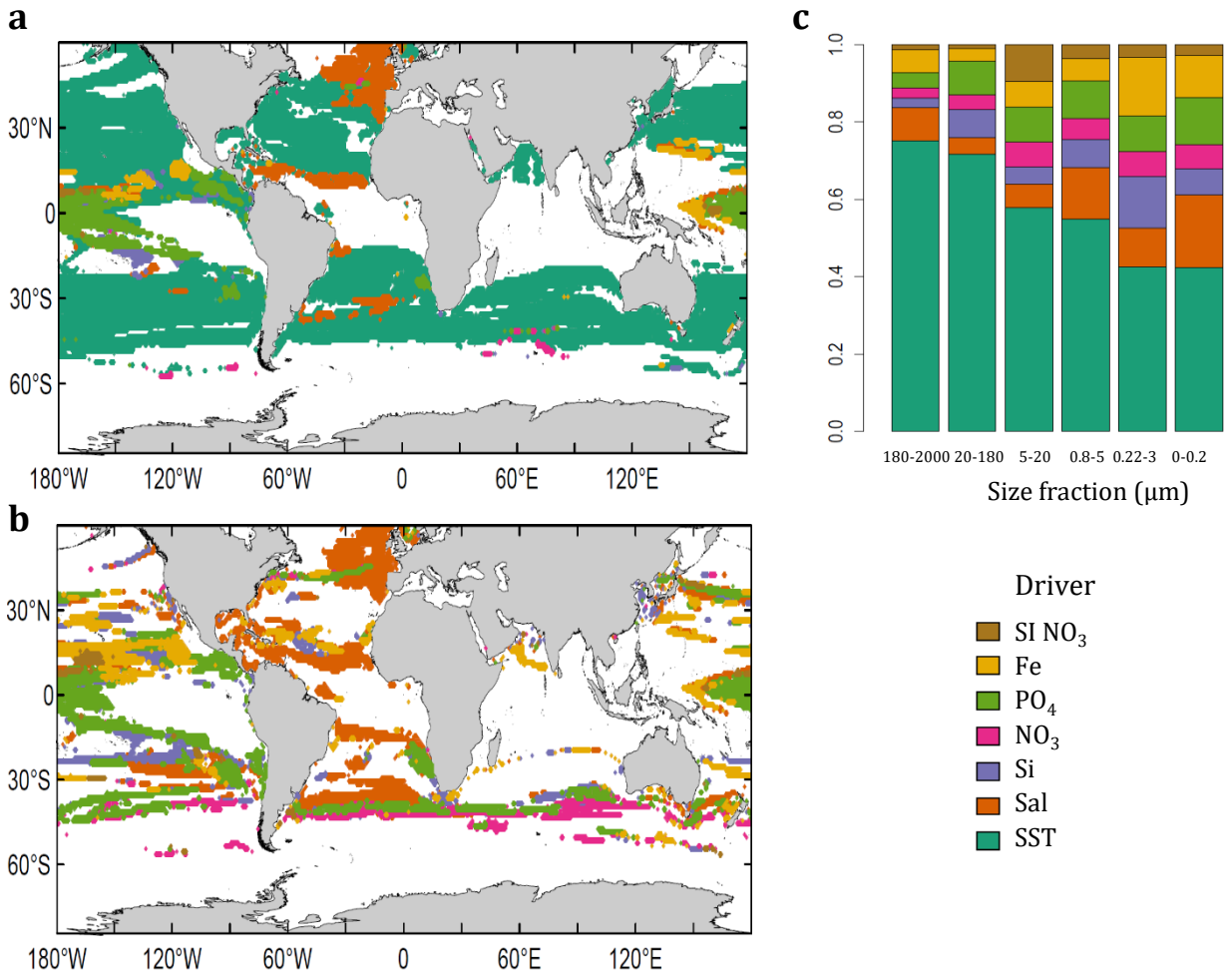
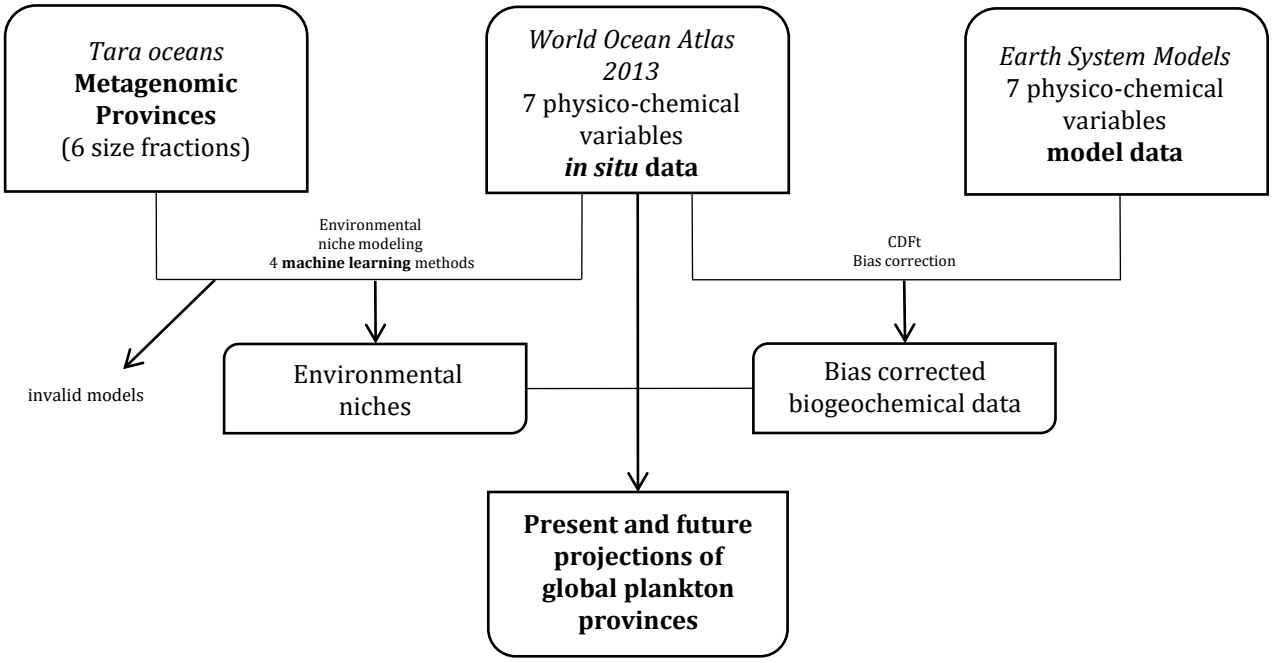
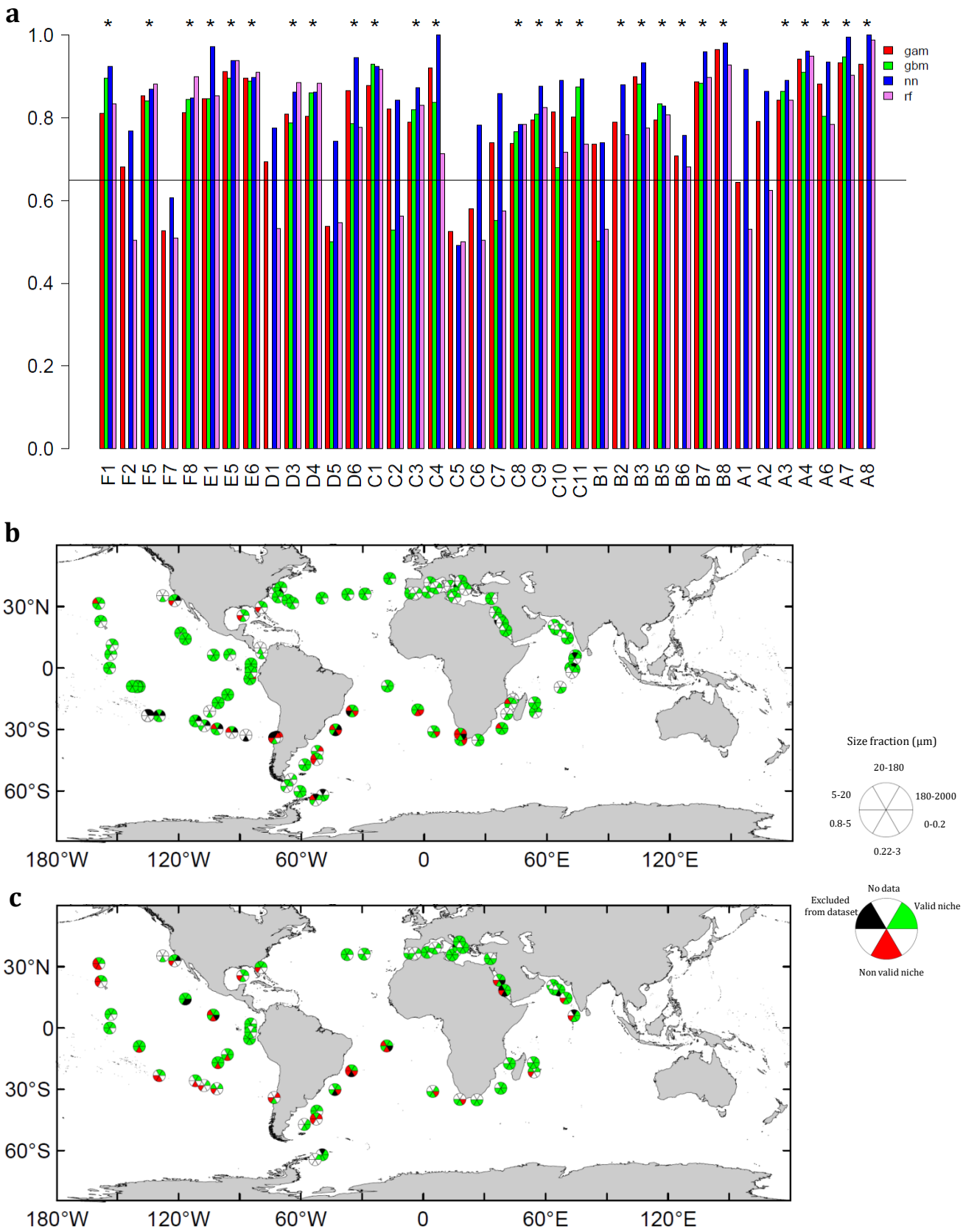


Fig. 5 | Map of most impacting drivers on dominant province changes (a), most impacting driver without considering temperature change (b) and relative importance of the drivers in the different size fractions (c). (**a**) Temperature appears as the top impacting driver on the majority of the projected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

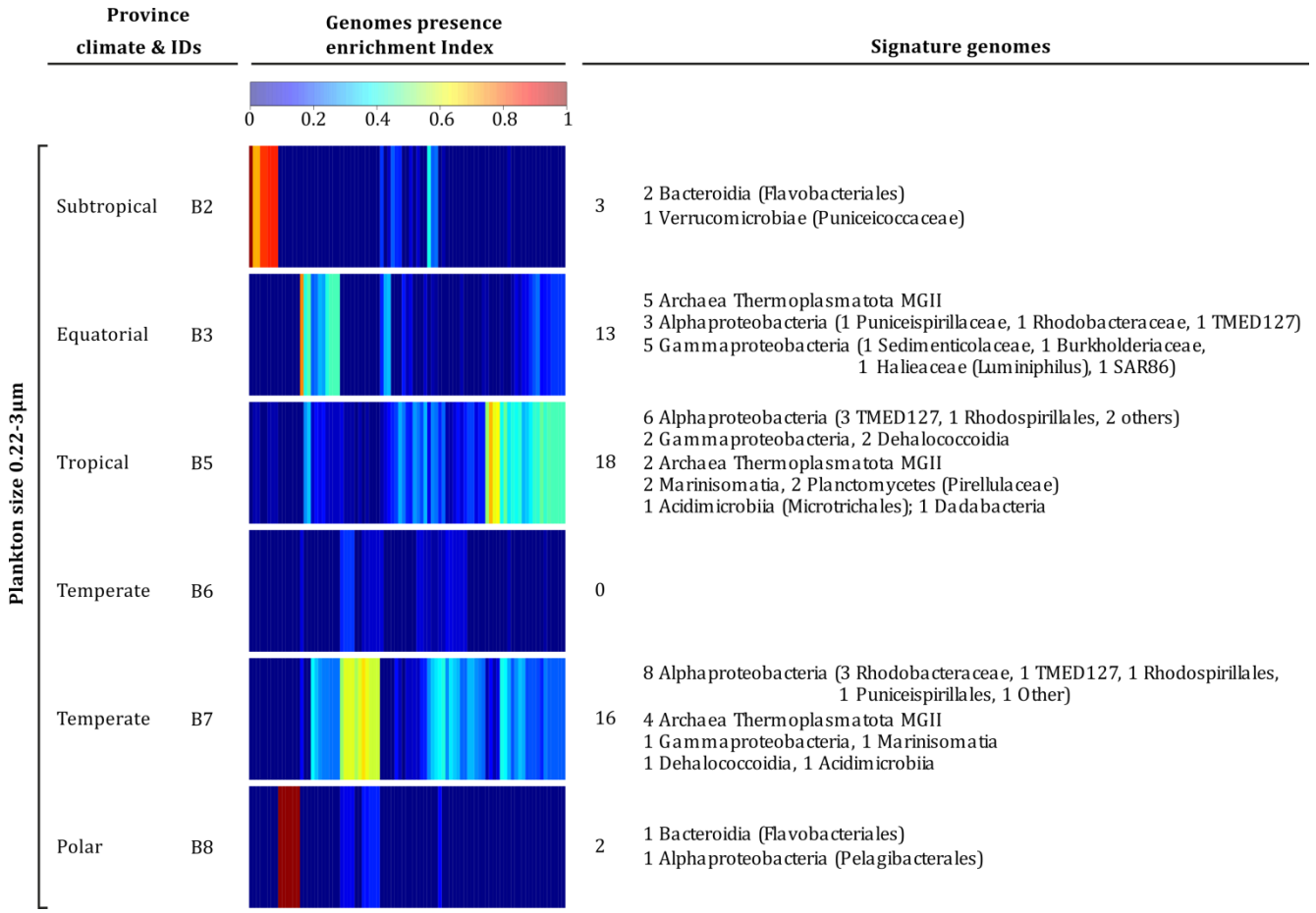


Supplementary Fig. 1 | Study pipeline. Metagenomic data from the 2009-2013 *Tara Oceans* expedition and *in situ* measurements of physicochemical variables (*World Ocean Atlas 2013*, WOA13)[1] are combined to define environmental niches at the plankton community level across 6 size fractions of the plankton realm. Bias corrected outputs from a mean model of 6 Earth System Models[2–7] and WOA13 data are then used to project global plankton provinces for present day conditions and end of the century conditions under a high emission climate change scenario (RCP8.5)[8]. Physico-chemical variables are Sea Surface Temperature (SST), Salinity (Sal), Dissolved silica (Si), Nitrate (NO_3), Phosphate (PO4), Iron (Fe) and a seasonality index of nitrate (SI NO_3).

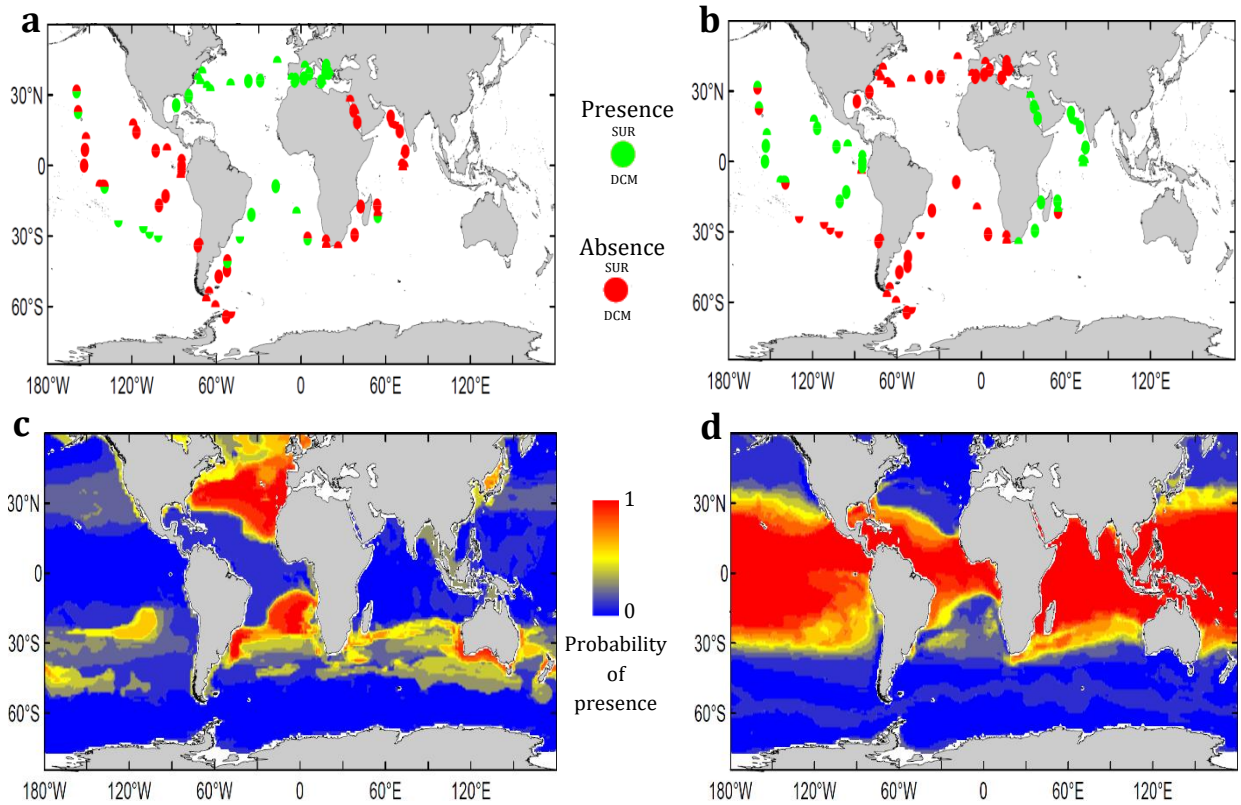


Supplementary Fig. 2 | (a) Barplot of mean AUC over 30-fold cross validation process of the 38 initial metagenomic provinces. (b,c) Map of validated and non validated stations across the six size fractions in surface samples (b) and DCM samples (c). (a) Mean AUC (Area Under the receiver operating Characteristic)[9] is plotted for the best hyperparameter combination of the 4 machine learning techniques used for each of the 38 metagenomic provinces. General Additive Models[10] (gam) are shown in red (no optimization), Gradient Boosting Machines[11] (gbm) in green, fully connected Neural Networks[12] (nn) in blue and Random Forest[13] models in purple (rf). A star for valid models is drawn at the top of each considered valid model. A model is validated when at least 3 out of the four models have a mean AUC superior to 0.65.

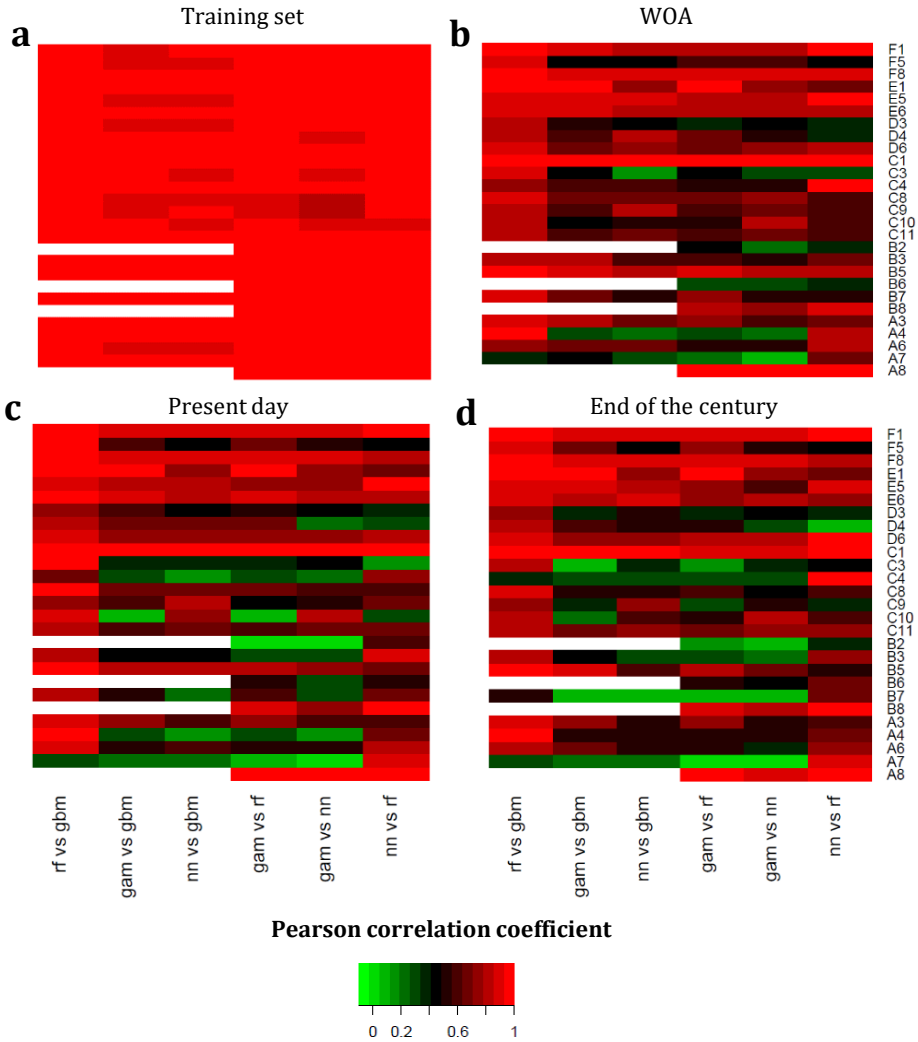
A valid model is found for 27 out of 38 initial provinces. Out of the 27 validated models 4 are not valid for the gbm method. (**b-c**) For each *Tara* sample present in the dataset at surface (**b**) or Deep Chlorophyll Maximum (DCM) (**c**) and for each size fraction, the filter (one size fraction at one location and one depth) belongs either to a validated niche (green), a non validated niche (red), an excluded filter (sea *Materials & Methods*) (black) or no data is available (white). Non validated niches represent only 11 % of filters present in the dataset (66 out of 595).



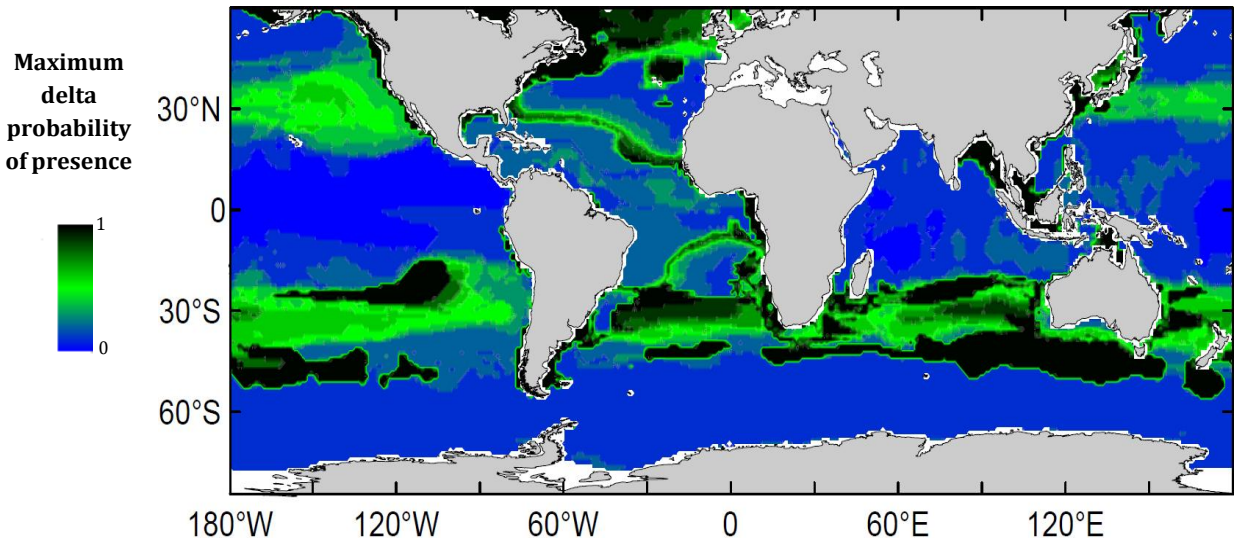
Supplementary Figure 3 | Signature genomes of provinces of the prokaryotes enriched size. Indexes of presence enrichment[14] for 523 genomes of prokaryotic plankton[15] in corresponding provinces are clustered and represented in a color scale. Signature genomes (see *Methods*) are found for almost all provinces, their number and taxonomies are summarized (detailed list in Supplementary Table 6).



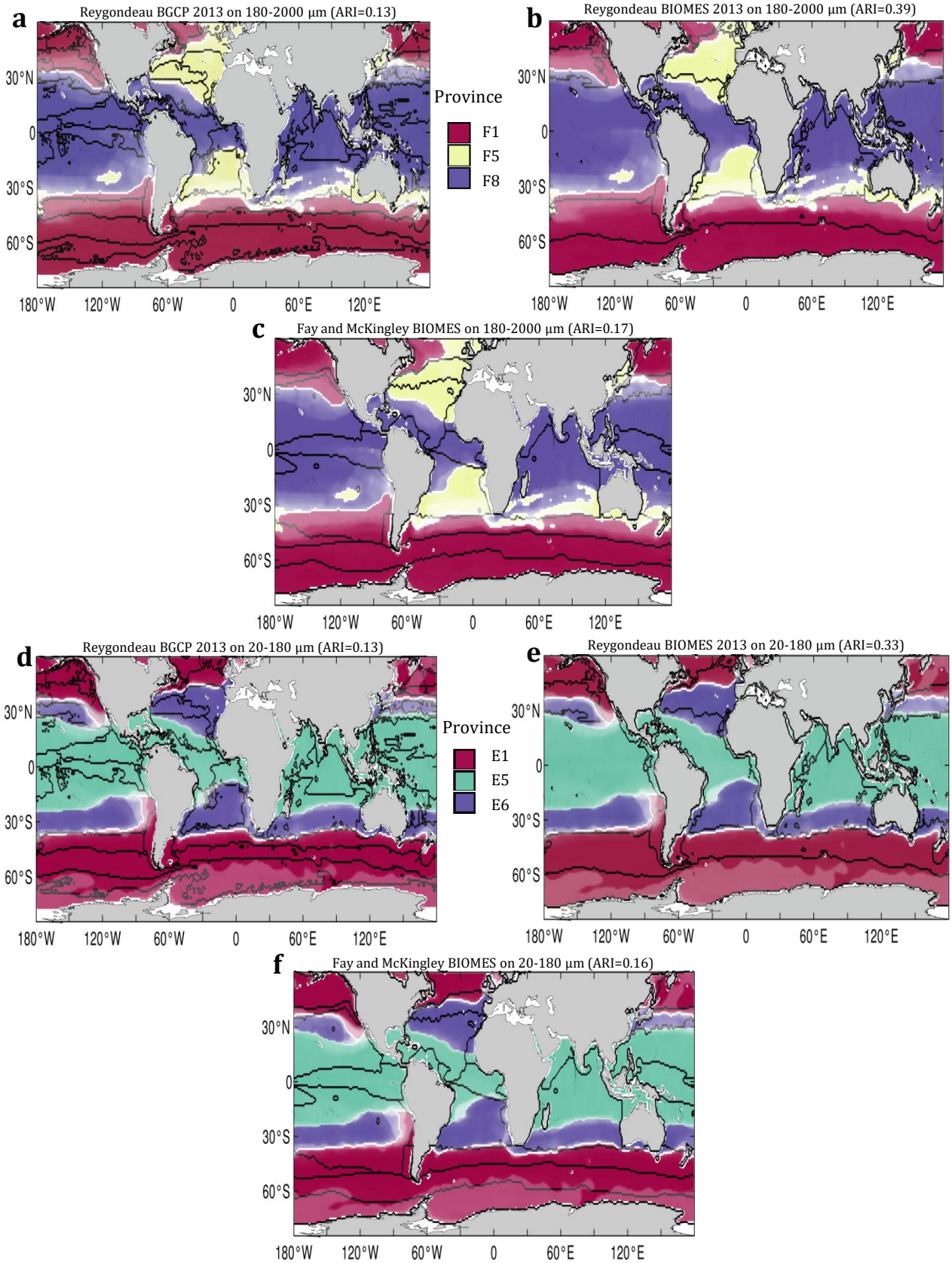
Supplementary Fig. 4 | Sampling and projection maps on WOA13 climatological data of two example provinces from size fraction 180-2000 µm. (**a**) Sampling map of province F5. (**b**) Province F8. (**c**) Projection map of province F5 on WOA13[1]. (**d**) Province F8. Qualitatively, projection maps are coherent with sampling maps of the two provinces with the highest probability of presence projected in the sampling regions. Other presence zones are also predicted by the projection. Sampling of these zones might be interesting to confirm our approach and projections such as South of Australia where a high probability of presence for province F5 is predicted.

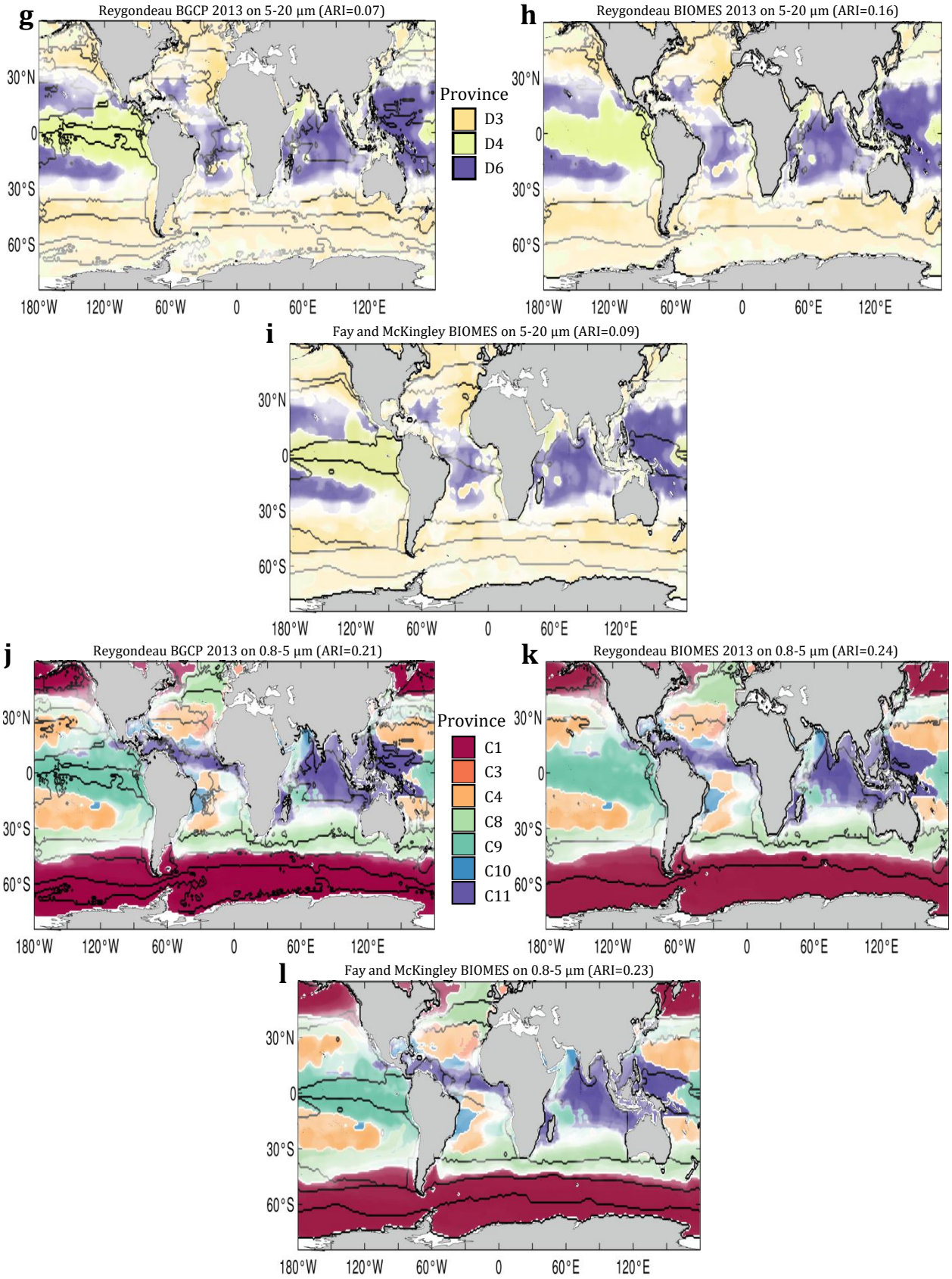


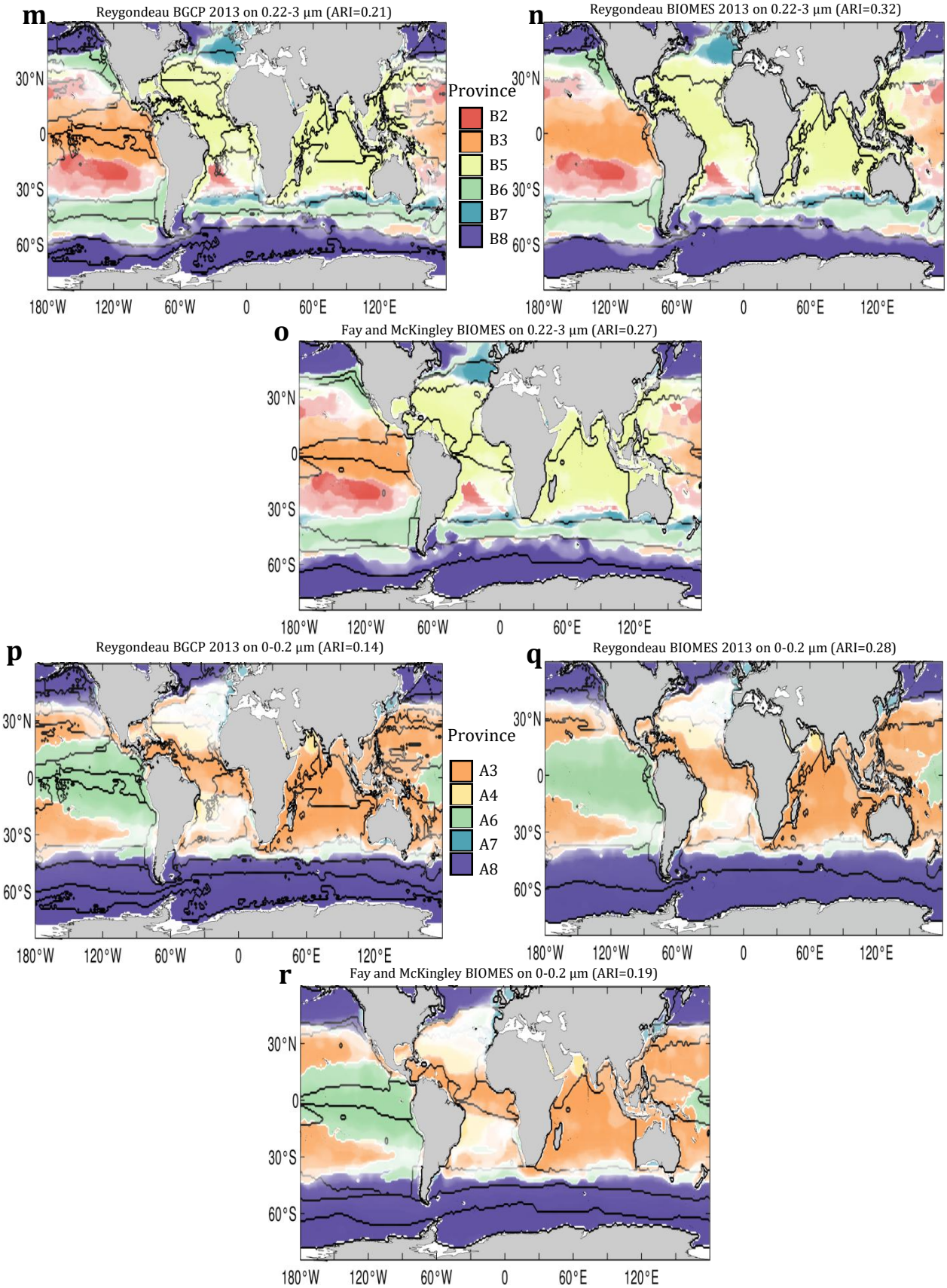
Supplementary Fig. 5 | Pairwise Pearson correlation coefficient heat maps between outputs of the 4 machine learning models. rf: Random Forest, gbm: gradient boosting machines, gam: general additive models, nn: neural networks. In rows are the provinces of the different size fractions. In columns are the pairwise comparisons of each machine learning technique. (**a**) Training set outputs. (**b**) WOA13 average data (except for Iron, PISCES-v2[16] is used) projection outputs. (**c**) Present day data projection outputs (bias corrected mean model of 6 Earth System Models). (**d**) End of the century projection outputs. On the training set, models are in agreement with most of the correlation coefficients superior to 0.9. A drop of correlation is observed for modeled data (**c**, **d**) especially in small size fractions. This shows modeled data are more distant from the training set than WOA data. Random Forest and Gradient Boosting Machine are in very good agreement (first columns) which could be expected as they are both based on multiple decision trees.



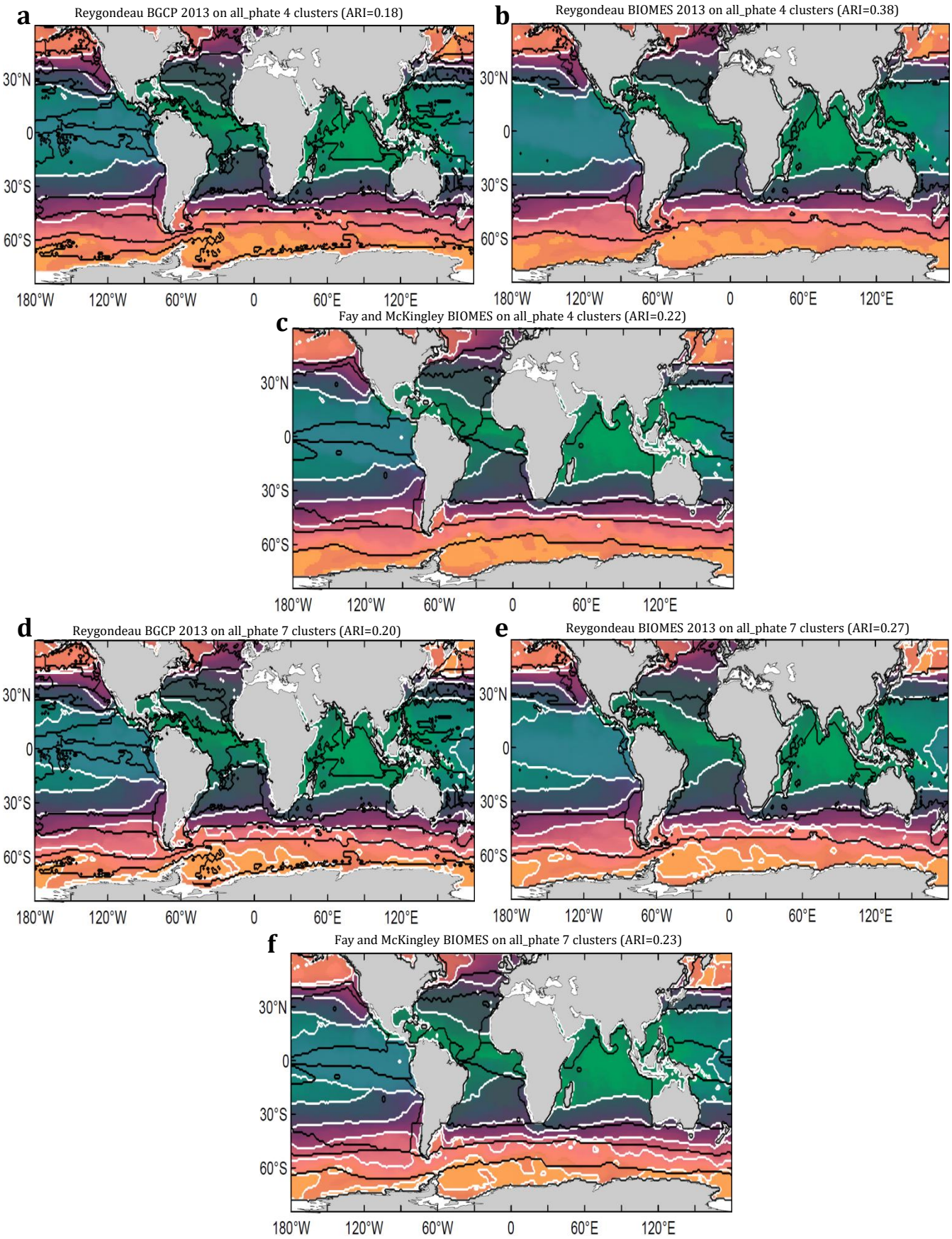
Supplementary Fig. 6 | Probability range map (WOA data) of province F5 of size fraction 180-2000 µm. At each point of the grid, the maximum delta probability of presence between the 4 machine learning projections is calculated. In black are the zones where two models completely disagree: one model predicts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


a
Reygondeau BGCP 2013 on 180-2000 µm (ARI=0.13)
b
Reygondeau BIOMES 2013 on 180-2000 µm (ARI=0.39)
Province
F1
F5
F8
c
Fay and McKingley BIOMES on 180-2000 µm (ARI=0.17)
d
Reygondeau BGCP 2013 on 20-180 µm (ARI=0.13)
e
Reygondeau BIOMES 2013 on 20-180 µm (ARI=0.33)
Province
E1
E5
E6
f
Fay and McKingley BIOMES on 20-180 µm (ARI=0.16)
30°N
0
30°S
60°S
180°W
120°W
60°W
0
60°E
120°E

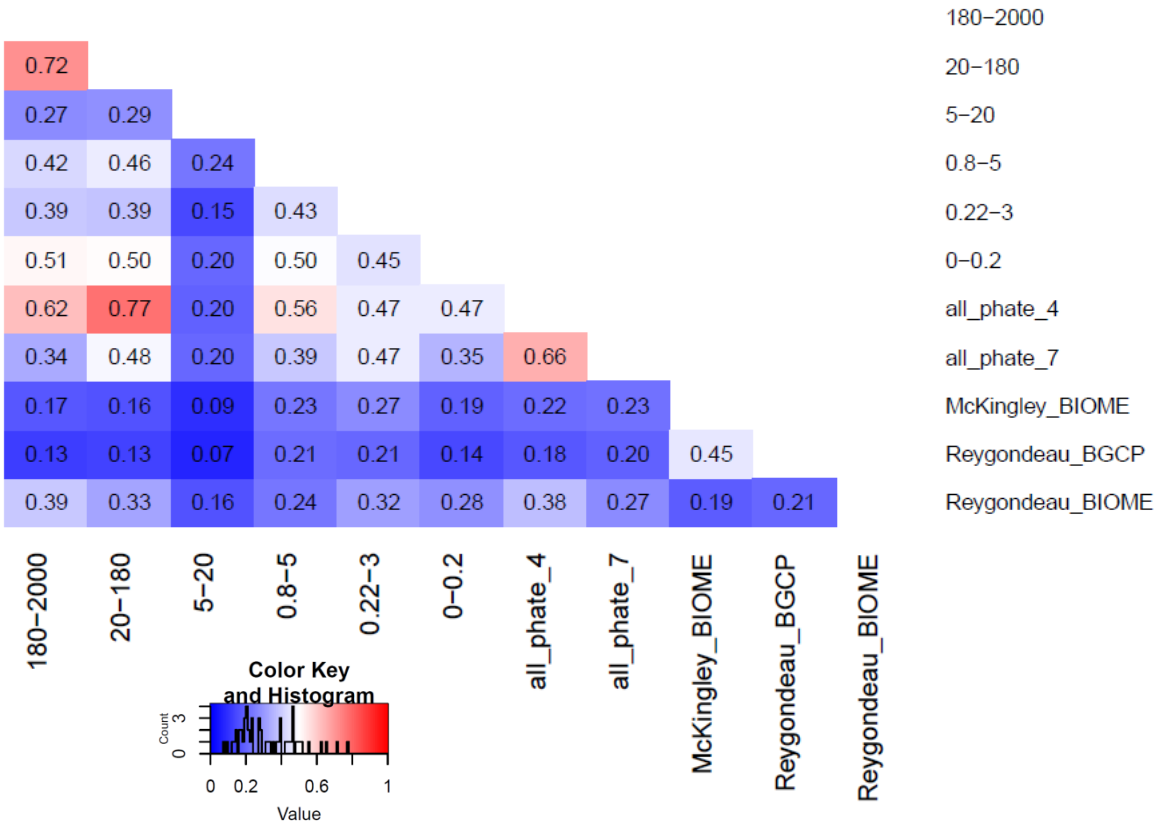

g
Reygondeau BGCP 2013 on 5-20 µm (ARI=0.07)
h
Reygondeau BIOMES 2013 on 5-20 µm (ARI=0.16)
Province
D3
D4
D6
i
Fay and McKingley BIOMES on 5-20 µm (ARI=0.09)
j
Reygondeau BGCP 2013 on 0.8-5 µm (ARI=0.21)
k
Reygondeau BIOMES 2013 on 0.8-5 µm (ARI=0.24)
Province
C1
C3
C4
C8
C9
C10
C11
l
Fay and McKingley BIOMES on 0.8-5 µm (ARI=0.23)
30°N
0
30°S
60°S
180°W
120°W
60°W
0
60°E
120°E



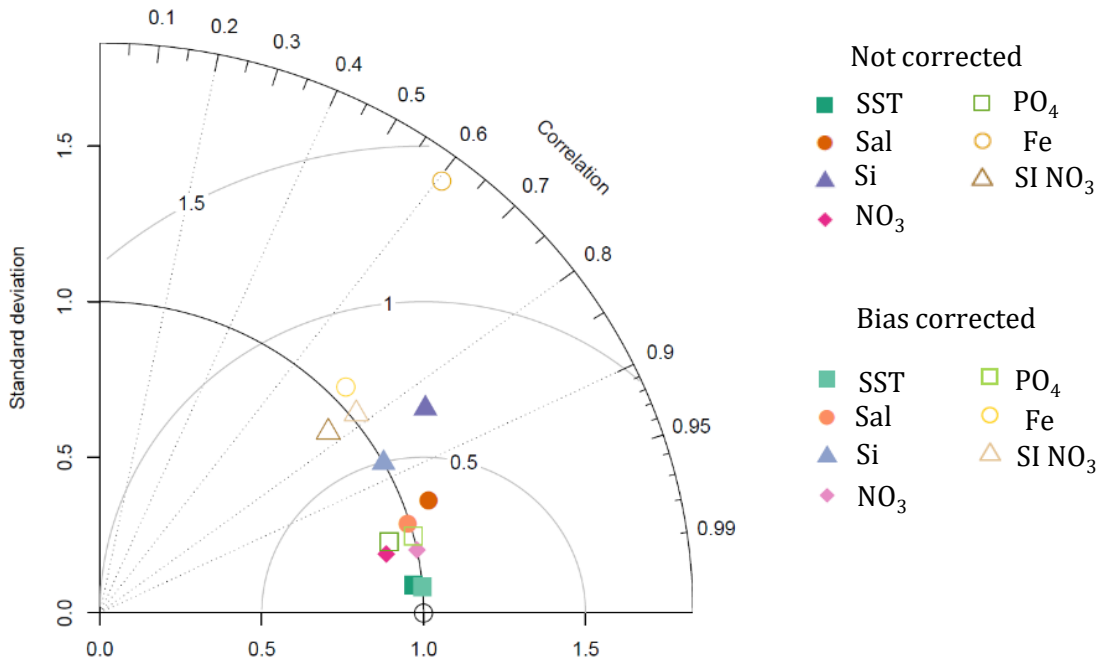
Supplementary Fig. 7 | Three existing oceanic partitions overlaid on top of plankton provinces for the six fractions and combined size fractions. The three oceanic partitions are Reygondeau et al. Biogeochemical Provinces (BGCP)[17]; Reygondeau et al. Biomes[17]; Fay and McKingley Biomes[18]. Each partitioning mask is overlaid in the above order on top of plankton 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 µm(**m-o**) 0.22-3 µm and (**p-r**) 0-0.2 µm. Above all maps, the adjusted rand index (ARI), an index used to compare partitions, comparing mapped biogeographies with the black lines mask is shown.



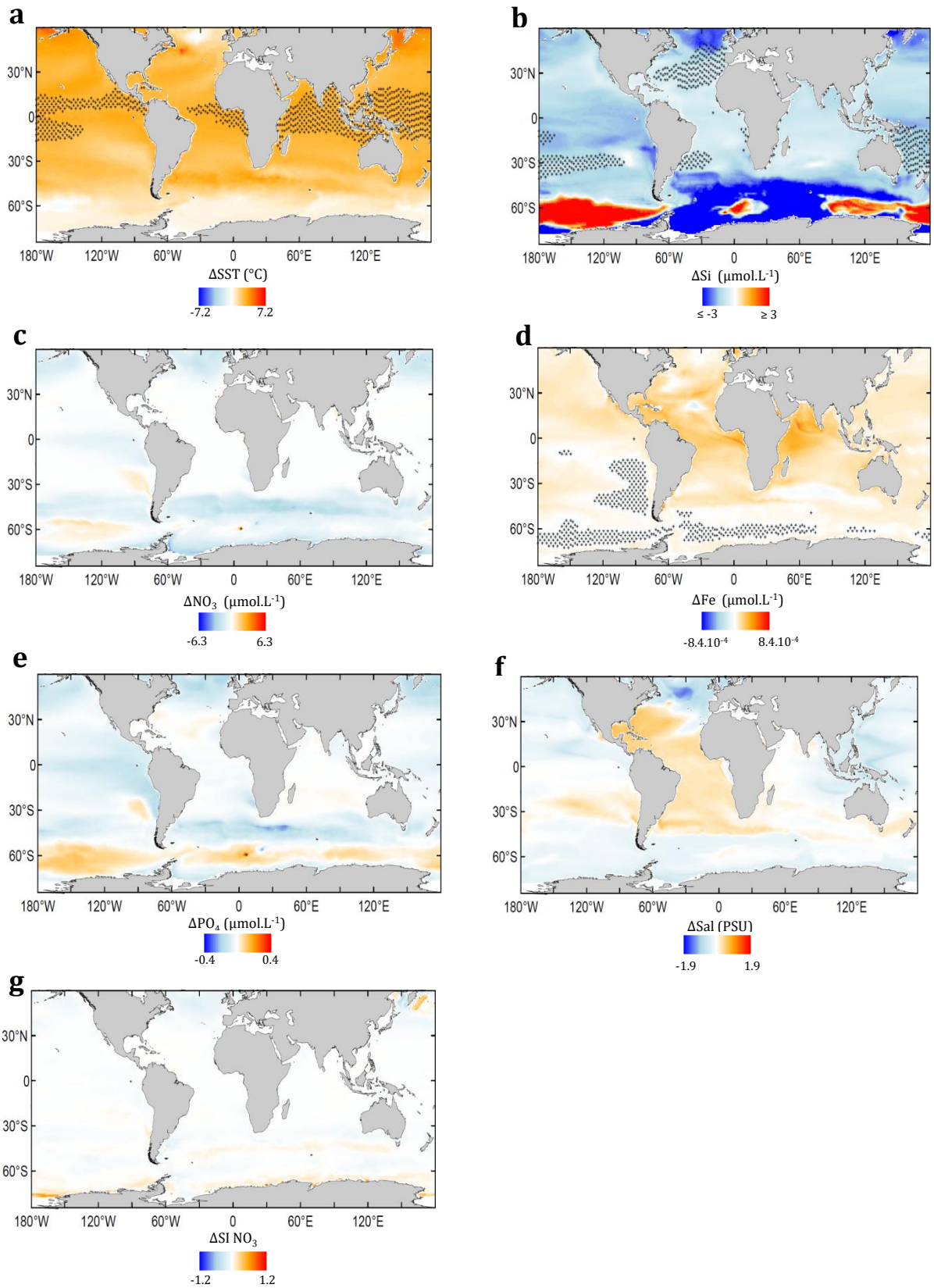
Supplementary Fig. 8 | Three existing oceanic partitions overlaid on top of plankton provinces for the combined size fractions. The three oceanic partitions are Reygondeau et al.[17] Biogeochemical Provinces (BGCP); Reygondeau et al. Biomes[17]; Fay and McKingley Biomes[18]. Each partitioning mask is overlaid in the above order on top of size fraction plankton provinces combined by the PHATE algorithm in (**a-c**) 4 clusters and (**d-f**) 7 clusters.



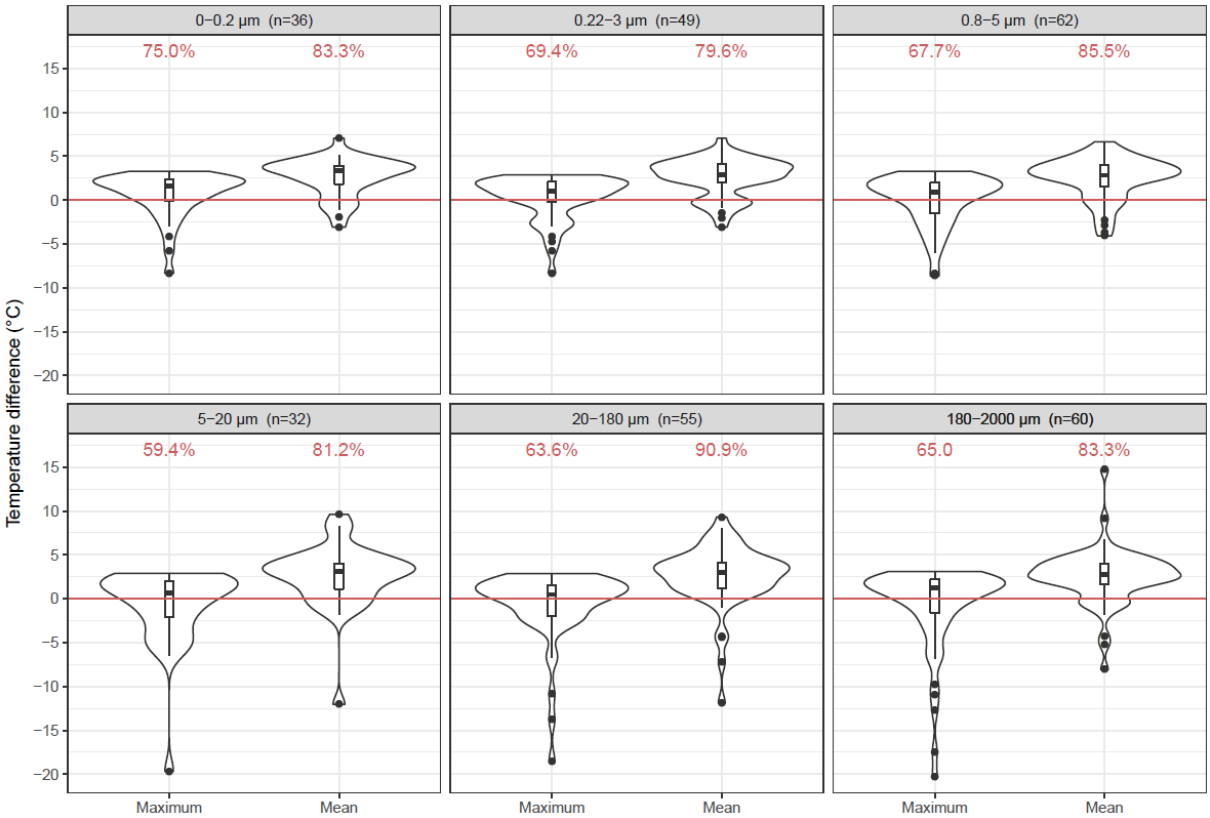
Supplementary Fig. 9 | Pairwise comparisons of ocean partitions based on plankton provinces and existing biogeochemical partitions. The three oceanic partitions are Reygondeau et al.[17] Biogeochemical Provinces (BGCP); Reygondeau et al. Biomes[17]; Fay and McKingley Biomes[18].



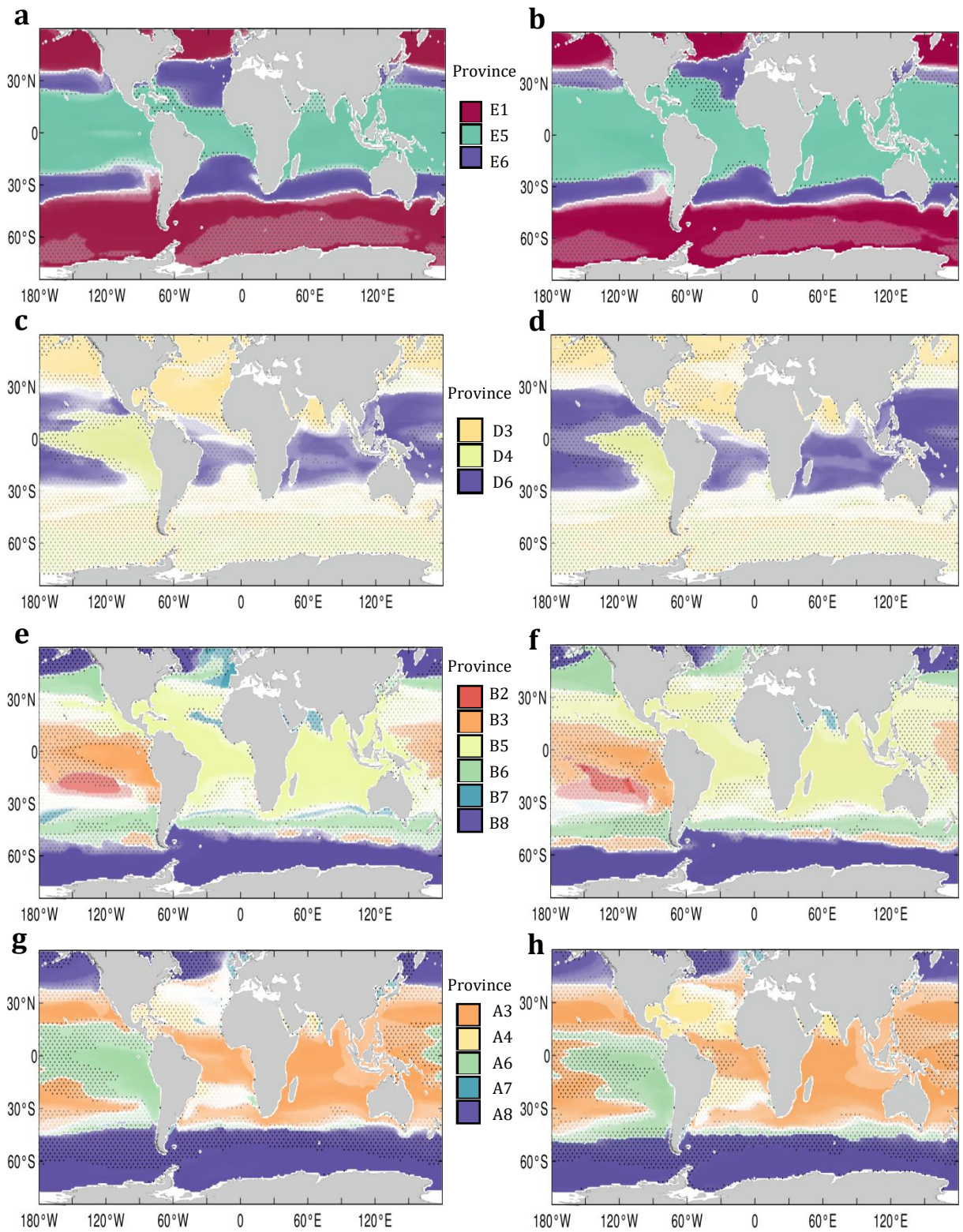
Supplementary Fig. 10 | Taylor diagram exhibiting statistical comparison of WOA 2013 observations and present day ESM model mean of the different drivers. Each variable is centered and scaled according to the mean and standard deviation of the observed variable (black circle point at standard deviation 1 on the x-axis). Dark color points are the ESM model mean without Cumulative Distribution Function transform (CDFt[19]) bias correction. They are to be compared with light color points for which the bias correction is performed. Overall, good spatial correlations are found between the model mean and the observations. CDFt bias correction performs well by bringing standard deviations of the model to the observed standard deviations without decreasing correlations.



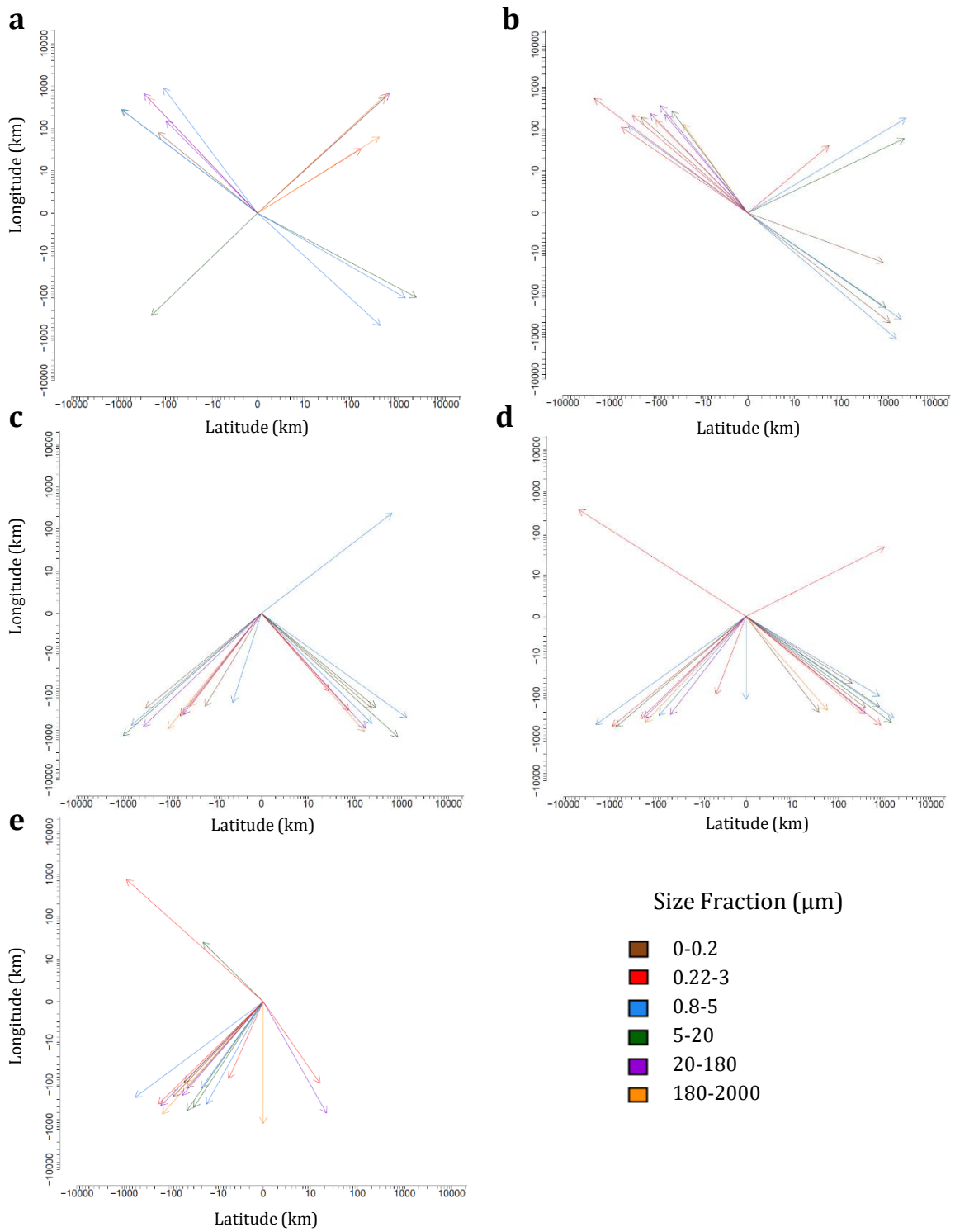
Supplementary Fig. 11 | Differences in drivers' intensity (2090/99-2006/15) in the bias corrected ESM model-mean under RCP8.5. (**a**) Sea Surface Temperature (SST). (**b**) Dissolved silica. (**c**) Nitrate. (**d**) Iron. (**e**) Phosphate. (**f**) Salinity (Sal). (**g**) Seasonality Index of Nitrate (SI NO_3). Note that the scale for dissolved silica variations is restricted to visualize small variations. Regions for which out of range values (*i.e.* inferior to the minimum or superior to the maximum found in 2006-15) are reached at the end of the century are highlighted with small stars reflecting high uncertainty zones for machine learning approaches.



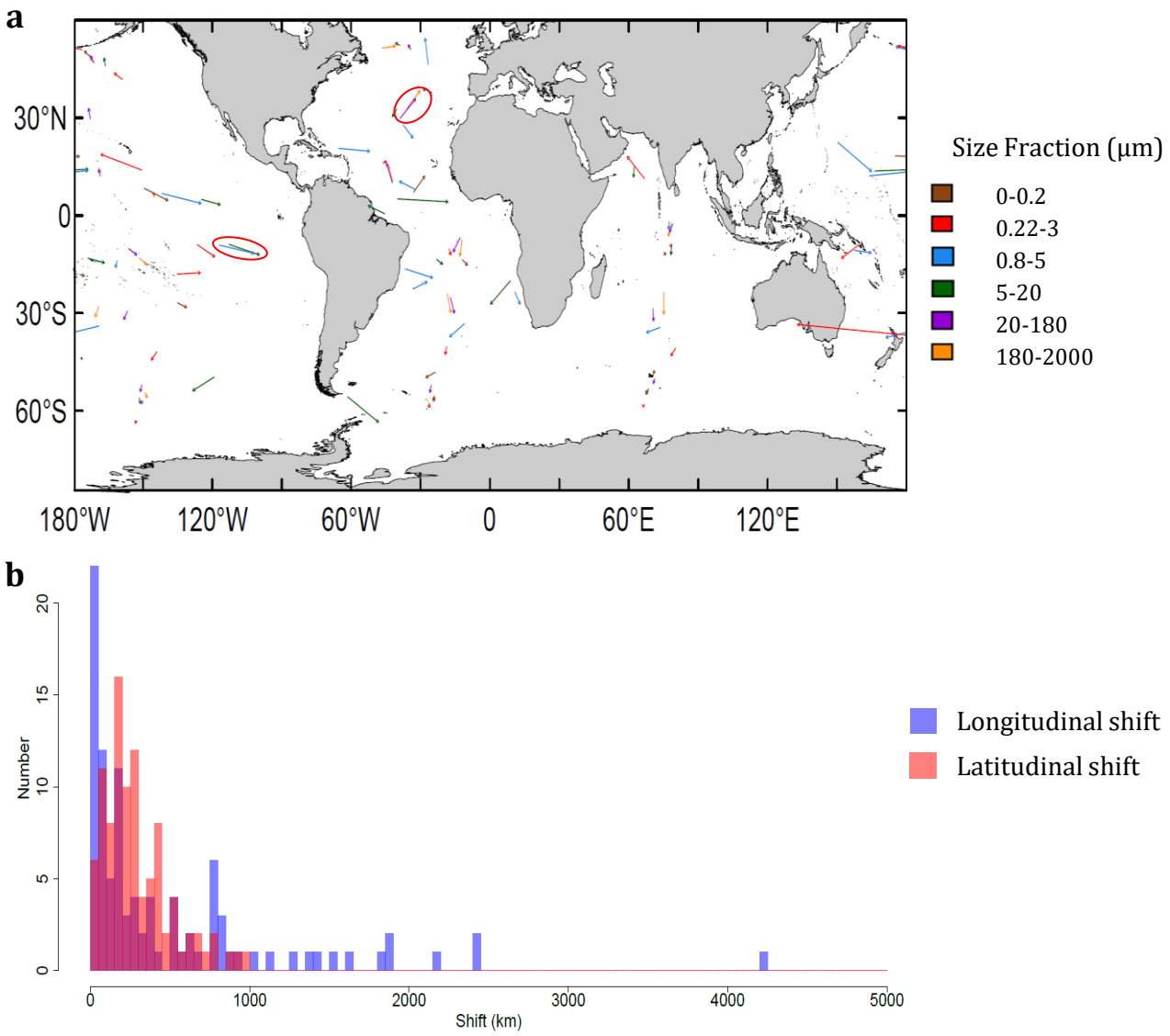
Supplementary Fig. 12 | Distribution of deltas between future temperature at each sampling site minus either the mean or maximum temperature within their contemporary genomic province. For most of the sites and across size fractions the future temperature projected by the bias adjusted ESM ensemble model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



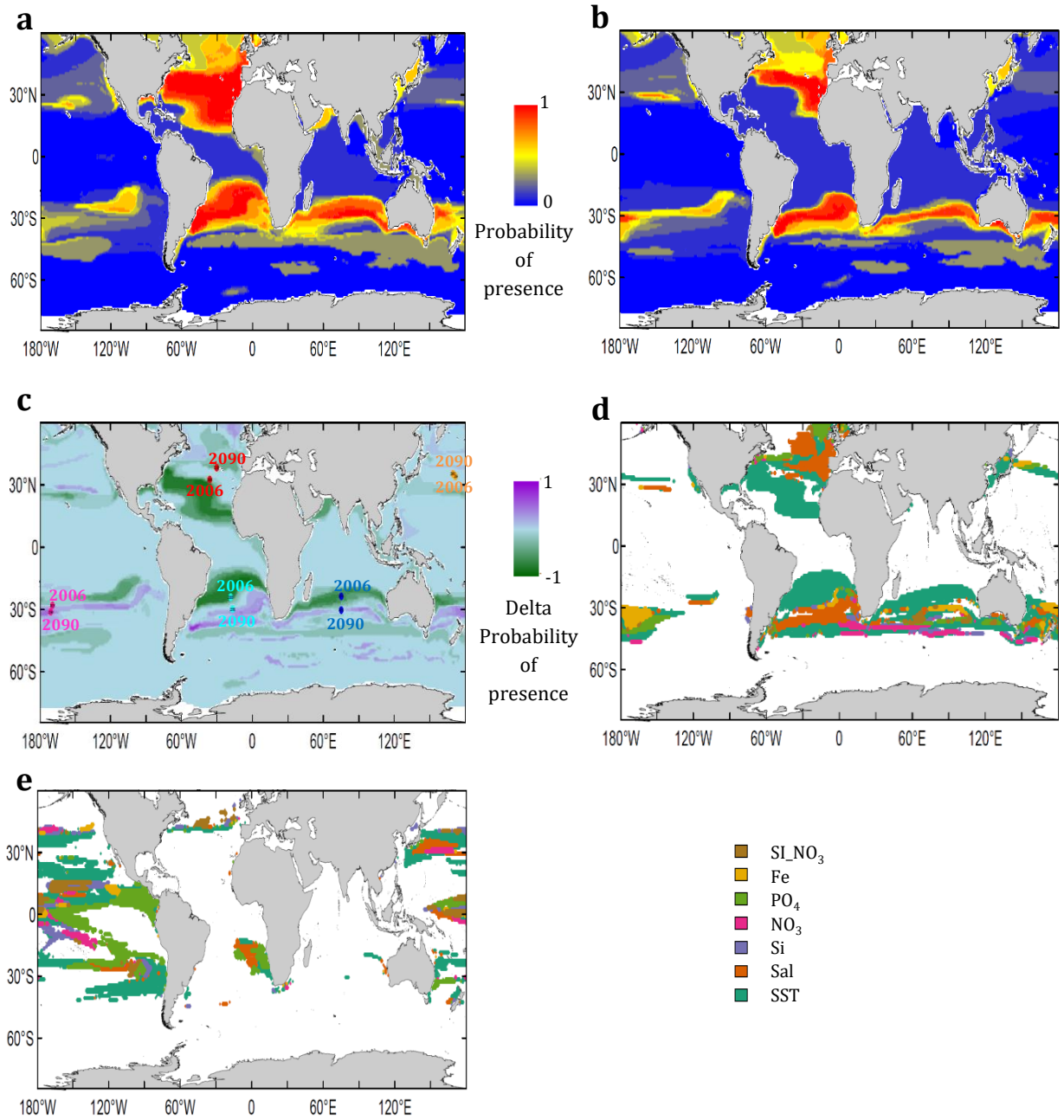
Supplementary Fig. 13 | Global geographical patterns for 20-180, 5-20, 0.22-3, 0-0.2 µm plankton size fractions in present day (a, c, e , g) and at the end of the century (b, d, f, h). The dominant province *i.e.* the one predicted to have the highest probability of presence is represented at each grid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



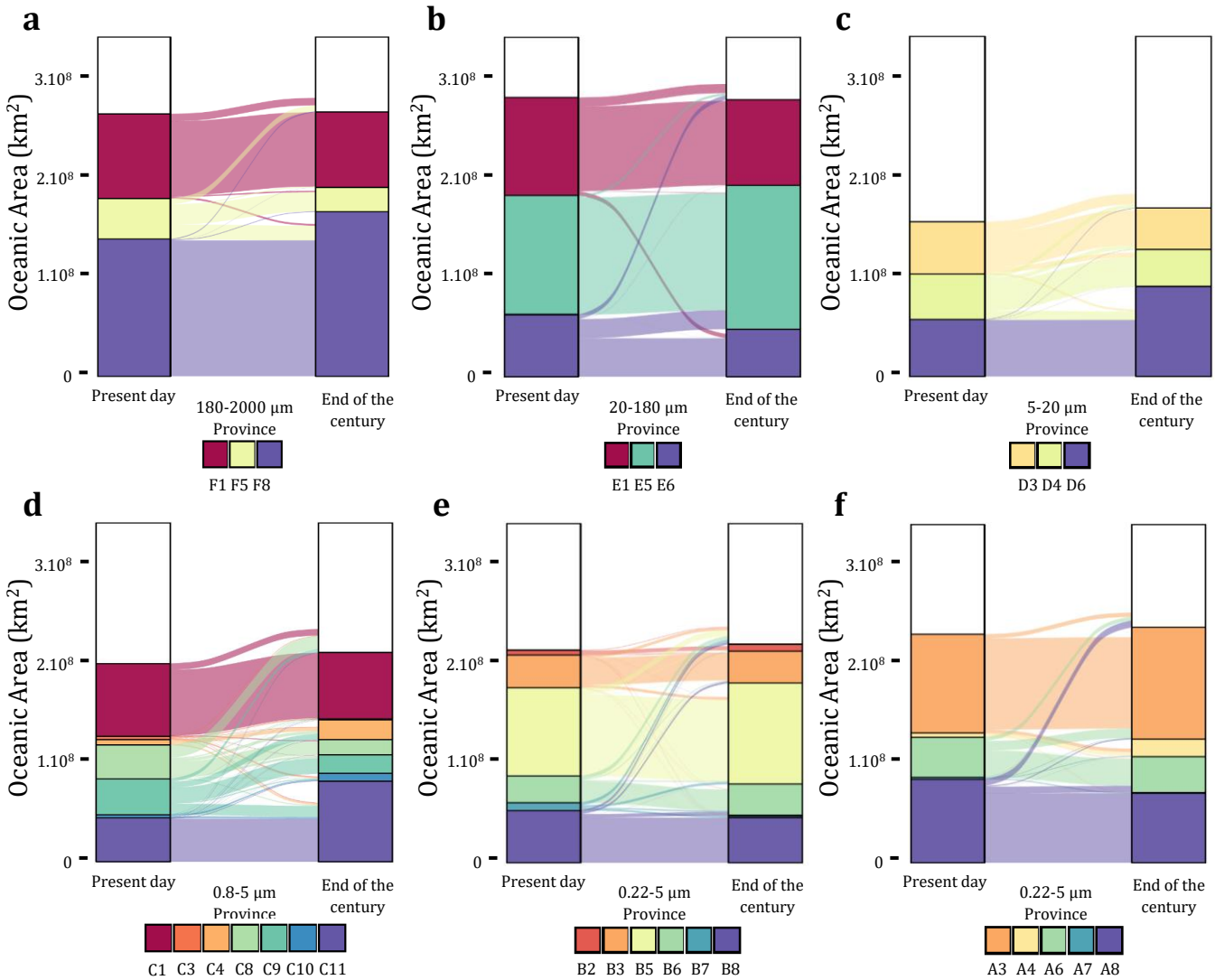
Supplementary Fig. 14 | Projected migration shifts of the 27 provinces between present day and end of the century. Predicted migration shifts are presented in 5 major ocean basins: (**a**) North Atlantic (**b**) North Pacific (**c**) South Atlantic (**d**) South Pacific and (**e**) Indian Ocean. 96 %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



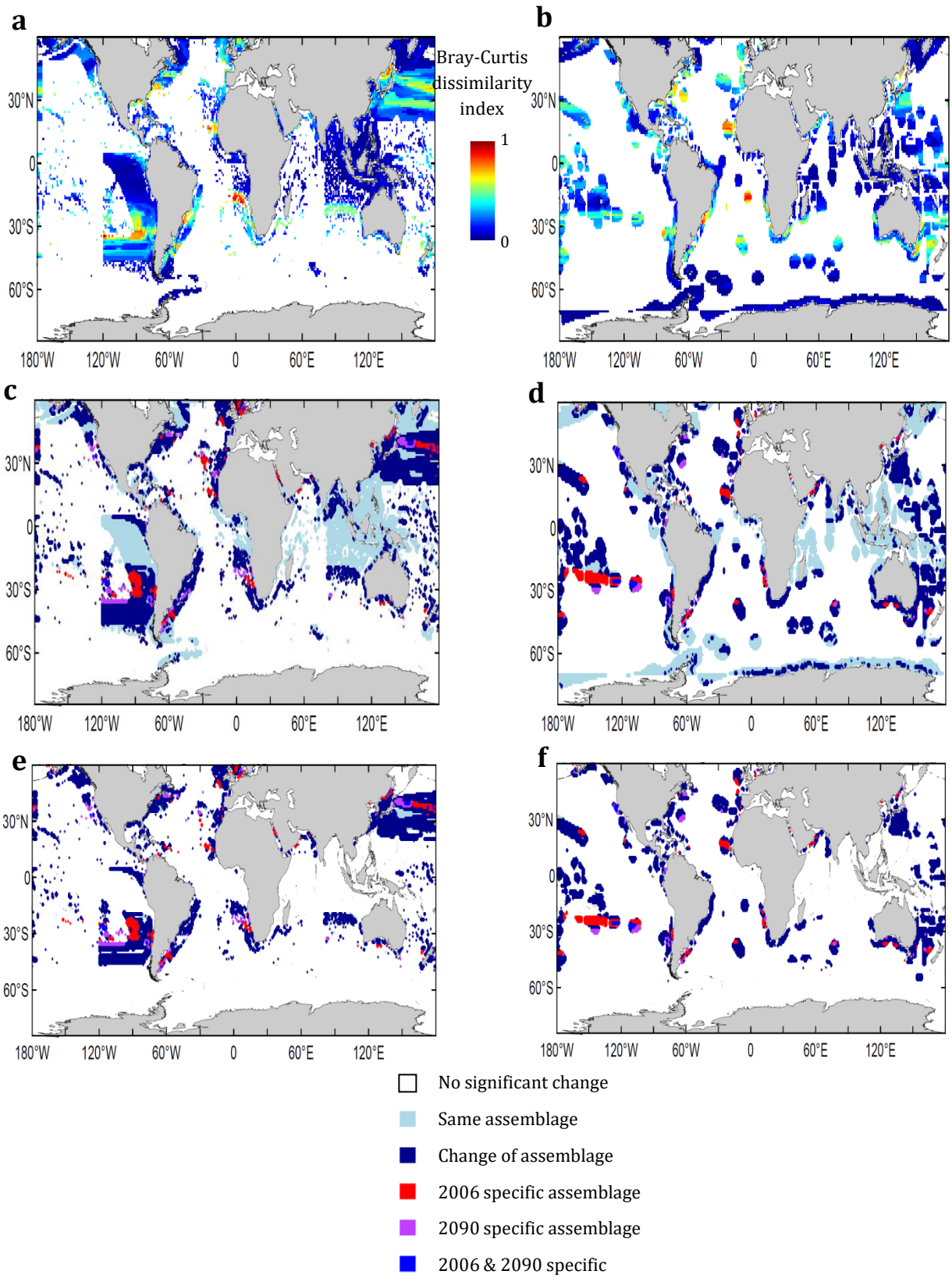
Supplementary Fig. 15 | (a) Projected migration shifts of provinces on the world map. (b) Latitudinal shift distribution (red bars) and longitudinal shift distribution (blue bars). (**a**) Migration shifts are represented as arrows pointing at the end of century centroid. Arrows are colored according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



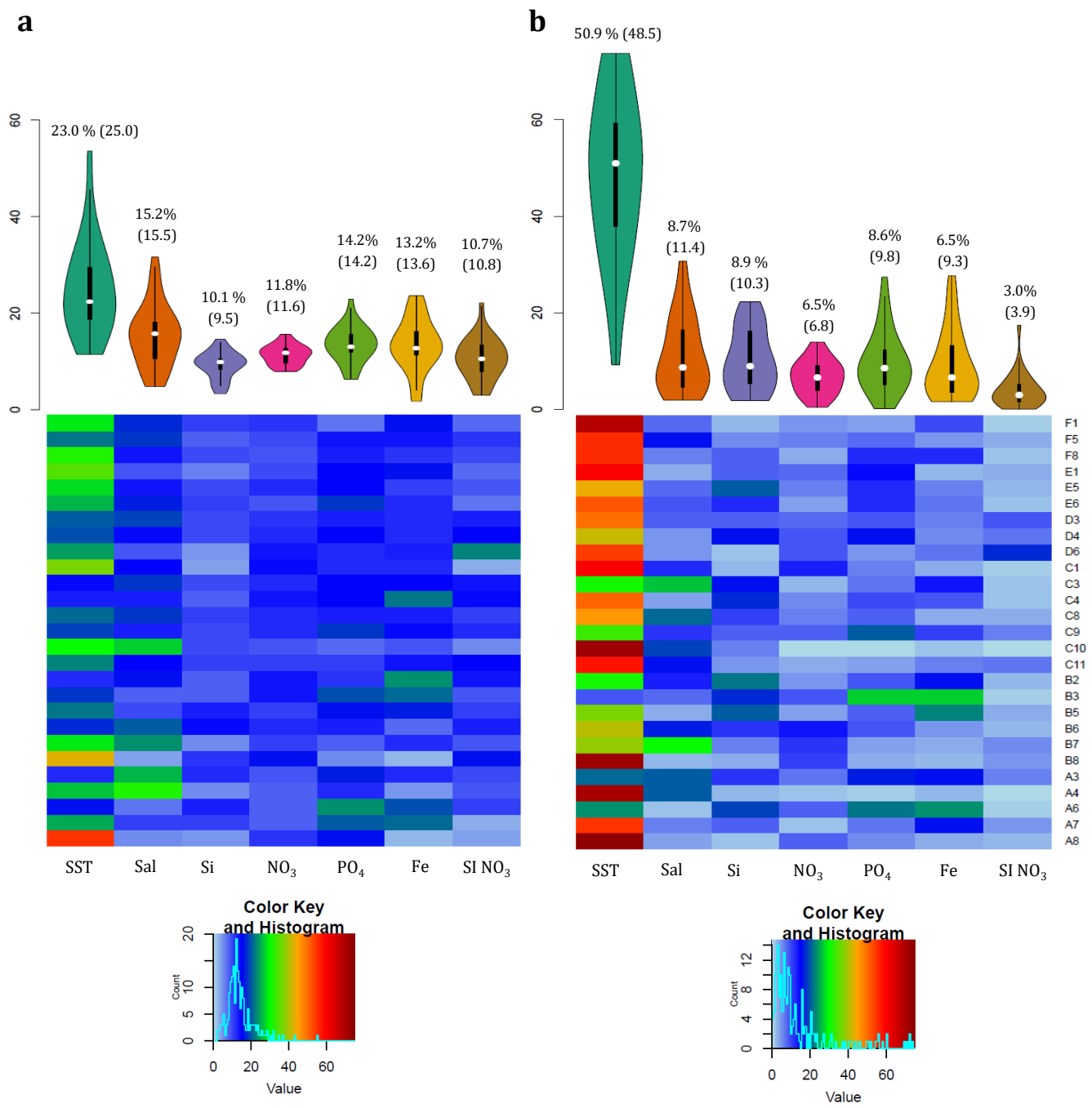
Supplementary Fig. 16 | Projection maps of province F5 of size fraction 180-2000 µm in present day (a) and at the end of the century (b). At each grid point, the probability of presence of the province is computed as the average of the predicted probability of each of the four machine learning techniques (gbm, nn, rf and gam). Red color indicates a high probability of presence. (**a**)The projected province corresponds to the sampled province (North Atlantic and South Atlantic) but several other places have high probabilities of presence such as South Australia where no sampling is available. (**b**) At the end of the century, the province is projected to reduce significantly in size. (**c**) Delta probability of presence map (2006/15 – 2090/99) and core range shift in the 5 major oceanic basins of province F5 of size fraction 180-2000 µm. In all the basins, the centroid of the province is projected to migrate poleward. (**d**) Main drivers associated with the projected changes. Changes are mainly driven by sea surface temperature (55%) followed by salinity (14%). (**e**) Main drivers associated with the shrinkage of the equatorial cluster C9 of size fraction 0.8-5. Considering only latitudes between the two tropics, changes are mainly driven by decreases in PO_4 (24 %) in addition to SST (27 %) (overall 34 % STT and 20 % PO_4).



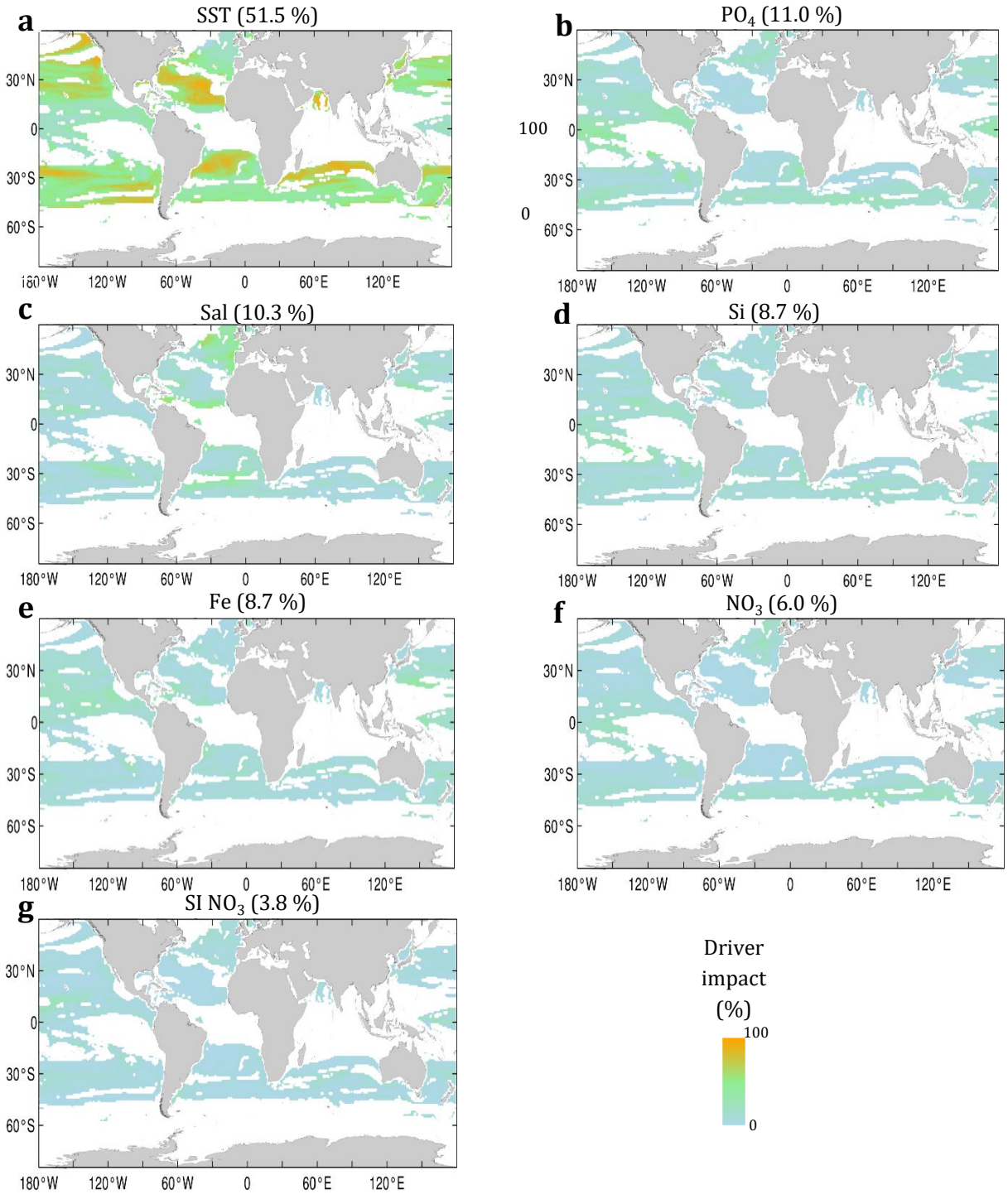
Supplementary Fig. 17 | Probabilistic covered areas of the provinces projected in present day and at the end of the century. (**a**) 180-2000 µm (**b**) 20-180 µm (**c**) 5-20 µm (**d**) 5-20 µm (**e**) 0.8-5 µm (**f**) 0-0.2 µm. The covered area by a province is defined as the area in which this province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



Supplementary Fig. 18 | Bray-Curtis dissimilarity index and assemblage changes maps comparing present day with end of the century projections of *dominant* provinces in (**a**) principal fisheries (4 last deciles[20]) and (**b**) Exclusive Economic Zones[21]. Assemblage changes in (**c**) Principal fisheries (**d**) Exclusive Economic Zones. Assemblage changes in (**e**) Principal fisheries (**f**) Exclusive Economic Zones with a Bray-Curtis dissimilarity index superior to 1/6.



Supplementary Fig. 19 | Distributions and heat maps of the relative influences of the different drivers in (a) defining single niches associated to the provinces from DALEX R package[22] (b) driving climate change associated reorganizations of single provinces. Median relative influence of Temperature is significantly higher than for all other environmental parameters (Pairwise Wilcoxon test $p<0.01$ for all parameters) in (**a**) defining the niches and (**b**) driving province reorganization**.** It is also the case within individual size fractions for (**b**) but not for (**a**). Respectively, Salinity (Sal) and Phosphate (PO_4), have second and third highest median relative influences far behind Temperature whereas dissolved Silica and seasonality index of Nitrate (SI NO_3) have the lowest median relative importance in (**a**)**.** Numbers above violins are median (and mean) relative influences.



Supplementary Fig. 20 | Maps of environmental drivers' relative influences in driving province reorganization. Each environmental parameter relative influence is quantified by considering only the variation of each parameter individually between present day and end of the century as defined in Barton et al.[23] (*Materials and Methods*) and where a significant change is projected. Importantly the mean impact of SST (51.5 %) is largely the highest. PO_4 is the second most impacting drivers (11.0 %). Contrary to supplementary Fig. 16, relative influence is calculated here by combining all provinces together. Therefore, mean relative influences slightly differ especially for dissolved silica (Si).

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

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

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%)

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

	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

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

References

1. Boyer, T. P. *et al.* WORLD OCEAN DATABASE 2013, NOAA Atlas NESDIS 72. *Sydney Levitus, Ed.; Alexey Mishonoc, Tech. Ed.* (2013). doi:10.7289/V5NZ85MT

2. Gent, P. R. *et al.* The community climate system model version 4. *J. Clim.* (2011). doi:10.1175/2011JCLI4083.1

3. Dunne, J. P. *et al.* GFDL's ESM2 global coupled climate-carbon earth system models. Part II: Carbon system formulation and baseline simulation characteristics. *J. Clim.* (2013). doi:10.1175/JCLI-D-12-00150.1

4. Collins, W. J. *et al.* Development and evaluation of an Earth-System model - HadGEM2. *Geosci. Model Dev.* (2011). doi:10.5194/gmd-4-1051-2011

5. Dufresne, J. L. *et al.* Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Clim. Dyn.* (2013). doi:10.1007/s00382-012-1636-1

6. Giorgetta, M. A. *et al.* Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.* (2013). doi:10.1002/jame.20038

7. Bentsen, M. *et al.* The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate. *Geosci. Model Dev.* (2013). doi:10.5194/gmd-6-687-2013

8. van Vuuren, D. P. *et al.* The representative concentration pathways: An overview. *Clim. Change* (2011). doi:10.1007/s10584-011-0148-z

9. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* (2006). doi:10.1016/j.patrec.2005.10.010

10. Wood, S. N. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Am. Stat. Assoc.* (2004). doi:10.1198/016214504000000980

11. Ridgeway, G. gbm: Generalized Boosted Regression Models. *R Packag. version 1.6-3.1* (2010).

12. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S Fourth edition by*. *World* (2002). doi:10.2307/2685660

13. Breiman, L. & Cutler, A. Breiman and Cutler's random forests for classification and regression. *Packag. 'randomForest'* (2012). doi:10.5244/C.22.54

14. Jaccard, P. Distribution comparée de la flore alpine dans quelques régions des Alpes occidentales et orientales. *Bull. la Murithienne* (1902).

15. Delmont, T. O. *et al.* Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* (2018). doi:10.1038/s41564-018-0176-9

16. Aumont, O., Ethé, C., Tagliabue, A., Bopp, L. & Gehlen, M. PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geosci. Model Dev.* (2015). doi:10.5194/gmd-8-2465-2015

17. Reygondeau, G. *et al.* Dynamic biogeochemical provinces in the global ocean. *Global Biogeochem. Cycles* (2013). doi:10.1002/gbc.20089

18. Fay, A. R. & McKinley, G. A. Global open-ocean biomes: Mean and temporal variability. *Earth Syst. Sci. Data* (2014). doi:10.5194/essd-6-273-2014

19. Michelangeli, P. A., Vrac, M. & Loukos, H. Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophys. Res. Lett.* (2009). doi:10.1029/2009GL038401

20. 20. Watson, R. A. A database of global marine commercial, small-scale, illegal and unreported fisheries catch 1950-2014. *Sci. Data* (2017). doi:10.1038/sdata.2017.39

21. Flanders Marine Institute (2018). Maritime Boundaries Geodatabase: Maritime Boundaries and Exclusive Economic Zones (200NM), version 10. (2018). doi:https://doi.org/10.14284/313.

22. Biecek, P. DALEX: explainers for complex predictive models. *J. Mach. Learn. Res.* **19**, 1–5 (2018).

23. Barton, A. D., Irwin, A. J., Finkel, Z. V. & Stock, C. A. Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. *Proc. Natl. Acad. Sci.* (2016). doi:10.1073/pnas.1519080113