Effect of cellular nutrient economy on the evolution of genome size in phytoplankton

Carlos Caceres[1,2,*], Marc Krasovec[1], Olivier Crispi[3],

Sebastien Gourbiere[2], Gwenael Piganeau[1,*]

[1] Sorbonne Université, CNRS, University of Perpignan, Microbial Biodiversity and Biotechnology Laboratory (UMR 8176), Banyuls sur Mer, France.

[2] CNRS, Genome and Plant Development Laboratory (UMR 5555), University of Perpignan, France

[3] CNRS, Oceanological Microbiology Laboratory (UMR 7621), Sorbonne Université, Banyuls sur Mer, France.

[*]Corresponding authors: carlos.caceres@univ-perp.fr, gwenael.piganeau@obs-banyuls.fr

Keywords: genome streamlining, genome size, natural selection, nutrient limitation, phytoplankton

Abstract:

The origin of genome size variation remains a central question in evolutionary biology. Previous analyses suggest that the energetic costs of small insertions and deletions (indels) influence genome size by altering the cellular energy budget. Here, we present a theoretical framework for genome evolution in phytoplankton—the main producers of the sunlit ocean—whose growth depends on nutrient availability. We derive an expression for the selection coefficient on indels as a function of four parameters, including the minimum amount of nutrient inside the cell. Parameter estimates indicate that 1-bp indels can be counterselected in phytoplankton with minimal nutrient requirements. We then test three predictions of nutrient-driven selection on genome evolution. Altogether, this model provides a quantitative and mathematical basis for the genome streamlining hypothesis.

Introduction

Genome size spans over five orders of magnitude, from 1.3 Mb in prokaryotes to more than 100 Gb in certain eukaryotes [1,2]. Such a variation is primarily driven by differences in the size of non-coding regions [3–5]. Present genome sizes result from past genome expansion and compaction dynamics driven by molecular processes generating "indel" mutations and evolutionary processes shaping their fixation [3]. Here, we use "indel" in a broad sense to refer to any mutation generating a genome size variation, which includes deletions, insertions, duplications, hybridizations or lateral gene transfers.

Many post-hoc explanations of the diversity of genome sizes have been suggested [5]. Of relevance for microorganisms, the "streamlining hypothesis" states that small genome sizes may have evolved to minimize the material cost of replication, leading to the selection of species with smaller genomes in resource-limited environments [1,6–9]. However, this hypothesis currently lacks a mechanistic theoretical framework linking genome size to the fixation probability of indels, which depends on both natural selection and genetic drift [3,4]. This framework requires estimating the selection coefficients of indel mutations. Previous approximations were based on the relative energetic cost of indels compared with a cell's ATP requirements from glucose [10,11]. Here, we provide an alternative estimation based on a model linking the molecular changes in the number of nitrogen (N) and phosphate (P) atoms in an indel to differences in the growth rate between the mutant carrying the indel and the wild type. In this way, the selection coefficients of indels can be explicitly linked to the changes in cellular nutrient requirements they cause.

The recent increase in genome resources in phytoplankton allows renewed investigation into the role of cellular nutrient requirements on genome evolution. Phytoplankton are mostly unicellular and inhabit ecosystems that show a wide range of nutrient availability. Low N and P levels limit phytoplankton growth in vast areas of the ocean and drive genomic adaptations that enhance survival and resource utilization in nutrient-limited environments [12–15]. Moreover, phytoplankton include some of the smallest genomes among free-living prokaryotes and eukaryotes: <2 Mb genomes in *Prochlorococcus* [16] and 13 Mb in *Ostreococcus* [17,18]. Importantly, the growth rate of phytoplanktonic organisms can be described as a function of the intracellular nutrient content, or quota, of the most limiting nutrient by using the quota model [19]. The model's key parameter, $Q_{\min}$, is the minimum amount of nutrient inside the cell, required for population growth, and has been estimated for several phytoplankton species [20]. Phytoplankton thus provide a unique opportunity to estimate selection coefficients (s) of indel mutations based on the changes in cellular nitrogen and phosphate content they cause.

Non-coding regions drive between-species genome size variations in phytoplankton 346

To investigate genome size variations in phytoplankton, we analyzed 173 genomes from 28 bacteria and 101 eukaryotic species, spanning 5 of the 8 major photosynthetic lineages [21]. Cyanobacterial genomes range from 1.6 to 8 Mb, whereas eukaryotic genomes span from 13 Mb in *Ostreococcus tauri* (Chlorophyta) to 4.1 Gb in *Prorocentrum cordatum* (Dinoflagellate) (Fig. 1A and fig. 1A). In eukaryotes, genome size variation is primarily driven by differences in non-coding DNA content, ranging from 2.2 Mb (17.6%) in *O. tauri* to >4 Gb (96.9%) in *P. cordatum* (Fig. 1A,B and fig. S1A). Both intergenic and intronic regions contribute substantially to genome size, averaging39% and 20% of eukaryotic genomes, respectively); in 80% of annotated species, intergenic DNA exceeds intronic DNA (39 of 49 fig. S1B). In fact, genome size correlates positively with the average size of both

intergenic and intronic regions (Fig. 1C, D). Lineage-specific patterns are also evident. In diatoms, coding and non-coding DNA contribute nearly equally to genome length (49% and 51% on average, respectively), whereas chlorophytes with similar genome sizes have a higher fraction of non-coding DNA (62% and 38%; Fig. 1A, B and fig. S1A). However, diatom genomes have a higher proportion of intergenic regions than Chlorophyta with a similar genome size (ANCOVA, $F_{1,14} = 7.41$, $p < 0.05$; Fig. f1B), reflecting a higher gene number and fewer introns per gene (ANCOVA, genes: $F_{1,14} = 8.98$, $p < 0.01$, Fig. 1E; introns per gene: $F_{1,14} = 30.17$, $p < 0.001$). Taken together, genome sizes in unicellular phytoplankton span four orders of magnitude, variation that largely reflects differencesin non-coding DNA content. This pattern is consistent with previous datasets spanning both unicellular and multicellular organisms [4,5,22–24].

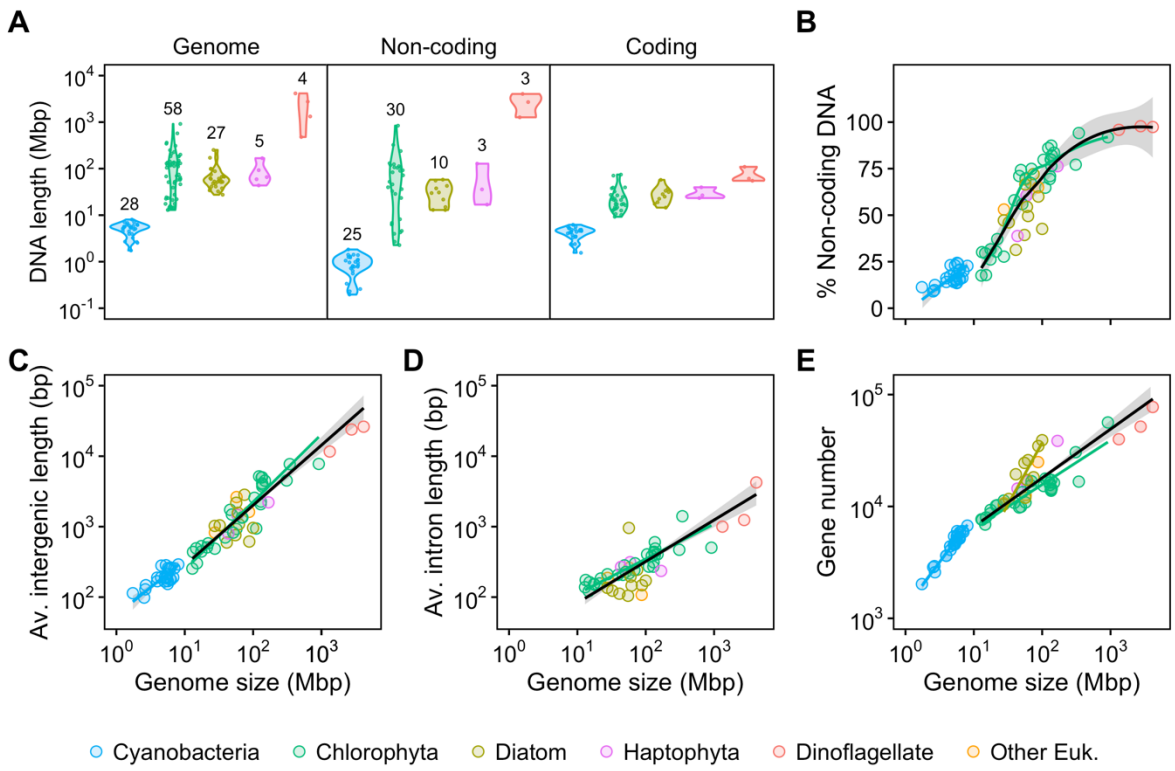

Fig. 1. Size of DNA regions and their relationship with genome size. **(A)** Violin plots showing the size distribution of genomes, non-coding regions, and coding regions. Numbers indicate the number of observations in each group; values for coding and non-coding regions are identical. **(B)** Relationship between genome size and the percentage of non-coding DNA. **(C)** Relationship between genome size and average intergenic length. **(D)** Relationship between genome size and average intron length. **(E)** Relationship between genome size and gene number. Each point corresponds to a different species. Black lines represent the best-fit regression when all eukaryotic species are pooled. Grey shading indicates the 95% confidence intervals for regression analyses of cyanobacteria and eukaryotes. Lines of best fit for individual taxonomic groups are shown when slopes differed significantly from zero and the number of observations exceeded four. For the percentage of non-coding DNA, local regression was used instead of standardized major axis regression.

Mathematical expression and estimation of selection coefficients of indels

The quota model [19] was used to derive the selection coefficients of indel mutations, s, from the nutrient cost of indels (ΔQ_{min}) in 83 species with available parameters on cellular minimum quota (Q_{min}) and the growth rate (see methods; figs. S2 and S3, and table S1 [20,25–28]):

$$s = \mu_m - \mu_{WT} = \mu_{inf}\left(\frac{Q_{min,wt}}{Q_{wt}} - \frac{Q_{min,wt} + \Delta Q_{min}}{Q_{wt} + \Delta Q_{min}}\right)$$

μ_{wt} : growth rate of wild type

μ_m : growth rate of mutant

μ_{inf} : asymptotic maximum growth rate

$Q_{min,wt}$: minimum quota of wild type

Q_{wt}: quota of wild type

ΔQ_{min}: nutrient cost of an indel mutation

The absolute value of selection coefficients can be derived for the effect of phosphate (s_P) and nitrogen (s_N) limitation (Supplementary Method section S1). As expected, the selection coefficient increases with the size of the indel, with deletions yielding positive s values and insertions yielding negative ones (Fig. 2A and fig. S4). When the contribution of indels to the DNA-packaging N cost of histones is included (s_{Nh}), the relationship with indel size remains monotonic but shows sharp changes whenever a new nucleosome is added or removed. Consequently, the same indel size may correspond to two different s_{Nh} values, depending on nucleosome addition (or removal) (Fig. 2A). Thus, s_{Nh} depends not only on indel size but also on chromosomal context, as nucleosome number can vary across loci [29].

The estimated selection coefficients allow comparison of the fixation probability of a non-transcribed (i.e. number of copies of mRNA, c_{mRNA}, = 0) 100 bp deletion across different species (Fig. 2A-C and figs. S4 and S5). Selection coefficients s_{P}, s_N, and s_{Nh} range six orders of magnitude, from 10^{-13} to 10^{-7} (Fig. 2B and fig. S5). Between-species variation in selection coefficients is primarily driven by Q_{min}, rather than by differences in μ_{inf}. This is because Q_{min} spans more than five orders of magnitude, whereas μ_{inf} spans less than two. This dominant effect of Q_{min} is further supported by strong correlations between s_P with Q_{minP} and s_N with Q_{minN} ($R^2 = 0.97$ and $R^2 = 0.98$, respectively, Fig. 2C). The minimum s value required for selection to be effective depends on the inverse of the effective population size, N_{e}, such that $N_{\mathrm{e}}\cdot s_{>}1$ ($2N_{\mathrm{e}}\cdot s_{>}1$ in diploids). In phytoplankton, N_{e} ranges from 10^6 to 10^8 and is typically around 10^7 (table S2) [11] Accordingly, we consider selection effective when $s > 10^{-7}$ for haploids and $s > 5\ 10^{-8}$ for diploids. These thresholds are met in 11 out of 70 for s_P, in 1 out of 47 for s_N) and in 4 out of 35 species for s_{Nh} (fig. S4). Selection coefficients can be approximated by the fractional cost of an indel, $\Delta Q_{min}/Q_{min}$. In species with low Q_{min} — such as *Prochlorococcus* and *Mamiellales* — 100 bp indel mutations have selection coefficients such that $N_{\mathrm{e}}\cdot s>1$ (fig. S4). Selection coefficients may exceed the conservative threshold of 10^{-6} for larger indels or those occurring within introns of highly expressed genes (Fig 2D and fig. S5). In contrast, in species with large Q_{min} — such as *Coscinodiscus* diatoms or large dinoflagellates — intergenic indels must exceed 1 Mb to reach selection coefficients on the order of $|s|\sim10^{-7}$. These estimates show that P and N limitation can significantly influence selection on indel mutations in dispensable non-coding regions,

which have little or no impact on cellular function, particularly in phytoplankton species with low cellular P and N requirements.

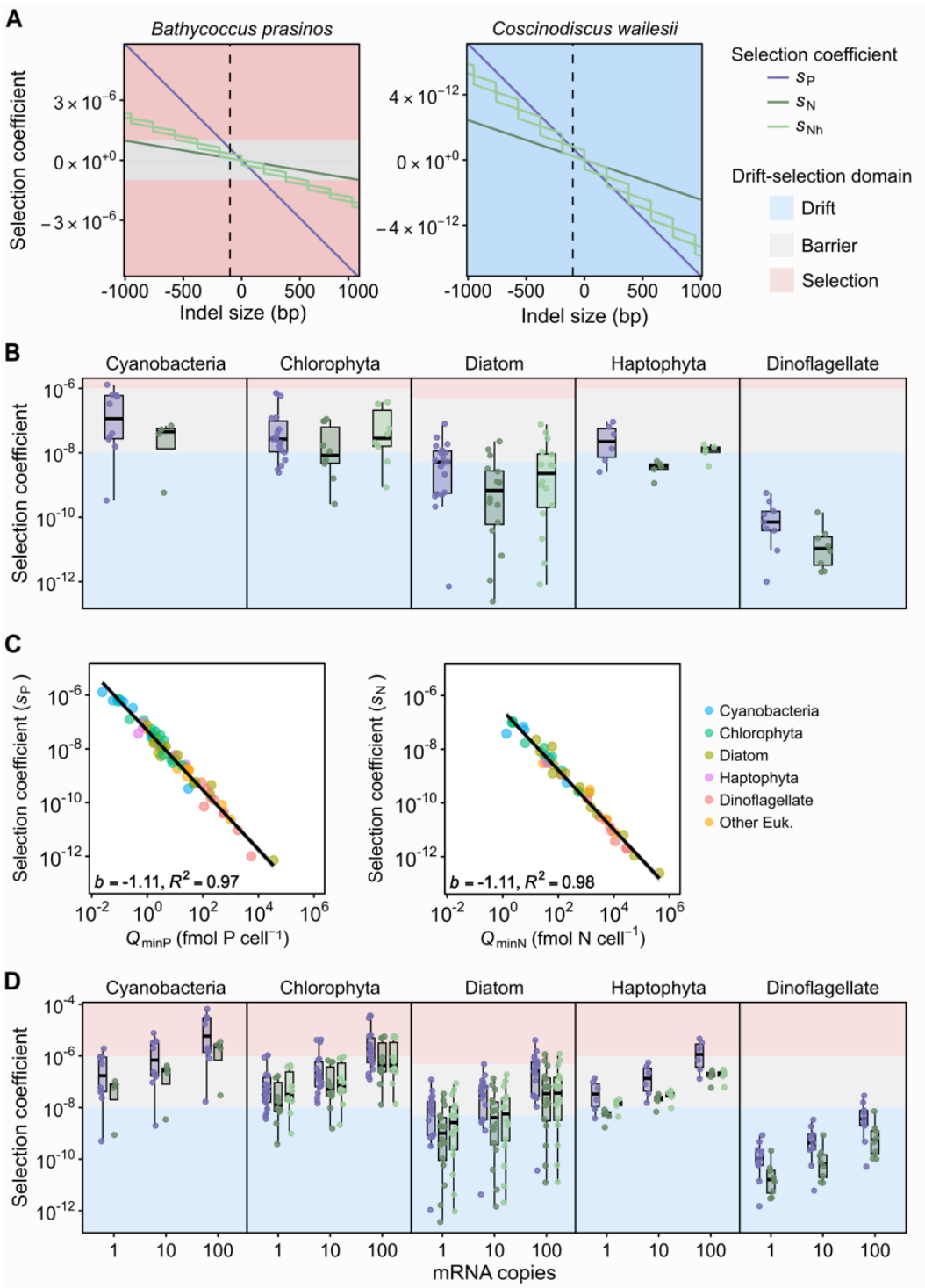

Fig. 2. Selection coefficient estimates from Eq1 and available quotas and growth rates in phytoplankton. (**A**) Selection coefficients as a function of indel size in untranscribed regions for the chlorophyte *Bathycoccus prasinos* (low minimum quotas) and the diatom *Coscinodiscus wailesii* (large minimum quotas). Selection coefficients reflect the impact of indels on P (s_P) or N (s_{Nh} and s_N, with or without considering the histone N) requirements under P- or N-limiting conditions. Panel background colors indicate the relative importance of drift (blue) versus natural selection (pink). The vertical dashed line marks a 100-bp deletion. (**B**) Selection coefficients for 100-bp deletions in untranscribed regions (each point represents one species). (**C**) Relationship between s_P and s_N and Q_{min} (Q_{minP} or Q_{minN}). Black lines show

best-fit regressions when all eukaryotic species are pooled, with grey shading indicating the 95% confidence intervals. (**D**) Selection coefficients for 100-bp deletions in transcribed regions at different expression levels (i.e., number of mRNA copies).

Stronger selection coefficients associated to P as compared to N limitation

Of note, the selection coefficient associated with phosphorus cellular requirements, s_P, was higher than that for nitrogen, s_N, in all species, and higher than s_{Nh} in 18 out of 24 species (fig. S5). Overall, phosphate requirements exert a stronger selective pressure on indels than nitrogen requirements. The relative impact of P versus N limitation on the selection of a given indel can be approximated by comparing the $\Delta Q_{min}/Q_{min}$ ratios for P and N. Thus, the relative impact of P versus N limitation will be particularly high in taxa with high N requirements in relative terms, such as cyanobacteria and Chlorophyta, whereas it will be more balanced in groups with low cellular N:P ratios, such as diatoms and Haptophyta [30]. Indeed, the six species with $s_P < s_{Nh}$ — four diatoms and two raphidophytes, all diploids — have Q_{minN}/Q_{minP} <13. In transcribed regions, s_P more frequently exceeds s_{Nh} (Fig. 2D and figs. S4 and S6), because the number of RNA copies amplifies changes in P (and N) in nucleotides, while histone-associated N remains unaffected. In addition, the s values are affected by the actual quota and growing conditions; s approaches 0 when $Q \approx Q_{\min}$, under severe nutrients limitation and negligible growth, or when $Q >> Q_{\min}$, under no nutrient limitation and maximal growth (fig. S7). Therefore, for most species, except those with low $Q_{\mathrm{minN}}/Q_{\mathrm{minP}}$ ratios, environments where P is the primary limiting nutrient exert stronger selection for genome reduction than those limited by N.

Estimation of the fate of duplications

Gene duplications are a major type of indel mutations (4). To investigate the fate of duplications, we compared the minimum length of indels required for selection to be effective, $l_{\min}$, to the lengths of non-coding regions. In eukaryotic species with available N_e estimates, the minimum indel length in non-transcribed regions ranged from 12–2,747 bp under phosphorus limitation and 75–4,341 bp under nitrogen limitation, with the lowest values observed in *Ostreococcus tauri* (fig. S8). In *Prochlorococcus marinus*, $l_{\min}$ was even smaller—only 5 bp under P-limitation. Interestingly, in most species, the estimated $l_{\min}$ values—calculated assuming $N_e = 10^8$ and $N_e = 10^6$—are similar to the median sizes of intergenic regions (fig. S8). Indeed, the 99th percentile of intergenic region sizes exceeded $l_{\min}$ in the majority of species. For transcribed regions, $l_{\min}$ was approximatively one-third lower than for non-transcribed regions, and, in most species, also aligned with the median intron size (fig. S8). This alignment between $l_{\min}$ and the average size of non-coding regions suggests that, on average, duplications larger than the average intergenic region are subject to effective purifying selection, or conversely, that deletions of such duplicated segments may be positively selected.

Positive correlation between genome size and the minimum cellular nutrient content Q_{min}

A corollary expectation of higher selection coefficients against insertions —or for deletions — in species with low Q_{min}, is a positive correlation between the current size of genomic regions and Q_{min}. Indeed, genome sizes, as well as the sizes of the different genomic regions (non-coding, coding, intergenic and intronic), correlate positively with Q_{min} measures across eukaryotic species and genera, both for Q_{minP} and Q_{minN} (Fig. 3, fig. S9 and table S3). These correlations remain significant after excluding the *Prorocentrum spp*. outlier observation and focusing on species-level data (p-value <

0.05, except for the average intron lengths), and are also significant for the green algae and cyanobacteria subsets (Fig. 3 and fig. S9, tables S4 and S6). To ensure these positive relationships were not soeley driven by the P and N content of DNA, we repeated the analyses using Q_{min} values corrected by subtracting the N and P content of the relevant genome compartment (fig. S10 and table S5). All correlations remained significant after correction, implying that DNA nitrogen and phosphorus content alone does not fully explain the positive relationship between DNA compartment lengths and Q_{min}.

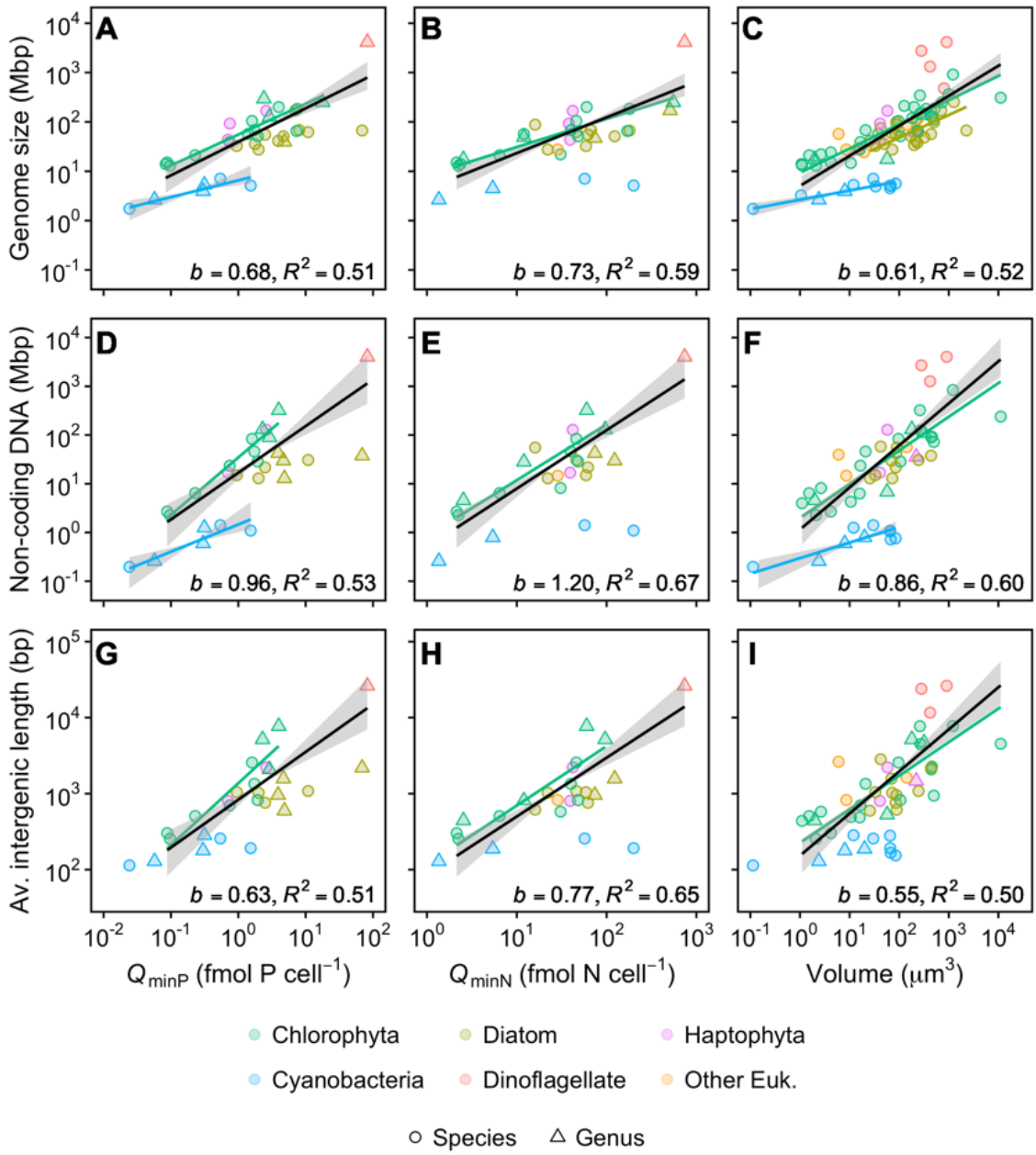

Fig. 3. Relationship between genomic traits and Q_{minP}, Q_{minN} and cell volume on a log-log scale. (A, B, C) Genome size. (D, E, F) Total non-coding DNA. (G, H, I) Average size of intergenic regions. Each point represents a different species or genus. Point color and shape indicate taxonomic affiliation and taxonomic level, respectively. Note that the set of species and the x-axis scales are not the same for Q_{minP}, Q_{minN}, and cell volume. Black lines represent best-fit regressions when all eukaryotic species are pooled, with slope coefficients, b, and explained variances, R^2, reported at the bottom of each panel. Grey shading indicates the 95% confidence interval for

regressions of cyanobacteria and eukaryotes. Best-fit regressions for individual taxonomic groups are shown when slopes were significantly different from zero and the number of observations exceeded four.

Given the long-standing interest in genome size–cell volume relationships [5,31–33], we investigated the influence of cell volume on genomic region size. The slope coefficients relating the size of DNA regions to cell volume were positive and similar to those observed for the relationships between DNA regions and Q_{min}, particularly Q_{minP}. This is as expected because cell volume and Q_{min} are positively correlated [20] (Fig. 3, fig. S9 and table S6). Furthermore, we estimated the variance in genome size explained by Q_{min} and cell volume from partial regression analysis (table S7). Most of the explained variation in genome and DNA regions size is shared between Q_{min} and cell volume. Thus, our model and these analyses suggest that cellular nutrient requirements, rather than cell volume *per se*, are the primary drivers of the relationship between genome size and cell volume in phytoplankton. Nevertheless, this does not exclude the possibility that selection of smaller cell volumes, that increase the surface to volume ratio and maximizes nutrient uptake in oligotrophic environments[34], may ultimately affect Q_{min} and thus indirectly promote genome reduction.

Observed intragenomic variations in intron and intergenic region lengths

The model further predicts that selection coefficients increase with transcription rate (ΔQ_{min} increases with c_{mRNA}; Eqs. S3-S5). In the low-Q_{min} *Ostreococcus* species, selection will be effective for non-transcribed indels longer than 12 bp under P-limiting conditions, which corresponds, in terms of the P budget, to a 1bp deletion with 22 mRNA copies. Consequently, shorter deletions are more likely to become fixed in highly transcribed regions, leading to a negative relationship between the length of introns and their transcription rates. Indeed, introns in the higher transcription-rate class are significantly shorter than those in regions with lower transcription rates in *Bathycoccus* and *Ostreococcus* (Fig. 4 and fig. S12). Compact genomes such as those of *Ostreococcus* and *Bathycoccus* show transcription at up to 99% of intergenic sites, with transcription rates spanning four orders of magnitude (figs. S11 and S12). Intergenic region size decreases with increasing expression, but shows a slight increase in the highest expression class, consistent with the presence of functional elements in these regions (figs. S11-12). The elemental cost of mRNA thus provides an indirect explanation for the origin of intragenomic variation in intron[35] and, to a lesser extent, in intergenic lengths.

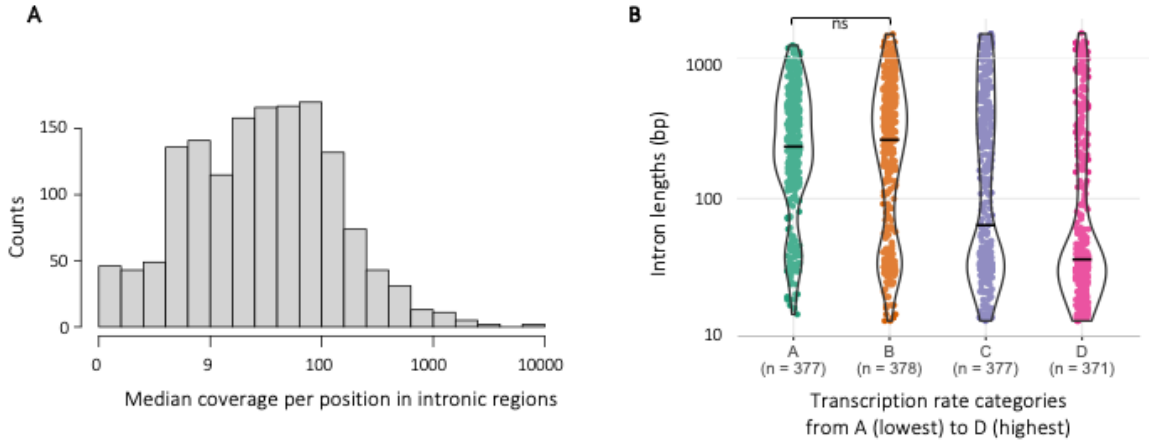

Fig. 4. Transcription rates of introns in *Bathycoccus prasinos*. (A) Distribution of median RNA coverage across introns. **(B)** Median RNA coverage across introns grouped by length category. All Pairwise Wilcoxon tests are significant, except the for A – B comparison.

Indirect evidence of selection against insertions in more oligotrophic environments

Selection may also leave signatures on mutations still segregating within natural populations [36]. Under prevalent limiting nutrient conditions, we expect insertions and deletions to be more and less deleterious, respectively, and all the more so in longer indels. This should be reflected in a lower minor allele frequency (MAF) of insertions as compared to deletions, as well as a lower MAF in longer insertions [37]. These predictions were tested by leveraging metagenomic and associated environmental data from the TARA Oceans pan-oceanic survey [38,39], focusing on species whose presence has been documented at 68 out of 133 sampling stations [40]. By mapping reads against the reference genomes, we identified 8260 indels with a coverage >= 20 reads, located in intergenic regions, and uniquely present in one metagenome (Fig. 5A, see methods section 9). Of these indels, 4871 (59%) and 2100 (25%) were affiliated to the cosmopolitan *Bathycoccus prasinos* and *Ostreococcus lucimarinus* species (fig. S14). Consistent with our hypothesis, MAF of long insertions were significantly lower than MAF of short insertions in lower P ($W = 23069$, $p < 0.001$; Fig. 5B) and lower N ($W = 24242$, $p < 0.05$; fig. S13B and table S8) environments. Moreover, the MAF of long insertions (≥10bp) was markedly lower in waters with a lower P concentration than in waters with a higher P concentration (Wilcoxon, $W = 91.5$, $p < 0.01$; Fig. 5B). MAF comparisons for long insertions between environments with lower and higher N concentrations were qualitatively similar but non-significant (fig. S13 and table S8). Similar results were obtained when analyses were performed for *B. prasinos* and *O. lucimarinus* separately, or when analyzing median MAF values per metagenome (fig. S14, S15 and table S8). Thus, selection on indels driven by phosphate economy may also be detected on shorter evolutionary timescales, that is, on polymorphisms segregating within natural populations.

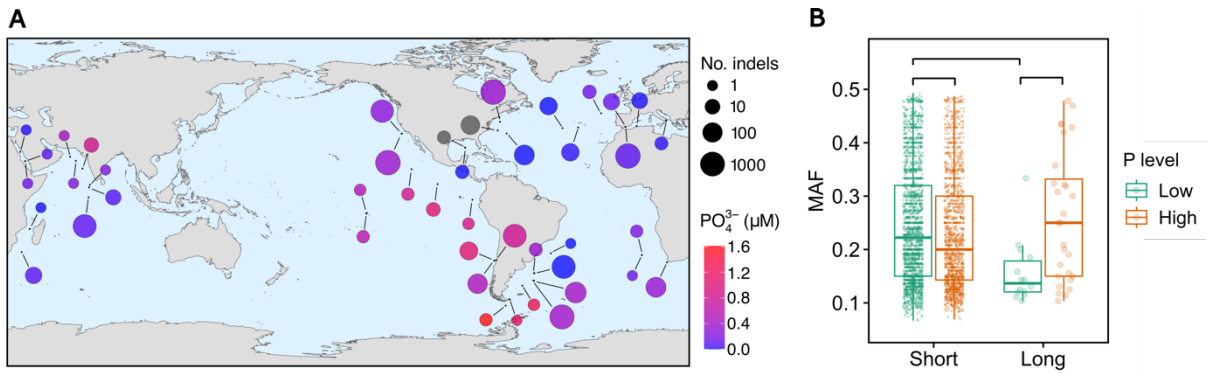

Fig. 5. Minor allele frequency (MAF) of insertions and deletions as a function of indel length and environmental phosphate concentration. (**A**) Worldwide distribution of metagenomes containing polymorphism for the target species. (**B**) MAF of short (<10bp) and long (>=10bp) indels in low (green) and high (orange) phosphorus environments. Results of the pairwise comparisons corresponding to panel B are provided in Table S8.

Conclusions

In conclusion, we provide an empirically informed theoretical background of the processes underlying genome streamlining observed in phytoplankton inhabiting oligotrophic waters (e.g. *Prochlorococcus* and *Pelagibacter*) [1,7,16,41]. Previous pan-oceanic metagenome analyses have reported shorter genome sizes in prokaryotes from tropical than polar oceans, which was mainly attributed to temperature effects [42]. Our alternative explanation highlights the role of nutrient limitation and the higher selective pressures to promote genome compactness of photoautotrophic organisms in oligotrophic subtropical waters.

Unlike heterotrophs, phytoplankton rarely experience carbon limitation, and its potential effect on genome size is likely lower than that of nitrogen (N) and phosphorus (P), given cellular and nucleotide stoichiometry [43]. Energy limitation, i.e. light in photoautotrophs, can certainly contribute to genome reduction, and light limiting conditions often alternate with nutrient limitation in oceans and lakes [10,44]. However, selective pressures exerted by N and especially P limitation may be stronger. To evaluate this, one may compare the main determinants of selection coefficients for indels under nutrient limitation ($\Delta Q_{min}/Q_{min}$) versus energy cost ($\Delta E/E$: cell energy budget) as reported in Lynch and Marinov (2015)[10]. Note that this comparison is relevant for the replication associated costs of non-coding regions. We found that, in eukaryotic species ≤100 μm^3, the fractional costs in terms of P and N were one order of magnitude and nearly five times higher, respectively, than energy costs. This corresponds to $\Delta Q_{minP}/Q_{minP} \approx 2\ \Delta Q_{minN}/Q_{minN} \approx 10\ \Delta E/E$ (fig. S16). This is particularly relevant under the current global change scenario, which might enhance P limitation due to anthropogenic N releases and a more marked stratification [45–48].

The model suggests several evolutionary implications of N and P constraints on genome evolution. First, higher s values in species with low Q_{min} reflect the higher relative costs of nucleotides in terms of N and P, which implies that each decrease in Q_{min} further strengthens selection for deletions. Second, any variation in Q_{min}, which can be promoted by a given indel or by any other cellular change, affect the strength of nutrient-mediated selective pressures on future mutations. This generates

positive feedback on genome size: a large insertion, such as a whole chromosome duplication, alters the selection efficiency against subsequent indels. Third, changes in Q_{min} can result from multiple small indels at different loci, not necessarily a single event, and thus permit effective selection against insertion-biased mutator alleles. Fourth, relaxed selection in high Q_{min} species increases the likelihood that duplications or horizontally transferred sequences are retained compared with low Q_{min} species(*4*). Thus, neofunctionalization and genome diversification following gene duplication are limited in low Q_{min} species or ecotypes thriving in oligotrophic environments, but more likely in nutrient-rich environments.

Material and methods are provided as supplementary material.

References

1. Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., Bibbs, L., Eads, J., Richardson, T.H., Noordewier, M., et al. (2005). Genetics: Genome streamlining in a cosmopolitan oceanic bacterium. Science (1979) *309*, 1242–1245. https://doi.org/10.1126/science.1114057.
2. Fernández, P., Amice, R., Bruy, D., Christenhusz, M.J.M., Leitch, I.J., Leitch, A.L., Pokorny, L., Hidalgo, O., and Pellicer, J. (2024). A 160 Gbp fork fern genome shatters size record for eukaryotes. iScience *27*. https://doi.org/10.1016/j.isci.2024.109889.
3. Petrov, D.A. (2001). Evolution of genome size: new approaches to an old problem. TRENDS in Genetics *17*, 23–28.
4. Lynch, M., and Conery, J.S. (2003). The Origins of Genome Complexity.
5. Gregory, T.R. (2001). Coincidence, coevolution, or causation? DNA content, cellsize, and the C-value enigma. Biological Reviews *76*, 65–101. https://doi.org/10.1111/j.1469-185X.2000.tb00059.x.
6. Swan, B.K., Tupper, B., Sczyrba, A., Lauro, F.M., Martinez-Garcia, M., Gonźalez, J.M., Luo, H., Wright, J.J., Landry, Z.C., Hanson, N.W., et al. (2013). Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc Natl Acad Sci U S A *110*, 11463–11468. https://doi.org/10.1073/pnas.1304246110.
7. Giovannoni, S.J., Cameron Thrash, J., and Temperton, B. (2014). Implications of streamlining theory for microbial ecology. Preprint at Nature Publishing Group, https://doi.org/10.1038/ismej.2014.60 https://doi.org/10.1038/ismej.2014.60.
8. Lauro, F.M., Mcdougald, D., Thomas, T., Williams, T.J., Egan, S., Rice, S., Demaere, M.Z., Ting, L., Ertan, H., Johnson, J., et al. (2009). The genomic basis of trophic strategy in marine bacteria https://doi.org/10.1073/pnas.0903507106.
9. Okie, J.G., Poret-Peterson, A.T., Lee, Z.M.P., Richter, A., Alcaraz, L.D., Eguiarte, L.E., Siefert, J.L., Souza, V., Dupont, C.L., and Elser, J.J. (2020). Genomic adaptations in information processing underpin trophic strategy in a whole-ecosystem nutrient enrichment experiment. Elife *9*. https://doi.org/10.7554/eLife.49816.
10. Lynch, M., and Marinov, G.K. (2015). The bioenergetic costs of a gene. Proc Natl Acad Sci U S A *112*, 15690–15695. https://doi.org/10.1073/pnas.1514974112.
11. Lynch, M.R. (2024). Evolutionary cell biology: the origins of cellular architecture (Oxford University Press).
12. Martiny, A.C., Coleman, M.L., and Chisholm, S.W. (2006). Phosphate acquisition genes in Prochlorococcus ecotypes: Evidence for genome-wide adaptation.

13. Ustick, L.J., Larkin, A.A., Garcia, C.A., Garcia, N.S., Brock, M.L., Lee, J.A., Wiseman, N.A., Moore, J.K., and Martiny, A.C. (2021). Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation. Science (1979) *372*, 287–291.
14. Lin, S., Litaker, R.W., and Sunda, W.G. (2016). Phosphorus physiological ecology and molecular mechanisms in marine phytoplankton. J Phycol *52*, 10–36.
15. Moore, C.M., Mills, M.M., Arrigo, K.R., Berman-Frank, I., Bopp, L., Boyd, P.W., Galbraith, E.D., Geider, R.J., Guieu, C., Jaccard, S.L., et al. (2013). Processes and patterns of oceanic nutrient limitation. Nat Geosci *6*, 701–710.
16. Dufresne, A., Garczarek, L., and Partensky, F. (2005). Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biol *6*, 1–10.
17. Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A.Z., Robbens, S., Dé Ric Partensky, F., Degroeve, S., Echeynié, S., Cooke, R., et al. (2006). Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features.
18. Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S., et al. (2007). The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation.
19. Droop, M.R. (1968). Vitamin B12 and marine ecology. IV. The kinetics of uptake, growth and inhibition in Monochrysis lutheri. Journal of the Marine Biological Association of the United Kingdom *48*, 689–733.
20. Edwards, K.F., Thomas, M.K., Klausmeier, C.A., and Litchman, E. (2012). Allometric scaling and taxonomic variation in nutrient utilization traits and maximum growth rate of phytoplankton. Limnol Oceanogr *57*, 554–566. https://doi.org/10.4319/lo.2012.57.2.0554.
21. Sibbald, S.J., and Archibald, J.M. (2020). Genomic insights into plastid evolution. Genome Biol Evol *12*, 978–990. https://doi.org/10.1093/GBE/EVAA096.
22. Gregory, T.R. (2005). The C-value enigma in plants and animals: A review of parallels and an appeal for partnership. In Annals of Botany, pp. 133–146. https://doi.org/10.1093/aob/mci009.
23. Michael, T.P. (2014). Plant genome size variation: Bloating and purging DNA. Brief Funct Genomic Proteomic *13*, 308–317. https://doi.org/10.1093/bfgp/elu005.
24. Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I.J. (2018). Genome size diversity and its impact on the evolution of land plants. Preprint at MDPI AG, https://doi.org/10.3390/genes9020088 https://doi.org/10.3390/genes9020088.
25. Ebenezer, V., Hu, Y., Carnicer, O., Irwin, A.J., Follows, M.J., and Finkel, Z. V. (2022). Elemental and macromolecular composition of the marine Chloropicophyceae, a major group of oceanic photosynthetic picoeukaryotes. Limnol Oceanogr *67*, 540–551. https://doi.org/10.1002/lno.12013.
26. Liefer, J.D., Garg, A., Fyfe, M.H., Irwin, A.J., Benner, I., Brown, C.M., Follows, M.J., Omta, A.W., and Finkel, Z. V. (2019). The macromolecular basis of phytoplankton C:N:P under nitrogen starvation. Front Microbiol *10*. https://doi.org/10.3389/fmicb.2019.00763.
27. Marañón, E., Cermeño, P., López-Sandoval, D.C., Rodríguez-Ramos, T., Sobrino, C., Huete-Ortega, M., Blanco, J.M., and Rodríguez, J. (2013). Unimodal size scaling of phytoplankton growth and the size dependence of nutrient uptake and use. Ecol Lett *16*, 371–379. https://doi.org/10.1111/ele.12052.
28. Chevin, L.M. (2011). On measuring selection in experimental evolution. Biol Lett *7*, 210–213. https://doi.org/10.1098/rsbl.2010.0580.
29. Almassalha, L.M., Carignano, M., Pujadas Liwag, E., Shun Li, W., Gong, R., Acosta, N., Dunton, C.L., Carrillo Gonzalez, P., Carter, L.M., Kakkaramadam, R., et al. (2025). Chromatin conformation, gene transcription, and nucleosome remodeling as an emergent system.

30. Ho, T.Y., Quigg, A., Finkel, Z. V., Milligan, A.J., Wyman, K., Falkowski, P.G., and Morel, F.M.M. (2003). The elemental composition of some marine phytoplankton. J Phycol *39*, 1145–1159. https://doi.org/10.1111/j.0022-3646.2003.03-090.x.
31. Cavalier-Smith, T. (2005). Economy, speed and size matter: Evolutionary forces driving nuclear genome miniaturization and expansion. In Annals of Botany, pp. 147–175. https://doi.org/10.1093/aob/mci010.
32. Connolly, J.A., Oliver, M.J., Beaulieu, J.M., Knight, C.A., Tomanek, L., and Moline, M.A. (2008). Correlated evolution of genome size and cell volume in diatoms (Bacillariophyceae). J Phycol *44*, 124–131. https://doi.org/10.1111/j.1529-8817.2007.00452.x.
33. Wang, H., Wu, P., Xiong, L., Kim, H.S., Kim, J.H., and Ki, J.S. (2024). Nuclear genome of dinoflagellates: Size variation and insights into evolutionary mechanisms. Preprint at Elsevier GmbH, https://doi.org/10.1016/j.ejop.2024.126061 https://doi.org/10.1016/j.ejop.2024.126061.
34. Litchman, E., Klausmeier, C.A., Schofield, O.M., and Falkowski, P.G. (2007). The role of functional traits and trade-offs in structuring phytoplankton communities: Scaling from cellular to ecosystem level. Ecol Lett *10*, 1170–1181. https://doi.org/10.1111/j.1461-0248.2007.01117.x.
35. Castillo-Davis, C.I., Mekhedov, S.L., Hartl, D.L., Koonin, E. V., and Kondrashov, F.A. (2002). Selection for short introns in highly expressed genes. Nat Genet *31*. https://doi.org/10.1038/ng940.
36. Shenhav, L., and Zeevi, D. Resource conservation manifests in the genetic code.
37. Sawyer*'+, S.A., and Hartlt, D.L. (1992). Population Genetics of Polymorphism and Divergence.
38. Sommeria-Klein, G., Watteaux, R., Ibarbalz, F.M., Karlusich, J.J.P., Iudicone, D., Bowler, C., and Morlon, H. (2021). Global drivers of eukaryotic plankton biogeography in the sunlit ocean. Science (1979) *374*, 594–599. https://doi.org/10.1126/science.abb3717.
39. Sunagawa, S., Acinas, S.G., Bork, P., Bowler, C., Babin, M., Boss, E., Cochrane, G., de Vargas, C., Follows, M., Gorsky, G., et al. (2020). Tara Oceans: towards global ocean ecosystems biology. Preprint at Nature Research, https://doi.org/10.1038/s41579-020-0364-5 https://doi.org/10.1038/s41579-020-0364-5.
40. Leconte, J., Benites, L.F., Vannier, T., Wincker, P., Piganeau, G., and Jaillon, O. (2020). Genome resolved biogeography of mamiellales. Genes (Basel) *11*. https://doi.org/10.3390/genes11010066.
41. Partensky, F., and Garczarek, L. (2010). Prochlorococcus: Advantages and limits of minimalism. Ann Rev Mar Sci *2*, 305–331. https://doi.org/10.1146/annurev-marine-120308-081034.
42. Ngugi, D.K., Acinas, S.G., Sánchez, P., Gasol, J.M., Agusti, S., Karl, D.M., and Duarte, C.M. (2023). Abiotic selection of microbial genome size in the global ocean. Nat Commun *14*. https://doi.org/10.1038/s41467-023-36988-x.
43. Sterner, R.W., and Elser, J.J. (2002). Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere: Robert W. Sterner, James J. Elser, Peter Vitousek: 9780691074917: Amazon.com: Books (Princeton University Press).
44. Longhurst, A. (2007). Ecological geography of the sea (Academic Press).
45. Karl, D.M. (2014). Microbially mediated transformations of phosphorus in the sea: New views of an old cycle. Ann Rev Mar Sci *6*, 279–337. https://doi.org/10.1146/annurev-marine-010213-135046.
46. Burson, A., Stomp, M., Akil, L., Brussaard, C.P.D., and Huisman, J. (2016). Unbalanced reduction of nutrient loads has created an offshore gradient from phosphorus to nitrogen

limitation in the North Sea. Limnol Oceanogr *61*, 869–888. https://doi.org/10.1002/lno.10257.
47. Duhamel, S., Diaz, J.M., Adams, J.C., Djaoudi, K., Steck, V., and Waggoner, E.M. (2021). Phosphorus as an integral component of global marine biogeochemistry. Preprint at Nature Research, https://doi.org/10.1038/s41561-021-00755-8 https://doi.org/10.1038/s41561-021-00755-8.
48. Gerace, S.D., Yu, J., Moore, J.K., and Martiny, A.C. (2025). Observed declines in upper ocean phosphate-to-nitrate availability. Proc Natl Acad Sci U S A *122*. https://doi.org/10.1073/pnas.2411835122.

Acknowledgments

We are grateful to all present and past Genophy group members for stimulating discussions and comments, especially Frederic Sanchez, Anais Labecot and Julien Ferrero for help with the cultures and pilot experimentations. We would like to thank the GenoToul bioinformatic platform for access to the computing facilities and David Pecqueur and Christophe Salmeron the BIOPIC platform, for access and support to the cytometry facilities.

Funding
This project was funded by the EU Horizon 2020 Marie Skłodowska- Curie program (H2020-MSCA-IF-2020) under grant number 101030734 to CC. Additional financial support was provided by the French Agence Nationale de la Recherche under grant agreement ELVIRA ANR-21-CE20-0041 and PHYTOMICS ANR-21CE02-0026.

Author contributions:

Conceptualization: CC, GP

Methodology: CC, MK, OC, SG, GP

Investigation: CC, MK, GP

Funding acquisition: CC, SG, GP

Writing – original draft: CC, GP

Writing – review & editing: CC, MK, SG, GP

Competing interests: The Authors declare that they have no competing interests.

Data and materials availability:

All data are available in the main text or the supplementary materials.