

Article

Codon Usage Bias in Phytoplankton

Marc Krasovec ^{1,*} and Dmitry A. Filatov ^{2,*}

¹ Sorbonne Université, CNRS, UMR 7232 Biologie Intégrative des Organismes Marins (BIOM), Observatoire Océanologique, 66650 Banyuls-sur-Mer, France

² Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK

* Correspondence: krasovec@obs-banyuls.fr (M.K.); dmitry.filatov@plants.ox.ac.uk (D.A.F.)

Abstract: Non-random usage of synonymous codons, known as “codon bias”, has been described in many organisms, from bacteria to *Drosophila*, but little is known about it in phytoplankton. This phenomenon is thought to be driven by selection for translational efficiency. As the efficacy of selection is proportional to the effective population size, species with large population sizes, such as phytoplankton, are expected to have strong codon bias. To test this, we measured codon bias in 215 strains from Haptophyta, Chlorophyta, Ochrophyta (except diatoms that were studied previously), Dinophyta, Cryptophyta, Ciliophora, unicellular Rhodophyta and Chlorarachniophyta. Codon bias is modest in most groups, despite the astronomically large population sizes of marine phytoplankton. The strength of the codon bias, measured with the effective number of codons, is the strongest in Haptophyta and the weakest in Chlorarachniophyta. The optimal codons are GC-ending in most cases, but several shifts to AT-ending codons were observed (mainly in Ochrophyta and Ciliophora). As it takes a long time to reach a new equilibrium after such shifts, species having AT-ending codons show a lower frequency of optimal codons compared to other species. Genetic diversity, calculated for species with more than three strains sequenced, is modest, indicating that the effective population sizes are many orders of magnitude lower than the astronomically large census population sizes, which helps to explain the modest codon bias in marine phytoplankton. This study represents the first comparative analysis of codon bias across multiple major phytoplankton groups.



Citation: Krasovec, M.; Filatov, D.A. Codon Usage Bias in Phytoplankton.

J. Mar. Sci. Eng. **2022**, *10*, 168.

<https://doi.org/10.3390/jmse10020168>

jmse10020168

Academic Editor: Linda Medlin

Received: 30 November 2021

Accepted: 26 January 2022

Published: 27 January 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: codon usage bias; phytoplankton; mutation bias; effective population size

1. Introduction

Codon usage bias (CUB)—the preferential use of one of the synonymous codons encoding the same amino acid—is an important feature of pro- and eukaryotic genomes [1,2]. This bias is known in many species of bacteria [3,4], plants [5], insects [6], green algae [7] and diatoms [8] and appears to be universal. However, it has not yet been systematically characterised across a wide range of phytoplankton groups. Phytoplankton includes species from different domains of the tree of life, and their diversity of life forms is striking. Many phytoplankton groups went through multiple rounds of secondary endosymbiosis [9,10] and acquired plastids independently of each other [11]. Phytoplankton cell sizes range over nine orders of magnitude [12] and their genomes span from only 13 Mb in the prasinophyte species *Ostreococcus tauri* [13] to over 200 Gb in dinoflagellate *Prorocentrum* sp. [14]. Phytoplankton genomes can have peculiar architectures, such as the division of the genome into macro- and micro-nuclei in *Paramecium* (Ciliophora), and tandemly repeated single-exon genes as reported for dinoflagellate *Polarella glacialis* [15]. The aim of this paper is to characterise and compare CUB across a wide diversity of phytoplankton groups.

CUB is thought to be caused by natural selection to optimise translation accuracy and efficiency [2,4,6,16]. The support for this hypothesis was found in bacteria, where it was shown that the most frequently used codons correspond to the most abundant tRNAs in the cell [17–19]. Direct analyses of protein elongation in *Escherichia coli* demonstrated that, for stretches of frequent (preferred) codons, the rate of elongation is six-times faster

compared to stretches of rare (unpreferred) codons [20]. A phenotypic effect of the artificial replacement of preferred with unpreferred synonymous codons was demonstrated in *Drosophila* [21]. The strength of CUB was shown to correlate with gene expression [22] and with effective population size across a wide range of organisms, which also supports the idea that CUB is driven by natural selection [1,2,23]. The selective advantage of optimal codons is generally modest, of the order of 10^{-6} [2], and a large effective population size is needed for selection to overcome drift. However, recent studies report that a small fraction of codons are under much stronger selection for synonymous codon usage [24].

Apart from selection for translational efficiency, CUB can be affected by neutral processes, such as mutation bias, which is present in almost all species studied previously [25]. The effect of this bias can go in the same or the opposite direction of the CUB. For example, in the haptophyte *Emiliania huxleyi*, the mutation bias is toward GC nucleotides and optimal codons are GC-ending [26], whereas, in yeast, the mutation bias is toward AT nucleotides [27,28] but the optimal codons are GC-ending. In most studied species, GC-ending codons are preferred to AT-ending codons [29], but this is not always the case and variations are observed even between closely related species. For example, in *Drosophila*, most species have GC-ending-preferred codons, whereas *Drosophila willistoni* shifted to AT-ending-preferred codons [30].

Another neutral directional process that affects GC content and CUB is GC-biased gene conversion (BGC)—the preferential resolution of (G or C):(A or T) DNA heteroduplexes towards G or C by the cellular DNA repair system [31]. Such heteroduplexes can be created by recombination at the heterozygous sites with one allele represented by A or T and the other allele by G or C. Resolution of these heteroduplexes is slightly biased towards GC and this bias was estimated to range from 0.014 in yeast [32] to 0.36 in humans [33]. BGC is thought to be a powerful force driving the evolution of higher GC content in actively recombining regions [23,34]. Such genomic regions are typically gene-rich and BGC can inflate the GC content at the third-codon positions, causing GC-biased codon usage.

Little is known about evolutionary processes in marine phytoplankton generally [35] and about CUB in particular [7,8]. A recent analysis of CUB in diatoms revealed wide variation in the extent and the direction of CUB, even within the same genera [8], illustrating the evolutionary lability of this trait. Here, we extend the analysis of CUB to other phytoplankton groups, taking advantage of the large dataset generated by the Microbial Eukaryote Transcriptome Sequencing Project (MMETSP; [36,37]). This dataset allowed us to describe the patterns and infer the evolution of codon bias in six major eukaryotic phytoplankton groups and one unicellular pigmented group and to compare them to one zooplanktonic group.

2. Materials and Methods

2.1. Data Preparation

The reassembled transcriptomes [37] from the MMETSP project [36] were used to explore the CUB in different eukaryotic groups of phytoplankton. We selected 8 groups with more than 5 strains sequenced: Haptophyta, Chlorophyta, Ochrophyta, Dinophyta, Cryptophyta, Ciliophora, Rhodophyta and Chlorarachniophyta. Diatoms were not included in the Ochrophyta group because they had already been analysed [8]. Before the CUB analysis, we excluded the sequences shorter than 300 bp to remove the non-coding RNA and fragmented coding sequences with uncertain open reading frames (ORFs). Then, we looked for the longest ORF in all of the six possible ORFs of each sequence and removed sequences that (i) contained an internal stop codon or (ii) had more than 1 possible ORF. The remaining set of transcripts was reverse-complemented, whenever necessary, to orient the ORF 5' to 3', and we trimmed the 5' end to ensure that the sequence started with the first-codon position in the ORF. Transcriptomes with less than 500 sequences remaining after filtering were discarded.

2.2. Codon Bias Analysis

We used CodonW (<http://codonw.sourceforge.net>, accessed on 26 January 2022) with each clean transcriptome to identify the preferred codons (by a 2-way Chi squared contingency test between 2 subsets of genes determined by a correspondence analysis [38]), and we calculated the frequency of optimal codons (FOP; [17]), the effective number of codons (ENC; [39]) and the GC and GC3 content of the whole transcriptome (-total option). Several strains had 2 to 4 different sequenced transcriptomes available at MMETSP. In those cases, the FOP, ENC, GC and GC3 contents of a strain were calculated as the average of the different values obtained from the different MMETSP transcriptomes; the differences between the transcriptomes for the same strain were negligible.

2.3. The Analysis of Genetic Diversity within Species

To estimate the genetic diversity within species, we measured silent pairwise divergence between strains from the same species when at least 2 strains per species were sequenced. This approach allows one to identify highly divergent strains (if any) and thus is preferable for phytoplankton, where cryptic species are not uncommon. We ran Orthofinder [40] on the cleaned transcriptomes and selected the single-copy orthologous genes. Then, we aligned the orthologous genes with Prank [41] and calculated the silent divergence between each strain separately with SNAP [42]. We analysed the species *Karenia brevis*, *Ostreococcus mediterraneus*, *Bigeloviella natans*, *Emiliania huxleyi*, *Hemiselmis andersenii* and *Heterosigma akashiwo*, *Lotharella globosa* and *Pycnococcus provasolii*. For *Ostreococcus mediterraneus*, we removed the strain MMETSP0932 because it highly limited the number of single-copy orthologous genes.

3. Results

3.1. Codon Bias Analysis

In total, we analysed 334 transcriptomes from 215 strains, corresponding to 184 species from six phytoplankton groups, one unicellular pigmented group and one zooplankton group. The number of strains representing each group ranged from 8 to 49. The number of genes analysed per transcriptome ranged (after filtering; see Methods) from 523 to 33,669 (Table 1).

The effective number of codons (ENC)—a measure of codon bias [39], ranging from $ENC = 61$ for random codon use, to $ENC = 20$ for extremely strong codon bias—was calculated for all groups analysed. The average ENC was modest in all groups, ranging from 46.4 in Haptophyta to 58.0 in Chlorarachniophyta (Tables 1 and S1). Moreover, the ENC variation within each group was high, with more than 10 points of difference between a lower and higher ENC, except Chlorarachniophyta ($54.4 < ENC < 59.7$) and unicellular Rhodophyta ($49.4 < ENC < 58.8$). However, these two groups were represented with only 12 and 8 strains in our dataset, respectively. In Chlorarachniophyta, the lowest ENC was 54.4, while all other species had an $ENC > 56.9$, with a high ENC ~ 59 in all strains of *Bigeloviella natans* and *Norrsiella sphaerica*, indicating that codon usage is nearly random in these species. More generally, considering all groups, 97 strains had an ENC between 40 and 50, 116 strains were above 50 and only four strains had an ENC lower than 40 (one Chlorophyta and three in Ochrophyta). There was a particularly low ENC of 30.2 in the case of *Paraphysomonas vestita* in Ochrophyta (MMETSP1107), which was surprisingly low compared to other species from this genus, which had an $ENC > 57$ (Table S1).

Table 1. Distribution of GC content at the 3rd codon position (GC3) and average codon bias statistics for eight phytoplankton groups. Data are provided in Table S1.

Group	N _{strains}	N _{species}	GC3 (%)							Average			
			<25	25–35	>35–45	>45–55	>55–65	>65–75	>75	GC3	GC	FOP	ENC
Haptophyta	24	19	0	0	0	0	3	10	11	72.9	64.7	0.49	45.9
Cryptophyta	19	16	0	0	0	0	8	10	1	66.7	58.2	0.45	51.6
Dinophyta	44	41	0	0	0	4	15	17	8	65.8	59.6	0.46	51.5
Chlorophyta	49	38	0	0	1	6	9	18	15	68.9	61.1	0.48	47.9
Unicellular Rhodophyta	8	8	0	0	1	3	3	1	0	55.4	54.1	0.45	54.2
Chlorarachniophyta	12	9	0	1	0	6	5	0	0	53.4	50.9	0.43	58.0
Ciliophora	19	19	0	1	6	4	6	2	0	51.0	48.7	0.44	52.9
Ochrophyta	40	34	3	4	3	3	10	13	4	56.6	54.3	0.44	50.2

Another measure of CUB—the frequency of optimal codons (FOP; [17])—ranged from the lowest FOP = 28.1% in Ochrophyta to the highest FOP = 54.1% in Chlorophyta. As with the ENC, the FOP values were modest in most of the groups, with no FOP values above 50% observed in Chlorarachniophyta, Ciliophora and Cryptophyta. Across all the strains analysed, only 35 strains had an FOP higher than 50% (Table S1). The highest FOP average for a phytoplankton group (49.2%) was calculated in Haptophytes, which also showed the lowest ENC average (45.9), indicating that codon bias in this group is particularly strong and possibly the strongest of all eukaryotic phytoplankton (Table 1).

As expected, the correlations between the CUB statistics were highly significant (Figure 1), implying a strong link between FOP, ENC and the GC composition at synonymous positions. However, for a dozen strains with FOP < 0.4, this correlation broke down (brown and yellow dots on Figure 1). Both ENC and FOP were low for these strains (as low as ENC = 30.2 for *Paraphysomonas vestita*), with the ENC indicating strong codon bias, whereas FOP indicated a low proportion of optimal codons. This disparity is likely driven by mutational biases or a recent shift in the set of optimal codons. Indeed, these low-FOP strains had AT-ending optimal codons (Table S1), whereas most of the strains had GC-ending codons.

3.2. GC Content and Optimal Codon Shift

GC content at the third-codon positions (GC3) was highly biased toward GC, with only 20 strains out of 215 being AT-rich (i.e., having GC3 < 45%). Only 3 out of 215 strains had GC3 lower than 25% (all in Ochrophyta) and six were between 25 and 35% (Table 1). Four groups contained AT-rich species (Ochrophyta, Ciliophora, unicellular Rhodophyta, Chlorarachniophyta), with a higher proportion of AT-rich strains in Ochrophyta (25.0%) and Ciliophora (36.8%). Ochrophyta species showed the largest variation in GC3 content, containing highly AT- and GC-rich species with GC3 < 25% and GC3 > 75%, respectively. In Chlorophyta, only one species was in the AT-rich category (*Picochlorum oklahomensis* CCMP2329), but it was a borderline case with GC3 = 44.3%. Highly GC-rich species constituted the majority, with 39 strains with GC3 > 75% and 72 with GC3 > 65%. Two groups (Haptophytes and Cryptophytes) contained only species with GC3 higher than 55% and were the only ones where no strains had a lower GC3 than the total GC%.

Twelve species were represented by at least two strains in the full dataset (Table S1). We compared the preferred codons between strains of the same species (Supplementary Material, file Archive.zip) to assess the consistency across the strains. In most cases, the preferred codons were exactly the same (*Karenia brevis*, *Phaeocystis antarctica*, *Hemiselmis andersenii*, *Bigelowiella natans*). However, in the case of *Ostreococcus mediterraneus*, we observed a small variation in the set of preferred codons. In particular, the preferred codon for the isoleucine was AUU in four *O. mediterraneus* strains and AUC in one of the strains of this species.

In general, the last nucleotide of the preferred codons was always a G or C for 284 (GC% = 100) out of the 334 transcriptomes analysed (Table S2), meaning that preferred AT-rich codons are very rare in phytoplankton. Moreover, only 16 transcriptomes had only A- or T-ending-preferred codons—less than 5% of the transcriptomes analysed. These 16 transcriptomes belonged to Ciliophora, Chlorarachniophyta and Ochrophyta. In the case of Chlorarachniophyta, this was MMETSP0110 *Gymnochlora* CCMP2014, where a single species out of nine contained 100% AT-ending-preferred codons, whereas the others were 100% GC-ending-preferred codons.

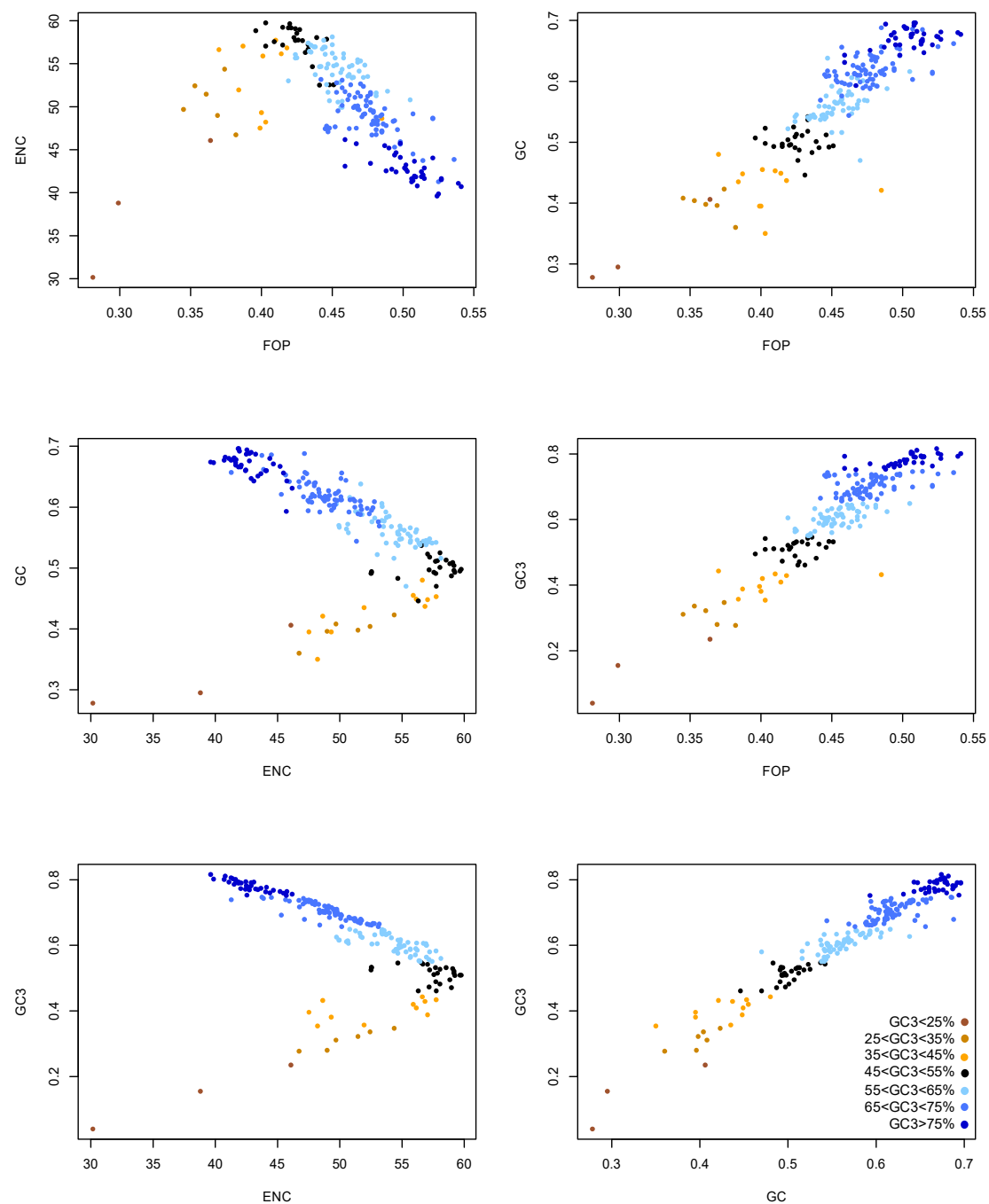


Figure 1. Correlation between codon usage bias and GC content across the phytoplankton species analysed. FOP, ENC, GC and GC3 are strongly correlated with each other, though the correlation is reversed for the species with AT-rich-preferred codons (orange and brown data points on the left three panels). All correlations are highly significant (Pearson correlation, p -value < 0.001). Original data are provided in Table S1.

3.3. Sequence Diversity within Species

To measure genetic variation within species, we calculated the pairwise silent divergence for all pairs of strains within eight species (Table S3), with 8 to 203 genes analysed per species. Pairwise silent divergence between the strains ranged from 0.00 to 0.03 (average 0.011) depending on the strain and the phytoplankton group analysed (Table S3). Surprisingly, we found no sequence variation between the sequenced strains of *Bigelowiella natans* and *Hemiselms andersenii*. The lack of genetic diversity suggests small effective population

sizes and a low efficacy of selection in these species. Indeed, codon usage was almost random in *B. natans* (ENC~59). Alternatively, low diversity may reflect the local prevalence of a single clone at the sampling location, e.g., for *B. natans*, the three strains were isolated in the same region—the northwest Atlantic Ocean.

4. Discussion

4.1. Patterns of CUB in Phytoplankton

Our analysis documented the extent of codon bias across a wide variety of phytoplankton, including representatives from eight major phytoplankton groups. This significantly expands the work on phytoplankton CUB, which was previously limited to diatoms [8] and Mamiellophyceae [7]. The current analysis revealed substantial variation in CUB among phytoplankton groups. Despite this variation, there were clear patterns in CUB across the groups. Codon bias was relatively modest in all groups, with the strongest overall CUB observed in Haptophyta (average FOP = 0.49 and ENC = 45.9), which was stronger than previously observed in diatoms but comparable to the CUB in terrestrial macro-organisms with large population sizes, such as *Drosophila* [30]. According to the observed patterns in CUB, groups can be classified into those having: (1) weak CUB (e.g., Chlorarachniophyta, where ENC~59); (2) stronger CUB with predominantly GC-ending optimal codons (e.g., Haptophyta) and (3) stronger CUB with both AT- and GC-ending optimal codons present (e.g., Ochrophyta).

4.2. The Prevalence of GC-Rich Optimal Codons

GC-ending optimal codons were prevalent in Haptophyta, Cryptophyta, Dinophyta and Chlorophyta (Table S1). This contrasts with prokaryotic phytoplankton (cyanobacteria), where the AT-ending codons are more frequent, except for the genus *Synechococcus* [43]. However, besides cyanobacteria, GC-ending optimal codons are prevalent in most other species studied in bacteria, plants, fungi and metazoans [1,23,44]. The reason that most of the species have GC-rich optimal codons is unknown, but selection for optimization of translation and GC-biased gene conversion are thought to be two main reasons for the prevalence of GC-ending codons [23]. It has been demonstrated in *Escherichia coli* that the CUB reduces the cost of missense and nonsense translation errors [4]. Although BGC is caused by a process unrelated to natural selection, the outcome of this process is almost identical to that of natural selection for G and C relative to A and T in GC-changing polymorphic sites—the ones where one allele is represented by G or C and the other allele by A or T. However, BGC affects coding and non-coding regions in a similar way, while selection for translational efficiency is expected to affect only coding regions, offering an opportunity to disentangle the effects of these processes. Detailed analysis of these two processes requires the availability of DNA polymorphisms with relatively large samples to reliably estimate allele frequencies [23], which was not available in our dataset.

4.3. Shifts between GC- and AT-Ending Preferred Codons

Despite the prevalence of species with GC-ending optimal codons, a number of shifts to AT-rich codons were observed in our dataset, as well as in other organisms [8,30]. Such shifts are mostly represented by one or a few species in a clade with predominantly GC-rich species, as was the case for a number of species in our dataset. In other cases, the entire clade had a high prevalence of species with optimal AT-ending codons, such as Ochrophyta in our dataset (Table S2). A previous study focused on diatoms [8] that belonged to Ochrophyta revealed that, among closely related species, e.g., in the genus *Chaetoceros*, some had AT-rich codons whereas others had GC-rich ones. In prokaryotic phytoplankton, such a shift was also reported, notably in *Prochlorococcus marinus*, where GC3 varies from 18 to 50% between strains [43]. The shifts to AT-rich codons were accompanied by a reduction in codon bias, particularly for the frequency of optimal codons. Such a reduction was reported for *D. willistoni* [30], as well as in the phytoplankton species with AT-rich optimal codons in our analysis. In particular, the shift to AT-rich codons may be a consequence of

mutation bias from GC to AT nucleotides [25,45], which can be very strong. For instance, a mutation rate from GC to AT more than six-times higher than the mutation rate from GC to AT was reported for *Arabidopsis thaliana* [46], *Mesoplasma florum* [47] and *Daphnia pulex* [48]. Furthermore, the AT-rich codons may have an advantage in particular conditions, causing a shift in selection from the usual GC-ending to AT-ending codons. For example, the particularly low GC3 content (44.3%) in *Picochlorum oklahomensis* (CCMP2329) may be associated with adaptations to the brackish conditions in the salt lake from which it was isolated. The comparison of GC content at third-codon positions and in the untranslated transcribed regions (UTRs) provides a way to disentangle the effects of selection and non-selective processes. However, the reasons for such shifts in selective pressure are unclear and many environmental variables may be considered.

4.4. Mutation Bias

Haptophytes, including the model species *Emiliana huxleyi* [49], have the strongest codon bias according to FOP and ENC statistics. One of the likely reasons for such strong CUB is the unusual GC bias in the mutation rate, as revealed by a mutation accumulation experiment in *E. huxleyi* [26]. Species with weaker CUB, such as diatoms or Chlorophyta (which are, in most cases, GC-ending species), have a higher GC to AT mutation rate compared to the rate in the opposite direction [45,50–53]. Overall, the AT-biased mutation rate opposes the effect of GC-biased gene conversion and the selection for GC-rich-preferred codons [54]. The observed GC3 is likely determined by the interplay of these opposing forces. However, in *E. huxleyi* [26], and possibly in Haptophytes generally, these forces act in the same direction, resulting in high GC3 and strong codon bias.

4.5. Low CUB and Genetic Diversity in Marine Phytoplankton

The silent divergence between strains of the same species (i.e., genetic diversity within a species) analysed here was low (~1%), indicating that their effective population sizes were relatively modest. This suggests that the efficacy of selection generally, and selection on CUB specifically, is not as high in marine phytoplankton as may be expected given their abundance in the oceans. However, both the modest codon bias and relatively low genetic diversity within species can be explained by other factors and may not reflect the effective population size (i.e., the rate of stochastic change in allele frequencies over generations). In particular, the genetic diversity in very large populations may grossly underestimate the true effective population size [55] because of past population size changes and constant adaptation across many sites [56,57]. The analysis of the specific causes limiting genetic diversity in a species requires detailed population genetic analysis [55], which cannot be done with the dataset used in this study because too few strains were analysed per species. Furthermore, the observed CUB may not accurately reflect the efficacy of selection, as discussed above. The expectations of high genetic diversity and strong codon bias in very large populations assume that populations are at an equilibrium state, which takes a long time to reach (of the order of $2N_e$ generations), and real populations hardly ever reach an equilibrium state.

4.6. Do the Peculiarities of the Genome or Cell Structure Affect CUB?

Some of the plankton groups analysed have peculiar genome features, such as the partitioning of the genome into macro- and micro-nuclei in ciliates, the huge genome sizes in dinoflagellates or the presence of a vestigial genome in Chlorarachniophyta and Cryptophyta [58]. Life cycles and ploidies are also very diverse, with species with an alternation of haploid and diploid phases in Haptophyta [59], diploid in diatoms [60] or mainly haploid in Chlorophyta [61]. Despite these genomic features, the extent and the patterns of codon bias in these groups did not stand out in any way from the rest of the phytoplankton groups analysed (Tables 1 and S1). Furthermore, different groups of phytoplankton studied here originated from different endosymbiosis events [9,10]. The green lineage evolved after the primary endosymbiosis when an eukaryotic cell engulfed and ‘domesticated’ a pho-

tosynthetic cyanobacterium. Then, secondary endosymbioses occurred multiple times in different eukaryotes kingdoms [11,62], when other eukaryotes engulfed and ‘domesticated’ photosynthetic eukaryotic cells. Such secondary endosymbiotic events leave a distinctive feature in the cell structure—additional membrane layers around the chloroplasts [9]. Haptophytes, cryptophytes, chlorarachniophytes and euglenophytes are thought to have originated via different secondary endosymbiosis events. Tertiary endosymbiotic events also have been reported, with multiple independent occurrences in dinoflagellates [63]. Regardless of the group considered, and the number of endosymbiotic events undergone, all groups show similar CUB patterns—predominantly GC-rich-preferred codons and modest codon bias. Thus, it appears unlikely that these genomic or cellular features affect CUB.

5. Conclusions

In this study, we characterised the extent and variation of codon usage bias in six phytoplankton groups, one unicellular pigmented group and one zooplanktonic group. Consistent with our previous analysis in diatoms [8], CUB was modest in all groups analysed, which is surprising as even very weak selection that is thought to drive codon bias is expected to be effective in very large populations of marine plankton. Our analysis revealed the predominance of GC-ending optimal codons in most species. Occasional shifts to AT-ending codons, observed in a few species (mainly in Ochrophyta and Ciliophora), led to a reduced frequency of optimal codons compared to other species, which likely reflects the recency of such shifts and the time needed for CUB to reach the new equilibrium. The unusual genomic features of some phytoplankton groups, such as the presence of macro- and micro-nuclei in Ciliophora, large genome sizes in Dinophyta or the involvement in secondary symbioses, do not appear to significantly affect CUB. On the other hand, genome-wide mutational bias appears to be a strong factor that can increase or reduce codon bias if it works in the same or the opposite direction to selection for codon usage. Overall, the extent and the patterns of codon usage bias in phytoplankton are fairly similar to those in other groups of organisms described previously [1–8].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jmse10020168/s1>, Table S1: Codon bias statistics from CodonW; Table S2: Last nucleotide of preferred codons; Table S3: Silent divergence between the strains of the same species. Archive.zip file includes files with preferred codons identified by CodonW.

Author Contributions: M.K. analysed the data, D.A.F. supervised the project, M.K. and D.A.F. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: This work only used previously published dataset generated by the Microbial Eukaryote Transcriptome Sequencing Project [36,37]. The MMETSP accession numbers are listed in Table S1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hershberg, R.; Petrov, D.A. Selection on Codon Bias. *Annu. Rev. Genet.* **2008**, *42*, 287–299. [[CrossRef](#)] [[PubMed](#)]
2. Duret, L. Evolution of Synonymous Codon Usage in Metazoans. *Curr. Opin. Genet. Dev.* **2002**, *12*, 640–649. [[CrossRef](#)]
3. Gouy, M.; Gautier, C. Codon Usage in Bacteria: Correlation with Gene Expressivity. *Nucleic Acids Res.* **1982**, *10*, 7055–7074. [[CrossRef](#)] [[PubMed](#)]
4. Stoletzki, N.; Eyre-Walker, A. Synonymous Codon Usage in *Escherichia coli*: Selection for Translational Accuracy. *Mol. Biol. Evol.* **2007**, *24*, 374–381. [[CrossRef](#)]
5. Qiu, S.; Bergero, R.; Zeng, K.; Charlesworth, D. Patterns of Codon Usage Bias in *Silene latifolia*. *Mol. Biol. Evol.* **2011**, *28*, 771–780. [[CrossRef](#)] [[PubMed](#)]
6. Shields, D.C.; Sharp, P.M.; Higgins, D.G.; Wright, F. “Silent” Sites in *Drosophila* Genes Are Not Neutral: Evidence of Selection among Synonymous Codons. *Mol. Biol. Evol.* **1988**, *5*, 704–716. [[CrossRef](#)]
7. Michely, S.; Toulza, E.; Subirana, L.; John, U.; Cognat, V.; Maréchal-Drouard, L.; Grimsley, N.; Moreau, H.; Piganeau, G. Evolution of Codon Usage in the Smallest Photosynthetic Eukaryotes and Their Giant Viruses. *Genome Biol. Evol.* **2013**, *5*, 848–859. [[CrossRef](#)]

8. Krasovec, M.; Filatov, D.A. Evolution of Codon Usage Bias in Diatoms. *Genes* **2019**, *10*, 894. [\[CrossRef\]](#)
9. Archibald, J.M. Endosymbiosis and Eukaryotic Cell Evolution. *Curr. Biol.* **2015**, *25*, 911–921. [\[CrossRef\]](#)
10. Reyes-Prieto, A.; Weber, A.P.M.; Bhattacharya, D. The Origin and Establishment of the Plastid in Algae and Plants. *Ann. Rev. Genet.* **2007**, *41*, 147–168. [\[CrossRef\]](#)
11. Baurain, D.; Brinkmann, H.; Petersen, J.; Rodríguez-Ezpeleta, N.; Stechmann, A.; Demoulin, V.; Roger, A.J.; Burger, G.; Lang, B.F.; Philippe, H. Phylogenomic Evidence for Separate Acquisition of Plastids in Cryptophytes, Haptophytes, and Stramenopiles. *Mol. Biol. Evol.* **2010**, *27*, 1698–1709. [\[CrossRef\]](#)
12. Marañón, E. Cell Size as a Key Determinant of Phytoplankton Metabolism and Community Structure. *Ann. Rev. Mar. Sci.* **2015**, *7*, 241–264. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Blanc-Mathieu, R.; Verhelst, B.; Derelle, E.; Rombauts, S.; Bouget, F.-Y.; Carré, I.; Château, A.; Eyre-Walker, A.; Grimsley, N.; Moreau, H.; et al. An Improved Genome of the Model Marine Alga *Ostreococcus tauri* Unfolds by Assessing Illumina de Novo Assemblies. *BMC Genom.* **2014**, *15*, 1103. [\[CrossRef\]](#) [\[PubMed\]](#)
14. LaJeunesse, T.C.; Lambert, G.; Andersen, R.A.; Coffroth, M.A.; Galbraith, D.W. *Symbiodinium* (Pyrrhophyta) Genome Sizes (Dna Content) Are Smallest Among Dinoflagellates. *J. Phycol.* **2005**, *41*, 880–886. [\[CrossRef\]](#)
15. Stephens, T.G.; González-Pech, R.A.; Cheng, Y.; Mohamed, A.R.; Burt, D.W.; Bhattacharya, D.; Ragan, M.A.; Chan, C.X. Genomes of the Dinoflagellate *Polarella glacialis* Encode Tandemly Repeated Single-Exon Genes with Adaptive Functions. *BMC Biol.* **2020**, *18*, 56. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Akashi, H. Synonymous Codon Usage in *Drosophila melanogaster*: Natural Selection and Translational Accuracy. *Genetics* **1994**, *136*, 927–935. [\[CrossRef\]](#)
17. Ikemura, T. Correlation between the Abundance of *Escherichia coli* Transfer RNAs and the Occurrence of the Respective Codons in Its Protein Genes: A Proposal for a Synonymous Codon Choice That Is Optimal for the *E. coli* Translational System. *J. Mol. Biol.* **1981**, *151*, 389–409. [\[CrossRef\]](#)
18. Ikemura, T. Codon Usage and tRNA Content in Unicellular and Multicellular Organisms. *Mol. Biol. Evol.* **1985**, *2*, 13–34. [\[CrossRef\]](#)
19. Yamao, F.; Andachi, Y.; Muto, A.; Ikemura, T.; Osawa, S. Levels of tRNAs in Bacterial Cells as Affected by Amino Acid Usage in Proteins. *Nucleic Acids Res.* **1991**, *19*, 6119–6122. [\[CrossRef\]](#)
20. Sørensen, M.A.; Kurland, C.G.; Pedersen, S. Codon Usage Determines Translation Rate in *Escherichia coli*. *J. Mol. Biol.* **1989**, *207*, 365–377. [\[CrossRef\]](#)
21. Carlini, D.B. Experimental Reduction of Codon Bias in the *Drosophila* Alcohol Dehydrogenase Gene Results in Decreased Ethanol Tolerance of Adult Flies. *J. Evol. Biol.* **2004**, *17*, 779–785. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Duret, L.; Mouchiroud, D. Expression Pattern and, Surprisingly, Gene Length Shape Codon Usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4482–4487. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Galtier, N.; Roux, C.; Rousselle, M.; Romiguier, J.; Figuet, E.; Glémin, S.; Bierne, N.; Duret, L. Codon Usage Bias in Animals: Disentangling the Effects of Natural Selection, Effective Population Size, and GC-Biased Gene Conversion. *Mol. Biol. Evol.* **2018**, *35*, 1092–1103. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Machado, H.E.; Lawrie, D.S.; Petrov, D.A. Pervasive Strong Selection at the Level of Codon Usage Bias in *Drosophila melanogaster*. *Genetics* **2020**, *214*, 511–528. [\[CrossRef\]](#)
25. Katju, V.; Bergthorsson, U. Old Trade, New Tricks: Insights into the Spontaneous Mutation Process from the Partnering of Classical Mutation Accumulation Experiments with High-Throughput Genomic Approaches. *Genome Biol. Evol.* **2019**, *11*, 136–165. [\[CrossRef\]](#)
26. Krasovec, M.; Rickaby, R.E.M.; Filatov, D.A. Evolution of Mutation Rate in Astronomically Large Phytoplankton Populations. *Genome Biol. Evol.* **2020**, *12*, 1051–1059. [\[CrossRef\]](#)
27. Lynch, M.; Sung, W.; Morris, K.; Coffey, N.; Landry, C.R.; Dopman, E.B.; Dickinson, W.J.; Okamoto, K.; Kulkarni, S.; Hartl, D.L.; et al. A Genome-Wide View of the Spectrum of Spontaneous Mutations in Yeast. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9272–9277. [\[CrossRef\]](#)
28. Zhu, Y.O.; Siegal, M.L.; Hall, D.W.; Petrov, D.A. Precise Estimates of Mutation Rate and Spectrum in Yeast. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 2310–2318. [\[CrossRef\]](#)
29. Sharp, P.M.; Cowe, E.; Higgins, D.G.; Shields, D.C.; Wolfe, K.H.; Wright, F. Codon Usage Patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*: A Review of the Considerable within-Species Diversity. *Nucleic Acids Res.* **1988**, *16*, 8207–8211. [\[CrossRef\]](#)
30. Vicario, S.; Moriyama, E.N.; Powell, J.R. Codon Usage in Twelve Species of *Drosophila*. *BMC Evol. Biol.* **2007**, *7*, 226. [\[CrossRef\]](#)
31. Duret, L.; Galtier, N. Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. *Annu. Rev. Genom. Hum. Genet.* **2009**, *10*, 285–311. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Mancera, E.; Bourgon, R.; Brozzi, A.; Huber, W.; Steinmetz, L.M. High-Resolution Mapping of Meiotic Crossovers and Non-Crossovers in Yeast. *Nature* **2008**, *454*, 479–485. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Halldorsson, B.V.; Hardarson, M.T.; Kehr, B.; Styrkarsdottir, U.; Gylfason, A.; Thorleifsson, G.; Zink, F.; Jonasdottir, A.; Jonasdottir, A.; Sulem, P.; et al. The Rate of Meiotic Gene Conversion Varies by Sex and Age. *Nat. Genet.* **2016**, *48*, 1377–1384. [\[CrossRef\]](#) [\[PubMed\]](#)

34. Lassalle, F.; Périan, S.; Bataillon, T.; Nesme, X.; Duret, L.; Daubin, V. GC-Content Evolution in Bacterial Genomes: The Biased Gene Conversion Hypothesis Expands. *PLoS Genet.* **2015**, *11*, e1004941. [[CrossRef](#)] [[PubMed](#)]
35. Rengefors, K.; Kremp, A.; Reusch, T.B.H.; Wood, A.M. Genetic Diversity and Evolution in Eukaryotic Phytoplankton: Revelations from Population Genetic Studies. *J. Plankton Res.* **2017**, *39*, 165–179. [[CrossRef](#)]
36. Keeling, P.J.; Burki, F.; Wilcox, H.M.; Allam, B.; Allen, E.E.; Amaral-Zettler, L.A.; Armbrust, E.V.; Archibald, J.M.; Bharti, A.K.; Bell, C.J.; et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biol.* **2014**, *12*, e1001889. [[CrossRef](#)]
37. Johnson, L.K.; Alexander, H.; Brown, C.T. Re-Assembly, Quality Evaluation, and Annotation of 678 Microbial Eukaryotic Reference Transcriptomes. *Gigascience* **2019**, *8*, giy158. [[CrossRef](#)]
38. Greenacre, M.J. *Heory and Applications of Correspondence Analysis*; Academic Press: London, UK, 1984.
39. Wright, F. The “Effective Number of Codons” Used in a Gene. *Gene* **1990**, *87*, 23–29. [[CrossRef](#)]
40. Emms, D.M.; Kelly, S. OrthoFinder: Solving Fundamental Biases in Whole Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy. *Genome Biol.* **2015**, *16*, 157. [[CrossRef](#)]
41. Löytynoja, A. Phylogeny-Aware Alignment with PRANK. *Methods Mol. Biol.* **2014**, *1079*, 155–170. [[CrossRef](#)]
42. Korber-Irrgang, B. HIV Signature and Sequence Variation Analysis. *Comput. Anal. HIV Mol. Seq.* **2000**, *4*, 55–72.
43. Prabha, R.; Singh, D.P.; Sinha, S.; Ahmad, K.; Rai, A. Genome-Wide Comparative Analysis of Codon Usage Bias and Codon Context Patterns among Cyanobacterial Genomes. *Mar. Genom.* **2017**, *32*, 31–39. [[CrossRef](#)] [[PubMed](#)]
44. Behura, S.K.; Severson, D.W. Codon Usage Bias: Causative Factors, Quantification Methods and Genome-Wide Patterns: With Emphasis on Insect Genomes. *Biol. Rev. Camb. Philos. Soc.* **2013**, *88*, 49–61. [[CrossRef](#)] [[PubMed](#)]
45. Krasovec, M.; Eyre-Walker, A.; Sanchez-Ferandin, S.; Piganeau, G. Spontaneous Mutation Rate in the Smallest Photosynthetic Eukaryotes. *Mol. Biol. Evol.* **2017**, *34*, 1770–1779. [[CrossRef](#)] [[PubMed](#)]
46. Ossowski, S.; Schneeberger, K.; Lucas-Lledó, J.I.; Warthmann, N.; Clark, R.M.; Shaw, R.G.; Weigel, D.; Lynch, M. The Rate and Molecular Spectrum of Spontaneous Mutations in *Arabidopsis thaliana*. *Science* **2010**, *327*, 92–94. [[CrossRef](#)]
47. Sung, W.; Ackerman, M.S.; Miller, S.F.; Doak, T.G.; Lynch, M. Drift-Barrier Hypothesis and Mutation-Rate Evolution. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 18488–18492. [[CrossRef](#)]
48. Flynn, J.M.; Chain, F.J.J.; Schoen, D.J.; Cristescu, M.E. Spontaneous Mutation Accumulation in *Daphnia pulex* in Selection-Free vs. Competitive Environments. *Mol. Biol. Evol.* **2017**, *34*, 160–173. [[CrossRef](#)]
49. Read, B.A.; Kegel, J.; Klute, M.J.; Kuo, A.; Lefebvre, S.C.; Maumus, F.; Mayer, C.; Miller, J.; Monier, A.; Salamov, A.; et al. Pan Genome of the Phytoplankton *Emiliania* Underpins Its Global Distribution. *Nature* **2013**, *499*, 209–213. [[CrossRef](#)]
50. Ness, R.W.; Morgan, A.D.; Vasanthakrishnan, R.B.; Colegrave, N.; Keightley, P.D. Extensive de Novo Mutation Rate Variation between Individuals and across the Genome of *Chlamydomonas reinhardtii*. *Genome Res.* **2015**, *25*, 1739–1749. [[CrossRef](#)]
51. López-Cortegano, E.; Craig, R.J.; Chebib, J.; Samuels, T.; Morgan, A.D.; Kraemer, S.A.; Böndel, K.B.; Ness, R.W.; Colegrave, N.; Keightley, P.D. De Novo Mutation Rate Variation and Its Determinants in *Chlamydomonas*. *Mol. Biol. Evol.* **2021**. [[CrossRef](#)]
52. Krasovec, M.; Sanchez-Brosseau, S.; Grimsley, N.; Piganeau, G. Spontaneous Mutation Rate as a Source of Diversity for Improving Desirable Traits in Cultured Microalgae. *Algal Res.* **2018**, *35*, 85–90. [[CrossRef](#)]
53. Krasovec, M.; Sanchez-Brosseau, S.; Piganeau, G. First Estimation of the Spontaneous Mutation Rate in Diatoms. *Genome Biol. Evol.* **2019**, *11*, 1829–1837. [[CrossRef](#)] [[PubMed](#)]
54. Weller, A.M.; Rödelberger, C.; Eberhardt, G.; Molnar, R.I.; Sommer, R.J. Opposing Forces of A/T-Biased Mutations and G/C-Biased Gene Conversions Shape the Genome of the Nematode *Pristionchus pacificus*. *Genetics* **2014**, *196*, 1145–1152. [[CrossRef](#)] [[PubMed](#)]
55. Filatov, D.A. Extreme Lewontin’s Paradox in Ubiquitous Marine Phytoplankton Species. *Mol. Biol. Evol.* **2019**, *36*, 4–14. [[CrossRef](#)] [[PubMed](#)]
56. Gillespie, J.H. Genetic Drift in an Infinite Population. The Pseudohitchhiking Model. *Genetics* **2000**, *155*, 909–919. [[CrossRef](#)]
57. Neher, R.A. Genetic Draft, Selective Interference, and Population Genetics of Rapid Adaptation. *Ann. Rev. Ecol. Evol. Syst.* **2013**, *44*, 195–215. [[CrossRef](#)]
58. Curtis, B.A.; Tanifuji, G.; Burki, F.; Gruber, A.; Irimia, M.; Maruyama, S.; Arias, M.C.; Ball, S.G.; Gile, G.H.; Hirakawa, Y.; et al. Algal Genomes Reveal Evolutionary Mosaicism and the Fate of Nucleomorphs. *Nature* **2012**, *492*, 59–65. [[CrossRef](#)]
59. Frada, M.J.; Bidle, K.D.; Probert, I.; de Vargas, C. In Situ Survey of Life Cycle Phases of the Coccolithophore *Emiliania huxleyi* (Haptophyta). *Env. Microbiol.* **2012**, *14*, 1558–1569. [[CrossRef](#)]
60. Bowler, C.; Allen, A.E.; Badger, J.H.; Grimwood, J.; Jabbari, K.; Kuo, A.; Maheswari, U.; Martens, C.; Maumus, F.; Otiillar, R.P.; et al. The *Phaeodactylum* Genome Reveals the Evolutionary History of Diatom Genomes. *Nature* **2008**, *456*, 239–244. [[CrossRef](#)]
61. Derelle, E.; Ferraz, C.; Rombauts, S.; Rouzé, P.; Worden, A.Z.; Robbens, S.; Partensky, F.; Degroev, S.; Echeynié, S.; Cooke, R.; et al. Genome Analysis of the Smallest Free-Living Eukaryote *Ostreococcus tauri* Unveils Many Unique Features. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11647–11652. [[CrossRef](#)]
62. Rogers, M.B.; Gilson, P.R.; Su, V.; McFadden, G.I.; Keeling, P.J. The Complete Chloroplast Genome of the Chlorarachniophyte *Bigelowiella natans*: Evidence for Independent Origins of Chlorarachniophyte and Euglenid Secondary Endosymbionts. *Mol. Biol. Evol.* **2007**, *24*, 54–62. [[CrossRef](#)] [[PubMed](#)]
63. Gagat, P.; Bodyl, A.; Mackiewicz, P.; Stiller, J.W. *Tertiary Plastid Endosymbioses in Dinoflagellates*. *Endosymbiosis*; Springer: Vienna, Austria, 2014.