UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

CAIO CESAR RIBEIRO

PROCHLOROCOCCUS AND _SYNECHOCOCCUS_ MARINE CYANOBACTERIA: A SCIENTOMETRICS REVIEW

DOIS VIZINHOS

2023

CAIO CESAR RIBEIRO

PROCHLOROCOCCUS AND *SYNECHOCOCCUS* MARINE CYANOBACTERIA: A SCIENTOMETRICS REVIEW

PROCHLOROCOCCUS E *SYNECHOCOCCU*S CIANOBACTÉRIAS MARINHAS: AVALIAÇÃO CIENTOMÉTRICA

Projeto de Trabalho de conclusão de curso de Especialização apresentado como requisito para obtenção do título de Especialização em Biologia Molecular – Habilitação Bioinformática da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientador(a): Nédia de Castilhos Ghisi

DOIS VIZINHOS

2023

CAIO CESAR RIBEIRO

PROCHLOROCOCCUS AND _SYNECHOCOCCUS_ MARINE CYANOBACTERIA: A SCIENTOMETRICS REVIEW

Projeto de Trabalho de conclusão de curso de Especialização apresentado como requisito para obtenção do título de Especialização em Biologia Molecular – Habilitação Bioinformática da Universidade Tecnológica Federal do Paraná (UTFPR).

Data de aprovação: 13/dezembro/2022

_____

Profa. Dra. Nédia de Castilhos Ghisi

Universidade Federal Tecnológica do Paraná

_____

Dra. Maristela Azevedo Linhares

TECPAR - Laboratório de Microbiologia e Toxicologia

_____

M.Sc. Gabriela Pustiglione Marinsek

Universidade Estadual Paulista Júlio de Mesquita Filho

DOIS VIZINHOS

2023

Dedico aos meus pais, amigos e professores que sempre me ensinaram tanto e torcem por mim sempre.

AGRADECIMENTOS

"O mestre da vida faz pouca distinção entre o seu trabalho e o seu lazer. Ele simplesmente persegue sua visão de excelência em tudo o que faz, deixando para os outros a decisão de saber se está trabalhando ou se divertindo. Mas ele sabe que está sempre fazendo as duas coisas simultaneamente."

Buda

RESUMO

As cianobactérias *Prochlorococcus* e *Synechococcus* são os produtores primários dominantes no oceano e realizam uma porcentagem significativa da fixação de carbono no oceano. No passado, essas células eram identificadas por citometria de fluxo, mas atualmente as técnicas genômicas trazem diversas análises que também identificam clados e ecótipos; a genômica comparativa ajudou a responder aos padrões de evolução e diversidade entre os microorganismos. Os dados genéticos e genômicos aumentaram o número de genomas de *Prochlorococcus* disponíveis e facilitaram a biologia evolutiva e a ecologia microbiana. Portanto, esta revisão resume os dados disponíveis na literatura (1992-2021; n = 480) usando o software CiteSpace para fornecer informações sobre as características específicas dos genes *Prochlorococcus* e *Synechococcus* que os cientistas podem usar para estabelecer futuras direções de pesquisa. De 1992 a 2021, a palavra-chave mais frequente e com maior centralidade foi "origem evolutiva múltipla" (0,43), seguida de citometria de fluxo (0,40), pois esse período foi o principal método utilizado para identificar cianobactérias. Em 2001-2010, as palavras-chave com maior centralidade foram "sequência" e "gene", ambas com 0,13 de centralidade; isso pode ser explicado pelo período em que o sequenciamento de genes foi desenvolvido. Nos últimos 10 anos, período 2011-2021, as palavras-chave com maior centralidade foram diversidade - 0,17; *Synechococcus* marinho - 0,10; Ecótipo *Prochlorococcus* - 0,08; expressão gênica - 0,14; fitoplâncton - 0,09; e leve - 0,07. Além disso, as características de migração de citometria de fluxo, métodos de identificação de genes, genomas, ecótipos de cianobactérias e diversidade de fitoplâncton indicam como os estudos futuros devem focar.

Palavras-chave: CiteSpace, Genoma, cianobactéria, cienciometria, *Prochlorococcus, Synechococcus*.

ABSTRACT

The cyanobacteria *Prochlorococcus* and *Synechococcus* are the dominant primary producers in the ocean and perform a significant percentage of ocean carbon fixation. In the past, these cells were identified in flow cytometry, but nowadays, the genomic techniques bring different analyses that also identify clades and ecotypes; comparative genomics has helped to answer patterns of evolution and diversity among microbes. The genetic and genomic data increased the number of available *Prochlorococcus* genomes and facilitated evolutionary biology, microbial ecology, and biological oceanography. Therefore, this review summarizes the available data in the literature (1992–2021; n=480) using CiteSpace software to provide insights into the specific characteristics of *Prochlorococcus* and *Synechococcus* genes that scientists can use for establishing future research directions. From 1992 to 2021, the most frequent and with higher centrality keyword was "multiple evolutionary origin" (0.43), followed by flow cytometry (0.40), because this period was the main method used to identify cyanobacteria. In 2001-2010, the keywords with higher centrality were "sequence" and "gene", both with 0.13 of centrality; this can be explained by the period when gene sequencing was developed. In the last 10 years period 2011-2021 the keywords with higher centrality were diversity - 0.17; marine *Synechococcus* - 0.10; *Prochlorococcus* ecotype - 0.08; gene expression - 0.14; phytoplankton - 0.09; and light - 0.07. Additionally, the migration characteristics of flow cytometry, genes identification methods, genomes, ecotypes of cyanobacteria, and phytoplankton diversity indicate how future studies should focus.

Keywords: CiteSpace, Genome, cyanobacteria, scientometrics *Prochlorococcus, Synechococcus.*

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1. INTRODUCTION

The genus *Prochlorococcus* contains the smallest photosynthetic organisms ever identified. In addition to their size, their red fluorescence in blue light distinguishes them from *Synechococcus*, which emit orange fluorescence derived from the excitation of phycobilin (Marie *et al*., 1997). Taxa of *Prochlorococcus* are present throughout the euphotic zone and are among the sea's most abundant forms of life. They are adapted to the oligotrophic ocean and inhabit areas with reduced light intensities by increasing their pigment concentration (Veldhuis & Kraay, 2000). These taxa have the most efficient carbon concentration mechanism (Hopkinson *et al*., 2014) and the highest available CO_2 fixation rates by photosynthetic pigment among phytoplankton (Hartmann *et al*., 2014). The cyanobacteria *Prochlorococcus marinus* is an abundant and significant primary producer in the open ocean, reaching over 82% of the North Pacific Subtropical Ocean (Liu *et al*., 1997).

Species of *Synechococcus* differ from typical prokaryotes due to their much more elaborate cell cycle, which, in turn, differs from that of species of *Prochlorococcus* in having higher growth rates. Taxa of *Synechococcus* are abundant in temperate and tropical euphotic zones and have a fundamental role in ocean primary productivity (Lalli & Parsons 1997). An inventory of the global distribution of *Synechococcus* reveals that it is the most abundant genus of bacteria and dominant in terms of biomass and productivity in all the (sub) tropical North Atlantic (Olson *et al*. 1990; Veldhuis & Kraay 2000).

Species of the genera *Prochlorococcus* and *Synechococcus* are among the smallest picoplankton and most abundant taxa in the collective photosynthetic activity

of cyanobacteria. They occupy deeper areas of the water column (Veldhuis & Kraay, 2000). Studies demonstrate that these two genera generally co-occur, with their depth distributions differing in absolute concentrations and vertical extension, except when the water column is homogeneous (Li *et al*. 1992, Campbell & Vaulot 1993, Veldhuis & Kraay 1993, 2000; Cesar-Ribeiro *et al.,* 2000).

For taxon-specific estimations of phytoplankton biomass, those of *Prochlorococcus* and *Synechococcus* are usually calculated separately because these two genera can be easily distinguished from pico-eukaryotes (Li *et al*. 1993, Campbell *et al*. 1994). The other most abundant prokaryotic group in marine environments is heterotrophic bacteria, which generally dominate the size range of 0.2 μm to 2 μm in cell number and biomass (Li *et al*. 1993). However, the members of this group are still little differentiated taxonomically due to the absence of pigments (Giovannoni *et al*. 1990). Still, current advances in metagenomics have made it easy to identify different taxa and even approximate their abundance.

Marine cyanobacteria within the genera *Prochlorococcus* and *Synechococcus* are estimated to be responsible for roughly 25% of ocean's net primary productivity. Because of their abundance and global distribution, *Prochlorococcus* and *Synechococcus* perform vital functions at the base of marine food webs, primarily the supply of fixed carbon to higher trophic levels (Flombaum *et al*., 2013).

Prochlorococcus is the numerically dominant phototroph in oligotrophic subtropical gyres, among the most extensive contiguous biomes on Earth (Berube *et al*., 2018). Although *Prochlorococcus* cells have the smallest genomes of known oxygenic phototrophs (~1.6–2.7 Mbp and ~2000–3000 genes), the global collective of this group harbors an immense diversity of protein-encoding genes. Recent estimates using 41 genomes of cultivated isolates suggested that the *Prochlorococcus* pan-

genome–the complete set of genes carried by all *Prochlorococcus*–contains more than 80,000 distinct genes, many of which presumably play a role in adaptation to local environmental conditions (Berube *et al*., 2018). An inventory of the global distribution of *Synechococcus* reveals that it is the most abundant genus of bacteria and dominant in terms of biomass and productivity in all the (sub) tropical North Atlantic (Olson *et al*., 1990, Veldhuis & Kraay, 2000).

Recent advances in the genomics of single cells have uncovered previously unknown marine microbial phyla and functions. They have identified a high degree of genome streamlining, mixotrophy, and metabolic specialization within bacterial cells of the surface ocean. Single-cell genomes of *Prochlorococcus* have revealed the existence of new clades with distinct ecological and physiological adaptations and a high degree of genomic and functional diversity among *Prochlorococcus* cells with nearly identical ribotypes (Berube *et al*., 2018).

Based on previous studies, we adopted CiteSpace software to make a visual bibliometric analysis of *Prochlorococcus* and *Synechococcus* research in 1992–2021. This study aims to comprehensively and systematically provide a scientometrics review. More specifically, our analysis aims to (1) summarize general characteristics of publication output and main subject categories; (2) evaluate the performance of countries, institutions, and journals. No previous studies have analyzed its research field to such depth to include aspects such as co-citation clusters, keywords, or research clusters. To bridge these gaps in extant literature by undertaking an in-depth scientometric review of global sustainability, we used CiteSpace software to conduct a visual scientometrics analysis. The most apparent advantage of CiteSpace is that it allows scholars to investigate specific research areas by analyzing the citations, co-citations, and geographical distribution, thus drawing a beneficial conclusion.

2.   MATERIALS & METHODS

Data were obtained from the Web of Science (WoS) database of Clarivate Analytics (Chen *et al.*, 2014). The searched terms included TS (Topic Search): *Prochlorococcus* AND *Synechococcus*. This database is regarded as the most important and frequently used scientific database in most fields. The whole records (cited references, titles, and abstracts) were exported to CiteSpace for subsequent analysis. CiteSpace is a Java-based scientific visualization software package used for analyzing and visualizing co-citation networks and provides various functions for understanding of network patterns, including identifying the major topic finding areas, and labeling clusters with terms from selected literature (Chen, 2006).

The search generated 1201 references between 1992 and 2021. Later, refining manually analyzed each publication by reading titles and abstracts to select only studies involving "genes." A total of 721 publications were excluded due to duplicated citations or because they were unrelated to genes, genomic, or even genetic; they were told to these cyanobacteria ecology, distribution, and identification (480 selected). Document type, publication year, source and impact factor, category, country, and author were identified. The impact factor was obtained from Journal Citation Reports (JCR) (clarivate.com/products/web-of-science). The results were analyzed using Microsoft Office Excel and CiteSpace software (Chen, 2014)*.*

We use the centrality of CiteSpace because of the comparison between data, which means the importance of each document in a co-citing network can be partially evaluated by the indicator betweenness centrality (Li *et al.*, 2017; Ouyang *et al.*, 2018). Centrality is defined in the following Equation (1).

$$\text{Centrality}\left(\text{node i}\right) = \sum_{i \neq j \neq k} \frac{\rho_{jk}\,(i)}{\rho_{jk}}$$

(1)Eq. (1), ρ_{jk} represents the shortest paths between node j and node k, and ρ_{jk} (i) is the number of those paths that pass-through node i.

The prism flowchart demonstrates the steps of identification, selection, eligibility, and inclusion of articles in this study; a total of 721 citations were excluded due to duplicated citations and publications that did not were related to genetic (Fig. 1).

3.   RESULTS AND DISCUSSION

An increase was observed in yearly publications, from one in 1992 to 33 in 2009 and 2015. These publications were classified into nine types of documents: article, procedure paper, abstract, review, letter, note, editorial material, book review, and others. Articles were predominant, comprising 93.33% of the total. The predominant language was English (100%).

Figure 1. PRISMA flowchart, showing the flow of data collection for inclusion in scientometrics.
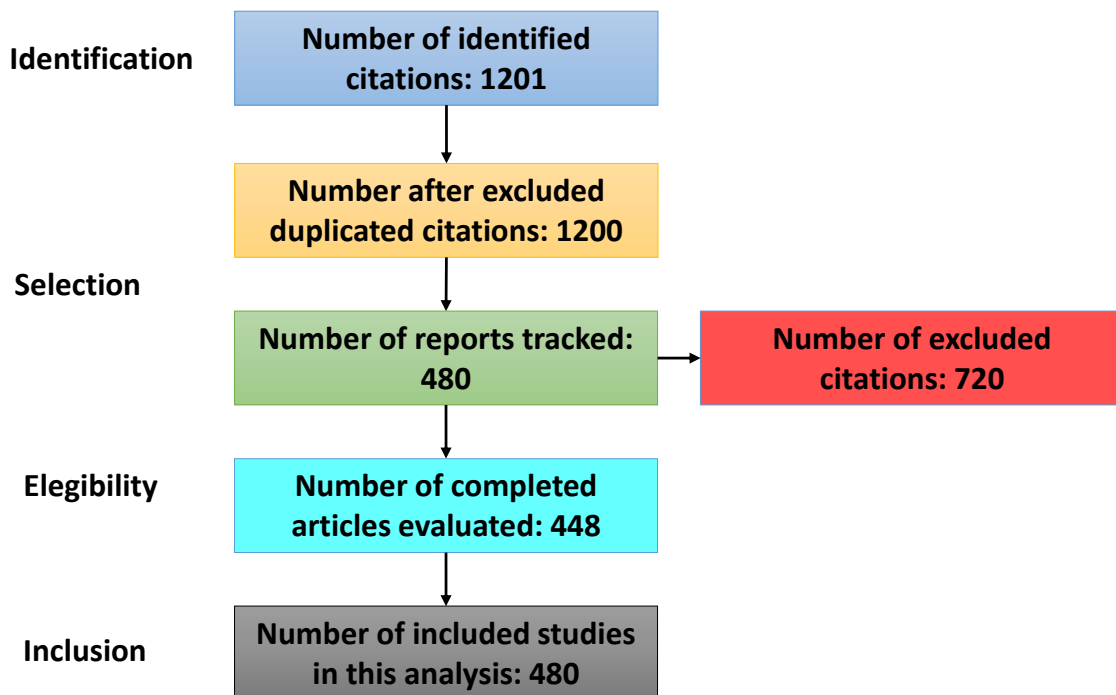
Source: Author

The articles in our refined dataset were cited 29,572 times, averaging 61.60 citations per item. The H-index is the maximum value of h for which an author or area has published H papers cited at least H times. The dispersion graph represents the mean of publications and citations per year (Figure 2). It shows that since 2002 there has been a considerable and growing increase in the publications and citations

regarding *Prochlorococcus* and *Synechococcus* genetics. In the whole period analyzed, it was not observed correlation between citations and publications (Pearson's correlation: $R^2=0.30$; $p > 0.05$).

Figure 2. Publication *vs.* Citation output performance between 1992 and 2021.
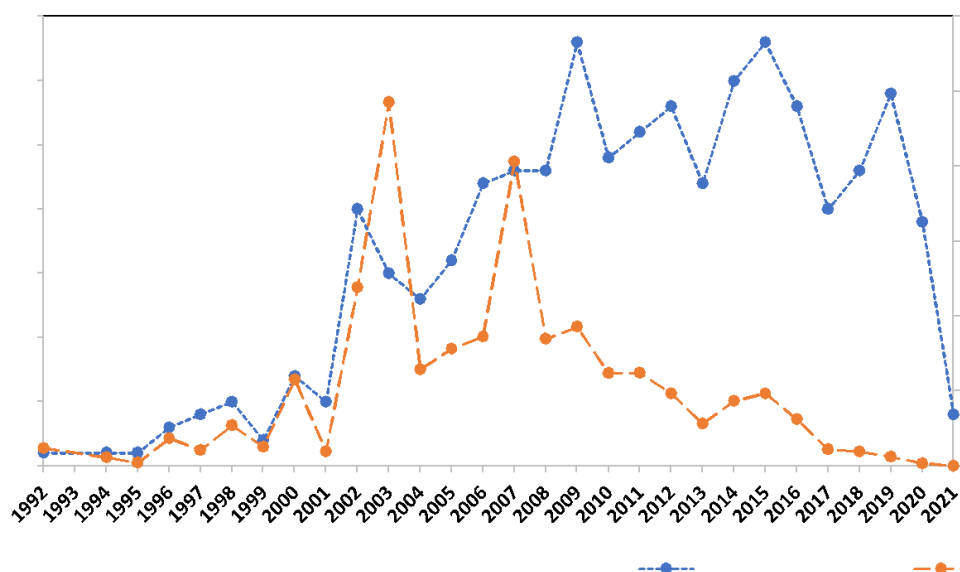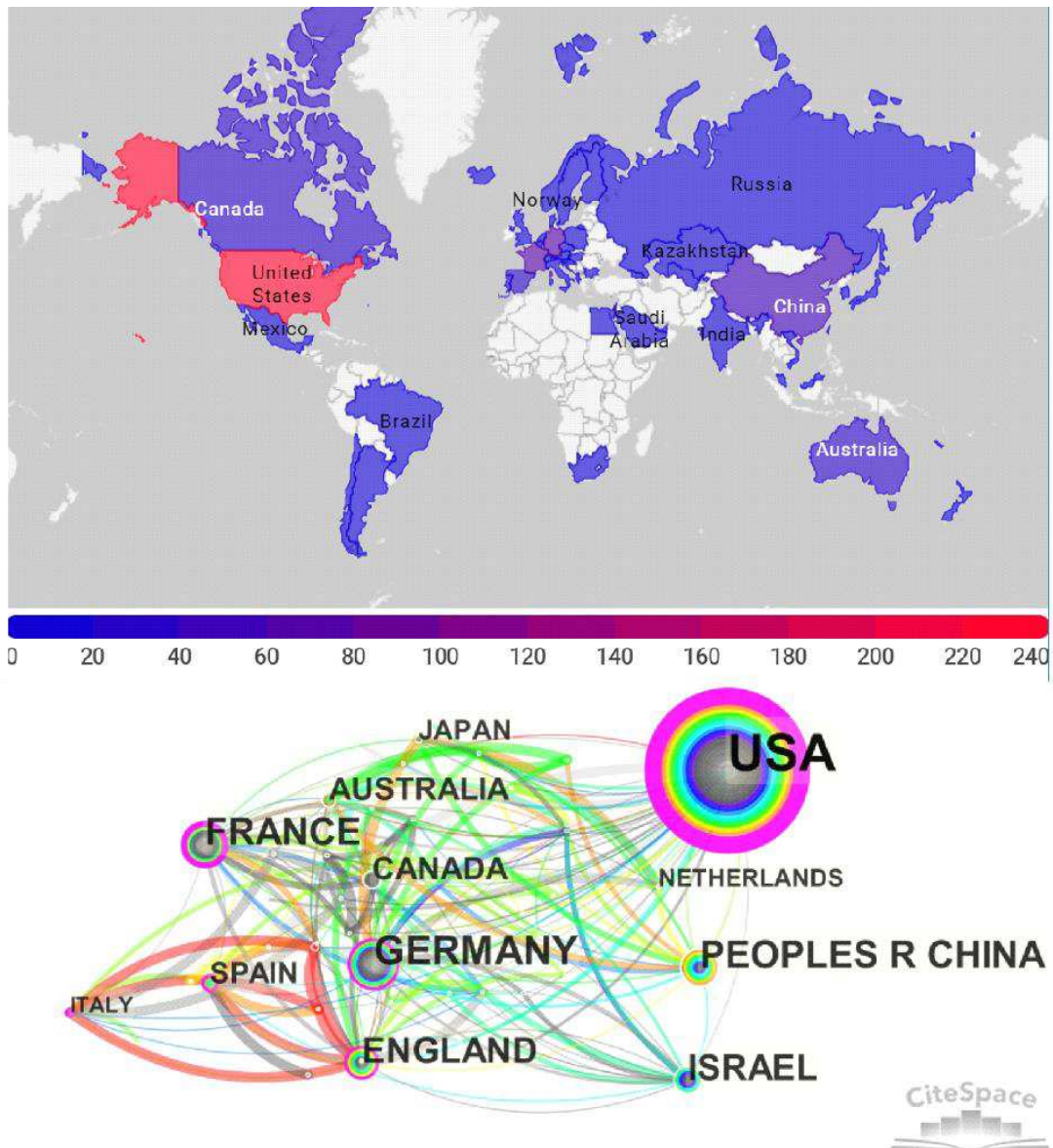

Source: Author

Information analysis based on the origin of the authors of the publications can be observed in Figure 3A. The publication leader was the USA, followed by Germany, France, and China. In Fig. 3B, information on the number of citations and cooperation networks between countries is summarized, showing the USA as the country that stands out in marine cyanobacteria studies.

The letter font size in Figure 3B is proportional to the volume of publication, while the thickness of the connecting lines between countries demonstrates the intensity of cooperation. The thicker the connection line, the higher the frequency of cooperation between two or more countries. The formation of clusters of countries separated by different colors can be seen in Figure 3B. The numbers represent the

citation counts of each country. The top-ranked item by citation counts is in order: USA, France, Germany, and China. Fig. 3B also presents the centrality of some countries (yellow circles). Centrality is an important measure that varies from 0 to 1 (Chen, 2015), representing the influence of a country. In this study, the USA has the most extraordinary centrality (0.85), followed by France (0.2), Germany and Italy (0.15), and England and Spain (0.14).

Moreover, some countries presented citation bursts. A citation burst is an indicator of a most active region (in this case), which can last for multiple years as well as a single year. A citation burst provides evidence that a publication/country is associated with a surge of citations. In other words, the item has attracted an extraordinary degree of attention from its scientific community (Chen, 2015). The top-ranked item in bursts is France (1995), with bursts of 7.13. The 2nd was Canada (2005) (5.15). The 3rd was Germany (1998) (4.19), and the 4th was China (2018) (3.16).

Figure 3. A - Localization of scientific papers published in peer-reviewed journals about genes of *Prochlorococcus* and *Synechococcus*. The geographic distribution of the reviewed papers is related to the color of each country, proportional to the publication number. **B -** The cooperation network of the producing countries. Different colors mean different clusters, and the numbers represent the citation counts.

Source: Author

After analysis in the CiteSpace software, the institutions with more frequency of publications were MIT, CNRS, Univ Warwick, and Univ Paris 06; according to the centrality, the institutions with the highest values were: MIT (0.21), Univ Warwick (0.19), CNRS (0.16), and Univ Calif San Diego (0.13) (Figure 4).

Figure 4. Frequency (number of publications) and centrality of publications per institution.
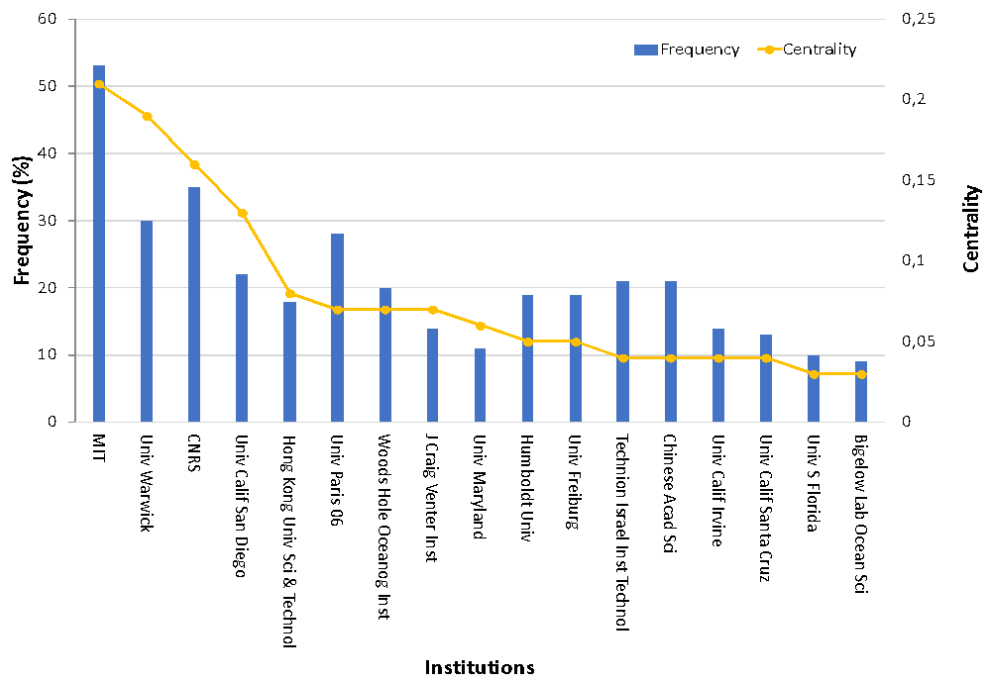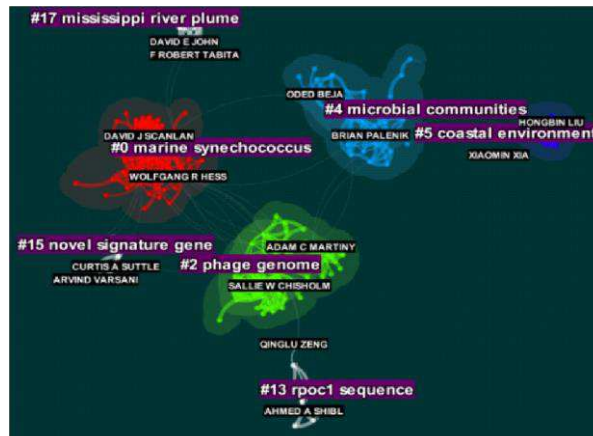

Source: Author

The principal authors that published more articles with higher citations are presented in Figure 5 and the keywords grouped in clusters related to these publications: cluster #0 marine *Synechococcus* is related to the authors D.J. Scanlan and W.R. Hess; the cluster #2 phage genome related to the author A.C. Martiny and S.W. Chisholm and the cluster #4 microbial communities associated with O. Beja and B. Palenik, this analysis had a high Silhouette 0.9846. The silhouette coefficient greater than 0.5 indicates reasonable data clustering (Kaufman and Rousseeuw, 1990). The silhouette coefficients of all five significant clusters exceed 0.9, which shows reasonable clustering results. The authors with the most robust Citation Bursts were in

order: S.W. Chisholm (1992) – 9.70; F. Palenik (1995) – 8.04; W.R. Hess (1995) – 6.12, and B. Palenik (2011) – 3.9.

The leading Journals that published more articles with higher citations are presented in Figure 6 A, and the keywords grouped in clusters related to these publications: the cluster #0 and #2 genome analysis is related to the Journal Environmental Sciences and Ecology; Biotechnology & Applied Microbiology, Genetics & Heredity, and Virology; the cluster #1 genome reduction related to the Journal Biochemistry & Molecular Biology, and the cluster #3 arctic ocean related to Engineering, Environmental Engineering, this analysis had a high Silhouette 0.9452, demonstrating that the representation is significant. The Journals with the most substantial Citation Bursts were in order: Plant Sciences (1995) – 6.99; Biotechnology & Applied Microbiology (1997) – 5.93; Genetics & Heredity (2007) – 4.04, Evolutionary Biology (2009) – 5.42, and Microbiology (2017) – 3.9. Figure 6 B represents the frequency (%) and centrality of publications by Journal. The Journals with the highest frequency of publications was also the same with the highest centrality, and they are in order (frequency – centrality): Microbiology (230 – 0.21), Environmental Sciences & Ecology (88 – 0.19), Science & Technology (70 – 0.16), and Multidisciplinary Sciences (70 – 0.13).

Table 1 expresses the publications with the highest number of citations between 1992-2021 regarding the genetics of *Prochlorococcus* or *Synechococcus,* and some authors of these publications were also the same with the most substantial Citation Burst expressed in figure 5.

Figure 5. CiteSpace representation of the principal authors and the keywords used in the publications.

Source: Author

Based on the analysis of the occurrence of a subject category, it is possible to describe which area is highlighted in the genetics of *Prochlorococcus* or *Synechococcus*. In this study, the most recurring areas for analysis of the characteristics of categories listed by WoS were chosen. Subject category information has been analyzed using CiteSpace.

Figure 7 represents the main keywords used in the publications and the themes in clusters. The keyword "cyanobacteria" were related to the clusters #0 *Prochlorococcus* ecotype, #2 thylakoid membrane, and #4 genome analysis, the keyword "gene" was related to #8 molecular component and #9 rRNA gene sequence analysis, the word "*Synechococcus*" was related to #1 lake Kinneret and #3 *Prochlorococcus* strain, and the word "phytoplankton" related to #5 single psba gene, #1 lake Kinneret and #6 ribulose bisphosphate carboxylase gene expression.

Keywords can be regarded as the soul of an article, and a keyword co-word analysis can be used to detect research topics, analyze research hotspots, and monitor the research frontier transitions of a specific knowledge domain (Yu *et al*., 2017;

Ouyang *et al*., 2018). CiteSpace conducted the keyword co-word analysis related to *Prochlorococcus* and *Synechococcus*. As shown in Figure 8, the evolution of the keywords in the publications in the periods of A – 1992-2000; B – 2001-2010, and C – 2011-2021. The migration characteristics of flow cytometry, genes identification methods, genomes, and clades of cyanobacteria indicate how future studies should focus. These results showed that the relevant studies of *Prochlorococcus* and *Synechococcus* were constantly expanded and developed. Table 2 shows the top 15 keywords in terms of frequency and centrality in the field of *Prochlorococcus* and *Synechococcus* over the past 30 years. The following is a typical analysis based on both tables and figures. The selection of the discussed keywords was generally based on frequency and centrality.

Figure 6. **A -** CiteSpace representation of the leading Journals and the keywords used in the publications and Strongest Citation Burst. **B –** Frequency (number of publications) and centrality of publications by Journal.
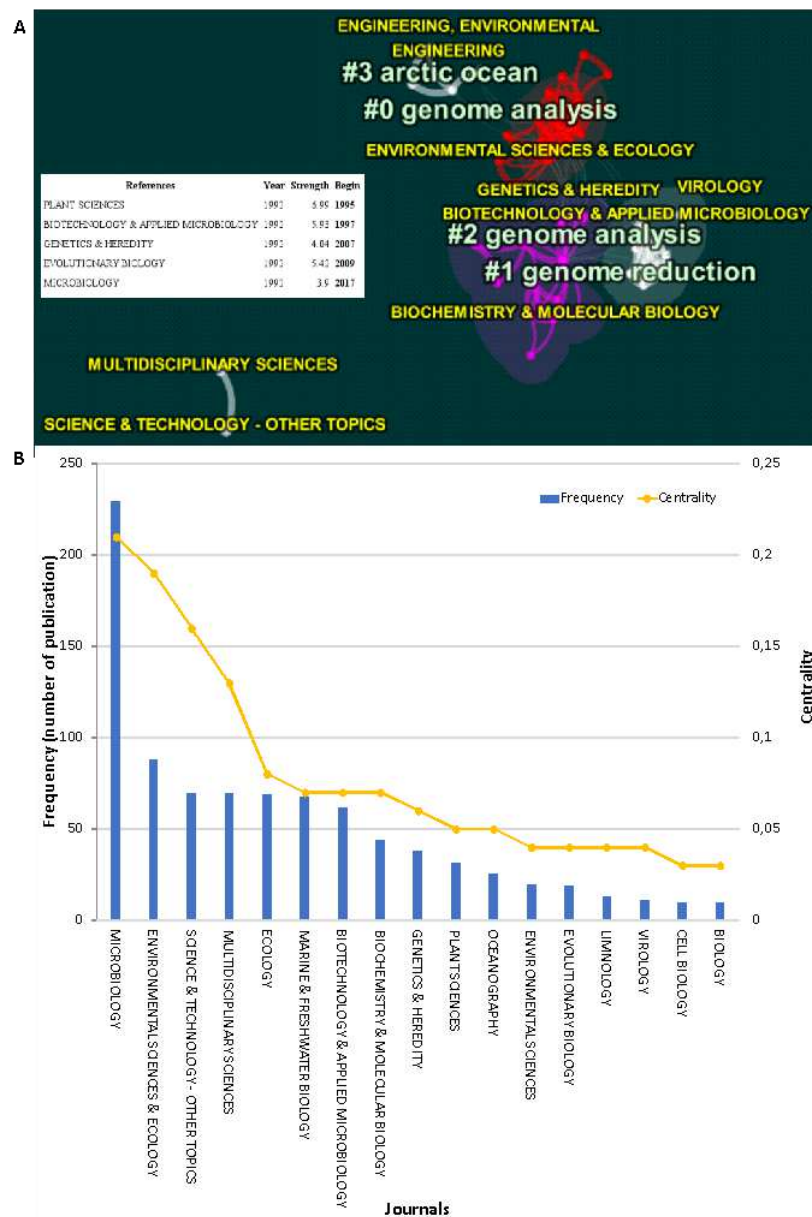
Source: Author

Comparing with period A (1992–2021), the most frequent and with higher centrality keyword was "multiple evolutionary origin" (centrality - 0.43). The phylogenetic analyses of cyanobacteria have gained in quantity over the past 20 years (Tomitani *et al*., 2006; Giovannoni *et al*., 1988; Blank *et al*., 2010). Analyses assessing characteristics of cyanobacterial ancestors provide essential information on the history of cyanobacteria and the evolution of life forms (Swingley *et al*., 2008). The second word with higher centrality is "flow cytometry" (0.40) because, in this period was the primary method used to identify *Prochlorococcus* and *Synechococcus* when *Synechococcus* cells are excited by the wavelength of blue light (and to a lesser degree UV), phycobilin has emitted a strong orange fluorescence that can be registered separately from the red fluorescence of their chlorophyll. As *Synechococcus* are significantly larger than *Prochlorococcus* (diameters: about 1 and 0.6 µm, respectively), these two cell types can be discriminated by their scatter signals in flow cytometry related to their size (Marie *et al*., 1997).

The period B (2001-2010), the keywords with higher centrality were "sequence" and "gene," both with 0.13 of centrality; this can be explained by the period when the gene sequencing was developed, and the scientific community discovered the genome of different species. The DNA analysis and the genomic development make the identification of cyanobacteria genes for phylogenetic analyses not without difficult and requires, an ideal case comparison of complete genome data. Genomic studies in cyanobacteria have emphasized that marine species of *Synechococcus* and *Prochlorococcus* are a particularly well-studied group (Hess *et al*., 2001; Rocap *et al*., 2003; Dufresne *et al*., 2008; Scanlan *et al*., 2009), as shown in Table 2, and according

to the Table 1 was the period with the publications more cited, due to the gene sequencing (Suzuki *et al.*, 2000; Moore *et al.*, 2002; Rocap *et al.*, 2002 and 2003; Badger and Price, 2003; Dufresne *et al.*, 2003; Palenik *et al.*, 2003; Sullivan *et al.*, 2003; Lindell *et al.*, 2004; Kettler *et al.*, 2007).

In the last ten years period C (2011-2021), the keywords with higher centrality were diversity - 0.17; marine *Synechococcus* - 0.10; *Prochlorococcus* ecotype - 0.08; gene expression - 0.14; phytoplankton - 0.09; and light - 0.07. The concern with the diversity and phytoplankton ecology and the discovery of ecotypes in *Prochlorococcus* were the most important activities regarding these cyanobacteria. The main distinction between ecotypes in *Prochlorococcus* refers to adaptations to low and high light (low light LL and high light - HL) evidenced by flow cytometry in the red fluorescence signal (Chl) (Campbell & Vaulot, 1993), confirmed by a physical examination, cell screening and growth of each subpopulation (Moore *et al.*, 1998) and supported at the genetic level, grouping HL into 16s rRNA or 16s 23s rDNA as opposed to the variable arrangements of LL lineages (West & Scanlan, 1999). Physiological differentiation into two HL clades and 4 LL clades allow them to occupy a wide range of luminosity spectrum and nutrient availability at low and mid-latitudes (Partensky *et al.*, 1999; Kettler *et al.*, 2007). The biogeography of *Synechococcus* genetic diversity is less clear. The current notion is that several ecotypes (I, IV, and perhaps several others) are abundant in cooler, nutrient-rich waters. Likewise, several ecotypes (classes II, III, V, VI, and VII) are frequent but variable in tropical and subtropical waters (Kent *et al.*, 2019).

Table 1. Publications regarding the genetics of *Prochlorococcus or Synechococcus* with the highest number of citations between 1992-2021.

|	Title	Author and year	Journal	Citation
1	Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*	Kettler *et al*., 2007	Plos Genetics	2737
2	Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation.	Rocap *et al*., 2003	Nature	1655
3	The genome of a motile marine *Synechococcus*	Palenik *et al*., 2003	Nature	935
4	Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5 '-nuclease assays	Suzuki *et al*., 2000	Appl. Environ. Microbiol.	823
5	Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome	Dufresne *et al*., 2003	PNAS	764
6	CO_2 concentrating mechanisms in cyanobacteria: molecular components, their diversity, and evolution	Badger and Price, 2003	J. Exp. Bot.	513
7	Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences	Rocap *et al*., 2002	Appl. Environ. Microbiol.	400
8	Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*	Moore *et al*., 2002	Limnol. Oceanogr.	383
9	Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*	Sullivan *et al*., 2003	Nature	379
10	Transfer of photosynthesis genes to and from *Prochlorococcus* viruses	Lindell *et al*., 2004	PNAS	361

Source: Author

Table 2. Top 15 keywords related to *Prochlorococcus and Synechococcus* for different periods of 1992 to 2021.

| A 1992 -2000			B 2001-2010			C 2011-2021		
Keyword	Frequency	Centrality	Keyword	Frequency	Centrality	Keyword	Frequency	Centrality
multiple evolutionary origin	9	0.43	sequence	39	0.13	diversity	61	0.17
prochlorophyte	5	0.16	diversity	31	0.11	photosynthesis gene	44	0.05
photosynthesis	4	0.04	genome	24	0.05	sequence	39	0.05
green chloroplast	4	0.33	evolution	23	0.05	marine Synechococcus	38	0.10
flow cytometry	3	0.40	photosynthesis gene	20	0.02	virus	35	0.06
rubisco	3	0.00	gene	19	0.13	cyanophage	26	0.07
phytoplankton	3	0.12	virus	18	0.04	evolution	25	0.06
gene	3	0.12	Prochlorococcus ecotype	18	0.07	Prochlorococcus ecotype	24	0.08
community structure	3	0.11	light	17	0.04	gene expression	22	0.14
Mediterranean sea	3	0.03	community structure	16	0.12	gene	22	0.08
bacterioplankton	3	0.00	marine Synechococcus	15	0.07	phytoplankton	21	0.09
mRNA	2	0.05	population	14	0.03	picocyanobacteria	20	0.02
			phytoplankton	13	0.07	genome	19	0.04
			phycoerythrin	12	0.04	light	18	0.07
			16s ribosomal RNA	11	0.05	ecotype	18	0.03

Source: Author

Figure 7. CiteSpace representation of the main keywords used in the publications and the themes in clusters.
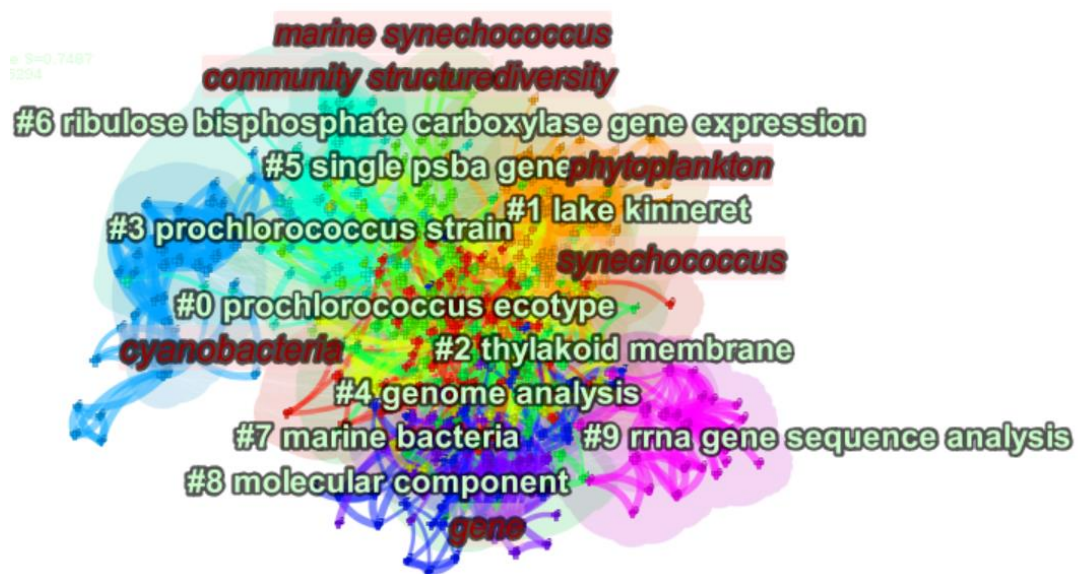
Source: Author

Figure 8. Evolution of the keywords in the publications in the period A − 1992-2000; B − 2001-2010; and C − 2011-2021.
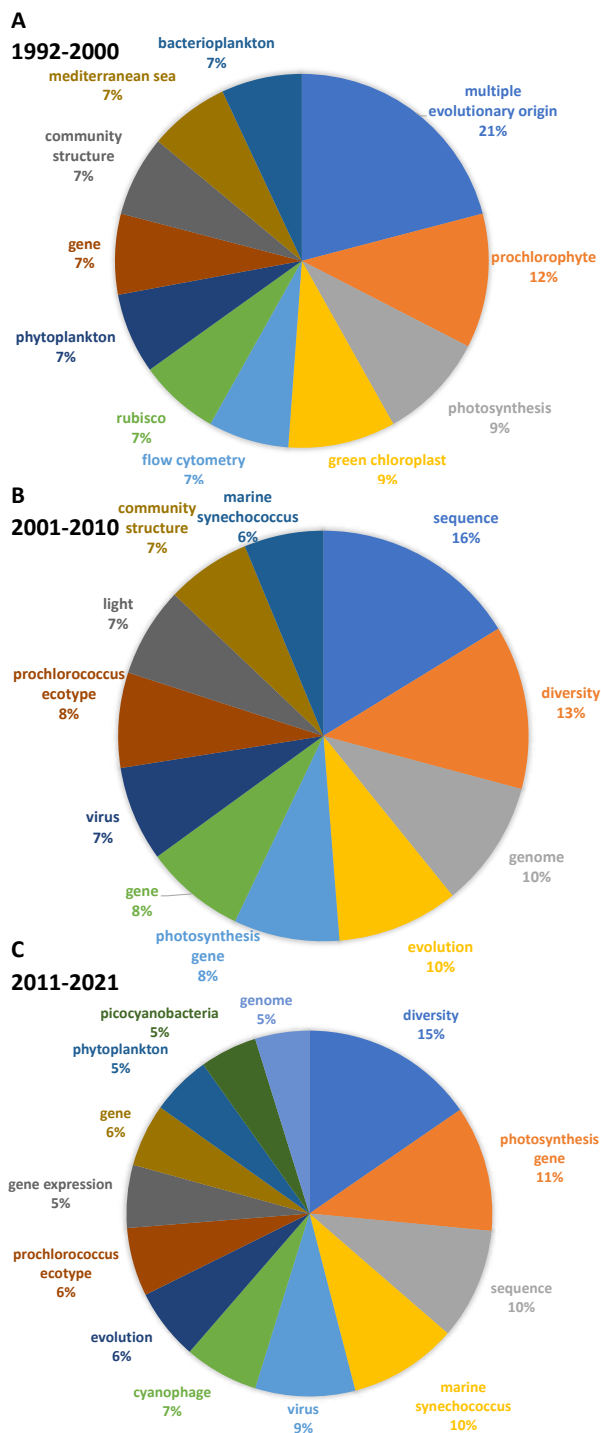
Source: Author

4.   CONCLUSION

Over the past 29 years, there were 480 journal articles related to genetics of cyanobacteria, and the number of articles increased steadily mainly concentrated in North America, Europe, and Asia. In addition, we examined the genetic of *Prochlorococcus* and *Synechococcus* characteristics from multiple perspectives. Meanwhile, international collaborations among North America, Europe, and Asia were relatively intimate. The USA was the most active contributor to most cooperation and publications among all the countries. In addition, the migration of keywords from "multiple evolutionary origin" and "flow cytometry" to "sequence" and "gene" to "diversity"; "marine *Synechococcus"* and "*Prochlorococcus* ecotype" indicate how future studies should focus, showing that the concern with phytoplankton diversity, ecotypes of cyanobacteria are the new way followed by the scientists regarding these cyanobacteria responsible for a considerable amount of atmospheric oxygen.

REFERENCES

Badger, M.R., and Price, G.D., (2003). CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity, and evolution. *Journal of experimental botany, 54*383, pp.609-622.

Berube, P. M., Biller, S. J., Hackl, T., Hogle, S. L., Satinsky, B. M., Becker, J. W., & Chisholm, S. W. 2018. Single-cell genomes of *Prochlorococcus, Synechococcus*, and sympatric microbes from diverse marine environments. *Scientific data*, 51, 1-11.

Campbell, L & Vaulot, D 1993 Photosynthetic picoplankton community structure in the subtropical North Pacific Ocean near Hawaii Station Aloha. *Deep-Sea Research Part I: Oceanographic Research Papers*, V 40, N 10, P 2043-2060

Campbell, L, Nolla, HA, Vaulot, D 1994 The importance of Prochlorococcus to community structure in the central North Pacific Ocean. *Limnology And Oceanography*, V 39, N 4, P 954-961

Cesar-Ribeiro C, Piedras FR, da Cunha LC, de Lima DT, Pinho LQ and Moser GAO 2020 Is Oligotrophy an Equalizing Factor Driving Microplankton Species Functional Diversity Within Agulhas Rings? *Front. Mar. Sci.* 7:599185. DOI: 10.3389/fmars.2020.599185

Chen, C., 2014. The citespace manual. College of Computing and Informatics, 1, pp.1-84.

Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I.M., Barbe, V., Duprat, S., Galperin, M.Y., Koonin, E.V., Le Gall, F. and Makarova, K.S., (2003). Genome sequence of the cyanobacterium Prochlorococcus marinus SS120, a nearly

minimal oxyphototrophic genome. *Proceedings of the National Academy of Sciences*, *100*17, pp.10020-10025.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., Karl, D.M., Li, W.K., Lomas, M.W., Veneziano, D. and Vera, C.S., 2013. Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proceedings of the National Academy of Sciences*, 11024, pp.9824-9829.

Giovannoni, SJ, Delong, EF, Schmidt, TM, Pace, NR 1990 Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. *Applied And Environmental Microbiology,* V 56, N 8, P 2572-2575

Hartmann, M, Gomez-Pereira, P, Grob, C, Ostrowski, M, Scanlan, DJ, Zubkov, MV 2014 Efficient CO_2 fixation by surface Prochlorococcus in the Atlantic Ocean. *The Isme Journal*, V 8, N 11, P 2280-2289

Hopkinson, BM, Young, JN, Tansik, AL, Binder, BJ 2014 The minimal CO_2-concentrating mechanism of *Prochlorococcus* spp Med4 Is effective and efficient. *Plant Physiology*, V 166, N 4, P 2205-2217

Kent, AG, Baer, SE, Mouginot, C, Huang, JS, Larkin, AA, Lomas, MW, Martiny, AC 2018 Parallel phylogeography of *Prochlorococcus* and *Synechococcus. The Isme Journal*, V 13, N 2, P 430-441

Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J. and Steglich, C., 2007. Patterns and implications of gene gain and loss in the evolution of Prochlorococcus. *PLoS genetics*, 312, p.e231.

Lalli, CM, Parsons, TR 1997 Biological Oceanography: an introduction Oxford: Butterworth-Heinemann, p. 320

Li, WKW, Lewis, MR, Lister, A 1992 Low cytometric detection of prochlorophytes and cyanobacteria in the Gulf of Policastro Archiv Für. *Hydrobiologie*, SL, V 42, P 1-8

Li, WKW, Subba, RAO, DV, Harrison, WG, Smith, JC, Cullen, LL, Irwin, B, Platt, T 1993 Autotrophic picoplankton in the tropical ocean. *Science,* 219,292-295

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W., 2004. Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proceedings of the National Academy of Sciences*, 10130, pp.11013-11018.

Liu, H, Nolla, HA, Campbell, L 1997 Prochlorococcus growth rate and contribution to primary production in the equatorial and subtropical north pacific ocean. *Aquatic Microbial Ecology*, V 12, P 39-47

Marie, D, Partensky, F, Jacquet, S, Vaulot, D 1997 Enumeration, and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain Sybr Green I applied and environmental. *Microbiology*, V 63, N 1, P 186-193

Moore, LR, Post, AF, Rocap, G, Chisholm, SW 2002 Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and Synechococcus. *Limnology and Oceanography*, V 47, N 4, P 989-996

Moore, LR, Rocap, G, Chisholm, SW 1998 Physiology and molecular phylogeny of coexisting *Prochlorococcus* Ecotypes, *Nature*, V 393, N 6684, P 464-467 1998

Olson, RJ, Chisholm, SW, Zettler, ER, Altabet, MA, Dusenberry, JA 1990 Spatial and temporal distributions of prochlorophyte picoplankton in the north Atlantic Ocean. *Deep-Sea Research Part A Oceanographic Research Papers*, V 37, N 6, P 1033-1051

Palenik, B., Brahamsha, B., Larimer, F.W., Land, M., Hauser, L., Chain, P., Lamerdin, J., Regala, W., Allen, E.E., McCarren, J. and Paulsen, I., (2003). The genome of a motile marine Synechococcus. *Nature*, *424*6952, pp.1037-1042.

Partensky, F, Blanchot, J, Vaulot, D 1999 Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* In Oceanic Waters: A Review In Loïc, C, Larkum, A W D **Marine Cyanobacteria**: Bulletin De L'Institut Océanographique De Monaco Paris: Colloque Ird/Université De Sydney, P 457-475

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, Arellano A, Coleman M, Hauser L, Hess WR, Johnson ZI, Land M, Lindell D, Post AF, Regala W, Shah M, Shaw SL, Steglich C, Sullivan MB, Ting CS, Tolonen A, Webb EA, Zinser ER, Chisholm SW, (2003): Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature*. 424 6952: 1042-1047. 10.1038/nature01947.

Rocap, G., Distel, D.L., Waterbury, J.B. and Chisholm, S.W., 2002. Resolution of Prochlorococcus and Synechococcus ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Applied and Environmental Microbiology*, 683, pp.1180-1191.

Sullivan, M.B., Waterbury, J.B. and Chisholm, S.W., (2003). Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. *Nature*, 4246952, pp.1047-1051.

Suzuki, M.T., Taylor, L.T. and DeLong, E.F., 2000. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5′-nuclease assays. *Applied and environmental microbiology*, *66*11, pp.4605-4614.

Swingley WD, Blankenship RE, Raymond J, 2008. Integrating Markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families. *Molecular Biology And Evolution*. 2008, 25 4: 643-654. 10.1093/molbev/msn034.

Veldhuis, MJW, Kraay, GW 1993 Cell abundance and fluorescence of picoplankton in relation to growth irradiance and nitrogen availability in the Red Sea Netherlands *Journal of Sea Research*, V 31, N 2, P 135-145

Veldhuis, MJW, Kraay, GW A 2000 Application Of Flow Cytometry In Marine Phytoplankton Research: Current Applications and Future Perspectives. *Scientia Marina*, V 64, N 2, P 121-134

West, NJ, Scanlan, D 1999 Niche-Partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean. *Applied And Environmental Microbiology*, V 65, N 6, P 2585-2591.