



UNIVERSIDADE FEDERAL DO CEARÁ  
INSTITUTO DE CIÊNCIAS DO MAR - LABOMAR  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MARINHAS TROPICAIS

**RENAN VANDRE DA SILVA TOSCANO SAES**

**DIVERSIDADE E ESTOQUE DE CARBONO EM CIANOBACTÉRIAS NA  
COSTA E NOS ARQUIPÉLAGOS DA PORÇÃO OESTE DO GIRO  
EQUATORIAL ATLÂNTICO**

FORTALEZA

2018

**RENAN VANDRE DA SILVA TOSCANO SAES**

**DIVERSIDADE E ESTOQUE DE CARBONO EM CIANOBACTÉRIAS NA  
COSTA E NOS ARQUIPÉLAGOS DA PORÇÃO OESTE DO GIRO  
EQUATORIAL ATLÂNTICO**

Tese de doutorado submetida ao Programa de Pós-graduação em Ciências Marinhas Tropicais da Universidade Federal do Ceará, como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências Marinhas Tropicais, área de concentração da Utilização e Manejo de Ecossistemas Marinhos e Estuarinos.

Orientador: Prof. Dr. Denis Moledo de Souza Abessa

FORTALEZA

2018

Sempre dedico qualquer trabalho acadêmico que eu tenha feito a quem encontrar utilidade para seu maior objetivo, que no atual caso, é a valorização, respeito e deslumbramento diante de paisagens divinamente esculpidas e particularmente importantes para a manutenção da vida local. Principalmente na nosso, a humana.

## AGRADECIMENTOS

Ao algo maior que fez o Sol, Jáh, o ar, a gravidade, as águas, as camadas terrestres, o cosmos, os animais cientistas que tentam entender isso tudo e todo o resto, que se representa na única coisa palpável dele(a): Deus! Este possibilita espíritos de plena luz para nos orientar na evolução de nossa luz interna natural, até voltarmos a vibração que superará a matéria algum dia. Mas por hora agradeço aos inesquecíveis momentos da minha materialização.

Também agradeço a tudo isso, por materializarem minha família de sangue e alma, os meus pais, irmã, vovós, vovôs, títias e primos - vivos ou mortos - por me proporcionarem a oportunidade de aprender algo com eles todos os dias (mesmo muito longe). Sendo os meus pais minhas referências de norte.

Ao mar, que modulado pelos ventos, é a maior representação da liberdade aos meus olhos, sempre me purificando e me enchendo de vida e sabedoria, pois mesmo caótico é perfeito.

Aos ventos e árvores que refrescam a minha vida, principalmente os Alísios e as que fazem grandes sombras, respectivamente.

A Leticia Costa-Lotufo que me adotou entre os seus e acreditou que eu poderia me expressar melhor em um projeto oceanográfico, obrigado pela confiança seguida de financiamento. Nos momentos mais solitários ou confusos vinculados a tese e a muito mais, fui devidamente amparado e acalmado. Obrigado!

Ao meu eterno orientador Denis Abessa, por ser um salvador e um norte sempre! No meu primeiro dia no lab do prof. Denis, ele me pediu, a simples tarefa de ajustar a salinidade da água do mar ao normal. Eu muito eficiente fiz muito mais, eu ajustei a uma salinidade de 80. Desde desse dia ele virou pra mim e disse que eu seria um grande cientista marinho! Obrigado professor, por sempre estar lá, aqui, acola, em qualquer lugar disposto a me ajudar e se preocupando comigo! Mais uma vez, valeu Denis pela força na finaleira punk-hardcore!

A Universidade Federal do Ceará (UFC) e aos Instituto de Ciências do Mar (Labomar/UFC) e Núcleo de Pesquisa e Desenvolvimento de Medicamentos (NPDM/UFC) que proveram o suporte logístico ao projeto, assim como seus professores e funcionários, que vem contribuindo a minha formação acadêmica. Também aos laboratórios que me auxiliaram na execução pratica deste trabalho, em específico: Laboratório de Bioprospecção e Biotecnologia Marinha (LaBBMar/UFC), responsável Prof. Dr. Diego Wilke; ao Laboratório de Ecologia Microbiana e Biotecnologia (LEMBiotech/UFC), responsável Profa. Dra. Vânia Melo; e ao Laboratorio de Cultura (LabCult/UERJ), responsável Profa. Geyci Moser.

Aos projetos de Prospecção Sustentável em Ilhas Oceânicas - Biodiversidade, Química, Ecologia e Biotecnologia (PROSPECMAR, P.I.: Dra. Leticia Veras Costa Lotufo) e ao Programa de Monitoramento de Longa Duração das Comunidades Recifais de Ilhas Oceânicas (PELD-ILOC, P.I.: Dr. Carlos Eduardo Leite Ferreira); que juntos sustentam a Rede Nacional de Pesquisa em Biodiversidade Marinha (SISBIOTA Mar). Estes financiaram toda a pesquisa e as expedições oceanográficas as ilhas e praias amostradas. Um agradecimento especial a Secretaria Interministerial para os Recursos do Mar (SECIRM), a Marinha do Brasil - Base Naval de Natal/RN e a Universidade Federal do Rio Grande do Norte (UFRN), pelo suporte logístico nas expedições.

A Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) por conceder uma bolsa acadêmica durante todo período do doutorado.

Aos parceiros que ajudaram com a “mão na massa” nos longos dias de coleta e/ou nas mais longas horas de bancadas, nos mais variados grupos: Bianca, Renato, Moisés e Lucas (SISBIOTA); Fernanda e Domênica (LabCult); Samuel, Geórgia e Frida, Thalita, Denise, Wal e Jonathan (LEMBiotech); Andréa, Elthon, Larissa e Thaís (NPDM).

## RESUMO

As corrente superficial Equatorial Sul - SEC e a corrente superficial Norte do Brasil - NBC estão diretamente conectadas ao giro Equatorial Atlântico. Estas correntes influenciam a movimentação e aporte de água em superfície nos arquipélagos oceânicos brasileiros e borda da plataforma continental nordeste do Brasil. O estoque de carbono de cianobactérias acompanha um gradiente longitudinal costa → arquipélagos oceânicos: PP > PA = IC > AC > FN > AS. O estoque de carbono na costa corresponde a  $1,04 \text{ mg C m}^{-3}$ , com o estoque relativo de 2,69% *Synechococcus* e 0,48% do POC em *Prochlorococcus*. Nos arquipélagos oceânicos, apenas estão estocados nas cianobactérias, com o estoque relativo de 0,38% em *Synechococcus* e 0,10% do POC em *Prochlorococcus* em águas superficiais. Embora as águas equatoriais sejam consideradas pobres no estoque de carbono, o carbono estocado na superfície se mantém ao longo do ano, sustentando a produtividade primária. O Arquipélago de São Pedro e São Paulo - AS apresenta maiores oscilações de POC devido a sua posição geográfica sendo diretamente influenciado pela sazonalidade da SEC. Ao todo, 1.158.901 sequências foram representadas em 56 táxons, com 48 pertencem às cianobactérias. Deste total, 37 táxons foram comuns entre as frações nano e picofitoplânctônicas. Isolando as frações do fitoplâncton pelo gradiente longitudinal, costa e ilhas, são compostos 4 cenários distintos, nos quais apenas 20 táxons são comuns entre si. O cenário de maior exclusividade taxonômica foi representado nos arquipélagos na fração nanofitoplânctônica, com 9 táxons exclusivos. Os gêneros *Prochlorococcus* e *Synechococcus* foram os dominantes, acumulando 98% das sequências totais em ambas frações fitoplânctônicas. As réplicas costeiras se correlacionam mais com *Synechococcus*, enquanto que as réplicas insulares foram dominadas por *Prochlorococcus*. Não foi possível estabelecer um gradiente costa → ilhas de diversidade de cianobactérias através de sequências 16S rRNA, pois os arquipélagos amostrados são muito distintos entre si. As localidades de FN, apresentaram os maiores índices de diversidade e muitos táxons raros, mesmo com abundância média de sequências, já as localidades do AS apresentaram padrões totalmente contrários. As forças ambientais que mais dominaram a distribuição de cianobactérias no gradiente superficial costa-ilhas foi a temperatura. As variáveis hidrodinâmica e tipo de substrato revelaram maior diversidade picofitoplânctônica nas planícies arenosas em condições expostas. Para a fração nanofitoplânctônica estas condições representaram a menor biodiversidade.

Palavras-chave: carbono; biomassa; metagenômica 16S rRNA; *Prochlorococcus*; *Synechococcus*

## ABSTRACT

The South Equatorial - SEC and North Brazillian - NBC surface currents are directly connected to the Atlantic Equatorial Gyre. These currents influence water movement and contribution on the surface water in the brazilian oceanic archipelagos and the border of Brazil's northeast continental platform. The carbon stock of the cyanobacterias follow a coast→oceanic archipelagos longitudinal gradient: PP>PA=IC>AC>FN>AS. The carbon stock along coast corresponds to 1,04mg C m<sup>-3</sup>, with relative stock of 2,69% *Synechococcus* and 0,48% of the POC in *Prochlorococcus*. On the oceanic archipelagos, only 0,15mg C m<sup>-3</sup> are stocked in the cyanobacterias, with a relative stock of 0,38% in *Synechococcus* and 0,10% of the POC in *Prochlorococcus* on surface waters. Although equatorial waters are considered poor when it comes to carbon stock, the carbon stocked on the surface keeps its levels throughout the year, sustaining the primary productivity. The Saint Peter and Saint Paul Archipelago - AS shows bigger POC oscillations due to its geographic position, being directly influenced by the SEC's seasonsality. Altogether, 1.158.901 sequences have been represented in 56 taxons, within 48 belonging to cyanobacterias. Out of amount of taxons, 37 have been common among nano and picophytoplanktonics fractions. Isolating the phytoplankton fractions by the longitudinal gradient, coasts and islands, 4 distinct cenarios are built, in which only 20 taxons are common among each other. The cenario with taxonomic exclusivity has been represented on the archipelagos, on the nanophytoplanktonic, with 9 exclusive taxons. The genres *Prochlorococcus* and *Synechococcus* have been dominants, accumulating 98% of the total of sequences in both phytoplanktonic fractions. The coastal replicates correlate more with *Synechococcus*, while insular replicates have been dominated by *Prochlorococcus*. It hasn't been possible to establish a cyanobacterias' diversity coast→island gradient through 16S rRNA sequences, because the sampled archipelagos are very distinct among each other. FN locations showed bigger diversity levels and several rare taxons, even with average sequence abundance, as far as AS locations showed fully opposite patterns. The environmental parameters that have mostly dominated the cyanobacterias distribution on the coast-island surface gradient was the temperature. The hydrodynamic variables and kind of substrate revealed bigger picophytoplanktonic diversity on the surfaces on exposed conditions. For the nanophytoplanktonic fraction, these conditions represented the smallest biodiversity.

Keywords: carbon; biomass; 16S rRNA metagenomic; *Prochlorococcus*; *Synechococcus*

## LISTA DE FIGURAS

Figura 1. Esquema dos principais representantes do grupo fitoplâncton (MOORE et al., 1995; ANDERSEN, 2005; REYNOLDS, 2006; HALLENBECK, 2017). Fonte: o próprio autor. .... 02

Figura 2. Classificação do fitoplâncton de acordo com a dimensão linear máxima (diâmetro) seguindo a nomenclatura proposta por Sieburth et al. (1978). Fonte: adaptado de Reynolds (1996). ..... 06

Figura 3: Esquema do processamento genômico da amostra desde a extração até as técnicas de purificação e isolamento dos genes que possibilitam o estudo da composição e filogenética microbiana. Análises que formam bibliotecas genômicas são mais específicas quanto a sequência exata de nucleotídeos nos fragmentos de DNA, podendo chegar a análises mais profundas como a diversidade filogenética e funcional. Análises de fingerprint são adequadas para se identificar a estrutura da comunidade microbiota, com um perfil de bandas específico a cada microbioma. Fonte: o próprio autor. .... 17

Figura 4. As estruturas de uma cianobactéria e de um cloroplasto são muito próximas. De acordo com a teoria da endossimbiose, os cloroplastos atualmente encontrados em células eucarióticas fotossintetizantes são provenientes de simbioses entre células fagocitárias nucleadas com cianobactérias (MARGULIS, 1981; PACE, 1997). Os tilacóides possuem as mesmas funções em ambos os grupos, responsáveis por captar a luz solar e transformá-la em energia química (ATP) através da hidrólise da água, sendo liberado oxigênio molecular ao final da reação (acceptor de elétrons) estabilizando o ATP e finalizando a fase clara da fotossíntese. A fase escura ou química é realizada no estroma (BARSANTI & GUALTIERI, 2006; KIRCHMAN, 2011). Fonte: adaptado de Ducat et al. (2011). ..... 23

Figura 5. As clorofilas-a e -b possuem dois picos de absorção no espectro do visível, um na região do azul-violeta (430nm) e outro no vermelho (665nm), por justamente absorver os comprimentos de onda azul e vermelho, o cloroplasto tende a ser representado pela cor esverdeada. No complexo antena, ocorrem as primeiras reações da fotossíntese, com a absorção do fóton luminoso pelos pigmentos mais externos do complexo a assim transferido é transferido aos outros pigmentos mais centrais por ressonância (se comportam como uma antena de captação de fótons), excitando-os até o centro de reação, o qual contém uma clorofila-a central. No centro de reação a energia é usada para transferir um elétron energético para um acceptor de elétrons. Cerca de 2500 moléculas de clorofila são necessárias para produzir uma molécula de oxigênio, e um mínimo de oito fótons de luz deve ser absorvido no processo. Um modelo representa cerca de 300 moléculas de clorofila e 40 ou mais de carotenóides e outros pigmentos acessórios (FALKOWSKI et al., 1981; MOORE et al., 1995; HUOT & BABIN, 2010; KIRCHMAN, 2011). Fonte: adaptado de Moore et al. (1995) e Lalli & Parsons (1997). ..... 24

Figura 6. O fotossistema I é mais complexo do que o fotossistema II, porque tem um sistema de antena mais complexo, mais subunidades na sua estrutura geral e pode exibir transferência de elétrons cíclicos ou não cíclicos. O esquema acima representa a via acíclica do transporte de elétrons gerados a partir hidrolise da água, desta forma o PS-II oxida água e o PS-I reduz o  $\text{NADP}^+$  à  $\text{NADPH}$ , constituindo os elementos necessários para a dar continuidade a fase escura (ciclo de Calvin) e gerar glicose. Em alguns momentos, quando o cloroplasto esta em desbalanço entre as quantidades de  $\text{NADPH}$  e  $\text{ATP}$ , ocorre a via cíclica apenas com a participação do PS-I. Nesse caso, ocorre uma inversão no transporte de elétrons, apos excitado pela energia solar o PS-I redireciona seus elétrons para o complexo citocromo  $b_6-f$  promovendo a entrada de prontos no lúmen do tilacóide e gerando  $\text{ATP}$  pela  $\text{ATP-ase}$ . Portanto, na fase cíclica ocorre a produção de  $\text{ATP}$ , mas não de  $\text{NADPH}$  e nem de  $\text{O}_2$  (o PS-II não participa) inviabilizando a produção de glicose no ciclo de Calvin. O centro de reação  $\text{P680}$  oxidado ( $\text{P680}^+$ ) do PS-II é o agente oxidante biológico mais forte conhecido na natureza, isto torna possível oxidar a água (molécula altamente estável):  $\text{H}_2\text{O} \rightarrow 2\text{H}^+ + \frac{1}{2}\text{O}_2 + 2e^-$ . Em contraste, no

PS-I o P700 quando excitado (P700<sup>+</sup>) se torna o agente redutor biológico mais forte (NITSCKE & RUTHERFORD, 1991; MOORE et al., 1995). Fonte: adaptado de Moore et al. (1995). .... 26

Figura 7. O esquema apresentado é uma simplificação do ciclo de Calvin. Considerando que o ciclo necessita ser mantido por moléculas produzidos por ele mesmo e ainda libere glicose para o uso celular, é necessário admitir a reação de seis moléculas de CO<sub>2</sub>, portanto, sendo consumidas 12 moléculas de ATP e 12 de NADPH, produzindo 12 moléculas de gliceraldeído-3-fosfato (G3P). Destas 12 moléculas, 2 vão formar uma glicose e as restantes 10 voltam para regenerar 6 moléculas de RuBP com o gasto adicional de mais 12 moléculas de ATP. O RuBisCo (ribulose-1,5-difosfato carboxilase/oxigenase) é a enzima mais importante e mais abundante na natureza, pois ela é a responsável da produção de toda a biomassa no planeta a partir de CO<sub>2</sub>. Calcula-se que existem 40 milhões de toneladas da enzima RuBisCo na biosfera (NITSCKE & RUTHERFORD, 1991; MOORE et al., 1995). Fonte: adaptado de Moore et al. (1995). ..... 28

Figura 8. Os pontos deste trabalho estão situados na costa e nos montes submarinos na zona equatorial do Atlântico oeste. As siglas dos pontos na costa correspondem a PA - Paracuru/CE, PP - Porto do Pecém/CE e IC - Icapuí/CE. A ilhas oceânicas correspondem a AR - Atol das Rocas/RN, FN - Fernando de Noronha/PE e AS - Arquipélago de São Pedro e São Paulo/PE. Com apenas um única amostragem na costa brasileira tropical, AC - Arraial do cabo/RJ. Fonte: o próprio autor. .... 34

Figura 9. Representação das correntes superficiais (linha escura) e sub-superficiais (linha tracejada) no perfil agrupado de 0-150m na zona equatorial e tropical do Atlântico Sul (PETERSON & STRAMMA, 1991; BUB & BROWN, 1996; BOURLÈS et al., 1999; STRAMMA, 1999; FRATANTONI & GLICKSON, 2002; TCHAMABI et al., 2017). A intensidade das correntes está de acordo com a espessura da linha. Os termos abreviados são: NEC - Corrente Superficial Equatorial Norte; NECC - Contra Corrente Superficial Equatorial Norte; NEUC - Contra Corrente Sub-Superficial Equatorial Norte; EUC - Contra Corrente Sub-Superficial Equatorial; SEC - Corrente Superficial Equatorial Sul; SECC - Contra Corrente Superficial Equatorial Sul; SEUC - Contra Corrente Sub-Superficial equatorial Sul; SAC - Corrente Superficial do Atlântico Sul; NBC - Corrente Superficial Norte do Brasil. Os giros do Atlântico Sul estão em destaque, com o sentidos representados. O perfil de profundidade e direção das correntes está em evidencia no porção inferior direita do mapa (PETERSON & STRAMMA, 1991; STRAMMA, 1999). Fonte: o próprio autor. .... 42

Figura 10. Desenho amostral do processamento desde a amostragem da água superficial até o armazenamento das amostras em galão de nitrogênio líquido (-196°C). As amostras fixadas foram utilizadas para os experimentos de citometria, enquanto que as amostras de membrana foram utilizadas para as análises de POC (0,45µm) e análises genômica (filtros de 3 e 0,22µm). Filtration - filtração por gravidade; TFF - filtração tangencial; VF - filtração a vácuo. Fonte o próprio autor. .... 47

Figura 11. Desenho amostral dos parâmetros físico-químicos e de variáveis categórica da coluna da água durante o processo de amostragem. Fonte o próprio autor. .... 48

Figura 12. Desenho amostral para obtenção do carbono orgânico particulado - POC através do método de digestão ácida (WALKLEY & BLACK, 1934; MENZEL & VACCARO, 1964; GRASHOFF et al., 1999). Fonte o próprio autor. .... 49

Figura 13. Desenho amostral do processamento das imagens baixadas pelo banco de dados Giovanni (NASA<sup>®</sup>) para extração dos valores médios de POC por pixel (16Km<sup>2</sup>). Sensor orbital MODIS-Aqua. Fonte o próprio autor. .... 51

Figura 14. Desenho amostral para obtenção da biomassa de cianobactérias, a fração nano e

- picoplantônicas através da densidade celular obtida pela citometria de fluxo (MARIE et al., 1997; PARTENSKY & VAULOT, 1999; DUBELAAR & JONKER, 2000; EIGEMANNA et al., 2013; MARIE et al., 2014). Fonte o próprio autor. .... 54
- Figura 15. Esquema do processamento genômico das membranas de 3,00 $\mu$ m e 0,22 $\mu$ m desde a extração até as técnicas de isolamento e identificação dos genes que possibilitam o estudo da composição e filogenética microbiana. Análises de fingerprint e sequenciamento através de bibliotecas rRNA 16s foram utilizadas para identificar a estrutura da comunidade microbiota. Fonte: o próprio autor. .... 57
- Figura 16. Agrupamentos pelas médias dos parâmetros físico-químicos de temperatura, salinidade, pH e oxigênio dissolvido. Os agrupamentos foram baseados em matrizes de dissimilaridade seguindo a distância euclidiana. Fonte: o próprio autor. .... 68
- Figura 17. A variação dos pontos seguiu a análise de componentes principais (PCA) considerando quatro parâmetros analisados durante as amostragem na coluna da água: pH; temperatura; salinidade; oxigênio dissolvido. Fonte: o próprio autor. .... 69
- Figura 18. Médias e desvios das réplicas de cada amostra para POC por digestão ácida e POC mensurado pelo sensor remoto MODIS-Aqua, ambas amostras para o período de outubro de 2015. O POC remoto foi mensurado apenas para as amostras dos arquipélagos de Fernando de Noronha - FN, Atol da Rocas - AT e São Pedro e São Paulo - AS. Fonte: o próprio autor. .... 71
- Figura 19. As imagens acima correspondem as concentrações de POC remoto via plataforma Giovanni (NASA<sup>®</sup>) adquiridas pelo sensor orbital MODIS-Aqua. As coordenadas do enquadramento das imagens são 55°O 5°N 15°E 10°S, com cada pixels equivalente a 16Km<sup>2</sup> (4 x 4Km). Cada pixel corresponde: A - Média mensal do mês de outubro de 2015, período de coleta das amostras nos arquipélagos de FN e AS; B - Média anual para o ano de 2015; C - Média da série histórica de 5 anos, de 2010-2015. Fonte: o próprio autor. .... 74
- Figura 20. Médias e desvios da concentração celular total (All) das réplicas de cada amostra. As concentrações foram estimadas por marcações celulares adquiridas pela técnica de citometria de fluxo. As faixas de tamanho delimitada pela beads marcadas (2 e 4 $\mu$ m) selecionou células de 0,22 a 6,00 $\mu$ m, correspondendo as frações nano e picofitoplanctônicas. Fonte: o próprio autor. .... 76
- Figura 21. Na parte superior da figura estão dispostas as porcentagens de cada uma das populações fitoplanctônicas identificadas em relação ao somatório total de células. Na parte inferior da figura estão os dotplots gerados pela citometria de fluxo, sendo que, cada ponto corresponde a uma célula distribuídas pelos eixos de fluorescência amarelo (FL2) e vermelho (FL3). Fonte: o próprio autor. .... 77
- Figura 22. Na parte superior da figura estão dispostas as porcentagens de cada uma das populações fitoplanctônicas identificadas em relação ao somatório total de células. Na parte inferior da figura estão os dotplots gerados pela citometria de fluxo, sendo que, cada ponto corresponde a uma célula distribuídas pelos eixos de fluorescência amarelo (FL2) e vermelho (FL3). Fonte: o próprio autor. Fonte: o próprio autor. .... 78
- Figura 23. Na parte superior da figura estão dispostas as porcentagens de cada uma das populações fitoplanctônicas identificadas em relação ao somatório total de células. Na parte inferior da figura estão os dotplots gerados pela citometria de fluxo, sendo que, cada ponto corresponde a uma célula distribuídas pelos eixos de fluorescência amarelo (FL2) e vermelho (FL3). Fonte: o próprio autor. Fonte: o próprio autor. .... 79
- Figura 24. A variação dos pontos seguiu uma análise de componentes principais (PCA)

considerando as cinco populações fitoplanctônicas identificadas por citometria: *Phyto*; *Synechococcus*; *Prochlorococcus*; *Nanoeukaryotics*; *Cyanochlorophyllates*. Os pontos insulares foram bem isolados dos pontos da costa. Os círculos agrupam as localidades de cada ponto. Fonte: o próprio autor. .... 80

Figura 25. Médias e desvios das réplicas de cada ponto para biomassa mensurada através da concentração celular e índices de biovolume. As barras indicam os valores reais de biomassa as populações de cianobactérias. Fonte: o próprio autor. .... 82

Figura 26. O escalonamento multidimensional (MDS) analisou a distribuição dos pontos em função das populações de cianobactérias. O baixo estresse revelou que a dimensionalidade dos pontos foi representada de maneira ideal. Fonte: o próprio autor. .... 83

Figura 27. Médias e desvios das réplicas a cada ponto para a riqueza observada para as frações do pico e nanofitoplâncton. na porção superior estão representadas a riqueza provinda dos dados de DGGE, abaixo, a riqueza dos dados de sequenciamento de amplicons 16S rRNA. Fonte: o próprio autor. .... 86

Figura 28. Curvas de rarefação de “espécies” considerando o índice de riqueza estimada Chao2 - presença/ausência. As replicas dos pontos de amostragem (13 replicas) representam as repetições de amostragem ao ambiente delimitado, no caso a costa. As barras amarelas indicam os desvios das replicas a cada ponto, enquanto que os traços pretos representam a media da riqueza total acumulada ao longo das amostragens. Os limites tracejados nas barra indicam a extrapolação para espécies estimadas. A área azul indica o ambiente de rarefação, contemplando os limites os máximos e miminhos da riqueza no ambiente costeiro estudado. Fonte: o próprio autor. .... 87

Figura 29. Agrupamentos pelas medias de beta diversidade das amostras costeiras. Os agrupamentos foram baseados em matrizes de dissimilaridade pelo índice binário de Jaccard. Fonte: o próprio autor. .... 89

Figura 30. Total de sequencias extraídas da biblioteca 16S rRNA para cianobactérias através da plataforma MiSeq (Illumina). Estas sequencias representam apenas as que passaram pelo processo de pré-tratamento e filtragem seguindo o QIIME com posterior alinhamento na base de dados do SILVA. Fonte: o próprio autor. .... 91

Figura 31. Representação do escalonamento multidimensional (MDS) das réplicas em função das sequencias extraídas da fração nanofitoplanctônica - 3,00µm. O baixo estresse revelou que a dimensionalidade dos pontos foi representada de maneira justa. Fonte: o próprio autor. .... 92

Figura 32. O baixo estresse encontrado na MDS das réplicas da fração picofitoplacntônica - 0,22µm - revelou uma distribuição ideal. Fonte: o próprio autor. .... 92

Figura 33. Médias e desvios das réplicas a cada ponto para a riqueza observada nas frações do pico e nanofitoplâncton a partir dos dados de sequenciamento de amplicons da biblioteca 16S rRNA. Fonte: o próprio autor. .... 97

Figura 34. Curvas de rarefação de “espécies” considerando o índice de riqueza estimada Chao1 - abundancia. Nas imagens superiores as replicas dos pontos de amostragem (37 réplicas) representam as repetições de amostragem na costa e nos arquipélagos. As barras amarelas indicam os desvios das replicas a cada ponto, enquanto que os traços pretos representam a media da riqueza total acumulada ao longo das amostragens. Os limites tracejados nas barra indicam a extrapolação para espécies estimadas. A área azul indica o ambiente de rarefação, contemplando os limites os máximos e miminhos da riqueza no ambiente costeiro estudado. Nas figuras abaixo estão demonstradas as curvas de acumulação de espécies a cada localidade,

estão identificadas por distintas cores. O numero de sequencias acumulado em cada replica esta representado no eixo x. Fonte: o próprio autor. .... 99

Figura 35. Agrupamentos pelas medias de beta diversidade das amostras insulares e costeiras das menores facções do fitoplâncton marinho em aguas superficiais da porção oeste do Atlântico Equatorial. Os agrupamentos foram baseados em matrizes de dissimilaridade pelo índice binário de Jaccard. Fonte: o próprio autor. .... 99

Figura 36. Relação da riqueza total observada aos diferentes cenários propostos, relacionando o ambiente costeiro e insular às frações do nano e picofitoplâncton. As caixas coloridas demonstram o numero de espécies raras a cada um dos 4 cenários: Nano - costa; nano - arquipélago; pico - costa; pico - arquipélago. Fonte: o próprio autor. .... 100

Figura 37. As correlações entre táxons e amostras da fraco nanofitoplanctônica (3.00 $\mu$ m) estão dispostas em um heatmap, com a barra de valores de correlação na parte superior. A distribuição de Jaccard foi utilizada para construir os dendogramas utilizados no heatmap, tanto o cluster das abundancias das sequencias agrupadas pelas localidades (réplicas), quanto o cluster dos táxons de maior abundancia. Na porção superior direita da figura esta ilustrada as correlações com suas respectivas significâncias entre os táxons mais abundantes. Os símbolos \*, \*\* e \*\*\* indicam significâncias de 0,05, 0,01 e 0,001, respectivamente. Fonte: o próprio autor. .... 103

Figura 38. As correlações entre táxons e amostras da fraco picofitoplanctônica (0,22 $\mu$ m) estão dispostas em um heatmap, com a barra de valores de correlação na parte superior. A distribuição de Jaccard foi utilizada para construir os dendogramas utilizados no heatmap, tanto o cluster das abundancias das sequencias agrupadas pelas localidades (réplicas), quanto o cluster dos táxons de maior abundancia. Na porção superior direita da figura esta ilustrada as correlações com suas respectivas significâncias entre os táxons mais abundantes. Os símbolos \*, \*\* e \*\*\* indicam significâncias de 0,05, 0,01 e 0,001, respectivamente. Fonte: o próprio autor. .... 104

Figura 39. A variação dos pontos seguiu uma análise de redundância (RDA) com as sequencias alinhadas da fração 3,00 $\mu$ m do fitoplâncton em cada replica. Apenas variáveis significativas foram utilizadas na analise de variação: pH; salinidade, diferentes substratos e oxigênio dissolvido. As populações melhores marcadas foram *Synechococcus* e *Prochlorococcus*. Fonte: o próprio autor. .... 106

Figura 40. A variação dos pontos seguiu uma análise de redundância (RDA) com as sequencias alinhadas da fração picoplanctônica - 0,22 $\mu$ m em cada replica. Apenas variáveis significativas foram utilizadas na analise de variação: pH; salinidade, diferentes substratos e oxigênio dissolvido. As populações melhores marcadas foram *Synechococcus* e *Prochlorococcus*. Fonte: o próprio autor. .... 107

Figura 41. Análise de dispersão da beta-diversidade em função das variáveis categóricas de substrato (fundo rochoso, planícies arenosa e poças de maré) e dinâmica (abrigado e exposto), tato para as frações do nano como do picofitoplâncton. Fonte: o próprio autor. .... 107

## LISTA DE TABELAS

- Tabela 1. Grupos taxonômicos representantes do fitoplâncton dulcícola e marinho (MOORE et al., 1995; ANDERSEN, 2005; REYNOLDS, 2006; HALLENBECK, 2017). Sendo a lacuna preenchida de cinza escuro considerada relevante; cinza claro considerada pouco relevante (baixa diversidade, distribuição ou biomassa); e a cor negra considerada ausente ou escasso. Fonte: o próprio autor. .... 02
- Tabela 2. Diferentes formas de vida dos grupos fitoplanctônicos. Espaços marcados de cinza escuro indicam que tal grupo apresenta esta forma de vida (MOORE et al., 1995; ANDERSEN, 2005; REYNOLDS, 2006; HALLENBECK, 2017). Fonte: o próprio autor. .... 05
- Tabela 3. Principais especializações do fitoplâncton para movimentação e sustentação na coluna da água. Espaços marcados de cinza escuro indicam que tal grupo apresenta esta especialização (MOORE et al., 1995; ANDERSEN, 2005; REYNOLDS, 2006; HALLENBECK, 2017). Fonte: o próprio autor. .... 05
- Tabela 4. Principais características dos gêneros cocóides (picofitoplâncton) de maior importância na produção primária nos giros oceânicos. Fonte: adaptado de Kirchman (2011). 13
- Tabela 5. Caracterização das amostras com descrições ambientais e geográficas aos respectivos pontos de coleta, assim como a descrição dos parâmetros físico-químicos da água superficial amostrada no início de cada coleta. Fonte: o próprio autor. .... 45
- Tabela 6. Algoritmos de conversão extraídos da literatura para a conversão do número celular do picoplâncton marinho em gramas de carbono por litro. As diversas referências listadas representam os artigos mais citados para a determinação das constantes de conversão. Estes valores estão devidamente adequados para transformar uma abundância celular ( $\text{cell L}^{-1}$ ) em biomassa de carbono ( $\text{pg C cell}^{-1} \text{L}^{-1}$ ). As referências marcadas em cinza tiveram os algoritmos de biomassa utilizados neste trabalho. Fonte: o próprio autor. .... 55
- Tabela 7. Os principais primers aplicados a filogenia de cianobactérias que codificam as regiões do DNA genômico entre as porções rRNA 16S e 23S. Os primers marcados em cinza foram os utilizados nos experimentos de DGGE e sequenciamento NGS neste trabalho. Fonte: o próprio autor. .... 59
- Tabela 8. Listagem das soluções e suas respectivas concentrações e quantidades para formar uma solução tampão ideal para o funcionamento da Go Taq e anelamento correto dos primers da PCR-DGGE 1. A cada amostra são somados 14,7 $\mu\text{L}$  do Mix, e portanto, 10,3 $\mu\text{L}$  de água para biologia molecular para formar o volume final de solução. Fonte: o próprio autor. .... 61
- Tabela 9. Condições ideais aplicadas ao termociclador durante o processo da PCR, considerando as particularidades físico-químicas do conjunto de primers e da enzima Taq durante a PCR semi-nexted 1. Foram as mesmas condições para as técnicas da PCR-DGGE e NGS. Fonte: o próprio autor. .... 61
- Tabela 10. Listagem das soluções e suas respectivas concentrações e quantidades para formar uma solução tampão ideal para a amplificação e anelamento correto dos primers da PCR-DGGE 2. A cada amostra são 27,2 $\mu\text{L}$  do Mix e portanto 22,8 $\mu\text{L}$  de água para biologia molecular para formar o volume final de solução. Fonte: o próprio autor. .... 62
- Tabela 11. Condições ideais aplicadas ao termociclador durante o processo da PCR, considerando as particularidades físico-químicas do conjunto de primers e da enzima Taq durante a PCR semi-nexted 2. Foram as mesmas condições para as técnicas da PCR-DGGE e NGS. Fonte: o próprio autor. Fonte: o próprio autor. .... 62

Tabela 12. Listagem das soluções e suas respectivas concentrações e quantidades para formar uma solução tampão ideal para o funcionamento da High Fidelity Taq e anelamento correto dos primers da PCR-NGS 1. A cada amostra são somados 8,7µL do Mix, e portanto, 16,3µL de água para biologia molecular para formar o volume final de solução. Fonte: o próprio autor. . 64

Tabela 13. Listagem das soluções e suas respectivas concentrações e quantidades para formar uma solução tampão ideal para o funcionamento da High Fidelity Taq e anelamento correto dos primers da PCR-NGS 1. A cada amostra são somados 7,7µL do Mix, e portanto, 17,3µL de água para biologia molecular para formar o volume final de solução. Fonte: o próprio autor. . 65

Tabela 14. Listagem das soluções e suas respectivas concentrações e quantidades para formar uma solução tampão ideal para o funcionamento da High Fidelity Taq e anelamento correto dos primers da PCR-Illumina Index. A cada amostra são somados 24,2µL do Mix, e portanto, 25,8µL de água para biologia molecular para formar o volume final de solução. Fonte: o próprio autor. .... 65

Tabela 15. Condições ideais aplicadas ao termociclador durante o processo da PCR, considerando as particularidades físico-químicas do conjunto de primers e da enzima Platinum Taq DNA High Fidelity durante a PCR-Illumina Index. Fonte: o próprio autor. .... 65

Tabela 16. Desvios dos valores observados dos parâmetros físico-químicos da água durante as amostragens nas diferentes localidades. Os valores azuis correspondem a desvios positivos, enquanto que os valores em vermelho correspondem aos desvios negativos. Fonte: o próprio autor. .... 67

Tabela 17. Desvios dos valores observados de POC e POC remote em relação as médias temporais consideradas normais para a áreas amostradas na costa e nas porções insulares. Os valores azuis correspondem a desvios positivos, enquanto que os valores em vermelho correspondem aos desvios negativos. Fonte: o próprio autor. .... 72

Tabela 18. Desvios entre as médias temporais de POC e POC remote para cada arquipélago. A serie temporal de 5 anos foi considerada a normal. Fonte: o próprio autor. .... 73

Tabela 19. Porcentagens da biomassa acumulada nas populações de cianobactérias em relação ao POC observado na coluna da água referente a cada local (estoque relativo). O POC do ambiente costeiro e dos arquipélagos foi estipulado pelas médias das respectivas localidades. Fonte: o próprio autor. .... 82

Tabela 20. Porcentagens da biomassa observada de cianobactérias entre as médias temporais de POC e POC remote para cada arquipélago. A biomassa media dos arquipélagos foi estipulado pelas medias das localidades de AS e FN. Fonte: o próprio autor. .... 83

Tabela 21. Desvios dos valores observados em relação a media da correspondente fração fitoplanctônica. Cada parâmetro ecológico possui um valor estipulado tanto pela técnica molecular da DGGE quanto para a técnica do sequenciamento. Os valores azuis correspondem a desvios positivos, enquanto que os valores em vermelho correspondem aos desvios negativos. Fonte: o próprio autor. .... 88

Tabela 22. Desvios dos valores observados em relação a media correspondente ao respectivo ambiente, costeiro ou insular, frente aos parâmetro ecológicos de riqueza e diversidade. Duas frações do fitoplâncton são comparadas, a fração do nanofitoplâncton e picofitoplâncton. As medias são baseados em matrizes de abundancia de sequências 16S rRNA para cianobactérias. Os valores azuis correspondem a desvios positivos, enquanto que os valores em vermelho correspondem aos desvios negativos. Fonte: o próprio autor. .... 98

Tabela 23. Grupos taxonômicos encontrados pela identificação das sequencias da biblioteca 16S rRNA para cianobactérias na porção superficial do Atlântico. Os grupos de trabalho estão demonstrados na variável Taxa, os quais reúnem as menores classificações taxonômicas encontradas, sendo gênero a menor delas. Cada grupo taxonômico possui o respectivo preenchimento dentre os 4 diferentes cenários analisados. Fonte: o próprio autor. .... 101

## SUMÁRIO

<b>1. Introdução</b> .....	<b>01</b>
<b>1.1. Fitoplâncton</b> .....	<b>01</b>
<b>1.2. Cianobactérias</b> .....	<b>06</b>
<b>1.3. Filogênia de microrganismos (16s rRNA)</b> .....	<b>07</b>
<b>1.3.1. Taxonomia das cianobactérias</b> .....	<b>09</b>
<b>1.3.2. Gêneros <i>Synechococcus</i> e <i>Prochlorococcus</i> (picofitoplâncton)</b> .....	<b>11</b>
<b>1.3.3. Ecologia microbiana</b> .....	<b>13</b>
<b>1.3.4. Técnicas moleculares aplicadas à ecologia molecular</b> .....	<b>14</b>
<b>1.4. Diversidade e biomassa do picofitoplâncton na zona equatorial</b> .....	<b>19</b>
<b>1.5. Fotossíntese</b> .....	<b>23</b>
<b>1.5.1. Produção primária</b> .....	<b>29</b>
<b>1.5.2. Determinação de biomassa (carbono)</b> .....	<b>31</b>
<b>2. Caracterização oceanográfica e ambiental da área de estudo</b> .....	<b>34</b>
<b>2.1. Circulação oceânica - Atlântico equatorial oeste</b> .....	<b>35</b>
<b>2.1.1. Ilhas oceânicas</b> .....	<b>37</b>
<b>2.2. Circulação costeira - correntes do Brasil e Norte do Brasil</b> .....	<b>39</b>
<b>2.2.1. Costa do CE</b> .....	<b>40</b>
<b>2.2.2. Arraial do Cabo</b> .....	<b>41</b>
<b>3. Objetivo</b> .....	<b>42</b>
<b>3.1. Objetivos específicos</b> .....	<b>43</b>
<b>4. Material e métodos</b> .....	<b>43</b>
<b>4.1. Amostragem, transporte e armazenamento</b> .....	<b>46</b>
<b>4.2. Forçantes ambientais</b> .....	<b>47</b>
<b>4.3. Biomassa na água</b> .....	<b>49</b>
<b>4.3.1. Oxidação por dicromato</b> .....	<b>49</b>
<b>4.3.2. Sensoriamento remoto</b> .....	<b>50</b>
<b>4.4. Biomassa das cianobactérias - citometria</b> .....	<b>51</b>
<b>4.5. Estrutura da comunidade</b> .....	<b>56</b>
<b>4.4.1. Extração de DNA</b> .....	<b>57</b>
<b>4.4.2. PCR-DGGE</b> .....	<b>60</b>
<b>4.4.3. PCR-sequenciamento em larga escala</b> .....	<b>62</b>
<b>4.6. Análises integradas</b> .....	<b>66</b>

<b>5. Resultados</b> .....	<b>66</b>
<b>5.1. Variáveis ambientais</b> .....	<b>66</b>
<b>5.2. Biomassa na água</b> .....	<b>69</b>
<b>5.3. Biomassa das cianobactérias</b> .....	<b>75</b>
<b>5.4. Estrutura da comunidade</b> .....	<b>83</b>
<b>5.4.1. Estudo preliminar - costa do CE</b> .....	<b>84</b>
<b>5.4.2. Sequencias alinhadas - costa e ilhas</b> .....	<b>89</b>
<b>5.5. Diversidade modulada pelas forçantes</b> .....	<b>102</b>
<b>6. Discussão</b> .....	<b>108</b>
<b>6.1. Concentração celular picofitoplanctônica</b> .....	<b>110</b>
<b>6.2. Carbono orgânico particulado e biomassa de cianobactérias</b> .....	<b>111</b>
<b>6.3. Diversidade de cianobactérias</b> .....	<b>117</b>
<b>6.3.1. Abundancia de sequencias</b> .....	<b>117</b>
<b>6.3.2. Diversidade taxonômica</b> .....	<b>120</b>
<b>6.4. Diversidade e variáveis oceanográficas</b> .....	<b>125</b>
<b>6.5. Comparação metodológica</b> .....	<b>126</b>
<b>7. Conclusão</b> .....	<b>127</b>
<b>8. Considerações finais</b> .....	<b>128</b>
<b>9. Referências</b> .....	<b>129</b>

# **Diversidade e estoque de carbono em cianobactérias na costa e nos arquipélagos da porção oeste do giro Equatorial Atlântico**

## **1. Introdução**

Os organismos do fitoplâncton desempenham a fundamental função ecológica de produtores primários, constituindo a base da cadeia trófica dos ambientes aquáticos (POMEROY, 1974; AZAM et al, 1982; WALTER, et al., 2017). Estas microalgas utilizam energia solar para sintetizar matéria orgânica e ATP (energia química), compostos que garantem sua própria manutenção e reprodução, e, como subproduto, geram oxigênio molecular (O<sub>2</sub>) liberado para a atmosfera, no processo conhecido como fotossíntese (POMEROY, 1974; AZAM et al, 1982; KIRCHMAN, 2011). Portanto, promovem a contínua renovação da matéria orgânica e garantem a manutenção da vida aeróbica na Terra, além de atuarem em processos biogeoquímicos, sendo a chave para a ciclagem de nutrientes nos mais diferentes biogeo Ciclos, como o do carbono, onde atuam fixando matéria orgânica e sequestrando carbono da atmosfera (MOORE et al., 1995; PROSSER et al, 2007; DELORENZO & FLEMING, 2008; TORRES et al., 2008).

### **1.1. Fitoplâncton**

Com base em características moleculares, a atual árvore da vida está composta por três grandes ‘domínios’: Eucarya (eucariontes), Bacteria (inicialmente chamado eubactérias) e Archaea (inicialmente chamado de arqueobactérias), sendo que estes últimos domínios eram denominados de procariotos (WOESE & FOX, 1977; PACE, 2007; HUG et al., 2016).

O termo fitoplâncton (Tabela 1) se refere a um agrupamento genérico de espécies, pois engloba diferentes táxons, contendo representantes de bactérias (cianobactérias) e eucariontes (microalgas e protozoários fotossintetizantes) (Figura 1). Todas as espécies que fazem parte desse grupo apresentam características comuns: (1) são microscópicas; (2) possuem capacidade de realizar fotossíntese; (3) estão adaptadas para passar parte ou todo o tempo de sua vida em suspensão em águas abertas oceânicas ou continentais (REYNOLDS, 2006). Vale ressaltar que constantemente estes táxons são atualizados e sofrem mudanças de enquadramento e nomenclatura, porém neste estudo considera-se a classificação empregada por Reynads (2006) (Tabela 1). No website livre do Registro Mundial de Espécies Marinhas (World Register of Marine Species - WoRMS, <http://www.marinespecies.org>), assim como na Lista de Algas do Mundo (Listing the World's Algae - AlgaeBase, <http://www.algaebase.org/>) é possível contextualizar futuras alterações taxonômicas.

Tabela 1. Grupos taxonômicos representantes do fitoplâncton dulcícola e marinho (MOORE et al., 1995; ANDERSEN, 2005; REYNOLDS, 2006; HALLENBECK, 2017). Sendo a lacuna preenchida de cinza escuro considerada relevante; cinza claro considerada pouco relevante (baixa diversidade, distribuição ou biomassa); e a cor negra considerada ausente ou escasso. Fonte: o próprio autor.

Domains (PACE, 1997)	Division/Phylum (REYNOLDS, 2006)	Freshwater	Seawater
Bacteria	Cyanobacteria		
	Glaucophyta		
	Prasinophyta		
	Chlorophyta		
	Euglenophyta		
	Cryptophyta		
	Raphidophyta		
	Xanthophyta		
	Eustigmatophyta		
	Chrysophyta		
	Bacillariophyta		
	Haptophyta		
	Dinophyta		

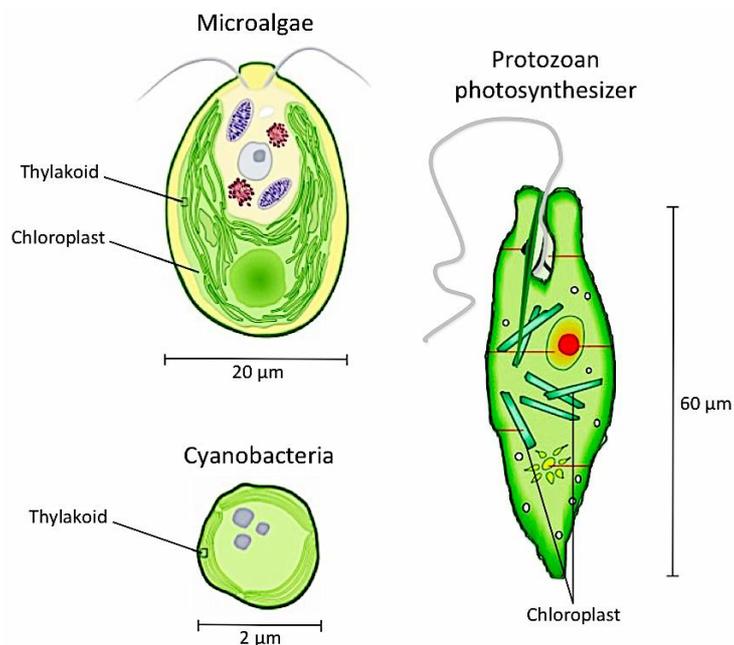


Figura 1. Esquema dos principais representantes do grupo fitoplâncton (MOORE et al., 1995; ANDERSEN, 2005; REYNOLDS, 2006; HALLENBECK, 2017). Fonte: o próprio autor.

As cianobactérias são um expoente no grupo fitoplanctônico, pois são as únicas representantes do domínio Bacteria e estão distribuídas em alta densidade nas águas continentais, oceanos e sedimentos.

As cianobactérias são seres primitivos, as quais habitam o planeta há mais de 3 bilhões de anos, inaugurando a transição da Terra anoxigênica às condições oxigenadas através da fotossíntese (SCHIRRMESTER et al., 2011; HALLENBECK, 2017). Apresentam diversas adaptações que as permitem ser cosmopolitas e apresentarem a maior biomassa dos ambientes aquáticos, sendo importantes e fundamentais no processo de remineralização do carbono (POMEROY, 1974; AZAM et al, 1982; VAN DEN HOEK et al., 1997; WHITTON & POTTS, 2000; VELDHUIS et al., 2005; KONOPKA, 2009; KIRCHMAN, 2011; THOMPSON et al., 2013b; HUG et al., 2016). Algumas espécies de cianobactérias apresentam compostos secundários chamados de cianotoxinas, e em eventos de florações elas podem promover a morte de peixes e outros organismos alterando a cadeia trófica da área afetada (WHITTON & POTTS, 2000; AZEVEDO & VASCONCELOS, 2006). Fluxos de matéria e energia nas teias alimentares oceânica são fortemente dependentes desses microorganismos, sendo o principal elemento da fixação de carbono pelo processo da fotossíntese e assim na produção primária (POMEROY, 1974; AZAM et al, 1982; ITURRIAGA & MITCHELL, 1986). Este grupo possui fundamental importância para a manutenção da vida em áreas de ressurgência costeira, áreas associadas à pesca e no centro dos giros oceânicos (áreas oligotróficas) (FIGUEIRAS et al., 2014). Além das cianobactérias outros grupos apresentam elevada diversidade e biomassa, como as clorófitas ou algas verdes (filo Chlorophyta), diatomáceas (filo Bacillariophyta) e dinoflagelados (filo Dinophyta). Estes grupos compõem grande parte da diversidade de microalgas.

As algas verdes (atuais clorófitas) receberam este nome por possuírem cloroplastos (organelas de característica esverdeada importantes no processo de fotossíntese) em abundância e em formas exageradas e bem variáveis (estrela, espiralado, parietal, etc.). Em sua maioria são representantes de ambientes continentais (lagos, rios, solos úmidos) (ANDERSEN, 2005; REYNOLDS, 2006).

As diatomáceas por sua vez, apresentam uma parede celular muito singular constituída de sílica e carboidratos (ausência de celulose), denominada frústula, que é composta por duas valvas que se encaixam e as cópulas ou bandas (encaixe das valvas). Com base na morfologia das valvas as diatomáceas podem ser divididas em cêntricas e penadas (TOMAS, 1997; REYNOLDS, 2006). As diatomáceas e outras duas ordens de microalgas marinhas, a Dictyochales - silicoflagelados (filo Ochrophyta - não abordado anteriormente, pois corresponde principalmente a macroalgas, possuindo apenas algumas classes unicelulares pertencentes ao fitoplâncton), e Coccolithophorales - cocolitoforídeos (filo Haptophyta) que são fundamentais para o biogeoquímico de silicatos e do carbonato de cálcio

nos oceanos, respectivamente (ANDERSEN, 2005; REYNOLDS, 2006; PROSSER et al, 2007).

Já os dinoflagelados são microalgas biflageladas fotossintetizantes, que podem possuir uma parede celular composta de celulose chamada de teca (placas poligonais) ou simplesmente serem células nuas (atecadas). Durante grandes florações, os gêneros *Prorocentrum*, *Ceratium*, *Cochlodinium*, *Gymnodinium*, *Gonyaulax* podem produzir substâncias potencialmente tóxicas, resultando em eventos de florações algais nocivas, anteriormente denominadas “marés vermelhas”, causando a morte de peixes e alterando a cadeia trófica (TOMAS, 1997; REYNOLDS, 2006).

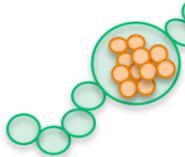
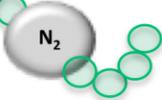
O filo Euglenophyta também apresenta peculiaridades, pois agrupa os protozoários fotossintéticos facultativos (flagelados - classe Euglenophyceae), que em condições apropriadas têm a capacidade de realizar fotossíntese, e na ausência de luz podem sobreviver como organismos heterotróficos (utilizando fagocitose para obtenção de alimento). Devido a essa amplitude metabólica, ocupa nichos ricos em matéria orgânica, sendo um excelente bioindicador de ambientes poluídos (TOMAS, 1997; ANDERSEN, 2005; REYNOLDS, 2006).

Com base na classificação taxonômica, o fitoplâncton pode apresentar particularidades morfofisiológicas, ou possuir diferentes formas de vida, pois podem ocorrer agrupados ou livres (Tabela 2). E ainda podem ter especializações para se movimentar/sustentar na coluna d'água (Tabela 3) (MOORE et al., 1995; REYNOLDS, 2006; HALLENBECK, 2017): (1) flagelo, estrutura alongada no exterior das células formado por um aparato proteico que propicia a locomoção ativa; (2) mucilagem, revestimento composto por polissacarídeos que envolvem as células com função de proteção, flutuação e/ou locomoção; (3) aerótopo, vacúolo de gás no interior das células que auxilia na flutuação; (4) acineto, é uma célula diferenciada e aumentada que acumula grânulos de reserva, formado em condições ambientais desfavoráveis entremeando cadeias filamentosas de cianobactérias, e que atua como um esporo de resistência de elevado peso, o qual decanta sobre o sedimento, gerando um novo organismo; (5) heterocisto, célula altamente especializada com a função de fixar nitrogênio atmosférico, formado em condições ambientais desfavoráveis entremeando cadeias filamentosas de cianobactérias, disponibilizando o composto fixado para o restante da cadeia, além de auxiliar na flutuação.

Tabela 2. Diferentes formas de vida dos grupos fitoplanctônicos. Espaços marcados de cinza escuro indicam que tal grupo apresenta esta forma de vida (MOORE et al., 1995; ANDERSEN, 2005; REYNOLDS, 2006; HALLENBECK, 2017). Fonte: o próprio autor.

Domains (PACE, 1997)	Division/Phylum (REYNOLDS, 2006)	Life forms		
		Unicellular	Colonial	Filamentous
				
Bacteria	Cyanobacteria			
	Glaucophyta			
	Prasinophyta			
	Chlorophyta			
	Euglenophyta			
	Cryptophyta			
	Raphidophyta			
	Xanthophyta			
	Eustigmatophyta			
	Chrysophyta			
	Bacillariophyta			
	Haptophyta			
	Dinophyta			

Tabela 3. Principais especializações do fitoplâncton para movimentação e sustentação na coluna da água. Espaços marcados de cinza escuro indicam que respectivo grupo apresenta esta especialização (MOORE et al., 1995; ANDERSEN, 2005; REYNOLDS, 2006; HALLENBECK, 2017). Fonte: o próprio autor.

Domains (PACE, 1997)	Division/Phylum (REYNOLDS, 2006)	Specializations				
		Flagella	Mucilage	Aerotope	Acineto	Heterocyst
						
Bacteria	Cyanobacteria					
	Glaucophyta					
	Prasinophyta					
	Chlorophyta					
	Euglenophyta					
	Cryptophyta					
	Raphidophyta					
	Xanthophyta					
	Eustigmatophyta					
	Chrysophyta					
	Bacillariophyta					
	Haptophyta					
	Dinophyta					

Além da divisão taxonômica, o fitoplâncton também pode ser classificado pelo ponto de vista ecológico, esta classificação não é modulada pelos grupos taxonômicos, uma vez que diferentes espécies dentro do mesmo clado podem apresentar formas (arredondadas, ovais, septiforme, com projeções, espinhos, etc.) e tamanhos muito variados (Figura 2).

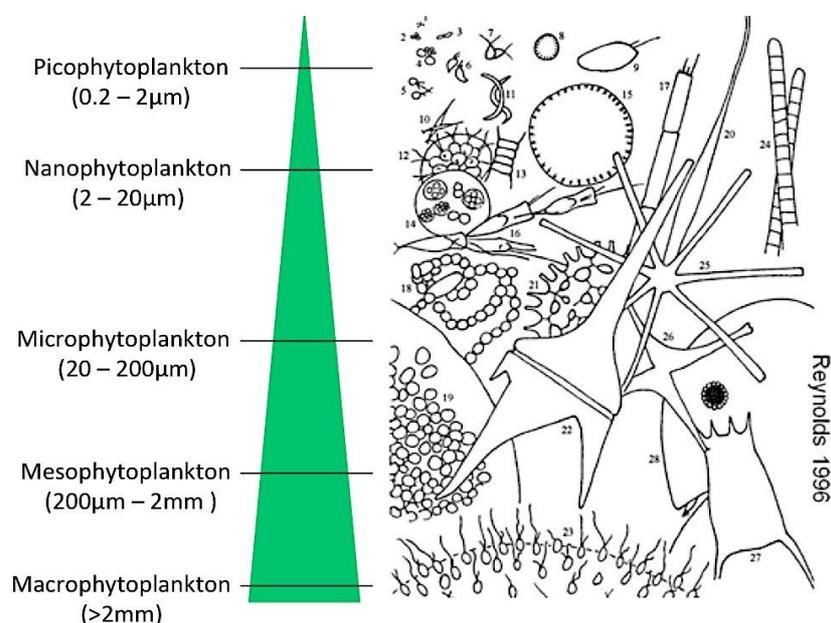


Figura 2. Classificação do fitoplâncton de acordo com a dimensão linear máxima (diâmetro) seguindo a nomenclatura proposta por Sieburth et al. (1978). Fonte: adaptado de Reynolds (1996).

## 1.2. Cianobactérias

As cianobactérias estão representadas por algumas espécies que estão dentro da fração nanofitoplanctônica e ainda englobam todos os representantes do picofitoplâncton, sendo estas frações correspondentes à maior parte da biomassa em águas oligotróficas, áreas nas quais os nutrientes estão em concentração extremamente baixa (<10nM) (ITURRIAGA & MITCHELL, 1986).

Esses microrganismos possuem características únicas e, portanto, papéis únicos no meio ambiente desempenhando a mesma função principal das algas eucarióticas, a produção primária. De fato, as cianobactérias compartilham muitos traços fisiológicos com algas eucarióticas e plantas superiores, usando o mesmo mecanismo para converter a energia da luz em energia química e convertem o carbono inorgânico em carbono orgânico pelo mesmo caminho, o ciclo de Calvin-Benson-Bassham (CBB), ou apenas, ciclo de Calvin (enzima rubisco) (COHEN & GUREVITZ, 2006; GAULT & MARLER, 2009; PALINSKA & SUROSZ, 2014). Mas em todos os outros aspectos, as cianobactérias são fieis ao domínio Bacteria, não possuindo cloroplastos ou organelas, e apresentando paredes celulares com componentes como ácido murâmico e lipopolissacarídeos que são encontrados apenas em

bactérias gram-negativas (WHITTON & POTTS, 2000; VELDHUIS et al., 2005; COHEN & GUREVITZ, 2006; GAULT & MARLER, 2009).

Assim como outros procariotas, as cianobactérias contêm duas estruturas de ácido desoxirribonucleico, o DNA cromossômico e uma subunidade menor autônoma, molécula extra-cromossômica chamada de plasmídeo. Ambos os tipos de DNA podem ser usados para o processo de gene-transferência e manipulação (ZHAXYBAYEVA et al., 2006). Inicialmente o genoma de cianobactérias foi escalonado pela medida de massa atômica, sendo que um trabalho pioneiro identificou um tamanho médio de  $1,6 \times 10^9$  daltons (Da) em cepas cultiváveis unicelulares, comparável com outros genomas bacterianos descritos até aquele momento. E cerca de  $8,6 \times 10^9$  Da para cepas filamentosas maiores, excedendo os maiores genomas descritos até o momento para procariotas (HERDMAN et al., 1979). Essa gigantesca massa de DNA levantou a possibilidade de que algumas sequências (regiões) de DNA não possuíssem uma capacidade de codificação, chamadas atualmente de íntrons (HERDMAN et al., 1979; COHEN & GUREVITZ, 2006). Técnicas moleculares modernas podem identificar de maneira precisa o comprimento real dos genomas e sua sequência exata de nucleotídeos (pares de base - pb). As cianobactérias, possuem uma variação no tamanho do genoma de 1,6Mb (*Prochlorococcus* sp. MIT9301) até 9,2Mb (*Nostoc punctiforme* PCC 73102) (SHIH et al., 2013). Gêneros unicelulares e não nitrogenados, como *Prochlorococcus* sp. e *Synechococcus* sp. possuem os menores tamanhos de genoma e representam biomassa significativa nos ecossistemas aquáticos (JIANG et al., 2010). Enquanto as cianobactérias filamentosas compreendem o maior tamanho do genoma, como *Nostoc* sp. PCC 7120, *Anabaena variabilis* ATCC 29413 e *Trichodesmium* sp. IMS 101 (JIANG et al., 2010). Os genomas de cianobactérias revelam uma história evolutiva complexa, que não pode ser representada por uma única árvore estritamente bifurcada à maioria dos genes (ZHAXYBAYEVA et al., 2006).

### **1.3. Filogenia de microrganismos (16S rRNA)**

As técnicas modernas de análise molecular, estão atualmente, se posicionando como definitivas para resolver varias questões de classificação taxonômica dentre os microrganismos, procariotas ou eucariotas. Anteriormente apenas cepas cultiváveis em meios artificiais podiam ser classificadas e enquadradas dentro de árvores filogenéticas. Portanto, a grande maioria dos microrganismos dos ambientes naturais estava inacessível ao advento das análises moleculares independentes de cultivo, as quais utilizam a heterogeneidade de sequências de genes para ter acesso à diversidade microbiana em um determinado ambiente (VANDAMME et al., 1996; BOYER et al., 2001; COLARES, 2014). Menos de 1% dos microrganismos encontrados em ambientes aquáticos, sedimentos e solos são recuperados em meios de cultivos, sendo estimado até o presente momento, que apenas 0,1% a 10% das espécies existentes foram descobertas e identificadas, dependendo do habitat estudado

(AMANN et al., 1995; BEZERRA, 2015). Sendo assim, o que não era representativo antes das técnicas independentes de cultivo, agora pode ser melhor mensurado obtendo uma diversidade mais próxima do real aos microrganismos presentes nos ambientes naturais, assim como não era representativo ao real agrupamento destas espécies.

Os primeiros genes a serem utilizados em tais estudos correspondiam aos ácidos ribonucleico ribossomais (rRNA), que codificam as estruturas dos ribossomos intracelulares. Estas macromoléculas possuem regiões que são consideradas conservadas estrutural e funcionalmente ao mesmo tempo em que apresentam regiões conservadas; regiões variáveis; e regiões hipervariáveis, tanto na estrutura primária quanto na secundária, suficientemente variáveis para determinar relações filogenéticas próximas (COLARES, 2010; BEZERRA, 2015). As sequências conservadas podem ser utilizadas para medir a distância evolutiva mesmo entre organismos mais distantemente relacionados, diferenciando índices de mudanças evolutivas pela sequência de nucleotídeos (PACE, 1997). E estas regiões não estão sujeitas a sofrer mudanças rápidas em sua sequência, ao menos pela sua função central na expressão de genes (WOESE et al., 1990; SINCLAIR et al., 2015). Desta forma, o gene do rRNA é um excelente cronômetro molecular, além de ser encontrado em todos os organismos, e serem essenciais para a síntese de proteínas.

Existem três distintos RNAr em procariotos, separados devido sua massa molecular estimada em ultra-centrífuga: o 5S (~120 nucleotídeos), o 16S (~1600 nucleotídeos) e o 23S (~3000 nucleotídeos). As porções 5S e 16S têm sido as mais utilizados para a maioria dos estudos em ecologia microbiana, principalmente por razões técnicas e históricas (PACE, 1997 COLARES, 2014). O 5S RNAr, foi largamente utilizado no final dos anos 1960 para análise de sequências, entretanto, a escassez de regiões de variabilidade limitou a sua utilização, devido a sua diminuta sequência. Já o 23S apresenta, tipicamente, diversas regiões de elevada variação, incluindo inserções e/ou deleções. Atualmente, o 16S RNAr é considerado com um tamanho apropriado para análises filogenéticas, por justamente possuir distintas regiões que compartimentalizam distintas informações, mais recentes (V1-V9) e concomitantemente sequencias mais conservadas (C1-C9) (PETROSINO et al., 2009). Particularmente para bactérias e arqueias, o gene 16S RNAr é usado para determinar as relações filogenéticas (BOYER et al., 2001; COLARES & MELO, 2013; VĚTROVSKÝ & BALDRIAN, 2013; SINCLAIR et al., 2015; HALLENBECK, 2017). A análise atual das árvores filogenéticas bacterianas muitas vezes dependem da construção de clusters de similaridade entre as espécies (OTUs - Operational Taxonomic Units) pelo genes 16S rRNA (VĚTROVSKÝ & BALDRIAN, 2013).

Os avanços recentes na genômica bacteriana permitem analisar o conjunto de dados existente para responder questões importantes sobre o potencial do 16S na ecologia e

taxonomia da comunidade bacteriana (ACINAS et al., 2008; VĚTROVSKÝ & BALDRIAN, 2013).

A necessidade de uma definição de espécie com base em genômica está emergindo de vários estudos independentes em todo o mundo, buscando uma definição de maneira unificada (THOMPSON et al., 2013a). Mas se tratando apenas da análise de sequências provindas do sequenciamento do rRNA e considerando a unidade taxonômica de espécie como a base da classificação filogenética, pelo menos 98,8% da sequência de 16S rRNA tem de ser distinta para configurar duas espécies diferentes (VANDAMME et al., 1996; WALTER, et al., 2017).

Os eucariotos, geralmente, possuem diversas cópias dos genes rRNA, porém organizados em repetições. As subunidades dos ribossomos eucariotos possuem quatro tipos de rRNA citoplasmático, as subunidades 28S; 5,8S; 18S e 5S. O 18S rRNA dos eucariotas tem a mesma função filogenética que o 16S rRNA aos microrganismos do domínio Bacteria (VAULOT et al., 2008; VĚTROVSKÝ & BALDRIAN, 2013).

A análise do DNA em si, é uma das formas mais íntegras de diferenciação de espécies. Devido, ainda, aos altos custos de sequenciamento de genomas inteiros, poucas comparações de genomas integrais suprem uma filogenia baseada apenas neste tipo de análise (VANDAMME et al., 1996). Primeiramente, uma diferença de nucleotídeos >70% para dois genomas integrais, poderia ser considerada válida para configurar duas distintas espécies pela técnica de hibridação do genoma-genoma (GGDH) *in silico* (via executada ou simulada por software) (VANDAMME et al., 1996). Outras abordagens mais recentes, ainda utilizando o DNA, complementam que para definir cepas das mesmas espécies microbianas, tem de compartilhar: >95% da identidade média de aminoácidos (AAI) e da identidade média de nucleotídeo (ANI); > 95% de identidade baseada em genes de alinhamento múltiplos; <10 para a assinatura genômica de Karlin (THOMPSON et al., 2013a). E ainda se aborda atualmente em taxonomia molecular que espécies do mesmo gênero formarão grupos monofiléticos com base em sequências de genes 16S rRNA, análise de sequências multilocus (MLSA) e análises supertree (uma única árvore filogenética gerada a partir de uma combinação de árvores filogenéticas menores, usando conjuntos de dados morfológicos e moleculares) (THOMPSON et al., 2013a).

Lembrado que as técnicas serão devidamente abordadas em um tópico mais abaixo.

### **1.3.1. Taxonomia das cianobactérias**

As cianobactérias são muito distintas metabólica e morfológicamente, sendo um dos grupos mais diversos dentre os procariotos, o qual é famoso pela classificação problemática e excepcionalmente tumultuada. Elas são, indiscutivelmente, um dos grupos mais antigos do planeta Terra, com alguns representantes fósseis muito semelhantes morfológicamente às espécies atuais (KOMÁREK et al., 2014; HALLENBECK, 2017). Inicialmente foram

classificadas juntos as plantas (eucariotos), e somente posteriormente foram reclassificadas como bactérias (STANIER et al., 1978), com base nas suas características morfológicas. Tal sistema de classificação continha 5 grandes ordens e foi bem sucedido, considerando apenas culturas puras (CASTENHOLZ, 2001; CAVALIER-SMITH, 2002; WATERBURY, 2006). As cinco ordens eram agrupadas em: Chroococcales (cianobactérias unicelulares com reprodução por fissão binária e formas de cocos, hastes (bastonetes) ou ainda agregados celulares); Pleurocapsales (reprodução por fissão múltipla, desde formas unicelulares a agregados celulares); Oscillatoriales (cianobactérias filamentosas ausentes de células especializadas e divisão celular por fissão binária com formas de haste e/ou disco); Nostocales (filamentosas com células especializadas, que se dividem por fissão binária); Stigonematales (filamentosas com mesmo grau de diferenciação celular que os membros da família Nostocales, e ainda são capazes de se dividir por fissão binária em múltiplos planos, com ramificações verdadeiras) (CASTENHOLZ, 2001; CAVALIER-SMITH, 2002; WATERBURY, 2006).

Avanços tecnológicos como, por exemplo, o sequenciamento de larga escala, permitiram uma rápida caracterização das comunidades de cianobactérias, sustentando cada vez mais a abordagem polifásica (informações genotípicas associadas às fenotípicas) na taxonomia de cianobactérias (PALINSKA, K.A.; SUROSZ, 2014).

O sistema de classificação vigente utiliza uma abordagem combinada com o fenótipo, composição ultraestrutural, métodos ecológicos, bioquímicos e moleculares com a aplicação de ‘iniciadores’ (primers específicos, pequenas sequências iniciadoras que codificam uma parte de interesse do DNA para posterior amplificação em PCR - reação em cadeia da polimerase) (URBACH et al., 1992; NÜBEL et al., 1997; LALLOUI et al., 2002; KOMÁREK et al., 2014). Komárek e colaboradores (2014) e Komárek (2016a; 2016b) propuseram 8 ordens com 45 famílias de cianobactérias (HALLENBECK, 2017), que estão devidamente descritas em anexo no final do documento (ANEXO 1). Com destaque para a ordem Synechococcales, que contem os gêneros mais abundantes dentro da cianobactérias. Esta grande ordem possui 11 famílias e ~70 gêneros, com representantes definidos como não monofiléticos, com tipos unicelulares e filamentosos. Este grupo está unido pela presença de tilacóides parietais. Com uma maior amostragem futura será possível revisar melhor este grupo. A família 2.1 Synechococcaceae (17 gêneros) ainda precisa de estudo e revisão em todos os níveis taxonômicos, devido à extensa complexidade. Apresentam formas coloniais e envelopes mucilaginosos, o gênero *Neosynechococcus* é uma linhagem independente recentemente descrita, morfológicamente semelhante ao gênero *Synechococcus*. O gênero *Synechococcus* é um dos mais amplos desta família com uma linhagem independente que ainda se apresenta indevidamente descrita (KOMÁREK et al., 2014; KOMÁREK, 2016a; 2016b). Outra família que necessita de uma extensa revisão é a 2.2 Prochloraceae com apenas

dois gêneros marinhos, caracterizada pela presença de clorofila-b. O gênero *Prochlorococcus*, é um dos mais importantes gêneros na produção primária dos oceanos, apresentando uma subdivisão em dois grupos ecologicamente distantes, os adaptados a alta luminosidade, e os adaptados a baixa luz. As relações taxonômicas deste último gênero com algumas espécies marinhas de *Synechococcus* sp. ainda não estão devidamente esclarecidas (KOMÁREK et al., 2014; KOMÁREK, 2016a; 2016b). As outras famílias deste grupo são: Merismopediaceae; Coelosphaeriaceae; Acaryochloridaceae; Chamaesiphonaceae; Romeriaceae; Pseudanabaenaceae; Leptolyngbyaceae; Heteroleibleiniaceae; Schizotrichaceae.

Um trabalho recente agrupou a análise ecológica com a abordagem genética baseada na distância genoma/genoma, propondo uma nova forma de classificação das cianobactérias fototróficas oxigenadas. Esta nova forma de classificação é denominada Taxonomia Genômica de Cianobactérias (CGT), que junta a taxonomia formal de cianobactérias acoplada a floresta de siglas existentes (OTUs) (WALTER, et al., 2017). Tal estratégia ecogenômica analisa a distribuição e abundância relativa de distintos nichos ecológicos e integra a taxonomia genômica. Como resultado, houve a distinção de três grandes grupos ecológicos de cianobactérias que foram coerentemente distribuídos pelos agrupamentos taxonômicos molecular. Considerando 99 genomas completos, estas foram distribuídas em: cianobactérias de baixa temperatura; cianobactérias copiotróficas (encontradas em ambientes ricos em nutrientes, particularmente em carbono, são o oposto das oligotróficas) de baixa temperatura; e cianobactérias oligotróficas de alta temperatura (WALTER, et al., 2017). Porém esta nova forma de agrupamento considerando a ecogenômica somente foi testada para bactérias que possuem genomas inteiros sequenciados, sendo ainda muito limitada.

### **1.3.2. Gêneros *Synechococcus* e *Prochlorococcus* (picofitoplâncton)**

Os gêneros cocóides *Synechococcus* e *Prochlorococcus* são os mais representativos dentre as cianobactérias para a produção primária em habitats naturais (ITURRIAGA & MITCHELL, 1986; PARTENSKY et al., 1999; KIRCHMAN, 2011; FIGUEIRAS et al., 2014). Um trabalho pioneiro na década de 1980 demonstra que organismos na fração de tamanho menor que 1µm (picoplâncton) são responsáveis pela fixação de 90% do <sup>14</sup>CO<sub>2</sub> nos primeiros metros da coluna da água no oceano Pacífico Tropical (LI et al., 1983). Uma análise de microscopia subsequente confirmou que uma significativa parcela dessa biomassa pertencia as cianobactérias cocóides, extrapolando que cerca de 25% da produção primária global é produzida por cianobactérias (LI et al., 1983; ITURRIAGA & MITCHELL, 1986; KIRCHMAN, 2011; LINDELL, 2014).

O gênero *Synechococcus* é rico em ficobiliproteínas, pigmentos facilmente visíveis quando excitados por microscopia de epifluorescência, sendo os primeiros a serem identificados devido a essa característica (PARTENSKY et al., 1999). As cianobactérias costumavam ser chamadas de algas azuis, e este termo cromático é muito preciso para o

gênero *Synechococcus*, uma vez que tal gênero apresenta o pigmento ficocianina que confere um tom azulado a célula. Este gênero também possui os pigmentos: clorofila-a (o de maior abundância na célula), ficobilina (coloração avermelhada) e ficoeritrina (alaranjada). Estes dois últimos pigmentos permitem que os organismos unicelulares flutuem em ambientes pouco iluminados e ocasionalmente superam as algas eucarióticas.

Ambos os gêneros discutidos neste tópico variam distintamente com a profundidade, apresentando migração vertical. *Synechococcus* geralmente é encontrado numa camada superior na coluna de água em relação à *Prochlorococcus*, embora os dois grupos de cianobactérias estejam adaptados à intensidade e qualidade da luz encontrada mais profundamente na coluna de água (HOFFMANN, 1999; LINDELL, 2014). A distribuição global de *Prochlorococcus* varia de 40°S a 40°N de latitude e as espécies podem ser encontradas em toda a zona eufótica. Em contraste, *Synechococcus* tem uma distribuição muito mais ampla que varia das águas do Ártico, até 70°N até 50°S, mas não é encontrada ao sul da frente polar (VELDHUIS et al., 2005).

Devido a ausência dos pigmentos acessórios (ficobiliproteínas) na maioria das espécies do gênero *Prochlorococcus*, estas somente foram descobertas posteriormente às *Synechococcus*, graças ao advento da técnica da citometria de fluxo (CHISHOLM et al., 1988), apresentando uma coloração esverdeada devido à presença majoritária de clorofila-a e clorofila-b, também chamada neste grupo em específico como divinil-clorofila (clorofila modificada com picos de absorção levemente superiores aos da clorofila de microalgas eucariotas e plantas superiores) (GOERICKE & REPETA, 1993; ITO & TANAKA, 2011). Tal gênero é atualmente um dos mais bem estudados microrganismos marinhos (CHISHOLM et al., 1988; PARTENSKY et al., 1999; HEYWOOD et al., 2006; JOHNSON et al., 2006; HALLENBECK, 2017). Até recentemente, a estrutura taxonômica do gênero *Prochlorococcus* estava mal definido, sendo que durante muitos anos apenas uma espécie havia sido formalmente descrita neste gênero, chamada *P. marinus* (CHISHOLM et al., 1988). Esta espécie foi descrita posteriormente com dois ecótipos, utilizando o nicho e os pigmentos como marcadores taxonômicos. Estes dois ecótipos habitam diferentes profundidades: um ecótipo de elevada luminosidade e uma baixa taxa clorofila-a / clorofila-b (chl-a / chl-b); e um ecótipo de baixa luminosidade e uma alta relação chl-a / chl-b. Como esperado, o ecótipo de baixa luz geralmente é encontrado mais profundo na coluna de água do que o ecótipo de alta luminosidade. Os genes do rRNA 16S isolados destes dois ecótipos diferem muito pouco com uma similaridade superior a 97%, apenas valores menores a essa porcentagem que podem ser separadas as espécies, por isso são chamados de ecótipos (ROCAP et al., 2002; 2003; KIRCHMAN, 2011; AHLGREN et al., 2006; BATMALLE et al., 2014). Até 2013 havia 13 genomas completos representativos de cepas cultivadas e mais 100 metagenomas marinhos (genes 16s rRNA) deste grupo, que foram investigados a fim de

determinar a estrutura taxonômica; sendo que *P. marinus* se enquadrou como espécie críptica a outra 10 espécies (THOMPSON et al., 2013a). Considerando os metagenomas e os genomas completos, estipula-se que podem haver um total de 35 espécies de *Prochlorococcus* nos oceanos do mundo, mas algumas espécies (por exemplo, *Prochlorococcus chisholmii* AS9601T) são muito abundantes, enquanto outras (por exemplo, *Prochlorococcus swingsii* MIT9313T) parecem estar restritas a menos locais (THOMPSON et al., 2013a). Thompson e colaboradores (2013) discutem que a incongruência entre a designação de ecótipo é dada a distintas populações dentro da mesma espécie, com sutis preferências de nicho. Enquanto a designação de espécie se refere a organismos que se diversificaram há muito tempo e, portanto, possuem distintos genomas e fenótipos (THOMPSON et al., 2013a).

Os dois gêneros são encontrados em abundância em oceanos abertos em baixas latitudes (água salgada), embora algumas espécies de *Synechococcus* também ocorram em lagos oligotróficos (águas continentais) (Tabela 4).

Tabela 4. Principais características dos gêneros cocóides (picofitoplâncton) de maior importância na produção primária nos giros oceânicos. Fonte: adaptado de Kirchman (2011).

Property	Two Major Coccoid Cyanobacterial Genera	
	<i>Synechococcus</i>	<i>Prochlorococcus</i>
Size (diameter)	1.5 - 1.0 $\mu\text{m}$	0.7 $\mu\text{m}$
Chlorophyll a	Yes	Yes (Modified)
Chlorophyll b	No	Yes
Phycobilins	Yes	Fill Strains
Habitat	Marine and freshwater	Marine
Distribution	Cosmopolitan	Oligotrophic Zones
N <sub>2</sub> Fixation	Fill Strains	No

### 1.3.3. Ecologia microbiana

O estudo da diversidade da microbiota pela aplicação das técnicas moleculares fundamenta a ecologia microbiana, esta ciência emergente tenta compreender as relações dos microrganismos e suas funções ecossistêmicas e metabólicas com o meio (forçantes ambientais), promovendo uma melhor descrição e compreensão das diferentes trocas energéticas e funcionais dos mais variados microbiomas (BARTON & NORTHUP, 2011). À medida que os métodos para determinar a diversidade molecular das comunidades microbianas se tornam mais acessíveis, aumenta o desafio dos ecologistas microbianos de vincular a estrutura das comunidades naturais com suas funções (SCHMIDT, 2006). A rápida acumulação de dados moleculares (sequências) está evidenciando uma diversidade ainda maior, além de novas funções e novos nichos (PROSSER et al., 2007). Devido à filogenia baseada no 16S rRNA, é possível atualmente estabelecer uma noção mais real da extensa diversidade de microrganismos que habitam o planeta, estimado recentemente que a Terra

está habitada por  $10^{11}$ - $10^{12}$  de espécies microbianas. Sendo que apenas  $\sim 10^4$  das espécies podem ser cultiváveis (LOCEY & LENNON, 2016).

A aplicação de análises estatísticas multivariadas tornou-se uma característica consistente na ecologia microbiana (BUTTIGIEG & RAMETTE, 2014). O agrupamento das unidades genômicas entre si (OTU clustering) e entre as variáveis ambientais demandam, obrigatoriamente, tais ferramentas estatísticas uni- e multivariadas (SINCLAIR et al., 2015). Existem atualmente inúmeros softwares, pacotes e plataformas para processamento de dados em bioinformática, a fim de estabelecer um script adequado para as investigações multivariadas (pipeline). As ferramentas atuais para o tratamento multivariado aplicado a bioinformática são: a plataforma MOTHUR, sendo a pioneira e ainda utilizada (SCHLOSS et al., 2009); a plataforma QIIME - Quantitative Insights Into Microbial Ecology (CAPORASO et al., 2010), é mais utilizada atualmente com o maior número de citações sendo específica para fragmentos 16S ou 18S rRNA (SINCLAIR et al., 2015); a plataforma UPARSE (EDGAR, 2013), é focada em genes amplificados e vem ganhando espaço no meio científico da bioinformática por ser a mais moderna e sensível (SINCLAIR et al., 2015); o laboratório virtual BIOVEL - Biodiversity Virtual e-Laboratory (HARDISTY et al., 2016); o laboratório virtual GUSTA ME - Guide to Statistical Analysis in Microbial Ecology (BUTTIGIEG & RAMETTE, 2014); e o pacote estatístico PHYLOSEQ para R scripts (MCMURDIE & HOLMES, 2014).

Embora todos os pipelines supracitados abordem análises multivariadas muito similares, não se pode ficar preso aos padrões estatísticos e ao fluxo de trabalho como se fossem 'receitas de bolo'. Para uma ideal conclusão dos dados, a bioinformática deve ser associada com extremo rigor às peculiaridades de cada ecossistema. Necessita-se de uma abordagem adequada caso-a-caso para os fenômenos investigados (BUTTIGIEG & RAMETTE, 2014; SINCLAIR et al., 2015).

Elementos importantes de pesquisa em ecologia da comunidade microbiana incluem, cada vez mais, a análise de vias funcionais associadas aos recursos nutricionais e fluxos de energia. Desse modo, é possível abordar uma compreensão mecanicista de interações entre populações microbianas o meio ambiente e os compostos bioativos emergentes, geralmente provenientes do metabolismo secundário em estratégias de redundância funcional e estabilidade do sistema (KONOPKA, 2009).

#### **1.3.4. Técnicas moleculares aplicadas à ecologia molecular**

A análise dos ecossistemas microbianos é atualmente um exercício taxonômico, que se utiliza de ferramentas experimentais para a obtenção de sequências de genes. A espécie é o elemento central da sistemática bacteriana, assim técnicas modernas se concentram em distinguir, de maneira eficiente, esta posição taxonômica.

Existem diversas técnicas de extração de DNA, sendo que ao final do processo todas isolam um peso muito rarefeito de macromoléculas de ácidos nucleicos, que são solvidas em um tampão de eluição (água ultrapura ou uma solução TRIS-HCl de baixa molaridade). Devido a esta baixa recuperação há necessidade de amplificação destas macromoléculas. A técnica de PCR (*Polymerase Chain Reaction*) possibilita milhões de cópias de sequências alvo-específicas e/ou aleatórias. Os métodos indiretos de análise destes fragmentos amplificados são chamados de *fingerprint* ou ribotipagem, e possuem um baixo poder de resolução aos perfis gerados pelos fragmentos. Este perfil é dado por uma imagem em um gel de eletroforese. São chamados de indiretos, pois analisam bandas estabelecendo uma composição de comunidade. Tais bandas, comumente, não passam pelo processo de identificação taxonômica (MUYZER et al., 1993). As técnicas dependentes de PCR mais comumente utilizadas nos estudos de diversidade microbiana são (VANDAMME et al., 1996): DGGE - Denaturing Gradient Gel Electrophoresis; RAPD - Random Amplified Polymorphic; T-RFLP - Terminal-Restriction Fragment Length Polymorfism; TGGE - Temperature Gradient Gel Electrophoresis; ARISA - Automated Ribosomal Intergenic Spacer Analysis; RISA - Ribosomal Intergenic Spacer Analysis; SSCP - Single-Strand Conformational Polymorphism; ASDRA - Amplified Ribosomal DNA Restriction Analysis (Figura 3).

A técnica da PCR-DGGE, especialmente, se tornou um meio rápido de se comparar e estudar a estrutura de comunidade sem o uso de metodologias de cultivo. Investigando aspectos básicos de comunidades (COLARES & MELO, 2013). Tal técnica consiste na eletroforese de fragmentos de genes do rRNA 16S amplificados previamente por PCR, onde cada um dos iniciadores apresenta uma região rica em G+C (grampo GC) com uma sequência 5'-CGCCCGCCGCGCGGGCGGGGCGGGGCACGGGGG-3' (MUYZER et al., 1993). Este grampo tem o objetivo de impedir a desnaturação total da dupla fita do DNA durante a eletroforese, em gel de poliacrilamida contendo um crescente gradiente de desnaturantes (uréia e formamida). Os compostos desnaturantes rompem as pontes de hidrogênio entre os nucleotídeos, possibilitando a separação de fragmentos do mesmo tamanho, mas com composição nucleotídica diferente. Portanto a DGGE é baseada na mobilidade eletroforética da molécula de DNA parcialmente desnaturada no gel de poliacrilamida, resultando em diferentes fragmentos de DNA migrando no gel em posições diferentes. Os diferentes padrões de bandas gerados são característicos de cada comunidade (MUYZER et al., 1993).

Outras estratégias baseiam-se no sequenciamento dos ácidos nucleicos ou fragmentos amplificados, analisando sequências tanto das bibliotecas de clones, quanto das bibliotecas genômicas e de 16S rRNA (Figura 3). O sequenciamento permite uma análise direta das sequências, tendo um maior poder de resolução (MACLEAN et al., 2009). O método de

Sanger foi pioneiro, envolvendo a extração de DNA total, amplificação por PCR de genes de interesse, clonagem e posterior sequenciamento.

Além da biblioteca de clones ser muito trabalhosa, ela não consegue abranger toda a diversidade presente em uma amostra. Esses fatores limitantes incentivaram o desenvolvimento de novas tecnologias de sequenciamento, chamadas de “nova geração”, as quais permitiram reduzir os custos, evitar a clonagem, promover a miniaturização das reações, alto rendimento, e uma análise em profundidade dos micro-organismos ou genes funcionais presentes nos ambientes (MACLEAN et al., 2009; SINCLAIR et al., 2015). Dentre as plataformas de sequenciamento chamadas de nova geração ou segunda geração, duas são as mais utilizadas em todo o mundo: plataforma 454 (pirosequenciamento) da Roche; e as plataformas Solexa, MiSeq e HiSeq da Illumina. Dentre as técnicas de sequenciamento de segunda geração, as plataformas Illumina se mostraram mais eficientes que a técnica do pirosequenciamento 454, apresentando melhores resultados de uniformidade e correspondência, mesmo com os protocolos de preparação de biblioteca muito semelhantes (SINCLAIR et al., 2015). Todas essas tecnologias de nova geração promovem o sequenciamento de DNA em plataformas capazes de gerar informação sobre milhões de pares de bases em uma única corrida (MACLEAN et al., 2009).

Atualmente, já existem os sequenciadores de terceira e quarta gerações com sequenciamento de moléculas individuais em tempo real, e análise da cinética da polimerase, como o: Ion Torrent da Life Technologies; Nanopore da GridIon/MiniIon; e SOLiD System da Applied Biosystems.

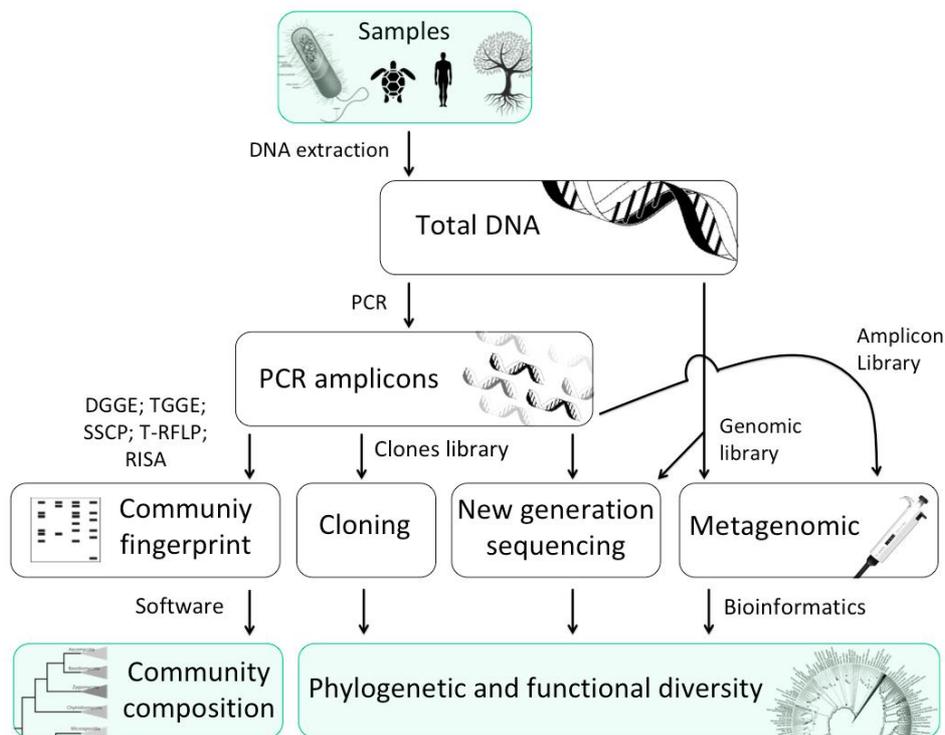


Figura 3: Esquema do processamento genômico da amostra desde a extração até as técnicas de purificação e isolamento dos genes que possibilitam o estudo da composição e filogenética microbiana. Análises que formam bibliotecas genômicas são mais específicas quanto a sequência exata de nucleotídeos nos fragmentos de DNA, podendo chegar a análises mais profundas como a diversidade filogenética e funcional. Análises de fingerprint são adequadas para se identificar a estrutura da comunidade microbiota, com um perfil de bandas específico a cada microbioma. Fonte: o próprio autor.

As técnicas citadas acima podem sequenciar bibliotecas de genomas inteiros e metagenomas. A metagenômica analisa a comunidade de microrganismos provenientes da extração do DNA total diretamente do ambiente e construção de uma biblioteca utilizando como substrato o genoma misto (DNA metagenômico) fragmentado aleatoriamente (shotgun). Sendo que grande parte destes genomas recuperados nem sempre estão descritos (OTUs). A obtenção de uma quantidade suficiente de DNA de alta qualidade, representando a comunidade microbiana inteira, é o primeiro passo na análise metagenômica de uma amostra ambiental (SINCLAIR et al., 2015). Estudos de diversidade taxonômica, diversidade funcional e de prospecção de genes e enzimas de interesse podem ser realizados em poucas horas devido à metagenômica associada aos métodos de sequenciamento de alto rendimento descritos acima (high throughput) (BEZERRA et al., 2015).

Tanto no sequenciamento pelo método de Sanger, quanto no sequenciamento das plataformas Illumina, ocorre a síntese de fragmentos a partir da DNA-polimerase e nucleotídeos terminadores marcados com diferentes fluoróforos. A inovação consiste numa superfície de clonagem *in vitro* dos fragmentos em uma plataforma sólida de vidro (flow cell).

Na plataforma MiSeq a flow cell possui uma linha, já na plataforma HiSeq possuem oito linhas que podem ser utilizadas para o sequenciamento de até oito bibliotecas.

O DNA é fragmentado aleatoriamente, sendo ligado a adaptadores em ambas as extremidades dos fragmentos. Em cada linha, os adaptadores são fixados à superfície pela extremidade 5', deixando a extremidade 3' livre para servir na iniciação da reação de sequenciamento dos fragmentos imobilizados no suporte por hibridização. As moléculas de DNA fita simples são aderidas por afinidade ao suporte sólido onde estão também aderidos, em alta densidade, oligonucleotídeos complementares aos adaptadores das extremidades do fragmento. Na etapa de anelamento, é formada uma estrutura em ponte, onde o adaptador da extremidade livre da molécula aderida ao suporte encontra seu oligonucleotídeo complementar também no suporte. Iniciando a PCR pela extremidade 3' livre do oligonucleotídeo com o iniciador. Na etapa de desnaturação, a estrutura de ponte é desfeita mediante elevação de temperatura. Então se repete a etapa de anelamento, formando novas estruturas em ponte e iniciando um novo ciclo de amplificação. Entre as etapas, ocorre uma lavagem para remoção dos reagentes excedentes e remoção do terminal 3' bloqueado e do fluoróforo do nucleotídeo incorporado no ciclo anterior (já lido por capturação de imagem) para que a reação de sequenciamento prossiga. Após uma série desses ciclos, são gerados clusters de moléculas idênticas (clones de fragmentos) ligadas ao suporte. Devido à incorporação de nucleotídeos terminadores marcados com diferentes fluoróforos, a excitação a laser gera um sinal, o qual é captado (imagem) e interpretado. O processo de incorporação de nucleotídeo marcado, excitação e leitura é repetido para cada nucleotídeo componente da sequência. Até 50 milhões de clusters podem ser produzidos por linha, correspondendo a uma representação satisfatória da biblioteca. A leitura é feita de forma sequencial, cerca de 25-35 são lidas em cada cluster, o que permite a montagem da sequência completa do cluster. (SHENDURE & JI, 2008; CARVALHO & SILVA, 2010; COLARES, 2014).

As plataformas de nova geração produzem um volume muito superior de dados de que o método de Sanger, porém os fragmentos são mais curtos, o que torna a montagem das sequências uma tarefa mais desafiadora (SHENDURE & JI, 2008). Este enorme volume de dados necessita de um robusto aparato de processamento. A bioinformática é a ciência que une as informações taxonômicas à montagem das sequências e assinala a taxonomia e ferramentas para análises de diversidade e estatística (análise dos dados), através de softwares e plataformas *in silico*. Os bancos de dados genômicos são fundamentais para a bioinformática e os agrupamentos filogenéticos adequados. Estes bancos podem ser particulares ou públicos. Uma colaboração internacional (INSDC - International Nucleotide Sequence Database Collaboration) une os três maiores e mais creditados bancos de dados genéticos públicos: o banco norte americano GenBank do NCBI (National Center for Biotechnology Information); banco japonês DDBJ (DNA DataBank of Japan); e o banco

inglês ENA (European Nucleotide Archive). Essas três organizações possuem dados de fragmentos (metagenomas) e genomas completos, ambos são conectados e trocam dados diariamente para todos os domínios da árvore da vida. Considerando apenas os microrganismos, os principais bancos de dados utilizados pelos programas de bioinformática são: GreenGenes - banco de dados 16S rRNA e de sequências quiméricas (DESANTIS et al., 2006); SILVA – banco de dados de genes ribossomais 16S e 18S (QUAST et al., 2012); OM-RGC - Ocean Microbial Reference Gene Catalog, banco gerado a partir do projeto Tara Oceans para fragmentos de rRNA apenas encontradas em amostras de água salgada (SUNAGAWA et al., 2015); MBGD - banco de dados de microrganismos (UCHIYAMA et al., 2015). Existem também bancos de proteínas, os quais armazenam as sequências de DNA e RNA que codificam os aminoácidos que geram as cadeias proteicas.

#### **1.4. Diversidade e biomassa do picofitoplâncton na zona equatorial**

Em ecologia uma amostragem garante a coleta de um número de organismos que representam diferentes espécies em um determinado local (área ou volume) (ODUM, 2007). O número de indivíduos coletados (processados) para cada espécie indica a abundância. A partir do número de espécies encontradas em uma amostra pode-se estimar a riqueza microbiana (riqueza observada). A contribuição de diferentes amostras implica num aumento do número de indivíduos (e espécies) permitindo gerar curvas de acumulação de espécies (riqueza estimada). Estas curvas permitem avaliar o quanto um estudo se aproximou de amostrar ou identificar todas as espécies do local.

O método da rarefação é o mais adequado para calcular o número esperado de espécies de cada amostra, pois permite a comparação de amostras com esforços e tamanhos amostrais diferentes, sendo a base para os índices de diversidade sem a necessidade de acessar todos os indivíduos de uma amostra. A estabilização da curva significa que nenhuma espécie nova é adicionada, ou seja, a riqueza total foi obtida. Curvas com menor inclinação da reta ( $<\alpha$ ) representam maior cobertura da diversidade, em relação a curvas de maior inclinação ( $>\alpha$ ) que apresentam uma diversidade mais vasta que ainda não foi toda representada (MAGURRAN, 1988; ODUM, 2007). Em regiões tropicais a estabilização da curva é bastante difícil, pois muitas espécies raras são constantemente amostradas (MARTINY et al., 2006), de modo que a curva tende a seguir aumentando, demorando para atingir um platô no qual se estabilize. A rarefação é expressa pelos índices de estimativa de riqueza. Tais medidas possibilitam estimar o número total de espécies numa determinada comunidade, a partir dos dados amostrais distintos. Entre estas estão os estimadores Jackknife (1 e 2), Chao (1 e 2) e Bootstrap. O método Chao 1 estima a riqueza total utilizando o número de espécies representadas por apenas um indivíduo nas amostras (singletons), e o número de espécies com apenas dois indivíduos nas amostras (doubletons). O Chao 2 é uma adaptação que utiliza os

uniques e os duplicates, entre todas as amostras tomadas na comunidade (MAGURRAN, 1988; ODUM, 2007).

Diferentemente da riqueza, que refere-se apenas à presença/ausência de espécies numa amostra, o termo biodiversidade relaciona a riqueza e a abundância relativa de espécies. A diversidade pode ser dividida em duas formas gerais: a diversidade em um ponto ( $\alpha$ -diversidade); e a diversidade ao longo de gradientes ( $\beta$ -diversidade), sendo que cada uma destas formas de avaliar a diversidade possui seus índices próprios, seja de riqueza ou agrupamento (MAGURRAN, 1988; ODUM, 2007). Os índices de diversidade se usam da riqueza, abundância e da igualdade de um conjunto complexo de espécies nos diferentes ambientes estudados, sendo gerado um valor correspondente a cada amostra ou ambiente.

A  $\alpha$ -div, ou diversidade local, corresponde ao número de espécies numa pequena área de habitat homogêneo. A alfa diversidade é a medida de diversidade mais utilizada em ecologia, considerando a uniformidade (equitabilidade) e a riqueza de espécies. O aumento do número de espécies ou o aumento da uniformidade das abundâncias aumenta a diversidade. Existem diversas formas de avaliar a  $\alpha$ -diversidade, sendo que os índices de Shannon-Wiener (H) e Simpson (S) são os mais clássicos para as assembleias microbianas (MARTINY et al., 2006). O índice de Shannon valoriza a importância relativa de cada espécie atribuindo um maior peso a espécies raras, o componente de riqueza de espécies é mais ponderado. Numericamente, o índice de Shannon-Wiener varia entre 0 e valor máximo qualquer. O índice de Shannon-Wiener será igual a zero somente quando houver uma única espécie na amostra e assumirá seu valor máximo, somente quando todas as espécies existentes na amostra apresentarem o mesmo número de indivíduos. Na prática, os valores assumidos pelo índice de Shannon-Wiener situam-se entre 1,5 e 3,5 e só raramente ultrapassam o valor de 4,5 (MAGURRAN, 1988). Já o índice S pondera mais a concentração de dominância, ou seja, a abundância das espécies mais comuns (ou de maior biomassa). Quanto maior o valor assumido pelo índice, menor é a diversidade de espécies estimada para a amostra (MAGURRAN, 1988).

A  $\beta$ -div revela a heterogeneidade da estrutura da comunidade, ou seja, o quanto as comunidades são distintas (ou similares) em termos de composição de espécies. Os índices de  $\beta$ -div utilizam a composição e as abundâncias relativas das diferentes espécies. Quanto menos espécies as diferentes comunidades compartilham, mais alta é a diversidade  $\beta$ . A maneira mais simples de medir a diversidade  $\beta$  entre pares de locais é pelo uso dos coeficientes de similaridade. Estes coeficientes comparam comunidades de forma qualitativa ou quantitativa. Os índices qualitativos, ou coeficientes binários, mais utilizados são o índice de Jaccard (Jac) e Sorensen (Sor). No caso de um gradientes sendo comparados simultaneamente, uma boa representação da diversidade  $\beta$  pode ser obtida através das análises de agrupamento (cluster) (MAGURRAN, 1988; ODUM, 2007). O índice Jac indica a proporção de espécies

compartilhadas entre amostras em relação ao total de espécies. Compara qualitativamente (presença/ausência) a semelhança de espécies que existe entre amostras sucessivas ao longo de um gradiente ambiental (temporal ou espacial). Quantitativamente, varia entre 0 (comunidades totalmente diferentes quanto à composição de espécies) e 1 (totalmente semelhantes). O índice Sor apresenta as mesmas propriedades do índice de Jaccard. Sendo tão similares que Sor também varia entre 0 (semelhança nula) e 1 (semelhança máxima) (MAGURRAN, 1988; ODUM, 2007).

Novas abordagens técnicas, como a amplificação do gene rRNA com primers inclinados para certos grupos, a construção de bibliotecas de genes, a contagem de células classificadas por citometria de fluxo com base na clorofila e fluorescência ajudam a explorar a extensão real da diversidade fotossintética do nano- picoplâncton (VAULOT et al., 2008). Embora uma classificação tradicional convencie 2mm o diâmetro celular máximo do picoplâncton (SIEBURTH et al., 1978 - Figura 2), esse limiar está sendo ampliado para 3mm em estudos de campo operacionais, devido as técnicas de filtração e inclusão dos eucariotos (VELDHUIS et al., 2005; VAULOT et al., 2008). Presume-se que estas menores frações do fitoplâncton, possuem uma vantagem ecológica de distribuição devido à grande relação superfície/volume, em termos de um menor requerimento de nutrientes (ferro) ou uma absorção mais eficiente do ferro em relação as espécies maiores (VELDHUIS et al., 2005).

O picofitoplâncton está presente em todos os sistemas oceânicos, de polo a polo, dominando em áreas de baixa biomassa de clorofila, como as regiões tropicais, mas também contribuem consideravelmente (até 20%) nas áreas de alta biomassa de clorofila (VELDHUIS et al., 2005). Enquanto a fração eucariótica (18S rRNA) é muito mais diversificada, o componente bacteriano (16S rRNA) do picofitoplâncton é dominado por dois gêneros, *Prochlorococcus* e *Synechococcus*. Estes dois gêneros são os mais abundantes e os menores componentes conhecidos das comunidades fitoplanctônicas, sendo a espécie *Prochlorococcus marinus* (SAR11) a bactéria mais abundante na Terra (LOCEY & LENNON, 2016).

Em termos de biomassa, o componente bacteriano da comunidade picofitoplancônica excede a dos eucariotos nos oceanos. Dentro das bactérias, a abundância numérica de *Prochlorococcus* excede a de *Synechococcus* em pelo menos uma ordem de grandeza (VELDHUIS et al., 2005). Os ecossistemas de ocorrência podem variar drasticamente em composição química, incluindo águas altamente misturadas e repletas de nutrientes, bem como uma zona eufótica estratificada empobrecida em nutrientes na superfície. Essas mudanças impõem uma forte pressão na composição da comunidade com distintas propriedades bioquímicas e fisiológicas das espécies presentes, em particular: a exigência de pigmentação e nutrientes (carbono, nitrogênio e fósforo); e a presença de uma coluna da água estratificada ou uma lâmina bem misturada (VELDHUIS et al., 2005). Sua principal distribuição está no oceano aberto (oligotrófico) onde dominam, principalmente em termos de

biomassa com média de  $1 \times 10^5$  células por mL (VELDHUIS et al., 2005). Os picoprocariontes dominam não apenas nas regiões tropicais, mas também nas regiões temperadas. Estudos de longo prazo mostram que no Atlântico Subtropical (zonas temperadas) os maiores números de *Synechococcus* são encontrados na primavera e *Prochlorococcus* no verão e no outono (VELDHUIS et al., 2005). *Prochlorococcus* podem representar de 21 a 43% da biomassa fotossintética em oceanos oligotróficos e 13 a 48% da produção primária líquida, podendo alcançar  $3 \times 10^5$  células por mL (JOHNSON et al., 2006). Os sistemas oceânicos abertos são deficientes em nutrientes, especialmente nitrogênio e fósforo, porém ainda são produtivos, este paradoxo pode ser explicado por uma efetiva ciclagem de nutrientes, pois novos nutrientes são gerados (ou reciclados) por meio da alça microbiana e remineralização da matéria orgânica (AZAM et al., 1982). A atividade da comunidade microbiota bacteriana é determinante para a transformação de nutrientes nestes ecossistemas. Na zona eufótica de oceanos abertos oligotróficos a biomassa de carbono das bactérias é comumente 2-3 vezes maior de que eucariotos, sendo que o carbono bacteriano compreende uma porção substancial (cerca de 40%) de carbono orgânico particulado ao longo da zona fótica (LI et al., 1992). Um transecto meridional das comunidades do picoplâncton no oceano Atlântico, em cinco províncias: temperado do norte, giro do Atlântico Norte, equatorial, giro do Atlântico Sul e temperado do sul. Sendo que *Prochlorococcus* foi o grupo fototrófico mais numeroso nas águas dos giros do norte e do sul e na zona equatorial, em concentrações superiores a  $1 \times 10^5$  células por mL, e com estoques mais elevados na primavera do que no outono nas águas dos dois giros (JOHNSON et al., 2006). Em águas temperadas os números e a biomassa de *Synechococcus* aumentaram, sendo essa espécie mais numerosa na primavera do que no outono. Os picos de biomassa foram um pouco deslocados ao Atlântico Sul, sendo maiores entre abril-maio em comparação com setembro-outubro (ZUBKOV et al., 2000).

Os processos ecológicos da fração microbiana na rede trófica podem ser de semanas, ou tão curto quanto em dias, o que significa que grandes oscilações de metabolismo e composição podem ocorrer em uma questão de horas (LALLI & PARSONS, 1997). Desta forma o inventário da diversidade microbiana é fundamental para entender processos biogeoquímicos e da ecologia microbiana nos oceanos. Apesar do papel crucial das menores frações do fitoplâncton, como as cianobactérias, ainda há um número limitado de estudos realizados até à data sobre a diversidade (HEYWOOD et al., 2006; THOMPSON et al., 2011b). Os trabalhos existentes são recentes e correspondem a transectos de amostragem em porções do oceano pacífico (CHOI et al., 2011; SATO et al., 2015), nas áreas mesoceânicas do Atlântico (ZWIRGLMAIER et al., 2008) e norte do oceano Índico (JYOTHIBABU et al., 2013). Informações sobre a diversidade das cianobactérias em costa e porções insulares ainda são muito precoces e recentes, ainda mais quando enquadradas a realidade brasileira (THOMPSON et al., 2011b; JUNIOR et al., 2015).

## 1.5. Fotossíntese

O processo de fotossíntese é extremamente complexo, envolvendo um aparato enzimático muito vasto e específico, mesmo nas menores frações fitoplanctônicas, como as cianobactérias.

Nos organismos eucariotos o processo de fotossíntese ocorre no cloroplasto, organela que contém os pigmentos e as enzimas fotossintéticas, já nas cianobactérias estes pigmentos não estão enovelados por uma organela específica e sim em envaginações (pregas) da membrana celular chamadas de lamelas (Figura 4) (MOORE et al., 1995; WHITTON & POTTS, 2000). No cloroplasto as lamelas se encapsulam recebendo o nome de tilacóides, portanto possuem uma membrana e uma parte interior chamada lúmen. O cloroplasto possui um sistema de membranas que o delimita como organela celular e assim uma parte interna chamada estroma, no qual está distribuído todo o aparato de tilacóides (MOORE et al., 1995; ANDERSEN, 2005; REYNOLDS, 2006).

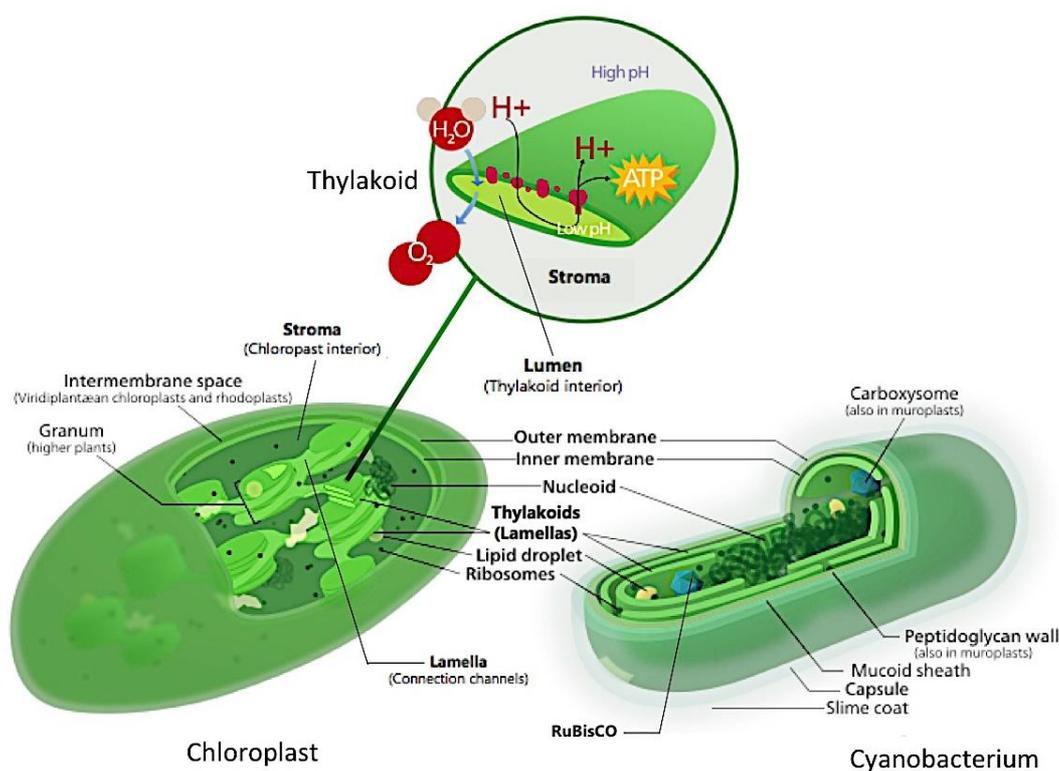


Figura 4. As estruturas de uma cianobactéria e de um cloroplasto são muito próximas. De acordo com a teoria da endossimbiose, os cloroplastos atualmente encontrados em células eucarióticas fotossintetizantes são provenientes de simbioses entre células fagocitárias nucleadas com cianobactérias (MARGULIS, 1981; PACE, 1997). Os tilacóides possuem as mesmas funções em ambos os grupos, responsáveis por captar a luz solar e transformá-la em energia química (ATP) através da hidrólise da água, sendo liberado oxigênio molecular ao final da reação (aceptor de elétrons) estabilizando o ATP e finalizando a fase clara da fotossíntese. A fase escura ou química é realizada no estroma (BARSANTI & GUALTIERI, 2006; KIRCHMAN, 2011). Fonte: adaptado de Ducat et al. (2011).

A energia obtida da luz solar é primeiramente absorvida pelas microalgas no complexo antena, este se encontra na parte externa da membrana do tilacóide e contém os pigmentos fotorreceptores. A energia do impulso luminoso absorvido por tais pigmentos é irradiada por ressonância para o centro da reação do complexo antena (Figura 5), sendo que o modelo de fluorescência sugeriria que cada fóton transferido tivesse um comprimento de onda mais longo e menor energia quântica pois um pouco é perdida na forma de calor. A clorofila-a consegue receber essa energia final porque consegue absorver fótons com comprimentos de onda mais longos de que outros pigmentos. Dois tipos de centros de reação com clorofila-a foram identificados e estão associados a dois complexos de proteínas identificados como fotossistema II (PS-II) e fotossistema I (PS-I) (MOORE et al., 1995; WHITTON & POTTS, 2000; KIRCHMAN, 2011).

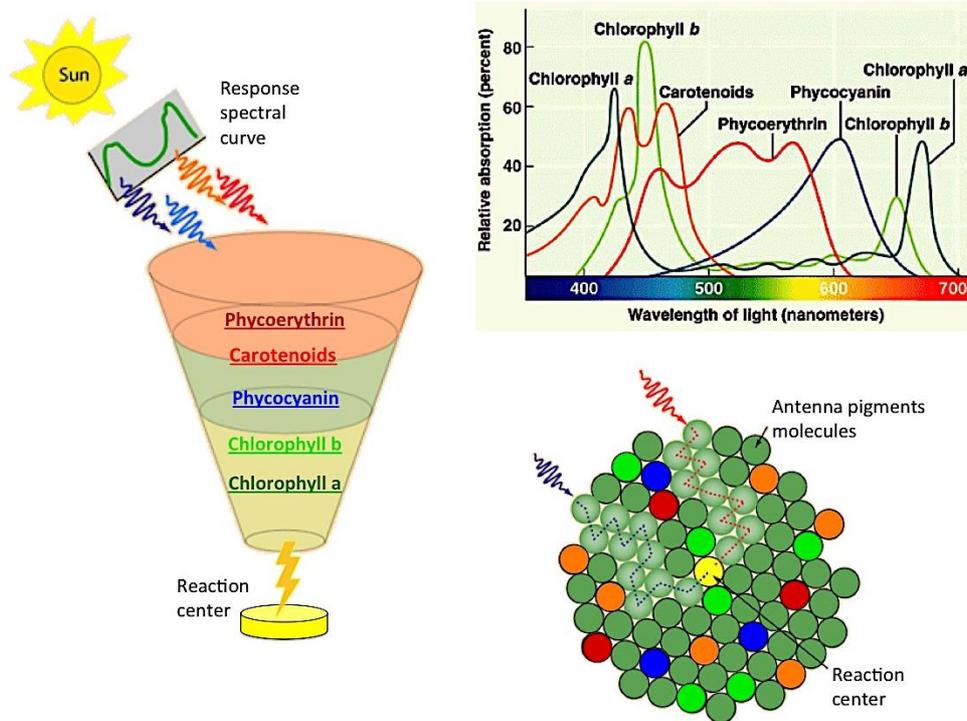


Figura 5. As clorofilas-a e -b possuem dois picos de absorção no espectro do visível, um na região do azul-violeta (430nm) e outro no vermelho (665nm), por justamente absorver os comprimentos de onda azul e vermelho, o cloroplasto tende a ser representado pela cor esverdeada. No complexo antena, ocorrem as primeiras reações da fotossíntese, com a absorção do fóton luminoso pelos pigmentos mais externos do complexo a assim transferido é transferido aos outros pigmentos mais centrais por ressonância (se comportam como uma antena de captação de fótons), excitando-os até o centro de reação, o qual contém uma clorofila-a central. No centro de reação a energia é usada para transferir um elétron energético para um aceptor de elétrons. Cerca de 2500 moléculas de clorofila são necessárias para produzir uma molécula de oxigênio, e um mínimo de oito fótons de luz deve ser absorvido no processo. Um modelo representa cerca de 300 moléculas de clorofila e 40 ou mais de carotenóides e outros pigmentos acessórios (FALKOWSKI et al., 1981; MOORE et al., 1995; HUOT & BABIN, 2010; KIRCHMAN, 2011). Fonte: adaptado de Moore et al. (1995) e Lalli & Parsons (1997).

Diversos pigmentos são descritos atualmente graças a modernas técnicas de cromatografia de alta performance (HPLC), como exemplo: carotenóides (pigmento amarelo-laranja), ficoeritrina (pigmento vermelho), ficocianina (pigmento azul), clorofila-b (verde), em especial a clorofila-a (verde), encontrada em todos os grupos fitoplanctônicos (MACKEY et al., 1996; HUOT & BABIN, 2010; KIRCHMAN, 2011). Estes outros pigmentos acessórios tem a função de capturar a maior quantidade de fótons possível, mesmo em uma baixa luminosidade. Os pigmentos ficocianina e ficoeritrina (pigmento vermelho) que absorvem a luz laranja e verde, respectivamente, estendendo a gama do espectro para a fotossíntese nas cianobactérias e algumas microalgas.

A fase clara da fotossíntese aeróbica corresponde aos aparatos enzimáticos dos fotossistemas II e I (PS-II e PS-I, respectivamente), responsáveis pela captação da energia solar pelos pigmentos do complexo antena, localizado na membrana externa dos tilacóides, até o centro de reação. As clorofilas-a e -b possuem dois picos de absorção no espectro do visível, um na região do azul-violeta (430nm) e outro no vermelho (665nm), por justamente absorver os comprimentos de onda azul e vermelho. No centro de reação a energia é usada para transferir um elétron energético para um aceptor de elétrons. Cerca de 2500 moléculas de clorofila são necessárias para produzir uma molécula de oxigênio, e um mínimo de oito fótons de luz deve ser absorvido no processo. Um modelo ideal do complexo antena representa cerca de 300 moléculas de clorofila e 40 ou mais de carotenóides e outros pigmentos acessórios (FALKOWSKI et al., 1981; MOORE et al., 1995; HUOT & BABIN, 2010; KIRCHMAN, 2011).

No PS-II, o centro de reação é chamado P680, sendo o “P” de pigmento e o número 680 sua absorção máxima aos fótons provenientes do complexo antena. Na presença do fóton ocorre a excitação do centro de reação. Uma vez excitado, o P680<sup>+</sup> usa o elemento Mg da Chl-a para acoplar a molécula da água e romper suas pontes de hidrogênio (foto-fosforilação da água), sendo reduzido por um elétron proveniente da água voltando ao estado estacionário P680. Todo esse processo ocorre na membrana externa do tilacóide, liberando o O<sub>2</sub> e prótons de H<sup>+</sup> para o estroma do cloroplasto enquanto que os elétrons são capturados pelo aceptor primário, uma molécula de feofitina localizada no interior do PS-II próximo ao de P680. Os elétrons formam um gradiente eletroquímico, seguindo uma via eletroquímica chamada de cadeia transportadora de elétrons. O transporte segue pelas plastoquinonas Q<sub>a</sub> e Q<sub>b</sub>, complexo citocromo b<sub>6</sub>-f a plastocianina (Pc), como consequência dessa transferência os prótons de H<sup>+</sup> são enviados para o lúmen do tilacóide. A plastocianina transferirá os elétrons para o centro de reação do PS-I (YOUVAN & MARRS, 1987; NITSCKE & RUTHERFORD, 1991; KIRCHMAN, 2011). As reações que ocorrem associados ao complexo b<sub>6</sub>-f acabam tornando-o uma bomba de prótons, transpondo os H<sup>+</sup> fornecidos pela cadeia de transporte de elétrons do estroma para o interior do tilacóide (baixo pH) (Figura 6). Desta forma é gerado

um potencial transmembranar utilizado pelo complexo da ATP-ase para fazer o fluxo contrário, usando o gradiente eletroquímico para enviar os prótons do lúmen para o estroma do cloroplasto. Neste processo ocorre a síntese do ATP (foto-fosforilação), a ATP-ase utiliza os prótons de  $H^+$  para complexar uma molécula de  $P_i$  a molécula de ADP, que será utilizada no ciclo de Calvin, fase escura da fotossíntese (NITSCKE & RUTHERFORD, 1991; KIRCHMAN, 2011).

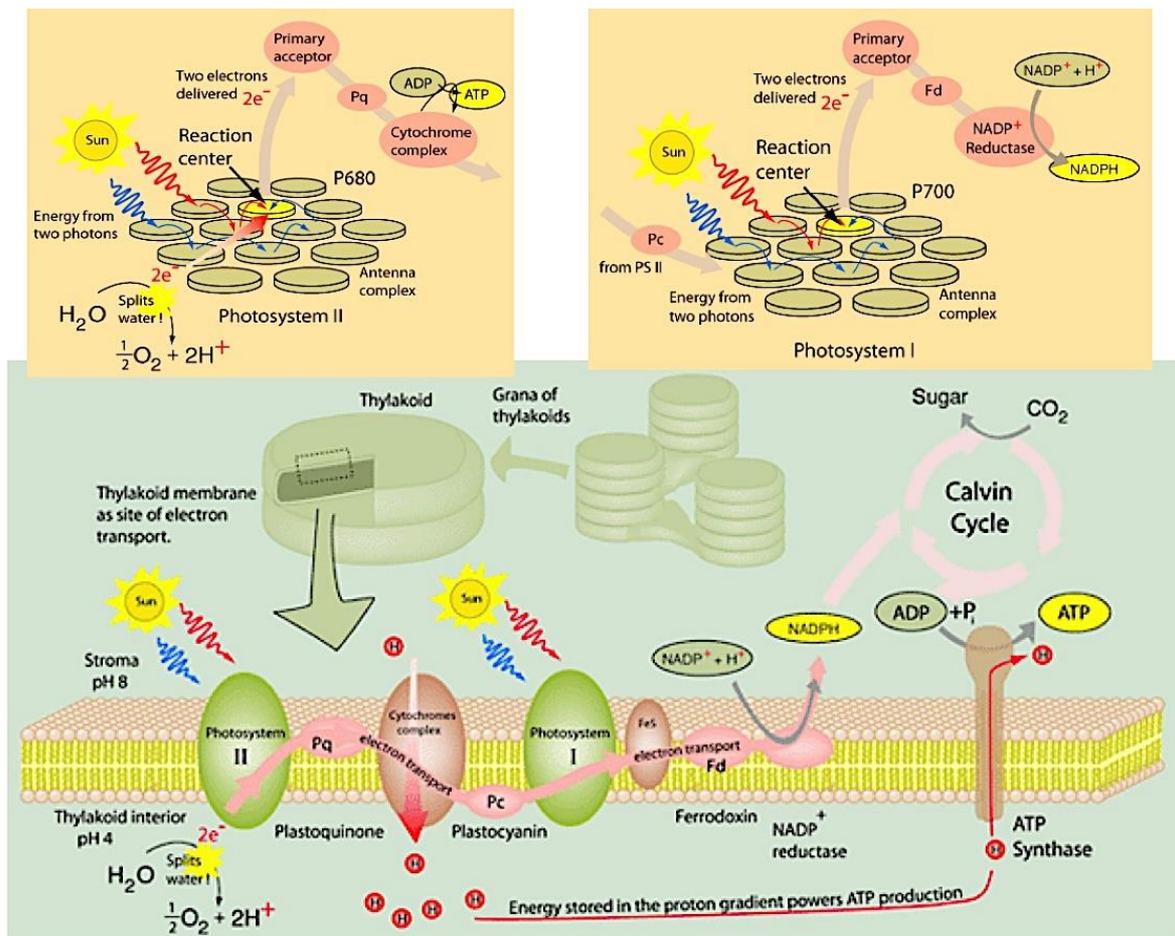


Figura 6. O fotossistema I é mais complexo do que o fotossistema II, porque tem um sistema de antena mais complexo, mais subunidades na sua estrutura geral e pode exibir transferência de elétrons cíclicos ou não cíclicos. O esquema acima representa a via acíclica do transporte de elétrons gerados a partir hidrólise da água, desta forma o PS-II oxida água e o PS-I reduz o  $NADP^+$  à NADPH, constituindo os elementos necessários para dar continuidade à fase escura (ciclo de Calvin) e gerar glicose. Em alguns momentos, quando o cloroplasto está em desequilíbrio entre as quantidades de NADPH e ATP, ocorre a via cíclica apenas com a participação do PS-I. Nesse caso, ocorre uma inversão no transporte de elétrons, após excitado pela energia solar o PS-I redireciona seus elétrons para o complexo citocromo  $b_6-f$  promovendo a entrada de prótons no lúmen do tilacóide e gerando ATP pela ATP-ase. Portanto, na fase cíclica ocorre a produção de ATP, mas não de NADPH e nem de  $O_2$  (o PS-II não participa) inviabilizando a produção de glicose no ciclo de Calvin. O centro de reação P680 oxidado ( $P680^+$ ) do PS-II é o agente oxidante biológico mais forte conhecido na natureza, isto torna possível oxidar a água (molécula altamente estável):  $H_2O \rightarrow 2H^+ + \frac{1}{2}O_2 + 2e^-$ . Em contraste, no PS-I o P700 quando excitado ( $P700^+$ ) se torna o agente redutor biológico mais forte (NITSCKE & RUTHERFORD, 1991; MOORE et al., 1995). Fonte: adaptado de Moore et al. (1995).

No fotossistema I a cadeia de transporte de elétrons completa-se com a transferência de  $e^-$  para o receptor final da fase clara da fotossíntese, o  $NADP^+$ , que é utilizado no ciclo de Calvin. O centro de reação do PS-I é chamado de P700, possui um maior limite de absorção máxima em relação ao P680, atinge picos a 700nm. O fotossistema I, assim como o PS-II, é excitado através dos pigmentos antena, que transferem a energia por ressonância até o centro de reação, resultando no P700+. O transporte segue pelo pigmento A0, filoquinona (A1), proteínas ferro-sulfuradas, e finalmente chega a ferredoxina (Fd), reduzindo-a. A Fd excitada se associa com o complexo  $NADP^+$ -redutase que protagoniza a função de reduzir o  $NADP^+$  em NADPH, que será utilizada no ciclo de Calvin, fase escura da fotossíntese. Na rota descrita acima (cíclica), o PS-II oxida água e o PS-I reduz o  $NADP^+$  (Figura 7) (NITSCKE & RUTHERFORD, 1991; KIRCHMAN, 2011).

A fase escura da fotossíntese é assim chamada porque não requer energia solar, embora ocorra também durante o dia. Nessa fase ocorre a síntese de carboidratos (ex. glicose - molécula orgânica estável utilizada como fonte de energia inicial na respiração celular) em um complexo ciclo de reações, chamado ciclo das pentoses ou ciclo de Calvin. De maneira geral, moléculas de  $CO_2$  unem-se umas às outras formando cadeias carbônicas que levam à produção de carboidratos, sendo que a energia necessária para o estabelecimento das ligações químicas é proveniente do ATP, e os hidrogênio do NADPH são obtidos pela fase clara (PS-II e -I). O ciclo de Calvin está dividido em duas partes, sendo que na primeira ocorre a fixação do  $CO_2$  pelo composto ribulose-1,5-bifosfato (RuBP), mediante ação da enzima ribulose-1,5-difosfato carboxilase/oxigenase (RuBisCo), esta etapa culmina com a formação de glicose; na segunda parte, ocorre a regeneração da RuBP (NITSCKE & RUTHERFORD, 1991; MOORE et al., 1995).

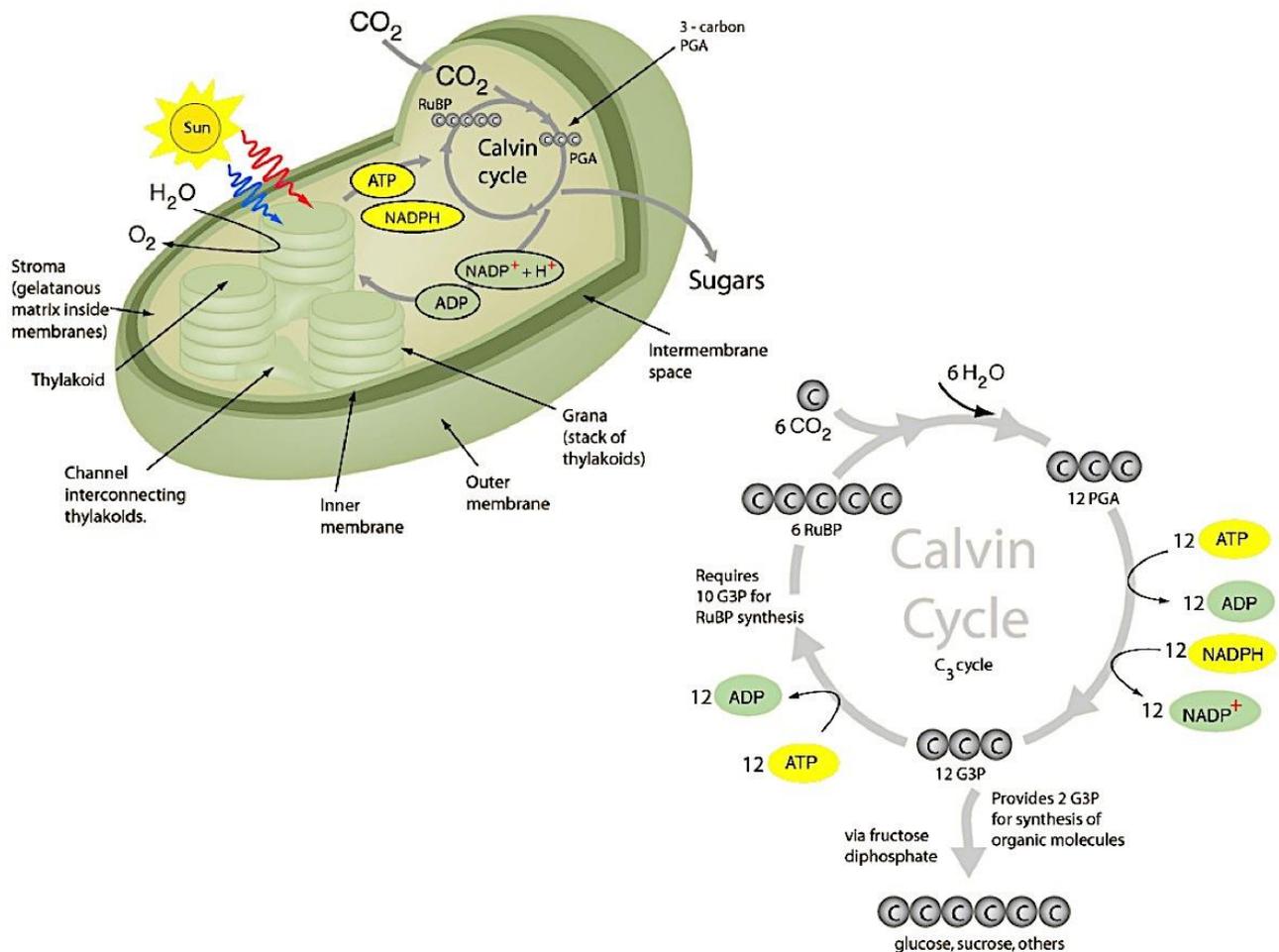


Figura 7. O esquema apresentado é uma simplificação do ciclo de Calvin. Considerando que o ciclo necessita ser mantido por moléculas produzidas por ele mesmo e ainda libere glicose para o uso celular, é necessário admitir a reação de seis moléculas de CO<sub>2</sub>, portanto, sendo consumidas 12 moléculas de ATP e 12 de NADPH, produzindo 12 moléculas de gliceraldeído-3-fosfato (G3P). Destas 12 moléculas, 2 vão formar uma glicose e as restantes 10 voltam para regenerar 6 moléculas de RuBP com o gasto adicional de mais 12 moléculas de ATP. O RuBisCo (ribulose-1,5-difosfato carboxilase/oxigenase) é a enzima mais importante e mais abundante na natureza, pois ela é a responsável da produção de toda a biomassa no planeta a partir de CO<sub>2</sub>. Calcula-se que existem 40 milhões de toneladas da enzima RuBisCo na biosfera (NITSCKE & RUTHERFORD, 1991; MOORE et al., 1995). Fonte: adaptado de Moore et al. (1995).

Em resumo, a fotossíntese oxigênica usa a energia luminosa captada pelos pigmentos acessórios sendo transferida para a clorofila-a no centro de reação onde a água é quebrada para sintetizar ATP e NADPH, liberando oxigênio molecular no processo de fase clara. O ATP e o NADPH são então utilizados para fixar CO<sub>2</sub> e sintetizar material orgânico (glicose) na reação escura pelo ciclo de Calvin ou Calvin-Benson-Bassham (CBB). Incluindo as reações claras e escuras, pode-se escrever a reação global da fotossíntese assim (KIRCHMAN, 2011):



### 1.5.1. Produção primária

A taxa de produção primária (produtividade primária) é o parâmetro mais importante para descrever um ecossistema a fim de compreender seus processos microbianos e biogeoquímicos. Nos ecossistemas ligados à presença de luz, tal produção pode ser definida pela equação apresentada no tópico acima. Pois estima a produção, e mensura o movimento dos elementos C e O, e as mudanças nas concentrações de O<sub>2</sub> e CO<sub>2</sub> ou CH<sub>2</sub>O. A produção primária nos oceanos está associada com a parte lábil do biogeo ciclo do carbono. O carbono total movimentado/remineralizado ao longo do processo de fotossíntese (produção primária bruta) seria a somatória: (1) carbono inorgânico liberado pela respiração (CO<sub>2</sub>); (2) ao carbono orgânico dissolvido liberado no meio pela exsudação celular (Dissolved Organic Carbon - DOC); (3) as partículas (macromoléculas) de carbono incorporadas como biomassa para o crescimento/duplicação celular, ou seja, a produção primária líquida (Particulate Organic Carbon - POC). A POC estará disponível como forma de energia para o consumidor secundário, no caso do ambiente aquático o pastoreio do fito pelo zooplâncton (LALLI & PARSONS, 1997; KIRCHMAN, 2011). O carbono orgânico total (Total Organic Carbon - TOC) é fracionado entre o POC e pelo DOC (<45µm). No ambiente marinho o POC é composto, majoritariamente, pela biomassa celular, e também por compostos (restos celulares, macromoléculas livres, compostos orgânicos, etc.) que estão em suspensão na coluna da água.

Uma das primeiras abordagens para calcular a produtividade primária nos ecossistemas aquáticos foi pelo "método de frasco claro-escuro", no qual as mudanças nas concentrações de O<sub>2</sub> ao longo do tempo em garrafas claras e escuras eram medidas encontrando o ponto de compensação. Em resumo, a produção primária nos oceanos é dada pela produção de carbono (biomassa), que orientada por luz é baseada na fotossíntese. A produtividade primária também pode ser medida através dos isótopos estáveis de <sup>14</sup>C e <sup>18</sup>C (LALLI & PARSONS, 1997; KIRCHMAN, 2011). A avaliação da fluorescência da clorofila por meio de fluorímetro com pulso de amplitude modulada (Phyto-PAM) é uma ferramenta atual, muito refinada e precisa para avaliar a produtividade das comunidades fitoplanctônicas (BERNARDI et al., 2016).

A produção líquida pode ser negativa em alguns curtos períodos na coluna da água ou em regiões com rápido e denso incremento de carbono proveniente de águas mais produtivas (KIRCHMAN, 2011). A respiração e decomposição excedem a produção bruta. Estes eventos negativos ocorrem em grandes regiões dos oceanos (giros), sendo necessários mais estudos para compreender tais taxas de troca de carbono entre oceano-atmosfera-picofitoplâncton. A queima de combustíveis fósseis leva a um ligeiro declínio no oxigênio atmosférico, e um aumento amplamente conhecido do CO<sub>2</sub> atmosférico. Uma produção líquida positiva leva ao aumento da pressão parcial de CO<sub>2</sub> (pCO<sub>2</sub>) na água em relação a atmosfera. Muitas regiões

dos oceanos são uma rede sumidouro de CO<sub>2</sub> atmosférico, cerca de metade de todos os CO<sub>2</sub> liberado pela queima de combustível fóssil termina nos oceanos (HOUGHTON, 2007).

Os oceanos abertos oligotróficos nas regiões tropicais recebem intensa radiação solar levando à estratificação termohalina da coluna d'água, as águas superficiais apresentam grande diferença de densidade em relação às águas profundas (picnoclina). Os detritos orgânicos da camada superficial ao afundarem para a camada abaixo da picnoclina, tendem a não retornarem à superfície. As taxas de produção primária em ecossistemas aquáticos tropicais podem rivalizar com os encontrados em florestas. A razão pela qual os sistemas de baixa biomassa podem ter elevada produtividade é devido às altas taxas de crescimento (duplicação celular microbiana). Os oceanos representam cerca de metade da produção primária global.

A fixação de carbono através de medições diárias de pigmentos do fitoplâncton foi avaliada para um transecto meridional pelo oceano Atlântico, o qual os valores de produtividade primária (taxa de fixação de carbono superficial) variou de 0,2 mmol C m<sup>-3</sup> d<sup>-1</sup> nos giroscópios subtropicais a 0,2-0,5mmol C m<sup>-3</sup> d<sup>-1</sup> no Atlântico equatorial (POULTON et al., 2006). Nas águas de superfície, embora o picofitoplâncton tenha representado a fração dominante em termos de fixação de carbono (50-70%) e de clorofila-a (80-90%), as contribuições do nanofitoplâncton para a fixação total de carbono (30-50%) foram maiores do que para clorofila-a total (10-20%). No entanto, no clorofila subsuperficial, o picoplâncton dominou tanto a fixação de carbono (70-90%) quanto a clorofila-a (70-90%). Essas baixas taxas de fixação de carbono normalizadas pela clorofila para o picofitoplâncton também podem refletir perdas de carbono fixo (exsudação ou respiração celular) (POULTON et al., 2006). No Atlântico equatorial, a comunidade permanece dominada por *Prochlorococcus*, embora *Synechococcus* e os piceucariotos tornam-se relativamente abundantes (KIRCHMAN, 2011; FIGUEIRAS et al., 2014).

O sinal da cor do oceano captado pelo satélite é influenciado pela composição, tamanho e concentração das partículas e possibilita um acompanhamento temporal e espacial da produção primária nos oceanos. Para fornecer estimativas precisas da concentração de POC pela cor do oceano por satélite algoritmos bem validados são exigidos (EVERS-KING et al., 2017). Comparação destas estimativas com as das reservas de carbono no fitoplâncton representam cerca de 1:3 do carbono do fito para o carbono POC (STRAMSKI et al., 2008; GRAFF et al., 2015; KOSTADINOV et al. 2016; EVERS-KING et al., 2017). Diversos algoritmos foram utilizados estimar o estoque global do POC, estoques que podem variar entre 0,77 e 1,30Pg C, seguindo o algoritmo da NASA a estimativa foi de 0,86Pg C (EVERS-KING et al., 2017).

As análises de carbono estocados em cianobactérias representam a parte lábil, produtiva, remineralizadora dos ciclos biogeoquímicos (POULTON et al., 2006; GRAFF et

al., 2015). Compreender a magnitude desta fixação deste carbono é fundamental para estimar as trocas de carbono entre os estoques, principalmente entre a interface oceano/atmosfera. Os estudos atuais estimam estes estoques de maneira generalizada, geralmente transectos, não individualizando áreas (ZUBKOV et al., 2000; HEYWOOD et al., 2006; POULTON et al., 2006; MILICI et al., 2016). Gerar dados de estoques regionais que particionem as mesmas características zonais é mais coerente. Agrupar áreas de mesmas características oceanográficas e correntes oceânicas, torna as taxas de fixação de carbono mais precisas. Informações do estoque de carbono no plâncton na plataforma costeira brasileira são escassas e imprecisas (EKAU & KNOPPERS, 1999). Considerando a costa nordestina essa realidade é ainda pior, e não existem trabalhos que discutam essa relação aos arquipélagos costeiros brasileiros.

### **1.5.2. Determinação de biomassa (carbono)**

O método tradicionalmente mais usado para transformar o volume celular em biomassa de carbono consiste na determinação do número celular (concentração). Sendo a contagem em microscópio óptico ou invertido o método clássico. Utilizando suportes de contagem para estabelecer o número de eventos visualizados (células ou macromoléculas marcadas), sendo uma maneira direta de contagem do número celular. Nestes casos a câmara de Neubauer (de vidro ou espelhada) é altamente adequada para a contagem do fitoplâncton. O microscópio invertido de epi-fluorescência também é utilizado, pois usa diferentes excitações luminosas e filtros para distinguir as células fitoplanctônicas por meio da fluorescência natural dos pigmentos fotossintetizantes.

A citometria de fluxo é um método de contagem celular eletrônico, o qual usa de uma fonte de excitação e diferentes *flags* (filtros) para a contagem de populações de microalgas, sendo que a observação depende dos sinais naturais dos pigmentos fotossintéticos ou corantes fluorescentes de DNA (MARIE et al., 1997; EIGEMANNA et al., 2013). O sistema ótico de um citômetro de fluxo, geralmente, é composto por um laser azul e um sistema de filtros e lentes que permitem a distinção de 2 parâmetros de tamanho: FSC (*forward scatter*) sinais diretor de tamanho; SSC (*side scatter*) sinais laterais de dispersão de luz que captam a complexidade/densidade da partícula. Os parâmetros de fluorescência são coletados pelo conjunto de lentes perpendiculares, e focados em uma série de filtros óticos. O primeiro espelho (*short pass*) deixa passar as fluorescências verdes e amarelo-esverdeadas e reflete comprimentos de onda maiores pelo tubo fotomultiplicador FL1 (verde - amarelo esverdeado). A luz refletida volta para um segundo espelho (*long pass*) que deixa passar os comprimentos de onda vermelhos para o FL3 (vermelho) e reflete as luzes amarela e laranja para o FL2 (amarelo - laranja) (PARTENSKY & VAULOT, 1999; DUBELAAR & JONKER, 2000). Desta forma é possível contar de maneira direta o número de células pigmentadas.

Rodríguez et al. (1998) foram os primeiros a classificar o fitoplâncton por meio da citometria de fluxo, não apenas a partir do tamanho, mas também com base na fluorescência dos pigmentos ficoeritrina e da clorofila-a presentes nos fotossistemas. Ainda considerando citômetros que se utilizam do laser azul (488nm) para excitação podemos distinguir pigmentos através dos *flags*: FL1 (emissão-máx. 520nm) para distinguir fluorocromos que marcam DNA, sendo o SYBR<sup>®</sup> Green um dos mais utilizados para fitoplâncton; FL2 (emissão-máx. 578nm) para ficoeritrina (*R-phycoerythrin*); FL3 (emissão-máx. 678nm) para clorofilas e carotenóides (*peridinin*), além do marcador iodeto de propídio (PI), também muito utilizados para células do fito fixadas. Alguns citômetros que apresentam mais de um laser emissor podem apresentar mais *flags*, por exemplo, os que possuem o laser excitador vermelho ( $\approx 635\text{nm}$ ) emitem no FL4 (emissão-máx. 660nm), determinante para as ficobilinas (*phycocyanin*) (MARIE et al., 1997; MARIE et al., 2014).

A partir do número celular pode ser calculado o biovolume celular e derivada a biomassa de carbono pela conversão por algoritmos pré-determinados através do estudo de culturas puras e amostras ambientais, representando médias as diferentes frações de diâmetro das comunidades picofitoplanctônicas (BJØRNSSEN 1986; SIMON & AZAM, 1989; BJØRNSSEN & KUPARINEN 1991; LI et al., 1992; CHRISTIAN & KARL, 1994; KROER, 1994; TROUSSELIER et al., 1997; ZUBKOV et al., 1998; 2000; HEYWOOD et al., 2006; LINACRE et al., 2015). Devido à forma circular cocóide das cianobactérias e de grande parte dos piceucariotos, o método do biovolume é ajustado ao formato celular esferoide (BJØRNSSEN 1986; SIMON & AZAM, 1989; CHRISTIAN & KARL, 1994; KROER, 1994). O fracionamento de tamanho não estima corretamente o biovolume das diferentes faixas de tamanho das comunidades do nanofitoplâncton e das frações superiores (micro, meso e macrofitoplâncton). Pois estas células são mais rebuscadas em formas e projeções, tendo de demandar uma infinita variação de algoritmos, sendo inviável a transformação direta da densidade destes organismos em carbono (SIMON & AZAM, 1989; LI et al., 1992; CHRISTIAN & KARL, 1994; KROER, 1994; TROUSSELIER et al., 1997; ZUBKOV et al., 1998).

Outro método muito utilizado para determinação de biomassa é pela quantidade de clorofila determinada pela espectrofotometria (excitação e emissão da concentração da molécula de cloro-a), sendo esta uma maneira indireta de contagem. Desde que se tenha uma curva de crescimento de uma determinada espécie, é possível determinar a quantidade de clorofila em cada ponto dessa curva, assim por matemática pode-se aferir o valor de clorofila pelo número de células (SUGGETT et al., 2010), permitindo estimativas da biomassa fitoplanctônica. Inicialmente o método foi descrito por Lorenzen (1967), o qual consistia na extração da clorofila em um solvente orgânico (acetona) e a captura do sinal a 664nm por um espectrofotômetro. Atualmente também é possível, com o uso do fluorímetro, avaliar a

quantidade de clorofila-a diretamente na amostra, sem a necessidade de extração com solventes (GIANNINI & CIOTTI, 2016).

O sensoriamento remoto se utiliza de sensores afastados da superfície terrestre preparados para absorver frequências emitidas a partir da superfície do planeta Terra. A fundamentação ocorre pela medida da radiação eletromagnética que chegam ao sensor a bordo de um satélite e podem caracterizar o estado ou inferir propriedades do mar (CIOTTI et al., 2002). Uma serie de filtros e softwares processam tais imagens e podem compor uma faixa espectral específica ao parâmetro observado (CARR et al., 2006; BRACHER et al., 2017). No caso do parâmetro clorofila, o pressuposto sugere que o teor de sua emissão difratada pela água do mar, resulte em uma frequência que pertença ao espectro de radiação emergente do oceano, também chamado por espectro de cor (O'REILLY et al., 1998). A concentração de organismos (biomassa) é proporcional à interação da energia solar com os pigmentos clorofilados (O'REILLY et al., 1998; CIOTTI et al., 2002; CARR et al., 2006; BRACHER et al., 2017). Relações que estimam a concentração de clorofila através da cor dos oceanos usam uma razão entre bandas do azul e do verde. Uma área de baixa abundância de fitoplâncton é correspondente a uma baixa concentração de clorofila. Medidas *in situ* são determinantes e necessárias para calibrar os modelos numéricos que relacionam os dados radiométricos com a concentração de clorofila nas bandas de 490 e 555nm (O'REILLY ET AL., 1998).

O sensoriamento remoto utilizado para determinar o POC também vem sendo cada vez mais creditado e utilizado para prever os estoques globais de POC e de carbono estocados no fitoplâncton (STRAMSKI et al., 2008; GRAFF et al., 2015; KOSTADINOV et al. 2016; EVERS-KING et al., 2017; ROY et al. 2017). Alguns métodos para estimar o carbono do fitoplâncton a partir dos dados das cores do oceano dependem da parametrização da relação carbono-clorofila. Ao passo que outros métodos são baseados na estimativa de partículas totais carbono orgânico (POC) e confiam na suposição de que uma fração conhecida de POC é constituída pelo carbono proveniente do fitoplâncton (STRAMSKI et al., 1999; STRAMSKI et al., 2008). Em um trabalho pioneiro usando amostras e imagens de satélite do Oceano Austral, pôde-se constatar que o POC está bem correlacionado com a retrodifusão óptica das partículas suspensas na água do mar. Esta relação, em conjunto com a recuperação do coeficiente de retrodifusão de rejeição de sensoriamento remoto, fornece um algoritmo para estimar o POC da superfície (STRAMSKI et al., 1999). A determinação de POC *in situ* (incorporação de C) é um dos parâmetros oceanográficos mais robustos e importantes, tão quanto á taxa fotossintética (liberação O<sub>2</sub>). Existem milhares de métodos fielmente creditados para tal avaliação, sendo a filtração retendo a matéria orgânica em uma membrana de fibra de vidro é o primeiro passo. Geralmente a variação do peso de antes e após á queima (física e/ou

química) determinam o resultado ( $\Delta$  OM - Organic Matter) (WALKLEY & BLACK, 1934; GRASHOFF et al., 1999).

## 2. Caracterização oceanográfica e ambiental da área de estudo

A área de interesse está situada na zona costeira e entorno dos montes submarinos (arquipélagos) do nordeste brasileiro. Esta região é controlada pelas massas de ar e de água da zona equatorial do Atlântico oeste. Priorizando massas da água oligotróficas, onde os processos microbianos são mais intensos, amostras superficiais foram coletadas em três pontos ao longo da costa cearense, nos municípios de Paracuru/CE (PA); Porto do Pecém - município de São Gonçalo do Amarante/CE (PP); Icapuí/CE (IC). Os arquipélagos de Fernando de Noronha/PE (FN) e São Pedro e São Paulo/PE (AS) se referem aos pontos insulares. Uma amostragem costeira na zona tropical foi realizada no município de Arraial do Cabo/RJ (AC) - zona de ressurgência, na costa sudeste do Brasil (Figura 8).

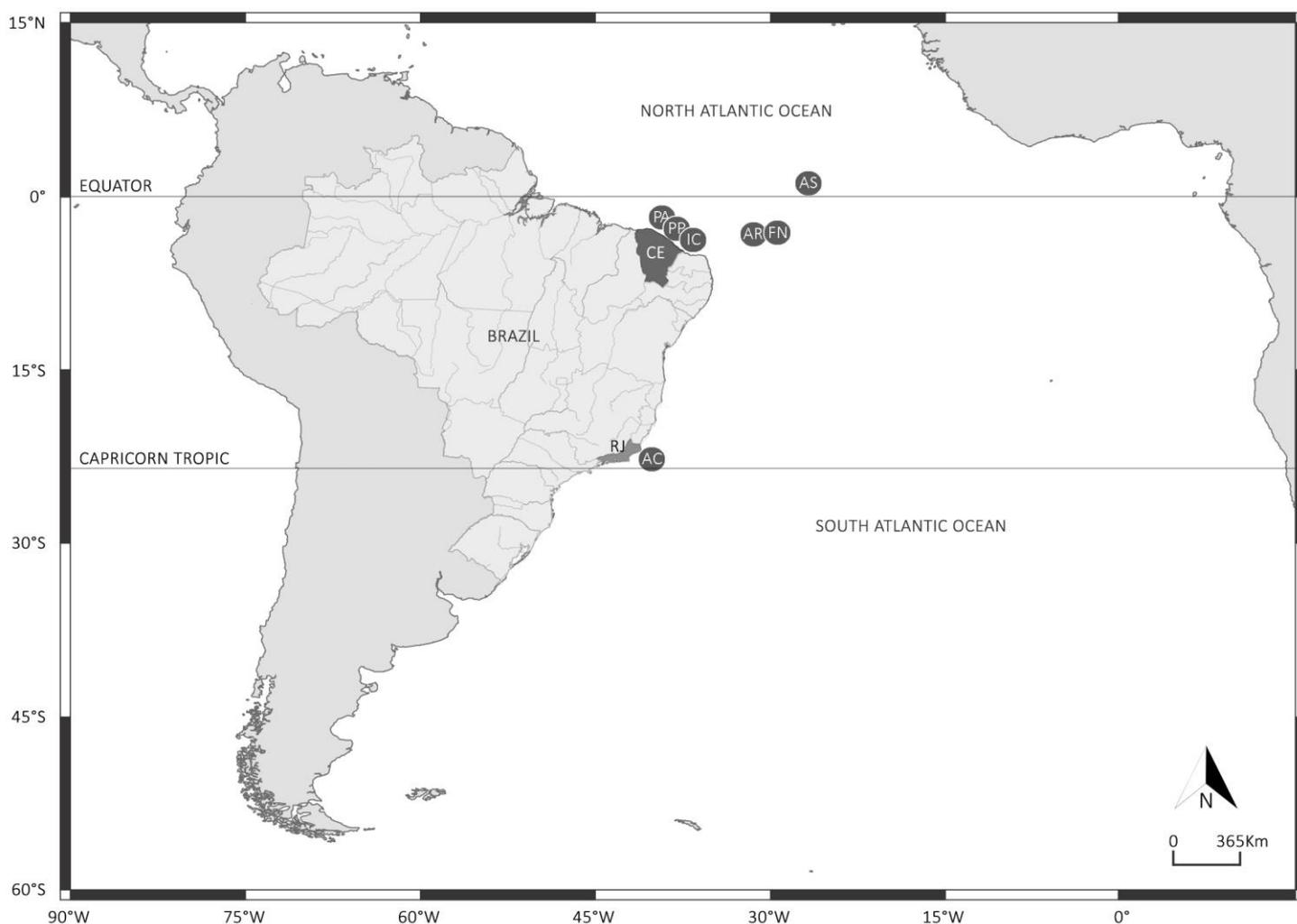


Figura 8. Os pontos deste trabalho estão situados na costa e nos montes submarinos na zona equatorial do Atlântico oeste. As siglas dos pontos na costa correspondem a PA - Paracuru/CE, PP - Porto do Pecém/CE e IC - Icapuí/CE. A ilhas oceânicas correspondem a AR - Atol das Rocas/RN, FN - Fernando de Noronha/PE e AS - Arquipélago de São Pedro e São Paulo/PE.

Com apenas uma única amostragem na costa brasileira tropical, AC - Arraial do cabo/RJ. Fonte: o próprio autor.

## 2.1. Circulação oceânica - Atlântico equatorial oeste

O movimento de energia observado nos padrões globais de circulação oceânica é transferido e, portanto, determinado pela atmosfera. A Zona de Convergência Inter-Tropical - ITCZ (Inter Tropical Convergence Zone) se apresenta como uma região de densas nuvens com grande desenvolvimento vertical entorno do globo, na baixa atmosfera acompanhando, geralmente, a linha do Equador. Isto ocorre devido ao encontro dos ventos alísios do hemisfério norte com os provenientes do sul, na zona equatorial. Os alísios de leste provenientes do hemisfério norte se movem de nordeste para sudoeste, enquanto que os ventos alísios do sul se movem de sudeste para noroeste. O deslocamento da ITCZ comanda a intensificação ou não dos ventos alísios em ambos os hemisférios, e portanto a intensificação ou não das correntes superficiais no oceano abaixo deles. As correntes oceânicas da zona equatorial do Atlântico oeste têm seus fluxos modulados diretamente pelo deslocamento da ITCZ (PETERSON & STRAMMA, 1991; TCHAMABI et al., 2017).

O clima de grande parte do Oceano Atlântico Sul é dominado por um sistema semi-permanente de alta pressão no centro do giro subtropical, que é semelhante ao encontrado em outras bacias oceânicas. No verão do hemisfério sul, o giro é centralizado perto de 32°S e 5-10°W com uma pressão central média de 1021mbar. No inverno, o centro se desloca cerca de 800km para noroeste, 27°S e 10-15°W com elevação da média para 1025mbar (PETERSON & STRAMMA, 1991). As diferenças de pressão entre o centro da alta e as linhas costeiras da América do Sul e África são maiores no verão, conseqüentemente, os ventos terrais da costa oeste do sul da África são mais fracos no verão. Isto explica o incremento no aporte de água da Corrente de Benguela para a Corrente Superficial Equatorial do Sul - SEC (South Equatorial Current). A SEC flui para o oeste em direção ao promontório oriental da América do Sul. O Cabo de São Roque/RN, é o ponto em que a SEC cisalha totalmente em duas correntes que margeiam a costa do Brasil. A Corrente Superficial do Norte do Brasil - NBC (North Brazil Current) e a Corrente Brasileira - BC (Brazilian Current) se movimentam ao longo das costas nordeste e sudeste, respectivamente (Figura 9). A NBC transporta maior volume em relação a BC, pois continua se deslocando para o Caribe contabilizando grande parte do transporte líquido do Atlântico Sul para o Norte em superfície. A SEC tem um abrangente fluxo de oeste (Figura 9), fluindo no sentido do giro, partindo do sul da África em direção ao NE brasileiro com médias de 0,3m s<sup>-1</sup>. O fluxo desta corrente diminui para o oeste (0,1-0,15m s<sup>-1</sup>) a partir de 5°S até próximo ao Equador (PETERSON & STRAMMA, 1991; STRAMMA, 1999; MARIN, 2009). A Contra Corrente Sub-Superficial Equatorial (EUC - Equatorial Under-Current), é um jato estreito com uma largura média de cerca de 1,5° de latitude e velocidades para leste, muitas vezes acima de 1,5m s<sup>-1</sup> na termoclina submersa, uma

centena de metros ou menos da superfície do mar. A maior parte da água da EUC vem da termoclina do Atlântico Sul por meio da NBC, estratificada na superfície pela pluma do Rio Amazonas. O Atlântico Norte provém muito pouco para tal corrente (PETERSON & STRAMMA, 1991; STRAMMA, 1999; COLES et al., 2013).

O Atlântico Central, também chamado de equatorial, está localizado nos limites dos giros oceânicos do Atlântico Norte e Sul. Por volta das latitudes 0-5°N encontra-se o centro de alta pressão do giro do Atlântico Equatorial, influenciado diretamente pela corrente de superfície SEC de oeste, com alta energia entre 5-10°S, e perdendo energia conforme avança ao hemisfério norte. A EUC flui para leste em sub-superfície retardando a corrente superficial SEC no avanço acima do Equador, promovendo vórtices que ativam as contra correntes de superfície (NECC - North Equatorial Counter Current) e de sub-superfície (NEUC - North Equatorial Under-Current). As contra correntes NECC e NEUC se desenvolvem na latitude de 5°N intensificando o giro equatorial (Figura 9) (FRATANTONI & GLICKSON, 2002; MARIN, 2009). O esquema dos giros oceânicos de superfície do Atlântico se direciona de acordo com os padrões de circulação forçada pelo vento presente no deslocamento da ITCZ (MAYER & WEISBERG, 1993; BUB & BROWN, 1996; BOURLÈS et al., 1999). Nas latitudes de 5-15°N se estabelece o giro do Atlântico Tropical (Figura 9). Este giro está delimitado pela Corrente Superficial Equatorial Norte (NEC - North Equatorial Current) de sentido oeste no limite superior. Enquanto que no limite inferior as correntes NECC e NEUC fluem para o leste (FRATANTONI & GLICKSON, 2002; MARIN, 2009). Nas latitudes próximas ao AS a média anual da SEC é 0,33m s<sup>-1</sup>, tal arquipélago se localiza no meio do giro equatorial do oceano Atlântico, resultando na entrada de ondulação na costa leste e na costa nordeste (LUIZ et al., 2015).

As latitudes referentes a FN e AR são diariamente afetadas pelo ramo central da SEC, com médias anuais variando de 0,2-0,15m s<sup>-1</sup> neste trecho, sendo a costa leste do arquipélago exposta, chamado de Mar de Fora. O Mar de Fora recebe ondulações sentido oeste intensificadas pelos alísios (ITCZ), é composto por 6 praias/baias. Enquanto que a costa leste é mais protegida, sendo reflexo da área de sombreamento da ilha, compondo o Mar de Dentro com calmas baías. O porto e a cidade em si, a vila dos Remédios, fazem parte da Baía de Santo Antônio, localizada no Mar de Dentro. A SEC nas latitudes 3-7°S e 10-30°W (áreas compreendidas por FN e AR) mostram uma variabilidade sazonal expressada por uma diminuição do início novembro até abril, com fortalecimento em meados de julho. Devido ao aquecimento da temperatura da superfície do mar (SST - Sea Surface Temperature), durante o verão austral perto do equador (fevereiro-março), uma maior umidade é carregada pelos alísios. Como consequência, o vento predominante em FN e AR são os ventos alísios sudeste, que se fortalecem entre julho e agosto (CORDEIRO et al., 2013; TCHAMABI et al., 2017).

Ilhas e montes submarinos estabelecem padrões de correntes locais, geralmente cisalhando a corrente superficial e originando vórtices na superfície que aprisionam água no entorno das ilhas. Já em correntes mais profundas podem promover o levantamento forçado de água abaixo da termoclina devido a perturbação física da montanha submersa (PETERSON & STRAMMA, 1991; FRATANTONI & GLICKSON, 2002; MARIN, 2009; LUIZ et al., 2015).

### **2.1.1. Ilhas oceânicas**

Os arquipélagos de FN e AS pertencem ao estado do Pernambuco, mas não necessariamente estão mais próximos a este estado, sendo o Rio Grande Norte o estado com porções de terra mais próximas as ilhas oceânicas do nordeste brasileiro. FN está a 540Km na direção sudoeste da capital pernambucana (Recife) e 375Km de Natal (Capital do RN). Considerando uma paralela longitudinal entre a ilha e a porção de terra continental, a ilha estaria distante da costa do estado do Ceará cerca de 675Km. O arquipélago de Fernando de Noronha é formado por ilhas e rochedos de origem vulcânica, sendo que a base desta montanha submarina está a 4km de profundidade (ANGULO et al., 2013). A atividade vulcânica final ocorreu a 1,8 milhões de anos atrás. Durante o período Quaternário, o intemperismo e a erosão costeira foram as principais responsáveis por esculpir a morfologia atual das ilhas. Os registros geológicos de Fernando de Noronha fornecem dados do oceano Atlântico ocidental equatorial, onde apenas duas outras massas de terra emergentes são encontradas. Estes são os arquipélagos do Atol das Rocas, 150 km a oeste e São Pedro e São Paulo, 630 km a nordeste (ANGULO et al., 2013). A área total das 21 ilhotas compreendem 26Km<sup>2</sup> dos quais 17Km<sup>2</sup> pertencem a ilha principal (Vila dos Remédios), com 10km de comprimento e 3,5km de largura (TEIXEIRA et al., 2003). O clima da ilha é o tropical típico, quente o ano todo com temperatura média anual de 26°C e chuvas concentradas entre fevereiro e julho, sendo abril o mês mais chuvoso (290mm). A vegetação atual é composta por restingas e uma mata pouco densa de médio porte secundária, uma vez que a ilha já foi extremamente alterada ao longo da história de ocupação/descoberta, que remonta desde a década de 1500 (TEIXEIRA et al., 2003). A maior parte do arquipélago foi declarada parque nacional marinho desde 1988, abrangendo uma área de 112,7Km<sup>2</sup> de preservação (parte marinha incluída). O plano de gestão aborda o turismo, pesquisa, educação ambiental, proteção e monitoramento da biodiversidade, levando em consideração a capacidade de carga de diferentes áreas dentro do parque e regula a navegação e o mergulho (MOHR et al., 2009).

A cerca de 150Km a oeste de FN situa-se a reserva biológica do Atol das Rocas, o único atol do Atlântico Sul. Pertence ao território do Rio Grande do Norte, com 7,5Km<sup>2</sup> abrigando a ilha do Farol e do Cemitério. Possui um formato oval, com 3,5km de comprimento e 2,5km de largura (PEREIRA et al., 2010). Em linha reta paralela estaria 505Km distante da costa cearense a oeste, e a sul-sudoeste distante 260Km de Natal/RN e

475Km de Recife/PE. O Atol das Rocas foi a primeira área protegida por uma reserva biológica marinha no Brasil, desde 1979. As únicas atividades humanas permitidas em seu interior são aquelas relacionadas à pesquisa científica. A gestão das áreas de preservação integral de FN e AR cabe atualmente ao ICMBio - Instituto Chico Mendes de Conservação da Biodiversidade. Esta duas áreas abrangem cerca de 422,7Km<sup>2</sup> de território preservado. Pela presença das unidades de conservação, em 2001, Fernando de Noronha e o Atol das Rocas foram designados pela UNESCO como Sítios do Patrimônio Marinho, reiterando a importância única destas áreas para a conservação da biodiversidade mundial (CORDEIRO et al., 2013; MOHR et al., 2009).

O arquipélago de São Pedro e São Paulo está a 1.160km na direção sudoeste de Recife/PE e cerca de 985Km de Natal/RN. Em linha paralela longitudinal, dista 2.305Km da costa do estado do Amapá, já no hemisfério norte. AS é composto por um conjunto de 5 ilhotas e mais 4 pequenos rochedos que compreendem 0,012Km<sup>2</sup>. A maior ilha é chamada Belmonte e ocupa metade da área total do arquipélago, com 100m de comprimento e aproximadamente 60m de largura (SICHEL et al., 2008; MOTOKI et al., 2009). Situada sobre a cordilheira submarina Dorsal Mesoatlântica, em uma das maiores falhas transformantes do oceano Atlântico, 630 quilômetros de descontinuidade dos segmentos da cadeia meso-oceânica. É a única exposição mundial do manto abissal acima do nível do mar composta de rochas peridotito (SICHEL et al., 2008). O ASPSP está situado no topo da elevação morfológica submarina mais profunda bacia oceânica a frente da costa norte e nordeste do Brasil, com sua base a aproximadamente 5Km de profundidade. É diferente de vulcões submarinos, pois é constituído de duas grandes elevações tabulares, que não possuem uma plataforma submarina de pequena profundidade ao entorno, ao sul existe uma escarpa vertical com desnível superior a 2Km (MOTOKI et al., 2009). Estas morfologias submarinas sugerem que o soerguimento é decorrente de tectonismo recente, com um processo ativo de ascensão com taxa anual de soerguimento de 0,15cm por ano (MOTOKI et al., 2009). O clima predominante é o equatorial, quente e úmido ao longo do ano. Apenas a maior ilha possui vegetação, rasteira e pouco abundante, nenhum dos rochedos dispõe de mananciais de água potável. A vegetação da ilha Belmonte não é natural, sendo transportada do continente durante as construção e manutenções do Farol. A amplitude térmica é muito pequena, característica da zona da linha do Equador. Ao longo do ano a temperatura da água do mar varia entre 26°C e 28°C e umidade do ar relativamente elevada, com médias mensais entre 70% e 90% (MOHR et al., 2009). O interesse pelo AS decorreu da assinatura da Convenção das Nações Unidas sobre o Direito do Mar, ratificado pelo Brasil em 1988, dando direito de 200 milhas náuticas em torno de qualquer território brasileiro pertencentes a Zona Econômica Exclusiva - ZEE. Em 1995 a Marinha do Brasil iniciou a construção de um novo farol e um ano após, a Marinha junto a Secretaria da Comissão Interministerial para os Recursos do Mar

- SECIRM iniciou o projeto “Pro-Arquipélago”. Em 1998 houve a implantação de uma estação científica de 45m<sup>2</sup> na ilha Belmonte. A presença permanente de cientistas na estação científica é necessária para provar a habitabilidade no arquipélago, que é fundamental para obter o seu reconhecimento internacional como território brasileiro, garantindo a incorporação de 450.000km<sup>2</sup> em área de pesca e exploração submarina a ZEE brasileira (MARINHO et al., 2010). Embora o ASPSP seja mantido pela Marinha uma unidade de conservação de uso sustentável também regula as atividades de pesca neste arquipélago, a área de proteção ambiental de Fernando de Noronha-Rocas-São Pedro e São Paulo - APA FN. Com o decreto assinado desde 1986, apenas anos depois foi criada a Secretaria do Meio Ambiente de Fernando de Noronha que gerencia a APA como órgão gestor. Cerca de 797Km<sup>2</sup> de área são regulamentadas pela APA (MOHR et al., 2009).

## **2.2. Circulação costeira - correntes do Brasil e Norte do Brasil**

A SEC origina a NBC e a BC, sendo estas duas caracterizadas como correntes superficiais de contorno oeste do Atlântico Sul (Figura 9).

A costa nordeste do Brasil recebe a influência direta da NBC, adentrando a plataforma continental no sentido oeste/noroeste margeando a costa. Na porção 3°S (correspondente a costa do CE) a morfologia da costa nordestina tem um padrão leste-oeste, justamente onde a NBC é mais intensa alcançando velocidades acima de 1,1m s<sup>-1</sup> durante a intensificação dos ventos alísios de leste durante o inverno austral (ITCZ mais ao norte do Equador). As médias anuais são próximas a 0,6-0,8m s<sup>-1</sup> (PETERSON & STRAMMA, 1991; STRAMMA, 1999; MARIN, 2009). Embora o Ceará possua vários estuários, estes são pequenos e pouco volumosos, ocorrendo eventos sazonais de oligotrofia ao longo do ano, além, de parcelas de hipersalinidade, intensificadas pelas elevadas temperaturas e o intenso padrão eólico (DIAS et al., 2013).

Uma característica já comentada é que a maioria do transporte de água pela SEC é enviada ao hemisfério sul pela NBC, e uma pequena parte alimenta a incipiente BC próximo a 10°S. A corrente superficial BC se intensifica com a movimentação geostrofica do giro sul, margeando a costa do sudeste e sul brasileiro na direção sul. Nas latitudes correspondentes a Região dos Lagos no litoral carioca, ocorre uma quebra na morfologia da costa compreendendo os municípios de Arraial do Cabo e Cabo Frio, latitudes próximas a 20,5°S. Nestas latitudes a costa tem sentido leste-oeste e a velocidade da corrente se mantém acelerada com 0,5-0,6m s<sup>-1</sup> (PETERSON & STRAMMA, 1991). De acordo com o diagrama TS (temperatura e salinidade), as águas transportadas pela BC são enquadradas como água Tropical - TW (Tropical Water), perfil típico do Atlântico Sul da superfície até 100m de profundidade, aproximadamente. Em Arraial do Cabo, ocorre o fenômeno da ressurgência costeira, aflorando águas de fundo, como a água central do Atlântico Sul - SACW (South Atlantic Central Water), que geralmente estaria abaixo dos 200-350m (DIAS et al., 2013). A SACW é

resultado do encontro da corrente superficial sentido sul BC com a Contra Corrente de Fundo das Maldivas - MC (Maldivas Current) (PETERSON & STRAMMA, 1991; STRAMMA, 1999). A ressurgência costeira é um fenômeno importante aos processos ecológicos marinhos, pois disponibiliza os nutrientes retrabalhados que estão aprisionados abaixo da termoclina. Os processos de ressurgência costeira são determinados pela morfologia costeira, a topografia de fundo e os campos de vento (COELHO-SOUZA et al., 2017). No caso de AC, a ressurgência costeira é muito sensível às mudanças do vento que acompanham duas estações perfeitamente definidas: uma temporada primavera-verão, com o estabelecimento de anticiclones marinhos tropicais com campos de ventos predominantes de E-NE, favoráveis a ressurgência. E uma temporada outono-inverno, com a passagem de frentes polares, com ventos sul promovem a subsidência de massas da água superficiais quentes, aprofundando a picnoclina (COELHO-SOUZA et al., 2017; LESSA et al., 2017).

### **2.2.1 Costa do CE**

O Ceará é fortemente influenciado pelos processos de deslocamento da ITCZ, sendo os ventos intensificados nos meses de agosto a outubro quando a ITCZ está a norte do Equador (MAIA et al., 2005). E com rajadas de ventos mais amenas nos meses de março a abril, quando a ITCZ está sob o Equador (JIMENEZ et al., 1999; MAIA et al., 2005). Pereira (2012) relata valores médios mensais de ventos (a 10m do solo) variando de 2,9 m/s no período de março a abril, com direção ESE (112,5°) e no período de agosto a outubro, mais seco, com ventos circulando a 6,1 m/s com direção predominante a SE (135°). Portanto, o clima regional, incluindo a variação sazonal das chuvas, é regulado pela ITCZ. Os fenômenos do El Niño influenciam na magnitude do deslocamento da ITCZ sobre o continente, promovendo seca na região e o aumento da SST na costa do NE (KAYANO et al., 2009). E as condições de La Niña promovem condições chuvosas (MARENGO et al., 2017). A costa cearense é marcada por um forte regime de precipitação sazonal, com dois períodos bem definidos. O clima é semi-árido tropical, e é influenciado por processos oceânicos e atmosféricos que determinam a distribuição de precipitação (MARENGO et al., 2017). Um período chuvoso se estende de Janeiro a junho, e um período seco de julho a dezembro. O litoral cearense apresenta resquícios de Mata Atlântica, sendo os tabuleiros pré-litorâneos caracterizados pela vegetação composta, predominantemente, por manguezais e restingas em coberturas de dunas e savanas (MORO et al., 2011).

O município de São Gonçalo do Amarante/CE abriga o maior porto offshore do Brasil, o Porto do Pecém. Apresenta uma planície praial tipicamente contornada por extensões de dunas cobertas por restinga esparsa, com clima predominante tropical atlântico. O Porto do Pecém é um terminal portuário distante 40Km a noroeste de Fortaleza (capital cearense). A Marinha do Brasil iniciou os estudos ecobatimétricos em 1995, e em 2002 foi completamente inaugurado. É o porto do NE com o menor tempo de trânsito entre o Brasil, os

Estados Unidos e a Europa, média de 7 dias para chegar ao destino, funciona como um dos atrativos para conquistar os armadores e impulsionar as exportações brasileiras.

O clima típico de Paracuru/CE é tropical semiúmido e uma paisagem praial de restinga recortada por recifes de arenitos expostos, que formam piscinas naturais nas bordas. Está distante 70Km a noroeste de Fortaleza.

Já Icapuí/CE dista 165Km em linha reta a sudeste de Fortaleza, sendo o município mais oriental do estado do Ceará e o ponto de amostragem mais seco, com clima típico semi-árido e falésias (tabuleiros litorâneos de arenitos) com recifes expostos. A vegetação é variada com restinga e caatinga.

### **2.2.2. Arraial do Cabo**

Arraial do Cabo/RJ está em uma porção de terra avançada para dentro do mar em relação à costa adjacente, extremamente recortado e com características de circulação de água bem específicas e variantes ao longo das estações do ano, com influencia da ressurgência costeira (COELHO-SOUZA et al., 2017; LESSA et al., 2017). É um município da Região dos Lagos, no estado do Rio de Janeiro distando 120Km da capital carioca. Com uma baixa média pluviométrica, é considerado de clima tropical litorâneo com muito vento e pouca chuva. Possui pequenas ilhotas no entorno das enseadas. A vegetação original é de mata atlântica, com característicos costões rochosos interrompendo as faixas de areias, resultando praias com uma estreita faixa de areia.

As informações geográficas acima foram obtidas através site público disponível pelo Instituto Brasileiro de Geografia e Estatística - IBGE, no portal cidades, clima e vegetação (IBGE, 2018).

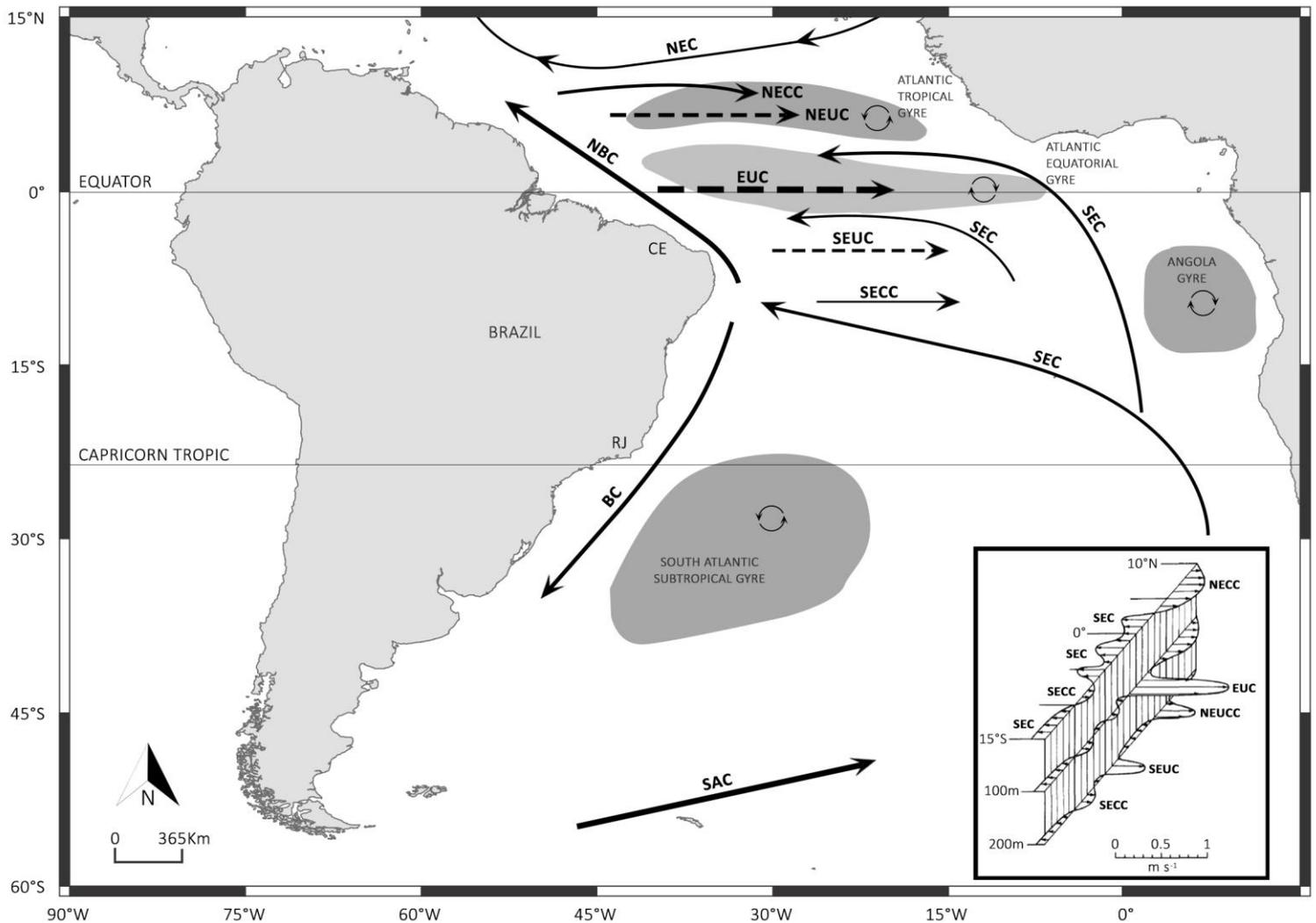


Figura 9. Representação das correntes superficiais (linha escura) e sub-superficiais (linha tracejada) no perfil agrupado de 0-150m na zona equatorial e tropical do Atlântico Sul (PETERSON & STRAMMA, 1991; BUB & BROWN, 1996; BOURLÈS et al., 1999; STRAMMA, 1999; FRATANTONI & GLICKSON, 2002; TCHAMABI et al., 2017). A intensidade das correntes está de acordo com a espessura da linha. Os termos abreviados são: NEC - Corrente Superficial Equatorial Norte; NECC - Contra Corrente Superficial Equatorial Norte; NEUC - Contra Corrente Sub-Superficial Equatorial Norte; EUC - Contra Corrente Sub-Superficial Equatorial; SEC - Corrente Superficial Equatorial Sul; SECC - Contra Corrente Superficial Equatorial Sul; SEUC - Contra Corrente Sub-Superficial equatorial Sul; SAC - Corrente Superficial do Atlântico Sul. Os giros do Atlântico Sul estão em destaque, com os sentidos representados. O perfil de profundidade e direção das correntes está em evidência na porção inferior direita do mapa (PETERSON & STRAMMA, 1991; STRAMMA, 1999). Fonte: o próprio autor.

### 3. Objetivo

Caracterização da diversidade e do estoque de carbono de cianobactérias em águas superficiais da costa cearense e dos arquipélagos na bacia oceânica a frente, seguindo um gradiente longitudinal através da biblioteca de metagenômica 16S rRNA. As correntes que atuam nestas áreas estão compreendidas na porção oeste dos giros equatoriais do oceano Atlântico.

### **3.1. Objetivos específicos**

. Determinar a concentração celular de distintas populações do pico e do nanofitoplâncton através de citometria de fluxo na costa cearense e ilhas oceânicas brasileiras localizadas na porção oeste do Atlântico.

. Estimar o volume de carbono presente nas águas superficiais em relação á biomassa específica às cianobactérias na costa e nas ilhas oceânicas na porção oeste do Atlântico equatorial;

. Descrever a diversidade de cianobactérias através da abundancia de sequencias metagenômica shotgun 16S rRNA em águas superficiais na costa e nos arquipélagos oceânicos ao longo do mesmo gradiente longitudinal na porção oeste Atlântico equatorial;

. Determinação da diversidade de cianobactérias em função das variáveis oceanográficas ph, oxigênio dissolvido, temperatura e salinidade, ao longo do gradiente longitudinal da costa e das ilhas oceânicas na zona equatorial da porção oeste do Atlântico.

. Comparação das técnicas moleculares DGGE e sequenciamento em larga escala de biblioteca 16S rRNA para determinação da extrutura da comunidade de cianobactérias na costa NE e sudeste do Brasil.

### **4. Material e métodos**

Foram realizadas 13 campanhas durante todo ano de 2015 (Tabela 5) nos arquipélagos e na costa. Com exceção de Icapuí (IC), a qual a amostragem foi realizada durante o período seco, os outros pontos da costa cearense foram amostrados no período chuvoso. Já em AC foi amostrada no outono. Nas ilhas oceânicas as coletas ocorreram entre primavera/verão, seca no Ceará (Tabela 5).

Na costa do Ceará foram amostrados 3 pontos, todos consistindo em praias expostas à dinâmica costeira, PP1 e PA1 no litoral oeste do estado e IC1 no litoral leste. Em Arraial do Cabo as coletas foram realizadas em dois pontos, AC1 uma praia exposta enquanto que AC2 corresponde a uma enseada protegida das ondulações (Tabela 5).

Em São Pedro e São Paulo ocorreram 4 amostragens compreendendo o entorno do arquipélago. Os dois pontos abrigados AS1 e AS2 foram amostrados na enseada e nas poças de maré ao lado da estação científica, respectivamente. As poças de maré situadas interior da ilha Belmonte são constantemente lavadas pelas ondas na maré alta. O ponto AS3 está localizado ao norte da ilha, numa área de sombreamento promovido pelo arquipélago, onde as ondulações mais amenas. O ponto AS4 esta ao sul da ilha, totalmente exposto ao elevado hidrodinamismo local (Tabela 5).

No arquipélago de Fernando de Noronha, a praia do Leão ao sul da ilha (FN2) possui uma elevada dinâmica sendo caracterizada como área exposta (Mar de Fora), assim como a enseada da Caieira (FN4) no leste da ilha. O ponto FN1 é abrigado, na Ponta da Sapata na

face norte da ilha. Também ao norte, o ponto FN3 está próximo às ilhotas Monte dos Dois Irmãos (Mar de Dentro), em local sombreado das ondulações.

Tabela 5. Caracterização das amostras com descrições ambientais e geográficas aos respectivos pontos de coleta, assim como a descrição dos parâmetros físico-químicos da água superficial amostrada no início de cada coleta. Fonte: o próprio autor.

Samples	Local	Site	Substrate	Dynamics	Date	GPS coordinates (datum WGS84)		Seawater physical-chemical parameters			
						Lat	Lon	pH	T (°C)	Salinity	DO (mg/L)
AC1	Arraial do Cabo/RJ	Grande Beach - Ponta da Cabeça	Sandy plain	Exposed	03/2015	22°58'30.00" S	42°02'05.70" W	8.12	25.45	33.15	6.10
AC2		Farol Island - Enseada do Maramutá	Rocky bottom	Sheltered	03/2015	22°59'25.30" S	41°59'55.20" W	8.29	25.62	34.21	5.40
PP1	Port of Pecém/CE	Pier 1 - Diving platform	Sandy plain	Exposed	06/2015	3°32'05.00" S	38°47'53.30" W	8.68	28.77	36.33	4.90
PA1	Paracuru/CE	Pedra Rachada Beach	Tide pools	Exposed	06/2015	3°23'52.80" S	39°00'50.30" W	8.33	29.50	35.67	4.70
IC1	Icapuí/CE	Ponta Grossa Beach	Sandy plain	Exposed	11/2015	4°37'35.90" S	37°30'26.80" W	8.12	28.70	35.50	5.80
AS1	Saint Peter and Saint Paul Archipelago/PE	Enseada	Rocky bottom	Sheltered	10/2015	0°55'02.20" N	29°20'46.40" W	8.24	27.25	32.49	8.20
AS2		Scientific station	Tide pools	Sheltered	10/2015	0°54'59.40" N	29°20'44.80" W	8.26	27.44	30.41	5.30
AS3		North - Open sea	Rocky bottom	Sheltered	10/2015	0°55'12.80" N	29°20'44.10" W	8.18	27.45	33.45	6.70
AS4		South - Dolphins family	Rocky bottom	Exposed	10/2015	0°54'50.70" N	29°20'42.70" W	7.19	27.14	31.28	5.90
FN1	Fernando de Noronha Archipelago/PE	Ponta da Sapata	Rocky bottom	Sheltered	10/2015	3°52'23.50" S	32°28'25.90" W	8.29	26.78	36.10	5.90
FN2		Leão Beach	Tide pools	Exposed	10/2015	3°52'14.20" S	32°26'19.20" W	8.50	27.32	36.20	6.80
FN3		Dois Irmãos Hill	Rocky bottom	Sheltered	10/2015	3°50'47.90" S	32°26'31.80" W	8.32	26.91	36.14	6.90
FN4		Enseada da Caieira	Rocky bottom	Exposed	10/2015	3°50'10.50" S	32°23'54.90" W	8.11	26.63	35.80	7.00

#### 4.1. Amostragem, transporte e armazenamento

Todas as amostras foram obtidas através de mergulho com cilindro. Nos locais com acesso direto ao mar, o mergulho para a coleta de água era realizado a partir do próprio local (praia ou encosta rochosa). Nos outros locais houve o uso de embarcações (bote inflável com motor) que serviam como plataforma de mergulho.

Amostras de água superficial foram coletadas entre 5-8m de profundidade em todos os locais de amostragem. Durante a coleta e transporte até os locais tratamento em campo, foram usadas garrafas térmicas garantindo a manutenção das características físico-químicas. Cada amostra (13 Samples - Tabela 5) possuía 5 réplicas, ou seja, 5 garrafas térmicas de plástico com o volume de 5L de água superficial destinadas às análises de citometria e genômica. E ainda mais 2 réplicas/garrafas destinadas à análise de POC, também com volume de 5L.

Ainda em campo, as réplicas pertencentes à citometria e genômica (5 garrafas de 5L) foram filtradas por ação da gravidade em malha de 10 $\mu$ m (Figura 10). No caso das duas réplicas para POC (2 garrafas de 5L), as amostras foram filtradas por gravidade em malha de 200 $\mu$ m (Figura 10), retirando as frações arenosas de sedimento em suspensão. Após a filtragem as garrafas eram transportadas, protegidas da incidência direta de luz solar direta, o mais rapidamente possível até as estações de tratamento locais para o processamento e armazenamento.

Nas estações de tratamento os conteúdos das duas garrafas destinadas ao POC foram novamente filtrados. Porém, utilizando membranas de fibra de vidro de 47mm de diâmetro e poro 0,45 $\mu$ m por bomba de filtração a vácuo (Figura 10), reterdo todo o material particulado nestas membranas. Imediatamente após a filtragem, cada membrana era alocada em um criotubo estéril e armazenada em galão de N<sub>2</sub> (nitrogênio líquido) a -195°C. Portanto as amostras de POC *in situ* foram analisadas em duplicata.

Para uma amostra, cada uma das 5 garrafas destinadas a citometria e genômica foram concentradas a 1,5L através da técnica da filtração tangencial - TFF (Tangencial Filtration Flow - Sartorius®) (Figura 10). Esta técnica usa uma malha interna de 0,22 $\mu$ m para reter as partículas da água sem danificá-las pelo excesso de pressão. Desta forma as células expostas ao sistema TFF não são corrompidas, permanecendo viáveis após este processo (BARTHEL et al., 1989; GIOVANNONI et al., 1990; RODRIGUEZ et al., 1998; BALZANO et al., 2012). O fluxo ideal padronizado previamente foi delimitado para 1L a cada 5min.

Para cada uma das cinco réplicas/garrafas concentradas a 1,5L foram retiradas 4 alíquotas fixadas para os experimentos de citometria. Cada alíquota continha 1,85mL da amostra concentrada e 0,15mL de glutaraldeído a 20% em um criotubo estéril, sendo que o volume final de 2mL de amostra estava fixada a glutaraldeído 1,5% (Figura 10).

Imediatamente após a fixação cada criotubo foi armazenado em galão de N<sub>2</sub>. Portanto as amostras de citometria foram analisadas em 5 réplicas, com cada réplica contendo 4 alíquotas.

Após a retirada das alíquotas para análise de citometria, o restante do concentrado de 1,5L de cada garrafa/réplica passou por uma filtração em série utilizando membranas inertes de éster-celulose com 47mm de diâmetro (Figura 10). Na primeira filtração a vácuo a membrana de 3µm separou a fração do nanofitoplâncton. A segunda filtração a vácuo reteve a fração do picofitoplâncton em uma membrana de 0,22µm. Imediatamente após a filtração cada membrana foi alocada em um criotubo estéril e armazenada em galão de N<sub>2</sub>. Portanto, para a análise genômica duas membranas distintas foram armazenadas para cada uma das 5 réplicas em cada amostra.

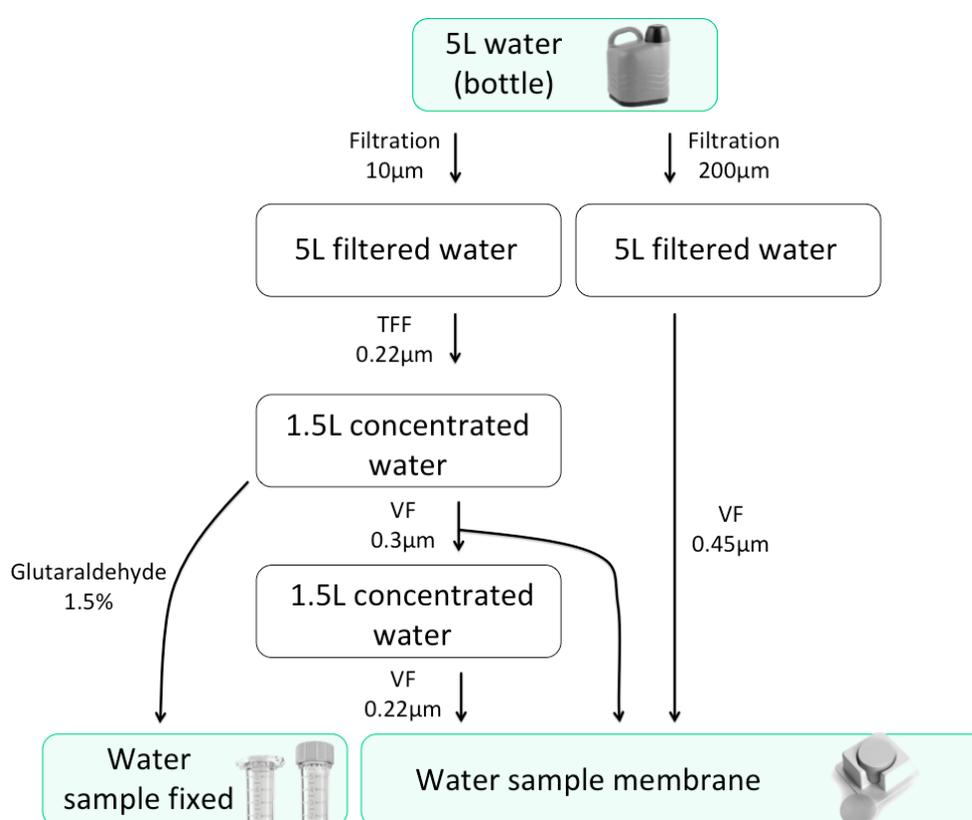


Figura 10: Desenho amostral do processamento desde a amostragem da água superficial até o armazenamento das amostras em galão de nitrogênio líquido (-196°C). As amostras fixadas foram utilizadas para os experimentos de citometria, enquanto que as amostras de membrana foram utilizadas para as análises de POC (0,45µm) e análises genômica (filtros de 3 e 0,22µm). Filtration - filtração por gravidade; TFF - filtração tangencial; VF - filtração a vácuo. Fonte o próprio autor.

#### 4.2. Variáveis ambientais

No início de cada amostragem, ou seja, assim que cada garrafa/réplica era recolhida. As variáveis físico-químicas como salinidade, pH, temperatura e teor de oxigênio dissolvido eram verificadas *in situ* através de uma sonda multi-paramétrica YSI (Hanna®). Antes da

coleta em cada ponto a sonda era devidamente calibrada de acordo com as especificações próprias (Figura 11).

Valores observados ponto a ponto foram descritos, assim como seus desvios em relação às medias costeiras e insulares, respectivamente. Uma análise de variância (ANOVA) multifatorial foi realizada entre os paramentos de temperatura, pH, salinidade e oxigênio dissolvido. Os dados foram normalizados pela função ‘decostand’ no software estatístico R. A partir destes parâmetros, foram gerados um cluster de similaridade pelas médias, usando distância euclidiana, agrupando os pontos pelas variáveis oceanográficas. Uma análise de componentes principais (PCA) e um escalonamento multidimensional (MDS) com distribuição euclidiana pelas médias também foram propostos para demonstrar a distribuição dos pontos pelas variáveis. Uma correlação de Spearman mostrou a significância para cada par de variáveis, demonstrando as forcas das relações entre si.

Descrições do tipo de fundo (Substrato - Substrate) e da hidrodinâmica local (Dinâmica - Dynamics) também foram utilizadas como variáveis categóricas a cada ponto de amostragem (Figura 11).

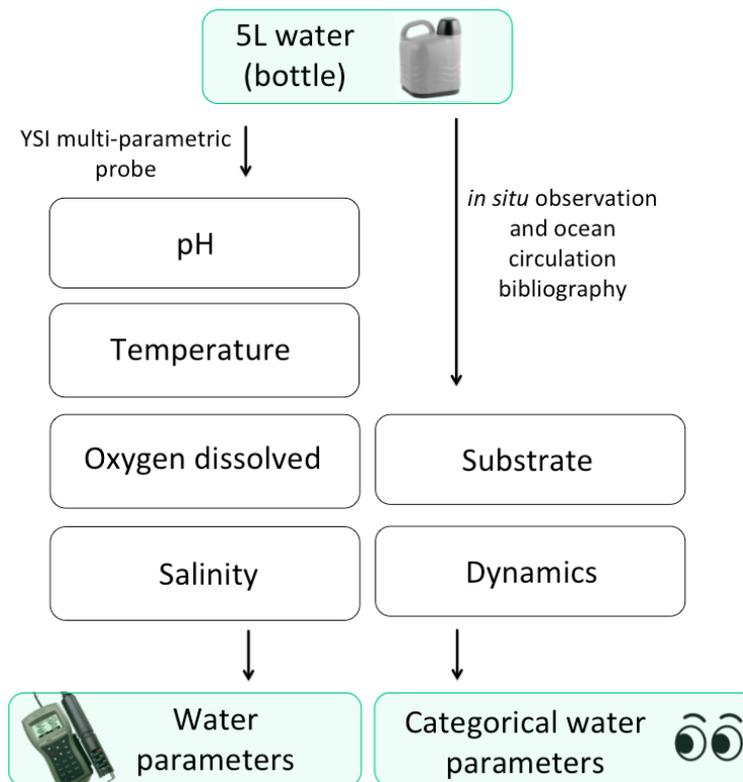


Figura 11: Desenho amostral dos parâmetros físico-químicos e de variáveis categóricas da coluna da água durante o processo de amostragem. Fonte o próprio autor.

### 4.3. Biomassa na água

A quantidade de carbono orgânico particulado presente na coluna da água no momento da coleta foi mensurado, assim como houve a preocupação de validar estes valores remotamente.

#### 4.3.1. Oxidação por dicromato

A análise de POC seguiu o método clássico de oxidação da matéria orgânica descrito por Walkley & Black (1934), e posteriormente adaptado a água do mar por Menzel & Vaccaro (1964) e Grashoff et al. (1999). Este consiste na filtração inicial da água e separação da matéria orgânica por membrana de vidro 0,45µm, e posterior pesagem das membranas liofilizadas. Após as membranas sofrerem digestão ácida úmida na presença de dicromato de potássio, sendo o excesso titulado com sulfato ferroso amoniacal (Figura 12). Após, cálculos de correção e normalização considerando o próprio erro de extração do método na recuperação da matéria orgânica foram aplicados (WALKLEY & BLACK, 1934; MENZEL & VACCARO, 1964; GRASHOFF et al., 1999; APHA, 2012). Esta análise de POC demonstra a quantidade de carbono particulado na água no momento das amostragens. Os dados foram ajustados em quantidades de carbono por volume (mg C m<sup>-3</sup>).

Uma análise de variância (ANOVA) foi realizada entre as amostras usando a função de normalização 'decostand' no software estatístico R.

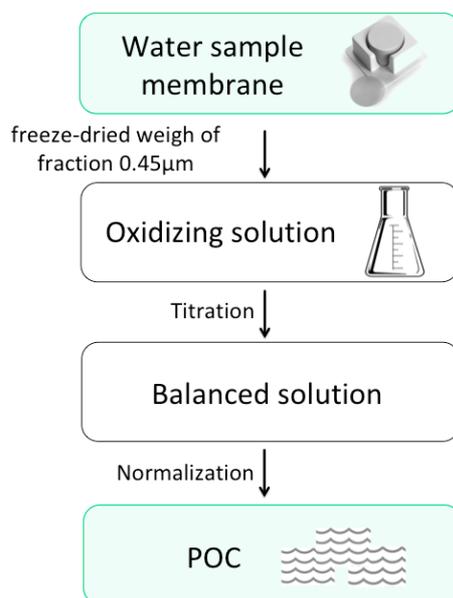


Figura 12: Desenho amostral para obtenção do carbono orgânico particulado - POC através do método de digestão ácida (WALKLEY & BLACK, 1934; MENZEL & VACCARO, 1964; GRASHOFF et al., 1999). Fonte o próprio autor.

#### 4.3.2. Sensoriamento remoto

A análise de POC determinada por sensoriamento remoto utilizou o modelo de resultados numéricos ajustados por sensores que medem a cor do oceano. Estes modelos foram retirados do banco de dados de imagens GIOVANNI v.4.24 (<https://giovanni.gsfc.nasa.gov/giovanni/>). Este modelo se utiliza dos modelos de correção de STRAMSKI et al. (2008) para a determinação de POC através da relação clorofila/carbono (KOSTADINOV et al. 2016). O sensor utilizado por tal plataforma é o MODIS-Aqua da Agencia Aeroespacial Norte-Americana - NASA (National Aeronautics and Space Administration). Sendo que o pixel desse sensor corresponde a um quadrado de 4 x 4Km, cobrindo uma área de 16Km<sup>2</sup> por pixel. A cada pixel uma média mensal pode ser exportada pela plataforma da NASA (Figura 13). O período pertinente para a coleta nas ilhas foi avaliado (out/2015 - Tabela 5), assim como a média anual do período e a série histórica dos últimos 5 anos. Uma imagem foi gerada para cada um destes períodos, revelando as médias dos teores de POC. Os valores de POC foram extraídos das imagens por meio de comandos pelo software MATLAB v. 9.1 (MathWorks<sup>®</sup>).

Devido à pouca acurácia dos sensores orbitais para carbono na zona costeira costa/oceano, somente foram descritos valores médios aos arquipélagos: FN, AT e AS. Para a determinação dos teores de POC através das imagens geradas, um retângulo foi estabelecido entorno das ilhas, a fim de obter um valor médio adequado a cada amostragem. No caso de FN 15 pontos/pixels foram estabelecidos preenchendo uma área de 128km<sup>2</sup>; para O AT foram avaliados 9 pontos/pixels cobrindo uma área de 64km<sup>2</sup>; e para o AS apenas 6 pontos cobriram uma área de 32km<sup>2</sup>. As médias de POC foram estabelecidas de acordo com os valores de pixel mais próximos a cada amostra em razão do volume (mg C m<sup>-3</sup>).

Uma análise de variância (ANOVA) foi realizada entre as amostras no software estatístico R usando a função de normalização 'decostand' no software estatístico R.

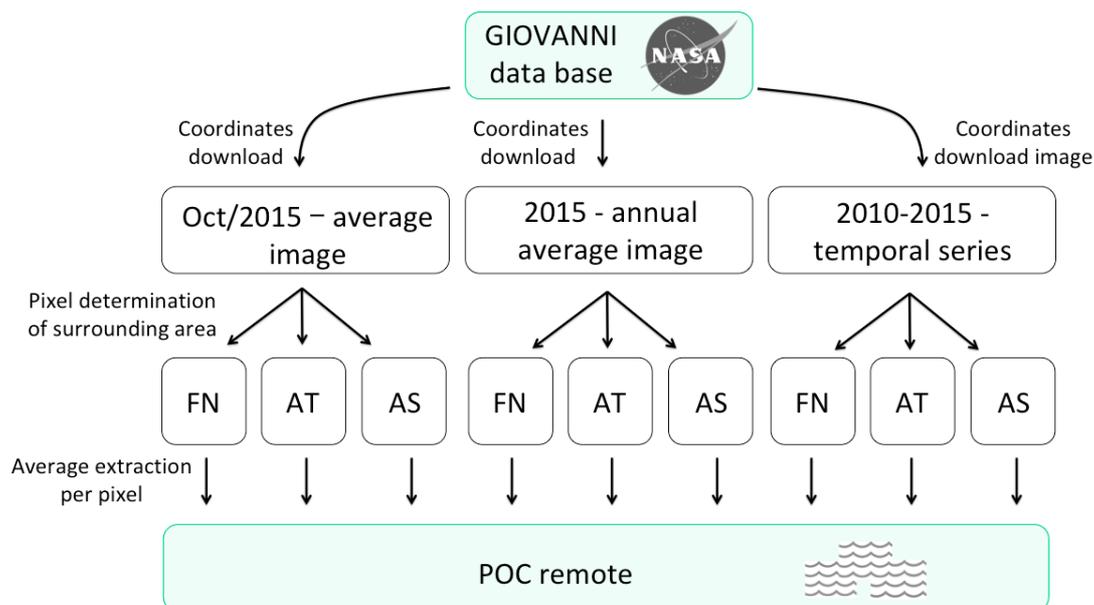


Figura 13: Desenho amostral do processamento das imagens baixadas pelo banco de dados Giovanni (NASA®) para extração dos valores médios de POC por pixel (16Km<sup>2</sup>). Sensor orbital MODIS-Aqua. Fonte o próprio autor.

Nas localidades da costa, a média de carbono particulado observado no momento da coleta (POC) das localidades amostradas foi considerada como o padrão normal. Desta forma, desvios puderam ser gerados para cada localidade em específico.

Nas amostras insulares, médias históricas obtidas pelo sensoriamento remoto (POC remote) puderam gerar médias que foram confrontadas entre si. Com base nestas médias, foram propostos desvios do POC observado em relação ao mês de amostragem (outubro de 2015), a série de um ano (2015) e a série temporal de 5 anos (2010-2015), sendo essa considerada a média normal (Figura 13).

Uma matriz de correlação de Pearson foi estabelecida entre as amostras insulares de POC e POC remote a fim de estabelecer alguma semelhança entre as distintas formas de amostragens.

#### 4.4. Biomassa das cianobactérias - citometria

As concentrações celulares das menores frações do fitoplâncton foram estimadas pelo uso da citometria de fluxo. As fluorescências naturais dos diferentes pigmentos fotossintéticos serviram como base para a diferenciação dos diferentes grupos, assim como a utilização dos parâmetros de FSC e SSC (MARIE et al., 1997; RODRÍGUEZ et al., 1998; PARTENSKY & VAULOT, 1999; DUBELAAR & JONKER, 2000; MARIE et al., 2014). As alíquotas fixadas com glutaraldeído 1,5% foram descongeladas e mantidas no escuro em refrigeração para adição de beads (esferas) marcadas. As esferas vermelho-fluorescentes (red beads) utilizadas eram marcadas para fluoescerem a 580/605nm (Thermofischer®) sob incidência do laser azul no FL3 do citômetro, como a clorofila natural. Foi utilizado o citômetro Accuri C6 (BD Biosciences®). As beads foram diluídas, para que ao final do experimento elas estivessem em

uma concentração de 1.000 eventos por uL na menor velocidade de aquisição do citômetro (velocidade baixa). Para atingir a concentração desejada das beads de 2µm (carboxiladas) houve uma diluição de 10.000X da solução estoque. Já as beads de 4µm (sulfonadas) foram diluídas 1.000X a partir da solução estoque. As diluições foram realizadas em água mili-q e ph 7,0. Uma solução de trabalho com as duas beads diluídas era realizada a cada amostragem (1 dia estabilizando em refrigeração antes do uso). A solução das duas beads não podiam variar mais que 5%, e respeitar o fluxo de 1.000 eventos por uL no equipamento. Após, 20uL desta solução de trabalho era adicionada a cada alíquota fixada (2mL com gluta 1,5%) e vortexada em máxima velocidade por 30s e posterior repouso de 30min. Cada uma destas alíquotas/replicatas foi vortexada novamente por 10s antes da aquisição pelo citômetro. Devido a testes preliminares, foi possível estabelecer os melhores padrões para a aquisição de dados pelo citômetro, que contemplasse tanto as amostras costeiras quanto as insulares, de uma maneira que não necessitassem de diluição. Após estes testes preliminares, o template de aquisição seguiu os limites (thresholds) no FSC-H de 8.000 e no FL3-H de 500, com tempo de aquisição de 2min na velocidade baixa eram adquiridos 27uL. Entre as amostras era realizada uma retrolavagem, para descontaminar possíveis interferentes de uma amostra para a outra.

A identificação das comunidades nano e picoplanctônicas autótrofas encontradas na zona equatorial do Atlântico foi avaliada através da emissão de fluorescência dos pigmentos fotossintéticos, assim como o tamanho relativo e a densidade celular. Os arquivos gerados pelo software de interface do próprio citômetro foram tratados pela plataforma FlowJo®. Os citogramas de identificação se basearam nas características intrínsecas dos pigmentos para grupo de interesse. As fluorescências no amarelo e no vermelho revelam os pigmentos das populações autotróficas naturais (MARIE et al., 1997; MARIE et al., 2001). Desta forma, foram construídos plots de: FSC-A × SSC-A - visão geral das marcações celulares da amostra (All) e determinação do Threshold residual; FSC-H × FSC-A - determinação das beads de 2µm e 4µm; FL2-A × FL3-A - grupos *Prochlorococcus* e Nanoeucariotos; FSC-A × FL3-A - Fito indeterminado e cianobactérias indiferenciadas; SSC-A × FL2-A - grupo *Synechococcus*. Os cinco grupos diferenciados (Phyto, Nanoeukaryotics, Cyanochlorophyllates, *Prochlorococcus* e *Synechococcus*) somente puderam ser distinguidos através da marcação dos sinais ao longo dos diferentes eixos e pela comparação com outros trabalhos de citometria (PARTENSKY & VAULOT, 1999; MARIE et al., 1997; MARIE et al., 2001; EIGEMANNA et al., 2013). Os resultados foram expressos pela concentração celular por litro de água (Cell L<sup>-1</sup>).

Os dados foram normalizados pela função ‘decostand’ no software estatístico R para melhor acomodação estatística. Através destes dados de concentração celular foi gerada uma análise de variância (ANOVA) multifatorial entre as populações das diferentes amostras.

Assim como, uma análise de componentes principais (Principal Component Analysis - PCA) e um escalonamento multidimensional (Metric Multidimensional Scaling - MDS) com distribuição euclidiana pelas médias.

A biomassa de carbono de cada grupo do picofitoplâncton foi inferida através da abundância/concentração de células e do uso de algoritmos de conversão (Figura 14). Estes algoritmos são extraídos da literatura e se baseiam em cálculos de área e volume celular, existem valores divergentes entre tais metodologias de conversão (LI et al., 1992). Um teor médio da densidade de carbono por célula é a base utilizada para converter os volumes de células em biomassa, usando os diferentes diâmetros celulares ( $\mu\text{m}$ ). O índice mais utilizado para células bacterianas é aferido a  $0,20\text{-}0,22\text{pg C } \mu\text{m}^{-3}$  (CHRISTIAN & KARL, 1994).

Existem diversos métodos para calcular os algoritmos de conversão através dos índices de volume acima. O foco para a escolha do índice deve estar na representatividade para a área de pesquisa e a população amostrada (Tabela 6) (LINACRE et al., 2015).

Zubkov e colaboradores (1998; 2000) realizaram um transecto norte-sul ao longo do oceano atlântico que definiu um algoritmo de conversão de abundância celular para biomassa de carbono para o picoplâncton natural em ambientes oligotróficos no oceano Atlântico Tropical/Equatorial (HEYWOOD et al., 2006; LINACRE et al., 2015). Para as amostras provenientes das ilhas oceânicas a conversão foi feita pelos fatores constantes propostos por Zubkov et al. (1998; 2000) (Tabela 6). E para as amostras provenientes da costa, os fatores constantes foram os propostos por Lee & Fuhrman (1987) (Tabela 6). Portanto, foi utilizada neste trabalho a abundância de células ( $\text{células L}^{-1}$ ) convertida em biomassa de carbono ( $\text{mg C m}^{-3}$ ) (Figura 14).

Uma análise de variância (ANOVA) bifatorial foi realizada entre os valores de biomassa das populações de cianobactérias, *Prochlorococcus* e *Synechococcus*. Os dados foram normalizados pela função 'decostand' no software estatístico R. Através destes valores de biomassa, também, foram geradas análises de componentes principais (PCA) e um escalonamento multidimensional (MDS) com distribuição euclidiana pelas médias.

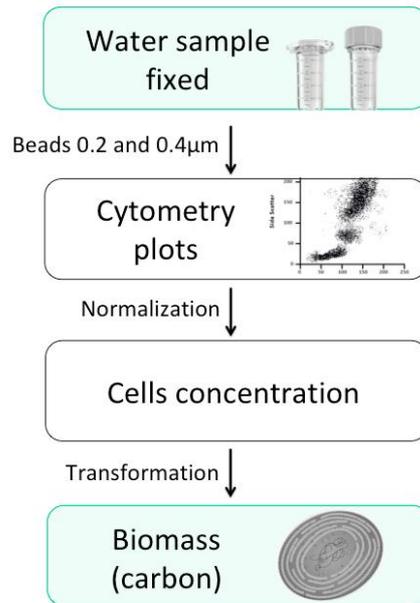


Figura 14: Desenho amostral para obtenção da biomassa de cianobactérias, a fração nano e picoplanctônicas através da densidade celular obtiva pela citometria de fluxo (MARIE et al., 1997; PARTENSKY & VAULOT, 1999; DUBELAAR & JONKER, 2000; EIGEMANNA et al., 2013; MARIE et al., 2014). Fonte o próprio autor.

Tabela 6. Algoritmos de conversão extraídos da literatura para a conversão do número celular do picoplâncton marinho em gramas de carbono por litro. As diversas referencias listadas representam os artigos mais citados para a determinação das constantes de conversão. Estes valores estão devidamente adequados para transformar uma abundancia celular (cell L<sup>-1</sup>) em biomassa de carbono (pg C cell<sup>-1</sup> L<sup>-1</sup>). As referencias marcados em cinza tiveram os algoritmos de biomassa utilizados neste trabalho. Fonte: o próprio autor.

Biomass conversion algorithms references									
Reference	(TROUSSELIER et al. 1997)	(KROER, 1994)	(LEE & FUHRMAN, 1987)	(GARRISON et al., 2000)	(CAMPBELL et al., 1994)	(LI et al., 1992)	(DURAND et al., 2001)	(ZUBKOV et al., 1998)	(ZUBKOV et al., 2000)
Samples	Pure cultures	Pure cultures	Natural population	Natural population	Natural population	Natural population	Natural population	Natural population	Natural population
Station	Lab	Estuaries - Florida/USA	Coast - New York /USA	Arabian sea	North Pacific ocean	Sargasso sea - Atlantic ocean	Sargasso sea - Atlantic ocean	Meridional transect - Atlantic ocean	Meridional transect - Atlantic ocean
Picoeukaryote	-	-	-	1.011	2.100	-	-	1.500	1.500
Heterotrophic bacteria	0.026	0.113	0.020	0.011	-	0.020	-	0.019	0.011
Cyanobacteria	-	-	-	-	-	0.250	-	-	-
<i>Prochlorococcus</i>	-	-	0.056	0.032-0.072	0.053	0,059	0.056	0.032	0.029
<i>Synechococcus</i>	-	-	0.112	0.101-0.152	0.250	-	0.112	0.103	0.100

Note: - does not have information

Em todas as localidades foram estabelecidas porcentagens de biomassa de cianobactérias em relação ao estoque de POC. As porcentagens foram estabelecidas para *Prochlorococcus*, *Synechococcus* e uma somatória destas, o grupo Cyanobacteria.

Somente nas amostras insulares, foram mensuradas porcentagens de biomassa ao longo das médias temporais do estoque de POC remote.

Uma matriz de correlação de Pearson foi estabelecida entre as populações de *Prochlorococcus* e *Synechococcus*, a fim de estabelecer uma semelhança de distribuição pelos pontos amostrados.

#### **4.5. Estrutura da comunidade**

Cada um dos 13 pontos de coleta gerou três membranas/réplicas de 3 $\mu$ m (nanofitoplâncton) e outras três membranas de 0,22 $\mu$ m (picofitoplâncton). Nestas membranas ficaram retidas células e detritos orgânicos presentes na coluna da água, inclusive as macromoléculas dos ácidos nucleicos. Este DNA foi recuperado e isolado para definir as diferentes composições de cianobactérias na costa e na ilhas oceânicas brasileiras. A análise laboratorial consistiu na extração de DNA; amplificação com primers rRNA 16s específicos para PCR-DGGE (grupo CG); amplificação com primers rRNA 16s e adaptadores específicos a preparação da biblioteca Illumina (MiSeq); e análise molecular *in silico* das técnicas (Figura 15).

A técnica da DGGE foi apenas utilizada as amostras de costa, enquanto que o sequenciamento através da biblioteca genômica 16S rRNA foi aplicado às amostras costeiras e insulares. Uma comparação experimental avaliando o refinamento das técnicas abordadas foi proposta considerando os dados qualitativos (presença e ausência - riqueza). Tal comparação experimental foi aplicada apenas as amostras costeiras nas frações do pico (0.22 $\mu$ m) e do nanofitoplâncton (3.00 $\mu$ m).

As amostras do material genético processado pelas técnicas de análise para DGGE foram convertidos em matrizes de dados de riqueza (presença/ausência) e assim comparados ecologicamente quanto às características qualitativas. Já as análises originadas a partir das amostras de sequenciamento geraram dados qualitativos e dados quantitativos (número de sequências por amostra a cada táxon - abundância).

Para os dados de DGGE as bandas encontradas foram enquadradas como OTUs. O método do sequenciamento em larga escala enquadrrou as sequências alinhadas aos bancos de dados com maior resolução taxonômica, identificadas até gênero.

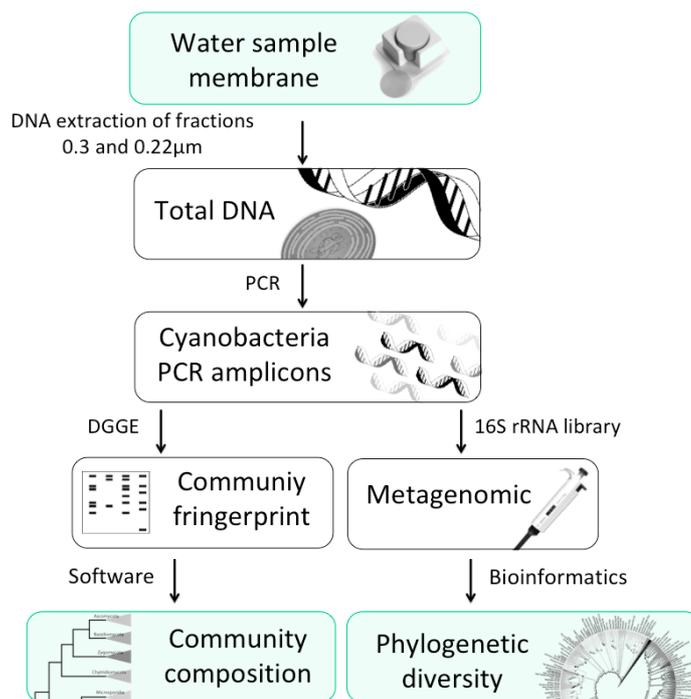


Figura 15: Esquema do processamento genômico das membranas de 3,00 $\mu$ m e 0,22 $\mu$ m desde a extração até as técnicas de isolamento e identificação dos genes que possibilitam o estudo da composição e filogenética microbiana. Análises de fingerprint e sequenciamento através de bibliotecas rRNA 16s foram utilizadas para identificar a estrutura da comunidade microbiota. Fonte: o próprio autor.

#### 4.4.1. Extração de DNA

A extração de DNA total das membranas, tanto da fração 3,00 $\mu$ m quanto para a fração 0,22 $\mu$ m foram iniciados com a lise celular com tampão CTAB (Hexadecyl-trimethyl-ammonium bromide) e separação com fenol/clorofórmio/álcool isoamílico (POREBSKI et al., 1997; WILSON, 2001; GARCIA et al., 2013). Após o processamento o material genético se encontrou isolado, disponível para a quantificação e posterior amplificação. Para maiores detalhamentos consultar os protocolos passo-a-passo em anexo no final do documento (ANEXO 2).

A amplificação por PCR garante uma multiplicação exponencial de uma região do DNA delimitada pelos primers forward e reverse. Ao decorrer das últimas três décadas, diversos trabalhos vêm codificando regiões entre o rRNA 16S e 23S aplicados a filogenia de cianobactérias (NÜBEL et al., 1997; ROCAP et al., 2002; JANSE et al., 2003; TATON et al., 2003; BOUTTE et al., 2006; LAVIN et al., 2008; MÜHLING et al., 2008). As áreas de variação V3 e V4 do rRNA 16s, assim como as regiões de hipervariação (ITS) entre elas e o rRNA 23S, são utilizadas (Tabela 7).

As duas distintas técnicas moleculares empregadas para definir a composição de cianobactérias nas amostras, a DGGE e o sequenciamento em larga escala, iniciam a partir da

amplificação por PCR. Uma PCR semi-nexted (isto é, uma amplificação em duas etapas que utilizam o mesmo iniciador - BOUTTE et al., 2006) a partir do DNA total isolado marca o início de ambas as técnicas. Porém, cada uma das duas etapas possui suas particularidades em relação às condições de PCR (BOUTTE et al., 2006; SINCLAIR et al., 2015). O primer reverse CYA781(b) utilizado na última etapa, para ambas as técnicas, prioriza como alvo cianobactérias unicelulares (NÜBEL et al., 1997; BOUTTE et al., 2006; MÜHLING et al., 2008). Todos os primers utilizados foram sintetizados comercialmente.

As réplicas de cada amostra passaram por diversas extrações, sendo que ao final de toda extração cada réplica possuía uma solução tampão com DNA total extraído suspenso. Todos estes DNA tiveram mensuradas suas concentrações em fluorímetro (DNA - 260nm), determinando as relações de pureza do material extraído. A relação 260nm/280nm é usada para avaliar a pureza de DNA, uma taxa de 1,8 é geralmente aceita como "pura" para o DNA. Se a relação é sensivelmente menor, pode indicar a presença de contaminantes que absorvem fortemente em ou perto de 280nm, como proteínas. Já a relação 260nm/230nm é usada como uma medida secundária de pureza de ácido nucleico. Os valores para DNA "puro" estão geralmente na faixa de 2,0 - 2,2. Se a relação é sensivelmente inferior ao esperado, pode indicar a presença de contaminantes que absorvem a 230nm, como: EDTA, carboidratos e fenol. Géis de agarose com posterior corrida em eletroforese também foram confeccionados com os DNA totais extraídos, a fim de identificar a integridade dos genes extraídos. Apenas os DNA extraídos com os padrões ótimos para as relações de pureza de DNA e integridade (sem esfregaço no gel de agarose) foram selecionados para as análises das sequências.

A técnica de DGGE foi empregada apenas para as amostras costeiras, enquanto que o sequenciamento foi empregado a todas as amostras. Portanto, uma comparação breve entre a resolução e similaridade das metodologias foi possível utilizando o exemplo das amostras costeiras.

Tabela 7. Os principais primers aplicados a filogenia de cianobactérias que codificam as regiões do DNA genômico entre as porções rRNA 16S e 23S. Os primers marcados em cinza foram os utilizados nos experimentos de DGGE e sequenciamento NGS neste trabalho. Fonte: o próprio autor.

Gene	Primer names	Sequence (5'-3')	References
rRNA 16s	CYA359F	GGG GAA TTT TCC GCA ATG GG	(NÜBEL et al., 1997)
	CYA359F-adapter 1	TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG GTG GGG AAT TTT CCG CAA TGG G	(NÜBEL et al., 1997; Illumina,2013)
	CYA784R	GAC TAC AGG GGT ATC TAA TCC C	(BOUTTE et al., 2006; MÜHLING et al., 2008)
	CYA781R(a)	GAC TAC TGG GGT ATC TAA TCC CAT T	(NÜBEL et al., 1997)
	CYA781R(b)	GAC TAC AGG GGT ATC TAA TCC CTT T	(NÜBEL et al., 1997)
	CYA781R-CG(b)	CGC CCG CCG CGC CCC GCG CCC GTC CCG CCG CCC CCG CCC GAC TAC AGG GGT ATC TAA TCC CTT T	(NÜBEL et al., 1997; BOUTTE et al., 2006)
	CYA781R-adapter 2(b)	GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGA CTA CAG GGG TAT CTA ATC CCT TT	(NÜBEL et al., 1997; Illumina,2013)
	CSIF-CG	GWC ACG CCC GAA GTC WTT AC	(JANSE et al., 2003)
	DiNAF	GAA TCT GCC CTC AGG AGG GGG	(LAVIN et al., 2008)
	DiNAR	GGG TTG CCC CAT TCG GAA AT	(LAVIN et al., 2008)
16S-1247F	CGTACTACAATGCTACGG	(ROCAP et al., 2002)	
ITS 16S	ITS-aF	GGA TCA CCT CCT AAC AGG GAG	(LAVIN et al., 2008)
	ITS-aR	GGA CCT CAC CCT TAT CAG GG	(LAVIN et al., 2008)
rRNA 23s	23S-30R	CTT CGC CTC TGT GTG CCT AGG T	(TATON et al., 2003)
	23S-241R	TTC GCT CGC CRC TAC T	(ROCAP et al., 2002)
	23S-1608R	CWA CCT GTG TCG GTT T	(ROCAP et al., 2002)
ITS 23S	ITS-bF	GTT GGT AGA GCG CCT GCT TTG	(LAVIN et al., 2008)
	ITS-bR	CCG TGA GCC CTT TGT AGC TTG	(LAVIN et al., 2008)
	ITS-AlaF	TWT AGC TCA GTT GGT AGA G	(ROCAP et al., 2002)
	ITS-AlaR	CTC TAC CAA CTG AGC TAW A	(ROCAP et al., 2002)
DGGE	CG gramp (F or R)†	[specific primer]-CGC CCG CCG CGC CCC GCG CCG GTC CCG CCG CCC CCG CCC G	(SHEFFIELD et al., 1989)
Illumina	Adapter 1F	TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG GT-[specific primer]	(Illumina,2013)
Nextera XT	Adapter 2R	GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G-[specific primer]	(Illumina,2013)

Notes: W - vase ajustada de acordo com o grupo de interesse; † R (reverso) ou F (iniciador) tem orientação designada de acordo com o rRNA, são 40 nucleotídeos ricos em uma sequencia GC (SHEFFIELD et al., 1989).

#### 4.4.2. PCR-DGGE

Para a técnica da DGGE, a primeira etapa da PCR semi-nested amplificou uma extensa região codificada pelos primers CYA359F e 23S30R, a partir do DNA total isolado após a extração (Tabelas 8 e 9). Na segunda etapa, a amplificação ocorreu a partir do produto de PCR da etapa anterior com o conjunto de primer CYA359F e CYA781R-CG(b) (Tabelas 10 e 11), sendo a região amplificada final com 422pb na região V4 do rRNA 16s (NÜBEL et al., 1997; YU & MORRISON, 2004; MÜHLING et al., 2008). Esses amplicons finais foram o input da técnica DGGE, sendo o grampo CG acoplado a um dos primers, uma característica fundamental para manter a estabilidade da dupla fita durante a execução da técnica (SHEFFIELD et al., 1989).

Os amplicons das amostras costeiras foram aplicados no gel desnaturante, para separação por bandeamento na DGGE. Géis de acrilamida 8% foram preparados com gradientes desnaturantes de 40 a 70%, usando uma solução desnaturante 100% (7M uréia e 40% formamida) e uma solução 0% (sem uréia e formamida). Em cada poço (slot) do gel foi aplicada uma alíquota de cerca de 10µL de produto de PCR final misturada com 2µL do tampão com azul de bromofenol. A aplicação das amostras foi realizada no gel com os slots previamente preenchidos com tampão TAE 0,5X. A eletroforese foi realizada a 60°C e a uma voltagem de 60V, mantida constantes em solução tampão TAE 0,5X por 16h. Após a eletroforese, o gel foi imerso em uma solução de corante com volume final de 45mL por 1h sob leve agitação (0,5mL de TAE 50X; 44,5mL de água ultrapura; 5µL SYBR® Green DNA). A imagem do gel foi capturada em um sistema de fotodocumentação ultravioleta (NÜBEL et al., 1997; JANSE et al., 2003; BOUTTE et al., 2006; MÜHLING et al., 2008).

Três réplicas de cada ponto localizado em região costeira foram aplicadas em três géis distintos para cada fração das membranas de 3,00 e 0,22µm. As amostras de AC tiveram apenas duas réplicas por localidade. Em cada gel eram aplicadas 13 amostras mais o branco e os marcadores. Os géis de cada fração foram analisados e comparados quanto a similaridade. Destes três géis de cada fração, o mais nítido quanto à separação das bandas foi o utilizado nas análises dos resultados.

As análises dos géis da DGGE foram realizadas pela plataforma GelCompar II do programa BioNumerics® (Applied Maths). Este programa foi fundamental para normalização, conversão, comparação e grupamento preliminar das amostras pelas imagens dos géis. Correções manuais foram feitas quando necessárias, sendo que a normalização e a seleção de bandas foram cuidadosamente avaliadas. A forma em que a abundância é expressa pelo método da DGGE seria através da espessura das bandas formadas para cada OTU. Devido ao gel ser muito dependente das condições de corrida e fotodocumentação, a abundância não foi aferida. Já a riqueza pode ser definida pela presença e ausência de distintas bandas ao longo do gradiente do gel.

Os parâmetros ecológicos foram estimados por diversos pacotes do software estatístico R, não houve normalização dos dados (dados qualitativos). A riqueza e a riqueza relativa foram dimensionadas a cada ponto, assim como a riqueza estimada e a curva de rarefação média pelo estimador de riqueza Chao2 (dados binários). A  $\alpha$ -diversidade (ou diversidade local) foi calculada usando o índice de Shannon-Wiener e Simpson. Médias de diversidade e riqueza da costa foram consideradas como os padrões normais, desta forma, desvios puderam ser gerados para cada localidade em específico. O índice de Jaccard foi utilizado para avaliar a  $\beta$ -diversidade e assim distribuir as amostras em clusters. Matrizes de presença/ausência foram originadas e resultaram em agrupamentos hierárquicos considerando o algoritmo UPGMA (média de valores) sendo avaliado o coeficiente de correlação cofenético (OKSANEN, 2015: 2018). Foi proposto um escalonamento multidimensional não métrico (Non-Metric Multidimensional Scaling - NMDS) utilizando a distribuição de Jaccard com dados binários pelas médias.

Tabela 8. Listagem das soluções e suas respectivas concentrações e quantidades para formar uma solução tampão ideal para o funcionamento da Go Taq e anelamento correto dos primers da PCR-DGGE 1. A cada amostra são somados 14,7 $\mu$ L do Mix, e portanto, 10,3 $\mu$ L de água para biologia molecular para formar o volume final de solução. Fonte: o próprio autor.

PCR-DGGE semi-nexted 1			
Reagents	Stock concentration	Final concentration	Final Volume (25 $\mu$ L)
Go Taq buffer	5X	1X	5 $\mu$ L
MgCl <sub>2</sub>	25mM	2mM	2 $\mu$ L
CYA359F	5 $\mu$ M	0.5 $\mu$ M	2.5 $\mu$ L
23S-30R	5 $\mu$ M	0.5 $\mu$ M	2.5 $\mu$ L
DNTP	10mM	0.2mM	0.5 $\mu$ L
Go Taq	5V/ $\mu$ L	1V	0.2 $\mu$ L
DNA	10ng/ $\mu$ L	20ng	2 $\mu$ L

Tabela 9. Condições ideais aplicadas ao termociclador durante o processo da PCR, considerando as particularidades físico-químicas do conjunto de primers e da enzima Taq durante a PCR semi-nexted 1. Foram as mesmas condições para as técnicas da PCR-DGGE e NGS. Fonte: o próprio autor.

PCR-DGGE and PCR-NGS semi-nexted 1			
	T (°C)	Time (min)	Process
10 cycles	94	05:00	Initial step of denaturing
	94	00:45	Denaturation
	54	00:45	Annealing
	72	02:00	Extension
	72	07:00	Extension stabilization

Tabela 10. Listagem das soluções e suas respectivas concentrações e quantidades para formar uma solução tampão ideal para a amplificação e anelamento correto dos primers da PCR-DGGE 2. A cada amostra são 27,2µL do Mix, e portanto, 22,8µL de água para biologia molecular para formar o volume final de solução. Fonte: o próprio autor.

PCR-DGGE semi-nexted 2			
Reagents	Stock concentration	Final concentration	Final Volume (50µL)
Go Taq buffer	5X	1X	10µL
MgCl <sub>2</sub>	25mM	2mM	4µL
CYA359F	5µM	0.5µM	5µL
CYA781R-CG(b)	5µM	0.5µM	5µL
DNTP	10mM	0.2mM	1µL
Go Taq	5V/µL	1V	0.2µL
Amplicon PCR-DGGE 1	-	-	2µL

Tabela 11. Condições ideais aplicadas ao termociclador durante o processo da PCR, considerando as particularidades físico-químicas do conjunto de primers e da enzima Taq durante a PCR semi-nexted 2. Foram as mesmas condições para as técnicas da PCR-DGGE e NGS. Fonte: o próprio autor.

PCR-DGGE and PCR-NGS semi-nexted 2		
T (°C)	Time (min)	Process
94	05:00	Initial step of denaturing
94	01:00	Denaturation
60	01:00	Annealing
72	01:00	Extension
4	∞	Storage

35 cycles {

#### 4.4.3. PCR-sequenciamento em larga escala

Para a técnica de sequenciamento em larga escala ou também chamado de nova geração - NGS (Next Generation Sequencing), os primers foram construídos utilizando os adaptadores específicos acoplados, que foram posteriormente pinçados aos indexes do MiSeq da Illumina nas etapas da preparação da biblioteca (ILLUMINA, 2013 - Tabela 7).

A primeira etapa da PCR semi-nexted amplificou uma extensa região codificada pelos primers CYA359F e 23S30R, a partir do DNA total isolado após a extração (Tabelas 12 e 9). Na segunda etapa, a amplificação ocorreu a partir do produto de PCR-NGS da etapa anterior com o conjunto de primer CYA359F-adapter 1 e CYA781R-adapter 2(b) (Tabelas 13 e 11). Sendo a região amplificada final próxima a 422pb na região V4 do rRNA 16s (ROCAP et al., 2002; YU & MORRISON, 2004). Esses amplicons finais foram o input para a preparação da biblioteca metagenômica shotgun de rRNA 16S. A preparação da biblioteca é realizada em

vários estágios, com sequências de purificações e ampliações até formar um pool único de todas as amostras. Este pool foi introduzido na plataforma de sequenciamento MiSeq utilizando o cartucho e os index Nextera XT V3 - 600 ciclos (ILLUMINA, 2013). Todas estas etapas estão devidamente descritas nos anexos finais (ANEXO 3). Na etapa 3 de amplificação (ANEXO 3) foram utilizados os conjuntos de índices S5 (F) e N7 (R) do Nextera XT, a partir dos amplicons purificados das etapas anteriores (Tabelas 14 e 15). Durante as etapas 7 e 8 da preparação da biblioteca (ANEXO 3), uma diluição de 17pM foi utilizada, tanto para o pool de amostras quanto ao controle positivo do próprio equipamento (PhiX). Na etapa seguinte o pool foi diluído em 20% pela solução de PhiX (ANEXO 3).

Foram aplicadas no cartucho três réplicas de cada ponto, considerando as duas frações fitoplanctônicas de 3 e 0,22µm, com exceção de AC, onde foram obtidas apenas duas réplicas. Portanto um pool com o total de 74 amostras mais o branco foram introduzidos no cartucho. Cada amostra possuía um par de indexes N7 e S5 específico para identificação na plataforma MiSeq. Toda esta informação genética das amostras (74 bibliotecas rRNA16S) foi analisada *in silico* com recursos de bioinformática e processadores de alta capacidade.

Os scripts adotados seguiram comandos dentro das plataformas QIIME (CAPORASO et al., 2010) e BMP - Brazilian Microbiome Project (PYLRO et al., 2014). Tais pipelines foram produzidos em linguagem de programação em Python, e posteriormente rodados em consoles, como o terminal do Linux e o software estatístico R. Os cromatogramas, assim que exportados da plataforma MiSeq, passaram por um pré-tratamento antes de seguir a análise ecológica no QIIME. Foram retirados das sequências os fragmentos correspondentes aos adaptadores e dos indexes do Nextera XT. Sequências com mais de 240pb e com mais de 80% de bases com QC>20 foram selecionadas para constituir o conjunto de dados final. A verificação da quimera foi realizada usando o algoritmo chimera.uchime e SILVA-Gold como referências (EDGAR, 2010; EDGAR et al., 2011; BOKULICH et al., 2013). O alinhamento dos fragmentos obtidos pelo sequenciamento ao banco de dados. As sequências foram atribuídas a unidades taxonômicas operacionais (OTU). Uma similaridade de 99% entre as sequências para notá-las em nível de espécie foi estabelecida. O banco de dados SILVA (atualização 128) (QUAST et al., 2012) adequando à plataforma QIIME foi utilizada ao alinhamento das sequências para classificar sequências do filo ao nível do gênero ([https://www.arb-silva.de/no\\_cache/download/archive/qiime/](https://www.arb-silva.de/no_cache/download/archive/qiime/)). Os pipelines das análises realizadas no programa Qiime v. 1.9 estão disponíveis em anexo no final do manuscrito (ANEXO 4).

Os números de sequências alinhadas para cada fração planctônica foram normalizados pela função 'decostand' no software estatístico R. Através destes dados foi gerada uma análise de variância (ANOVA) para cada um dos gêneros *Prochlorococcus* e *Synechococcus*. Uma ANOVA bifatorial entre os gêneros também foi estabelecida, assim

como uma análise de componentes principais (PCA) e um escalonamento multidimensional (MDS) com distribuição euclidiana de dados binários pelas médias.

As sequências alinhadas pelo QIIME ao banco de dados do SILVA foram readequadas quanto à classificação taxonômica, pois neste trabalho foi usada a classificação para cianobactéria de Komárek e colaboradores (2014) e Komárek (2016a; 2016b) (HALLENBECK, 2017).

Devido ao ensaio experimental entre as técnicas de DGGE e biblioteca 16S rRNA aplicado as amostras costeiras, dados qualitativos (presença/ausência) foram extraídos. Os dados quantitativos foram identificados a partir da abundância das sequências, tanto para as amostras insulares como costeiras. Os parâmetros ecológicos foram estimados a partir destas sequências. A identificação dos fragmentos se deu por diversos pacotes do software estatístico R em conjunto com comandos do QIIME. Estas plataformas também foram utilizadas para construir clusters com base na abundância relativa em diferentes níveis hierárquicos. Os dados qualitativos não foram normalizados, já os quantitativos sofreram normalização.

A riqueza e a riqueza relativa foram dimensionadas a cada ponto, assim como a riqueza estimada e a curva de rarefação média pelo estimador de riqueza Chao1. A  $\alpha$ -diversidade (ou diversidade local) foi calculada usando o índice de Shannon-Wiener e Simpson. Médias de diversidade e riqueza da costa e arquipélago, a cada fração (3,00 e 0,22 $\mu$ m), foram consideradas como os padrões normais. Desta forma, os valores observados a cada localidade puderam gerar desvios. O índice de Jaccard foi utilizado para avaliar a  $\beta$ -diversidade e assim distribuir as amostras em clusters, resultando em agrupamentos hierárquicos considerando o algoritmo UPGMA (média de valores) sendo avaliado o coeficiente de correlação cofenético (OKSANEN, 2015: 2018). Foram propostas uma PCA e um escalonamento multidimensional não métrico (Non-Metric Multidimensional Scaling - NMDS) utilizando a distribuição de Jaccard com dados binários pelas médias. Os táxons considerados abundantes foram definidos como os que continham mais de 100 sequências/OTU. Os táxons raros foram definidos entre 2 e 10 sequências/OTU (LOGARES et al., 2013). Um heatmap ou ‘mapa de calor’ foi proposto considerando apenas os táxons mais abundantes, transformando os dados em porcentagem das sequências e removendo os gêneros menores que 0,01% de abundância relativa máxima. Uma correlação de Spearman atribuiu a significância estatística a estes táxons selecionados como mais abundantes, demonstrando as forças das relações entre si.

Tabela 12. Listagem das soluções e suas respectivas concentrações e quantidades para formar uma solução tampão ideal para o funcionamento da High Fidelity Taq e anelamento correto dos primers da PCR-NGS 1. A cada amostra são somados 8,7 $\mu$ L do Mix, e portanto, 16,3 $\mu$ L de água para biologia molecular para formar o volume final de solução. Fonte: o próprio autor.

PCR-NGS semi-nexted 1			
Reagents	Stock concentration	Final concentration	Final Volume (25µL)
High Fidelity Taq buffer	10X	1X	2.5µL
MgCl <sub>2</sub>	25mM	1.5mM	1.5µL
CYA359F	5µM	0.2µM	1µL
23S-30R	5µM	0.2µM	1µL
DNTP	10mM	0.2mM	0.5µL
High Fidelity Taq	5V/µL	1V	0.2µL
DNA	10ng/µL	20ng	2µL

Tabela 13. Listagem das soluções e suas respectivas concentrações e quantidades para formar uma solução tampão ideal para o funcionamento da High Fidelity Taq e anelamento correto dos primers da PCR-NGS 1. A cada amostra são somados 7,7µL do Mix, e portanto, 17,3µL de água para biologia molecular para formar o volume final de solução. Fonte: o próprio autor.

PCR-NGS semi-nexted 2			
Reagents	Stock concentration	Final concentration	Final Volume (25µL)
High Fidelity Taq buffer	10X	1X	2.5µL
MgCl <sub>2</sub>	25mM	1.5mM	1.5µL
CYA359F-adapter 1	10µM	0.2µM	0.5µL
CYA781R-adapter 2	10µM	0.2µM	0.5µL
DNTP	10mM	0.2mM	0.5µL
High Fidelity Taq	5V/µL	1V	0.2µL
Amplicon PCR-NGS 1	-	-	2µL

Tabela 14. Listagem das soluções e suas respectivas concentrações e quantidades para formar uma solução tampão ideal para o funcionamento da High Fidelity Taq e anelamento correto dos primers da PCR-Illumina Index. A cada amostra são somados 24,2µL do Mix, e portanto, 25,8µL de água para biologia molecular para formar o volume final de solução. Fonte: o próprio autor.

PCR-Illumina index			
Reagents	Stock concentration	Final concentration	Final Volume (50µL)
High Fidelity Taq buffer	10X	1X	5µL
MgCl <sub>2</sub>	25mM	1.5mM	3µL
Index B	-	-	5µL
Index L	-	-	5µL
DNTP	10mM	0.2mM	1µL
High Fidelity Taq	5V/µL	1V	0.2µL
Amplicon purified	-	-	5µL

Tabela 15. Condições ideais aplicadas ao termociclador durante o processo da PCR, considerando as particularidades físico-químicas do conjunto de primers e da enzima Platinum Taq

DNA High Fidelity durante a PCR-Illumina Index.  
Fonte: o próprio autor.

PCR-Illumina index			
T (°C)	Time (min)	Process	
12 cycles	95	03:00	Initial step of denaturing
	95	00:20	Denaturation
	55	00:30	Annealing
	72	00:30	Extension
	72	05:00	Extension stabilization
	4	∞	Storage

#### 4.6. Análises integradas

A distribuição dos parâmetros físico-químicos da água foi contraposta contra a distribuição das abundâncias das sequências alinhadas pela biblioteca 16S rRNA por meio da análise de redundância (RDA). Desta forma a significância de cada parâmetro foi testada, sendo esta significância confirmada pelo teste de Mantel. Tal teste é utilizado para investigar as correlações entre as duas matrizes de distribuição mencionadas acima. Os dados ambientais foram padronizados pela função ‘decostand’ do software R.

Variáveis descritivas como os tipos de substrato (fundo rochoso, planície arenosa e poça de maré) e a hidrodinâmica local (abrigada ou exposta) foram utilizadas para aferir médias de riqueza observadas e diversidade local ( $\alpha$ -diversidade) pelos índices de Shannon e Simpson, porém estabelecendo estas médias pelas condições ambientais e não pela localidade de amostragem. Uma distribuição de beta-diversidade utilizando a distribuição de Jaccard para as abundâncias de sequências alinhadas também foi proposta. Esta distribuição foi modulada pelas variáveis categóricas da água assim como ponderadas pelos centroides.

Todos os pacotes do software R utilizados diferentes tópicos da metodologia estão devidamente listados nos anexos no final do manuscrito. As linhas de comando a cada análise estatística também estão descritas (ANEXO 5).

### 5. Resultados

O resultados a seguir sempre estarão representados em relação a posição longitudinal, da costa aos arquipélagos oceânicos. No caso das análises genéticas ainda houve o fracionamento nas frações fitoplânctônicas 3,00 $\mu$ m (nano) e 0,22 $\mu$ m (pico).

#### 5.1. Variáveis ambientais

Desvios dos valores observados dos parâmetros físico-químicos em relação às médias costeiras foram estimados para cada localidade. De maneira geral, os parâmetros da água apresentaram médias costeiras próximas às das médias obtidas para os arquipélagos (Tabela 16). Os valores de pH das amostras costeiras tiveram pequenos desvios, com destaque para PP1 com 4,28%. Nas amostras insulares, as localidades de FN apresentaram maiores desvios

positivos no pH do que as localidades de AS, sendo em AS4 o menor valor de pH amostrado, com 7,19 (Tabela 16). Em relação à temperatura as amostras costeiras apresentaram maiores desvios em PA e menores em AC com uma média de 25,54°C. As amostras insulares apresentaram desvios muito próximos em relação a média (Tabela 16). Os desvios positivos de salinidade mais elevados foram em FN e PP. Já os desvios negativos maiores foram em AS e AC, com destaque para AS2 que apresentou uma salinidade estuarina, com 30,41 (Tabela 16). Em relação ao oxigênio dissolvido as localidades costeiras tiveram os maiores desvios positivos nas amostras de AC e IC. PA apresentou os maiores desvios negativos. Nas amostras insulares os desvios foram muito próximos, com destaque para AS1 com 8,20mg/L (Tabela 16). A ANOVA multifatorial revelou uma significativa diferença entre todas as contraposições dos parâmetros amostrados na água ( $p_{ANOVA}=0,0001$ ).

A análise de cluster seguiu a distribuição euclidiana com agrupamento pelas médias, em todos os casos as triplicatas de cada localidade foram agrupadas em conjunto, demonstrando uma fidelização e homogeneidade dos parâmetros da água em cada local (Figura 16). A reação cofenética foi ajustada ( $\text{cof}=0,88$ ) demonstrando uma boa validação do agrupamento. O ponto AS4 foi o mais dissimilar em relação aos demais com uma dissimilaridade de 99%. As amostras costeiras também foram dissimilares ao restante das amostras insulares, com exceção de AC (Figura 16). Embora AC tenha sido altamente similar entre si foi agrupado juntamente com as amostras insulares. FN apresentou um agrupamento coeso entre suas localidades, com massas da água bem misturadas (Figura 16).

A PCA apresentou 4 eixos, sendo que o primeiro foi significativo estatisticamente, e representou sozinho 45,74% das variações dos dados. As variáveis que apresentaram maior correlação com este eixo (axis) foram DO e temperatura (Figura 17). As amostras da costa nordestina se distribuíram principalmente em função da temperatura, as quais tiveram as mais elevadas e coincidentes médias entre si. Enquanto que AC foi contrária a essa variável, apresentando as menores temperaturas (Figura 17 e Tabela 16). As localidades de FN se distribuíram com maior uniformidade, em conjunto, enquanto que as amostras de AS foram bem dissimilares entre si. A MDS confirmou uma dispersão justa das localidades com um estresse de 0,0984 em apenas 2 dimensões. Já a análise de correlação de Spearman confirmou a relação das variáveis, houve uma correlação positiva ( $\text{corSpearman}=0,75$ ) entre salinidade e pH com elevada relevância estatística ( $p_{\text{Spearman}}=0,01$ ). E uma correlação negativa ( $\text{corSpearman}= -0,75$ ) entre DO e temperatura com relevância estatística ( $p_{\text{Spearman}}<0,05$ ), demonstrando um efeito de distribuição contrária das variáveis.

Tabela 16. Desvios dos valores observados dos parâmetros físico-químicos da água durante as amostragens nas diferentes localidades. Os valores azuis correspondem a desvios positivos, enquanto que os valores em vermelho correspondem aos desvios negativos. Fonte: o próprio autor.

		Water Parameters			
		pH	T (°C)	Salinity	DO (mg/L)
Coast	Average	8.32	27.93	35.17	5.32
AC1	Observed	8.12	25.45	33.15	6.10
	Deviation (%)	-2.45	-8.87	-5.75	14.60
AC2	Observed	8.29	25.62	34.21	5.40
	Deviation (%)	-0.41	-8.26	-2.73	1.45
PP1	Observed	8.68	28.77	36.33	4.90
	Deviation (%)	4.28	3.02	3.30	-7.95
PA1	Observed	8.33	29.50	35.67	4.70
	Deviation (%)	0.07	5.63	1.42	-11.71
IC1	Observed	8.12	28.70	35.50	5.80
	Deviation (%)	-2.45	2.77	0.94	8.96
Archipelagos	Average	8.14	27.12	33.98	6.59
AS1	Observed	8.24	27.25	32.49	8.20
	Deviation (%)	1.28	0.50	-4.40	24.48
AS2	Observed	8.26	27.44	30.41	5.30
	Deviation (%)	1.52	1.20	-10.52	-19.54
AS3	Observed	8.18	27.45	33.45	6.70
	Deviation (%)	0.54	1.24	-1.57	1.71
AS4	Observed	7.19	27.14	31.28	5.90
	Deviation (%)	-11.63	0.09	-7.96	-10.44
FN1	Observed	8.29	26.78	36.10	5.90
	Deviation (%)	1.89	-1.24	6.23	-10.44
FN2	Observed	8.50	27.32	36.20	6.80
	Deviation (%)	4.47	0.76	6.52	3.23
FN3	Observed	8.32	26.91	36.14	6.90
	Deviation (%)	2.26	-0.76	6.34	4.74
FN4	Observed	8.11	26.63	35.80	7.00
	Deviation (%)	-0.32	-1.79	5.34	6.26

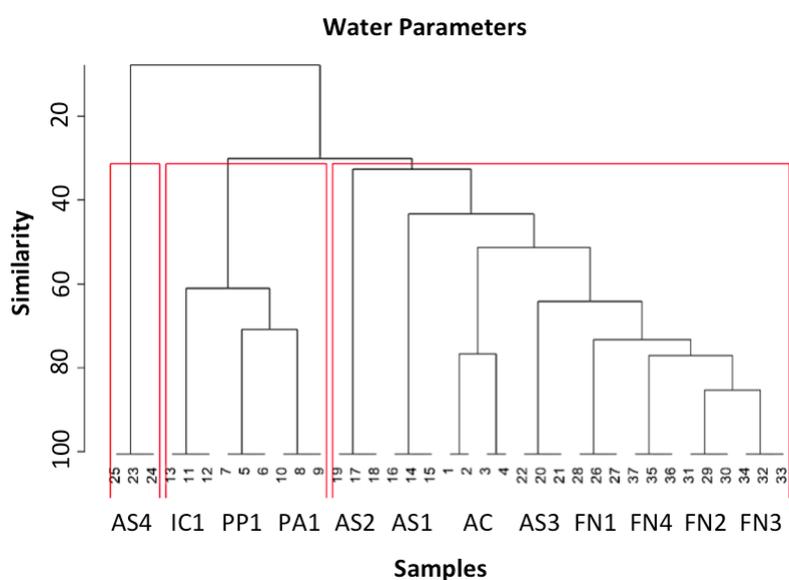


Figura 16. Agrupamentos pelas médias dos parâmetros físico-

químicos de temperatura, salinidade, pH e oxigênio dissolvido. Os agrupamentos foram baseados em matrizes de dissimilaridade seguindo a distância euclidiana. Fonte: o próprio autor.

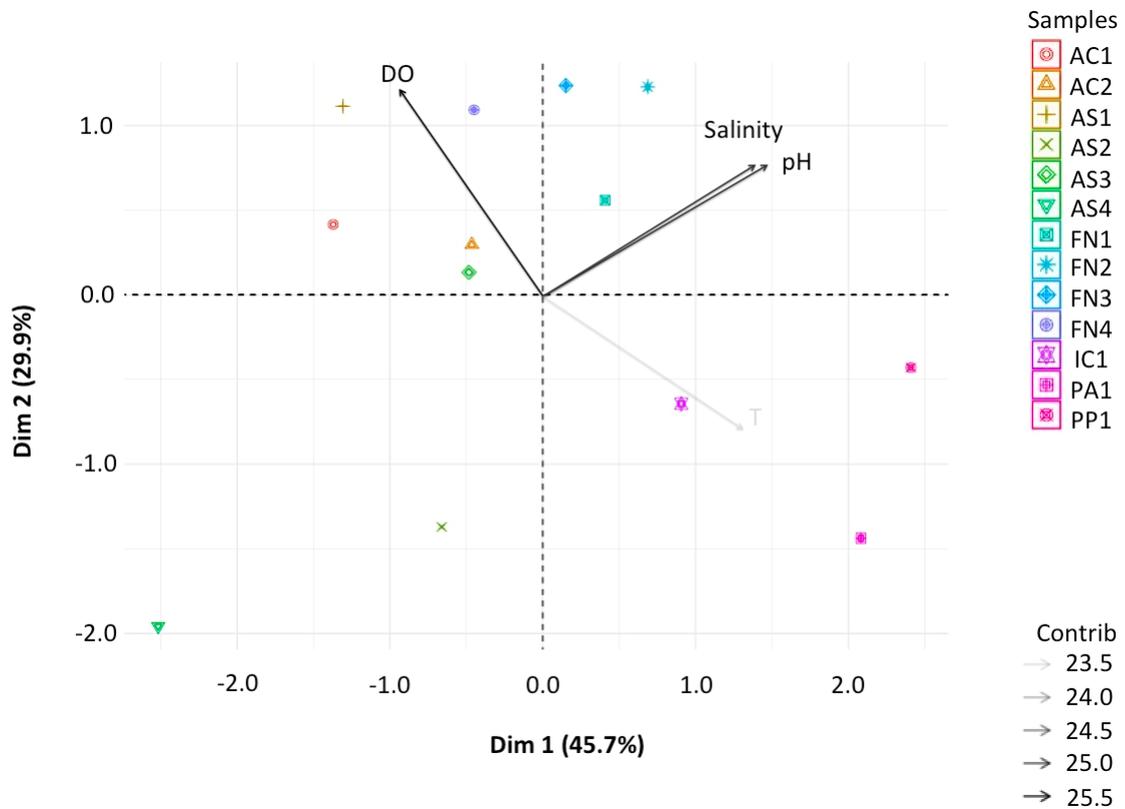


Figura 17. A variação dos pontos seguiu a análise de componentes principais (PCA) considerando quatro parâmetros analisados durante as amostragens na coluna da água: ph; temperatura; salinidade; oxigênio dissolvido. Fonte: o próprio autor.

## 5.2. Biomassa na água

A terminologia POC adotada neste trabalho se refere aos valores de carbono orgânico particulado suspenso na coluna da água extraído pela técnica de digestão ácida, observado no dia da amostragem. O POC foi avaliado tanto nas amostras insulares, quanto na costa. A média de POC das amostras dos arquipélagos ( $31,65 \pm 1,08$ ) foi muito próxima da média das amostras da costa ( $32,96 \pm 0,83$ ) e também da média geral de todas as localidades, com  $32,16 \text{ mg C m}^{-3}$ . O PP apresentou a maior média de POC com  $34,03 \text{ mg C m}^{-3}$ , sendo a menor média de  $31,11 \text{ mg C m}^{-3}$  para as amostras do AS (Figura 18). A análise de variância apresentou uma relevância estatística ( $p_{ANOVA}=0,0343$ ), porém o teste posteriori de Tukey não distinguiu diferença alguma entre as amostras considerando as duplicatas. O agrupamento pelo dendograma com distribuição euclidiana não agrupou de maneira conclusiva os pontos de coleta.

Pequenos desvios foram notados no POC observado nos pontos da costa e nos arquipélagos, quando comparados a suas respectivas médias. Sendo PP1 e FN4 os maiores desvios positivos com 3,26 e 6,07%, respectivamente (Tabela 17).

O carbono orgânico particulado determinado pelo sensor remoto MODIS-Aqua foi mensurado apenas para as amostras insulares - POC remote. A média do AS foi muito próxima à média de FN, com 41,57 ( $\pm 0,17$ ) e 42,11 ( $\pm 1,41$ ) mg C m<sup>-3</sup>, respectivamente. As amostras de FN tiveram maior variação, com FN3 apresentando os maiores valores de POC remote dentre todas as amostras com 44,00mg C m<sup>-3</sup>. Enquanto que FN4 teve os menores valores dentre todas as amostras com 40,94mg C m<sup>-3</sup> (Figura 18). A ANOVA não distinguiu diferença entre as amostras dos arquipélagos considerando as duplicatas dos pixels (pANOVA=0,0599). O teste à posteriori de Tukey indicou uma diferença moderada entre as amostras FN3 e FN2 (pTukey=0,0666), e uma diferença menos expressiva entre FN3 e AT (pTukey=0,0893). Já entre FN3 e FN4 a diferença foi significativa com pTukey=0,0394.

As amostras insulares tiveram mensurados o POC no momento da coleta e ainda médias mensal, anual e uma série histórica (POC remote). O POC observado na coleta foi menor que todas as médias de POC remote, com apenas metade do estoque anual de carbono particulado, desvio de 49,66% para os arquipélagos (Tabela 18). Considerando apenas as localidades do AS, o POC observado tem um desvio negativo de mais de 60% em relação à média anual. Já FN apresenta um desvio negativo constante entre o POC observado e as médias de POC remote (Tabela 18).

O AS apresenta as maiores médias anual e temporal em relação às outras ilhas, com 64,71 e 85,06mg C m<sup>-3</sup>, respectivamente. Isso corresponde ao dobro de carbono particulado nas águas do AS em relação a FN e AT em 2015. E 1,5 vezes mais quando comparados com a média temporal de 5 anos, estabelecida como normal. Para o mês de amostragem AT e FN foram superiores a AS, com 43,87, 42,11 e 41,57mg C m<sup>-3</sup>, respectivamente (Tabela 18). Considerando a série temporal de 5 anos como a normal para os arquipélago, AS apresenta as maiores variações durante o ano de 2015 com um desvio positivo de 31,45%. Já para o mês de amostragem houve um desvio negativo (-35,75%), o período apresentou o POC remote muito inferior ao esperado no AS (Tabela 18). Esse desvio é ainda mais negativo quando a média de outubro de 2015 é comparada a média anual, com -51,12%. Para FN e AT os desvios foram muito diminutos para todos os períodos avaliados.

Para a série temporal de 5 anos, a média de todas as amostras insulares resultou em 52,88mg C m<sup>-3</sup>. A média anual foi ligeiramente superior (62,87mg C m<sup>-3</sup>), já a média mensal foi inferior, com 42,07mg C m<sup>-3</sup> (Tabela 18). Todas as localidades de AS apresentaram expressivos desvios positivos tanto para a média anual, quanto para a série temporal. Para os mesmos períodos, as ilhas de FN e AT apresentaram somente desvios negativos em relação às médias (Tabela 18).

As imagens geradas com base no software Giovanni foram as bases para todas as análises de POC remote, estas imagens expressam os padrões das concentrações de carbono particulado na água em cada período temporal. Em outubro de 2015 a concentração de POC

apresenta-se mais turbulenta e com maiores áreas de baixa concentração - situação oligotrofia (Figura 19A). A média anual de 2015 apresenta valores bem elevados em relação aos outros períodos e demonstra um padrão de maior carbono na água no AS em relação a AT e FN. Esse padrão é melhor definido pela série temporal de 5 anos (Figura 19B e C). O software MatLab (MathWorks®) foi utilizado para adequar as imagens baixadas e realizar os plots pelas médias.

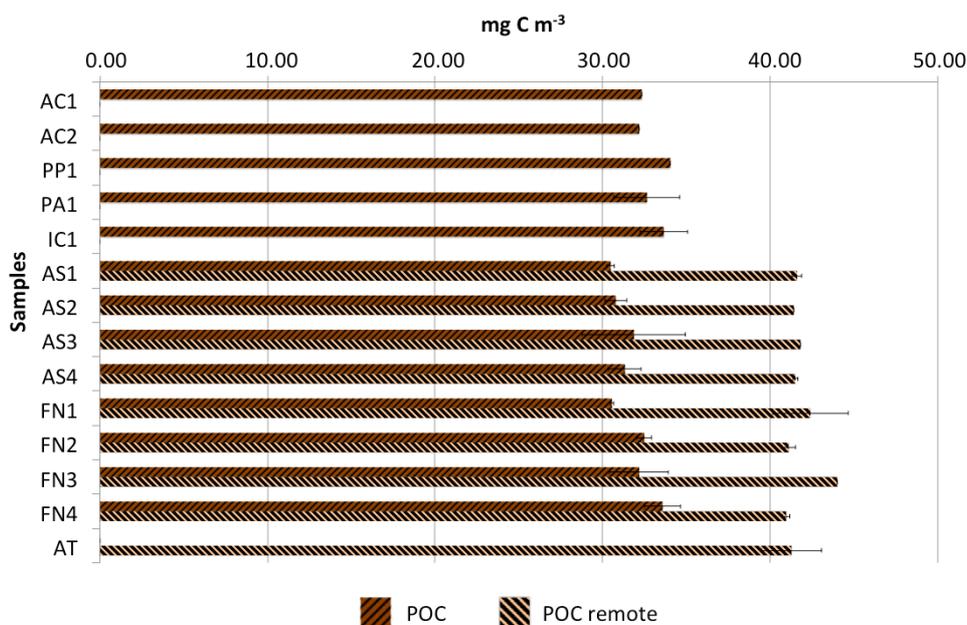


Figura 18. Médias e desvios das réplicas de cada amostra para POC por digestão ácida e POC mensurado pelo sensor remoto MODIS-Aqua, ambas amostras para o período de outubro de 2015. O POC remoto foi mensurado apenas para as amostras dos arquipélagos de Fernando de Noronha - FN, Atol das Rocas - AT e São Pedro e São Paulo - AS. Fonte: o próprio autor.

Tabela 17. Desvios dos valores observados de POC e POC remote em relação às médias temporais consideradas normais para as áreas amostradas na costa e nas porções insulares. Os valores azuis correspondem a desvios positivos, enquanto que os valores em vermelho correspondem aos desvios negativos. Fonte: o próprio autor.

Samples	Analyses	POC remote			POC
		2010-2015 five years	2014-2015 one year	2015 October one month	2015 October sampling day
Coast	Average (mg C m <sup>-3</sup> )				32.96
AC1	Observed (mg C m <sup>-3</sup> )	-	-	-	32.33
	Deviation (%)				-1.91
AC2	Observed (mg C m <sup>-3</sup> )	-	-	-	32.17
	Deviation (%)				-2.38
PP1	Observed (mg C m <sup>-3</sup> )	-	-	-	34.03
	Deviation (%)				3.26
PA1	Observed (mg C m <sup>-3</sup> )	-	-	-	32.64
	Deviation (%)				-0.97
IC1	Observed (mg C m <sup>-3</sup> )	-	-	-	33.65
	Deviation (%)				2.08
Archipelagos	Average (mg C m <sup>-3</sup> )	52.88	62.87	42.07	31.65
AS1	Observed (mg C m <sup>-3</sup> )	65.03	85.95	41.60	30.47
	Deviation (%)	22.98	36.70	-1.11	-3.73
AS2	Observed (mg C m <sup>-3</sup> )	64.70	86.42	41.40	30.78
	Deviation (%)	22.35	37.45	-1.59	-2.75
AS3	Observed (mg C m <sup>-3</sup> )	65.36	85.47	41.80	31.86
	Deviation (%)	23.61	35.95	-0.64	0.68
AS4	Observed (mg C m <sup>-3</sup> )	63.73	82.39	41.50	31.32
	Deviation (%)	20.52	31.05	-1.35	-1.04
FN1	Observed (mg C m <sup>-3</sup> )	44.80	45.82	42.40	30.55
	Deviation (%)	-15.28	-27.11	0.79	-3.49
FN2	Observed (mg C m <sup>-3</sup> )	42.77	45.11	41.10	32.48
	Deviation (%)	-19.12	-28.25	-2.30	2.64
FN3	Observed (mg C m <sup>-3</sup> )	45.08	46.07	44.00	32.17
	Deviation (%)	-14.75	-26.73	4.59	1.66
FN4	Observed (mg C m <sup>-3</sup> )	43.17	45.41	40.95	33.57
	Deviation (%)	-18.37	-27.78	-2.66	6.07
AT	Observed (mg C m <sup>-3</sup> )	41.27	43.20	43.87	-
	Deviation (%)	-21.95	-31.29	4.28	

Note: - not sampled

Tabela 18. Desvios entre as médias temporais de POC e POC remota para cada arquipélago. A série temporal de 5 anos foi considerada a normal. Fonte: o próprio autor.

		Samples				
		Archipelagos	AS	FN	AT	
POC remota	Five years (2010-2015)	Temporal series average (mg C m <sup>-3</sup> )	52.88	64.71	43.95	41.27
	One year (2015)	Year average (mg C m <sup>-3</sup> )	62.87	85.06	45.60	43.20
		Temporal series deviation (%)	18.90	31.45	3.75	4.68
	One month (October 2015)	Month average (mg C m <sup>-3</sup> )	42.07	41.57	42.11	43.87
		Temporal series deviation (%)	-20.44	-35.75	-4.19	6.30
		Year deviation (%)	-33.09	-51.12	-7.65	1.55
POC	Sampling day (October 2015)	POC (mg C m <sup>-3</sup> )	31.65	31.11	32.19	-
		Temporal series deviation (%)	-40.14	-51.92	-26.75	
		Year deviation (%)	-49.66	-63.43	-29.40	
		Month deviation (%)	-24.76	-25.18	-23.55	

Note: - not sampled

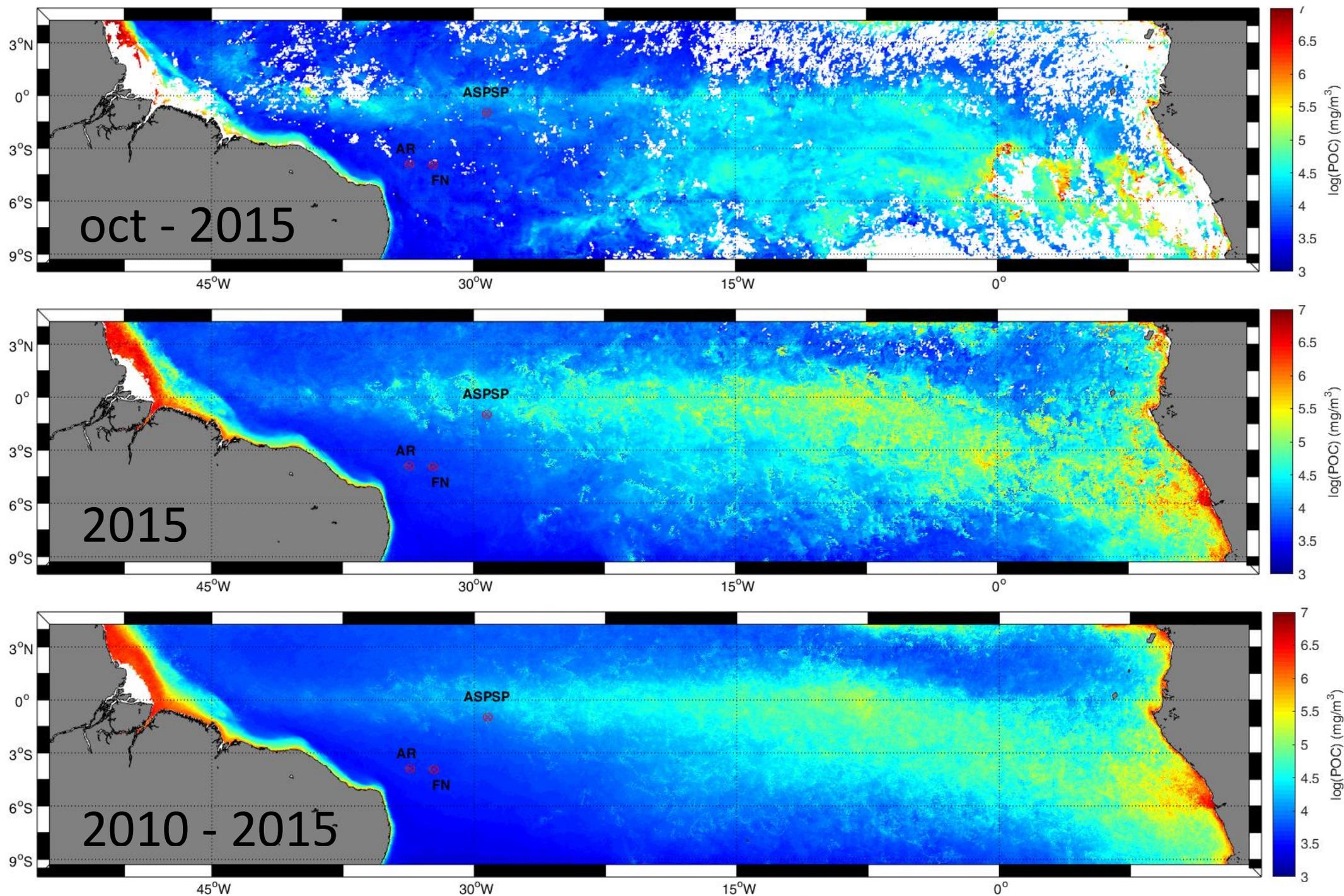


Figura 19. As imagens acima correspondem as concentrações de POC remote via plataforma Giovanni (NASA®) adquiridas pelo sensor orbital MODIS-Aqua. As coordenadas do enquadramento das imagens são 55°O 5°N 15°E 10°S, com cada pixels equivalente a 16Km<sup>2</sup> (4 x 4Km). Cada pixel corresponde: A - Média mensal do mês de outubro de 2105, período de coleta das amostras nos arquipélagos de FN e AS; B - Média anual para o ano de 2105; C - Média da série histórica de 5 anos, de 2010-2015. Fonte: o próprio autor. 74

A análise de correlação de Pearson entre as localidades insulares de POC e POC remota foi negativa ( $\text{corPearson} = -0,1273$ ) com uma baixa relevância estatística ( $\text{pPearson} = 0,6512$ ), demonstrando uma baixa relação entre as amostras observadas durante a coleta e o próprio mês de outubro, avaliado remotamente.

### 5.3. Biomassa das cianobactérias

Considerando todas as populações (nano e picoplâncton - All), a concentração celular das amostras dos arquipélagos foi muito menor ( $5,85 \times 10^6$ ) que a média das amostras da costa ( $4,52 \times 10^7$ ) e também que a média geral de todas as localidades, com  $2,10 \times 10^7$  células por litro. O PP apresentou a maior concentração celular com  $8,24 \times 10^7 \text{ Cell L}^{-1}$ . A menor concentração média com  $3,11 \times 10^6 \text{ Cell L}^{-1}$  ocorreu nas localidades do AS, seguido de FN com concentração média de  $8,58 \times 10^6 \text{ Cell L}^{-1}$  (Figura 20). Dentre as populações isoladas por citometria (costa e ilhas) o grupo Phyto apresentou a maior porcentagem celular (67,51%) do total de células analisados (All) para todos os pontos. Seguido de *Synechococcus* (17,98%), *Prochlorococcus* (14,21%), Cyanochlorophyllates (0,50%) e Nanoeukaryotics (0,08%). A ANOVA multifatorial revelou uma discreta diferença significativa entre todas as populações marcadas por citometria ( $\text{pANOVA} = 0,0614$ ), com exceção do grupo Cyanochlorophyllates que não foi diferenciado como população isolada ( $\text{pANOVA} = 0,6527$ ).

Considerando as populações isoladas por citometria dentro do total de células analisados (All) para as amostras da costa, o padrão observado foi predominante do Phyto. O segundo grupo mais abundante foi *Synechococcus* e depois *Prochlorococcus*. As localidades de AC foram consideradas exceção, sendo que AC1 apresentou uma concentração ligeiramente superior de Prochloro em relação a Syenco, enquanto que AC2 não apresentou células de Prochloro (Figura 21). Em IC as cianobactérias dominaram a amostra (Figura 21). Já para as amostras insulares, o padrão foi diferente em cada arquipélago. As localidades de AS apresentaram dominância do grupo Phyto, seguido por Prochloro e Syenco. Sendo a concentração de Prochloro é de 1,5 a 2X maior que Syenco nas diferentes localidades do arquipélago austral brasileiro (Figura 22). Em AS3 55% da amostra era composta somente por cianobactérias (Figura 22). Para FN, a maior concentração celular também estava distribuída dentro o grupo Phyto, porém as populações de Syenco e Prochloro não se destacaram uma sobre a outra, sendo muito próximas as concentrações (Figura 23). FN3 apresentou uma maior concentração de cianobactérias (Figura 23).

Os dois primeiros eixos da PCA foram significativos estatisticamente, e explicaram 86,97% das variâncias, sendo que o primeiro eixo representou sozinho mais da metade das variações dos dados. As variáveis que apresentam maior correlação com este eixo são as populações de *Synechococcus* e *Prochlorococcus*, com 95% de contribuição. As amostras insulares se distribuíram em função do eixo 1, separando algumas localidades de FN das outras ilhas. Este eixo também isola as amostras insulares das amostras continentais (Figura

24). O eixo 2 está associado, principalmente, com a variação de Cyanochlorophyllates e Phyto, com 95 e 85% de contribuição, respectivamente. Este eixo separa as localidades de IC e agrupa as outras localidades (Figura 24). A menor contribuição foi do grupo Nanoeukaryotics (<80%). Um outro método multivariado de redução de dimensionalidade foi utilizado, a MDS, que comprovou a dispersão das localidades com um estresse excelente de 0,0437 em apenas 2 dimensões.

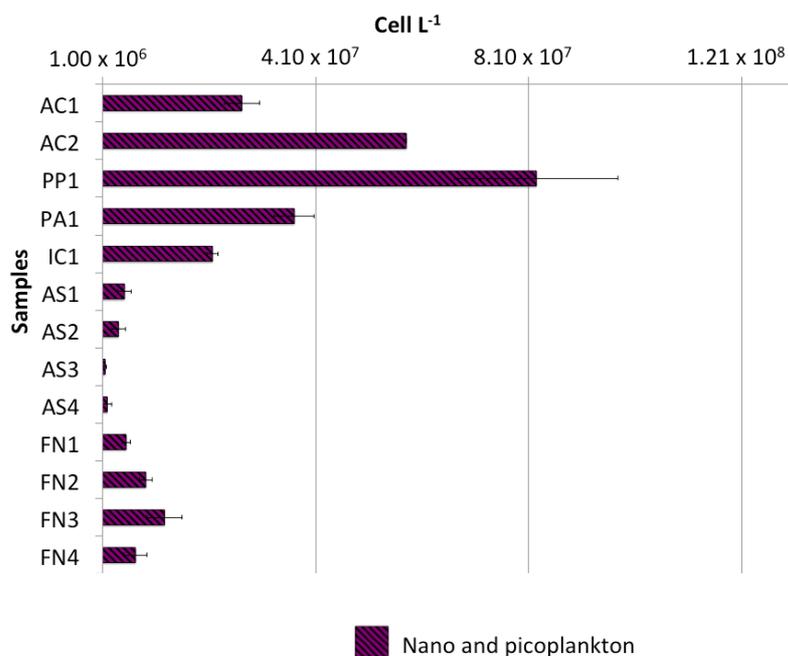


Figura 20. Médias e desvios da concentração celular total (All) das réplicas de cada amostra. As concentrações foram estimadas por marcações celulares adquiridas pela técnica de citometria de fluxo. As faixa de tamanho delimitada pela beads marcadas (2 e 4µm) selecionou células de 0,22 a 6,00µm, correspondendo as frações nano e picofitoplanctônicas. Fonte: o próprio autor.

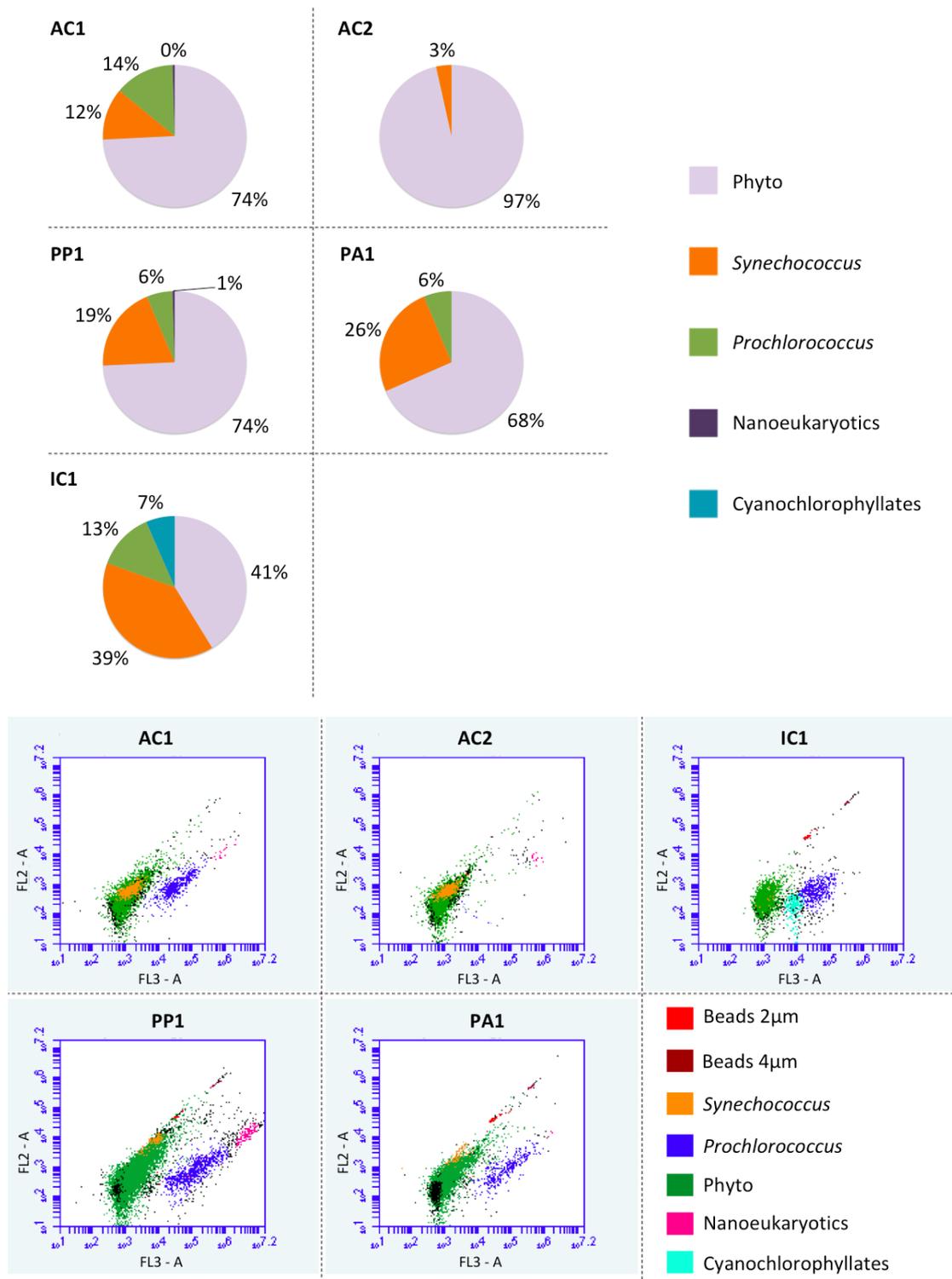


Figura 21. Na parte superior da figura estão dispostas as porcentagens de cada uma das populações fitoplanctônicas identificadas em relação ao somatório total de células. Na parte inferior da figura estão os dotplots gerados pela citometria de fluxo, sendo que, cada ponto corresponde a uma célula distribuída pelos eixos de fluorescência amarelo (FL2) e vermelho (FL3). Fonte: o próprio autor.

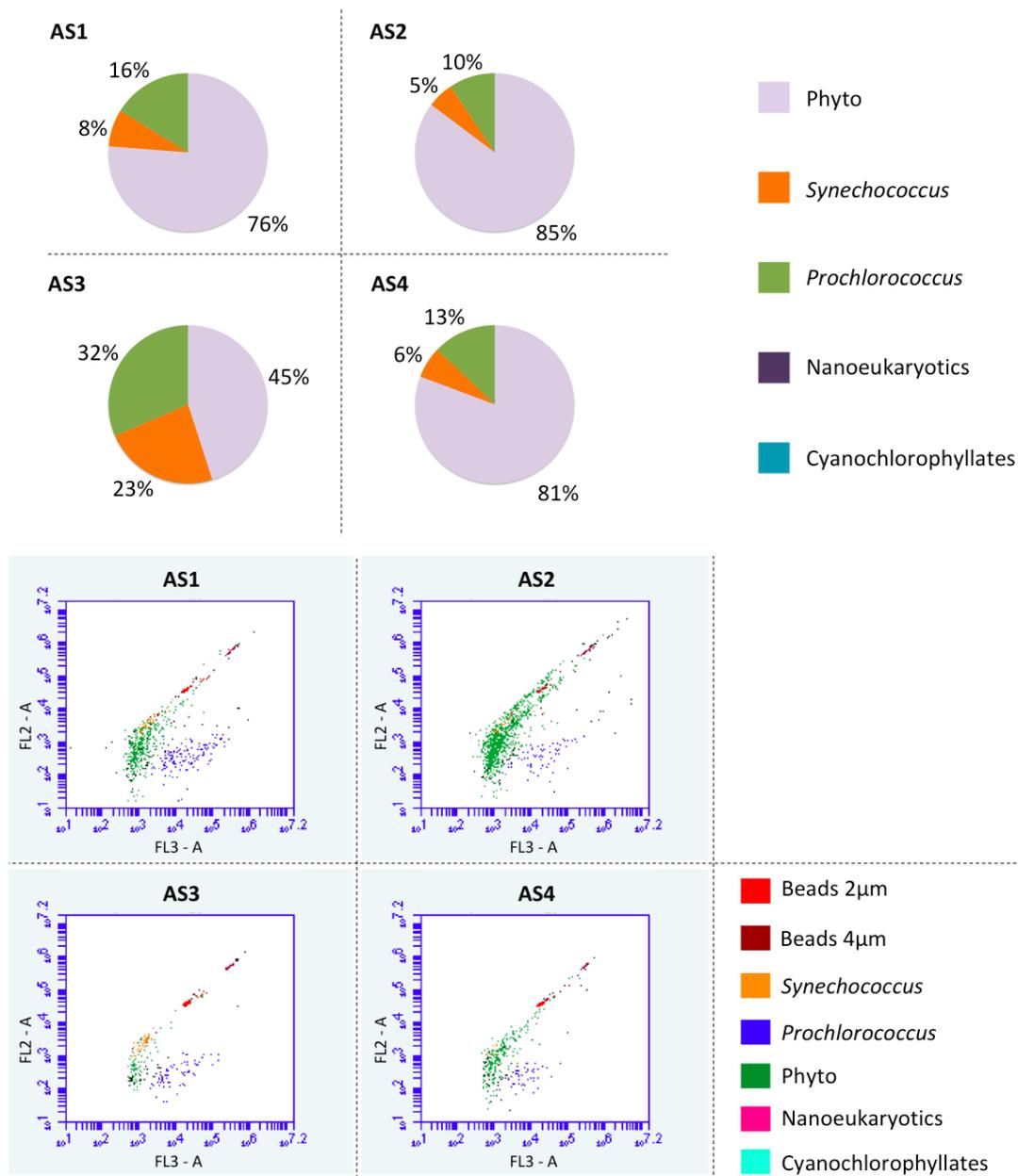


Figura 22. Na parte superior da figura estão dispostas as porcentagens de cada uma das populações fitoplanctônicas identificadas em relação ao somatório total de células. Na parte inferior da figura estão os dotplots gerados pela citometria de fluxo, sendo que, cada ponto corresponde a uma célula distribuída pelos eixos de fluorescência amarelo (FL2) e vermelho (FL3). Fonte: o próprio autor. Fonte: o próprio autor.

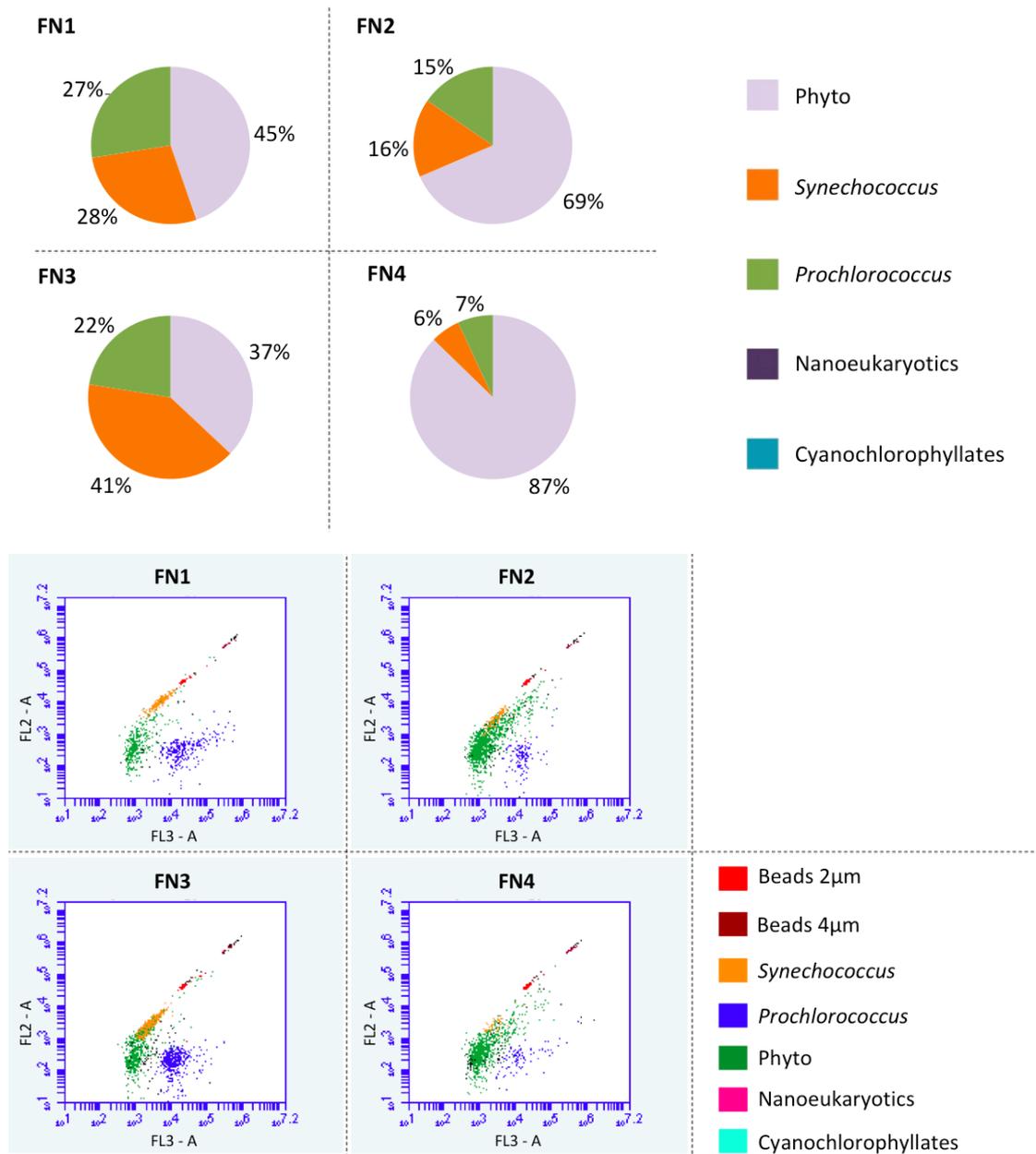


Figura 23. Na parte superior da figura estão dispostas as porcentagens de cada uma das populações fitoplanctônicas identificadas em relação ao somatório total de células. Na parte inferior da figura estão os dotplots gerados pela citometria de fluxo, sendo que, cada ponto corresponde a uma célula distribuída pelos eixos de fluorescência amarelo (FL2) e vermelho (FL3). Fonte: o próprio autor. Fonte: o próprio autor.

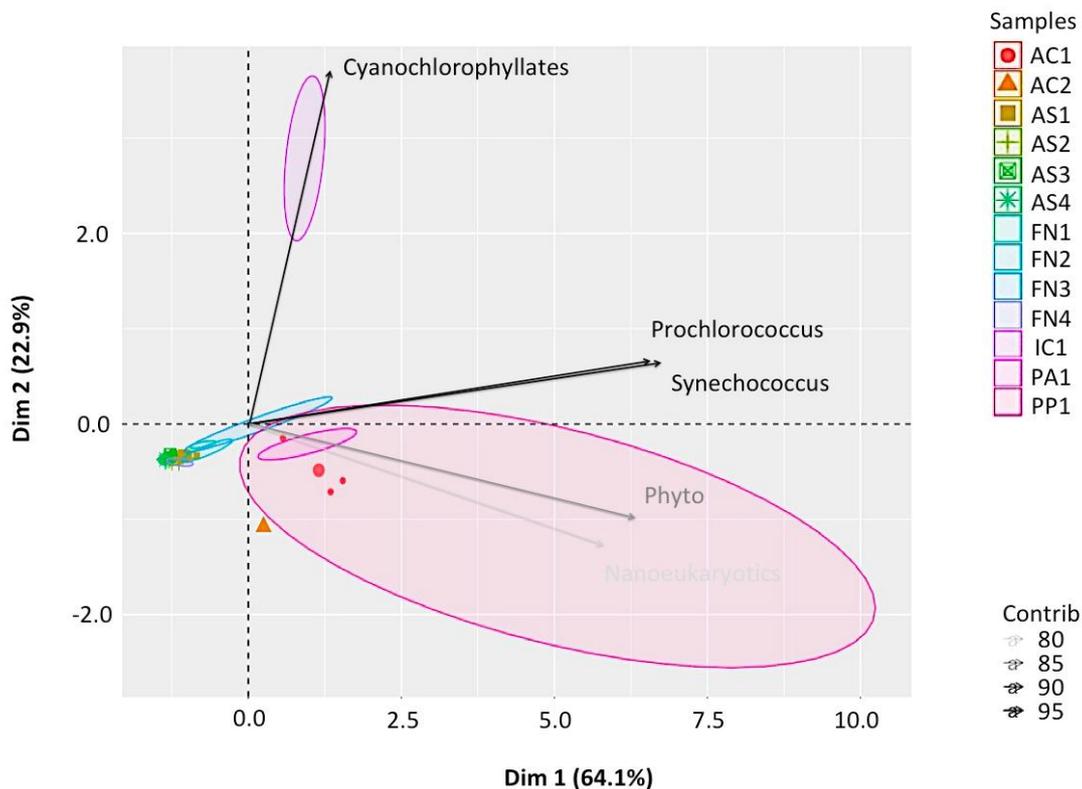


Figura 24. A variação dos pontos seguiu uma análise de componentes principais (PCA) considerando as cinco populações fitoplanctônicas identificadas por citometria: Phyto; *Synechococcus*; *Prochlorococcus*; Nanoeukaryotics; Cyanochlorophyllates. Os pontos insulares foram bem isolados dos pontos da costa. Os círculos agrupam as localidades de cada ponto. Fonte: o próprio autor.

As biomassas dos grupos *Synechococcus* e *Prochlorococcus* foram estimadas e analisadas individualmente e somadas, revelando a biomassa de cianobactérias presentes nas amostras da água (Figura 25).

A biomassa da fração de *Synechococcus* foi proeminente em PP1 (1,8592mg C m<sup>-3</sup>), PA1 (1,0517mg C m<sup>-3</sup>) e IC1 (0,9486mg C m<sup>-3</sup>). As localidades do AS apresentaram as menores concentrações, com média de 0,0271mg C m<sup>-3</sup>. As médias das localidades de AC e FN exibiram uma biomassa semelhante com 0,2912 e 0,2158mg C m<sup>-3</sup>, respectivamente (Figura 25). A biomassa média de *Synechococcus* nas amostras da costa correspondeu a 2,69% de todo carbono orgânico particulado presente na coluna da água nestas localidades. Nos arquipélagos, a biomassa de Syneco correspondeu a apenas 0,38% (Tabela 19).

Para *Prochlorococcus*, PP1 se destacou com a maior biomassa (0,2793mg C m<sup>-3</sup>), já as médias das localidades do AS e FN apresentaram os menores estoques, com 0,0145 e 0,0474mg C m<sup>-3</sup>, respectivamente. AC2 apresentou uma quantidade traço de células de Procloro (Figura 21), resultando numa aparente ausência do estoque de biomassa (Figura 26). A média da biomassa de *Prochlorococcus* nas amostras da costa, assim como a média dos arquipélagos, correspondem a uma baixa porcentagem de carbono estocada dentro de todo o

POC nas respectivas localidades. Sendo que na costa (0,47%) o estoque é ligeiramente maior que nos arquipélagos (0,10%) (Tabela 19).

A somatória das frações Procloro e Syneco corresponde à fração denominada Cyanobacteria. As cianobactérias estocam 3,16% do carbono particulado presente na água costeira do NE brasileiro. Já nos arquipélagos equatoriais da bacia oceânica à frente, as cianobactérias representam apenas 0,48% (Tabela 19). Destaque para o ponto PP1 que representa 6,28% do estoque de carbono da costa. As localidades do AS correspondem à menor biomassa com apenas 0,13% do carbono estocado em cianobactérias (Tabela 19).

Como apenas as amostras insulares possuem as séries históricas de POC remota na coluna da água, as proporções do estoque de carbono convertido em cianobactérias foi limitado apenas aos arquipélagos. FN contribuiu mais ao estoque de ciano no momento da coleta do que suas contribuições mensais e anuais. O AS apresentou um baixo estoque de ciano observado, e este padrão se mantém similar ao longo das médias históricas (Tabela 20). As cianobactérias contribuíram mais ao estoque de carbono na coluna da água no período da coleta com 0,48%, em relação à contribuição anual (0,24%) (Tabela 20).

Uma ANOVA entre estas duas variáveis demonstrou que as variações das biomassas ao longo dos pontos são distintas, com significância estatística ( $p_{ANOVA} < 0,001$ ). A variação dos dados de biomassa foi exibida em apenas duas dimensões pela PCA, com 96,58% da variação no eixo 1, enquanto que o eixo 2 explica o restante de 3,42%. Ambos os eixos estão relacionados às variáveis *Synechococcus* e *Prochlorococcus*, por isso possuem a mesma contribuição de 99,42% ( $p < 0,001$ ). Uma MDS confirmou a redução de dimensionalidade dos dados, distribuindo as biomassas das localidades pelas frações Syneco e Procloro com um estresse ideal menor que 0,0001 em apenas 2 dimensões (Figura 26). A MDS demonstra melhor a maior variabilidade dentro de cada local, sendo AS e IC os pontos de menor variação entre as próprias réplicas. Enquanto que PP e AC apresentaram uma grande distribuição das réplicas. Estas grandes distâncias são justificadas pelos elevados desvios das médias em PP1 (Figura 25). Em FN houve uma separação clara entre as réplicas de FN3 das demais localidades (Figura 26), com um comportamento muito semelhante à distribuição dos pontos pela abundância celular (Figura 23), porém considerando apenas as variáveis Syneco e Procloro. Na distribuição da concentração celular haviam outras populações fitoplanctônicas, sendo as populações de cianobactérias (*Synechococcus* e *Prochlorococcus*) as de maior peso.

Uma correlação de Pearson também revelou uma elevada similaridade entre ambas populações ( $cor_{Pearson} = 0,9316$ ) com uma robusta relevância estatística ( $p_{Pearson} < 0,0001$ ), variando com o mesmo padrão.

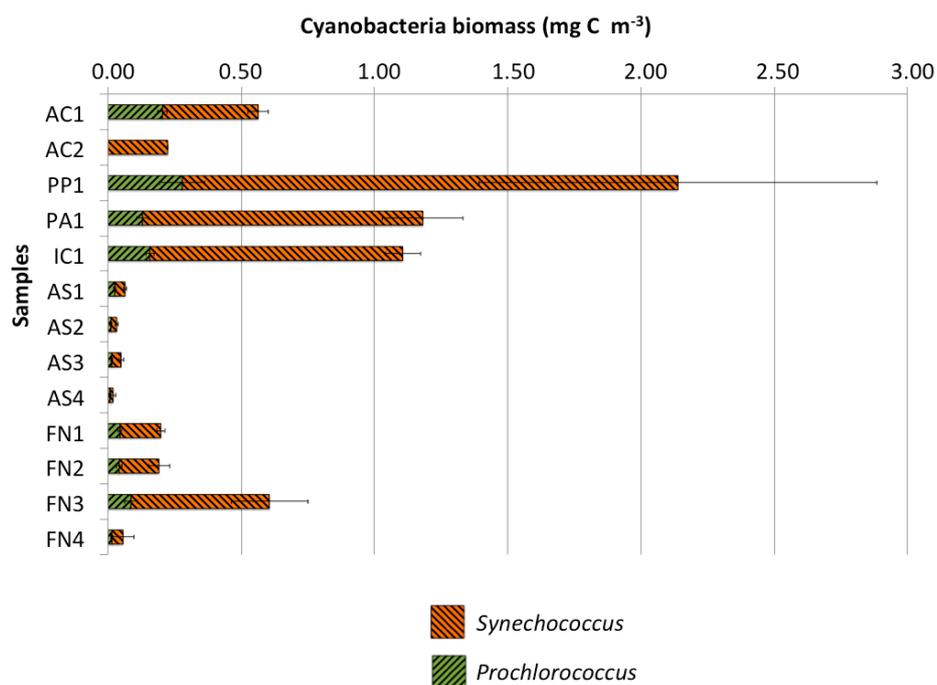


Figura 25. Médias e desvios das réplicas de cada ponto para biomassa mensurada através da concentração celular e índices de biovolume. As barras indicam os valores reais de biomassa das populações de cianobactérias. Fonte: o próprio autor.

Tabela 19. Porcentagens da biomassa acumulada nas populações de cianobactérias em relação ao POC observado na coluna da água referente a cada local (estoque relativo). O POC do ambiente costeiro e dos arquipélagos foi estipulado pelas médias das respectivas localidades. Fonte: o próprio autor.

		Samples							
		Coast	Archipelagos	AC	PP	PA	IC	AS	FN
<i>Synechococcus</i>	POC (mg C m <sup>-3</sup> )	32.96	31.65	32.25	34.03	32.64	33.65	31.11	32.19
	Biomass (mg C m <sup>-3</sup> )	0.89	0.12	0.29	1.86	1.05	0.95	0.03	0.22
	%	2.69	0.38	0.90	5.46	3.22	2.82	0.09	0.67
<i>Prochlorococcus</i>	POC (mg C m <sup>-3</sup> )	32.96	31.65	32.25	34.03	32.64	33.65	31.11	32.19
	Biomass (mg C m <sup>-3</sup> )	0.15	0.03	0.10	0.28	0.13	0.16	0.01	0.05
	%	0.47	0.10	0.32	0.82	0.40	0.47	0.05	0.15
<i>Cyanobacteria</i>	POC (mg C m <sup>-3</sup> )	32.96	31.65	32.25	34.03	32.64	33.65	31.11	32.19
	Biomass (mg C m <sup>-3</sup> )	1.04	0.15	0.39	2.14	1.18	1.11	0.04	0.26
	%	3.16	0.48	1.22	6.28	3.62	3.29	0.13	0.82

Tabela 20. Porcentagens da biomassa observada de cianobactérias entre as médias temporais de POC e POC remota para cada arquipélago. A biomassa média dos arquipélagos foi estipulado pelas médias das localidades de AS e FN. Fonte: o próprio autor.

	Cyanobacteria		
	Archipelagos	AS	FN
Temporal series average (mg C m <sup>-3</sup> )	52.88	64.71	43.95
Biomass (mg C m <sup>-3</sup> )	0.15	0.04	0.26
%	0.28	0.06	0.60
Year average (mg C m <sup>-3</sup> )	62.87	85.06	45.60
Biomass (mg C m <sup>-3</sup> )	0.15	0.04	0.26
%	0.24	0.05	0.58
Month average (mg C m <sup>-3</sup> )	42.07	41.57	42.11
Biomass (mg C m <sup>-3</sup> )	0.15	0.04	0.26
%	0.36	0.10	0.63
Observed POC (mg C m <sup>-3</sup> )	31.65	31.11	32.19
Biomass (mg C m <sup>-3</sup> )	0.15	0.04	0.26
%	0.48	0.13	0.82

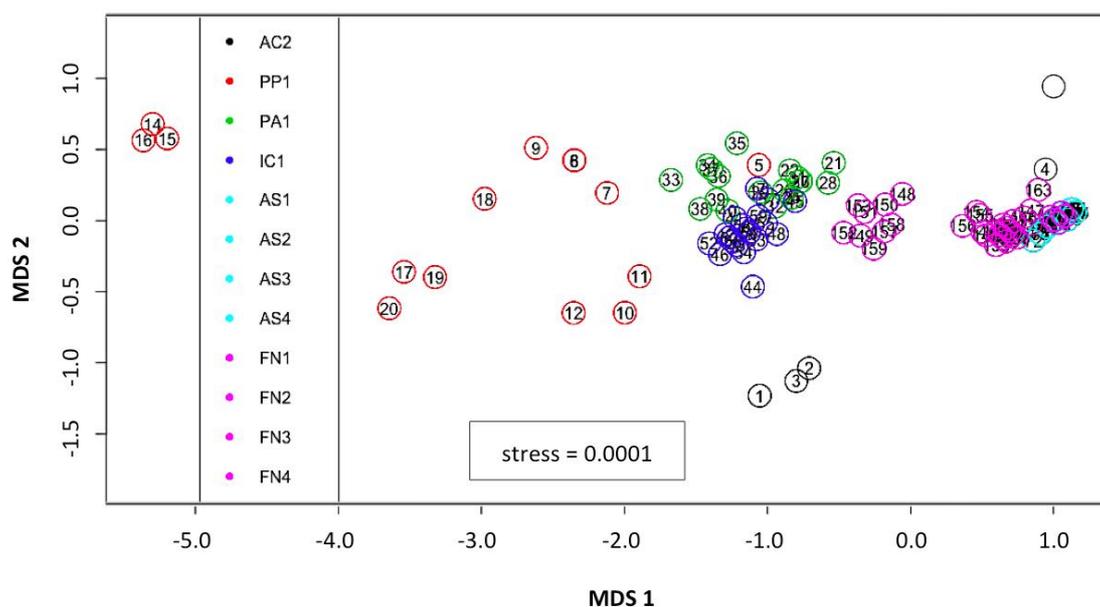


Figura 26. O escalonamento multidimensional (MDS) analisou a distribuição dos pontos em função das populações de cianobactérias. O baixo estresse revelou que a dimensionalidade dos pontos foi representada de maneira ideal. Fonte: o próprio autor.

#### 5.4. Estrutura da comunidade

Avaliações ecológicas das amostras costeiras foram comparadas pelas distintas metodologias (DGGE e sequenciamento) empregadas para estudo da ecologia de cianobactérias na coluna da água. As sequências alinhadas extraídas do sequenciamento foram analisadas quanto aos dados brutos e também extraídos índices ecológicas para costa e arquipélagos.

#### 5.4.1. Estudo preliminar - costa do CE

A análise dos parâmetros ecológicos ao estudo preliminar considerou as frações 3 e 0,22 $\mu$ m em separado, tanto para as técnicas de DGGE quanto para as sequências alinhadas. Portanto 4 possíveis cenários experimentais serão constantemente comparados: amostras da fração do nanofitoplâncton avaliadas pela técnica de DGGE (DGGE - 3,00 $\mu$ m); Sequenciamento - 3,00 $\mu$ m; DGGE - do picofitoplâncton; Sequenciamento - 0,22 $\mu$ m.

Considerando a técnica de DGGE, a fração nanofitoplanctônica apresentou uma riqueza total observada de 22 OTUs (bandas), sendo que a média das amostras costeiras foi de 8,08( $\pm$ 3,57) (Figura 27). As poças de maré do Paracuru, na costa cearense, apresentaram a maior média com 12 bandas identificadas, 53,03% da riqueza total observada. Já AC apresentou a menor média, com apenas 3 bandas (14,77%). Uma estimativa de riqueza utilizando o método binário Chao2 registrou 23 bandas, valor muito próximo da riqueza total encontrada. As análises de sequenciamento através de biblioteca 16S rRNA encontraram uma riqueza observada e estimada muito superior compradas as riquezas encontradas pela DGGE. Pelo sequenciamento foram registrados 39 táxons (Chao2 = 46 OTUs estimados), com uma média costeira de 14,31( $\pm$ 6,30). Assim como encontrado na DGGE, as poças do Paracuru apresentaram a maior riqueza observada, com 37,33% (19 táxons) da riqueza observada. PP apresentou a menor média com apenas 17,33% (Figura 27).

A fração picofitoplacntônica apresentou uma riqueza costeira de apenas 18 bandas pela técnica da DGGE, e uma média de 10,15( $\pm$ 5,26) por amostra. A riqueza estimada foi igual à riqueza observada pelo método de rarefação Chao2. O Porto do Pecém apresentou 88,89% do total de bandas observado, uma média de 16 bandas registradas. AC apresentou a menor média com apenas 16,67% (3 bandas). Pelo sequenciador o padrão foi bem diferente, pois o PP foi amostra de menor riqueza, com apenas 11,63% (5 OTUs) de um total de 34 táxons observados (Figura 27). Pelo método Chao2 foram estimados 42 OTUs.

Considerando as réplicas de cada amostra, uma curva de acumulação de OTUs pelo método da rarefação pelo índice Chao2 foi ajustada, para ambas as técnicas e frações do fitoplâncton. Pela técnica da DGGE houve uma baixa resolução da cobertura de espécies em relação às amostras sequenciadas. Na fração DGGE-0,22 $\mu$ m a riqueza das microalgas atingiu o total observado em 50% dos quartis de amostragem, estabilizando a curva como se não houvessem mais espécies raras a serem reveladas ao picofitoplâncton das amostras costeiras (Figura 28). Já pela técnica do sequenciamento 16S rRNA, o picoplâncton não apresentou estabilização da curva, revelando uma riqueza estimada superior ao que pode ser amostrado (Figura 28). Através da biblioteca de amplicons, o picofitoplâncton apresentou uma riqueza estimada 2,3x superior às mesmas amostras avaliadas pela DGGE.

As OTUs consideradas como raras estão presentes em apenas um ponto, sem compartilhamento com outra localidade. Lembrando que são dados de presença e ausência.

Arraial do Cabo e Paracuru apresentaram o maior número de “espécies raras” dentre todas as avaliações metodológicas. Para a DGGE, todos os pontos apresentaram ao menos uma OTUs rara na fração 3,00 $\mu$ m, sendo que nenhuma apresentou na fração 0,22 $\mu$ m. As amostras sequenciadas apresentaram OTUs raros na fração 3,00 $\mu$ m apenas em AC (2 OTUs) e PA (1 OTU). E na fração 0,22 $\mu$ m 5 OTUs raros em AC (2 OTUs - AC1; 2 - AC2; 1 - AC1 e AC2) e mais 6 em PA.

As diversidades de Shannon e Simpson apresentaram os mesmos padrões, porém cada índice respeitando a sua escala. No entanto, as diferenças foram melhor elucidadas pelo índice de Shannon-Wiener (Tabela 21). Pela técnica da DGGE a diversidade foi menor nas amostras de AC em ambas frações (3 e 0,22 $\mu$ m). E as maiores médias foram nas amostras de PA e PP nas frações 3 e 0,22 $\mu$ m, respectivamente. As mesmas amostras analisadas pelo sequenciador apresentaram as maiores diversidades em AC e PA e menor  $\alpha$ -div para o PP, nas frações do pico e do nanofitoplâncton (Tabela 21).

Para analisar a diversidade entre os distintos pontos que caracterizaram as amostras costeiras, a  $\beta$ -diversidade foi expressa pela média e por clusters de similaridade entre os pontos.

Para a fração 3 $\mu$ m a média de beta diversidade foi maior para a DGGE com 0,70 quando comparado à média de 0,60 do sequenciamento, utilizando o índice de Jaccard. As duas técnicas moleculares separaram as amostras de AC das outras amostras costeiras, com exceção de uma réplica de AC2 que ficou agrupada com PA1 nas amostras sequenciadas (Figura 29). Já para o picofitoplâncton a média de beta diversidade encontrada pelo sequenciamento foi superior a DGGE, como 0,65 e 0,53 respectivamente. Tais amostras da fração 0,2 $\mu$ m tiveram o mesmo comportamento, separando as amostras de AC das outras de maneira ainda mais clara, com menos de 10% de similaridade. Paracuru também se mostrou distinta das outras amostras cearenses (Figura 29). As correlações cofenéticas e os estresses das NMDS foram satisfatórios validando os agrupamentos em todos os casos.

De maneira geral para ambas as técnicas não houve uma diferença muito grande entre os as médias ecológicas encontradas entre as frações do pico e do nanofitoplâncton para as amostras costeira do atlântico oeste. Porém a dominância dos padrões foi inversa. No caso da DGGE a média da riqueza do nanofito foi inferior em relação ao pico, ocorrendo o mesmo com a diversidade (Tabela 21). Para o sequenciamento, os índices de alfa diversidade e riqueza foram superiores na fração do nanofito (Figura 29).

Os desvios dos valores observados em cada ponto, em relação à média costeira, apresentaram valores mais elevados na técnica de DGGE, em relação à técnica de sequenciamento (Tabela 21). Para a porção do nanofito, apenas o ponto AC1 apresentou desvios negativos em relação às médias, em ambas as metodologias. Os maiores desvios foram encontrados no parâmetro da riqueza observada, o qual a técnica de DGGE demonstrou

uma diferença negativa de -62,86% e o sequenciamento de -33,60% (Tabela 21). O ponto PA1 apresentou somente desvios positivos nas duas técnicas, exibindo a maior diversidade pela DGGE e a segunda maior pelo sequenciamento. Pelo sequenciamento a maior diversidade foi encontrada em AC2, sendo que esta mesma amostra analisada pela DGGE apresentou somente desvios negativos (Tabela 21). Para o picofito o ponto PA1 também apresentou desvios positivos em relação às respectivas médias em todos os parâmetros ecológicos. Para o pico fito AC2 apresentou desvios negativos em todos os índices ecológicos nas duas metodologias, com exceção da alfa diversidade de Simpson nas amostra sequenciadas que apresentou um leve desvio positivo (1,37%) (Tabela 21).

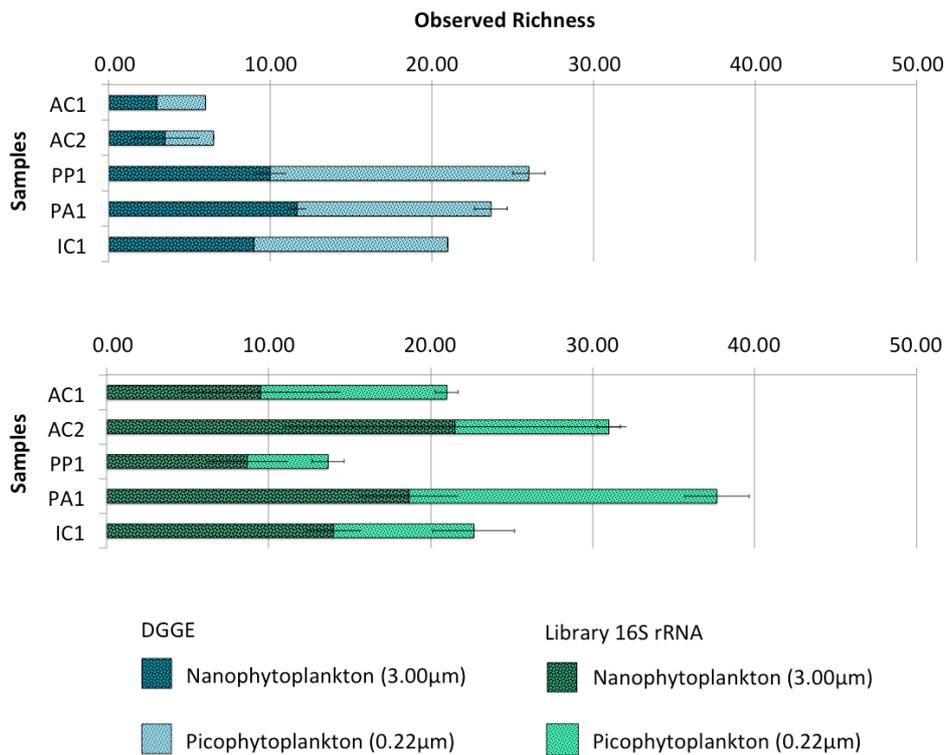


Figura 27. Médias e desvios das réplicas a cada ponto para a riqueza observada para as frações do pico e nanofitoplâncton. Na porção superior estão representadas a riqueza provinda dos dados de DGGE, abaixo, a riqueza dos dados de sequenciamento de amplicons 16S rRNA. Fonte: o próprio autor.

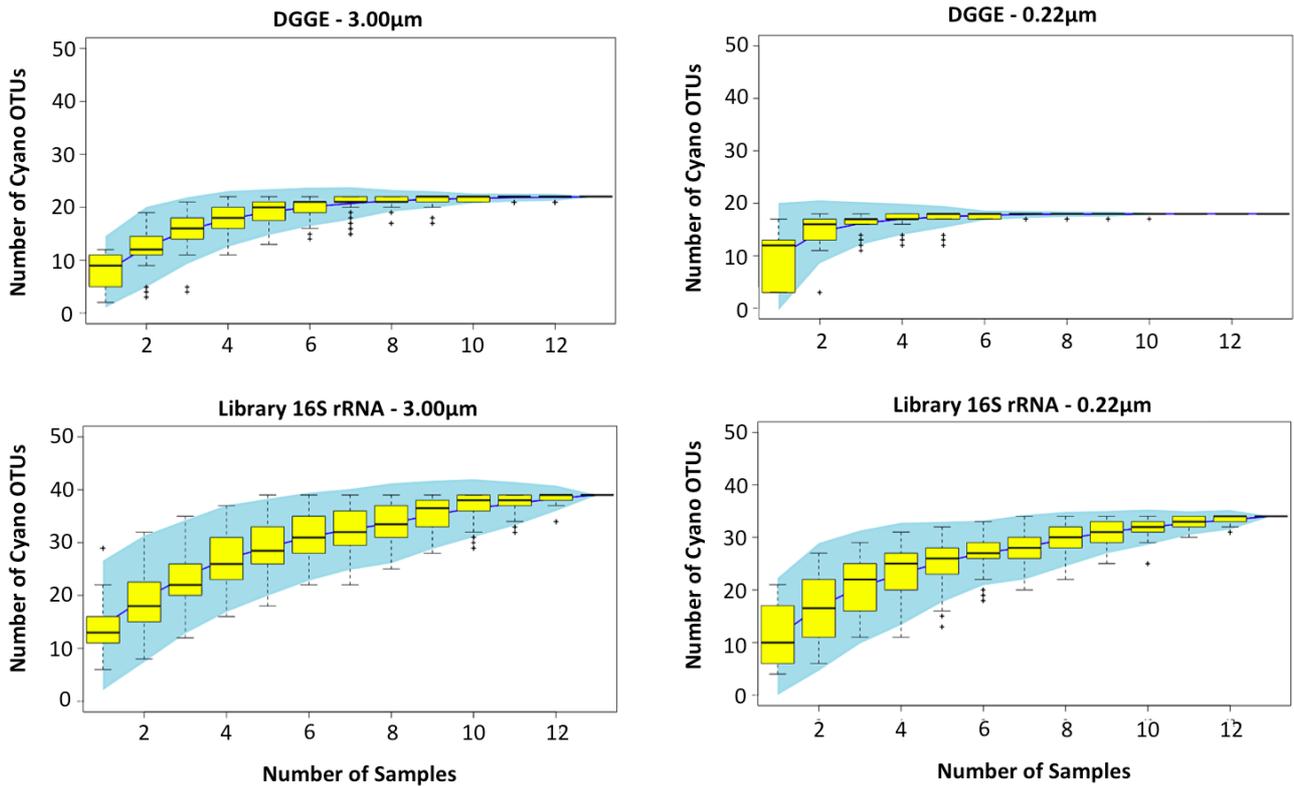


Figura 28. Curvas de rarefação de “espécies” considerando o índice de riqueza estimada Chao2 - presença/ausência. As replicas dos pontos de amostragem (13 replicas) representam as repetições de amostragem ao ambiente delimitado, no caso a costa. As barras amarelas indicam os desvios das replicas a cada ponto, enquanto que os traços pretos representam a media da riqueza total acumulada ao longo das amostragens. Os limites tracejados nas barras indicam a extrapolação para espécies estimadas. A área azul indica o ambiente de rarefação, contemplando os limites os máximos e mínimos da riqueza no ambiente costeiro estudado. Fonte: o próprio autor.

Tabela 21. Desvios dos valores observados em relação à média da correspondente fração fitoplanctônica. Cada parâmetro ecológico possui um valor estipulado tanto pela técnica molecular da DGGE quanto para a técnica do sequenciamento. Os valores azuis correspondem a desvios positivos, enquanto que os valores em vermelho correspondem aos desvios negativos. Fonte: o próprio autor.

		Observed Richness		$\alpha$ -Div Shannon		$\alpha$ -Div Simpson		
		DGGE	Library 16s rRNA	DGGE	Library 16s rRNA	DGGE	Library 16s rRNA	
Nanophytoplankton	3.00 $\mu$ m	Average Coast	8.08	14.31	1.951	2.570	0.826	0.916
	AC1	Observed	3	10	1.099	2.178	0.667	0.878
		Deviation (%)	-62.86	-33.60	-43.68	-15.23	-19.31	-4.08
	AC2	Observed	4	22	1.151	3.003	0.650	0.947
		Deviation (%)	-56.67	50.27	-40.98	16.87	-21.33	3.44
	PP1	Observed	10	9	2.299	2.129	0.899	0.877
		Deviation (%)	23.81	-39.43	17.88	-17.15	8.85	-4.20
	PA1	Observed	12	19	2.456	2.918	0.914	0.945
		Deviation (%)	44.44	30.47	25.91	13.55	10.65	3.27
	IC1	Observed	9	14	2.197	2.634	0.889	0.928
Deviation (%)		11.43	-2.15	12.65	2.51	7.59	1.35	
Picophytoplankton	0.22 $\mu$ m	Average Coast	10.15	10.77	2.124	2.260	0.844	0.882
	AC1	Observed	3	12	1.099	2.441	0.667	0.913
		Deviation (%)	-70.45	6.79	-48.27	8.03	-21.05	3.46
	AC2	Observed	3	10	1.099	2.250	0.667	0.894
		Deviation (%)	-70.45	-11.79	-48.27	-0.44	-21.05	1.37
	PP1	Observed	16	5	2.771	1.596	0.937	0.794
		Deviation (%)	57.58	-53.57	30.48	-29.39	11.00	-9.96
	PA1	Observed	12	19	2.483	2.941	0.916	0.947
		Deviation (%)	18.18	76.43	16.89	30.12	8.51	7.33
	IC1	Observed	12	9	2.485	2.129	0.917	0.877
Deviation (%)		18.18	-19.52	17.00	-5.80	8.56	-0.59	

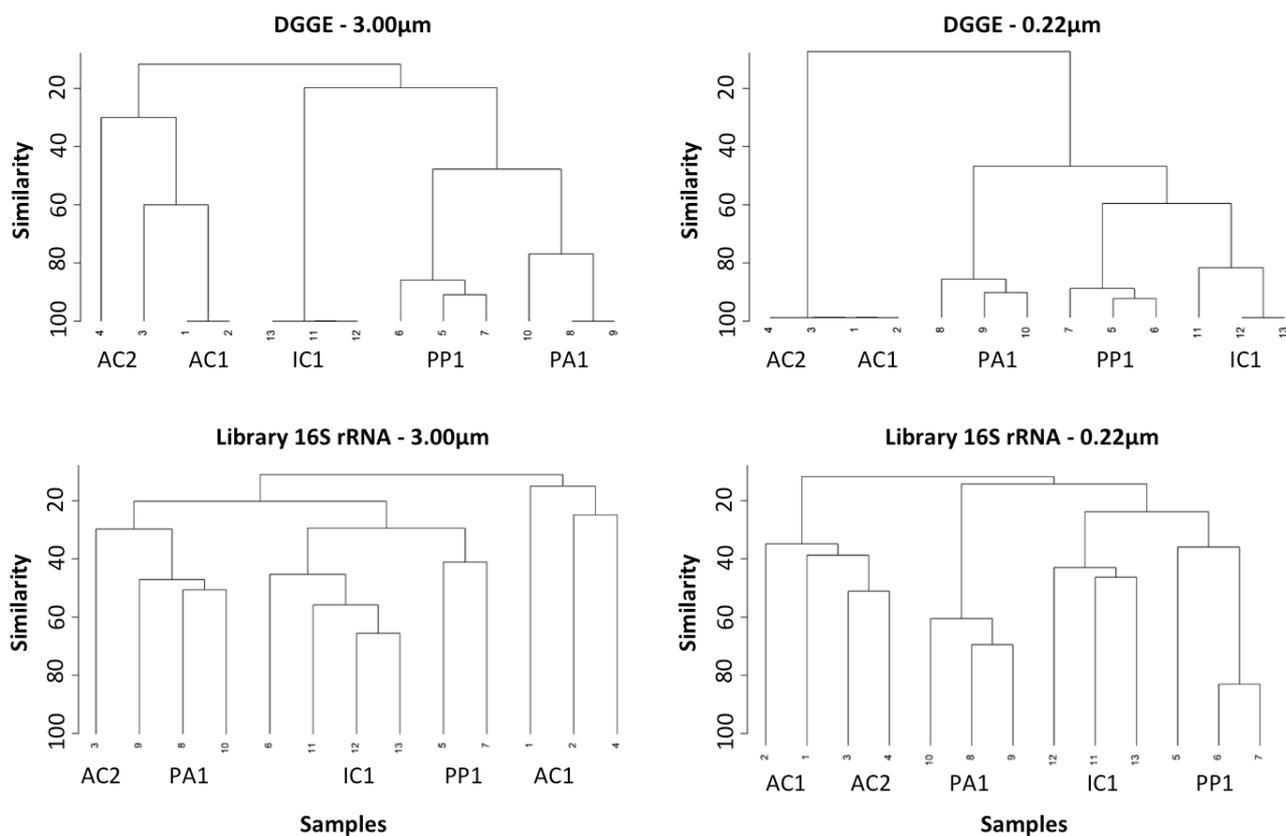


Figura 29. Agrupamentos pelas medias de beta diversidade das amostras costeiras. Os agrupamentos foram baseados em matrizes de dissimilaridade pelo índice binário de Jaccard. Fonte: o próprio autor.

#### 5.4.2. Sequencias alinhadas - costa e ilhas

Os controles internos do sequenciador MiSeq foram satisfatórios durante a corrida, com um índice de qualidade Q30 aos 70%. A soma de todas sequências alinhadas (nano e picoplâncton - All) registrou 1.158.901 sequências (fragmentos 16S rRNA). As amostras dos arquipélagos somaram 73,48% do total de sequências, sendo que as costeiras somaram os outros 26,52%. O AS e FN tiveram as maiores porcentagens do total, com 45,63% e 27,86%, respectivamente. O PP apresentou a maior porcentagem dentre as amostras costeiras com 9,14%, enquanto que AC apenas 5,18% do total de sequências. A média de sequências por amostra foi de aproximadamente 15 mil sequências ( $15.661 \pm 9.295$ ). O tamanho médio dos fragmentos foi de 240pb.

As amostras costeiras e insulares da porção nanofitoplanctônica somaram 467.497 sequências (40,34% do All). Considerando apenas as sequências da fração 3,00µm, aquelas alinhadas dos arquipélagos foram 2,3 vezes mais representativas do que as costeiras. As amostras do AS foram responsáveis por quase 50% do total, com destaque pra AS1 (Figura 30). O menor número de sequências alinhadas foi em AC, e a localidade AC1 teve menor contribuição com apenas 1,66% da somatória da fração 3,00µm (Figura 30). As variações das sequências da fração 3,0µm ao longo dos táxons observados foram expostas em 9 dimensões

para acumular 80% de variância pela PCA, sendo inconclusiva graficamente para demonstrar a importância de algum táxon em especial. A MDS propicia a redução de dimensionalidade dos dados, no caso seguindo a distribuição euclidiana. Um estresse de distribuição de 0,0905 foi encontrado em apenas 2 dimensões, representando uma distribuição considerada justa a todas as amostras de cada local (Figura 31). As réplicas mais distribuídas dos demais e ao mesmo tempo similares entre si corresponderam aos pontos FN4, PA1 e AC2 (uma única réplica - n° 3). As amostras de AS e as outras amostras costeiras foram menos individualizadas, apresentando abundâncias de sequências mais próximas aos mesmos táxons (Figura 31).

Analisando apenas o gênero *Synechococcus* em cada uma das localidades, e apenas a fração 3,00µm, o maior número médio de sequências foi encontrado nas amostras costeiras. O PP obteve a maior média com aproximadamente 20.000 sequências. Uma ANOVA (pTukey<0,05) distinguiu PP1 de todas as localidades insulares. Para o gênero *Prochlorococcus* as localidades insulares foram mais proeminentes, sendo as amostras do AS mais abundantes. A amostra AS1 apresentou uma média de aproximadamente 30.000 sequências. Uma ANOVA (pTukey<0,05) entre os pontos deste gênero diferenciou AS1 de todas as outras amostras, exceto AS3 que apresentou a segunda maior abundância com aproximadamente 18.000 sequências em média. Uma ANOVA bifatorial comparou ambos os gêneros demonstrando uma clara diferença entre eles (pANOVA=0,0001).

A porção picofitoplanctônica somou 691.404 sequências (59,66% do All). Considerando apenas esta fração como o total, as sequências alinhadas dos arquipélagos foram representativas mais de duas vezes em relação às costeiras. Assim como na fração 3,00µm, as médias das localidades do AS foram responsáveis por quase 50% do total, com destaque para AS2 (Figura 30). O menor número de sequências alinhadas foi em AC, e a localidade AC2 apresentou a menor contribuição, com apenas 1,63% da somatória da fração 0,22µm (Figura 30). Assim como para o nanofito, as sequências da fração picofitoplanctônica foram expostas em 9 dimensões para acumular 80% de variância pela PCA, não concluindo nem demonstrando de maneira gráfica a importância de algum táxon em especial. A MDS seguindo a distribuição euclidiana apresentou um estresse de distribuição de 0,0689, em apenas 2 dimensões, sendo considerada satisfatória (Figura 32). As réplicas mais dissimilares das demais e ao mesmo tempo similares entre si, corresponderam aos pontos FN4, PA e AC. As amostras insulares foram menos individualizadas, apresentando abundâncias similares aos mesmos gêneros (Figura 32).

O gênero *Synechococcus* presente na fração 0,22µm demonstrou maiores médias de sequências nas amostras costeiras, assim como na fração 3,00µm. O PP obteve a maior média, seguida de IC com aproximadamente 18.000 e 16.000 sequências, respectivamente. Uma ANOVA (pTukey<0,05) distinguiu as amostras insulares de todas as amostras costeiras,

porém as amostras insulares e costeiras não apresentaram diferença entre si. Para o gênero *Prochlorococcus* as localidades insulares somaram quase todas as sequências, sendo as amostras do AS mais abundantes. A amostra AS2 apresentou uma média de aproximadamente 39.000 sequências. Uma ANOVA (pTukey<0,05) entre os pontos deste gênero diferenciou as quatro localidades do AS de todas as outras amostras costeiras. Uma ANOVA bifatorial comparou ambos os gêneros demonstrando uma clara diferença entre os dois (pANOVA=0,0001).

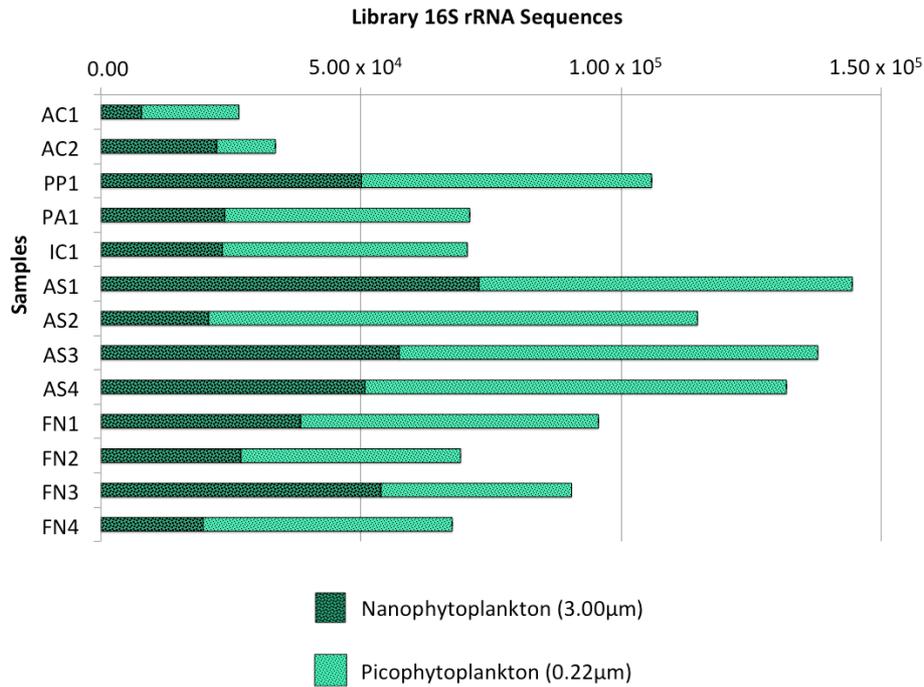


Figura 30. Total de sequências extraídas da biblioteca 16S rRNA para cianobactérias através da plataforma MiSeq (Illumina). Estas sequências representam apenas as que passaram pelo processo de pré-tratamento e filtragem seguindo o QIIME com posterior alinhamento na base de dados do SILVA. Fonte: o próprio autor.

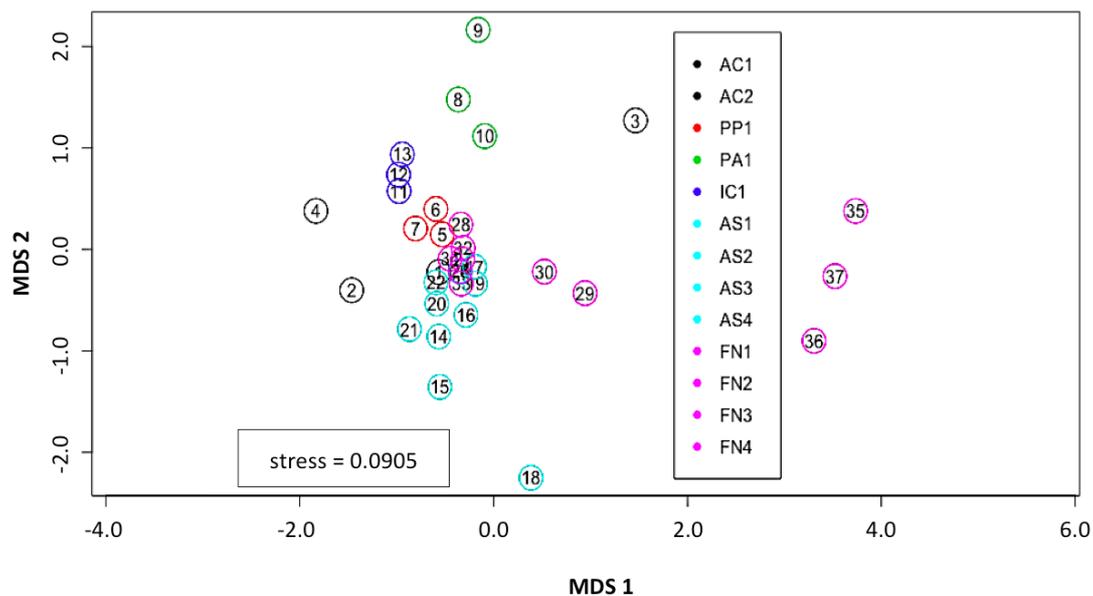


Figura 31. Representação do escalonamento multidimensional (MDS) das réplicas em função das seqüências extraídas da fração nanofitoplanctônica - 3,00µm. O baixo estresse revelou que a dimensionalidade dos pontos foi representada de maneira justa. Fonte: o próprio autor.

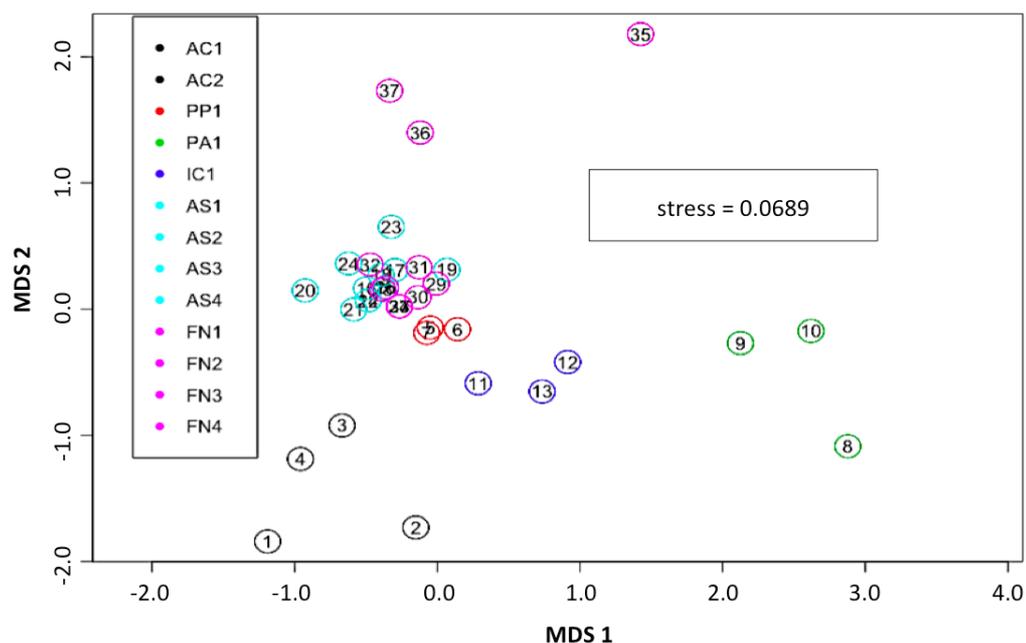


Figura 32. O baixo estresse encontrado na MDS das réplicas da fração picofitoplanctônica - 0,22µm - revelou uma distribuição ideal. Fonte: o próprio autor.

Como visto anteriormente, a comparação metodológica entre a DGGE e o sequenciamento foi baseada em matrizes ecológicas de presença/ausência. A seguir, serão descritas as análises ecológicas restritas apenas a técnica do sequenciamento, aplicada às amostras insulares e costeiras. Tais análises foram baseadas em matrizes de abundância de seqüências a cada táxon analisado. Para facilitar a descrição, as frações 3,00 e 0,22µm separadas. Portanto 4 possíveis cenários serão constantemente comparados: amostras da costa da fração nanofitoplanctônica avaliadas pela técnica do sequenciamento em larga escala (Seq

3,00µm - costa); amostras dos arquipélagos da fração nanofitoplanctônica (Seq 3,00µm - arquipélagos); amostras do picofitoplâncton na costa (Seq 0,22µm - costa); amostras do picofitoplâncton nas ilhas (Seq 0,22µm - arquipélagos). Tais cenários juntos representam grande parte do fitoplâncton na camada superficial das águas do contorno oeste do oceano Atlântico Equatorial. O total de 56 táxons pôde ser identificado, sobrepondo as frações do pico e do nanofitoplâncton da costa e nas ilhas através do sequenciamento.

Considerando somente a fração nanofitoplanctônica, uma riqueza total de 50 táxons foi observada com média de 13,54(±7,59) por amostra. Isolando-se apenas a fração costeira foram observados 39 táxons com uma média de 14,31 por amostra (Tabela 22). As amostras de AC2 e PA1 apresentaram as maiores riquezas observadas, com 43,00% e 37,33% da riqueza relativa total (Figura 33). Uma estimativa de riqueza utilizando o método de rarefação Chao1 registrou 41 OTUs, valor muito próximo da riqueza total encontrada na costa. Nos arquipélagos a riqueza foi superior: 45 táxons foram observados, com média de 13,13 por amostra. A riqueza estimada pelo método Chao1 foi muito similar à observada, com 48 OTUs especulados. FN4 e FN2 apresentaram as maiores riquezas com 27 (54,00% da riqueza relativa total) e 22 (44,67%) dos táxons observados, respectivamente (Figura 33). Os táxons considerados como raros estão presentes em apenas um ponto, sem compartilhamento com outra localidade apenas considerando a fração 3,00µm (LOGARES et al., 2013). Para a costa apenas Arraial do Cabo e Paracuru apresentaram táxons raros, com apenas 1 em PA1 (New Oscillatoriophycideae); 1 em AC1 (Unassigned); e 1 em AC1 e AC2 (Chlorophyta). Para os arquipélagos 5 táxons foram considerados raros, sendo 1 em AS2 (Lyngbya); 1 em FN2 (Symploca); e 3 em FN4 (SM1D11, Nostocaceae, Spirulina). Em contrapartida, os táxons que estiveram presentes a todas as localidades da fração nanofitoplanctônica são Stramenopiles (0,21% do total de sequências desta fração), Haptophyceae (0,38%), Synechococcus (29,66%), Prochlorococcus (68,70%).

O picofitoplâncton apresentou uma riqueza total observada de 43 táxons, sendo que a média por amostra foi de 7,57(±4,92). Na costa apenas 34 táxons foram observados, com uma média de 10,77 a cada amostra costeira (Tabela 22). O ponto PA apresentou 44,19% da riqueza total observada, com 2-4 vezes mais táxons observados em relação às outras amostras costeiras (Figura 33). A riqueza estimada por Chao1 as amostras costeira foi 36 OTUs. Para as amostras insulares uma menor riqueza foi encontrada, quando comparadas as amostras costeiras, com apenas 30 táxons com media de 5,83 a cada amostra. A riqueza estimada foi igual à riqueza observada nas amostras insulares. As riquezas de FN e AS são muito próximas, com destaque para FN4 que apresenta 32,56% da riqueza relativa total (Figura 33). Os táxons considerados raros estão mais adensados nas amostras costeiras do que nas amostras insulares considerando apenas a fração 0,22µm. Para a costa são 2 em AC1 (New Rodophyta e *Phascolarctobacterium*); 2 em AC2 (New Streptophyta e New CL50015);

1 em AC1 e AC2 (Chlorophyta); e 6 em PA1 (New SM1D11, New Oscillatoriothycidae, New Chroococcales, New Synechococcophycidae, *Leptolyngbya*, New Pirellulaceae). Para as amostras insulares 5 táxons foram considerados raros, sendo 2 em AS2 (*Microcoleus* e New Pseudanabaenaceae) e 3 em FN4 (New Cyano 2, New Nostocales, *Phormidium*). Já os táxons que estiveram presentes em todas as localidades da fração picofitoplancônica são apenas os gêneros *Synechococcus* (26,05% do total de sequências desta fração) e *Prochlorococcus* (73,43%).

Uma curva de acumulação de espécies, considerando o acúmulo de sequências, e uma curva de rarefação pelo índice Chao1 foram ajustadas, para ambas as frações do fitoplâncton (Figura 34). Considerando o número de réplicas como amostragens, a fração nanofitoplancônica foi melhor retratada com uma curva de rarefação bem ajustada com número de espécies amostradas similar ao número estimado (Figura 34). As amostras de PP apresentaram a menor equitatividade com diversos táxons expressos em baixas abundâncias, indicado presença de espécies raras (Linha vermelha - fração 3,00 $\mu$ m - Figura 34). Uma elevada equitatividade foi encontrada nas réplicas de AC e algumas localidades dos arquipélagos, indicando dominâncias de alguns grupos em específico (Figura 34). Embora a fração picofitoplancônica tenha sido mais abundante em número de sequências observadas, apresentou uma menor riqueza em relação à porção do nanofito (Figura 34). A curva de rarefação para a fração 0,22 $\mu$ m apresentou uma menor inclinação da curva ( $\alpha$ ) e uma maior área de rarefação de espécies demonstrando uma amostragem não muito representativa da riqueza total estimada (elevados desvios - área azul - 0,22 $\mu$ m - Figura 34). Réplicas de FN3 e PA1 e AC2 apresentaram as menores equitatividades (Figura 34).

Os índices de  $\alpha$ -diversidade Shannon e Simpson apresentaram os mesmos padrões, porém cada índice respeitando a sua escala. Com as diferenças melhor elucidadas pelo índice Shannon-Wiener (Tabela 22). O nanofitoplâncton apresentou maiores médias que a fração do picofitoplâncton. Considerando apenas a fração do nanofito houve maiores médias na porção insular, embora as médias costeiras tenham sido ligeiramente menor. AC e FN apresentaram as maiores médias, com destaque para FN4. PP1 e AS4 apresentaram as menores médias (Tabela 22). A fração do picofito apresentou uma grande diferença entre a diversidade da costa em relação ao ambiente insular, sendo a costa muito mais diversa que os ambientes insulares. Nesta fração 0,22 $\mu$ m AC apresentou as maiores médias, enquanto que PP apresentou a menor. FN e AS foram muito similares entre si (Tabela 22).

Clusters pela média foram gerados através do índice de Jaccard para analisar a  $\beta$ -diversidade entre os distintos pontos que caracterizaram as amostras costeiras e insulares da biblioteca de cianobactérias 16S rRNA. Para a fração 3 $\mu$ m a média de beta diversidade foi ligeiramente menor que a média para a fração do pico com 0,64 e 0,68, respectivamente. As correlações cofenéticas e os estresses das NMDS foram satisfatórios, validando os

agrupamentos em todos os casos. Para o nano, PA1 e IC1 apresentaram as maiores similaridades entre réplicas, com aproximadamente 50%. PA1 foi agrupada junto com FN4 e algumas réplicas de FN e AS, resultando em um agrupamento altamente dissimilar a todos os outros pontos. AS2 também se diferenciou drasticamente das outras amostras (Figura 35). O agrupamento da fração 0,22 $\mu$ m foi mais claro que a fração do nanofito, ou seja, as réplicas das mesmas amostras ficaram mais agrupadas indicando uma maior similaridade (Figura 35). Na fração 0,22 $\mu$ m as réplicas de PA1 também apresentaram as maiores similaridades com 60%, seguidas de IC1 e AC com aproximadamente 40%. PA1 e FN4 ficaram bem distribuídas em relação às outras amostras. A réplica 19, pertencente ao ponto AS2, foi muito dissimilar as outras amostras. As réplicas de AC e IC apresentaram muita similaridade entre si, assim como FN2 e AS1 (Figura 35).

De maneira geral as diferenças entre a fração nanofitoplanctônica e a fração picofitoplacntônica são maiores nos arquipélagos de que na costa, tanto para as médias de riqueza quanto para  $\alpha$ -diversidade (Tabela 22). Para o nanofito houve uma maior riqueza na costa, enquanto a diversidade foi maior nos arquipélagos. Já para o picofito tanto a riqueza quanto a diversidade foram maiores no ambiente costeiro (Tabela 22).

Analisando ambas as frações do fitoplâncton conjuntamente, AC apresentou os maiores desvios positivos em relação à média costeira, de acordo com índices ecológicos avaliados. PP1 seguido de IC1 apresentam os menores desvios (Tabela 22). Já para as médias ecológicas insulares, FN apresentou maiores desvios positivos em relação ao AS. As localidades de FN tiveram destaque positivo em FN4 e negativo em FN1. No AS os desvios positivos mais acentuados foram em AS3, e os desvios mais negativos em AS2 e AS4 (Tabela 22).

Ao todo, 56 táxons foram identificados considerando as frações 3,00 $\mu$ m e 0,22 $\mu$ m em conjunto, sendo que um destes são sequências genéticas não assinadas dentro do banco de dados SILVA (Unassigned). A máxima resolução taxonômica alcançada foi em nível de gênero, com 21 táxons identificados. Os outros 35 ‘táxons’ correspondem a novas sequências encontradas, sendo denominadas com o prefixo “New” e seguidas pelo nome do grupo taxonômico mais próximo identificado (Tabela 23). Como a biblioteca foi de 16S rRNA o termo novas sequências não deve ser confundido com novas espécies. Apenas primers específicos ao filo das cianobactérias foram utilizados neste trabalho, porém mesmo assim representante de outros 5 filios puderam ser amplificados e assim identificados. Dos 56 táxons, 48 pertencem às cianobactérias, estando compreendidos em 32 famílias distintas. As famílias de maior representatividade foram Phormidiaceae, Xenococcaceae, Pseudanabaenaceae, Synechococcaceae, Cyanobacteriaceae e Nostocaceae (Tabela 23). O maior número de novas sequências (New) foi dentre as classes Chloroplast, Synechococcophycideae e Oscillatoriohyphycideae (Tabela 23).

Isolando as frações 3 e 0,22 $\mu$ m apenas 37 táxons continuam sendo compartilhados, com a maior parte desta diversidade distribuída entre as famílias Pseudanabaenaceae, Cyanobacteriaceae, Xenococcaceae e Synechococcaceae (Figura 36). Apenas 2 táxons foram coincidentes entre as amostras costeiras de ambas as frações (New Chlorophyta e New Oscillatoriophycidae) e apenas 1 entre as frações insulares (New Pseudanabaenaceae 1). Isolando as frações do fitoplâncton pelas localidades “costa e ilhas” são compostos 4 cenários distintos, nos quais apenas 20 táxons são comuns entre eles (Figura 36). Cada cenário possui seus táxons exclusivos, sendo que a somatória destes compreendem 17 táxons. Os táxons exclusivos ao nanofito da costa são 2 (New Gomphosphaeriaceae e *Chroococidiopsis*); exclusivos ao nano dos arquipélagos são 9 (New Nostocaceae, New Prochloraceae, *Spirulina*, *Lyngbya*, New Phormidiaceae 1, *Oscillatoria*, *Symploca*, New Pseudanabaenales, New BS119); exclusivos ao picofito da costa são 4 (New Rhodophyta, New Streptophyta, *Phascolarctobacterium*, New CL50015); e exclusivos ao pico dos arquipélagos são apenas 2 táxons (*Microcoleus* e *SGSH944*).

Utilizando a abundância relativa do número que sequências em cada amostra, foram retirados os táxons com abundâncias abaixo de 0,01% (ou seja, raros), restando apenas os táxons mais abundantes. Para a fração do nanofito (3,00 $\mu$ m) 12 táxons foram considerados os mais relevantes em abundância: New Chloroplast, *Cyanobacterium*, *Halomicronema*, New Haptophyceae, *Microcystis*, New Xenococcaceae 1, *Prochlorococcus*, New Pseudanabaenaceae 2, *Rivularia*, New Stramenopiles, *Synechococcus* e New Xenococcaceae 2 (Figura 37). Um heatmap (‘mapa de calor’) correlacionou o dendograma (jaccard) dos táxons de maior abundância descritos acima, com o dendograma (jaccard) da matriz geral das abundâncias das sequências pelos pontos (fração 3,00 $\mu$ m - Figura 35). Desta forma os pontos de coleta puderam ser correlacionados com os grupos de maior relevância. As amostras costeiras apresentaram uma clara correlação positiva com o gênero *Synechococcus* e correlação negativa a todos os outros grupos (Figura 37). As amostras insulares apresentaram correlação positiva com o gênero *Prochlorococcus*. Destaque para as várias réplicas das localidades de FN que apresentaram fracas correlações negativas com *Synechococcus*, indicando uma importância na contribuição deste grupo para a abundância nas amostras de FN (Figura 37). A correlação de Spearman realizada entre os táxons de maior abundância revela a relação antagônica entre *Synechococcus* e *Prochlorococcus*, apresentando entre eles uma alta correlação negativa e com elevada significância estatística ( $p_{\text{Spearman}} < 0,001$ ) (Figura 37 - lado direito).

Na fração picofitoplantônica (0,22 $\mu$ m) um número menor de táxons foi considerado de elevada abundância relativa: New Haptophyceae, New Chroococcales 1, New Phormidiaceae 2, *Prochlorococcus*, *Synechococcus*, New Xenococcaceae 2, *Xenococcus* (Figura 38). Nesta fração as correlações positivas e negativas demonstram de maneira clara a importância da

contribuição da abundância de *Synechococcus* para as amostras costeiras, e a importância da abundância de *Prochlorococcus* às amostras insulares no heatmap (Figura 38). A correlação de Spearman realizada entre os táxons de maior abundância revelou uma forte correlação negativa de elevada significância estatística ( $p_{\text{Spearman}} < 0,001$ ) entre *Synechococcus* e *Prochlorococcus* (Figura 38 - lado direito).

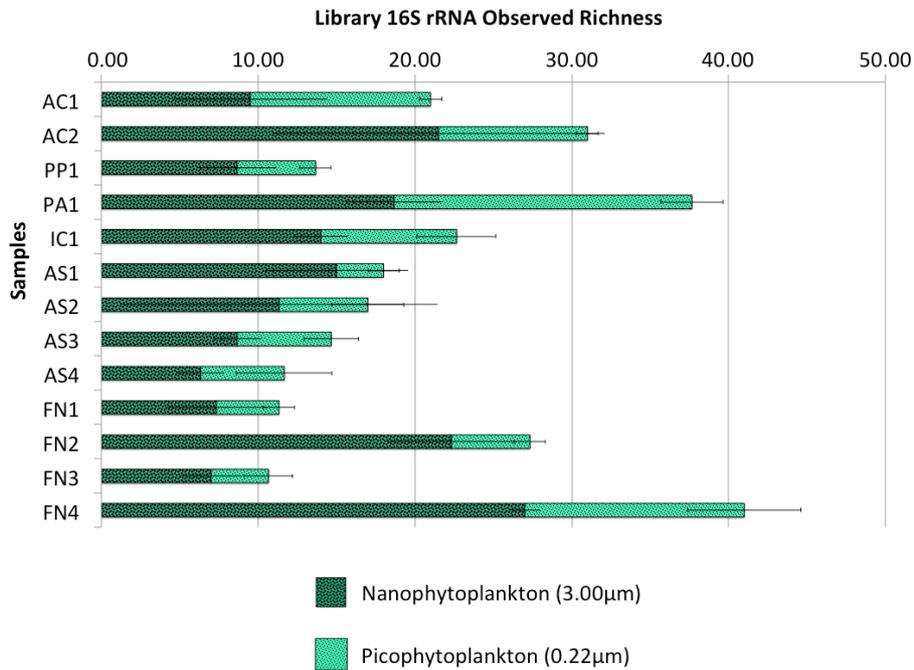


Figura 33. Médias e desvios das réplicas a cada ponto para a riqueza observada nas frações do pico e nanofitoplâncton a partir dos dados de sequenciamento de amplicons da biblioteca 16S rRNA. Fonte: o próprio autor.

Tabela 22. Desvios dos valores observados em relação à média correspondente ao respectivo ambiente, costeiro ou insular, frente aos parâmetros ecológicos de riqueza e diversidade. Duas frações do fitoplâncton são comparadas, a fração do nanofitoplâncton e picofitoplâncton. As médias são baseadas em matrizes de abundância de sequências 16S rRNA para cianobactérias. Os valores azuis correspondem a desvios positivos, enquanto que os valores em vermelho correspondem aos desvios negativos. Fonte: o próprio autor.

		Library 16S rRNA					
		Richness		α-Div Shannon		α-Div Simpson	
		Nano	Pico	Nano	Pico	Nano	Pico
Coast	Average	14.31	10.77	0.2401	0.2000	0.1011	0.0779
AC1	Observed	10	12	0.3208	0.6246	0.1266	0.2701
	Deviation (%)	-33.60	6.79	33.60	212.27	25.21	246.69
AC2	Observed	22	10	0.6162	0.4281	0.3246	0.1699
	Deviation (%)	50.27	-11.79	156.68	114.02	220.96	118.15
PP1	Observed	9	5	0.0450	0.0171	0.0123	0.0044
	Deviation (%)	-39.43	-53.57	-81.25	-91.47	-87.83	-94.38
PA1	Observed	19	19	0.1424	0.1224	0.0385	0.0336
	Deviation (%)	30.47	76.43	-40.68	-38.80	-61.93	-56.87
IC1	Observed	14	9	0.2283	0.0255	0.0866	0.0062
	Deviation (%)	-2.15	-19.52	-4.92	-87.26	-14.35	-91.99
Archipelagos	Average	13.13	5.83	0.3383	0.0569	0.1502	0.0202
AS1	Observed	15	3	0.1037	0.0250	0.0333	0.0076
	Deviation (%)	14.29	-48.57	-69.35	-56.01	-77.80	-62.61
AS2	Observed	11	6	0.1695	0.0283	0.0610	0.0083
	Deviation (%)	-13.65	-2.86	-49.90	-50.25	-59.40	-59.00
AS3	Observed	9	6	0.2014	0.1258	0.0849	0.0530
	Deviation (%)	-33.97	2.86	-40.45	120.88	-43.50	162.49
AS4	Observed	6	5	0.0436	0.0237	0.0130	0.0069
	Deviation (%)	-51.75	-8.57	-87.10	-58.44	-91.37	-65.92
FN1	Observed	7	4	0.1366	0.0273	0.0513	0.0081
	Deviation (%)	-44.13	-31.43	-59.60	-52.00	-65.85	-59.76
FN2	Observed	22	5	0.7380	0.0711	0.4247	0.0256
	Deviation (%)	70.16	-14.29	118.17	24.82	182.77	26.84
FN3	Observed	7	4	0.2878	0.0725	0.1466	0.0276
	Deviation (%)	-46.67	-37.14	-14.93	27.31	-2.36	36.70
FN4	Observed	27	14	1.0255	0.0818	0.3868	0.0245
	Deviation (%)	105.71	140.00	203.17	43.69	157.52	21.26

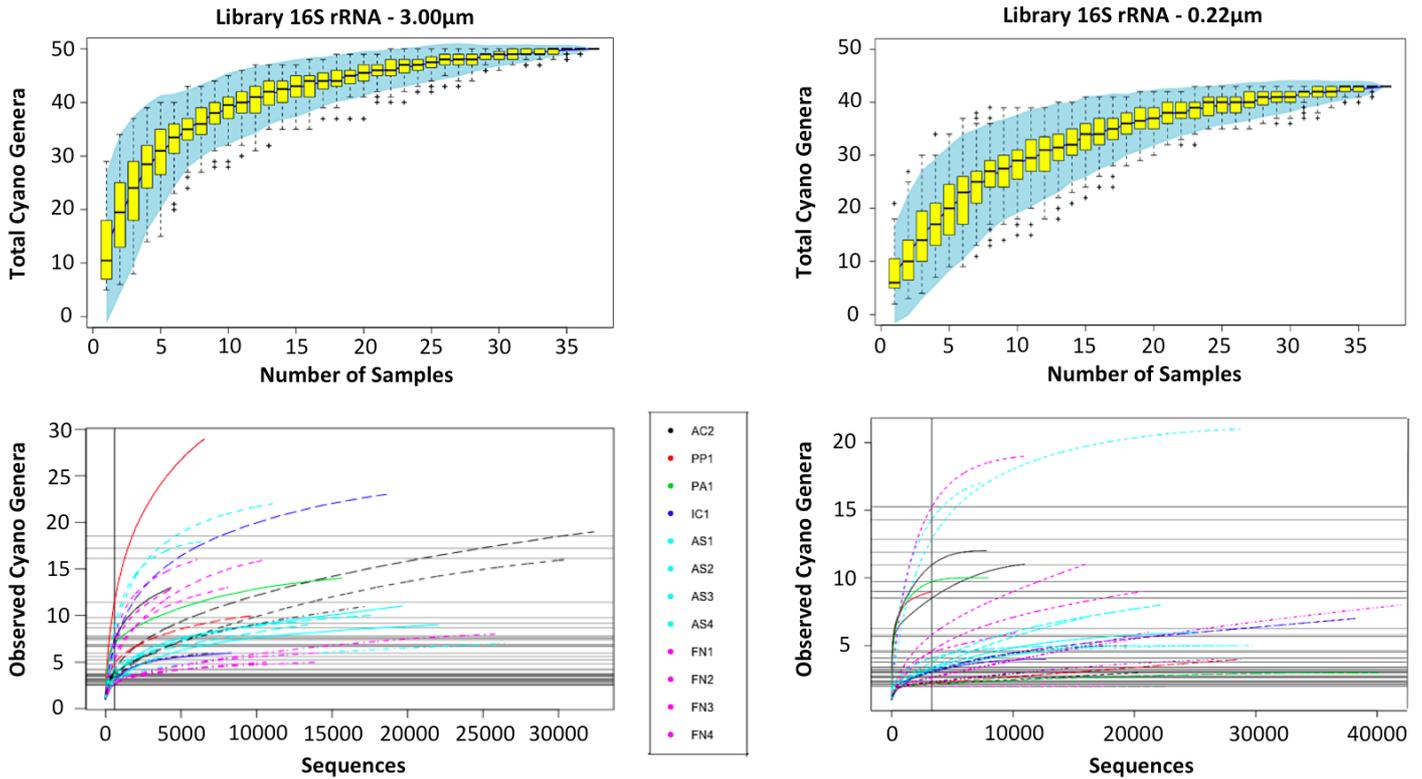


Figura 34. Curvas de rarefação de “espécies” considerando o índice de riqueza estimada Chao1 - abundância. Nas imagens superiores, as réplicas dos pontos de amostragem (37 réplicas) representam as repetições de amostragem na costa e nos arquipélagos. As barras amarelas indicam os desvios das réplicas em cada ponto, enquanto que os traços pretos representam a média da riqueza total acumulada ao longo das amostragens. Os limites tracejados nas barras indicam a extrapolação para espécies estimadas. A área azul indica o ambiente de rarefação, contemplando os limites os máximos e mínimos da riqueza no ambiente costeiro estudado. Nas figuras abaixo estão demonstradas as curvas de acumulação de espécies a cada localidade, estão identificadas por distintas cores. O número de sequências acumulado em cada réplica esta representado no eixo x. Fonte: o próprio autor.

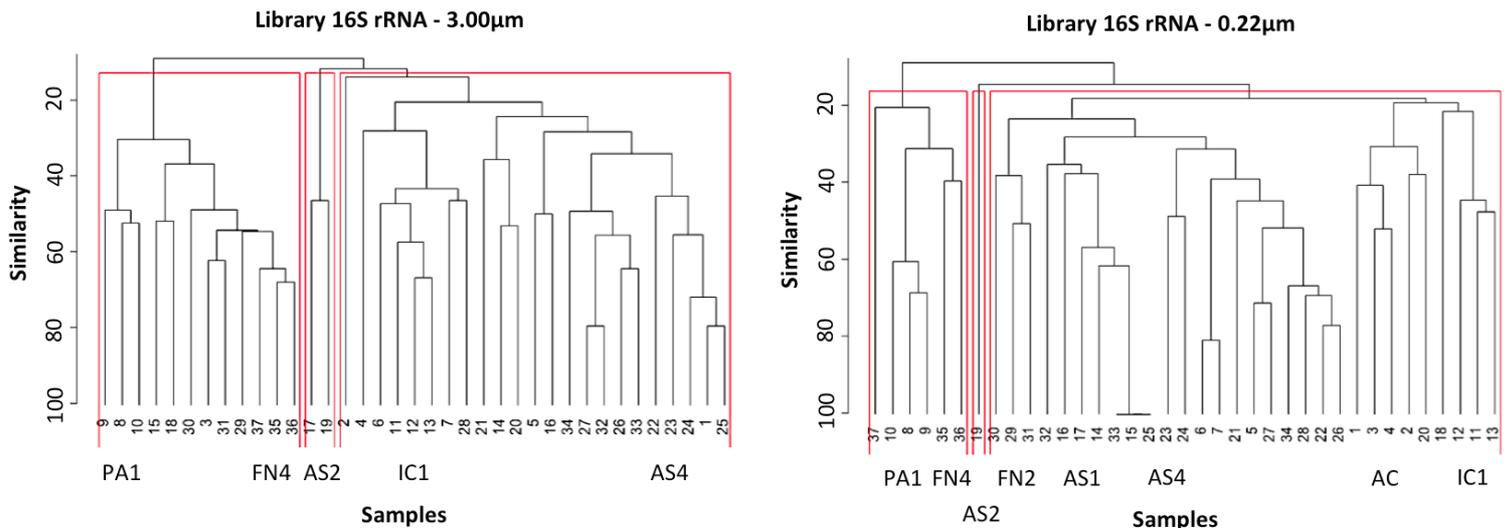


Figura 35. Agrupamentos pelas médias de beta diversidade das amostras insulares e costeiras das menores frações do fitoplâncton marinho em águas superficiais da porção oeste do Atlântico Equatorial. Os agrupamentos foram baseados em matrizes de dissimilaridade pelo índice binário de Jaccard. Fonte: o próprio autor.

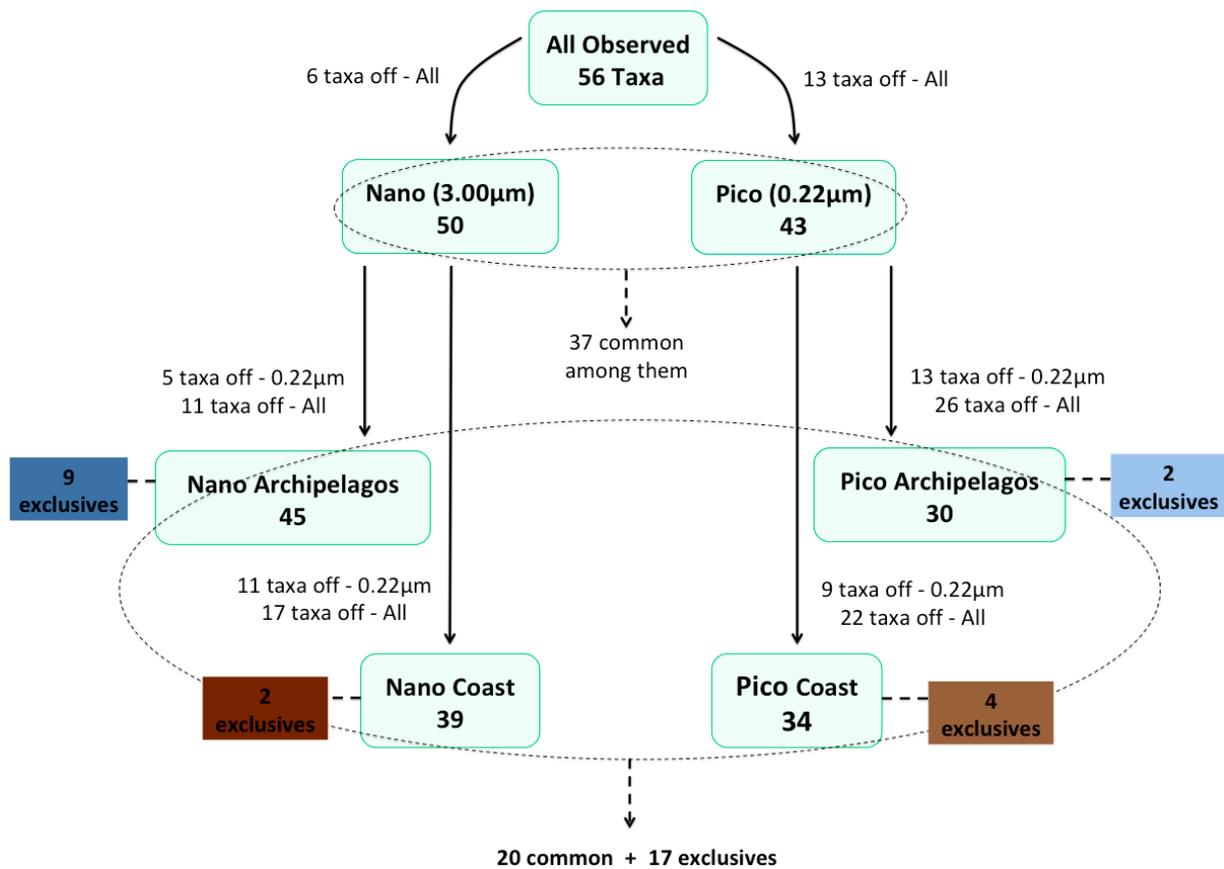


Figura 36. Relação da riqueza total observada aos diferentes cenários propostos, relacionando o ambiente costeiro e insular às frações do nano e picofitoplâncton. As caixas coloridas demonstram o número de espécies raras a cada um dos 4 cenários: Nano - costa; nano - arquipélago; pico - costa; pico - arquipélago. Fonte: o próprio autor.

Tabela 23. Grupos taxonômicos encontrados pela identificação das sequências da biblioteca 16S rRNA para cianobactérias na porção superficial do Atlântico. Os grupos de trabalho estão demonstrados na variável Taxa, os quais reúnem as menores classificações taxonômicas encontradas, sendo gênero a menor delas. Cada grupo taxonômico possui o respectivo preenchimento dentre os 4 diferentes cenários analisados. Fonte: o próprio autor.

Phylum	Class	Order	Family	Taxa (OTUs + Genus)	Nano (3.00µm)		Pico (0.22µm)	
					Coast	Archipelagos	Coast	Archipelagos
Unassigned	Unassigned	Unassigned	Unassigned	Unassigned				
Cyanobacteria	Other	Other	Other	New Cyano 1				
Cyanobacteria	Other	Other	Other	New Cyano 2				
Cyanobacteria	4C0d-2	MLE1-12	Other	New MLE112				
Cyanobacteria	4C0d-2	SM1D11	Other	New S M1D11				
Cyanobacteria	Chloroplast	Other	Other	New Chloroplast				
Cyanobacteria	Chloroplast	Cercozoa	Other	New Cercozoa				
Cyanobacteria	Chloroplast	Chlorophyta	Other	New Chlorophyta				
Cyanobacteria	Chloroplast	Cryptophyta	Other	New Cryptophyta				
Cyanobacteria	Chloroplast	Haptophyceae	Other	New Haptophyceae				
Cyanobacteria	Chloroplast	Rhodophyta	Other	New Rhodophyta				
Cyanobacteria	Chloroplast	Stramenopiles	Other	New Stramenopiles				
Cyanobacteria	Chloroplast	Streptophyta	Other	New Streptophyta				
Cyanobacteria	Gloeobacterophycideae	Gloeobacterales	Gloeobacteraceae	<i>Gloeobacter</i>				
Cyanobacteria	Nostocophycideae	Nostocales	Other	New Nostocales				
Cyanobacteria	Nostocophycideae	Nostocales	Nostocaceae	New Nostocaceae				
Cyanobacteria	Nostocophycideae	Nostocales	Nostocaceae	<i>Cylindrospermopsis</i>				
Cyanobacteria	Nostocophycideae	Stigonematales	Rivulariaceae	<i>Rivularia</i>				
Cyanobacteria	Oscillatoriophycideae	Other	Other	New Oscillatoriophycideae				
Cyanobacteria	Oscillatoriophycideae	Chroococcales	Other	New Chroococcales 1				
Cyanobacteria	Oscillatoriophycideae	Chroococcales	Other	New Chroococcales 2				
Cyanobacteria	Oscillatoriophycideae	Chroococcales	Cyanobacteriaceae	New Cyanobacteriaceae 1				
Cyanobacteria	Oscillatoriophycideae	Chroococcales	Cyanobacteriaceae	New Cyanobacteriaceae 2				
Cyanobacteria	Oscillatoriophycideae	Chroococcales	Cyanobacteriaceae	<i>Cyanobacterium</i>				
Cyanobacteria	Oscillatoriophycideae	Chroococcales	Gomphosphaeriaceae	New Gomphosphaeriaceae				
Cyanobacteria	Oscillatoriophycideae	Chroococcales	Microcystaceae	<i>Microcystis</i>				
Cyanobacteria	Oscillatoriophycideae	Chroococcales	Prochloraceae	New Prochloraceae				

Cyanobacteria	Oscillatoriophycideae	Chroococcales	Spirulinaceae	<i>Spirulina</i>				
Cyanobacteria	Oscillatoriophycideae	Chroococcales	Xenococcaceae	New Xenococcaceae 1				
Cyanobacteria	Oscillatoriophycideae	Chroococcales	Xenococcaceae	New Xenococcaceae 2				
Cyanobacteria	Oscillatoriophycideae	Chroococcales	Xenococcaceae	<i>Chroococciopsis</i>				
Cyanobacteria	Oscillatoriophycideae	Chroococcales	Xenococcaceae	<i>Xenococcus</i>				
Cyanobacteria	Oscillatoriophycideae	Oscillatoriales	Oscillatoriaceae	<i>Lyngbya</i>				
Cyanobacteria	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	New Phormidiaceae 1				
Cyanobacteria	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	New Phormidiaceae 2				
Cyanobacteria	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	<i>Microcoleus</i>				
Cyanobacteria	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	<i>Oscillatoria</i>				
Cyanobacteria	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	<i>Phormidium</i>				
Cyanobacteria	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	<i>Symploca</i>				
Cyanobacteria	Synechococcophycideae	Other	Other	New Synechococcophycideae				
Cyanobacteria	Synechococcophycideae	Pseudanabaenales	Other	New Pseudanabaenales				
Cyanobacteria	Synechococcophycideae	Pseudanabaenales	Pseudanabaenaceae	New Pseudanabaenaceae 1				
Cyanobacteria	Synechococcophycideae	Pseudanabaenales	Pseudanabaenaceae	New Pseudanabaenaceae 2				
Cyanobacteria	Synechococcophycideae	Pseudanabaenales	Pseudanabaenaceae	<i>Halomiconema</i>				
Cyanobacteria	Synechococcophycideae	Pseudanabaenales	Pseudanabaenaceae	<i>Leptolyngbya</i>				
Cyanobacteria	Synechococcophycideae	Synechococcales	Acaryochloridaceae	<i>Acaryochloris</i>				
Cyanobacteria	Synechococcophycideae	Synechococcales	Synechococcaceae	New Synechococcaceae				
Cyanobacteria	Synechococcophycideae	Synechococcales	Synechococcaceae	<i>Prochlorococcus</i>				
Cyanobacteria	Synechococcophycideae	Synechococcales	Synechococcaceae	<i>Synechococcus</i>				
Firmicutes	Clostridia	Clostridiales	Veillonellaceae	<i>Phascolarctobacterium</i>				
Planctomycetes	OM190	CL500-15	Other	New CL50015				
Planctomycetes	Planctomycetia	Pirellulales	Pirellulaceae	New Pirellulaceae				
Proteobacteria	Alphaproteobacteria	Other	Other	New Alphaproteobacteria				
Proteobacteria	Gammaproteobacteria	Alteromonadales	OM60	New OM60				
SAR406	AB16	Arctic96B-7	A714017	<i>SGSH944</i>				
ZB3	BS119	Other	Other	New BS119				

### Library 16S rRNA - 3.00µm

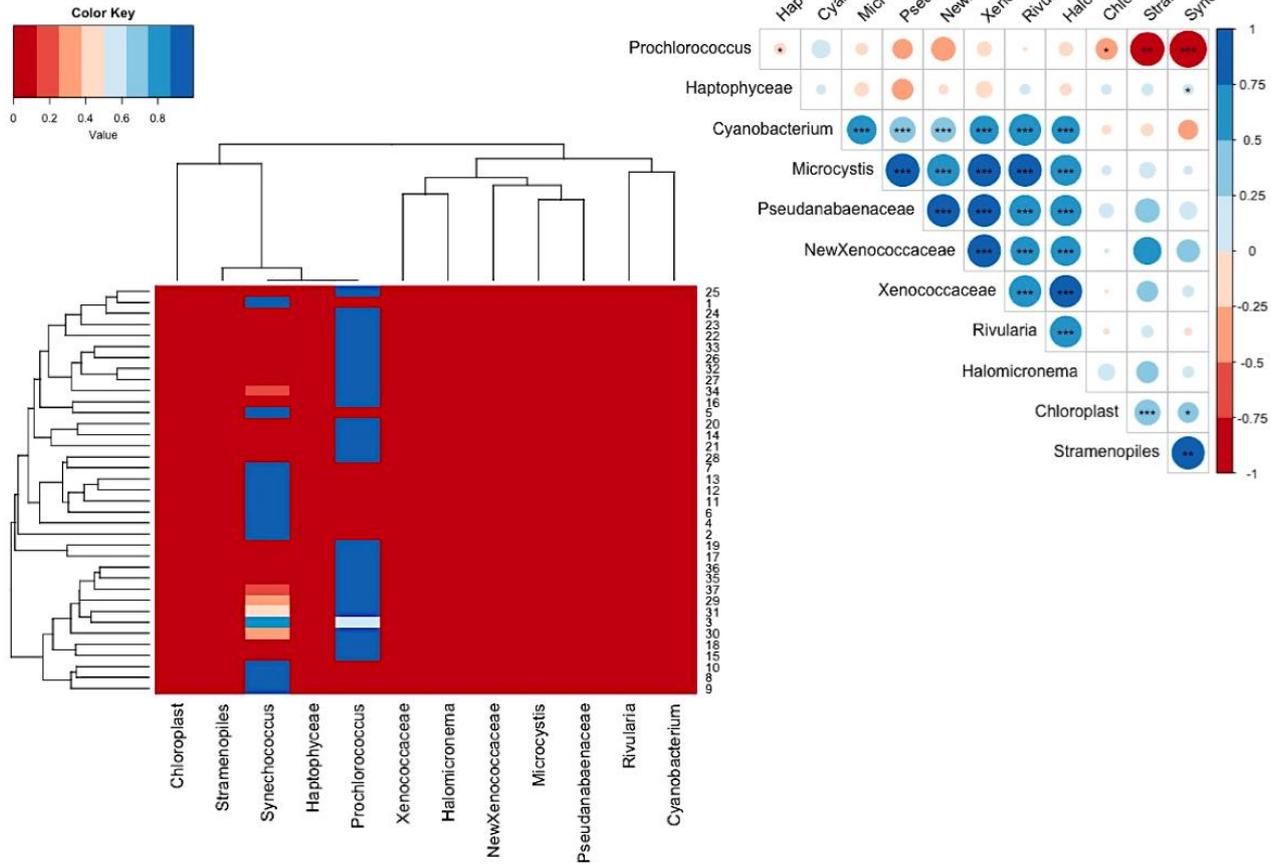


Figura 37. As correlações entre táxons e amostras da fraco nanofitoplanctônica (3.00µm) estão dispostas em um heatmap, com a barra de valores de correlação na parte superior. A distribuição de Jaccard foi utilizada para construir os dendogramas utilizados no heatmap, tanto o cluster das abundâncias das sequencias agrupadas pelas localidades (réplicas), quanto o cluster dos táxons de maior abundância. Na porção superior direita da figura estão ilustradas as correlações com suas respectivas significâncias entre os táxons mais abundantes. Os símbolos \*, \*\* e \*\*\* indicam significâncias de 0,05, 0,01 e 0,001, respectivamente. Fonte: o próprio autor.

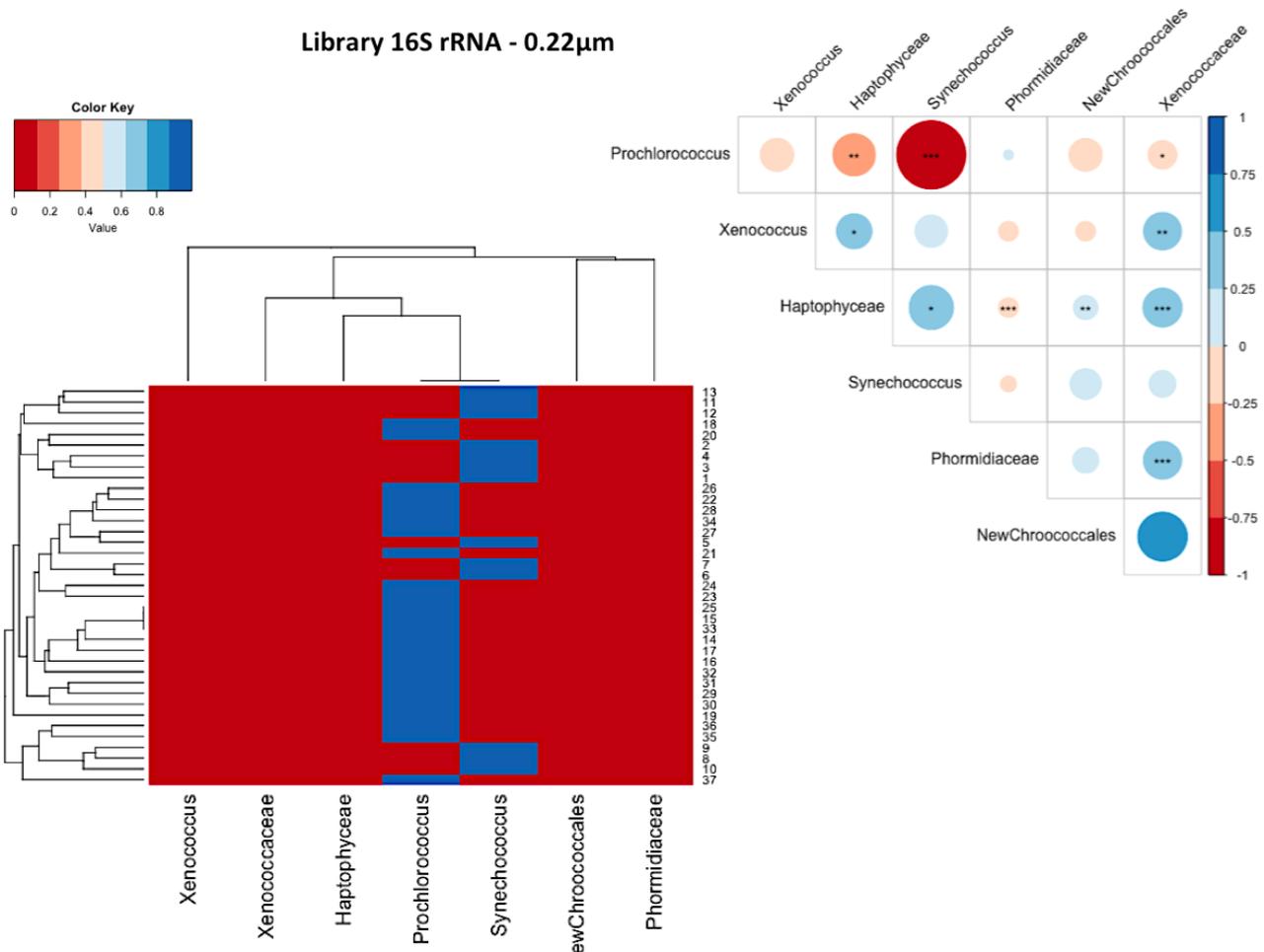


Figura 38. As correlações entre táxons e amostras da fração picofitoplantônica (0,22µm) estão dispostas em um heatmap, com a barra de valores de correlação na parte superior. A distribuição de Jaccard foi utilizada para construir os dendogramas utilizados no heatmap, tanto o cluster das abundâncias das sequências agrupadas pelas localidades (réplicas), quanto o cluster dos táxons de maior abundância. Na porção superior direita da figura estão ilustradas as correlações com suas respectivas significâncias entre os táxons mais abundantes. Os símbolos \*, \*\* e \*\*\* indicam significâncias de 0,05, 0,01 e 0,001, respectivamente. Fonte: o próprio autor.

### 5.5. Diversidade modulada pelas forçantes

Uma PCA foi proposta analisando apenas as variações dos dados abióticos da água, representando graficamente a distribuição dos pontos por estas variáveis (Figura 17), porém tais variáveis não possuem o mesmo peso quando contrapostos aos dados bióticos - abundância de sequências alinhadas. Uma RDA foi proposta a cada fração do fitoplâncton para determinar as principais variáveis que as modulam, assim como um teste de Mantel para confrontar a correlação das diferentes matrizes bióticas e abióticas. Neste caso as variáveis numéricas e também as categóricas (Substrate e Dynamics) também foram consideradas.

Para a fração 3,00µm uma RDA preliminar considerando todas as variáveis demonstrou apenas significância ( $p_{ANOVA} < 0,05$ ) entre as variáveis temperatura, salinidade, DO e os diferentes tipos de substrato. Sendo a temperatura a variável mais significativa. A RDA final composta apenas das variáveis significativas também apresentou significância ( $p < 0,01$ ), e obteve um  $R^2 = 0,60$  indicando que o modelo explicou de maneira parcial toda a variabilidade

dos dados (Figura 39). Esse  $R^2$  mediano pode ser explicado pela baixa correlação entre as matrizes, uma vez que o teste de Mantel apresentou uma baixa correlação positiva ( $corMantel=0,216$ ;  $pMantel=0,006$ ). *Synechococcus* apresentou maior relação com as variáveis temperatura e salinidade, assim como planícies arenosas; enquanto *Prochlorococcus* demonstrou maior relação com DO e substratos de fundo rochosos (Figura 39).

A RDA preliminar para a fração do picofito também demonstrou significância para os parâmetros de temperatura, salinidade, DO e substrato (Figura 40), a temperatura foi a variável de maior significância. A RDA final apresentou variações muito similares à fração 3,00 $\mu$ m, sendo os gêneros *Syneco* e *Procloro* os mais variáveis e melhor distribuídos em relação aos outros táxons (Figura 40). A RDA da fração 0,22 $\mu$ m também foi significativa ( $p<0,01$ ) e com um  $R^2=0,64$ . Até mesmo o teste de Mantel foi similar entre as duas frações, com uma baixa correlação positiva ( $corMantel=0,201$ ;  $pMantel=0,021$ ), sendo que neste caso as matrizes abióticas e bióticas foram ainda menos correspondentes às variáveis mais significativas.

As análises ecológicas já haviam sido bem detalhadas nos tópicos acima quanto às diferenças de localidade, ilhas e costa. Porém neste tópico as relações ecológicas serão demonstradas quanto às variáveis categóricas ambientais: Substrate e Dynamics.

Na fração nanofitoplanctônica a maior riqueza observada, com 47 táxons, se deu nas amostras de fundo rochoso. Das 37 réplicas que caracterizam 13 pontos de amostragem, 20 delas foram caracterizadas como fundo rochoso (Tabela 5). As poças de maré representaram apenas 9 réplicas dentro de 37, mas possuem 42 espécies observadas e a maior alfa-diversidade, com 0,3500 (Shannon) e 0,1747 (Simpson). As planícies arenosas correspondem a apenas 8 réplicas, com apenas 26 espécies observadas e uma baixa diversidade de 0,1827 (Shannon) e 0,0687 (Simpson). A variável hidrodinâmica (Dynamics) foi classificada em exposta e abrigada. As áreas expostas apresentaram 45 táxons observados e 17 réplicas dentro das 37 totais. A diversidade foi maior nas áreas expostas (0,3897 - Shannon, 0,1693 - Simpson) do que nas áreas abrigadas (0,2307 - Shannon, 0,1020 - Simpson). Uma matriz relacionando os substratos com a dinâmica, demonstrou que ambientes de fundo rochoso e expostos são os mais diversos (0,5346 - Shannon e 0,1999 - Simpson), enquanto as planícies arenosas de áreas expostas são as menos diversas (0,1553 - Shannon e 0,0580 - Simpson).

A beta-diversidade foi calculada pela distância de Jaccard, seguindo uma distribuição em função do substrato e da hidrodinâmica. No caso da fração 3,00 $\mu$ m, não houve um agrupamento isolado entre as amostras, em nenhum dos casos (Figura 41). Uma ANOVA entre os diferentes tipos de substrato foi proposta, porém não houve diferença estatística entre os pontos ( $pANOVA=0,22$ ). Também não houve diferença significativa entre as populações expostas e abrigadas ( $pANOVA=0,71$ ) (3,00 $\mu$ m - Figura 41).

Considerando apenas a fração picofitoplacntônica, as localidades com fundos rochosos foram as de maior riqueza observada com 32 táxons identificados. As poças de maré possuem

28 espécies observadas e as menores alfa-diversidades, com 0,0739 (Shannon) e 0,0225 (Simpson). As planícies arenosas correspondem a menor riqueza com apenas 22 táxons identificados, em contrapartida possuem a maior  $\alpha$ -diversidade com 0,1721 (Shannon) e 0,0715 (Simpson). De acordo com a hidrodinâmica, as áreas expostas apresentaram 38 táxons observados e uma diversidade levemente maior (0,1292 - Shannon, 0,0485 - Simpson). As áreas abrigadas possuem 26 táxons observados e uma menor diversidade (0,0885- Shannon e 0,0336 - Simpson). Uma matriz relacionando os diferentes substratos com áreas exposta e abrigadas, revelou que a maior  $\alpha$ -div ocorreu em planícies arenosas expostas (0,2601- Shannon e 0,1107 - Simpson), enquanto a menor foi observada em planícies arenosas abrigadas (0,0255 - Shannon e 0,0062 - Simpson).

Avaliando a beta-diversidade da fração 0,22 $\mu$ m utilizando a distância de Jaccard, não houve agrupamentos isolados entre as amostras, nem em relação ao função do substrato ou à hidrodinâmica (Figura 41). Uma ANOVA entre os diferentes tipos de substrato foi proposta, porém não houve diferença estatística entre os pontos (pANOVA= 0,40), tampouco entre as populações expostas e abrigadas (pANOVA= 0,07) (0,22 $\mu$ m - Figura 41).

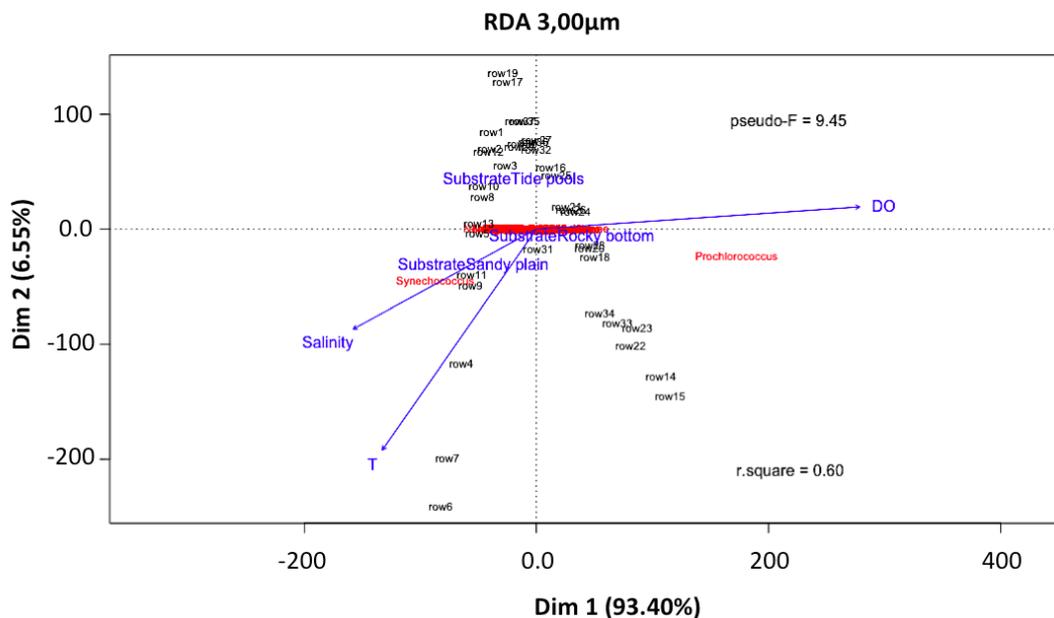


Figura 39. A variação dos pontos seguiu uma análise de redundância (RDA) com as sequências alinhadas da fração 3,00 $\mu$ m do fitoplâncton em cada replicação. Apenas variáveis significativas foram utilizadas na análise de variação: pH; salinidade, diferentes substratos e oxigênio dissolvido. As populações destacadas foram *Synechococcus* e *Prochlorococcus*. Fonte: o próprio autor.

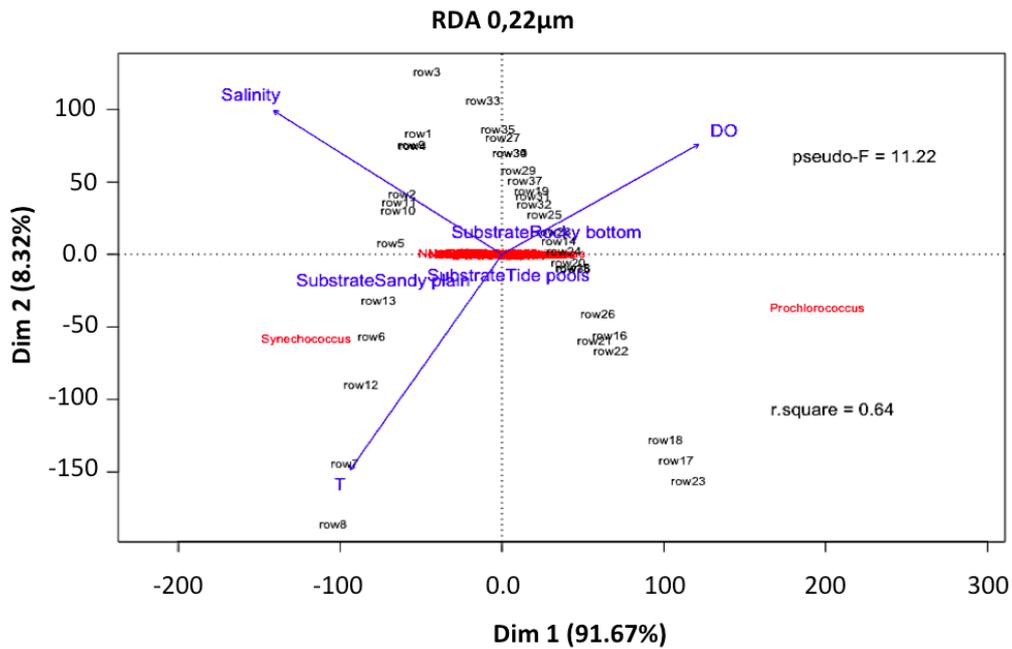


Figura 40. A variação dos pontos seguiu uma análise de redundância (RDA) com as sequências alinhadas da fração picoplânctônica - 0,22µm em cada réplica. Apenas variáveis significativas foram utilizadas na análise de variação: pH; salinidade, diferentes substratos e oxigênio dissolvido. As populações em destaque foram *Synechococcus* e *Prochlorococcus*. Fonte: o próprio autor.

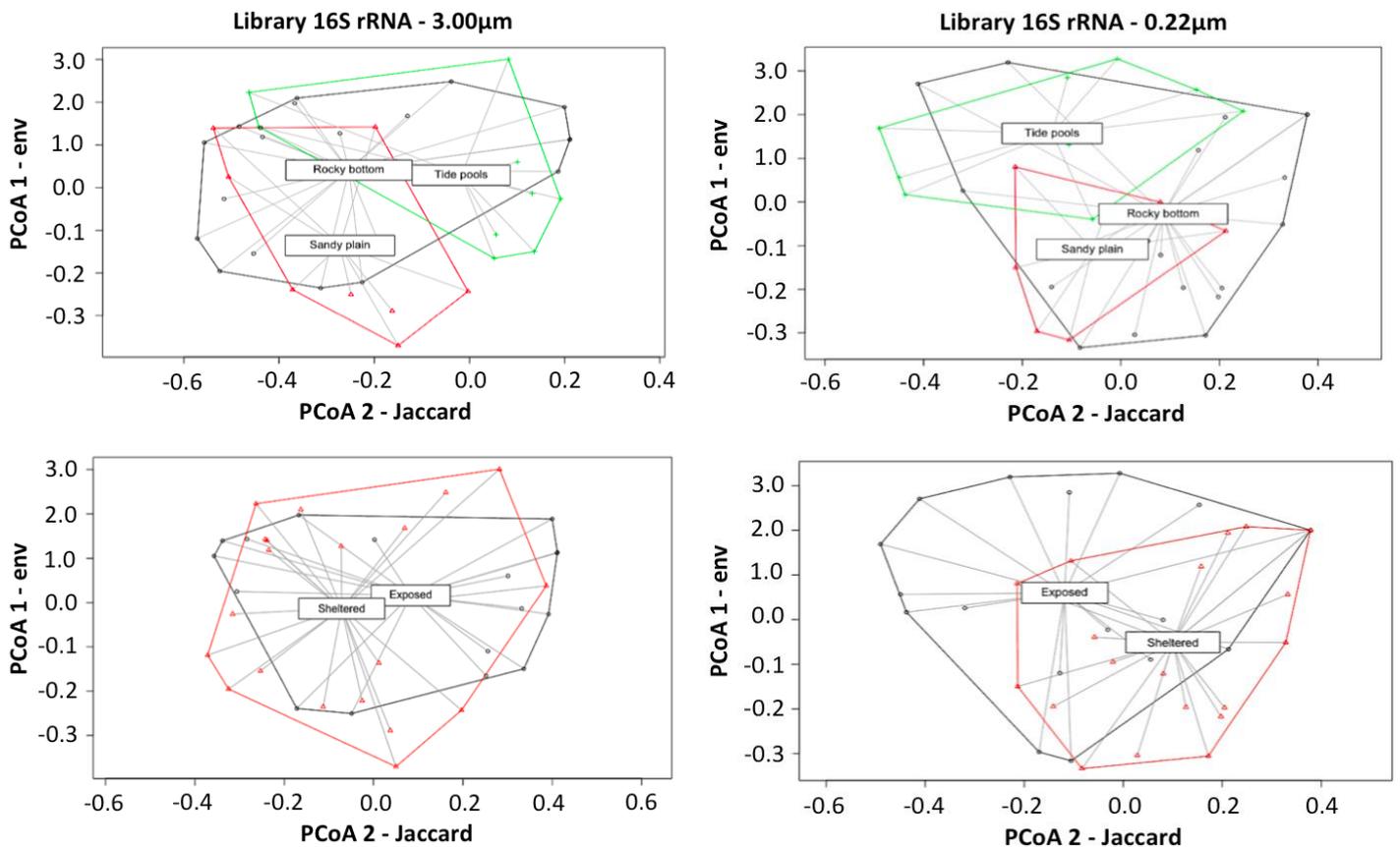


Figura 41. Análise de dispersão da beta-diversidade em função das variáveis categóricas de substrato (fundo rochoso, planícies arenosa e poças de maré) e dinâmica (abrigado e exposto), tanto para as frações do nano como do picofitoplâncton. Fonte: o próprio autor.

## 6. Discussão

Atualmente, já são conhecidos os giros oceânicos da porção Equatorial do Atlântico, que resultam num sistema de circulação oceânica e atmosférica muito localizado, energético e complexo (FRATANTONI & GLICKSON, 2002; MARIN, 2009; LUIZ et al., 2015). Tais giros se formam em poucas latitudes ao norte e ao sul do Equador, são resultados do encontro dos grandes giros Norte e Sul do oceano Atlântico e da circulação dos ventos Elíseos (MAYER & WEISBERG, 1993; BUB & BROWN, 1996; BOURLÈS et al., 1999; FRATANTONI & GLICKSON, 2002). Uma das correntes de maior abrangência na porção equatorial/central do Atlântico é a corrente superficial Equatorial Sul - SEC, que flui na direção Leste-Oeste do oceano Atlântico Sul. Esta corrente se divide ao se aproximar da plataforma nordeste do Brasil, resultando em duas correntes que margeiam a plataforma do Brasil (Figura 10). A Corrente Superficial Norte do Brasil (NBC) margeia a plataforma costeira do nordeste do Brasil com maior intensidade entre os estados do RN e MA, sendo a plataforma do CE uma região clássica com inserções da NBC (FRATANTONI & RICHARDSON, 2006; MARIN, 2009; MORAES, 2011). As massa da água deslocadas pelas correntes equatoriais na costa nordeste e nos arquipélagos oceânicos brasileiros são diretamente influenciados pelos efeitos do giro Equatorial Atlântico. Quanto a SEC quanto a NBC são caracterizadas como zonas de pouca produtividade associada ao fitoplâncton, nos arquipélagos (VIANA et al., 2015) e nas costa (EKAU & KNOPPERS, 1999), respectivamente. Porém, pouco se sabe sobre a influência destas correntes na comunidade ou distribuição fitoplanctônica (MARTINY et al., 2006).

O principal objetivo deste trabalho é a caracterização da comunidade de cianobactérias e seu estoque de carbono nas correntes descritas acima, seja na plataforma (costa) e fora da plataforma (nos arquipélagos oceânicos). O termo equatorial foi empregado neste trabalho por que os locais de amostragem ocorrem em baixas latitudes sul muito próximas, ou coincidentes, a latitude zero (Tabela 5). Os arquipélagos brasileiros de FN e AS foram escolhidos para caracterizar os efeitos da SEC, uma corrente quente superficial, assim como três pontos amostrados na costa do CE para caracterizar os efeitos costeiros da NBC (Figura 9). Em todos os locais amostrados foram mensurados parâmetros físico-químicos para determinar as condições das massas da água (Figura 12).

Este trabalho busca encontrar semelhanças e diferenças entre as localidades amostradas a fim de estabelecer um gradiente costa-ilhas por diferentes metodologias. De acordo com os parâmetros físico-químicos mensurados na coluna da água durante as coletas, as médias costeiras foram muito semelhantes às médias dos arquipélagos (Tabela 16). Considerando que as correstes superficiais que irrigam a zona equatorial do Atlântico Oeste são quentes, e a NBC deriva da SEC, esperasse que elas tivessem características em comum. Mesmo as médias costeiras e insulares sendo muito similares aos parâmetros da água, elas apresentaram agrupamentos particulares entre as réplicas e puderam ser distinguidos padrões de agrupamento

entre as localidades. As amostras costeiras foram agrupadas juntas, muito similares entre si, principalmente PP1 e PA1, que na geograficamente são muito próximas entre si (Figuras 9 e 17). As amostras insulares também ficaram próximas entre si, principalmente FN e AT, que também são muito próximos geograficamente. As localidades do AS se distribuíram de maneira menos linear, concluindo que as diferentes localidades amostradas apresentam condições mais singulares dos parâmetros da água. Já em FN as porções das água de diferentes localidades foram coincidentes entre si, com agrupamentos próximos (Figura 18). Este agrupamento das localidades de FN também se seguiu na PCA, em que os pontos ficaram distribuídos próximos (Figura 19). As réplicas de AC foram muito particulares, distintas as amostras costeiras, porém, também distintas das amostras insulares. Nas duas localidades de AC foram observadas as menores temperaturas, com uma média de 25,54c (Tabela 16). O ponto de Arraial do Cabo foi amostrado com o objetivo de representar uma condição diferenciada aos pontos do NE, em uma área tipicamente tropical afetada pelo fenômeno da ressurgência. Durante este fenômeno que ocorre com a intensificação de ventos de NE, as massas de água profundas (ACAS - Água Central do Atlântico Sul) afloram e resultam em temperaturas entre 15-21°C, geralmente menores que 20°C (COELHO-SOUZA et al., 2017; LESSA et al., 2017). Considerando o mês de amostragem em AC1 e AC2 (março - Tabela 5), uma água mais fria seria encontrada seria encontrada nestas localidades. Porém, por mais que o ponto AC tenha sido o mais frio em relação aos outros pontos costeiros, não foi constatado o efeito da ressurgência nesta amostragem.

Já nas amostras insulares, merece destaque a baixa salinidade encontrada em AS (Tabela 16). O arquipélago de São Pedro e São Paulo está localizado logo acima da linha do Equador, aonde a intensidade de nuvens é conhecidamente elevada, assim como a precipitação no Equador (JIMENEZ et al., 1999; MAIA et al., 2005). Desta forma a elevada precipitação pode ter influenciado na redução direta da salinidade na extrema superfície nesta zona (COELHO-SOUZA et al., 2017; LESSA et al., 2017). Nesta zona, as temperaturas são mais elevadas e maiores salinidades são notadas nos meses de agosto e novembro com valores de 36,5 ou mais de salinidade (PETERSON & STRAMMA, 1991; STRAMMA, 1999; TCHAMABI et al., 2017). A amostragem no AS ocorreu outubro, período no qual a temperatura da superfície do oceano está mais fria nestas águas (~26,5°C), sendo um dos motivos à expeça cobertura de nuvens e a influencia da EUC (PETERSON & STRAMMA, 1991; STRAMMA, 1999; TCHAMABI et al., 2017). Sendo esperado um aumento da temperatura de setembro a abril, com ápice a estação mais calorosa entre março a maio (SILVA et al., 2009; CINTRA et al., 2015). A EUC é mais intensa no Atlântico Oeste e enfraquece à medida que flui para o leste, com meandros entre 0°50'N e 0°50'S (PETERSON & STRAMMA, 1991; STRAMMA, 1999; TCHAMABI et al., 2017). Portanto a EUC realiza meandros justamente nas áreas que estão situadas as montanhas submarinas do AS, com relevante contribuição para a circulação sazonal.

Embora uma amostragem *in situ* não tenha sido realizada no AT consideramos neste trabalho que as mesmas correntes que atuam em FN atuam de maneira muito similar no AT. Além desses arquipélagos estarem muito próximos entre si, e formarem uma área de preservação em comum (MOHR et al., 2009), as médias históricas de POC remote avaliadas neste trabalho demonstram que estão sobre a mesma área de interferência e são constantes (Figura 20). Avaliando as correntes que margeiam FN e AT foi demonstrado que as correntes de superfície e sub-superfície apresentam os mesmos padrões em ambas as ilhas com o mesmo padrão sazonal (REVERDIN et al., 1991; SOUZA et al., 2013; TCHAMABI et al., 2017). Com destaque a contracorrente SEUC que se mostra mais forte de julho a setembro no Atlântico Oeste justamente nas mesmas coordenadas que estão AT e FN. Diferente do AS, os arquipélagos de FN e AT apresentam outro sistema de correntes e contracorrentes.

### **6.1. Concentração celular picofitoplantônica**

Cinco populações foram isoladas de acordo com a fluorescência natural dos pigmentos fotossintéticos e escalas de complexidade e tamanho celular. O principal interesse deste trabalho são as cianobactérias, portanto estas populações foram agrupadas, individualizando de maneira inequívoca apenas células com características morfológicas e estruturais dos gêneros *Synechococcus* e *Prochlorococcus*. As outras três populações foram agrupadas apenas para não serem ruído ou promoverem interferência no isolamento de *Synechococcus* e *Prochlorococcus*, e o agrupamento ocorreu por similaridade funcional. Portanto somente ambos gêneros de cianobactérias tiveram suas abundâncias celulares convertidas em biomassa de carbono (Figura 14). A somatória total (All) destas populações caracterizam as menores frações fitoplanctônicas, nesse caso, células com diâmetros menores que 10µm para viabilização da contagem por citometria de fluxo (Figura 10).

Considerando um gradiente longitudinal, a abundância costeira foi uma ordem de grandeza maior que a abundância das ilhas oceânicas, com  $4,52 \times 10^7$  Cell L<sup>-1</sup> e  $5,85 \times 10^6$  Cell L<sup>-1</sup>, respectivamente (Figura 20). A produtividade primária costeira tende a ser superior em áreas costeiras, principalmente quando comparadas com águas oligotróficas (LALLI & PARSONS, 1997; KIRCHMAN, 2011).

A principal distribuição do bacterioplâncton marinho se encontra nos oceanos abertos do planeta, águas comumente oligotróficas. O bacterioplâncton é maior contribuinte da fração picoplanctônica, com abundância celular superior a  $1 \times 10^8$  Cell L<sup>-1</sup> (VELDHUIS et al., 2005). Já para o Atlântico Sul, um gradiente latitudinal costa brasileira → oceano na cadeia submarina Vitória-Trindade aos 20,5°S, encontrou uma abundância picoplanctônica variando entre  $4,68 \times 10^8$  a  $1,09 \times 10^9$  Cell L<sup>-1</sup> (MEIRELLES et al., 2015). Nesta fração, as cianobactérias representam as bactérias clorofiladas, com os gêneros *Prochlorococcus* e *Synechococcus*, representando a maior parte desta abundância. Em ambientes costeiros, estas concentrações podem ser ainda mais elevadas, devido ao constante aporte de nutrientes (KIRCHMAN, 2011).

Em oceanos abertos a abundância celular de *Prochlorococcus* pode exceder em pelo menos uma ordem de grandeza a abundância de *Synechococcus*, podendo alcançar  $3 \times 10^8$  Cell L<sup>-1</sup> nas latitudes de 40°N a 40°S (VELDHUIS et al., 2005; JOHNSON et al., 2006). Considerando apenas as áreas oligotróficas do giro do Atlântico Norte, foram encontradas uma abundância de  $\sim 3,75 \times 10^8$  Cell L<sup>-1</sup> para *Prochlorococcus* e  $\sim 4,75 \times 10^6$  Cell L<sup>-1</sup> de *Synechococcus* (VELDHUIS et al., 2005). No Atlântico equatorial, a comunidade tende a permanecer dominada por *Prochlorococcus*, embora *Synechococcus* e os piceucariotos tornem-se relativamente abundantes (KIRCHMAN, 2011; FIGUEIRAS et al., 2014). Os arquipélagos oceânicos avaliados neste estudo não seguiram um padrão entre si quanto a dominância dos gêneros, em AS a abundância média de *Prochlorococcus* ( $4,77 \times 10^5$  Cell L<sup>-1</sup>) foi de 1,5-2x superior em relação a *Synechococcus* ( $2,67 \times 10^5$  Cell L<sup>-1</sup>), mas dentro da mesma ordem de grandeza (Figura 22). Já em FN as duas populações foram mais semelhantes, com médias de abundâncias similares entre Procloro e Syneco, com  $1,56 \times 10^6$  Cell L<sup>-1</sup> e  $2,13 \times 10^6$  Cell L<sup>-1</sup>, respectivamente (Figura 23).

No caso das amostras costeiras, a abundância de *Synechococcus* foi superior a *Prochlorococcus* em todos os pontos da costa cearense, com  $7,93 \times 10^6$  Cell L<sup>-1</sup> e  $2,76 \times 10^6$  Cell L<sup>-1</sup> (Figura 21). Estas duas abundâncias somadas representam, em grande parte, a abundância de cianobactérias da costa do NE brasileiro com influência de correntes equatoriais. Um estudo realizado na quebra da plataforma brasileira entre as latitudes 8-23°S revelou uma abundância celular mais acentuada, com  $8,9 \times 10^8$  Cell L<sup>-1</sup>, porém esta abundância está associada às cianobactérias mais as bactérias heterotróficas/mixotróficas (JUNIOR, et al., 2015). *Synechococcus* é uma das cianobactérias mais abundantes nos ecossistemas marinhos, exibindo uma ampla diversidade de pigmentos colonizando desde águas oligotróficas azuis até águas costeiras verde turva. A influência de pigmentação no particionamento de nicho das diferentes espécies e linhagens de *Synechococcus* spp. permanece pouco compreendida, mas percebe-se um melhor aproveitamento destes pigmentos acessórios em águas costeiras ou de baixa salinidade (HUMILY et al., 2014). Alguns ecótipos de *Synechococcus* são mais abundantes nas águas costeiras que nas águas abertas, como na costa da Califórnia, do oceano Pacífico Sul e na porção oeste do Atlântico (HOFFMANN, 1999; HUMILY et al., 2014).

A abundância de *Synechococcus* foi superior na costa em relação a *Prochlorococcus*, e nas amostras insulares (FN + AS) ambos os gêneros apresentaram médias muito próximas, com *Synechococcus* ligeiramente superior. Portanto não houve um gradiente longitudinal quanto a abundância celular, pois *Synechococcus* foi dominante nas amostras costeiras e nas ilhas oceânicas em mar aberto.

## **6.2. Carbono orgânico particulado e biomassa de cianobactérias**

A determinação do estoque de carbono em cada fração fitoplanctônica pode ser estabelecida pela razão da biomassa em cada fração, pela quantidade de carbono orgânico particulado na coluna da água - POC.

Para determinar o POC nas águas superficiais foram utilizados dois métodos distintos, a oxidação com amostras *in situ* e as médias temporais por sensoriamento remoto. O método da oxidação por dicromato de potássio é considerado extremamente robusto para avaliar o carbono orgânico particulado, com um limite de detecção de  $0-500\mu\text{g C L}^{-1}$  ( $1,0\mu\text{g C L}^{-1} = 1,0\text{mg C m}^{-3}$ ), com precisão de  $\pm 10\mu\text{g C L}^{-1}$  (MENZEL & VACCARO, 1964; GRASHOFF et al., 1999). Portanto, os limites encontrados neste estudo estão de acordo com os valores encontrados neste estudo (Figura 18). Wangersky (1976) determinou que a amostragem mais próxima a superfície da água para a determinação de POC *in situ* deveriam ser tomadas a 5m, evitando a contaminação por materiais provenientes de perturbações antrópicas diretas (derramamento de óleo combustível, por exemplo), assim como adotado no presente estudo.

O POC avaliado no momento da coleta apresentou valores médios muito semelhantes entre a costa e os arquipélagos, não apresentando um gradiente longitudinal costa  $\rightarrow$  ilhas definido (Tabela 17). Esta ausência de gradiente foi confirmada por uma ANOVA, não havendo diferença estatística entre as réplicas costeiras e as réplicas insulares, assim como a análise de cluster não agrupou as réplicas seguindo um gradiente longitudinal (Figura 19). A amostra costeira PP1 apresentou o maior valor, com  $34\text{mg C m}^{-3}$ . Esta área se localiza numa região portuária, aonde o fluxo de navios é alto e o fundo esta sendo constantemente ressuspensionado, aportando matéria orgânica do fundo a coluna da água. O arquipélago AS apresentou o menor valor de POC com apenas  $31\text{mg C m}^{-3}$ , esta área se encontra muito distante da costa, sem contato algum com a plataforma continental. As correntes atuantes nesta área carregam POC provenientes da costa da África pela SEC, sendo sua contribuição moduladas pela sazonalidade, apresentando máximas anuais de  $85\text{mg C m}^{-3}$  em 2015 (Figura 19 e Tabela 18).

Um estudo pioneiro na década de 1970 determinou o POC desde a camada superficial até o assoalho oceânico em distintos transectos nos oceanos Atlântico Sul e Pacífico Sul, determinando que a porção superficial em ambos oceanos possuem elevados valores de POC e são semelhantes. A região de maior concentração se estendeu do Equador à  $10^{\circ}\text{S}$  com valores entre  $60-40\text{mg C m}^{-3}$  em superfície durante o verão do Hemisfério Sul, com media de  $54\text{mg C m}^{-3}$  ao Atlântico Sul (WANGERSKY, 1976; KRISHNASWAMI & LAL, 1977). Estes valores estão ligeiramente abaixo dos valores encontrados para as amostras costeiras e aos arquipélagos de AS e FN. Esta ligeira distinção pode ser devido aos diferentes períodos sazonais das amostragens, ou até mesmo à diferenças intrínsecas de metodologia. Embora os trabalhos envolvendo a análise de POC observado *in situ* apresentem um viés pelo emprego de diferentes métodos e distintas coletas e análises, é possível estabelecer comparações desde que estejam na mesma região (WANGERSKY, 1976).

A fim de estabelecer médias temporais, a quantidade de carbono orgânico particulado na coluna d'água também foi estimada pelo sensor orbital MODIS-Aqua - POC remote. Considerando as series mensal, anual e de 5 anos, o POC remote através do desempenho dos algoritmos para a coloração da superfície da água. A interpretação de dados de satélite através da coloração dos oceanos progrediu nos últimos anos, além das estimativas clássicas de concentração de clorofila. A concentração de carbono orgânico em partículas um dos produtos mais recentes da cor do oceano (ŚWIRGOŃ & STRAMSKA, 2015).

O algoritmo da NASA utilizado neste trabalho para analisar as médias e as imagens por sensoriamento remoto para o sensor MODIS vem sendo comprovado de maneira muito eficiente para diversas amostras do oceano Atlântico (STRAMSKI et al., 2008; STRAMSKA, 2014; KOSTADINOV et al. 2016; EVERS-KING et al., 2017). Análises costeiras não foram incentivadas devido as incertezas espectrais que ainda existem perto da zona de saída do rio e perto da costa, pois, para estas amostras costeiras o POC tende a ser muito elevado, principalmente em regiões estuarinas, superando as médias dos algoritmos (MISHONOV et al., 2003; ŚWIRGOŃ & STRAMSKA, 2015). Devido a este fato não foram traçadas as médias das amostras costeiras. Embora este trabalho pretendesse demonstrar a concentração de POC remote apenas em águas superficiais, as análises utilizando a coloração dos oceanos sobrepõem sempre uma parcela significativa do estoque permanente localizado abaixo de uma profundidade de atenuação, posando como um erro constante da metodologia (MISHONOV et al., 2003).

As análises de POC remote indicaram que a media mensal avaliada durante a coleta dos arquipélagos, no mês de outubro, apresentou valores muito semelhantes tanto em AS como em FN, sem diferença estatística entre as réplicas dos dois arquipélagos. Considerando a media dos arquipélagos, houve uma correlação negativa entre o POC observado e o POC remote do mês de outubro. Considerando a media anual dos arquipélagos essa diferença foi ainda maior, sendo que o POC correspondeu apenas metade do volume de carbono particulado que normalmente estaria na água ao longo do ano. Considerando o padrão anual de cada ilha, FN continua com valores próximos ao POC observado e a media mensal, demonstrando a ausência do efeito da sazonalidade sobre o POC. Enquanto que a media anual de AS foi superior a media mensal, e muito superior ao POC observado. A media anual do AS é 2X maior que em FN, assim como pela serie histórica de 5 anos (Tabela 18). Considerando as séries históricas, a região do AS sofre um aporte muito superior de águas vindo da costa africana em superfície, diferente de FN e AT que estão fora desta área de aporte (Figura 19B e 19C).

Através de diversos transectos longitudinais e latitudinais em águas superficiais do Atlântico Sul, tanto da porção oeste como leste, análises por sensoriamento remoto determinaram variações sazonais de 20-30% na concentração calculada de POC, e com valores de POC entre 30-50mg C m<sup>-3</sup> na zona equatorial do Atlântico durante o verão (MISHONOV et al., 2003). Embora uma avaliação sincrônica entre a coleta e análise do POC por técnicas

químicas e pelos sensores orbitais, foi possível avaliar a distribuição em toda a bacia com uma precisão aceitável, menor que 10% de variação (MISHONOV et al., 2003). Os valores dos transectos acima são muito similares aos encontrados em AS e FN, sendo que as amostragens de ambos os trabalhos nessas áreas foi similar outubro/novembro.

Nas camadas de superfície os valores POC estão distribuídos em função da produtividade e, assim, exibem uma variação sazonal, bem como uma variabilidade determinada pela hidrografia local (WANGERSKY, 1976). Cerca de 0,6-1,5% do POC encontrado na zona equatorial do Atlântico foi proveniente do estoque de clorofila no oceano aberto do Atlântico Sul em superfície durante o verão. Esta diferença pode indicar as áreas as quais a produção primária é nova (elevada POC pela clorofila) ou áreas em a produção seja mais velha, com baixa clorofila e determinar onde o POC esta sendo acumulado (MISHONOV et al., 2003). A concentração do POC varia mais significativamente em função da composição da comunidade planctônica, de que com a profundidade. Desvios negativos de 11% no POC foram associados em áreas de dominância de diatomáceas em relação a áreas com a comunidade de reciclagem (picofitoplâncton) (CETINIĆ et al., 2012). Recentemente, algoritmos bio-ópticos estão estimando a relação carbono-clorofila e os estoques permanentes de carbono de fitoplâncton divididos em várias classes de tamanho, pela cor do oceano. A abordagem combina as relações alométricas empíricas da estrutura e tamanho fitoplânctônico, com um algoritmo baseado na absorção estimada em espectros desenvolvidos anteriormente para o tamanho fitoplânctônico (ROY et al. 2017). No presente estudo, o estoque de carbono encontrado nas populações picofitoplânctônicas não foi estimado por algoritmos bio-ópticos, mas sim por algoritmos de biovolume que utilizam da concentração celular para estimar a biomassa.

As biomassas de *Prochlorococcus* e *Synechococcus* foram estimadas tanto para as amostras costeiras como para as insulares. Somadas as populações de *Prochlorococcus* e *Synechococcus* se estima a biomassa costeira e insular das cianobactérias na zona equatorial do oceano Atlântico Oeste. As cianobactérias estocam 1,04mg C m<sup>-3</sup> na costa e apenas 0,15mg C m<sup>-3</sup> nas ilhas oceânicas equatoriais brasileiras, apresentando um claro gradiente longitudinal costa → arquipélagos oceânicos: PP>PA=IC>AC>FN>AS (Figura 25 e Tabela 19). Avaliando a biomassa costeira junto com a dos arquipélagos, as populações de Syneco e Procloro apresenta diferença significativa entre suas médias. Porem, a variação ponto a ponto semelhante entre ambas populações, tanto nas amostras costeiras quanto nas insulares. Essa variação foi confirmada por uma significativa e elevada correlação (corPearson=0,9316), confirmando a proposição do gradiente.

Regiões de plataforma tropical com o aporte de rios estão entre os ambientes produtivos do oceano, já aqueles com prateleiras costeiras estreitas afetadas por correntes de fronteira estão entre os menos produtivos, o caso da costa cearense (EKAU & KNOPPERS, 1999). A plataforma tropical e subtropical do oeste do Atlântico Sul, dentro do território brasileiro, é mais

produtiva na plataforma amazônica no norte e ao sul do Brasil. Pouco se sabe sobre a produtividade do restante das prateleiras brasileiras, governadas por correntes de influências equatorial (EKAU & KNOPPERS, 1999).

Na plataforma NE foi estimada por avaliações sinópticas uma baixa produção primária, devido as águas pobres da SEC e portanto uma biomassa de planctônica baixa em relação as outras áreas da costa brasileira (EKAU & KNOPPERS, 1999). Sabe-se mais sobre os estoques de zooplâncton da área de que a composição fitoplanctônica (EKAU & KNOPPERS, 1999; SOUZA et al., 2016). No Nordeste, áreas com enriquecimento de nutrientes por processos de ressurgência ocorrem apenas na borda da prateleira, longe da linha de costa. Uma termoclina estável entre 100 e 150m em mar aberto a nordeste do Brasil, resultando na formação de uma massa de água estratifica chamada de Água de Superfície Equatorial (ESW), esta é carregada pela costa do NE através da ação da SEC (EKAU & KNOPPERS, 1999). Todo este cenário reflete em espécies fitoplanctônicas dominantes e biomassas mais altas restritas as águas próximas da costa e com pouca mudança no domínio oceânico. A composição de espécies e concentrações de biomassa são típicas de águas oceânicas oligotróficas (EKAU & KNOPPERS, 1999), assim como encontrado para a costa cearense em comparação com as áreas subtropicais.

O estoque relativo também foi estimado considerando a biomassa de cianobactérias na costa e nos arquipélagos. A porcentagem de carbono orgânico em suspensão (POC) na coluna da água que esta estocado em células viáveis correspondem ao estoque relativo (Tabela 19). O estoque relativo na costa esta concentrado mais em *Synechococcus* de que em *Prochlorococcus*, com 2,69% e 0,48% do POC em águas superficiais, respectivamente. Nas amostras insulares o estoque relativo é muito inferior em relação a costa, e *Synechococcus* (0,38%) também estoca mais carbono que *Prochlorococcus* (0,10%). Embora exista um gradiente de biomassa de cianobactérias, o gênero *Synechococcus* garante o maior estoque de carbono ao longo do gradiente.

A biomassa máxima de cianobactérias avaliada através de transectos latitudinais na porção mesoceânicas do Atlântico, estimou que em latitudes médias (borda da zona tropical) são encontradas concentrações de  $\sim 1,39 \text{ mg C m}^{-3}$  na porção superficial. Com dominância de *Prochlorococcus* nos giros norte e sul do Atlântico e na zona equatorial, dando lugar a *Synechococcus* em águas mais frias nas extremidades dos transectos (ZUBKOV et al., 1998; 2000).

Heywood e colaboradores (2006) realizaram um grande inventário sobre a biomassa estocada nas cianobactérias em distintos transectos latitudinais na porção mesoceânicas do Atlântico, dos giros oligotróficos do norte e do sul e na região equatorial. Houve uma semelhança muito grande entre as metodologias abordadas entre Heywood et al. (2006) e este trabalho, pois ambos utilização a citometria e os algoritmos de bio-volume para estimar a biomassa. Porem os dados foram expostos por área ( $\text{m}^{-2}$ ) ao invés de volume ( $\text{m}^{-3}$ ). A biomassa

de *Prochlorococcus* na camada superficial foi de 173mg C m<sup>-2</sup> no giro oligotrófico do norte, 190mg C m<sup>-2</sup> no giro oligotrófico sul e 141mg C m<sup>-2</sup> na região equatorial. Para *Synechococcus* a biomassa foi menor em cada uma das três províncias com 18, 17 e 32mg C m<sup>-2</sup>, respectivamente. A zona equatorial, apresentou a menor biomassa de Procloro em relação aos giros oligotróficos, em contra partida Syneco apresentou maior abundancia no equador de que nos giros. Os dados não mostraram uma variabilidade interanual estatisticamente significativa na biomassa deste gêneros (HEYWOOD et al., 2006). No oceano Pacífico Norte central, a biomassa integrada em área de *Prochlorococcus* foi muito superior (973mg C m<sup>-2</sup>) que a de *Synechococcus* (58mg C m<sup>-2</sup>), indicando uma dominância no estoque de Procloro nestas áreas oligotróficas (CAMPBELL et al., 1994), diferente do que foi encontrado nos estoques costeiros e insulares da zona equatorial do Atlântico Oeste (Tabela 19).

Estimativas zonais, como a proposta neste estudo, fomentam os bancos de dados mais abrangentes, que conseguem somatizar a biomassa estocadas nas frações planctônicas ao longo eixo horizontal e vertical dos oceanos. Uma estimativa da variação de carbono foi realizada considerando um transecto meridional no oceano Atlântico, revelando valores de POC variando de 9-229 µg L<sup>-1</sup>; 4-58 µg L<sup>-1</sup> para o carbono estocado no fitoplâncton; e 0,026-1,13 µg L<sup>-1</sup> para clorofila-a (GRAFF et al., 2015). Uma média de 44% do volume de POC encontrado no Atlântico está estocada no fitoplâncton (GRAFF et al., 2015). Estimativas baseadas em satélites calculadas usando um conjunto diferente de valores de profundidade da camada mista, com estimativas derivadas de Stramski et al. (2008), sugerem um estoque global de carbono em torno de 0,24Pg C para o fitoplâncton, podendo variar de 85-40% do POC total no Atlântico (KOSTADINOV et al. 2016). Outro estudo recente encontrou um valor muito próximo ao supracitado, 0,26Pg C estocados no fitoplâncton mundial (ROY et al. 2017). Fracionando o fito por classe de tamanho cerca de 54% do estoque global de carbono de fitoplâncton está dentro da fração picoplanctônica (~0,14Pg C); cerca de 31% para o nanoplâncton (~0,08Pg C); e 15% (~0,14Pg C) para o micropoplâncton (ROY et al. 2017).

A profundidade da remineralização dos fluxos de carbono orgânico particulado (POC) exportados da superfície dos oceanos exercem um grande controle sobre os níveis atmosféricos de CO<sub>2</sub>. De acordo com um paradigma de longa data, a maioria do POC exportado para o assoalho oceânico está associado a partículas grandes. No entanto, linhas recentes de evidências sugerem que POC de menor tamanho (dimensões semelhantes a dos picofito) são de dissipação lenta e possuem um papel mais importante de que se previa (BAKER et al., 2017). Analisando altas e médias latitudes norte e sul, a concentração médias do POC de menor tamanho (5,5mg C m<sup>-3</sup> é 13 vezes menor que as concentrações de POC em superfície e 75 vezes maior do que as concentrações de POC maiores de rápido afundamento. Os modelos biogeoquímicos geralmente não reproduzem esse comportamento, o que significa que provavelmente superestimam o armazenamento de carbono oceânico a longo prazo. Para corrigir estes erros, caracterizações

locais do POC locais devem ser empregadas, principalmente em áreas como poucas informações como as zona equatoriais dos oceanos (BAKER et al., 2017). Os reservatórios de carbono de biomassa organicamente ligados e vivos nos oceanos são significativamente menores do que os reservatórios inorgânicos. No entanto, o crescimento contínuo do plâncton na superfície mantém a concentrações muito inferiores de CO<sub>2</sub> ao que seria sem eles na camada superficial dos oceanos (HEINZE et al., 2015). Todos estes estudos são apenas concentrados no Atlântico e Pacífico Norte, sendo muito generalistas, nem sequer citando ou isolando a porção equatorial.

Considerando que os oceanos regulam as concentrações atmosféricas de CO<sub>2</sub>, sabe-se atualmente que cerca de 25% das emissões antrópicas anuais de carbono na atmosfera estão sendo absorvidas nas camadas superficiais dos oceanos (HEINZE et al., 2015). Com o aumento no estoque de carbono oceânico afetará seus ecossistemas e seus serviços ecológicos. A acidificação dos oceanos está intimamente relacionada ao ciclo do carbono e a remineralização, uma vez que o CO<sub>2</sub> dissolvido na coluna da água pode ser utilizado para formar biomassa pela fotossíntese ou ser concentrado e aumentar a pressão de dissolução deste gás. Uma vez concentrado, o CO<sub>2</sub> dissolvido na água aumenta a acidificação e reduz a bomba de carbonato de cálcio (CaCO<sub>3</sub>). Portanto os efeitos do aquecimento global são diretamente sentidos pelos estoques lábeis das distintas populações fitoplanctônicas (HEINZE et al., 2015).

### **6.3. Diversidade de cianobactérias**

#### **6.3.1. Abundância de sequências**

Diversos primers são citados para amplificar regiões do 16S rRNA em cianobactérias (Tabela 7). Nübel e colaboradores (1997) foram alguns dos pioneiros ao determinarem iniciadores que retratassem de maneira eficiente a composição das cianobactérias na estrutura de comunidades. Alguns conjuntos de primers (forward e reverse) foram determinados para cianobactérias cocóides unicelulares, muito propício para trabalhos que o objetivo de amplificação dos gêneros *Prochlorococcus* e *Synechococcus*, CYA359F-CYA781R(b) (NÜBEL et al., 1997). Estes conjuntos de primers que amplificam a região conservada V4 do rRNA continuam a serem utilizados até o presente momento, comprovando sua eficiência tanto em técnicas moleculares de DGGE com adaptadores de grampo CG (BOUTTE et al., 2005; MÜHLING et al., 2008), assim como nos sequenciamentos de nova geração com adaptadores específicos a cada método (ROCAP et al., 2002; YU & MORRISON, 2004). Porém, no próprio trabalho original de Nübel et al. (1997), seus conjuntos de primers foram muito eficientes em amplificar cianobactérias, mas também amplificou outros grupos taxonômicos, como exemplares do filo Proteobacteria. Tentando corrigir este erro, foi incorporada uma amplificação em duas etapas, PCR semi-nested, primeiro amplificando uma área maior, depois restringindo a porção de interesse a partir do mesmo iniciador forward (BOUTTE et al., 2005; MÜHLING et al., 2008). Esta técnica de amplificação possibilitou uma resolução ainda melhor na recuperação

com 91-83% de recuperação aos grupos alvos, no entanto, também aumentou a porcentagem de grupos não alvos com 16-42% pela técnica de DGGE (MÜHLING et al., 2008). Também foi notada uma melhor recuperação na diversidade com uso de PCR semi-nexted em amostra sequenciadas no Illumina e 454 (pirosequenciamento) (SINCLAIR et al., 2015). As frações 3 e 0,22µm analisadas no presente estudo também se utilizou da PCR semi-nexted, dentre dos 56 táxons identificados apenas 7 não eram do filo das cianobactérias (Tabela 23). Outros trabalhos veem descrevendo novos conjuntos de primers as cianobactérias, incluindo regiões de ITS (23S-16S) e até mesmo outras regiões não variadas como a V6 (ROCAP et al., 2002; LAVIN et al., 2008; THOMPSON et al., 2011).

Um passo fundamental na análise de sequenciamento de microbiomas é a atribuição de leitura das unidades taxonômicas. Uma comparação entre os principais bancos de dados para 16S rRNA, os bancos SILVA, RDP (Ribosomal Database Project) e GG (Greengenes) mapeiam bem nos banco de dados maiores como o NCBI e OTT (Open Tree of Life Taxonomy) (BALVOČIŪTĖ & HUSON, 2017). Todos os bancos de dados atribuem fileiras a seus nós: domínios, filo, classe, ordem, família, gênero e espécie. No entanto, RDP e SILVA descem apenas para o nível do gênero, enquanto NCBI, GG e OTT descem para o nível de espécie e abaixo. O banco de dados SILVA, possui um maior acervo em relação aos demais bancos de 16S (GG e RDP), além de atualizações mais frequentes (BALVOČIŪTĖ & HUSON, 2017). Tal banco foi utilizado para identificar as cianobactérias das frações 3 e 0,22µm na costa e nos arquipélagos neste trabalho. E uma das maiores vantagens é o processo manual de incorporação de sequências, curadas manualmente (YILMAZ et al., 2014). A taxonomia do SILVA veem sendo amplamente utilizada para estudos de 16S (BALVOČIŪTĖ & HUSON, 2017), até mesmo em projetos de escala mundial aplicados as cianobactérias, como no Tara Ocean Project (SUNAGAWA et al., 2015; FARRANT et al., 2016).

A filogenia de cianobactérias sempre foi considerada muito complicada, tanto que foi recentemente utilizada como estudo de caso a um novo software de identificação e qualidade de sequências, o SATIVA (Semi-Automatic Taxonomy Improvement and Validation Algorithm) (KOZLOV et al., 2016). Relativo ao estudo de caso das cianobactérias, houve a comparação do SATIVA em comparação com os outros bancos de dados 16S, como o GG, SILVA e RDP. O RDP parte do agrupamento em 13 famílias, enquanto que o Greengenes reconhece seis ordens. Já o SILVA reconhece 5 subsecções, que seriam congruentes as ordens acima, porém com grupos polifiléticos (KOZLOV et al., 2016; BALVOČIŪTĖ & HUSON, 2017). O SATIVA classificou 9 sequências erradas (1%) para o SILVA, 0 para RDP e 1 (0,1%) para Greengenes. Tais porcentagens foram consideradas baixas em todas as taxonomias, a maioria dos erros de rotulagem foi identificada no nível do gênero (KOZLOV et al., 2016).

Outra questão importante a ser considerada para uma taxonomia correta junto aos bancos de dados é a filtragem das sequências quiméricas, realizada no passo anterior ao

alinhamento taxonômico (WRIGHT et al., 2012). Quimeras são sequências não naturais formadas a partir de duas ou mais sequências biológicas distintas, formadas frequentemente durante o processo de PCR. Embora as quimeras sejam raras durante sequenciamento shotgun, elas ocorrem (HAAS et al., 2011). Devido a utilização do banco de dados SILVA acoplado a as plataformas QIIME e BMP durante as análises taxonômicas das frações 3 e 0,22 $\mu$ m, foi possível retirar as sequências quiméricas utilizando específicos ao software UCHIME (Robert C. Edgar - Pesquisador Independente; CAPORASO et al., 2010; KIM et al., 2013; PYLRO et al., 2014).

Ao final do alinhamento, os fragmentos 16S rRNA recuperadas do sequenciador e posteriormente tratadas e alinhadas pelo banco de dados SILVA, somou um total de 1.158.901 sequências em ambas frações fitoplanctônicas (3 e 0,22 $\mu$ m). A técnica shotgun foi utilizada para compor a biblioteca metagenômica 16S rRNA destas frações. Deste total, 73,48% foi proveniente das amostras insulares, com o AS dominando em relação a FN e a todas as outras amostras costeiras. Apenas 26,52% das sequências total estavam contidas nas amostras costeiras, com PP sendo superior. Arraial do Cabo apresentou a menor quantidade de sequências dentre outras amostras costeiras e insulares. Deste total de sequências alinhadas, uma grande representavam novas sequências, ainda não registradas no banco de dados. Estas sequências foram chamadas de New (Tabela 23). Um trabalho na cadeia de montes marinhos Vitória-Trindade (gradiente costa-oceano no Atlântico Sudoeste) encontrou 53,12% das sequências metagenômicas identificadas pela taxonômica de bancos de dados, sendo que quase metade ainda tem potencial desconhecido de diversidade (MEIRELLES et al., 2015). Um transecto longitudinal na porção sudoeste do Atlântico Sul, após a quebra da plataforma brasileira foram encontrados mais de 400 mil metagenomas, 36,7% não tiveram homólogos contra as 17 milhões de sequências comparadas com o Atlântico Norte, Atlântico Sul, Oceano Pacífico Norte, Pacífico Sul e Índico. Esse percentual pode ser considerado baixo, demonstrando a elevada e desconhecida diversidade encontrada nesta porção do Atlântico que margeia a costa brasileira (JUNIOR, et al., 2015).

A falta de informação sobre os números de cópias do 16S rRNA em relação aos tamanhos dos genomas podem tornam as estimativas de abundância relativa imprecisas em populações bacterianas complexas. Supõe-se que as cópias dos genes rRNA dentro de um organismo são sujeitas a homogeneização, mesmo assim, as sequências 16S da mesma espécie (OTU) se apresentam com um comportamento diferente, não padronizado ou em repetição. Resultando em uma quantidade de 16S cerca de 2,5 vezes maior do que o número de espécies bacterianas (ACINAS et al., 2008). Desta forma considerar que a abundância de sequências seja exatamente igual a abundância de células seria um erro, pois o número de sequências está limitado pelas multicópias intrínsecas da própria célula e pela variabilidade intragenômica do marcador molecular. Mas é fato considerar que ambas abundâncias devam seguir razões e padrões similares, respeitando as suas grandezas próprias (VĚTROVSKÝ & BALDRIAN,

2013). Já as análises de dados de metagenômica shotgun amplificados por PCR não são diretamente afetadas pelos números de cópias do 16S rRNA (VĚTROVSKÝ & BALDRIAN, 2013). No presente trabalho, tanto a abundância celular como a abundância de sequências foram realizadas. As abundâncias apresentam padrões semelhantes nas amostras costeiras, guardadas as respectivas proporções, já nas amostras insulares não houveram semelhança. Nas localidades do AS foram estimadas as menores concentrações celulares (Figura 20), enquanto que no sequenciamento revelaram a maior abundância de fragmentos (Figura 30).

As análises estatísticas de redução de dimensionalidade das sequências, testes multivariados de PCA e MDS, possibilitaram uma estimativa prévia de diversidade local. Isso é possível observando a distribuição das réplicas de cada localidade em função da variação dos táxons pela abundância das sequências. Nas frações 3 e 0,22 $\mu$ m, os pontos de AC e FN tiveram as suas réplicas melhor distribuídas entre si e ainda dissimilares das outras (Figuras 32 e 33). Enquanto que as réplicas de AS agruparam suas abundâncias de sequências aos mesmos táxons, se sobrepondo, revelando maiores dominâncias e menor diversidade local (Figuras 32 e 33). Estas observações estatísticas foram posteriormente confirmadas pelos índices de  $\alpha$ -diversidade, onde AC e FN apresentam elevada diversidade, e AS uma baixa diversidade (Tabela 22).

Dentre todos os táxons que tiveram sequências alinhadas em ambas frações 3 e 0,22 $\mu$ m, o gênero *Prochlorococcus* apresentou as maiores médias de abundância de sequências nas amostras insulares. Destaque para AS que apresentou médias superiores a todas as outras localidades, estatisticamente diferente à todas. Já o gênero *Synechococcus* apresentou as maiores médias costeiras, sendo PP significativamente diferente entre todas as localidades costeiras.

Sequências “não assinadas” (unassigned) ocorrem com frequência quando se realiza o alinhamento de sequências 16S aos bancos de dados, inclusive em alguns trabalhos estas sequências representam grande parte da abundâncias alinhadas (THOMPSON et al., 2011). Uma nova metodologia de alinhamento e teste de qualidade de sequências, a plataforma SATIVA, isolou 1050 sequências de cianobactérias para a inferência filogenética aos outros bancos de dados abertos para comparação. Destas apenas 900 foram utilizados, pois o restante foram classificadas como Unassigned a algum dos bancos (KOZLOV et al., 2016). Foram encontradas em somente algumas amostras da costa e dos arquipélagos grandes abundâncias de sequências não assinadas, nas frações 3 e 0,22 $\mu$ m. Desta forma esta sequências unassigned também foram utilizadas para diferenciação das amostras quanto a taxonomia (Tabela 23).

### **6.3.2. Diversidade taxonômica**

A biodiversidade de cianobactérias nos ecossistemas marinhos tropicais é maior nas zonas litorâneas, quando comparadas ao oceano aberto, onde formam esteiras intertidais na coluna da água, vivem como endólitos em substratos carbonáticos ou formam associações simbióticas. Existem diversos nichos e zonas ao longo do planeta que ainda não possuem dados suficientes

sobre a ecologia das cianobactérias, principalmente em áreas equatoriais e zonas costeiras infralitorais menos acessíveis (HOFFMAN, 1999). Dentre as poucas informações de diversidade de cianobactérias em águas superficiais na zona equatorial do oceano Atlântico, a maioria descreve médias gerais as porções mesoceânicas (ZWIRGLMAIER et al., 2008).

Uma análise global demonstrou que a distribuição de *Prochlorococcus* e *Synechococcus* são notavelmente semelhantes em diferentes sistemas oceânicos desde que tenham condições ambientais similares. Quatro distintos biomas oceânicos foram estabelecidos: o Domínio Polar, o Domínio das Fronteiras Costeiras, o Domínio dos Ventos Alísios (zona equatorial e tropical) e o Domínio dos Ventos Ocidentais (subtropical). *Prochlorococcus* é mais abundante em águas oligotróficas quentes (>15°C), enquanto o *Synechococcus* domina em áreas costeiras e (ZWIRGLMAIER et al., 2008). O domínio costeiro e as águas mesotróficas do oceano aberto temperadas são dominados por *Synechococcus*. O domínio dos ventos alísios é dominada por *Prochlorococcus*, e com presença de *Synechococcus* em baixa e moderada abundância (ZWIRGLMAIER et al., 2008).

A mesma defasagem encontrada aos ambientes equatoriais, se aplica a plataforma costeira brasileira e as bacias oceânicas adjacentes, contudo, alguns trabalhos veem descobrindo feições gerais da comunidade do bacterioplâncton marinho nestes destes ambientes (THOMPSON et al., 2011; JUNIOR, et al., 2015).

O perfil de comunidade do bacterioplâncton da costa ao longo de um gradiente latitudinal entre Porto Rico e Argentina computou um surpreendente número de 937 à 1946 OTUs (16S rRNA) correspondentes às espécies. Aproximadamente 80% da abundância total de sequências corresponderam ao filos Proteobacteria, Bacteroidetes, Actinobacteria e Cyanobacteria. Um enriquecimento de cianobactérias foi observado em Porto Rico (THOMPSON et al., 2011), região influenciada pela corrente NBC, a mesma que margeia a costa cearense. Ao longo de sete zonas definidas pelo trabalho, seguindo um gradiente latitudinal, a riqueza observada foi inversamente proporcional a abundância das sequências. A zona de interferência da pluma do rio Amazonas (plataforma norte do Brasil) apresentou as menores abundâncias, em contrapartida apresentou as maiores riquezas (THOMPSON et al., 2011).

Estudos recentes demonstram que os gêneros *Prochlorococcus* e *Synechococcus* podem ter sua biodiversidade melhor representadas por análise de ecótipos, ao invés da classificação clássica de diversidade e riqueza atribuídas ao elemento taxonômico fundamental de espécie (THOMPSON et al., 2013a; BATMALLE et al., 2014; LINDELL, 2014; FARRANT et al., 2016).

O projeto Tara Oceans revelou uma complexa e ampla distribuição biogeográfica/biodiversidade dos distintos ecótipos de *Prochlorococcus* e *Synechococcus*. Foram consideradas leituras de 109 metagenomas da expedição numa escala oceânica global (FARRANT et al., 2016). Para definir um novo significado aos ecótipos, o termo OUT foi

substituído por ESTU - Unidade Taxonômica Ecologicamente Significativa (Ecologically Significant Taxonomic Units), uma proposição mais ecossistêmica e menos taxonômica. Os ESTUs são organismos pertencentes a um mesmo clado e que ocupam um nicho oceânico comum. Três principais assembleias de ESTUs foram identificados para *Prochlorococcus* e oito para *Synechococcus*. Todos estes ecótipos foram muito vinculados a intensidade luminosa e temperatura (FARRANT et al., 2016).

Os ecótipos de *Prochlorococcus* e *Synechococcus*, representam a nova riqueza de espécies ao grupo das cianobactérias. O estudo da estrutura comunidade estão tomando novos formatos após as recentes incorporação da técnicas metagenômica. Cada vez mais pequenas nuances, entre a variação de linhagens, dentro de um mesmo gênero estão sendo melhor relacionadas e explicadas as condições ambientais de seus nichos. Um bom exemplo é o estudo do microbioma associado a raízes de uma única espécie de árvore (COLARES et al., 2013). Estes estudos utilizam da interpolação com bioinformática para explicar os padrões ecossistêmicos de maneira multivariada, anteriormente os estudos somente podiam comparar localidades através dos parâmetros clássicos de riqueza, diversidade e equitabilidade. No entanto, esta visão ecológica moderna somente pode avançar graças aos estudos clássicos pretéritos que disponibilizaram os primeiros registros/padrões de biodiversidade nos mais diversos ambientes, subsidiados com os clássicos índices de riqueza e diversidade (ODUM, 2007). Devido ao escasso registro de informações ecológicas das frações fitoplantônicas, o presente estudo utilizou técnicas moleculares e ferramentas de bioinformática para descrever o registro pioneiro clássico da estrutura da comunidade de cianobactérias num gradiente longitudinal da costa cearense aos arquipélagos oceânicos influenciados pelo regime de correntes equatoriais do Atlântico.

A riqueza observada da fração nanofitoplantônica aumentou ao longo do gradiente costa (39 táxons)→ilhas (45 táxons), enquanto que a diversidade local foi similar. Já na fração picofitoplantônica, houve perda de riqueza e diversidade local no gradiente costa (34 táxons)→ilhas (30 táxons). Para ambas frações, as análises de  $\beta$ -diversidade distinguiram PA na costa e FN4 nas ilhas como as localidades mais diversas de cada ambiente (Figura 36).

O gradiente longitudinal costa → arquipélagos oceânicos da zona equatorial do Atlântico Oeste é notado apenas para a abundância celular e biomassa (Figuras 20 e 25). Para a abundância de sequências e diversidade de cianobactérias um gradiente costa → ilhas não pode ser definido, pois os arquipélagos são muito distintos entre si. As localidades de FN, apresentaram elevada diversidade e muitos táxons raros, mesmo com abundância média de sequências. (Figura 30 e Tabela 22). Tchamabi et al. (2017) também salienta uma notável produtividade e biodiversidade com muitas espécies endêmicas, pronunciando um "efeito de massa na ilha" em FN.

Os arquipélagos oceânicos do Brasil podem ser considerados "oásis de vida em um deserto oceânico" na porção oeste do Atlântico Equatorial (HAZIN, 1993). Esse aumento do biomassa é

resultado das maiores turbulências ao redor das ilhas e montes submarinos, que conduzem águas subterrâneas enriquecidas à superfície, aumentando a produção primária e consequentemente aumentando os fluxos de massa e energia em toda a cadeia trófica (HAZIN, 1993). Regiões emergentes, de pequena escala, com maiores concentrações de nutrientes foram encontradas nas proximidades das ilhas oceânicas do Brasil (MACEDO et al., 1998). A produção primária e a clorofila integrada com mais de 100 m de profundidade na sequência de FN e AT foram mais significativas nas áreas próximas (CORDEIRO et al., 2013).

Integrando o gradiente longitudinal, a fração 3,00 $\mu$ m teve uma maior  $\alpha$ -diversidade de que a fração 0,22 $\mu$ m (Tabela 22). As curvas de rarefação também revelaram que a fração 3 apresentou uma maior cobertura de que a fração 0,22 $\mu$ m, faltando amostragens para atingir o platô da curva nesta última fração (Figura 35). Vaillot e col. (2008) também consideraram o limiar de 3 $\mu$ m para separar o pico do nanofitoplâncton em amostras naturais. Tais autores também relataram uma maior diversidade nesse limiar, porém, de organismos eucariotos (18S rRNA). Confirmando a ideia de que as cianobactérias sejam o grupo mais abundante nas águas superficiais dos oceanos em qualquer gradiente, mas não necessariamente o mais diverso (VAULOT et al., 2008).

Ao todo, 1.158.901 sequências foram representadas em 56 táxons, com 48 pertencem às cianobactérias. Deste total, 37 táxons foram comuns entre as frações 3 e 0,22. Isolando as frações do fitoplâncton pelas localidades “costa e ilhas” são compostos 4 cenários distintos, nos quais apenas 20 táxons são comuns entre eles (Figura 36). Comprovando que ambas localidades em ambas frações são muito semelhantes, e possuem muita dominância de poucas espécies. Apenas 2 táxons foram coincidentes entre as amostras costeiras de ambas as frações e apenas 1 entre as frações insulares (Tabela 23). As amostras dos arquipélagos na fração nanofitoplanctônica foram as mais “VIPs” dentre os 4 cenários, com 9 táxons exclusivos.

Os táxons de maior abundância relativa de sequências, ou seja, táxons com abundâncias acima de 0,01%, foram selecionados a cada fração fitoplanctônica. Na fração 3,00 $\mu$ m, 12 táxons foram considerados os mais relevantes em abundância. *Prochlorococcus* englobou quase 70% do total de 467.497 sequências atribuídas a esta fração. Seguido de *Synechococcus* com 29% das sequências, New Haptophyceae (0,38%), New Stramenopiles (0,21%), *Rivularia* (0,16%), *Halomicronema* (0,14%), New Xenococcaceae 1 (0,10%), New Xenococcaceae 2 (0,09%), New Pseudanabaenaceae 2 (0,07%), Cyanobacterium (0,06%), *Microcystis* (0,06%), New Chloroplast (0,05%). Um total de 691.404 sequências foram identificadas na fração 0,22 $\mu$ m, superior ao total da fração 3,00 $\mu$ m. Este maior montante de sequências foi distribuído em apenas 7 táxons considerado pela abundância relativa. Isto sugere que os grupos da fração picofitoplanctônica demonstraram maior dominância de que a fração nanofitoplanctônica, além de uma maior abundância de sequências. Na fração 0,22 $\mu$ m, *Prochlorococcus* apresentou 73% das sequências, seguido de *Synechococcus* com 26%, New Haptophyceae (0,24%), New

Xenococcaceae 2 (0,07%), New Chroococcales 1 (0,01%), New Phormidiaceae 2 (0,02%), *Xenococcus* (0,02).

Embora alguns táxons tenham sido relativamente abundantes, os gêneros *Prochlorococcus* e *Synechococcus* foram os dominantes, acumulando cerca de 98% das sequências totais em ambas frações fitoplanctônicas. A abundância de sequências foi tão dominante nestes gêneros, que somente houveram correlações entre as localidades e os gêneros *Prochlorococcus* e *Synechococcus* pelos heatmaps de ambas frações (Figuras 37 e 38). As correlações de Spearman entre os táxons de maiores abundâncias relativas, são significativamente negativos entre os gêneros *Prochlorococcus* e *Synechococcus* tanto na fração 3, como na 0,22 $\mu$ m. Esta oposição é demonstrada nos heatmaps, com as réplicas costeiras se correlacionam positivamente com *Synechococcus* e negativamente com *Prochlorococcus*, enquanto que as réplicas insulares são oposto, sendo positivamente correlacionadas com a dominância de *Prochlorococcus*.

Doze diferentes localidades alinhadas num transecto latitudinal no Atlântico Sudoeste, talude e bacia oceânica com influência da corrente do Brasil (de 8 a 23°S), confirmou uma grande diversidade microbiana. Porém, uma elevada cobertura permitiu recuperar os grupos dominantes de procariotos nas massas de água, sugerindo que a comunidade microbiana foi dominada pelos gêneros *Prochlorococcus* e *Synechococcus* (JUNIOR, et al., 2015). Apesar dos tremendos avanços na oceanografia microbiana, pouco se sabe ainda sobre a diversidade microbiota no oceano Atlântico Sudoeste, uma vez que a maioria dos estudos se concentram nos oceanos do Atlântico Norte e Pacífico (JUNIOR, et al., 2015).

Muito recentemente uma pesquisa, também pioneira, que revelou a estrutura da comunidade de cianobactérias em águas superficiais de um estuário oligotrófico na porção sul da costa nordestina (riqueza e abundância). Biblioteca metagenômica shotgun 16S rRNA e o kit Nextera XT V3 do Illumina para o sequenciamento foram as metodologias taxonômicas utilizadas (AFFE et al., 2018). Situações geográficas e metodológicas muito semelhantes as analisadas no presente estudo as frações 3 e 0,22 $\mu$ m da costa nordestina e nos arquipélagos oceânicos. Foi encontrada uma elevada riqueza associada à elevada hidrodinâmica da área, com 219 táxons de cianobactérias observados, com as maiores riquezas atribuídas ao picofitoplâncton, nos gêneros *Prochlorococcus* (58 OTUs) e *Synechococcus* (34 OTUs). *Synechococcus* englobou quase 90% das sequências atribuído ao nível do gênero recuperado em todas as amostras, seguido de *Prochlorococcus* com 5% das sequências, *Microcystis* (0,3%), *Leptolyngbya* (0,6%), *Cyanobium* (0,7%), *Trichodesmium* (0,8%), e outros gêneros também foram observados, compreendendo 2% das sequências (AFFE et al., 2018). Uma avaliação sazonal em dois diferentes períodos de chuvas (estações chuvosa e seca), determinou uma contribuição no aumento da riqueza na estação chuvosa (água doce), causando uma variação na comunidade de cianobactérias. Porém, com prevalência de *Synechococcus* em maior abundância em todos os períodos. O índice de diversidade de Shannon também foi similar durante todo o

ciclo das marés. No entanto, em ambas as estações chuvosa e seca, um aumento na abundância relativa de *Prochlorococcus* ocorreu durante a maré alta, demonstrando a influência da água salgada oligotrófica externa a baía (AFFE et al., 2018).

#### **6.4. Diversidade e variáveis oceanográficas**

Os parâmetros da água apresentaram médias costeiras próximas às médias dos arquipélagos, porém diferenças sutis foram suficientes para agrupar de maneira correta as réplicas das devidas localidades, estresse baixo de 0,0984. As variáveis salinidade e pH estabelecem uma variação semelhante ao longo dos pontos, com uma elevada correlação positiva, enquanto que uma correlação negativa ocorre entre a temperatura e o oxigênio dissolvido (Figura 17).

As frações 3 e 0,22 $\mu$ m apresentam as mesmas variáveis modulando a abundância de sequências. A temperatura, salinidade, OD e os tipos de substratos foram as variáveis significativas. As correlações entre as matrizes de sequências e a matriz das variáveis significativas foram positivas nas duas frações. A temperatura foi a variável de maior significância nas duas RDAs (Figuras 39 e 40). A algum tempo já se sabe que a temperatura superficial da água do mar é o fator limitante da distribuição geográfica de cianobactérias nos ecossistemas marinhos tropicais e equatoriais (HOFFMAN, 1999). Meirelles et al. (2015) avaliou que o gradiente vertical de temperatura foi melhor relacionado ao gradiente de abundância de sequências costa-oceano do que apenas o gradiente horizontal da lamina da água. O gradiente longitudinal tropical Vitória-Trindade apresentou uniformidade de riqueza e de famílias do bacterioplâncton marinho (MEIRELLES et al., 2015). Para a porção sudoeste do Atlântico Sul a abundância e a diversidade da comunidade picofitoplancctônica parece ser fortemente influenciada pela temperatura, carbono orgânico dissolvido, e profundidade, sendo identificados três grupos distintos: águas superficiais; clorofila sub-superficial máxima (48-82m); e águas profundas (236-1200m) de acordo com a composição microbiana (JUNIOR, et al., 2015). O projeto Tara Ocean também ressalva a temperatura como um dos mais importantes fatores para a composição taxonômica das comunidades microbianas aquáticas, sendo a abundância de sequências de cianobactérias muito maior em superfície do que nas outras profundidades, como a zona fótica (SUNAGAWA et al., 2015). E ainda, observou que a temperatura e a intensidade luminosa regulam os diferentes ecótipos de *Prochlorococcus* e *Synechococcus* e, portanto, a produtividade do planeta (FARRANT et al., 2016).

A evolução genética associada a imensa e desconhecida biodiversidade de cianobactérias, promove uma readequação contínua da taxonomia, pelos contínuo descobrimento de novas sequências. Averiguar a diversidade com novos olhares, integrando os parâmetros ambientais e os ecótipos é cada vez mais fundamental (WALTER et al., 2017). Analisando 191 metagenomas marinhos de cianobactérias dentro dos bancos de dados de três grandes grupos ecológicos de cianobactérias variando pela temperatura. Vê-se que a

temperatura é central para regular a composição e funcionamento das comunidades de cianobactérias, portanto o aquecimento global pode afetar as taxas de crescimento e potencial de florescimento e abundâncias de muitos táxons dentro das pico-cianobactérias (WALTER et al., 2017). Alguns modelos baseados em nichos avaliam a abundância de células com base em temperatura e radiação fotossinteticamente ativa, predizem alterações significativas nas abundâncias de *Prochlorococcus* e *Synechococcus* devido ao aquecimento global (FLOMBAUM et al., 2013). Atualmente as abundâncias globais médias anuais de *Prochlorococcus* e *Synechococcus* estão especuladas em  $2,9 \times 10^{27}$  e  $7,0 \pm \times 10^{26}$  células, respectivamente. Usando projeções para o aumento da temperatura da superfície do mar como resultado da concentração dos gases do efeito estufa no final do século 21, foram estimados um aumento celular de 29% e 14% para *Prochlorococcus* e *Synechococcus*, respectivamente. Por causa das elevadas abundâncias e contribuições para a produção primária serão alteradas ocorrendo grandes impactos nos ecossistemas oceânicos e ciclos biogeoquímicos (FLOMBAUM et al., 2013).

A análise de beta-diversidade não encontrou diferença das variáveis categóricas de dinâmica e tipos de substrato entre si, e em nenhuma das duas frações (Figura 41). Correlacionando as variáveis categóricas, a menor diversidade local ocorreu nas planícies arenosas em condições expostas na fração nanofitoplanctônica. Enquanto que na fração picofitoplancntônica a maior diversidade local ocorreu em planícies arenosas expostas. Isso pode ser explicado devido a própria biologia dos gêneros que dominam a abundancia dentro do picofitoplâncton, *Prochlorococcus* e *Synechococcus*. As planícies arenosas amostradas neste trabalho encontram-se em áreas com águas superficiais abertas e quentes com alta luminosidade, condições ideais para a proliferação de *Prochlorococcus* e *Synechococcus* em águas oligotróficas (THOMPSON et al., 2013a; BATMALLE et al., 2014; LINDELL, 2014; FARRANT et al., 2016).

### **6.5. Comparação metodológica**

Para as amostras costeiras, a riqueza observada pelas sequencias 16S rRNA foi aproximadamente duas vezes superior a riqueza determinada pela DGGE, comparando as respectivas frações celulares. Mesmo com essa diferença de grandeza, o mesmo padrão de riqueza foi encontrado em ambas as técnicas moleculares, sendo a riqueza da fração nano superior a riqueza da fração picofitoplancntônica. Quanto a  $\alpha$ - e  $\beta$ -diversidade, não houveram similaridade de padrões entre as respectivas frações entre as técnicas comparadas. Portanto, as técnicas não conseguiram estabelecer um mesmo perfil de estrutura da comunidade (Figura 29). A técnica de DGGE, mesmo com uma riqueza muito inferior a observada pelas sequencias determinou uma cobertura adequada, atingindo um platô na curva de acumulação, nas duas frações. Diferente do sequenciamento que, mesmo obtendo uma cobertura real mais elevada, ainda especulou uma maior diversidade estimada, com uma maior rarefação (Figura 28).

Durante muito tempo a técnica de DGGE foi considerada uma técnica válida para indicar a composição ecológica de cianobactérias (NÜBEL et al., 1997; BOUTTE et al., 2006; MÜHLING et al., 2008). Mas considerando as recentes descobertas da complexidade filogenética do clado de cianobactérias e o avanço das técnicas moleculares, esta técnica não é mais representativa ao perfil de comunidade de cianobactérias. E neste estudo comprovamos que além de não revelar o real perfil das amostras costeiras, mascarou a diversidade local de cianobactérias. Uma comparação metodológica entre DGGE e T-RFLP para populações de cianobactérias marinhas com regiões amplificadas entre 16S-23S rRNA, concluiu que é possível identificar linhagens de *Prochlorococcus* e *Synechococcus*, porém o T-RFLP apresentou respostas mais claras sendo melhor indicado para estudar a diversidade de populações naturais de cianobactérias. Tais dados foram confirmados por estimativas em citometria de fluxo. O T-RFLP apresentou semelhança de *Prochlorococcus* e *Synechococcus* pelas duas técnicas, 34,8 e 65,2% por T-RFLP e 30,6 e 69,4% por citometria de fluxo, respectivamente (LAVIN et al., 2008).

## 7. Conclusão

As correntes SEC e NBC atuam diretamente na movimentação das massas de água superficiais nos arquipélagos oceânicos brasileiros, assim como na costa NE do Brasil. Tal área é dominada pela circulação oeste do giro Equatorial Atlântico. O estoque de carbono em cianobactérias acompanhou um gradiente longitudinal costa → arquipélagos oceânicos: PP>PA=IC>AC>FN>AS. O estoque de carbono na costa corresponde a 1,04mg C m<sup>-3</sup>, com o estoque relativo de 2,69% *Synechococcus* e 0,48% do POC em *Prochlorococcus*. Nos arquipélagos oceânicos apenas 0,15mg C m<sup>-3</sup> estão estocados nas cianobactérias, com o estoque relativo de 0,38% em *Synechococcus* e 0,10% do POC em *Prochlorococcus* em águas superficiais. O AS apresentou maiores oscilações de biomassa pois é diretamente influenciado pela sazonalidade da SEC. No total, 1.158.901 sequências foram representadas em 56 táxons, com 48 pertencem às cianobactérias. Deste total de táxons, 37 foram comuns entre as frações nano e picofitoplanctônicas. Isolando estas frações pelo gradiente longitudinal costa e ilhas, 4 cenários distintos foram compostos, nos quais apenas 20 táxons comuns entre si. O cenário de maior exclusividade taxonômica foi representado nos arquipélagos na fração nanofitoplanctônica, com 9 táxons exclusivos. Os gêneros *Prochlorococcus* e *Synechococcus* foram os dominantes, acumulando 98% das sequências totais em ambas frações fitoplanctônicas. As réplicas costeiras se correlacionam melhor à *Synechococcus*, enquanto que as réplicas insulares foram dominadas por *Prochlorococcus*. Não foi possível estabelecer um gradiente costa → ilhas pela diversidade de cianobactérias através da abundância de sequências 16S rRNA, devido as disparidades de abundância e riqueza entre os arquipélagos. As localidades de FN, apresentaram os maiores índices de diversidade e muitos táxons raros, mesmo com abundância média de sequências, já no AS os padrões se invertem totalmente. As

forçantes ambientais que mais influenciaram a distribuição de cianobactérias no gradiente superficial costa → ilhas foi a temperatura. As variáveis tipo de substrato e hidrodinâmica local revelaram maior diversidade picofitoplancntônica nas planícies arenosas em condições expostas. Para a fração nanofitoplanctônica estas condições representaram a menor biodiversidade.

## 8. Considerações finais

. Ampliar a malha de amostragem em POC remota para a zona costeira seria uma boa maneira deixar estas amostras comparáveis quanto médias temporais, assim como estabelecido pelas amostras insulares. Mesmo que exista uma margem de erro em amostras costeiras, pode-se utilizar o sinal remoto par POC em áreas costeiras para identificar padrões (MISHONOV et al., 2003).

. O uso do fluoro cromo SYBER-Green como marcador de DNA poderia ter revelado outros grupos planctônicos, ao apenas ter isolado as frações *Prochlorococcus e Synechococcus* pela fluorescência natural.

. Um sequenciamento HiSeq teria possibilitado o nível de identificação de espécie, pois pela cobertura atual apenas o nível de gênero foi atingindo em apenas alguns OTUs.

. Realizar o upload das novas sequências 16S rRNA o mais breve possível no banco de dados NCBI.

. Investigar da maneira mais profunda possível a relação das sequencias alinhadas as duas frações fitoplanctônicas as diferentes vias metabólicas ligadas as vias de fixação do nitrogênio, podendo revelar novos gradientes costa→arquipélagos oceânicos.

## 9. Referências

- ACINAS, S.G.; MARCELINO, L.A.; KLEPAC-CERAJ, V.; POLZ, M.F. 2004. Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *Journal of Bacteriology*, v. 186, p. 2629-2635.
- AFFE, H.M.J.; RIGONATO, J.; NUNES, J.M.C.; MENEZES, M. 2018. Metagenomic analysis of cyanobacteria in an oligotrophic tropical estuary, South Atlantic. *Frontiers in Microbiology*, v. 9, p.1-13.
- AHLGREN, N.A.; ROCAP, G.; CHISHOLM, S.W. 2006. Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environmental Microbiology*, v. 8, p. 441-54.
- AMANN, R.I.; LUDWIG, W.; SCHLEIFER, K.H. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, v. 59, p. 143-169.
- ANDERSEN, R.A., 2005. *Algal culturing techniques*. London: Elsevier Academic Press, 589 p.
- ANGULO, R.J.; SOUZA, M.C.; FERNANDES, L.A.; DISARÓ, S.T. 2013. Quaternary sea-level changes and aeolianites in the Fernando de Noronha archipelago, northeastern Brazil. *Quaternary International*, v. 305, p. 15-30.
- APHA; AWWA; WEF. 2012. Standard methods for the examination of water e wastewater. In: RICE, E.W.; BAIRD, R.B.; EATON, A.D.; CLESCERI, L.S. (Eds.) 5310 Total organic carbon (TOC). 22. ed. Washington: APHA, p. 5-18-5-25.
- AZAM, F.; FENCHEL, T.; FIELD, J.G.; GRAY, J.S.; MEYER-REIL, L.A.; THINGSTAD, F. 1982. The ecological role of water-column microbes in the sea. *Marine Ecology Progress*, v. 10, p. 257-263.
- AZEVEDO, S.M.F.O.; VASCONCELOS, V. 2006. Toxinas de cianobactérias: Causas e consequências para a saúde pública. In: ZAGATTO, P.A.; BERTOLETTI, E. eds. *Ecotoxicologia aquática: Princípios e aplicações*. São Carlos: Rima, pp. 433-452.
- BAKER, C.A.; HENSON, S. A.; CAVAN, E. L.; GIERING, S. L. C.; YOOL, A.; GEHLEN, M.; BELCHER, A.; RILEY, J.S.; SMITH, H.E.K.; SANDERS, R. 2017. Slow-sinking particulate organic carbon in the Atlantic Ocean: magnitude, flux, and potential controls, *Global Biogeochemical Cycles*, v. 31, p. 1051-1065.
- BALVOČIŪTĒ, M. HUSON, D.H. 2017. SILVA, RDP, Greengenes, NCBI and OTT: how do these taxonomies compare? *BMC Genomics*, v. 18, p. 1-8.

- BALZANO, S.; MARIE, D.; GOURVIL, P.; VAULOT, D. 2012. Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. *The ISME Journal*, v. 6, p. 1480-1498.
- BARSANTI, L.; GUALTIERI, P. 2006. *Algae: anatomy, biochemistry, and biotechnology*. Boca Raton: Taylor & Francis Group. 320 p.
- BARTHEL, K.G.; SCHNEIDER, G.; GRADINGER, R.; LENZ, J. 1989. Concentration of live pico- and nanoplankton by means of tangential flow filtration. *Journal of Plankton Research*, v. 11, p. 1213-1221.
- BARTON, L.L.; NORTHUP, D.E. 2011. *Microbial ecology*. Hoboken: Wiley-Blackwell. 407 p.
- BATMALLE, C.S.; CHIANG, H.I.; ZHANG, K.; LOMAS, M.W.; MARTINY, A.C. 2014. Development and bias assessment of a method for targeted metagenomic sequencing of marine cyanobacteria. *Applied and Environmental Microbiology*, v. 80, p. 1116-1125.
- BECK, C.; KNOOP, H.; AXMANN, I.M.; STEUER, R. 2012. The diversity of cyanobacterial metabolism: genome analysis of multiple phototrophic microorganisms. *BMC Genomics*, v. 13, p. 1-17.
- BERNARDI, A.; NIKOLAOU, A.; MENEGHESSO, A.; MOROSINOTTO, T.; CHACHUAT, B.; BEZZO, F. 2016. High-fidelity modelling methodology of light-limited photosynthetic production in microalgae. *Plos One*, v. 11, p.1-20.
- BEZERRA, W.M. 2015. *Microbioma de sedimentos de manguezais brasileiros e seu potencial biotecnológico*. Tese (Doutorado) - Centro de Ciências, Universidade Federal do Ceará, 153 p.
- BJØRNSSEN, P.K. 1986. Automatic determination of bacterioplankton biomass by image analysis. *Applied and Environmental Microbiology*, v. 51, p. 1199-1204.
- BJØRNSSEN, P.K.; KUPARINEN, J. 1991. Growth and herbivory by heterotrophic dinoflagellates in the Southern Ocean studied by microcosm experiments. *Marine Biology*, v. 109, p. 397-405.
- BOKULICH, N. A.; SUBRAMANIAN, S.; FAITH, J. J.; GEVERS, D.; GORDON, J. I.; KNIGHT, R.; MILLS, D. A.; CAPORASO, J. G. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature methods*, v. 10, p. 57-59.

- BOUMAN, H.A.; ULLOA, O.; SCANLAN, D.J.; ZWIRGLMAIER, K.; LI, W.K.W. PLATT, T.; STUART, V.; BARLOW, R.; LETH, O.; CLEMENTSON, L.; VIVIAN LUTZ, V.; FUKASAWA, M.; WATANABE, S.; SATHYENDRANATH, S. 2006. Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science*, v. 312, 918-921.
- BOURLÈS, B.; GOURIOU, Y.; CHUCHLA, R. 1999. On the circulation in the upper layer of the western equatorial Atlantic. *Journal of Geophysical Research*, v. 104, p. 21,151-21,170.
- BOUTTE, C.; GRUBISIC, S.; BALTHASART, P.; WILMOTTE, A. 2006. Testing of primers for the study of cyanobacterial molecular diversity by DGGE. *Journal of Microbiological Methods*, v. 65, p. 542-550.
- BOYER, S.L.; FLECHTNER, V.R.; JOHANSEN, J.R. 2001. Is the 16S–23S rRNA internal transcribed spacer region a good tool for use in molecular systematics and population genetics? A case study in cyanobacteria. *Molecular Biology and Evolution*, v. 18, p. 1057-1069.
- BRACHER, A.; BOUMAN, H.A.; BREWIN, R.J.W.; BRICAUD, A.; BROTHAS, V.; CIOTTI, A.M.; CLEMENTSON, L.; DEVRED, E.; DI CICCO, A.; DUTKIEWICZ, S.; HARDMAN-MOUNTFORD, N.J.; HICKMAN, A. E.; HIERONYMI, M.; HIRATA, T.; LOSA, S.N.; MOUW, C.B.; ORGANELLI, E.; RAITOSOS, D.E.; UITZ, J.; VOGT, M.; WOLANIN, A. 2017. Obtaining phytoplankton diversity from ocean color: a scientific roadmap for future development. *Frontiers in Marine Science*, v. 4, p. 1-55.
- BUB, F.L.; BROWN, W.S. 1996: Intermediate layer water masses in the Western Tropical Atlantic Ocean. *Journal of Geophysical Research*, v. 101, p. 11,903-11,922.
- BUTTIGIEG, P.L.; RAMETTE, A. 2014. A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. *FEMS Microbiology Ecology*, v. 90, p. 543-550.
- CAMPBELL, L.; NOLLA, H.A.; VAULOT, D. 1994. The importance of *Prochlorococcus* to community structure in the central North Pacific Ocean. *Limnology and Oceanography*, v. 39, p. 954-961.
- CAPORASO, J.G.; KUCZYNSKI, J.; STOMBAUGH, J.; BITTINGER, K.; BUSHMAN, F.D.; COSTELLO, E.K.; FIERER, N.; PEÑA, A.G.; GOODRICH, J.K.; GORDON, J.I.; HUTTLEY, G.A.; KELLEY, S.T.; KNIGHTS, D.; KOENIG, J.E.; LEY, R.E.; LOZUPONE, C.A.; MCDONALD, D.; MUEGGE, B.D.;

- PIRRUNG, M.; REEDER, J.; SEVINSKY, J.R.; TURNBAUGH, P.J.; WALTERS, W.A.; WIDMANN, J.; YATSUNENKO, T.; ZANEVELD, J.; KNIGHT, R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, v. 7, p. 335-336.
- CARR, M.E.; FRIEDRICH, M.A.M.; SCHMELTZ, M.; AITA, M.N.; ANTOINE, D.; ARRIGO, K.R.; ASANUMA, I.; AUMONT, O.; BARBER, R.; BEHRENFELD, M.; BIDIGARE, R.; BUITENHUIS, E.T.; CAMPBELL, J.; CIOTTI, A.; DIERSSEN, H.; DOWELL, M.; DUNNE, J.; ESAIAS, W.; GENTILI, B.; GREGG, W.; GROOM, S.; HOEPFFNER, N.; ISHIZAKA, J.; KAMEDA, T.; LE QUÉRÉ, C.; LOHRENZ, S.; MARRA, J.; MÉIN, F.; MOORE, K.; MOREL, A.; REDDY, T.E.; RYAN, J.; SCARDI, M.; SMYTH, T.; TURPIE, K.; TILSTONE, G.; WATERS, K.; YAMANAKA, Y. 2006. A comparison of global estimates of marine primary production from ocean color. *Deep-Sea Research Part II*, v. 53, p. 741-770.
- CARVALHO, M.C.C.G.; SILVA, D.C.G. 2010. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. *Ciência Rural*, v. 40, p. 735-744.
- CASTENHOLZ, R.W. 2001. Phylum BX. cyanobacteria: oxygenic photosynthetic bacteria. In: BOONE, D.R.; CASTENHOLZ, R.W.; GARRITY, G.M. eds. *Bergey's manual of systematic bacteriology*. 2 ed. New York: Springer, p. 473-599.
- CAVALIER-SMITH, T. 2002. The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial megaclassification. *International Journal of Systematic and Evolutionary Microbiology*, v. 52, p. 7-76.
- CETINIĆ, I.; PERRY, M.J.; BRIGGS, N.T.; KALLIN, E.; D'ASARO, E.A.; LEE, C.M. 2012. Particulate organic carbon and inherent optical properties during 2008 North Atlantic Bloom Experiment. *Journal of Geophysical Research*, v. 117, p. 1-18.
- CHISHOLM, S.W.; OLSON, R.J.; ZETTLER, E.R.; GOERICKE, R.; WATERBURY, J.B.; 1988. A novel free living prochlorophyte abundant in the oceanic euphotic zone. *Nature*, v. 334, p. 340-343.
- CHRISTIAN, J.R.; KARL, D.M. 1994. Microbial community structure at the U.S.-Joint Global Ocean Flux Study Station ALOHA: inverse methods for estimating biochemical indicator ratios. *Journal of Geophysical Research*, v. 99, p. 14269-14276.
- CINTRA, M.; LENTINI, C.A.D.; SERVAIN, J.; ARAÚJO, M.; MARONE, E. 2015. Physical processes that drive the seasonal evolution of the southwestern tropical atlantic warm pool. *Dynamics of Atmospheres and Oceans*, v. 72, p. 1-11.

- CIOTTI, A.M.; LEWIS, M.R.; CULLEN, J.J. 2002. Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient. *Limnology and Oceanography*, v. 47, p. 404-417.
- COELHO-SOUZA, S.A.; PEREIRA, G.C.; LOPEZ, M.S.; GUIMARAES, J.R.D.; COUTINHO, R. 2017. Seasonal sources of carbon to the Brazilian upwelling system. *Estuarine, Coastal and Shelf Science*, v. 194, p. 162-171.
- COHEN, Y.; GUREVITZ, M. 2006. The cyanobacteria: ecology, physiology and molecular genetics. In: DWORKIN, M.; FALKOW, S.; ROSENBERG, E.; SCHLEIFER, K.H.; STACKEBRANDT, E., eds. *The prokaryotes: bacteria: firmicutes, cyanobacteria*. New York: Springer-Verlag, p. 1074-1098.
- COLARES, G.B. 2010. Diversidade e estrutura de comunidades microbianas associadas à rizosfera de *Rhizophora mangle* do manguezal do rio Pacoti, zona leste da costa cearense. Dissertação (Mestrado) - Instituto De Ciências Do Mar, Universidade Federal Do Ceará, 59 p.
- COLARES, G.B. 2014. Ecologia microbiana em sedimentos de manguezais do estado do Ceará. Tese (Doutorado) - Instituto de Ciências do Mar, Universidade Federal do Ceará, 162 p.
- COLARES, G.B.; MELO, V.M.M. 2013. Relating microbial community structure and environmental variables in mangrove sediments inside *Rhizophora mangle* L. habitats. *Applied Soil Ecology*, v. 64, p. 171-177.
- COLES, V.J.; BROOKS, M.T.; HOPKINS, J.; STUKEL, M.R.; YAGER, P.L.; HOOD, R.R. 2013. The pathways and properties of the Amazon River Plume in the tropical North Atlantic Ocean. *Journal of Geophysical Research*, v. 118, p. 6894-6913.
- CORDEIRO, T.A.; BRANDINI, F.P.; ROSA, R.S.; SASSI, R. 2013. Deep chlorophyll maximum in Western Equatorial Atlantic: how does it interact with islands slopes and seamounts? *Marine Science*, v. 3, p. 30-37.
- DELILLE, D.; FIALA, M.; KUPARINEN, J.; KUOSA, H.; PLESSIS, C. 2002. Seasonal changes in microbial biomass in the first-year ice of the Terre Adélie area (Antarctica). *Aquatic Microbial Ecology*, v. 28, p. 257-265.
- DELORENZO, M.E.; FLEMING, J. 2008. Individual and mixture effects of selected pharmaceuticals and personal care products on the marine phytoplankton species *Dunaliella tertiolecta*. *Archives of Environmental Contamination and Toxicology*, v. 54, p. 203-210.

- DESANTIS, T.Z.; HUGENHOLTZ, P.; LARSEN, N.; ROJAS, M.; BRODIE, E.L.; KELLER, K.; HUBER, T.; DALEVI, D.; HU, P.; ANDERSEN, G.L. 2006. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, v. 72, p. 5069-5072.
- DETMER, A.E.; BATHMANN, U.V. 1997. Distribution patterns of autotrophic picoplankton and nanoplankton and their relative contribution to algal biomass during spring in the Atlantic sector of the Southern Ocean. *Deep-Sea Research Part II*, v. 44, p. 299-320.
- DIAS, F.J.S.; CASTRO, B.M.; LACERDA, L.D. 2013. Continental shelf water masses off the Jaguaribe River (4S), northeastern Brazil. *Continental Shelf Research*, v. 66, p. 123-135.
- DUBELAAR, G.B.J.; JONKER, R.R. 2000. Flow cytometry as a tool for the study of phytoplankton. *Scientia Marina*, v. 64, p. 135-156.
- DUBELAAR, G.B.J.; JONKER, R.R., 2000. Flow cytometry as a tool for the study of phytoplankton. *Scientia Marina*, v. 64, p. 135-156.
- DUCAT, D.C.; WAY, J.C.; SILVER, P.A. 2011. Engineering cyanobacteria to generate high-value products. *Trends Biotechnology*, v. 29, p. 95-103.
- DURAND, M.M.D.; OLSON, R.R.J.; CHISHOLM, S.S.W.S. 2001. Phytoplankton population dynamics at the Bermuda Atlantic time-series station in the Sargasso Sea. *Deep-Sea Research Part II*, v. 48, p. 1983-2003.
- EDGAR, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, v. 26, p. 2460-2461.
- EDGAR, R. C.; HAAS, B. J.; CLEMENTE, J. C.; QUINCE, C.; KNIGHT, R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, v. 27, p. 2194-2200.
- EDGAR, R.C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, v. 10, p. 996-1000.
- EIGEMANNA, F.; HILT, S.; SCHMITT-JANSEN, M. 2013. Flow cytometry as a diagnostic tool for the effects of polyphenolic allelochemicals on phytoplankton. *Aquatic Botany*, v. 104, p. 5-14.
- EKAU, W.; KNOPPERS, B. 1999. An introduction to the pelagic system of the Northeast and East Brazilian shelf. *Archive of Fishery and Marine Research*, v. 47, p. 5-24.

- EVERS-KING, H.; MARTINEZ-VICENTE, V.; BREWIN, R.J.W.; OLMO, G.D.; HICKMAN, A.E.; JACKSON, T.; KOSTADINOV, T.S.; KRASEMANN, H.; LOISEL, H.; RÖTTGERS, R.; ROY, S.; STRAMSKI, S.; THOMALLA, S.; PLATT, T.; SATHYENDRANATH, S. 2017. Validation and Intercomparison of Ocean Color Algorithms for Estimating Particulate Organic Carbon in the Oceans. *Frontiers in Marine Science*, v. 4, p. 1-20.
- FALKOWSKI, P.G.; OWENS, T.G.; LEY, A.C.; MAUZERALL, D.C. 1981. Effects of growth irradiance levels on the ratio of reaction centers in two species of marine phytoplankton. *Plant Physiology*, v. 68, p. 969-973.
- FARRANT, G. K., DORÉ, H., CORNEJO-CASTILLO, F. M., PARTENSKY, F., RATIN, M., OSTROWSKI, M., PITT, F.D.; WINCKER, P.; SCANLAN, D.J.; IUDICONE, D.; ACINAS, S.G.; GARCZAREK, L. 2016. Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proceedings of the National Academy of Sciences*, v. 113, p. E3365-E3374.
- FIGUEIRAS, F.G.; ESPINOZA-GONZÁLEZ, O.; ARBONES, B.; GARRIDO, J.L.; TEIXEIRA, I.G.; CASTRO C.G. 2014. Estimating phytoplankton size-fractionated primary production in the northwestern Iberian upwelling: Is mixotrophy relevant in pigmented nanoplankton? *Progress in Oceanography*, v. 128, p. 88-97.
- FRATANTONI, D.M.; GLICKSON, D.A. 2002. North Brazil Current ring generation and evolution observed with SeaWiFS. *Journal of Physical Oceanography*, v. 32, p. 1058-1074.
- FRATANTONI, D.M.; RICHARDSON, P.L. 2006. The evolution and demise of North Brazil Current rings. *Journal Of Physical Oceanography*, v. 36, p. 1241-1264.
- GARCIA, G.D.; GREGORACCI, G.B.; SANTOS, E.O.; MEIRELLES, P.M.; SILVA, G.G.Z.; EDWARDS, R.; SAWABE, T.; GOTOH, K.; NAKAMURA, S.; IIDA, T.; & MOURA, R.L.; THOMPSON, F.L. 2013. Metagenomic analysis of healthy and white plague-affected *Mussismilia braziliensis* corals. *Microbial Ecology*, v. 65, p. 1076-1086.
- GARRISON, D.L.; GOWING, M.M.; HUGHES, M.P.; CAMPBELL, L.; CARON, D.A.; DENNETT, M.R.; SHALAPYONOK, A.; OLSON, R.J.; LANDRY, M.R.; BROWN, S.L.; LIU, H.B.; AZAM, F.; STEWARD, G.F.; DUCKLOW, H.W.; SMITH, D.C. 2000. Microbial food web structure in the Arabian Sea: a US JGOFS study. *Deep-Sea Research Part II*, v. 47, p. 1387-1422.

- GAULT, P.M.; MARLER, H.L. 2009. HANDBOOK on cyanobacteria: biochemistry, biotechnology and applications. New York: Nova Science Publishers, Inc. 538 p.
- GIANNINI, M.F.C.; CIOTTI, Á.M. 2016. Parameterization of natural phytoplankton photo-physiology: effects of cell size and nutrient concentration. *Limnology Oceanography*, v. 61, p. 1495-1512.
- GIOVANNONI, S.J.; DELONG, E.F.; SCHMIDT, T.M.; PACE, N.R. 1990. Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. *Applied and Environmental Microbiology*, v. 56, p. 2572-2575.
- GOERICKE, R.; REPETA, D.J. 1993. Chlorophyll-a and chlorophyll-b and divinyl chlorophyll-a and chlorophyll-b in the open subtropical North-Atlantic ocean. *Marine Ecology Progress Series*, v. 101, p. 307-313.
- GRAFF, J.R.; MILLIGAN, A.J.; BEHRENFELD, M.J. 2012. The measurement of phytoplankton biomass using flow-cytometric sorting and elemental analysis of carbon. *Limnology and Oceanography Methods*, v. 10, p. 910-920.
- GRAFF, J.R.; WESTBERRY, T.K.; MILLIGAN, A.J.; BROWN, M.B.; OLMO, G.D., VAN DONGEN-VOGELS, V., REIFEL, K.M.; BEHRENFELD, M.J. 2015. Analytical phytoplankton carbon measurements spanning diverse ecosystems. *Deep-Sea Research Part I*, v. 102, p. 16-25.
- GRASSHOFF, K.; EHRHARDT, R.; KREMLING, K. 1999. *Methods of seawater analysis*. Weinheim: Wiley-VCH. 600 p.
- HAAS, B.J.; GEVERS, D.; EARL, A.M.; FELDGARDEN, M.; WARD, D.V.; GIANNOUKOS, G.; CIULLA, D.; TABBAA, D.; HIGHLANDER, S.K.; SODERGREN, E.; METHÉ, B.; DESANTIS, T.Z.; HUMAN MICROBIOME CONSORTIUM; PETROSINO, J.F.; KNIGHT, R.; BIRREN, B.W. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, v. 21, p. 494-504.
- HALLENBECK, P.C. 2017. *Modern topics in the phototrophic prokaryotes: environmental and applied aspects*. Cham: Springer, 492 p.
- HARDISTY, A.R.; BACALL, F.; BEARD, N.; BALCÁZAR-VARGAS, M.P.; BALECH, B.; BARCZA, Z.; BOURLAT, S.J.; GIOVANNI, R.; JONG3, Y.; LEO, F.D.; DOBOR, L.; DONVITO, G.; FELLOWS, D.; GUERRA, A.F.; FERREIRA, N.; FETYUKOVA, Y.; FOSSO, B.; GIDDY, J.; GOBLE, C.; GÜNTSCH, A.; HAINES, R.; ERNST, V.H.; HETTLING, H.; HIDY, D.; HORVÁTH, F.; ITTZÉS, D.; ITTZÉS, P.; JONES, A.; KOTTMANN, R.; KULAWIK, R.; LEIDENBERGER,

- S.; LYYTIKÄINEN-SAARENMAA, P.; MATHEW, C.; MORRISON, N.; NENADIC, A.; HIDALGA, A.N.; OBST, M.; OOSTERMEIJER, G.; PAYMAL, E.; PESOLE, G.; PINTO, S.; POIGNÉ, A.; FERNANDEZ, F.Q.; SANTAMARIA, M.; SAARENMAA, H.; SIPOS, G.; SYLLA, K.H.; TÄHTINEN, M.; VICARIO, S.; VOS, R.A.; WILLIAMS, A.R.; YILMAZ, P. 2016. BioVeL: a virtual laboratory for data analysis and modelling in biodiversity science and ecology. *BMC Ecology*, v. 16, p. 1-16.
- HAZIN, F.H.V. 1993. Fisheries: oceanographical study of tunas, billfishes and sharks in the Southwestern Equatorial Atlantic Ocean. Unpublished Ph.D. Dissertation, Tokyo University of Fisheries, Tokyo, 286 p.
- HEINZE, C.; MEYER, S.; GORIS, N.; ANDERSON, L.; STEINFELDT, R.; CHANG, N.; LE QUÉRÉ, C.; BAKKER, D.C.E. 2015. The ocean carbon sink: impacts, vulnerabilities and challenges. *Earth System Dynamics*, v. 6, p. 327-358.
- HERDMAN, M.; JANVIER, M.; RIPPKA, R.; STANIER, R.Y. 1979. Genome Size of cyanobacteria. *Journal of General Microbiology*, v. 111, p. 73-85.
- HEYWOOD, J.L.; ZUBKOV, M.V.; TARRAN, G.A.; FUCHS, B.M.H.P. 2006. Prokaryoplankton standing stocks in oligotrophic gyre and equatorial provinces of the Atlantic Ocean: evaluation of inter-annual variability. *Deep-Sea Research Part II*, v. 53, p. 1530-1547.
- HOFFMANN, L.; 1999. Marine cyanobacteria in tropical regions: diversity and ecology. *European Journal of Phycology*, v. 34, p. 371-379.
- HOUGHTON, R.A. 2007. Balancing the global carbon budget. *Annual Review of Earth and Planetary Sciences*, v. 35, p. 313-347.
- HUG, L.A.; BAKER, B.J.; ANANTHARAMAN, K.; BROWN, C.T.; PROBST, A.J.; CASTELLE, C.J.; BUTTERFIELD, C.N.; HERNSDORF, A.W.; AMANO, Y.; ISE, K.; SUZUKI, Y.; DUDEK, N.; RELMAN, D.A.; FINSTAD, K.M.; AMUNDSON, R.; THOMAS, B.C.; BANFIELD, J.F. 2016. "A new view of the tree of life". *Nature Microbiology*, v. 1, p. 1-6.
- HUMILY, F.; FARRANT, G.K.; MARIE, D.; PARTENSKY, F.; MAZARD, S.; PERENNOU, M.; LABADIE, K.; AURY, J.M.; WINCKER, P.; SEGUI, A.N.; SCANLAN, D.J.; GARCZAREK, L. 2014. Development of a targeted metagenomic approach to study a genomic region involved in light harvesting in marine *Synechococcus*. *FEMS Microbiology Ecology*, v. 88, p. 231-249.

- HUOT, Y.; BABIN, M. 2010. Overview of fluorescence protocols: theory, basic concepts, and practice. In: SUGGETT, D.J.; PRÁŠIL, O.; BOROWITZKA, M.A. eds. Chlorophyll-a fluorescence in aquatic sciences: Methods and applications. London: Springer, pp. 31-74.
- IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA: Censo 2010: Cidades. Disponível em: <<https://cidades.ibge.gov.br>>. Acesso em: março de 2018.
- ILLUMINA. 2013. 16S Metagenomic sequencing library preparation: preparing 16S ribosomal RNA gene amplicons for the Illumina MiSeq system. 16S Metagenomic Sequencing Library Preparation (15044223 B), p. 1-28.
- ITO, H.; TANAKA, A. 2011. Evolution of a divinyl chlorophyll-based photosystem in *Prochlorococcus*. *Proceedings of the National Academy of Sciences*, v. 108, p. 18014-18019.
- ITURRIAGA, R.; MITCHELL, B.G. 1986. Chroococcoid cyanobacteria: a significant component in the food web dynamics of the open ocean. *Marine Ecology Progress Series*, v. 28, p. 291-297.
- JANSE, I.; MEIMA, M.; KARDINAAL, W.E.A.; ZWART, G. 2003. High-resolution differentiation of cyanobacteria by using rRNA-internal transcribed spacer denaturing gradient gel electrophoresis. *Applied and Environmental Microbiology*, v. 69, p. 6634-6643.
- JIANG, Q.; QIN, S.; WU, Q.Y. 2010. Genome-wide comparative analysis of metacaspases in unicellular and filamentous cyanobacteria. *BMC Genomics*, v. 11, p. 1-11.
- JIMENEZ, J.; MAIA, L.P.; RAVENTOS, J.S.; MORAIS, J.O. 1999. Dune migration along the Ceará coast, Northeastern Brazil. *Sedimentology*, v. 46, p. 689-701.
- JOHNSON, Z.I.; ZINSER, E.R.; COE, A.; MCNULTY, N.P.; WOODWARD, E.M.S.; CHISHOLM, S.W. 2006. Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science*, v. 311, p. 1737-1740.
- JUNIOR, N.A.; MEIRELLES, P.M.; SANTOS, E.O.; DUTILH, B.; SILVA, G.G.Z.; PARANHOS, R.; CABRAL, A.S.; REZENDE, C.; IIDA, T.; MOURA, R.L.; KRUGER, R.H.; PEREIRA, R.C.; VALLE, R.; SAWABE, T.; THOMPSON, C.; THOMPSON, F. 2015. Microbial community diversity and physical-chemical features of the Southwestern Atlantic Ocean. *Archives of Microbiology*, v. 197, p. 165-179.

- JYOTHIBABU, R.; MOHAN, A.P.; JAGADEESAN, L.; ANJUSHA, A.; MURALEEDHARAN, K.R.; LALLU, K.R.; KIRAN, K.; ULLAS N. 2013. Ecology and trophic preference of picoplankton and nanoplankton in the Gulf of Mannar and the Palk Bay, southeast coast of India. *Journal of Marine Systems*, v. 111-112, p. 29-44.
- KAYANO, M.T.; OLIVEIRA, C.P.; ANDREOLI, R.V. 2009. Interannual relations between South American rainfall and tropical sea surface temperature anomalies before and after 1976. *International Journal of Climatology*, v. 29, p. 1439-1448.
- KIM, M.; LEE, K.H.; YOON, S.W.; KIM, B.S.; CHUN, J.; YI, H. 2013. Analytical Tools and Databases for Metagenomics in the Next-Generation Sequencing Era. *Genomics & Informatics*, v. 11, p. 102-113.
- KIRCHMAN, D.L., 2011. Microbial primary production and phototrophy. In: KIRCHMAN D.L., ed. *Processes in microbial ecology*. Oxford: Oxford University Press, pp. 55-78.
- KOMÁREK, J. 2016a. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. *European Journal of Phycology*, v. 51, p. 1-8.
- KOMÁREK, J. 2016b. Review of the cyanobacterial genera implying planktic species after recent taxonomic revisions according to polyphasic methods: state as of 2014. *Hydrobiologia*, v. 764, p. 259-270.
- KOMÁREK, J.; KAŠTOVSKÝ, J.; MAREŠ, J.; JOHANSEN, J.R. 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia*, v. 86, p. 295-335.
- KONOPKA, A. 2009. What is microbial community ecology? *International Society for Microbial Ecology Journal*, v. 3, p. 1223-1230.
- KOPRIVNJAK, J.F.; BLANCHETTE, J.G.; BOURBONNIERE, R.A.; CLAIR, T.A.; HEYES, A.; LUM, K.R.; MCCREA, R.; MOORE, T. R. 1995. The underestimation of concentrations of dissolved organic carbon in freshwaters. *Water Research*, v. 29, p. 91-94.
- KOSTADINOV, T.S.; MILUTINOVIĆ, S.; MARINOV, I.; CABRÉ, A. 2016. Carbon based phytoplankton size classes retrieved via ocean color estimates of the particle size distribution. *Ocean Science*, v. 12, p. 561-575.
- KOZLOV, A. M., ZHANG, J., YILMAZ, P., GLÖCKNER, F. O., STAMATAKIS, A. 2016. Phylogeny-aware identification and correction of taxonomically mislabeled sequences. *Nucleic Acids Research*, v. 44, p. 5022-5033.

- KRISHNASWAMI, S.; LAL, D. 1977. Particulate organic carbon in Atlantic surface waters. *Nature*, v. 266, p. 713-716.
- KROER, N. 1994. Relationships between biovolume and carbon and nitrogen content of bacterioplankton. *FEMS Microbiology Ecology*, v. 13, p. 217-224.
- LALLI, C.; PARSONS, T.R. 1997. *Biological oceanography: An introduction*. 2. ed. Oxford: Butterworth-Heinemann. 320 p.
- LALOUI, W.; PALINSKA, K.A.; RIPPKA, R.; PARTENSKY, F.; MARSAC, N.T.; HERDMAN, M.; ITEMAN, I. 2002. Genotyping of axenic and non-axenic isolates of the genus *Prochlorococcus* and the OMF-‘*Synechococcus*’ clade by size, sequence analysis or RFLP of the Internal Transcribed Spacer of the ribosomal operon. *Microbiology*, v. 148, p. 453-465.
- LAVIN, P.; GÓMEZ, P.; GONZÁLEZ, B.; ULLOA, O. 2008. Diversity of the marine picocyanobacteria *Prochlorococcus* and *Synechococcus* assessed by terminal restriction fragment length polymorphisms of 16S-23S rRNA internal transcribed spacer sequences. *Revista Chilena de Historia Natural*, v. 81, p. 515-531.
- LEE, S.; FUHRMAN, J.D. 1987. Relationships between biovolume and biomass of naturally derived marine bacterioplankton. *Applied and Environmental Microbiology*, v. 53, p. 1298-1303.
- LESSA, D.V.O.; SANTOS, T.P.; VENANCIO, I.M.; ALBUQUERQUE, A.L.S. 2017. Offshore expansion of the Brazilian coastal upwelling zones during Marine Isotope Stage 5. *Global and Planetary Change*, v.158, p. 13-20.
- LI, W.K.W.; DICKIE, P.M.; IRWIN, B.D.; WOOD, A.M. 1992. Biomass of bacteria, cyanobacteria, prochlorophytes and photosynthetic eukaryotes in the Sargasso Sea. *Deep-Sea Research Part I*, v. 39, p. 501-519.
- LI, W.K.W.; SUBBA RAO, D.V.; HARRISON, W.G.; SMITH, J.C.; CULLEN, J.J.; IRWIN, B.; PLATT, T. 1983. Autotrophic picoplankton in the tropical ocean. *Science*, v. 219, p. 292-295.
- LINACRE, L.; LARA-LARA, R.; CAMACHO-IBAR, V.; HERGUERA, J.C.; BAZÁN-GUZMÁN, C.; VICENTE FERREIRA-BARTRINA, V. 2015. Distribution pattern of picoplankton carbon biomass linked to mesoscale dynamics in the southern gulf of Mexico during winter conditions. *Deep-Sea Research Part I*, v. 106, p. 55-67.
- LINDELL, D. 2014. The genus *Prochlorococcus*: phylum Cyanobacteria. In: ROSENBERG E.; DELONG E.F.; LORY S.; STACKEBRANDT E.; THOMPSON F. eds. *The prokaryotes*. Berlin: Springer-Verlag, p. 829-845.

- LOCEY, K.J.; LENNON, J.T. 2016. Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, v. 113, p. 5970-5975.
- LOGARES, R.; LINDSTRÖM, E.S.; LANGENHEDER, S.; LOGUE, J.B.; PATERSON, H.; LAYBOURN-PARRY, J.; RENGEFORS, K.; TRANVIK, L.; BERTILSSON, S. 2013. Biogeography of bacterial communities exposed to progressive long-term environmental change. *The ISME journal*, v. 7, p. 937-948.
- LORENZEN, C.J. 1967. Determination of chlorophyll and pheo-pigments: spectrophotometric equations. *Limnology Oceanography*, v. 12, p. 343-346.
- LUIZ, O.J.; MENDES, T.C.; BARNECHE, D.R.; FERREIRA, C.G.W.; NOGUCHI, R.; VILLAC, R.C.; RANGEL, C.A.; GASPARINI, J.L.; FERREIRA, C.E.L. 2015. Community structure of reef fishes on a remote oceanic island (St Peter and St Paul's Archipelago, equatorial Atlantic): the relative influence of abiotic and biotic variables. *Marine and Freshwater Research*, v. 66, 739-749.
- LYNCH, J.M.; BENEDETTI, A.; INSAM, H.; NUTI, M.P.; SMALLA, K.; TORSVIK, V.; NANNIPIERI, P. 2004. Microbial diversity in soil: ecological theories, the contribution of molecular techniques and the impact of transgenic plants and transgenic microorganisms. *Biology and Fertility of Soils*, v. 40, p. 363-385.
- MACEDO, S.J.; MONTES, M.J.F.; LINS, I.C.; COSTA, K.M.P. 1998. REVIZEE: programa da avaliação do potencial sustentável dos recursos vivos da Zona Econômica Exclusiva. SCORE/NE. In: *Relatório da Oceanografia Química*. UFPE, Recife, p. 37.
- MACKEY, M.D.; MACKEY, D.J.; HIGGINS, H.W.; WRIGHT, S.W. 1996. CHEMTAX - a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. *Marine Ecology Progress Series*, v. 144, p. 265-283.
- MACLEAN, D.; JONES, J.D.; STUDHOLME, D.J. 2009. Application of 'next-generation' sequencing technologies to microbial genetics. *Nature Reviews Microbiology*, v. 7, p. 287-296.
- MAGURRAN, A.E. 1988. *Ecological diversity and its measurement*. Cambridge: Cambridge University Press, 179 p.
- MAIA L.P.; FREIRE, G.S.S.; LACERDA L.D. 2005. Accelerated dune migration and sand transport during El Niño events along the NE Brazilian coast. *Journal of Coastal Research*, v. 21, p. 1121-1126.

- MARENGO, J.A.; TORRES, R.R.; ALVES, L.M. 2017. Drought in Northeast Brazil - past, present, and future. *Theoretical and Applied Climatology*, v. 129, p. 1189-1200.
- MARGULIS, L. 1981. Symbiosis in cell evolution: Life and its environment on the early Earth. Boston: NASA Technical Reports Server (NTRS). 438 p.
- MARIE, D., RIGAUT-JALABERT, F., VAULOT, D., 2014. An improved protocol for flow cytometry analysis of phytoplankton cultures and natural samples. *Cytometry Part A*, v. 85A, p. 962-968.
- MARIE, D.; PARTENSKY, F.; JACQUET, S.; VAULOT, D. 1997. American society for microbiology enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Applied Environmental Microbiology*, v. 63, p. 186-193.
- MARIE, D.; PARTENSKY, F.; VAULOT, D.; BRUSSAARD, C. 2001. Enumeration of Phytoplankton, Bacteria, and Viruses in Marine Samples. *Current Protocols in Cytometry*, v. 10, p. 11.11.1-11.11.15.
- MARIN, F.O. 2009. A Subcorrente Norte do Brasil ao Largo da Costa do Nordeste. Dissertação (Mestrado) - Instituto Oceanográfico, Universidade de São Paulo, 115p.
- MARINHO, A.; JANSEN, R.; MARINHO, S. 2010. nos limites da Amazônia Azul: as ilhas de São Pedro e São Paulo + Trindade. Rio de Janeiro: Casa da Palavra, 160 p.
- MARTINY, J.B.H.; BOHANNAN, B.J.M.; BROWN, J.H.; COLWELL, R.K.; FUHRMAN, J.A.; GREEN, J.L.; HORNER-DEVINE, M.C.; KANE, M.; KRUMINS, J.A.; KUSKE, C.R.; MORIN, P.J.; NAEEM, S.; ØVREÅS, L.; REYSENBACH, A.L.; SMITH, V.H.; STALEY, J.T. 2006. Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, v. 4, p. 102-112.
- MAXWELL, K.; JOHNSON, G.N. 2000. Chlorophyll fluorescence: a practical guide. *Journal Experimental Botany*, v. 51, 659-668.
- MAYER, D.A.; WEISBERG, R.H. 1993. A description of COADS surface meteorological fields and the implied sverdrup transports for the atlantic ocean from 30°S to 60°S. *Journal of Physical Oceanography*, v. 23, p. 2201-2221.
- MCMURDIE, P.J.; HOLMES, S. 2014. Waste not, want not: why rarefying microbiome data is inadmissible. *Plos Computational Biology*, v. 10, p. 1-12.
- MEIRELLES, P.M.; AMADO-FILHO, G.M.; PEREIRA-FILHO, G.H.; PINHEIRO, H.T.; MOURA, R.L.; JOYEUX, J.C.; MAZZEI, E.F.; BASTOS, A.C.; EDWARDS, R.A.; DINDALE, E.; PARANHOS, R.; SANTOS, E.O.; IIDA, T.; GOTOH, K.; NAKAMURA, S.; SAWABE, T.; REZENDE, C.E.; GADELHA JR, L.M.R.,

- FRANCINI-FILHO, R.B.; THOMPSON, C.; THOMPSON, F.L. 2015. Baseline assessment of mesophotic reefs of the Vitória-Trindade seamount chain based on water quality, microbial diversity, benthic cover and fish biomass data. *Plos One*, v. 10, p. 1-22.
- MENZEL, D.W.; VACCARO, R.F. 1964. The measurement of dissolved organic and particulate carbon in seawater. *Limnology and Oceanography*, v. 9, p. 138-42.
- MILICI, M.; TOMASCH J.; WOS-OXLEY, M.L.; DECELLE, J.; JÁUREGUI, R.; WANG, H.; DENG, Z.D.; PLUMEIER, I.; GIEBEL, H.A.; BADEWIEN, T.H.; WURST, M.; PIEPER, D.H.; SIMON, M.; WAGNER-DÖBLER, I. 2016. Bacterioplankton Biogeography of the Atlantic Ocean: A Case Study of the Distance-Decay Relationship. *Frontiers in Microbiology*, v. 7, p. 1-15.
- MISHONOV, A.V.; GARDNER, W.D.; RICHARDSON, M.J. 2003. Remote sensing and surface POC concentration in the South Atlantic. *Deep-Sea Research Part II*, v. 50, p. 2997-3015.
- MOHR, L.V.; CASTRO, J.W.A.; COSTA, P.M.S.; ALVES, R.J.V. 2009. *Ilhas oceânicas brasileiras: da pesquisa ao manejo*. Brasília: Secretaria de Biodiversidade e Florestas, 501 p.
- MOORE, R.; CLARK, W.D.; KINGSLEY, R.S.; VODOPICH, D. 1995. *Botany*. Dubuque: William C. Brown Pub., 832 p.
- MORAES, S.O.C. 2011. Caracterização da corrente norte do Brasil na região da retroflexão. Dissertação (Mestrado) - Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro, 66 p.
- MORAIS, R.A.; FERREIRA, C.E.L.; FLOETER, S.R. 2017. Spatial patterns of fish standing biomass across Brazilian reefs. *Journal of Fish Biology*, v. 91, p. 1642–1667.
- MORO, M.F.; CASTRO, A.S.F.; ARAÚJO, F.S. 2011. Composição florística e estrutura de um fragmento de vegetação savânica sobre os tabuleiros pré-litorâneos na zona urbana de Fortaleza, Ceará. *Rodriguésia*, v. 62, p. 407-423.
- MOTA, E.M.T.; GARCIA, T.M.; FREITAS, J.E.P.; SOARES, M.O. 2017. Composition and cross-shelf distribution of ichthyoplankton in the Tropical Southwestern Atlantic. *Regional Studies in Marine Science*, v. 14, p. 27-33.
- MOTOKI, A.; SICHEL, S.E.; CAMPOS, T.F.C.; SRIVASTAVA, N.K.; SOARES, R. 2009. Present-day uplift rate of the Saint Peter and Saint Paul Islets, Equatorial Atlantic Ocean. *REM International Engineering Journal*, v. 62, p. 331-342.

- MÜHLING, M.; WOOLVEN-ALLEN, J.; MURRELL, J.C.; IAN JOINT, I. 2008. Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *The ISME Journal*, v. 2, p. 379-392.
- MUYZER, G.; DE WAAL, E.C.; UITTERLINDEN, A.G. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*, v. 59, p. 695-700.
- NAKATSU, C.H. 2007. Soil microbial community analysis using denaturing gradient gel electrophoresis. *Soil Science Society of America Journal*, v. 71, p. 562-571.
- NITSCKE, W.; RUTHERFORD, A.W. 1991. Photosynthetic reaction centres: variations on a common structural theme? *Trends in Biochemical Sciences*, v. 16, p. 241-245.
- NÜBEL, U.; GARCIA-PICHEL, F.; MUYZER, G. 1997. PCR primers to amplify 16S rRNA genes from cyanobacteria. *Applied and Environmental Microbiology*, v. 63, p. 3327-3332.
- NÜBEL, U.; GARCIA-PICHEL, F.; MUYZER, G. 1997. PCR primers to amplify 16S rRNA genes from cyanobacteria. *Applied and Environmental Microbiology*, v. 63, p. 3327-3332.
- O'REILLY, J.E.; MARITORENA, S.; MITCHELL, B.G.; SIEGEL, D.A.; CARDER, K.L.; GARVER, S.A.; MCCLAIN, C.R. 1998. Ocean color chlorophyll algorithms for SeaWiFS, *Journal of Geophysical Research*, v. 103, p. 24,937-24,953.
- ODUM, E.P. 2007. *Fundamentos de ecologia*. 5. ed. São Paulo: Thomson, 612 p.
- OKSANEN, J. 2015. *Multivariate Analysis of Ecological Communities in R: vegan tutorial*. eBook, 43 p.
- OKSANEN, J. 2018. *Vegan: ecological diversity*. eBook, 12 p.
- PACE, N.R. 1997. A molecular view of microbial diversity and the biosphere. *Science*, v. 276, p. 734-740.
- PALINSKA, K.A.; SUROSZ, W. 2014. Taxonomy of cyanobacteria: a contribution to consensus approach. *Hydrobiologia*, v. 740, p. 1-11.
- PARTENSKY, F.; HESS, W.R.; VAULOT, D. 1999. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiology and Molecular Biology Reviews*, v. 63, p. 106-127.

- PARTENSKY, F.; VAULOT, D. 1999. Current protocols in cytometry: supplement 10. In: ROBINSON, J.P., ed. Microbiological applications. Nova Iorque: John Wiley & Sons, Inc., p. 11.11.1-11.11.15.
- PARTENSKY, F.; VAULOT, D., 1999. Current protocols in cytometry: supplement 10. In: ROBINSON, J.P., ed. Microbiological applications. Nova Iorque: John Wiley & Sons, Inc., p. 11.11.1-11.11.15.
- PEREIRA, N.S.; MANSO, V.A.V.; SILVA, A.M.C.; SILVA, M.B. 2010. Geomorphological mapping and morphodynamic of Rocas atoll, South Atlantic. *Journal of Integrated Coastal Zone Management*, v. 10, p. 331-345.
- PEREIRA, S.P. 2012. Modelagem da qualidade bacteriológica das águas costeiras de Fortaleza (Nordeste do Brasil). 2012. 174 p. Tese (Doutorado) - Universidade Federal do Ceará.
- PETERSON, R.G.; STRAMMA, L. 1991. Upper-level circulation in the South Atlantic Ocean. *Progress in Oceanography*, v. 26, p. 1-73.
- PETROSINO, J.F.; HIGHLANDER, S.; LUNA, R.A.; GIBBS, R.A.; VERSALOVIC, J. 2009. Metagenomic Pyrosequencing and Microbial Identification. *Clinical Chemistry*, v. 55, p. 856-866.
- POMEROY, L.R. 1974. The ocean's food web, a changing paradigm. *BioScience*, v. 24, p. 499-504.
- POREBSKI, S.; BAILEY, L.G. BAUM, B.R. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter*, v. 15, p. 8-15.
- POULTON, A.J.; HOLLIGAN, P.M.; HICKMAN, A.; KIM, Y.N.; ADEY, T.R.; STINCHCOMBE, M.C.; HOLETON, C.; ROOT, S.; WOODWARD, E.M.S. 2006. Phytoplankton carbon fixation, chlorophyll-biomass and diagnostic pigments in the Atlantic Ocean. *Deep-Sea Research Part II*, v. 53, p. 1593-1610.
- PROSSER, J.I.; BOHANNAN, B.J.M.; CURTIS, T.P.; ELLIS, R.J.; FIRESTONE, M.K.; FRECKLETON, R.P.; GREEN, J.L.; GREEN, L.E.; KILLHAM, K.; LENNON, J.J.; OSBORN, A.M.; SOLAN, M.; VAN DER GAST, C.J.; YOUNG, J.P.W. 2007. The role of ecological theory in microbial ecology. *Nature*, v.5, p. 384-392.
- PYLRO, V.S.; ROESCH, L.F.; MORAIS, D.K.; CLARK, I.M.; HIRSCH, P.R.; TÓTOLA, M.R. 2014. Data analysis for 16S microbial profiling from different

- benchtop sequencing platforms. *Journal of Microbiological Methods*, v. 107, p. 30-37.
- QUAST, C.; PRUESSE, E.; YILMAZ, P.; GERKEN, J.; SCHWEER, T.; YARZA, P.; PEPLIES, J.; GLÖCKNER, F.O. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, v. 41, p. D590–D596.
- RASSE, R.; OLMO, G.D.; GRAFF, J.; WESTBERRY, T.K.; VAN DONGEN-VOGELS, V.; BEHRENFELD, M.J. 2017. Evaluating optical proxies of particulate organic carbon across the surface Atlantic Ocean. *Frontiers in Marine Science*, v. 4, p. 1-18.
- REVERDIN, G.; RUAL, P.; DU PENHOAT, Y.; GOURIOU, Y. 1991. Vertical structure of the seasonal cycle in the central equatorial Atlantic Ocean: XBT sections from 1980 to 1988. *Journal of Physical Oceanography*, v. 21, p. 27-291.
- REYNOLDS, C.S. 2006. *The Ecology of Phytoplankton: ecology, biodiversity, and conservation*. Cambridge: Cambridge University Press, 535 p.
- ROCAP, G.; DISTEL, D.L.; WATERBURY, J.B.; CHISHOLM, S.W. 2002. Resolution of *Prochlorococcus* and *Synechococcus* eco-types by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Applied and Environmental Microbiology*, v. 68, p. 1180-1191.
- ROCAP, G.; DISTEL, D.L.; WATERBURY, J.B.; CHISHOLM, S.W.; 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Applied and Environmental Microbiology*, v. 68, p. 1180-1191.
- ROCAP, G.; LARIMER, F.W.; LAMERDIN, J.; MALFATTI, S.; CHAIN, P.; AHLGREN, N.A.; Arellano, A.; Coleman, M.; Hauser, L.; Hess, W.R.; Johnson, Z.I.; Land, M.; Lindell, D.; Post, A.F.; Regala, W.; Shah, M.; Shaw, S.L.; Steglich, C.; Sullivan, M.B.; Ting, C.S.; Tolonen, A.; Webb, E.A.; Zinser, E.R.; Chisholm, S.W. 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature*, v. 424, p. 1042-1047.
- RODRIGUEZ, E.; AZEVEDO, R.; COSTA, A.; SERÔDIO, J.; SANTOS, C. 2012. Chloroplast functionality assessment by flow cytometry: case study with pea plants under Paraquat stress. *Photosynthetica*, v. 50, p. 197-205.
- RODRIGUEZ, V.; BAUTISTA, B.; JIMENEZ-GOMEZ, F.; GUERRERO, F.; BLANCO, J.M.; RODRIGUEZ, J. 1998. Loss of pico- and nanoplankton cells during

- concentration by tangential flow filtration. *Journal of Plankton Research*, v. 20, p.1087-1097.
- RODRÍGUEZ, V.; BAUTISTA, B.; JIMÉNEZ-GÓMEZ, F.; GUERRERO, F.; BLANCO, J.M.; RODRÍGUEZ, J. 1998. Loss of pico- and nanoplankton cells during concentration by tangential flow filtration. *Journal of Plankton Research*, v. 20, p. 1087-1097.
- RODRÍGUEZ, V.; BAUTISTA, B.; JIMÉNEZ-GÓMEZ, F.; GUERRERO, F.; BLANCO, J.M.; RODRÍGUEZ, J. 1998. Loss of pico- and nanoplankton cells during concentration by tangential flow filtration. *Journal of Plankton Research*, v. 20, p. 1087-1097.
- ROY, S.; SATHYENDRANATH, S.; PLATT, T. 2017. Size-partitioned phytoplankton carbon and carbon-to-chlorophyll ratio from ocean colour by an absorption based bio-optical algorithm. *Remote Sensing of Environment*, v. 194, p. 177-189.
- SATO, M.; KODAMA, T.; HASHIHAMA, F.; FURUYA, K. 2015. The effects of diel cycles and temperature on size distributions of pico- and nanophytoplankton in the subtropical and tropical Pacific Ocean. *Plankton & Benthos Research*, v. 10, p. 26-33.
- SCHIRRMEISTER, B.E.; ANTONELLI, A.; BAGHERI, H.C. 2011. The origin of multicellularity in cyanobacteria. *BMC Evolutionary Biology*, v. 11, p. 1-21.
- SCHLOSS, P.D.; WESTCOTT, S.L.; RYABIN, T.; HALL, J.R.; HARTMANN, M.; HOLLISTER, E.B.; LESNIEWSKI, R.A.; OAKLEY, B.B.; PARKS, D.H.; ROBINSON, C.J.; SAHL, J.W.; STRES, B.; THALLINGER, G.G.; VAN HORN, D.J.; CAROLYN F. WEBER, C.F. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, v. 75, p. 7537-7541.
- SCHMIDT, T.M. 2006. The maturing of microbial ecology. *International Microbiology*, v. 9, p. 217-223.
- SCHREIBER, U. 1998. Chlorophyll fluorescence: New instruments for special applications. In: GARAB, G., ed. *Photosynthesis: Mechanisms and effects*. Dordrecht: Kluwer Academic Publishers, pp. 4253-4258.
- SERÔDIO, J.; VIEIRA, S.; CRUZ, S.; COELHO, H. 2006. Rapid light-response curves of chlorophyll fluorescence in microalgae: relationship to steady-state light curves and non-photochemical quenching in benthic diatom-dominated assemblages. *Photosynthesis Research*, v. 90, p. 29-43.

- SHEFFIELD, V.C.; COX, D.R.; LERMAN, L.S.; MYERS, R.M. 1989. Attachment of a 40-base-pair G1C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proceedings of the National Academy of Sciences of the United States of America*, v. 86, p. 232-236.
- SHENDURE, J.; JI, H. 2008. Next-generation DNA sequencing. *Nature Biotechnology*, v. 26, p.1135-1145.
- SHIH, P.M.; WU, D.; LATIFI, A.; AXEN, S.D.; FEWER, D.P.; TALLA, E.; CALTEAU, A.; CAI, F.; MARSAC, N.T.; RIPPKA, R.; HERDMAN, M.; SIVONEN, K.; COURSIN, T.; LAURENT, T.; GOODWIN, L.; NOLAN, M.; DAVENPORT, K.W.; HAN, C.S.; RUBIN, E.M.; EISEN, J.A.; WOYKE, T.; GUGGER, M.; KERFELD, C.A. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, v. 110, p. 1053-1058.
- SICHEL, S.E.; ESPERANÇA, S.; MOTOKI, A.; MAIA, M.; HORAN, M.F.; SZATMARI, P.; ALVES, E.C.; MELLO, S.L.M. 2008. Geophysical and geochemical evidence for cold upper mantle beneath the Equatorial Atlantic Ocean. *Brazilian Journal of Geophysics*, v. 26, p. 69-86.
- SICHEL, S.E.; ESPERANÇA, S.; MOTOKI, A.; MAIA, M.; HORAN, M.F.; SZATMARI, P.; ALVES, E.C.; MELLO, S.L.M. 2008. Geophysical and geochemical evidence for cold upper mantle beneath the equatorial atlantic ocean. *Revista Brasileira de Geofísica*, v. 26, p. 69-86.
- SIEBURTH, J.M.N.; SMETACEK, V.; LENZ, J. 1978. Pelagic ecosystem structure: heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnology Oceanography*, 23, 1256-1263.
- SIKORSKI, J. 2015. The prokaryotic biology of soil. *Soil Organisms*. v. 87, p. 1-28.
- SILVA, E.M.; CASTRO, M.A.H. 2006. Uma análise preliminar da distribuição espacial da climatologia de precipitação pluviométrica simulada em bacias hidrográficas no estado do Ceará - Brasil. *Revista Tecnológica*, v. 27, p. 7-18.
- SILVA, M.; ARAUJO, M.; SERVAIN, J.; PENVEN, P.; LENTINI, C.A.D. 2009. High-resolution regional ocean dynamics simulation in the southwestern tropical Atlantic. *Ocean Modelling*, v. 30, p. 25-26.
- SIMON, M.; AZAM, F. 1989. Protein content and protein synthesis rates of planktonic marine bacteria. *Marine Ecology Progress Series*, v. 51, p. 201-213.

- SINCLAIR, L.; OSMAN, O.A.; BERTILSSON, S.; EILER, A. 2014. Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the Illumina platform. *Plos One*, v. 10, p. 1-18.
- SINCLAIR, L.; OSMAN, O.A.; BERTILSSON, S.; EILER, A. 2015. Microbial Community composition and diversity via 16S rRNA gene amplicons: evaluating the Illumina platform. *Plos One*, v. 10, p. 1-18.
- SISSINI, M.N.; BARRETO, M.B.B.B.; SZECHY, M.T.M.; LUCENA, M.B.; OLIVEIRA, M.C.; GOWER, J.; LIU, G.; BASTOS, E.O.; MILSTEIN, D.; GUSMÃO, F.; MARTINELLI-FILHO, J.E.; ALVES-LIMA, C.; COLEPICOLO, P.; AMEKA, G.; JOHNSON, K.G.; GOUVEA, L.; TORRANO-SILVA, B.; NAUER, F.; NUNES, J.M.C.; BARUFI, J.B.; RÖRIG, L.; RIOSMENA-RODRÍGUEZ, R.; MELLO, T.J.; LOTUFO, L.V.C.; HORTA, P.A. 2017. The floating Sargassum (Phaeophyceae) of the South Atlantic Ocean: likely scenarios. *Phycologia*, v. 56, p. 321-328.
- SOUZA, C.S.; CONCEIÇÃO, L.R.; JUNIOR, P.O.M. 2016. Hyperiid amphipods around the seamounts and islands off northeastern Brazil. *Brazilian Journal of Oceanography*, v. 64, p. 339-352.
- SOUZA, C.S.; LUZ, J.A.G.; MACÊDO, S.J.; MONTES, M.J.F.; MAFALDA, P.J.R. 2013. Chlorophyll a and nutrient distribution around seamounts and islands of the tropical south-western Atlantic. *Marine and Freshwater Research*, v. 64, p. 168-184.
- STANIER, R.Y.; SISTROM, W.R.T.; HANSEN, A.; WHITTON, B.A.; CASTENHOLZ, R.W.; PFENNIG, N.; GORLENKO, W.N.; KONDRATIEVA, E.N.; EIMHJELLEN, K.E.; WHITTENBURY, R.; GHERNA, R.L.; TRIIPER, H.G. 1978. Proposal to placethe nomenclature of the cyanobacteria (blue-green algae) under the rules of the Internationa Clode of Nomenclature of Bacteria. *International Journal of Systematic Bacteriology*, v. 28, p. 335-336.
- STRAMMA, L. 1999. On the water masses and mean circulation of the South Atlantic Ocean. *Journal Of Geophysical Research*, v. 104, p. 20,863-20,883.
- STRAMSKA, M. 2014. Particulate organic carbon in the surface waters of the North Atlantic: spatial and temporal variability based on satellite ocean colour. *International Journal of Remote Sensing*, v. 35, p. 4717-4738.
- STRAMSKI, D.; REYNOLDS, R. A.; BABIN, M.; KACZMAREK, S.; LEWIS, M. R.; RÖTTGERS, R.; SCIANDRA, A.; STRAMSKA, M.; TWARDOWSKI, M.S.; FRANZ, B.A.; CLAUSTRE, H. 2008. Relationships between the surface

- concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans. *Biogeosciences*, v. 5, p.171-201.
- STRAMSKI, D.; REYNOLDS, R.A.; BABIN, M.; KACZMAREK, S.; LEWIS, M.R.; RÖTTGERS, R.; SCIANDRA, A.; STRAMSKA, M.; TWARDOWSKI, M.S.; FRANZ, B.A.; CLAUSTRE, H. 2008. Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans. *Biogeosciences* 5, 171-201.
- STRAMSKI, D.; REYNOLDS, R.A.; KAHRU, M.; MITCHELL, B.G. 1999. Estimation of particulate organic carbon in the ocean from satellite remote sensing. *Science*, v. 285, p. 239-242.
- SUGGETT, D.J.; PRÁŠIL, O.; BOROWITZKA, M.A. 2010. Chlorophyll-a fluorescence in aquatic sciences: Methods and applications. London: Springer. 323 p.
- SUNAGAWA, S.; COELHO, L.P.; CHAFFRON, S.; KULTIMA, J.R.; LABADIE, K.; SALAZAR, G.; DJAHANSCHIRI, B.; ZELLER, G.; MENDE, D.R.; ALBERTI, A.; CORNEJO-CASTILLO, F.M.; COSTEA, P.I.; CRUAUD, C.; OVIDIO, F.; ENGELEN, S.; FERRERA, I.; GASOL, J.M.; GUIDI, L.; HILDEBRAND, F.; KOKOSZKA, F.; LEPOIVRE, C.; LIMA-MENDEZ, G.; POULAIN, J.; POULOS, B.T.; ROYO-LLONCH, M.; SARMENTO, H.; VIEIRA-SILVA, S.; DIMIER, C.; PICHERAL, M.; SEARSON, S.; KANDELS-LEWIS, S.; BOWLER, C.; VARGAS, C.; GORSKY, G.; GRIMSLEY, N.; HINGAMP, P.; IUDICONE, D.; JAILLON, O.; NOT, F.; OGATA, H.; PESANT, S.; SPEICH, S.; STEMMANN, L.; SULLIVAN, M.B.; WEISSENBACH, J.; WINCKER, P.; KARSENTI, E.; RAES, J.; ACINAS, S.G.; BORK, P. 2015. Structure and function of the global ocean microbiome. *Science*, v. 348, p. 1-9.
- ŚWIRGOŃ, M.; STRAMSKA, M. 2015. Comparison of in situ and satellite ocean color determinations of particulate organic carbon concentration in the global ocean. *Oceanologia*, v. 57, p. 25-31.
- TATON, A.; GRUBISIC, S.; BRAMBILLA, E.; DE WIT, R.; WILMOTTE, A. 2003. Cyanobacterial diversity in natural and artificial microbial mats of Lake Fryxell (McMurdo Dry Valleys, Antarctica): a morphological and molecular approach. *Applied and Environmental Microbiology*, v. 69, p. 5157-5169.

- TCHAMABI, C.C.; ARAUJO, M.; SILVA, M.; BOURLÈS, B. 2017. A study of the Brazilian Fernando de Noronha island and Rocas atoll wakes in the tropical Atlantic. *Ocean Modelling*, v. 111, p. 9-18.
- TEIXEIRA, W.; CORDANI, V.G.; MENOR, E.A.; TEIXEIRA, M.G.; LINSKER, R. 2003. Arquipélago Fernando de Noronha: o paraíso do vulcão. São Paulo: Terra Virgem, 167 p.
- THOMAS, Y.; GAREN, P.; COURTIES, C.; CHARPY, L. 2010. Spatial and temporal variability of the pico- and nanophytoplankton and bacterioplankton in a deep Polynesian atoll lagoon. *Aquatic Microbial Ecology*, v. 59, p. 89-101.
- THOMPSON, C.C.; CHIMETTO, L.; EDWARDS, R.A.; SWINGS, J.; STACKEBRANDT, E.; THOMPSON, F.L. 2013a. Microbial genomic taxonomy. *BMC Genomics*, v. 14, p. 1-8.
- THOMPSON, C.C.; SILVA, G.G.Z.; VIEIRA, N.M.; EDWARDS, R.; VICENTE, A.C.P.; THOMPSON, F.L. 2013b. Genomic taxonomy of the genus *Prochlorococcus*. *Microbial Ecology*, v. 66, p. 752-762.
- THOMPSON, F.L.; BRUCE, T.; GONZALEZ, A.; CARDOSO, A.; CLEMENTINO, M.; COSTAGLIOLA, M.; HOZBOR, C.; OTERO, E.; PICCINI, C.; PERESSUTTI, S.; SCHMIEDER, R.; EDWARDS, R.; SMITH, M.; TAKIYAMA, L.R.; VIEIRA, R.; PARANHOS, R.; ARTIGAS, L.F. 2011. Coastal bacterioplankton community diversity along a latitudinal gradient in Latin America by means of V6 tag pyrosequencing. *Archives of Microbiology*, v. 193, p. 105-114.
- TOMAS, C.R. 1997. Identifying marine phytoplankton. San Diego: Academic Press. 858 p.
- TORRES, M.A.; BARROS, M.P.; CAMPOS, S.C.G.; PINTO, E.; RAJAMANI, S.; SAYRE, R.T.; COLEPICOLO, P. 2008. Biochemical biomarkers in algae and marine pollution: a review. *Ecotoxicology Environmental Safety*, v. 71, p. 1-15.
- TROUSSELIER, M.; BOUVY, M.; COURTIES, C.; DUPUY, C. 1997. Variation of carbon content among bacterial species under starvation condition. *Aquatic Microbial Ecology*, v. 13, p. 113-119.
- UCHIYAMA, I.; MIHARA, M.; NISHIDE, H.; HIROKAZU CHIBA, H. 2015. MGD update 2015: microbial genome database for flexible ortholog analysis utilizing a diverse set of genomic data. *Nucleic Acids Research*, v. 43, p. D270-D276.

- URBACH, E.; ROBERTSON, D.L.; CHISHOLM, S.W. 1992. Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. *Nature*, v. 355, p. 267-270.
- VAN DEN HOEK, C.; MANN, D.G.; JAHNS, H.M. 1997. *Algae: An introduction to phycology*. Cambridge: Cambridge University Press, 627 p.
- VANDAMME, P.; POT, B.; GILLIS, M.; DE VOS, P.; KERSTERS, K.; SWINGS, J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological Reviews*, v. 60, p. 407-438.
- VAULOT, D.; EIKREM, W.; VIPREY, M.; MOREAU, H. 2008. The diversity of small eukaryotic phytoplankton ( $\leq 3\mu\text{m}$ ) in marine ecosystems. *FEMS Microbiology Reviews*, v. 32, p. 795-820.
- VELDHUIS, M.J.W.; TIMMERMANS, K.R.; CROOT, P.; VAN DER WAGT, B. 2005. Picophytoplankton: a comparative study of their biochemical composition and photosynthetic properties. *Journal Sea Research*, v. 53, p. 7-24.
- VĚTROVSKÝ, T.; BALDRIAN, P. 2013. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *Plos One*, v. 8, p. 1-10.
- VIAENE, K.P.J.; LAENDER, F.; BRINK, P.J.V.; JANSSEN, C.R. 2013. Using additive modelling to quantify the effect of chemicals on phytoplankton diversity and biomass. *Science of the Total Environment*, v. 449, p. 71-80.
- VIANA, D.F.; HAZIN, F.H.V.; ANDRADE, H.A.; NUNES, D.M.; VIANA, D.L. 2015. Fisheries in the Saint Peter and Saint Paul Archipelago: 13 years of monitoring. *Boletim do Instituto de Pesca*, v. 41, p. 239-248.
- WALKLEY, A.; BLACK, I.A. 1934. An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science*, v. 37, p. 29-38.
- WALTER, J.M.; COUTINHO, F.H.; DUTILH, B.E.; SWINGS, J.; THOMPSON, F.J.; CRISTIANE C. THOMPSON, C.C. 2017. Ecogenomics and taxonomy of cyanobacteria phylum. *Frontiers in Microbiology*, v. 8, p. 1-18.
- WANGERSKY, P.J. 1976. Particulate organic carbon in the Atlantic and Pacific oceans. *Deep-Sea Research*, v. 23, p. 457-465.
- WATERBURY, J.B. 2006. *The cyanobacteria: isolation, purification and identification*. In: DWORKIN, M.; FALKOW, S.; ROSENBERG, E.; SCHLEIFER, K.H.;

- STACKEBRANDT, E., eds. The prokaryotes: bacteria: firmicutes, cyanobacteria. New York: Springer-Verlag, p. 1053–1073.
- WHITTON, B.; POTTS, M. 2000. The ecology of cyanobacteria: Their diversity in time and space. New York: Kluwer Academic Publishers, 669 p.
- WILSON, K. 2001. Preparation of genomic DNA from bacteria. *Current Protocols in Molecular Biology*, v. 2.4, p. 2.4.1-2.4.5.
- WOESE, C.R.; FOX, G.E. 1977. Evolution phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, v. 74, p. 5088-5090.
- WOESE, C.R.; KANDLER O.; WHEELIS, M.L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, v. 87, p. 4576-4579.
- WRIGHT, E.S.; YILMAZ, L.S.; NOGUERA, D.R. 2012. DECIPHER, a Search-Based Approach to Chimera Identification for 16S rRNA Sequences. *Applied and Environmental Microbiology*, v. 78, p. 717-725.
- YILMAZ, P.; PARFREY, L.W.; YARZA, P.; GERKEN, J.; PRUESSE, E.; QUAST, C.; SCHWEER, T.; PEPLIES, J.; LUDWIG, W.; GLÖCKNER, F.O. 2014. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Research*, v. 42, p. D643-D648.
- YOUVAN, D.C.; MARRS, B.L. 1987. Molecular mechanisms of photosynthesis. *Scientific American* v.256, p. 42-48.
- YU, Z.; MORRISON, M. 2004. Comparisons of different hypervariable regions of rrs genes for use in fingerprinting of microbial communities by PCR-Denaturing Gradient Gel Electrophoresis. *Applied and Environmental Microbiology*, v. 70, p. 4800-4806.
- ZHAXYBAYEVA, O.; GOGARTEN, J.P.; CHARLEBOIS, R.L.; DOOLITTLE, W.F.; PAPKE, R.T. 2006. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Research*, v. 16, p.1099-1108.
- ZUBKOV, M.V., SLEIGH, M.A., TARRAN, G.A., BURKILL, P.H., & LEAKEY, R.J.G. 1998. Picoplankton community structure on an Atlantic transect from 50°N to 50°S. *Deep-Sea Research Part I*, v. 45, p. 1339-1355.
- ZUBKOV, M.V.; SLEIGH, M.A.; BURKILL, P.H.; LEAKEY, R.J.G. 2000. Picoplankton community structure on the Atlantic Meridional transect: a comparison between seasons. *Progress in Oceanography*, v. 45, p. 369-386.

ZWIRGLMAIER, K.; JARDILLIER, L.; OSTROWSKI, M.; MAZARD, S.;  
GARCZAREK, L.;VAULOT, D.; NOT, F.; MASSANA, R.; ULLOA, O.;  
SCANLAN, D.J. 2008. Global phylogeography of marine *Synechococcus* and  
*Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes.  
*Environmental Microbiology*, v. 10, p. 147-161.

## CLASSIFICAÇÃO TAXONÔMICA DE CIANOBACTÉRIAS - 8 ORDENS COM 45 FAMÍLIAS ATUAIS (KOMÁREK et al., 2014; KOMÁREK, 2016a; 2016b; HALLENBECK, 2017).

### 1. Gloeobacterales

Esta ordem contém apenas uma família monofilética, 1.1 Gloeobacteraceae, e um gênero, *Gloeobacter*, com uma clara posição taxonômica independente na base de todas as cianobactérias, sendo os únicos representantes que não possuem tilacóides.

### 2. Synechococcales

Uma grande ordem, com 11 famílias e ~70 gêneros, com representantes definidos como não monofiléticos, com tipos unicelulares e filamentosos. Este grupo está unido pela presença de tilacóides parietais. Com uma maior amostragem futura será possível revisar melhor este grupo. A família 2.1 Synechococcaceae (17 gêneros) ainda precisa de estudo e revisão em todos os níveis taxonômicos, devido à extensa complexidade. Apresentam formas coloniais e envelopes mucilaginosos, o gênero *Neosynechococcus* é uma linhagem independente recentemente descrita, morfologicamente semelhante ao gênero *Synechococcus*. O gênero *Synechococcus* é um dos mais amplos desta família com uma linhagem independente que ainda se apresenta indevidamente descrita (KOMÁREK et al., 2014; KOMÁREK, 2016a; 2016b). Outra família que necessita de uma extensa revisão é a 2.2 Prochloraceae com apenas dois gêneros marinhos, caracterizada pela presença de clorofila-b. O gênero *Prochlorococcus*, é um dos mais importantes gêneros na produção primária dos oceanos, apresentando uma subdivisão em dois grupos ecologicamente distantes, os adaptados a alta luminosidade, e os adaptados a baixa luz. As relações taxonômicas deste último gênero com algumas espécies marinhas de *Synechococcus* sp. ainda não estão devidamente esclarecidas (KOMÁREK et al., 2014; KOMÁREK, 2016a; 2016b). As outras famílias deste grupo são: 2.3 Merismopediaceae; 2.4 Coelosphaeriaceae; 2.5 Acaryochloridaceae; 2.6 Chamaesiphonaceae; 2.7 Romeriaceae; 2.8 Pseudanabaenaceae; 2.9 Leptolyngbyaceae; 2.10 Heteroleibleiniaceae; 2.11 Schizotrichaceae.

### 3. Spirulinales

Esta ordem tem uma posição filogenética especial e é caracterizada por tricomas (filamentos, com sequência linear de células envolvidas por uma bainha de mucilagem) diferenciados neste grupo, pois são em forma de parafuso e sem bainha, com uma família, 3.1 Spirulinaceae.

### 4. Chroococcales

Organismos de forma cocóide e ausentes de baeócitos (endósporos), distribuídos nas famílias: 4.1. Microcystaceae; 4.2. Aphanothecaceae; 4.3. Cyanobacteriaceae; 4.4. Cyanotrichaceae; 4.5. Stichosiphonaceae; 4.6. Chroococcaceae; 4.7. Gomphosphaeriaceae; 4.8. Entophysalidaceae.

### 5. Pleurocapsales

Baseado nas poucas sequências disponíveis atualmente este grupo parece ser monofilético. Faltam muitos dados sobre esta ordem, pois os membros são de difícil cultivo e os dados genéticos são escassos, as atuais famílias são: 5.1. Hydrococcaceae; 5.2. Dermocarpellaceae; 5.3. Xenococcaceae; 5.4. Pleurocapsaceae.

### 6. Oscillatoriales

Filamentosas com tricomas relativamente estreitos e tilacóides parietais. Exceção para o gênero filogeneticamente fechado, *Cyanothece*, que possui formas cocóides. Esta ordem, eventualmente, exigirá revisão se uma maior taxa monofilética for alcançada, pois possuem organismos de transição evolutiva. Famílias: 6.1 Cyanotecaceae; 6.2. Borziaceae; 6.3 Coleofasciculaceae; 6.4 Microcoleaceae, o vasto gênero marinho *Trichodesmium*, está distribuído nesta família. As espécies de *Trichodesmium* sp. de água doce são molecularmente distintas dos tipos marinhos; 6.5. Homoeotrichaceae; 6.6. Oscillatoriaceae; 6.7. Gomontiellaceae.

### 7. Chroococciopsidales

Este grupo de organismos (única família 7.1 Chroococciopsidaceae e único gênero *Chroococciopsis*), vive em habitats extremos (*C. thermalis*). Esta ordem se aglomera, surpreendentemente, na vizinhança das cianobactérias heterocíticas (que possuem heterocisto - Tabela 3) (KOMÁREK et al., 2014).

### 8. Nostocales

Esta ordem representa um grande conjunto monofilético de cianobactérias filamentosas, e possuem células especializadas proeminentes (Tabela 3). Esta ordem agrupa o maior número de famílias: 8.1. Scytonemataceae; 8.2. Symphyonemataceae; 8.3. Rivulariaceae; 8.4. Tolypotrichaceae; 8.5. Godleyaceae; 8.6. Chlorogloeopsidaceae; 8.7. Hapalosiphonaceae; 8.8. Capsosiraceae; 8.9. Stigonemataceae; 8.10. Gloeotrichiaceae; 8.11. Aphanizomenonaceae; 8.12. Nostocaceae (KOMÁREK et al., 2014; KOMÁREK, 2016a; 2016b).

## PROTOCOLO DE EXTRAÇÃO DNA/CTAB - MEMBRANAS DE FILTRAÇÃO (POREBSKI et al., 1997; WILSON, 2001; GARCIA et al., 2013)

### Protocolo resumido

- 1 – Processar as membranas do criotubo imediatamente a retirada do freezer/N<sub>2</sub>
- 2 – Macerar em cadinho estéril.
- 3 – Add 1ml de tampão CTAB (de EDTA) com proteinase (5 µg).
- 4 – Congela/descongela – 3min -80°C / 3min 65°C (ou incubar no banho, por pelo menos 1h).
- 5 – Fenol-clorofórmio-álcool isoamílico (25:24:1) 1X e clorofórmio 1X.
- 6 – Purificação de DNA com precipitação - Nh<sub>4</sub>Cl + Isopropanol. Se a precipitação clássica não funcionar (degradação de DNA ou DNA impuro), realizaremos a opção B, com a coluna do kit de purificação da QIAGEN.

### Material

- . Criotubo de polipropileno (resistente a imersão em nitrogênio líquido)
- . Galão de nitrogênio líquido (p/ coleta)
- . Cadinho e pistão
- . Espátula
- . CTAB (Sigma Aldrich)
- . EDTA
- . Tris-HCL
- . NaCl
- . proteinase K
- . 2-mercaptoethanol (Sigma Aldrich)
- . Colunas do kit DNeasy Blood & Tissue Kit (QIAGEN)
- . Phenol:Chloroform:Isoamyl Alcohol (25:24:1) Saturated with 10 mM Tris, pH 8.0, 1 mM EDTA (Sigma Aldrich)

### Tampão de lise

- . CTAB buffer [2% CTAB (m/v; Sigma Aldrich), 1.4M NaCl, 100mM EDTA, 100mM Tris-HCl (pH 8.0)] and freshly added 5µg proteinase K (Sigma Aldrich) and 0.5% 2-mercaptoethanol (v/v; Invitrogen).

### Soluções STOCK e condição de estocagem

- . Fenol/clorofórmio/álcool isoamílico (24:23:1 v:v:v) → geladeira
- . Clorofórmio → geladeira
- . 7.5M Acetato de amônia → geladeira
- . CTAB, solução stock → 10% (p/v), estoque a RT, MAS aquecer a 60°C antes de usar (para solubilizar cristais)
- . 1M Tris-HCl pH 8.0 (pesar 12.114g /100ml) → RT
- . 0.5M EDTA (pH 8.0 com NaOH) (pesar 18.172g /100ml) → RT
- . Etanol absoluto → -4°C
- . Isopropanol → -4°C

### Antes de começar!

- . Cadinho e pistão (estéril), espátula e pinça (limpos com destilada e etanol 70%).
- . Rotular os tubos eppendorf que serão usados para receber o pó macerado.
- . Separar o nitrogênio líquido no isopor pequeno.
- . Preparar o tampão de lise (adição da PtnaseK e do 2-mercaptoetanol).
- . Se não for fazer STOP POINT - lembrar de ligar o banho seco a 60°C e o banho maria (úmido) a 56°C, necessários para as etapas de lise.

### Passo-a-passo - extração DNA/CTAB (Adaptado de Garcia et al., 2013)

1 – Macerar a membrana de filtração (imersa em nitrogênio líquido), equivalente, em cadinho com nitrogênio líquido. No processo de maceração adiciona-se constantemente nitrogênio líquido ao cadinho para não deixar a amostra descongelar. Após cada processamento ir aclimatando a amostra (tubo no freezer - 20°C, até terminar todas as amostras, esperar ainda uns 5' após a última amostra congelada, para estabilizar a temperatura).

*Obs1.: Protospalythoa variabilis – aproximadamente 1 pólipo de uma colônia.*

*Palythoa caribeorum – tecido contendo vários pólipos e muco protetor, tentar pegar mais tecido do que muco.*

2 – Todo o conteúdo macerado foi transferido para um eppendorf. Adicionou-se **1mL de tampão de lise** pré-aquecido a 60°C e **colocando-se na hora** ~5µg de proteinase K (**10uL de proteinase K para cada 1mL de tampão** – estoque [10mg/mL], proteinase em pó) e 0,5% de 2-mercaptoetanol (~5uL de mercaptoetanol para cada 1mL de tampão – 0,5% de 1000uL) (coloca-se o 2-mercaptoetanol na capela!!!).

*Obs2a.: Para cada 10 amostras, são 10mL de tampão de lise + 500uL de PtnaseK + 52,5uL (~50uL) de mercapto, disso colocaremos 1mL em cada amostra (tubo).*

*Obs2b.: Deve-se adicionar em 1 amostra de cada vez, assim apenas ir descongelando de poucas em poucas amostras. Enquanto isso as outras amostras esperam no -20oC.*

3 – Os tubos foram **incubados por 30 min a 56°C**. A cada 10' de incubação homogeneizar os tubos por inversão manual (1 vez, gentilmente).

4 – Após a lise inicial os tubos foram **congelados no -80°C**.

Pode ser **STOP POINT!!!**

5 – Os tubos são então submetidos a **três ciclos de congela-descongela: 10min freezer -70oC e 10min no banho seco a 60°C**, começando-se com banho

a 60°C. Entre as etapas homogeneizar os tubos por inversão manual (3 vezes, muito gentilmente).

6 – **Divide-se o volume do lizado de cada amostra em dois eppendorfs (~500ul em cada tubo).** Extração Fenol/Clorofórmio: Adiciona-se igual volume, ou seja, **500ul de Fenol:clorofórmio:álcool-isoalcoólico (25mL:24mL:1mL).** Homogeneizar por inversão (3 vezes, muito gentilmente). Deixar ocorrer a separação das fases na bancada da capela por cerca de 10' (importante para desproteínia melhor).

7 – Centrifugação por **10min a 11.000 RPM a 4°C.**

8 – **Recupera-se a fase aquosa de cada amostra em novos eppendorfs (fase de cima, onde está o valioso DNA total)** muito cuidado para não perturbar a camada proteica (película esbranquiçada) que fica entre a fase orgânica e aquosa. Descarta-se a fase orgânica (em garrafa para descarte de fenol, dentro da capela).

9 – Extração com **400uL de clorofórmio.**

10 – Centrifugação por **10min a 11.000RPM a 4°C.**

11 – **Recupera-se novamente a fase aquosa de cada amostra em novos eppendorfs, considerando um volume recuperado de cerca de ~400uL (fase de cima, onde está o valioso DNA total).** Novamente cuidado para não perturbar a camada proteica e **continuar conforme passos 12-18. Caso a extração esteja muito contaminada (coloração esbranquiçada devido a contaminação proteica/fenólica) realizar o passo 11<sup>A</sup> e somente após pular direto para o passo 18.**

11<sup>A</sup> – Transfere-se o sobrenadante para a coluna do kit DNeasy Blood & Tissue Kit (QIAGEN) e procede-se com as lavagens e eluição conforme sugerido pelo fabricante. Utilizando as colunas e os tampões wash buffer do kit (lembrando de adicionar etanol Merk absoluto nas soluções do kit, recomendação da QIAGEN), inicie o passo 1d.4 do protocolo do fabricante.

11<sup>A</sup>. 1 – Colocar o conteúdo recuperado no centro da coluna de sílica (acopla a um eppendorf) e centrifugamos (spin) de 1min a 8.000 RPM. Recupera o que passou pela coluna e armazena no antigo eppendorf no -80°C, apenas como forma de backup.

11<sup>A</sup>. 2 – Descarta o que passou pela coluna e recoloca no centro da coluna no eppendorf (acopla a coluna) e adiciona 500uL do tampão de lavagem 1 na coluna. E centrifuga por 1min a 8.000 RPM.

11<sup>A</sup>. 3 – Descarta o que passou pela coluna e recoloca no centro da coluna no eppendorf (acopla a coluna) e adiciona 500uL do tampão de lavagem 2 na coluna. E centrifuga por 3min a 14.000 RPM.

11<sup>A</sup>. 4 – Descarta o que passou pela coluna e recoloca no centro da coluna no eppendorf (acopla a coluna) e realiza outra centrifugação sem nada na coluna por 1min a 14.000 RPM, apenas para garantir a volatilização completa do etanol.

11<sup>A</sup>. 5 – Descarta o que passou pela coluna assim como o eppendorf base, recoloca a coluna em um novo eppendorf (acopla a coluna).

11<sup>A</sup>. 6 – Adiciona 50uL do tampão de eluição (TE) do próprio kit (ou o tampão do protocolo CTAB) no centro da coluna e deixa incubando 3min a temperatura ambiente. E centrifuga por 1min a 12.000 RPM. Recupera o que passou pela coluna e repete o passo 11<sup>A</sup>. 6 mas com apenas 1min de incubação.

11<sup>A</sup>. 7 – Recupera o que passou pela coluna, onde está o valioso DNA total (~50uL).

12 – Precipitação do DNA total – adiciona-se em cada tubo eppendorf 0.75X vol. de acetato de amônia 5M (**considerando um volume recuperado de cerca de ~400uL, são 300uL de acetato de amônia 5M**) e 1X vol. de isopropanol gelado (**considerando um novo volume de cerca de ~700uL, são 800uL de isopropanol gelado**). **Deixa em repouso por 01:30h no freezer (-20°C), afim de precipitar.** O volume fica perto do volume máximo de 1.5mL do tubo.

13 – Centrifugação por **30min a 14.000 RPM a 4°C.**

14 – **Descarta-se o sobrenadante,** e lava-se o pellet **200uL de etanol absoluto (Merk) gelado (-20°C).**

15 – Centrifugação por **10min a 14.000 RPM a 4°C.**

16 – **Descarta-se o sobrenadante e seca o pellet** na bancada invertido. Por ~40min-1h, **até secar totalmente.**

17 – Eluir em **50uL de tampão TE (10 mM Tris-HCL) sem agitar.** Colocar na **geladeira overnight.**

18 – Armazenar no **-80°C.**

*Obs18.: Sempre armazenar junto com as amostras um eppendorfs com 50uL de tampão TE no -80°C, pois será o branco nas leituras futuras do NanoDrop. No caso das amostras que passaram pelo passo 11<sup>A</sup> também armazenar uma alíquota do tampão TE do kit QIAGEN.*

## PROTOCOLO DE QUANTIFICAÇÃO DNA

### Protocolo resumido

- 1 – Manipular as amostras de DNA imediatamente após a retirada do freezer.
- 2 – Quantificar pelo NanoDrop e achar definir as relações 260/280 e o valor total de DNA.
- 3 – Quantificar pelo Qubit e achar a concentração de DNA dupla fita disponíveis para ao sequenciamento.
- 4 – Fazer o gel de agarose para ver a qualidade (fragmentação) do DNA extraído.
- 5 – Caso as amostras tenham ficado inadequadas tentar recupera-las utilizando o kit de purificação com a coluna do kit DNA (QIAGEN).

### Material

- . Agarose
- . Água Mili-q autoclavada
- . Tampão TE
- . Cuba de eletroforese

### Soluções STOCK e condição de estocagem

- . Kit Qubit dsDNA BR assay e o kit Qubit dsDNA HS assay → geladeira
- . Blue Juice → geladeira
- . DNA Ladder → -4°C
- . Solução de trabalho Qubit → RT
- . Agarose → RT
- . Cyber GREEN → RT, MAS armazenado no escuro

- . Brometo de etídio → RT, MAS armazenado no escuro
- . TBE 1X → RT

**Antes de começar!**

- . Ligar o NanoDrop e realizar limpeza com água mili-q.
- . Ligar o Qubit e conferir as soluções a serem utilizadas de acordo com a concentração de DNA expressa pelo NanoDrop.
- . Deixar os tampões do Qubit fora da geladeira para estarem a temperatura ambiente no momento da leitura, assim como as amostras. **Deixar no escuro**
- . Limpar a cuba e certificar se há necessidade de fazer mais tampão TBE para a corrida de eletroforese.

**Passo-a-passo - NanoDrop**

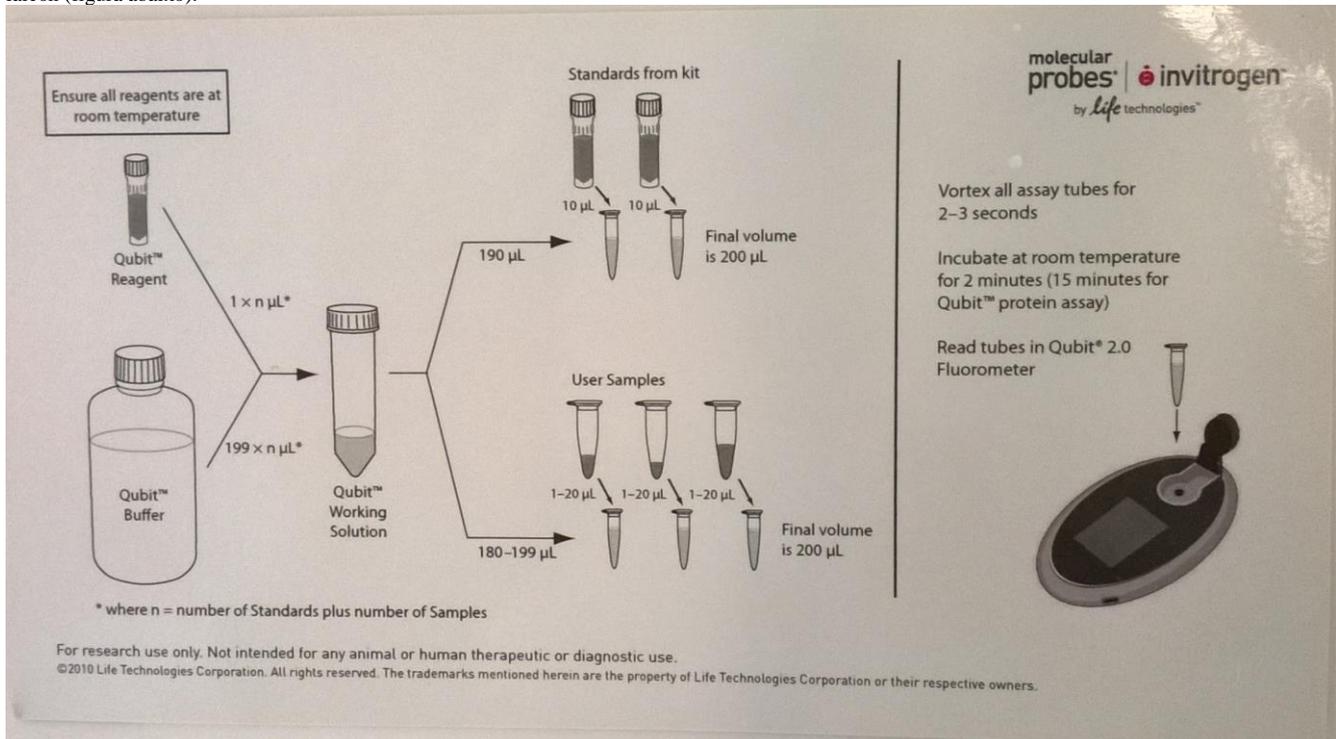
- 1 – Após ligar o NanoDrop abre o programa do aparelho no PC e marca a opção ‘add to report’ para depois conseguir selecionar os dados. Colocar **5uL de água mili-q no leitor e deixar por 5min para limpar** (solver) resíduos antigos.
- 2 – **Passar 2uL de água mili-q** como branco por **duas vezes** (marcar a opção blank), afim de calibrar o espectro. Após, passar o **branco final com 2uL do tampão de eluição** em que as amostras a serem passadas a seguir estão eluídas (TE) (marcar a opção blank). Ler o tampão TE como uma amostra, esse tem de zerar a quantidade de DNA, se isso não ocorrer repetir o passo 2 até que a leitura para o branco TE esteja zerada.
- 3 – As amostras descongelam muito rapidamente após retirar do -80°C. **Não vortexar, apenas homogeneizar com a pipeta eletrônica** em movimentos circulares, e pegar a amostra do fundo do eppendorf. Passar **2uL de amostra**, e realizar a leitura em duplicata logo após. Se a diferença for muito grande, ler em triplicata. Colocar o nome da amostra e apertar ‘measure’ para registrar a concentração de DNA (ng/uL) e as relações 260/280 e 260/230 dos 2uL de amostra avaliados.
- 4 – Entre as amostras limpar com um papel dupla folha (Neve) ou papel específico (Kimtech).
- 5 – Após ler todas as amostras, fazer uma limpeza final com **5uL de água mili-q no leitor e deixar por 5min para limpar**. E selecionar no programa do PC as amostras em ‘report’ e depois clicar em ‘export’ para o Excel.
- 6 – Imediatamente retornar as amostras ao armazenamento em **-80°C**.

**Passo-a-passo - Qubit**

1 – A análise do Qubit está associada a quantificação do NanoDrop, pois dependendo da concentração da amostra diferentes kits do Qubit tem de ser utilizado. Os kits da Molecular Probes (Life Technologies) de Labeling & Detection podem ser divididos pela quantidade de DNA quantificado. O kit Qubit dsDNA BR assay é usado para amostras entre 2 - 1.000 ng/uL. Para amostras entre 0,2 - 100 ng/uL se utiliza o kit Qubit dsDNA HS assay. Em ambos os kits o procedimento é o mesmo, mudam apenas as concentrações padrões da curva.

*Obs1.: Todos os tubos que serão lidos pelo fluorímetro Qubit tem de ser específicos para o aparelho, sendo tubos transparentes do Qubit assay tubes de 500uL, portanto não pode escrever nada nem danificar o tubo (manuseá-lo somente pela parte de cima). Tanto as amostras como as soluções da curva serão preparadas a partir de uma solução de trabalho (preparada em falcon estéril).*

2 – A **solução de trabalho** é composta pela diluição do Qubit reagente em 1:200 no Qubit buffer, ou seja, **1uL de reagente para 199uL de tampão para cada amostra. Multiplica pelo o número de amostra que serão lidas e mais as duas amostras da curva** (amarela e vermelha), resultando na solução em um falcon (figura abaixo).



- 3 – Após preparar a solução de trabalho, iniciar as **soluções da curva**, sendo **190uL da solução de trabalho para 10uL da solução da curva**, são apenas duas soluções de curva (a vermelha e a amarela). Para as **amostras são 198uL da solução de trabalho para 2uL de DNA+TE** (figura acima).
- 4 – Após, **vortexar** todos os tubos **entre 2 - 3seg e deixar incubando por 2min em temperatura ambiente** no escuro. Fazer apenas 4 amostras de cada vez (de 4 em 4), pois como tem o tempo de incubação evita tumultuar.

5 – Após a incubação, colocar os tubos no fluorímetro Qubit e pedir para **ler DNA na unidade de ng/uL** e anotar os valores.

*Obs5.: No sequenciamento pelo MiSeq (Illumina) é necessário deixar as amostras com uma **concentração** por volta dos **2 ng de DNA**, portanto uma **concentração entre 0,5 e 2 ng/uL é aceitável**, considerando que podemos botar até 5uL da solução DNA+TE no MiSeq, considerando que as amostras foram diluídas 100X pela solução de trabalho ao expressarem o valor no Qubit.*

6 – Imediatamente retornar as amostras ao armazenamento em **-80°C**.

#### Passo-a-passo - Gel de agarose

1 – A malha do gel para ver a integridade do DNA extraído foi delimitado para fragmentos de 400bp-10Kb (ideias para o sequenciamento fragmentos com 1Kb), isso corresponde a um gel de agarose 1%. No caso, a cuba do LaBBmar é pequena, assim **50mL do tampão TBE 1X adicionados com 0,5g de agarose** (Invitrogen) resulta no **gel 1%**.

*Obs1.: Misturar os 50mL de TBE 1X com a agarose em um erlenmeyer e levar ao micro-ondas durante 40s, retira e agita e coloca novamente mais alguns segundos até a solução ficar transparente. **Não deixar ferver!***

2 – Após a diluição completa pelo aquecimento, esperar esfriar até conseguir segurar o erlenmeyer com a palma da mão e neste momento adicionar o corante do gel. Para um gel de 50mL, adicionar 2,5uL de brometo de etídio a uma [10mg/mL] e agitar cuidadosamente para não formar bolhas.

*Obs2.: Devido a fotossensibilidade do brometo é necessário apagar luzes que estejam diretamente incididas sobre esta substancia.*

3 – Depois basta despejar o volume de erlenmeyer na cuba e colocar as grades (poços) e esperar solidificar (polimerizar), cerca de 10-15min.

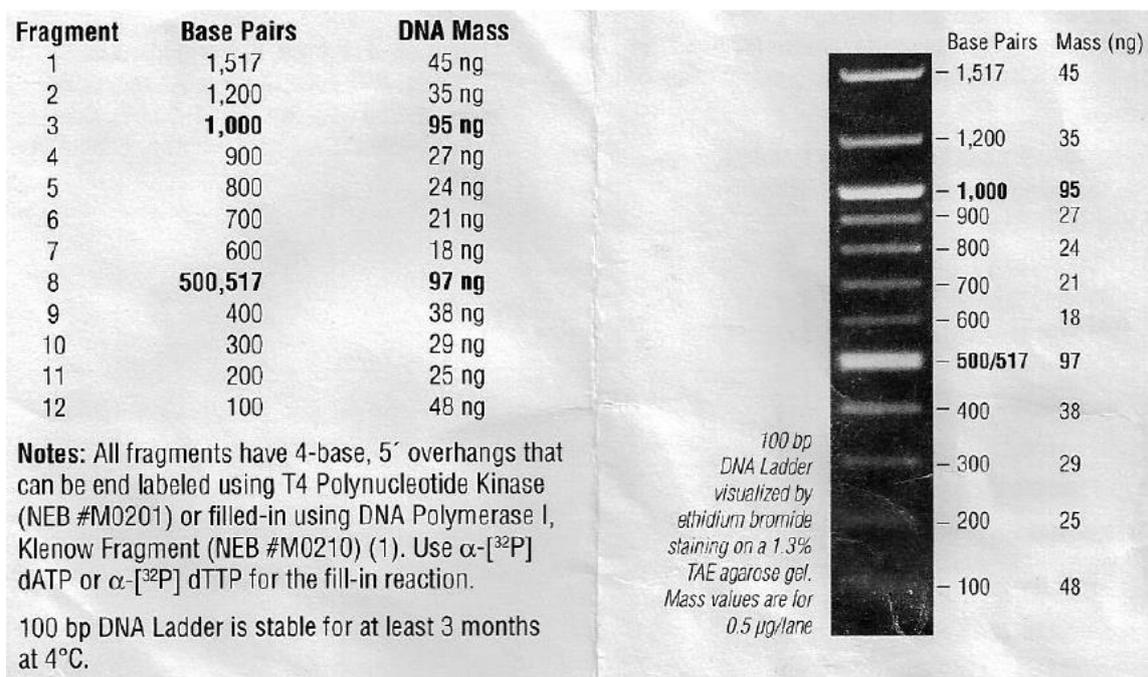
*Obs3.: Esperar solidificar no escuro, usar um revestimento de papel alumínio.*

4 – Retira as grades após solidificar e coloca o gel na posição de corrida da cuba eletroforética. Cobre o gel com o tampão TBE 1X (mesma concentração do utilizados para fazer o gel).

*Obs4.: Usar um revestimento de papel alumínio para proteger o sistema da luz. Enquanto isso manusear as amostras a serem adicionadas nos poços do gel.*

5 – Nos poços do gel serão colocados amostras e o DNA Leader, este será a escala de tamanho dos bandeamentos dos fragmentos de DNA, sempre colocar **1uL do DNA Leader no primeiro poço** de cada linha (grampo) do gel (imagem abaixo).

*Obs5.: tanto para colocar o DNA Leader tanto para colocar as amostras no gel, ter muito cuidado para não perfurar o poço com a ponteira.*



6 – Após iniciar a preparação das amostras que serão adicionadas nos poços subsequentes. Cada amostra será colocada **em um poço**, usa-se **4uL da amostra (DNA+TE) com 2uL de Blue Juice** (precipitador da amostra no poço do gel).

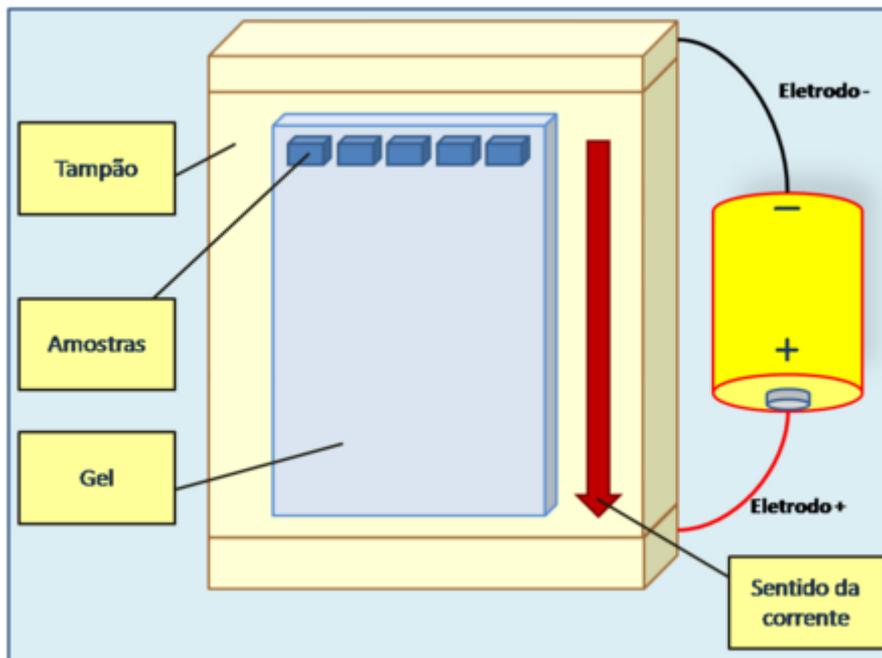
*Obs6.: Usar um papel Parafilm para juntar a amostra mais o Blue Juice (imagem abaixo). Primeiro bota a gota do Blue Juice e depois adiciona a essa gota o DNA, homogeneiza bem com a ponteira (3X) e põe tudo no poço.*



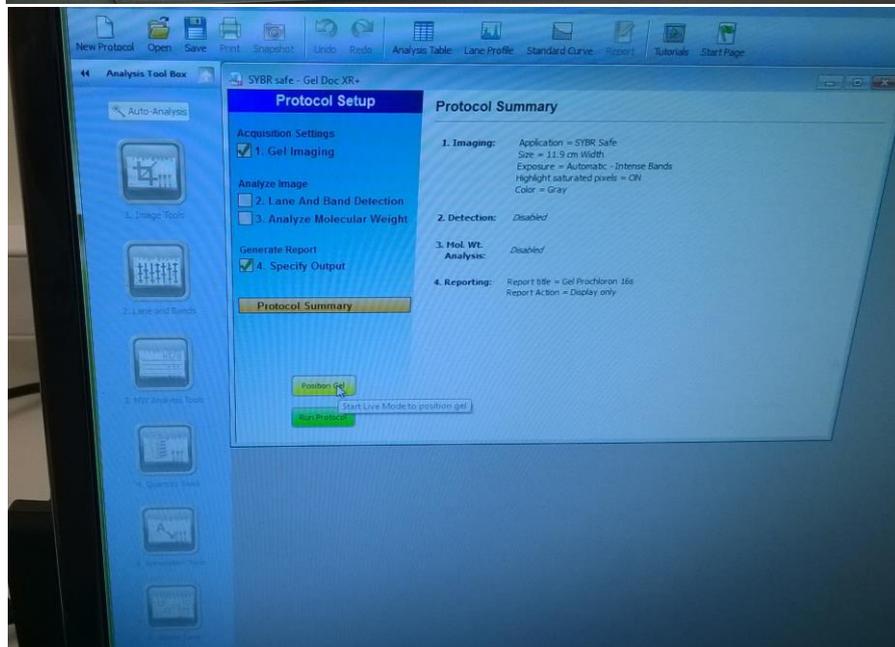
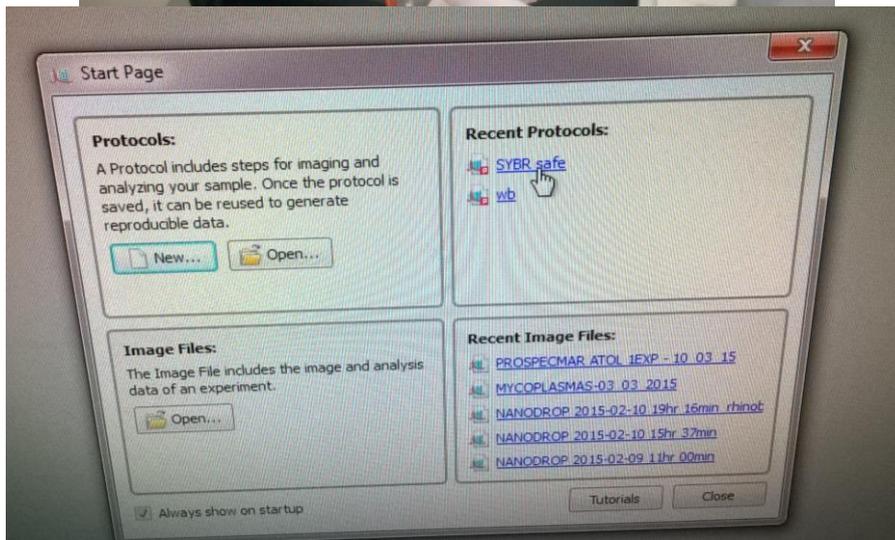
7 – Imediatamente retornar as amostras ao armazenamento em **-80°C**.

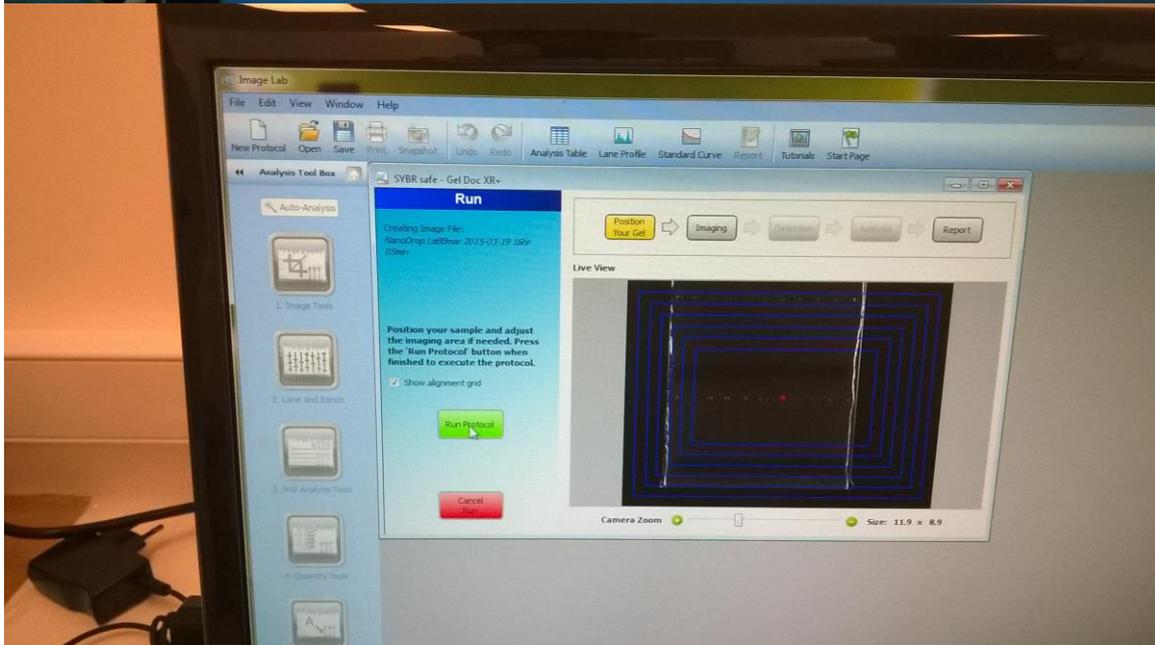
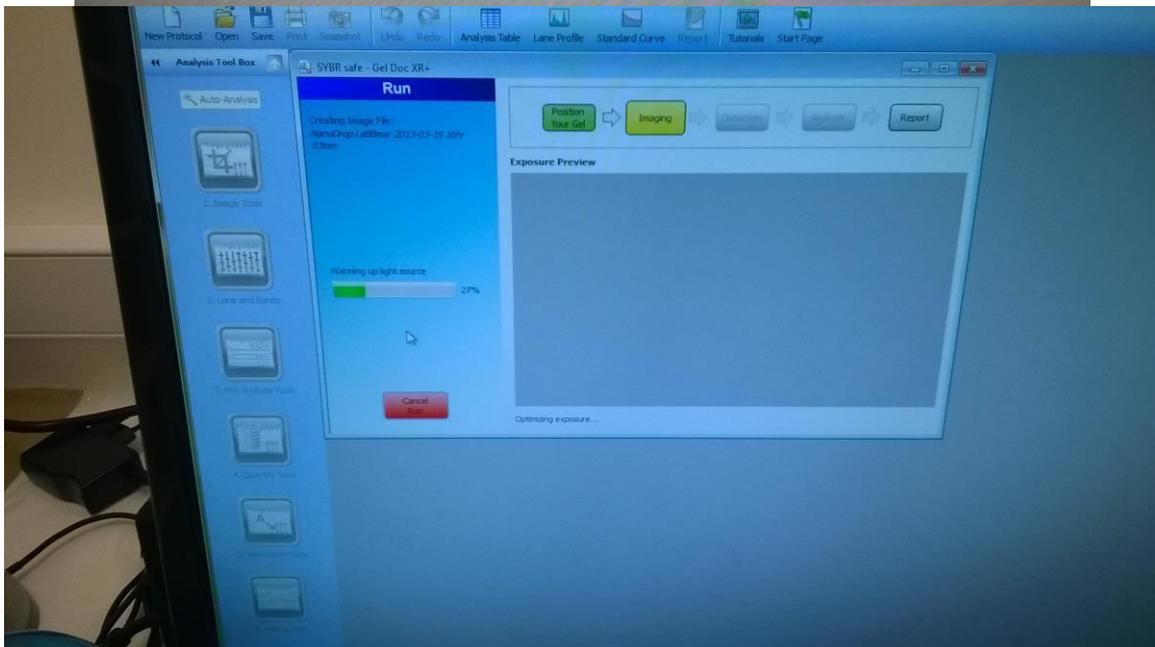
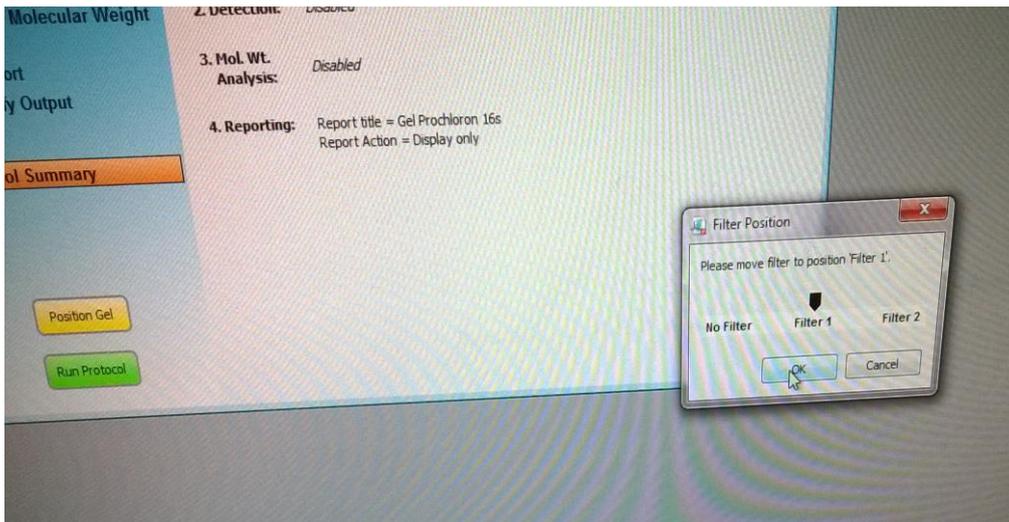
8 – Após as amostras estarem todas no gel, iniciar a **corrida**, com **50min** a uma voltagem de **80V** e uma amperagem de **100mA**. A corrida ocorre sempre do polo negativo (geralmente eletrodo preto) para o polo positivo (geralmente eletrodo vermelho), portanto os poços têm de estar próximos ao polo negativo. Como a polaridade do DNA é negativa este acompanha o sentido da corrente (imagem abaixo).

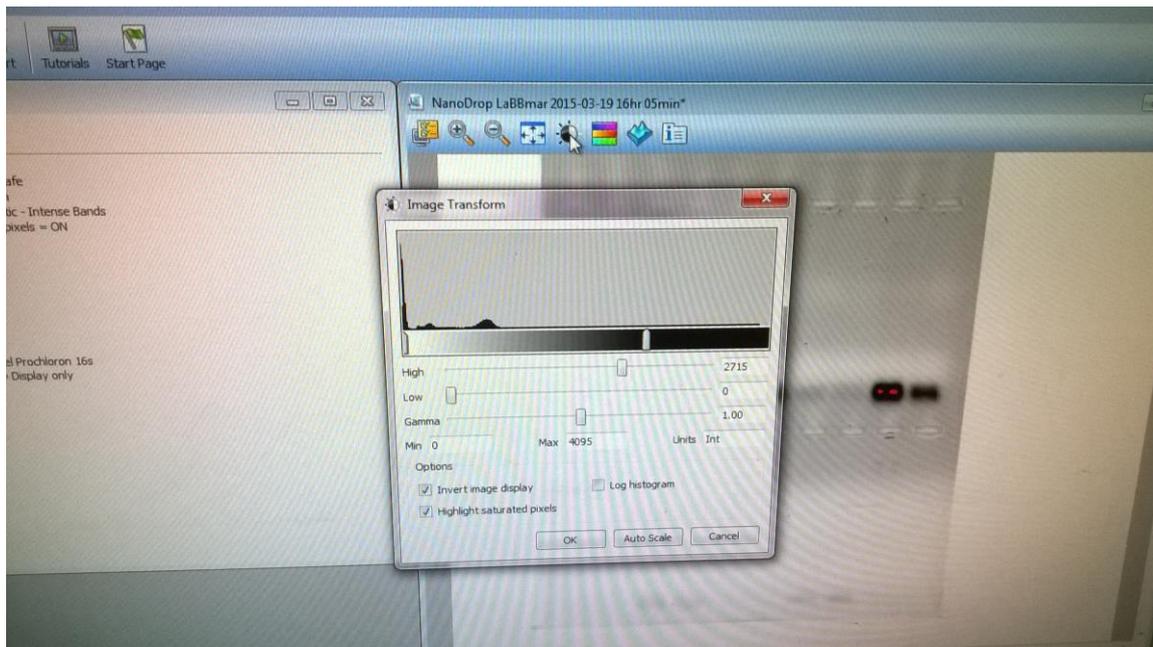
*Obs8.: Usar novamente o revestimento de papel alumínio para proteger o sistema da luz durante a corrida.*



9 – A fotodocumentação do gel após a corrida é realizada pelo programa Image Lab (isso no laboratório LaBBmar). Para tal basta colocar o gel dentro do fotodocumentador e operar o programa. As fotos em sequência abaixo indicam como operar o programa.







10 –Após, descartar em local adequado (Brometo de etídio é altamente carcinogênico) o gel e limpar o fotodocumentador. Importar as fotos do gel para o Power Point.

## PROTOCOLO DE PREPARAÇÃO DA BIBLIOTECA rRNA 16S - PLATAFORMA MiSeq ILLUMINA (ILLUMINA 2013; BEZERRA, 2015)

### Passo-a-passo - Biblioteca rRNA 16S

#### 1-PCR dos amplicons

. Primeira etapa de PCR, que amplifica o fragmento de interesse e adiciona às extremidades do amplicon a sequência (tag) complementar aos identificadores (indexes).

#### Consumíveis

DNA metagenômico (5 a 10 ng/uL)

Iniciadores específicos – Forward e Reverse(10 uM)

DNA polimerase (preferencialmente High Fidelity), junto com o tampão apropriado

Água para PCR

Plásticos (ponteiras, tubos ou placa para PCR)

#### Reação

. Para cada enzima que for utilizar, ajustar a reação para que os constituintes fiquem nas condições ótimas para o melhor rendimento

. Para reações com os primers 515F/806R, 10 ng de DNA por reação é suficiente; para os primers de Eucarioto (EUKFor/EUK-BR), pode ser necessário aumentar para 20 ng

#### Ciclagem

. As temperaturas ótimas para desnaturação e extensão devem ser ajustadas de acordo com a polimerase usada

#### 2-PCR clean up

. Purificação dos produtos da PCR para posterior indexação

#### Materiais

Solução de Tris 10 mM, pH 8,5	52,5 uL/amostra
AMPure XP beads	20 uL/amostra
Etanol 80%, recém preparado	400 uL/amostra
Placa de 96 poços de fundo cônico	2
Adesivo para placa	
Suporte magnético para placa	
Plásticos (ponteiras)	8 caixas de P200/placa
Pipetas multicanal (P20, P200)	

#### Preparação

. Retirar a solução de beads magnéticos da geladeira e deixar equilibrar com a temperatura ambiente

#### Procedimento

- 1 – Centrifugar a placa (spin) após o PCR para coletar a condensação, e remover o adesivo com cuidado
- 2 – Transferir 25 uL da reação para uma nova placa, tomando o cuidado de trocar as ponteiras a cada amostra
- 3 – Agitar a solução de beads magnéticos (VORTEX) por 30 seg, para a completa homogeneização dos beads
- 4 – Adicionar 20 uL da solução de beads magnéticos em cada amostra, e pipetar gentilmente 10x para homogeneização
- 5 – Incubar em temperatura ambiente sem agitação, por 5 min
- 6 – Acoplar a placa no suporte magnético e deixar por 2 min, sem agitação, ou até que o sobrenadante esteja límpido
- 7 – Sem retirar a placa do suporte, remover e descartar o sobrenadante utilizando pipeta multicanal
- 8 – Sem retirar a placa do suporte, lavar os beads com 200 uL de etanol 80%
- 9 – Sem retirar a placa do suporte, remover e descartar o etanol utilizando pipeta multicanal
- 10 – Repetir os passos 8 e 9
- 11 – Se necessário, remover o excesso de etanol utilizando P20 (pipetar com ponteiras finas)
- 12 – Deixar a placa no suporte em repouso por 10 min, para completa evaporação do etanol
- 13 – Retirar a placa do suporte e adicionar 52,5 uL de Tris 10 mM pH 8,5 em cada amostra, e pipetar gentilmente 10x para homogeneização

- 14 – Incubar em temperatura ambiente por 2 min
- 15 – Acoplar a placa no suporte magnético e deixar por 2 min, sem agitação, ou até que o sobrenadante esteja límpido
- 16 – Transferir 50 uL do sobrenadante para uma nova placa.

Pode ser **STOP POINT!!!**

. O produto de PCR purificado pode ser armazenado em freezer (-15 a -20 °C) por até uma semana

### 3-Indexação

. Esse passo adiciona os indexes (ou identificadores) que serão utilizados para posterior recuperação das sequências provenientes de cada amostra

#### Consumíveis

Produto do PCR purificado

Nextera XT Indexes (N7xx e S5xx)

DNA polimerase (preferencialmente High Fidelity), junto com o tampão apropriado

Água para PCR

Plásticos (ponteiros, tubos ou placa para PCR)

#### Reação

. Para cada enzima que for utilizar, ajustar a reação para que os constituintes fiquem nas condições ótimas para o melhor rendimento

. Utilizar o suporte de placas TruSeq (Illumina) para auxiliar a distribuição dos indexes

#### Ciclagem

. As temperaturas ótimas para desnaturação e extensão devem ser ajustadas de acordo com a polimerase usada

### 4-Clean up

. Purificação dos produtos da indexação

#### Materiais

Solução de Tris 10 mM, pH 8,5	27,5 uL/amostra
AMPure XP beads	56 uL/amostra
Etanol 80%, recém preparado	400 uL/amostra
Placa de 96 poços de fundo cônico	2
Adesivo para placa	
Suporte magnético para placa	
Plásticos (ponteiros)	8 caixas de P200/placa
Pipetas multicanal (P20, P200)	

#### Preparação

. Retirar a solução de beads magnéticos da geladeira e deixar equilibrar com a temperatura ambiente

#### Procedimento

- 1 – Centrifugar a placa (spin) após o PCR para coletar a condensação, e remover o adesivo com cuidado
- 2 – Agitar a solução de beads magnéticos (VORTEX) por 30 seg, para a completa homogeneização dos beads
- 3 – Adicionar 56 uL da solução de beads magnéticos em cada amostra, e pipetar gentilmente 10x para homogeneização
- 4 – Incubar em temperatura ambiente sem agitação, por 5 min
- 5 – Acoplar a placa no suporte magnético e deixar por 2 min, sem agitação, ou até que o sobrenadante esteja límpido
- 6 – Sem retirar a placa do suporte, remover e descartar o sobrenadante utilizando pipeta multicanal
- 7 – Sem retirar a placa do suporte, lavar os beads com 200 uL de etanol 80%
- 8 – Sem retirar a placa do suporte, remover e descartar o etanol utilizando pipeta multicanal
- 9 – Repetir os passos 7 e 8
- 10 – Se necessário, remover o excesso de etanol utilizando P20 (pipetar com ponteiros finas)
- 11 – Deixar a placa no suporte em repouso por 10 min, para completa evaporação do etanol
- 12 – Retirar a placa do suporte e adicionar 27,5 uL de Tris 10 mM pH 8,5 em cada amostra, e pipetar gentilmente 10x para homogeneização
- 13 – Incubar em temperatura ambiente por 2 min
- 14 – Acoplar a placa no suporte magnético e deixar por 2 min, sem agitação, ou até que o sobrenadante esteja límpido
- 15 – Transferir 25 uL do sobrenadante para uma nova placa.

Pode ser **STOP POINT!!!**

. O produto de PCR indexado purificado pode ser armazenado em freezer (-15 a -20 °C) por até uma semana

### 5-Quantificação

. Realizar a quantificação das bibliotecas purificadas, por fluorimetria, e misturar quantidades equimolares das bibliotecas para sequenciamento

#### Materiais

Kit de quantificação de DNA para Qubit (dsDNA BR Assay para Qubit)

Fluorímetro – Qubit

Tubos para quantificação

Ponteiras e Pipetas

Solução de Tris 10 mM, pH 8,5

Solução de NaOH 0,2N, recém preparada

Solução HT1, Cartucho e flowcell

#### Quantificação - Qubit

1 – Separar a quantidade necessária de tubos de quantificação (2 padrões + 1 para cada amostra) e identificar os tubos com a menor quantidade de informação necessária

2 – Preparar a solução de quantificação, diluindo 1 uL da solução reagente com 199 uL da solução diluente num frasco de plástico (falcon)

3 – Para a leitura dos padrões, misturar 190 uL da solução de quantificação com 10 uL de cada padrão

4 – Homogeneizar com vortex por 3 segundos

5 – Incubar a mistura por 2 min

6 – Realizar a leitura no equipamento

7 – Para a leitura das amostras, misturar 198 uL da solução de quantificação com 2,0 uL de cada amostra

8 – Homogeneizar com vortex por 3 segundos

9 – Incubar a mistura por 2 min

10 – Realizar a leitura no equipamento

#### 6-Normalização

1 – Calcular a concentração do DNA em nM, baseado no tamanho dos amplicons

$$\frac{[\text{conc. em ng/\u00b5l}]}{660 \text{ g/mol} \times \text{m\u00e9dia do tamanho dos fragmentos}} \times 10^6 = [\text{conc. em nM}]$$

. Para bibliotecas de V4 de procariotos (515F/806R) – tamanho m\u00e9dio dos fragmentos = 300 pb

. Para bibliotecas de V9 de eucariotos (EUK F/EUK BR) – tamanho m\u00e9dio dos fragmentos = 200 pb

. Considerar o tamanho medio do seu fragmento, confirmar o tamanho medio com um gel de agarose \u00e9 o mais apropriado

2 – Diluir as bibliotecas purificadas para uma concentra\u00e7\u00e3o final de 4 nM com solu\u00e7\u00e3o de Tris 10 mM, pH 8,5

#### 7-Pooling

1 – Juntar 5,0 uL de cada biblioteca normalizada num \u00fanico eppendorf e homogeneizar

2 – Num novo eppendorf, adicionar 5,0 uL da mistura das bibliotecas a 5,0 uL de NaOH 0,2 N

3 – Homogeneizar e incubar em temperatura ambiente por 5 min

4 – Adicionar 990 uL da solu\u00e7\u00e3o HT1

5 – Manter no gelo at\u00e9 o uso

. Ap\u00f3s essa dilui\u00e7\u00e3o, o pool das bibliotecas estar\u00e1 numa concentra\u00e7\u00e3o de 20 pM, o sequenciamento deve ser realizado com o pool entre 10 e 17 pM

. Para 10 pM, misturar 300 uL do pool de bibliotecas e 300 uL de solu\u00e7\u00e3o HT1

. Para 12 pM, misturar 360 uL do pool de bibliotecas e 240 uL de solu\u00e7\u00e3o HT1

. Para 17 pM, misturar 510 uL do pool de bibliotecas e 90 uL de solu\u00e7\u00e3o HT1

#### 8-Prepara\u00e7\u00e3o do PhiX (controle positivo do sequenciamento)

PhiX (10 nM) 2,0 uL

Tris 10 mM pH 8,5 3,0 uL

NaOH 0,2N 5,0 uL

1 – Homogeneizar com vortex, 3 segundos

2 – Incubar em temperatura ambiente por 5 min

3 – Adicionar 990 uL de solu\u00e7\u00e3o HT1

4 – Mater no gelo at\u00e9 o uso

. Após essa diluição, a solução de PhiX estará numa concentração de 20 pM, o sequenciamento deve ser realizado com essa solução entre 10 e 17 pM (depende da concentração do pool das bibliotecas)

. Para 10 pM, misturar 300 uL da solução de PhiX e 300 uL de solução HT1

. Para 12 pM, misturar 360 uL da solução de PhiX e 240 uL de solução HT1

. Para 17 pM, misturar 510 uL da solução de PhiX e 90 uL de solução HT1

#### **9-Desnaturação das bibliotecas e aplicação no cartucho**

. Com as bibliotecas já equilibradas para a concentração ideal, dosar a quantidade de controle positivo que será utilizada no sequenciamento

. Para bibliotecas de baixa diversidade (como bibliotecas de amplicons) a Illumina recomenda 20% de PhiX; o EMP recomenda entre 10-15%.

Para 20% de spike in:

Pool das bibliotecas                    480 uL

PhiX    120 uL

1 –Incubar a mistura a 96 °C por 2 min

2 –Homogeneizar por inversão do tubo 3x

3 –Imediatamente depois da inversão, incubar em banho de gelo por 5 min

4 –Imediatamente após o banho de gelo, carregar a solução no cartucho de sequenciamento

#### **10-Sequenciar plataforma MiSeq**

. Inserir o pool de amostras no cartucho Illumina Nextera XT V3 e sequenciar

**PIPELINE DAS ANÁLISES PELO SOFTWARE QIIME v. 1.9 (<http://qiime.org/scripts/>)  
(PYLRO et al., 2014)**

```
#####acionando o qiime no directorio

source activate qiime1

print_qiime_config.py -t

conda install -c bioconda seqprep

#####baixa o usearch, coloca no directorio e executa a permissao

chmod +x /home/bif/Daku/usearch7.0.1090_i86linux32

#####testa seu mapping file

validate_mapping_file.py -m mapping_file_Daku -o mapping_output

#####os passos abaixo sao exemplos para cada amostra, portanto para cada amostra realizar uma linha de comando. Uma
programação em Phyton resolveria este problema criando um looping de comandos

#####colocar todos os arquivos bmp dentro do directorio de trabalho

SeqPrep -f ../AC1FNA_R1_001.fastq -r ../AC1FNA_R2_001.fastq -1 ../1.seqprep/AC1FNA_R1_001_output.fastq.gz -2
../1.1.seqprep/AC1FNA_R2_001_output.fastq.gz -s ../1.seqprep/AC1FNA_merged.fastq.gz -A
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGT -B GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

mkdir ../2.trim_galore
trim_galore --stringency 5 --nextera --fastqc --length 20 ../1.seqprep/AC1FNA_merged.fastq.gz -o ../2.trim_galore

split_libraries_fastq.py -i ../2.trim_galore/AC1FNA_merged_trimmed.fq.gz -o ../3.demulti/AC1FNA -m mapping_file_Daku.txt --sample_ids
AC1FNA --barcode_type 'not-barcoded' --store_demultiplexed_fastq -q 20 -r 0

mkdir ../4.filterUsearch
/home/bif/Daku/usearch7.0.1090_i86linux32 -fastq_filter ../3.demulti/AC1FNA/seqs.fastq -fastq_maxee 0.8 -fastq_truncLen 240 -fastaout
../4.filterUsearch/AC1FNA.fasta

mkdir ../5.qiime/
cat ../4.filterUsearch/*.fasta > ../5.qiime/cyano.fasta

count_seqs.py -i ../5.qiime/cyano.fasta

#####apos concatenar todas as amostras em apenas um arquivo fasta, iniciam as linhas de comando do qiime

perl /home/bif/Daku/bmp-Qiime2Uparse.pl -i cyano.fasta -o reads_uparse.fa

/home/bif/Daku/usearch7.0.1090_i86linux32 -derep_fulllength reads_uparse.fa -output cyano_derep.fasta -sizeout

/home/bif/Daku/usearch7.0.1090_i86linux32 -sortbysize cyano_derep.fasta -output cyano_sorted.fa -minsize 2

/home/bif/Daku/usearch7.0.1090_i86linux32 -cluster_otus cyano_sorted.fa -otus cyano_otus1.fa

/home/bif/Daku/usearch7.0.1090_i86linux32 -uchime_ref cyano_otus1.fa -db /home/bif/Daku/gold.fa -strand plus -nonchimeras cyano_otus2.fa

fasta_formatter -i cyano_otus2.fa -o cyano_formated_otus2.fa

perl /home/bif/Daku/bmp-otuName.pl -i cyano_formated_otus2.fa -o cyano_otus.fa

/home/bif/Daku/usearch7.0.1090_i86linux32 -usearch_global cyano.fasta -db cyano_otus.fa -strand plus -id 0.99 -uc cyano_map.uc

#####alinhamento com o banco genomico

assign_taxonomy.py -i cyano_otus.fa -o cyano_assign -r /home/bif/Daku/SILVA_128_QIIME_release/rep_set/99_otus_16S.fasta -t
/home/bif/Daku/SILVA_128_QIIME_release/taxonomy/16S_only/99/consensus_taxonomy_all_levels.txt

align_seqs.py -i cyano_otus.fa -o cyano_rep_set_align -t /home/bif/Daku/SILVA_128_QIIME_release/rep_set_aligned/99_otus_aligned.fasta

filter_alignment.py -i cyano_rep_set_align/cyano_otus_aligned.fasta -o cyano_filtered_alignment
```

```
make_phylogeny.py -i cyano_filtered_alignment/cyano_otus_aligned_pfiltered.fasta -o cyano_rep_set.tre
python /home/bif/Daku/bmp-map2qiime.py cyano_map.uc > cyano_otu_table.txt
make_otu_table.py -i cyano_otu_table.txt -t cyano_assign/cyano_otus_tax_assignments.txt -o cyano_otu_table.biom
biom summarize-table -i cyano_otu_table.biom -o cyano_results_biom_table_quantitativo
biom summarize-table -i cyano_otu_table.biom --qualitative -o cyano_results_biom_table_qualitativo
core_diversity_analyses.py -i cyano_otu_table.biom -m mapping_file_Daku.txt -t cyano_rep_set.tre -e 23256 -c Replicate -o cyano_core_output
-aO 35
core_diversity_analyses.py -i cyano_otu_table.biom -m mapping_file_Daku.txt -t cyano_rep_set.tre -e 23256 -c Location -o cyano_core_output
-aO 35
#core_diversity_analyses.py -i cyano_otu_table.biom -m mapping_file_Daku.txt -t cyano_rep_set.tre -e 23256 -c Replicate,Location -o
cyano_core_output -aO 35
beta_diversity.py -i cyano_otu_table.biom -m euclidean -o beta_div
upgma_cluster.py -i beta_div/euclidean_cyano_otu_table.txt -o beta_div_cluster.tre
#####desacionando o qiime no directorio
source deactivate
```

## SCRIPTS DAS ANÁLISES PELO SOFTWARE RSTUDIO x86\_64-apple-darwin15.6.0

## Pacotes utilizados ao longo das análises

```

library(vegan)
library(lattice)
library(outliers)
library(BiodiversityR)
library(ade4)
library(MASS)
library(cluster)
library(gclus)
library(graphics)
library(labdsv)
library(asbio)
library(moments)
library(car)
library(biotools)
library(usdm)
library(VGAM)
library(lawstat)
library(dplyr)
library(dendextend)
library(ggrepel)
library(abind)
library(ca)
library(gplots)
#os pacotes abaixo necessitam de versoes anteriores especificas para rodar no script ajustado abaixo, assim nao serve as atualizacoes atuais que sao baixadas
automaticamente pelo comando "install.packages", portanto procuramos no CRAN e baixamos versoes especificas anteriores nos sites descritos (.zip)
library(FactoMineR) #versao 1.33 # https://cran.r-project.org/src/contrib/Archive/FactoMineR/
library(devtools) #versao 1.12.0 #https://cran.r-project.org/src/contrib/Archive/devtools/
library(factoextra) #versao 1.0.3 #https://cran.r-project.org/src/contrib/Archive/factoextra/
library(ggplot2) #versao 2.2.1 #https://cran.r-project.org/web/packages/ggplot2/index.html
library(fossil)
library(corrplot)
library(Heatplus)
source("http://bioconductor.org/biocLite.R")
library(RColorBrewer)

```

## POC técnica de digestão ácida

```

POC <- read.table("~/Documents/R operacional/POC.txt", header = T, dec=',')

# paso 1 - entendendo a disposição dos dados:
View(POC) #lendo no RStudio
POC #lendo no console
attach(POC) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(POC) #resumo pra saber as significancias e valores da planilha
attributes(POC) #mostra os criterios

hist(POC$mg) #funcao histograma para a distribuicao, da variavel em funcao da planilha
#nao normal

#Associar duas variaveis a um parametro
tapply(POC$mg, POC$Samples, mean) #separando as medias All pela variavel samples
MedSamples<-tapply(POC$mg, POC$Samples, mean) #nomeando
MedSamples #vendo o arquivo

tapply(POC$mg, POC$Samples, sd) #separando os desvios All pela variavel samples
SdPOC<-tapply(POC$mg, POC$Samples, sd) #nomeando
SdPOC #vendo o arquivo

#graficos de media com desvio padrao, funcao BOXPLOT
boxplot(POC$mg~POC$Samples) #da um ~ com um espaco, eh um comando que da a media e o desvio de uma variavel por outra variavel, o ~ separa os
dados

shapiro.test(POC$mg) #teste de normalidade p<0,05 para ser normal
#nao deu normal (tem de transformar em log, e agora usar essa planilha transformada pra tudo

POCed <- POC[-16,]
View(POCed)
shapiro.test(POCed$mg) #teste de normalidade p<0,05 para ser normal

```

```

POClog<-log(POCed$mg)#dados de tam em log para ficarem normal
View(POClog)
shapiro.test(POClog)

POCstand <- decostand(POCed$mg, "standardize")#dados de tam em log para ficarem normal
View(POCstand)
shapiro.test(POCstand)

#ANOVA unifactorial
POCANOVA<-aov(POCstand~POCed$Samples, POCed) #fit e o nome que eu dei para aov (analysis of variance) do que eu quero
POCANOVA
anova(POCANOVA) #Análise de variancia, probabilidades
TukeyHSD(POCANOVA) #Teste Tukey, teste de tratamentos pareados a posteriori
plot(POCANOVA)

# paso 2 - matriz de dissimilaridade (agrupamento) - distancia euclidiana:
POCeU<-vegdist(POCstand, method='eu', binary=F, diag = F, upper = F) #apenas dei um nome, indice de similaridade (distância euclidiana pura para análises
quantitativas)
POCeU
POCeUcluster<-hclust(POCeU, method = "average") # nomeando e gerando a analise de cluster
plot(POCeUcluster, hang = -1) #gera o grafico de cluster
POCeUcluster

#Correlação linear - POC e POC remote
POCremote <- read.table("~/Documents/R operacional/POCremote.txt", header = T, dec=',')

POCremoteed.0 <- POCremote[1:16,]
POCremoteed.0
POCremoteed <- POCremoteed.0[-5,]
POCremoteed

POCremotestand <- decostand(POCremoteed$mg, "standardize")#dados de tam em log para ficarem normal
View(POCremotestand)
shapiro.test(POCremotestand)

POCstand
POCstanded <- POCstand[11:25,] # deixa apenas a partir da linha 11, para ficar com os mesmos pontos de POC remote
View(POCstanded)

poccor <- cor.test(POCremotestand, POCstanded)
poccor

```

## POC remote sensoriamento remoto

```

POCremote <- read.table("~/Documents/R operacional/POCremote.txt", header = T, dec=',')

# paso 1 - entendendo a disposição dos dados:
View(POCremote) #lendo no RStudio
POCremote #lendo no console
attach(POCremote) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(POCremote) #resumo pra saber as significancias e valores da planilha
attributes(POCremote) #mostra os criterios

hist(POCremote$mg) #funcao histograma para a distribuicao, da variavel em funcao da planilha
#nao normal

#Associar duas variaveis a um parametro
tapply(POCremote$mg, POCremote$Samples, mean) #separando as medias All pela variavel samples
MedSamples<-tapply(POCremote$mg, POCremote$Samples, mean) #nomeando
MedSamples #vendo o arquivo

tapply(POCremote$mg, POCremote$Samples, sd) #separando os desvios All pela variavel samples
SdPOC<-tapply(POCremote$mg, POCremote$Samples, sd) #nomeando
SdPOC #vendo o arquivo

#graficos de media com desvio padrao, funcao BOXPLOT
boxplot(POCremote$mg~POCremote$Samples) #da um ~ com um espaco, eh um comando que da a media e o desvio de uma variavel por outra variavel, o ~
separa os dados

shapiro.test(POCremote$mg) #teste de normalidade p<0,05 para ser normal
#sim deu normal (normal caso nao tem de transformar em log, e agora usar essa planilha transformada pra tudo
#POClog<-log(POCremote$mg)#dados de tam em log para ficarem normal
#View(POCremotelog)
#shapiro.test(POCremotelog)

POCstand <- decostand(POCremote$mg, "standardize")#dados de tam em log para ficarem normal
View(POCstand)

```

```

shapiro.test(POCstand)

#ANOVA unifactorial
POCANOVA<-aov(POCstand~POCremote$Samples) #fit e o nome que eu dei para aov (analysis of variance) do que eu quero
POCANOVA
anova(POCANOVA) #Análise de variancia, probabilidades
TukeyHSD(POCANOVA) #Teste
plot(POCANOVA)

# passo 2 - matriz de dissimilaridade (agrupamento) - distancia euclidiana:
POCremote
POCEu<-vegdist(POCstand, method='eu', binary=F, diag = F, upper = F) #apenas dei um nome, indice de similaridade (distância euclidiana pura para analises
quantitativas)
POCEu
POCEucluster<-hclust(POCEu, method = "average") # nomeando e gerando a analise de cluster
plot(POCEucluster, hang = -1) #gera o grafico de cluster
POCEucluster

% % POC
clear
clc
close all

% % % % % % Figura 1
% % % % % % Grande 2 meses
ncdisp('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20150901-20151031.55W_10S_15E_5N.nc')
poc=double(ncread('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20150901-20151031.55W_10S_15E_5N.nc','MODISA_L3m_POC_2014_poc'));
lon=double(ncread('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20150901-20151031.55W_10S_15E_5N.nc','lon'));
lat=double(ncread('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20150901-20151031.55W_10S_15E_5N.nc','lat'));
[Lat,Lon] = meshgrid(lat,lon);
flag=max(max(poc));
id=find(poc==flag);
poc(id)=nan;

figure(1)
colormap('jet')
set(gca,'fontsize',12)
[mlon,mlat]=meshgrid(lon,lat);% usa o "" para transpor a matriz e alinhar ela corretamente
m_proj('mercator','lon',[min(lon) max(lon)],'lat',[min(lat) max(lat)])
m_pcolor(mlon,mlat,log(poc))% nao usa log pois eh temperatura
shading interp
m_gshhs_f('patch',[0.5 0.5 0.5])
m_grid('box','fancy','fontsize',12)
c=colorbar
ylabel(c,'log(POC) (mg/m^3)','fontsize',12)
set(c,'fontsize',12)
m_text(-29.33,-9, '\bf \otimes','fontsize',11,...
'HorizontalAlignment','center','color','r');
m_text(-29.33,.2, '\bf ASPSP','fontsize',12,...
'HorizontalAlignment','center');
m_text(-32.4,-3.86, '\bf \otimes','fontsize',11,...
'HorizontalAlignment','center','color','r');
m_text(-32,-4.8, '\bf FN','fontsize',12,...
'HorizontalAlignment','center');
m_text(-33.8,-3.85, '\bf \otimes','fontsize',11,...
'HorizontalAlignment','center','color','r');
m_text(-34,-3, '\bf AR','fontsize',12,...
'HorizontalAlignment','center','color','k');
set(gca,'fontsize',12)
caxis([3 7])

%definição da área
latid_fn_fig1 = find(lat<-3.8 & lat>-3.9);
lonid_fn_fig1 = find(lon<-32.45 & lon>-32.55);
poc_fn_fig1 = poc(lonid_fn_fig1,latid_fn_fig1);
lat_fn_fig1 = lat(latid_fn_fig1);
lon_fn_fig1 = lon(lonid_fn_fig1);
latid_aspfig1 = find(lat>0.87 & lat<0.97);
lonid_aspfig1 = find(lon<-29.25 & lon>-29.35);
poc_aspfig1 = poc(lonid_aspfig1,latid_aspfig1);
lat_aspfig1 = lat(latid_aspfig1);
lon_aspfig1 = lon(lonid_aspfig1);
latid_at_fig1 = find(lat>-3.90 & lat<-3.80);
lonid_at_fig1 = find(lon>-33.85 & lon<-33.75);
poc_at_fig1 = poc(lonid_at_fig1,latid_at_fig1);
lat_at_fig1 = lat(latid_at_fig1);
lon_at_fig1 = lon(lonid_at_fig1);

```

Figuras de POC remote (MATLAB v. 9.1)





```

%% % % % % % % % % Figura 5
%% % % % % % % % % ASPSP (setembro)
clear
clc

% close all
ncdisp('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20150901-20150930.30W_0S_28W_2N.nc')
poc=double(ncread('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20150901-20150930.30W_0S_28W_2N.nc','MODISA_L3m_POC_2014_poc'));
lon=double(ncread('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20150901-20150930.30W_0S_28W_2N.nc','lon'));
lat=double(ncread('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20150901-20150930.30W_0S_28W_2N.nc','lat'));
[Lat,Lon] = meshgrid(lat,lon);
flag=max(max(poc));
id=find(poc==flag);
poc(id)=nan;

figure(5)
colormap('jet')
set(gca,'fontsize',12)
[mlon,mlat]=meshgrid(lon,lat);
m_proj('mercator','lon',[min(lon) max(lon)],'lat',[min(lat) max(lat)])
m_pcolor(mlon,mlat,log(poc))
shading interp
m_gshhs_f('patch',[0.5 0.5 0.5])
m_grid('box','fancy','fontsize',12)
c=colorbar
ylabel(c,'log(POC) (mg/m^3)','fontsize',12)
set(c,'fontsize',12)
m_text(-29.33,.917, '\bf \otimes','fontsize',11,...
    'HorizontalAlignment', 'center','color','r');
m_text(-29.33,1, '\bf ASPSP','fontsize',12,...
    'HorizontalAlignment', 'center');
caxis([3 7])

%% % % % % % % % % Figura 6
%% % % % % % % % % Grande 1 ano
clear
clc

% close all
ncdisp('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20141101-20151031.55W_10S_15E_5N.nc')
poc=double(ncread('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20141101-20151031.55W_10S_15E_5N.nc','MODISA_L3m_POC_2014_poc'));
lon=double(ncread('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20141101-20151031.55W_10S_15E_5N.nc','lon'));
lat=double(ncread('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20141101-20151031.55W_10S_15E_5N.nc','lat'));
[Lat,Lon] = meshgrid(lat,lon);
flag=max(max(poc));
id=find(poc==flag);
poc(id)=nan;

figure(6)
colormap('jet')
set(gca,'fontsize',12)
[mlon,mlat]=meshgrid(lon,lat);
m_proj('mercator','lon',[min(lon) max(lon)],'lat',[min(lat) max(lat)])
m_pcolor(mlon,mlat,log(poc))
shading interp
m_gshhs_f('patch',[0.5 0.5 0.5])
m_grid('box','fancy','fontsize',12)
c=colorbar
ylabel(c,'log(POC) (mg/m^3)','fontsize',12)
set(c,'fontsize',12)
m_text(-29.33,-.9, '\bf \otimes','fontsize',11,...
    'HorizontalAlignment', 'center','color','r');
m_text(-29.33,.2, '\bf ASPSP','fontsize',12,...
    'HorizontalAlignment', 'center');
m_text(-32.4,-3.86, '\bf \otimes','fontsize',11,...
    'HorizontalAlignment', 'center','color','r');
m_text(-32,-4.8, '\bf FN','fontsize',12,...
    'HorizontalAlignment', 'center');
m_text(-33.8,-3.85, '\bf \otimes','fontsize',11,...
    'HorizontalAlignment', 'center','color','r');
m_text(-34,-3, '\bf AR','fontsize',12,...
    'HorizontalAlignment', 'center','color','k');
set(gca,'fontsize',12)
caxis([3 7])

%definiÃ, o da rea
latid_fn_fig6 = find(lat<-3.8 & lat>-3.9);
lonid_fn_fig6 = find(lon<-32.45 & lon>-32.55);

```

```

poc_fn_fig6 = poc(lonid_fn_fig6,latid_fn_fig6);
lat_fn_fig6 = lat(latid_fn_fig6);
lon_fn_fig6 = lon(lonid_fn_fig6);
latid_aspfig6 = find(lat>0.87 & lat<0.97);
lonid_aspfig6 = find(lon<-29.25 & lon>-29.35);
poc_aspfig6 = poc(lonid_aspfig6,latid_aspfig6);
lat_aspfig6 = lat(latid_aspfig6);
lon_aspfig6 = lon(lonid_aspfig6);
latid_at_fig6 = find(lat>-3.90 & lat<-3.80);
lonid_at_fig6 = find(lon>-33.85 & lon<-33.75);
poc_at_fig6 = poc(lonid_at_fig6,latid_at_fig6);
lat_at_fig6 = lat(latid_at_fig6);
lon_at_fig6 = lon(lonid_at_fig6);

%% % % % % % % % % Figura 7
%% % % % % % % % % Grande 5 anos
clear
clc

% close all
ncdisp('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20101101-20151031.55W_10S_15E_5N.nc')
poc=double(ncread('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20101101-20151031.55W_10S_15E_5N.nc','MODISA_L3m_POC_2014_poc'));
lon=double(ncread('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20101101-20151031.55W_10S_15E_5N.nc','lon'));
lat=double(ncread('g4.timeAvgMap.MODISA_L3m_POC_2014_poc.20101101-20151031.55W_10S_15E_5N.nc','lat'));
[Lat,Lon] = meshgrid(lat,lon);
flag=max(max(poc));
id=find(poc==flag);
poc(id)=nan;

figure(7)
colormap('jet')
set(gca,'fontsize',12)
[mlon,mlat]=meshgrid(lon,lat);
m_proj('mercator','lon',[min(lon) max(lon)],'lat',[min(lat) max(lat)])
m_pcolor(mlon,mlat,log(poc))
shading interp
m_gshhs_f('patch',[0.5 0.5 0.5])
m_grid('box','fancy','fontsize',12)
c=colorbar
ylabel(c,'log(POC) (mg/m^3)','fontsize',12)
set(c,'fontsize',12)
m_text(-29.33,-.9, '\bf \otimes','fontsize',11,...
'HorizontalAlignment','center','color','r');
m_text(-29.33,.2, '\bf ASPSP','fontsize',12,...
'HorizontalAlignment','center');
m_text(-32.4,-3.86, '\bf \otimes','fontsize',11,...
'HorizontalAlignment','center','color','r');
m_text(-32,-4.8, '\bf FN','fontsize',12,...
'HorizontalAlignment','center');
m_text(-33.8,-3.85, '\bf \otimes','fontsize',11,...
'HorizontalAlignment','center','color','r');
m_text(-34,-3, '\bf AR','fontsize',12,...
'HorizontalAlignment','center','color','k');
set(gca,'fontsize',12)
caxis([3 7])

%definiÁ, o da área
latid_fn_fig7 = find(lat<-3.8 & lat>-3.9);
lonid_fn_fig7 = find(lon<-32.45 & lon>-32.55);
poc_fn_fig7 = poc(lonid_fn_fig7,latid_fn_fig7);
lat_fn_fig7 = lat(latid_fn_fig7);
lon_fn_fig7 = lon(lonid_fn_fig7);
latid_aspfig7 = find(lat>0.87 & lat<0.97);
lonid_aspfig7 = find(lon<-29.25 & lon>-29.35);
poc_aspfig7 = poc(lonid_aspfig7,latid_aspfig7);
lat_aspfig7 = lat(latid_aspfig7);
lon_aspfig7 = lon(lonid_aspfig7);
latid_at_fig7 = find(lat>-3.90 & lat<-3.80);
lonid_at_fig7 = find(lon>-33.85 & lon<-33.75);
poc_at_fig7 = poc(lonid_at_fig7,latid_at_fig7);
lat_at_fig7 = lat(latid_at_fig7);
lon_at_fig7 = lon(lonid_at_fig7);

```

Citometria das amostras de água

```

cyto <- read.table("~/Documents/R operacional/cytometry.txt", header = T, dec=',')
#ou
cyto<-read.table("~/Documents/R operacional/cytometry.txt", header=T, row.names=1, sep='\t', dec=',')

```

```

# paso 1
#entendendo a disposiçao dos dados:
View(cyto) #lendo no RStudio
cyto #lendo no console
attach(cyto) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(cyto) #resumo pra saber as significancias e valores da planilha
attributes(cyto) #mostra os criterios

#####
#Syneco

hist(cyto$Synecococcus) #funcao histograma para a distribuicao, da variavel em funcao da planilha
#nao normal

histogram(~ Synecococcus|factor(Samples),
  data = cyto,
  layout = c(1, 13),
  nint = 30,
  xlab = list("Cell / L", cex.lab = 1.5),
  strip = FALSE,
  strip.left = TRUE,
  ylab = "Frequencies")

#Associar duas variaveis a um parametro
tapply(cyto$Synecococcus, cyto$Samples, mean) #separando as medias All pela variavel samples
MedAll<-tapply(cyto$Synecococcus, cyto$Samples, mean) #nomeando
MedAll #vendo o arquivo

tapply(cyto$Synecococcus, cyto$Samples, sd) #separando os desvios All pela variavel samples
SdAll<-tapply(cyto$Synecococcus, cyto$Samples, sd) #nomeando
SdAll #vendo o arquivo

#graficos de media com desvio padrao, funcao BOXPLOT
par(mar = c(5,5,2,2), cex.lab = 1.5)
boxplot(cyto$Synecococcus) #media e desvios do tamanho apenas
boxplot(cyto$Synecococcus~cyto$Samples) #da um ~ com um espaco, eh um comando que da a media e o desvio de uma variavel por outra variavel, o ~
separa os dados

shapiro.test(cyto$Synecococcus) #teste de normalidade p<0,05 para ser normal
#deu normal (tem de transformar em log, e agora usar essa planilha transformada pra tudo
#cytolog<-log(cyto$Synecococcus)#dados de tam em log para ficarem normal
#View(cytolog)
#shapiro.test(cytolog)

#ANOVA unifactorial
cytoANOVA<-aov(cyto$Synecococcus~cyto$Samples, cyto) #fit e o nome que eu dei para aov (analysis of variance) do que eu quero
cytoANOVA
anova(cytoANOVA) #An?lise de variancia, probabilidades
TukeyHSD(cytoANOVA) #Teste Tukey
plot(cytoANOVA)
plot(TukeyHSD(cytoANOVA))

#####
#Phlocoo

hist(cyto$Prochlorococcus) #funcao histograma para a distribuicao, da variavel em funcao da planilha
#nao normal

histogram(~ Prochlorococcus|factor(Samples),
  data = cyto,
  layout = c(1, 13),
  nint = 30,
  xlab = list("Cell / L", cex.lab = 1.5),
  strip = FALSE,
  strip.left = TRUE,
  ylab = "Frequencies")

#Associar duas variaveis a um parametro
tapply(cyto$Prochlorococcus, cyto$Samples, mean) #separando as medias All pela variavel samples
MedAll<-tapply(cyto$Prochlorococcus, cyto$Samples, mean) #nomeando
MedAll #vendo o arquivo

tapply(cyto$Prochlorococcus, cyto$Samples, sd) #separando os desvios All pela variavel samples
SdAll<-tapply(cyto$Prochlorococcus, cyto$Samples, sd) #nomeando
SdAll #vendo o arquivo

#graficos de media com desvio padrao, funcao BOXPLOT
par(mar = c(5,5,2,2), cex.lab = 1.5)

```

```

boxplot(cyto$Prochlorococcus) #media e desvios do tamanho apenas
boxplot(cyto$Prochlorococcus~cyto$Samples) #da um ~ com um espaco, eh um comando que da a media e o desvio de uma variavel por outra variavel, o ~
separa os dados

shapiro.test(cyto$Prochlorococcus) #teste de normalidade p<0,05 para ser normal
#nao deu normal (tem de transformar em log, e agora usar essa planilha transformada pra tudo
#cytolog<-log(cyto$Prochlorococcus)#dados de tam em log para ficarem normal
#View(cytolog)
#shapiro.test(cytolog)

cytoed <- cyto[-163,]
cytoed
hist(cytoed$Prochlorococcus)
shapiro.test(cytoed$Prochlorococcus)

cytoedstan.pro <- decostand(cytoed$Prochlorococcus, "standardize")#dados de tam em log para ficarem normal
View(cytoedstan.pro)
shapiro.test(cytoedstan.pro)

#ANOVA unifactorial
cytoANOVA<-aov(cytoedstan.pro~cytoed$Samples, cytoed) #fit e o nome que eu dei para aov (analysis of variance) do que eu quero
cytoANOVA
anova(cytoANOVA) #An?lise de variancia, probabilidades
TukeyHSD(cytoANOVA) #Teste Tukey, teste de tratamentos pareados a posteriori
plot(cytoANOVA)
plot(TukeyHSD(cytoANOVA))

#####
#All
#cyto ajustada - estatisticas:
cyto.mds.0 <- cytoed[,-1]
cyto.mds.0
cyto.mds.1 <- cyto.mds.0[-9,]
cyto.mds.1
cyto.mds.2 <- cyto.mds.1[-12,]
cyto.mds.2
cyto.mds <- cyto.mds.2[-16,]
cyto.mds
shapiro.test(cyto.mds$Synechococcus)

cytostand <- decostand(cyto.mds, "standardize")#dados de tam em log para ficarem normal
View(cytostand)
shapiro.test(cytostand$Synechococcus)

#ANOVA multifatorial

fait<-
aov(cyto.mds$Phyto~cyto.mds$Synechococcus*cyto.mds$Prochlorococcus*cyto.mds$Cyanochlorophyllates*cyto.mds$Nanoekaryotics*cyto.mds$Phyto,
cyto.mds)
anova(fait)
TukeyHSD(fait)

fruit<-
aov(cytostand$Phyto~cytostand$Synechococcus*cytostand$Prochlorococcus*cytostand$Cyanochlorophyllates*cytostand$Nanoekaryotics*cytostand$Phyto,
cytostand)
anova(fruit)
TukeyHSD(fruit)

# paso 2
#PCA 1
pca.cyto.pad<- prcomp(cytostand, center=F) #gera a PCA, method false indica que os dados nao vao ser ajustados por esta funcao, porem como eu ja
padronizei antes, a matriz gerada sera de co-relacao (mesmo eu nao usando o T, pois eu ja padronizei)
pca.cyto.pad
loadings(pca.cyto.pad) #verifica os pesos encontrados
summary(pca.cyto.pad)
attributes(pca.cyto.pad)
biplot(pca.cyto.pad, main='Titulo', xlab='Eixo 1', ylab='Eixo 2', bstick=T) #criterio de selcao de eixos, para saber quais vao ser os criterios usados para fazer a
matriz de agrupamento. Sendo a funcao brokenstick a mais usada, portanto para os eixos onde os autovalores observado (barra do grafico) for maior que o
esperado (pontos ligados), deve-se usar estes eixos apenas. A partir dos eixos escolhidos vc vai la na tabela da pca e ve quais sao os criterios que influenciaim
estes eixos e serao estes criterios que vc vai usar para fazer finalmente a matriz de agrupamento

cyto.mds.3 <- cytoed[-9,]
cyto.mds.3
cyto.mds.4 <- cyto.mds.3[-12,]
cyto.mds.4
cyto.mds.5 <- cyto.mds.4[-16,]
cyto.mds.5

```

```

x1 <- pca.cyto.pad$scores
x1
y1 <- pca.cyto.pad$scores~cyto.mds.5$Samples
y1

fit1 <- envfit(pca.cyto.pad, cytostand, perm = 9999)
fit1

#PCA 2
pca.cyto.pad.comp<- princomp(cytostand, cor=F) #calcula os pesos de cada variavel dos eixos da PCA, usa F pois ja fiz a padronizacao dos dados
anteriormente, caso contrario usaria TRUE
pca.cyto.pad.comp
loadings(pca.cyto.pad.comp) #verifica os pesos encontrados
summary(pca.cyto.pad.comp)
attributes(pca.cyto.pad.comp)
biplot(pca.cyto.pad.comp, main='Titulo', xlab='Eixo 1', ylab='Eixo 2', bstick=T)

x2 <- pca.cyto.pad.comp$scores
x2
y2 <- pca.cyto.pad.comp$scores~cyto.mds.5$Samples
y2

fit2 <- envfit(pca.cyto.pad.comp, cytostand, perm = 9999)
fit2

#PCA 3 - grafico chique que agrupa
iris.pca <- PCA(cytostand, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

fit2 <- envfit(pca.cyto.pad.comp, cytostand, perm = 9999)
fit2

fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=cyto.mds.5$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=cyto.mds.5$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=cyto.mds.5$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(1,3), label="var", habillage=cyto.mds.5$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(2,3), label="var", habillage=cyto.mds.5$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()

#PCA 4 - grafico chique que agrupa sem decostand
iris.pca.1 <- PCA(cyto.mds, scale.unit = FALSE, graph = FALSE)
iris.pca.1
loadings(iris.pca.1) #verifica os pesos encontrados
summary(iris.pca.1)
attributes(iris.pca.1)

pca.cyto.pad.comp.1<- princomp(cyto.mds, cor=F) #calcula os pesos de cada variavel dos eixos da PCA, usa F pois ja fiz a padronizacao dos dados
anteriormente, caso contrario usaria TRUE
pca.cyto.pad.comp.1
loadings(pca.cyto.pad.comp.1) #verifica os pesos encontrados
summary(pca.cyto.pad.comp.1)
attributes(pca.cyto.pad.comp.1)
biplot(pca.cyto.pad.comp.1, main='Titulo', xlab='Eixo 1', ylab='Eixo 2', bstick=T)

x3 <-pca.cyto.pad.comp.1$scores
x3
y3 <- pca.cyto.pad.comp.1$scores~cyto.mds.5$Samples
y3

fit3 <- envfit(pca.cyto.pad.comp.1, cyto.mds, perm = 9999)
fit3

fviz_pca_ind(iris.pca.1, label="none")
fviz_pca_ind(iris.pca.1, label="none", habillage=cyto.mds.5$Samples)
fviz_pca_ind(iris.pca.1, label="none", habillage=cyto.mds.5$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca.1, axes = c(1,2), label="var", habillage=cyto.mds.5$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(1,3), label="var", habillage=cyto.mds.5$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(2,3), label="var", habillage=cyto.mds.5$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()

```

```
# MDS - (distancia euclidiana metrica - abiotico)
cyto.mds.eu<- metaMDS(cytostand, distance='eu', k=2) #ficar em somente em 2 dimensoes
cyto.mds.eu
attributes(cyto.mds.eu)
plot(cyto.mds.eu, cex=3, display=c('sites')) #deixa apenas os pontos
text(cyto.mds.eu, display='sites')
points(cyto.mds.eu$points[1:3,], col=1, cex=3)
points(cyto.mds.eu$points[4,], col=1, cex=3)
points(cyto.mds.eu$points[5:20,], col=2, cex=3)
points(cyto.mds.eu$points[21:40,], col=3, cex=3)
points(cyto.mds.eu$points[41:60,], col=4, cex=3)
points(cyto.mds.eu$points[61:75,], col=5, cex=3)
points(cyto.mds.eu$points[76:90,], col=5, cex=3)
points(cyto.mds.eu$points[91:105,], col=5, cex=3)
points(cyto.mds.eu$points[106:120,], col=5, cex=3)
points(cyto.mds.eu$points[121:135,], col=6, cex=3)
points(cyto.mds.eu$points[136:147,], col=6, cex=3)
points(cyto.mds.eu$points[148:159,], col=6, cex=3)
points(cyto.mds.eu$points[160:170,], col=6, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.0437'))
```

```
#sem decostand
cyto.mds.eu.1<- metaMDS(cyto.mds, distance='eu', k=2) #ficar em somente em 2 dimensoes
cyto.mds.eu.1
attributes(cyto.mds.eu.1)
plot(cyto.mds.eu.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(cyto.mds.eu.1, display='sites')
points(cyto.mds.eu.1$points[1:3,], col=1, cex=3)
points(cyto.mds.eu.1$points[4,], col=1, cex=3)
points(cyto.mds.eu.1$points[5:20,], col=2, cex=3)
points(cyto.mds.eu.1$points[21:40,], col=3, cex=3)
points(cyto.mds.eu.1$points[41:60,], col=4, cex=3)
points(cyto.mds.eu.1$points[61:75,], col=5, cex=3)
points(cyto.mds.eu.1$points[76:90,], col=5, cex=3)
points(cyto.mds.eu.1$points[91:105,], col=5, cex=3)
points(cyto.mds.eu.1$points[106:120,], col=5, cex=3)
points(cyto.mds.eu.1$points[121:135,], col=6, cex=3)
points(cyto.mds.eu.1$points[136:147,], col=6, cex=3)
points(cyto.mds.eu.1$points[148:159,], col=6, cex=3)
points(cyto.mds.eu.1$points[160:170,], col=6, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.0755'))
```

## Biomassa das cianobactérias

```
biomass <- read.table("~/Documents/R operacional/biomass.txt", header = T, dec=',')
```

```
# paso 1
#entendendo a disposição dos dados:
View(biomass) #lendo no RStudio
cyto #lendo no console
attach(biomass) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(biomass) #resumo pra saber as significancias e valores da planilha
attributes(biomass) #mostra os criterios
```

```
#####
#Syneco
```

```
hist(biomass$Synecococcus) #funcao histograma para a distribuicao, da variavel em funcao da planilha
#nao normal
```

```
histogram(~ Synecococcus|factor(Samples),
  data = biomass,
  layout = c(1, 13),
  nint = 30,
  xlab = list("Cell / L", cex.lab = 1.5),
  strip = FALSE,
  strip.left = TRUE,
  ylab = "Frequencias")
```

```
#Associar duas variaveis a um parametro
tapply(biomass$Synecococcus, biomass$Samples, mean) #separando as medias All pela variavel samples
MedAll<-tapply(biomass$Synecococcus, biomass$Samples, mean) #nomeando
MedAll #vendo o arquivo
```

```
tapply(biomass$Synecococcus, biomass$Samples, sd) #separando os desvios All pela variavel samples
SdAll<-tapply(biomass$Synecococcus, biomass$Samples, sd) #nomeando
SdAll #vendo o arquivo
```

```

#graficos de media com desvio padrao, funcao BOXPLOT
par(mar = c(5,5,2,2), cex.lab = 1.5)
boxplot(biomass$Synechococcus) #media e desvios do tamanho apenas
boxplot(biomass$Synechococcus~biomass$Samples) #da um ~ com um espaco, eh um comando que da a media e o desvio de uma variavel por outra variavel,
o ~ separa os dados

shapiro.test(biomass$Synechococcus) #teste de normalidade p<0,05 para ser normal
#deu normal (nao precisa transformar)
#biomassstan.syn <- decostand(biomass$Prochlorococcus, "standardize")#dados de tam em log para ficarem normal
#View(biomassstan.syn)
#shapiro.test(biomassstan.syn)

#ANOVA unifactorial
biomassANOVA<-aov(biomass$Synechococcus~biomass$Samples, biomass) #fit e o nome que eu dei para aov (analysis of variance) do que eu quero
biomassANOVA
anova(biomassANOVA) #An?lise de variancia, probabilidades
TukeyHSD(biomassANOVA) #Teste Tukey
plot(biomassANOVA)
plot(TukeyHSD(biomassANOVA))

#####
#Phlocoro

hist(biomass$Prochlorococcus) #funcao histograma para a distribuicao, da variavel em funcao da planilha
#nao normal

histogram(~ Prochlorococcus|factor(Samples),
  data = biomass,
  layout = c(1, 13),
  nint = 30,
  xlab = list("Cell / L", cex.lab = 1.5),
  strip = FALSE,
  strip.left = TRUE,
  ylab = "Frequencies")

#Associar duas variaveis a um parametro
tapply(biomass$Prochlorococcus, biomass$Samples, mean) #separando as medias All pela variavel samples
MedAll<-tapply(biomass$Prochlorococcus, biomass$Samples, mean) #nomeando
MedAll #vendo o arquivo

tapply(biomass$Prochlorococcus, biomass$Samples, sd) #separando os desvios All pela variavel samples
SdAll<-tapply(biomass$Prochlorococcus, biomass$Samples, sd) #nomeando
SdAll #vendo o arquivo

#graficos de media com desvio padrao, funcao BOXPLOT
par(mar = c(5,5,2,2), cex.lab = 1.5)
boxplot(biomass$Prochlorococcus) #media e desvios do tamanho apenas
boxplot(biomass$Prochlorococcus~biomass$Samples) #da um ~ com um espaco, eh um comando que da a media e o desvio de uma variavel por outra
variavel, o ~ separa os dados

shapiro.test(biomass$Prochlorococcus) #teste de normalidade p<0,05 para ser normal
#deu normal (nao precisa transformar)
#biomassstan.pro <- decostand(biomass$Prochlorococcus, "standardize")#dados de tam em log para ficarem normal
#View(biomassstan.pro)
#shapiro.test(biomassstan.pro)

#ANOVA unifactorial
biomassANOVA<-aov(biomass$Prochlorococcus~biomass$Samples, biomass) #fit e o nome que eu dei para aov (analysis of variance) do que eu quero
biomassANOVA
anova(biomassANOVA) #An?lise de variancia, probabilidades
TukeyHSD(biomassANOVA) #Teste Tukey, teste de tratamentos pareados a posteriori
plot(biomassANOVA)
plot(TukeyHSD(biomassANOVA))

#####
#All
#biomass ajustada - estatisticas:
biomass.ed <- biomass[,-1]
View(biomass.ed)
shapiro.test(biomass.ed$Synechococcus)
shapiro.test(biomass.ed$Prochlorococcus)

biomassstand <- decostand(biomass.ed, "standardize")#dados de tam em log para ficarem normal
View(biomassstand)
shapiro.test(biomassstand$Synechococcus)
shapiro.test(biomassstand$Prochlorococcus)

#ANOVA multifatorial

```

```

fait<-aov(biomass.ed$Prochlorococcus~biomass.ed$Synechococcus*biomass.ed$Prochlorococcus, biomass.ed)
anova(fait)
TukeyHSD(fait)

fruit<-aov(biomassstand$Prochlorococcus~biomassstand$Synechococcus*biomassstand$Prochlorococcus, biomassstand)
anova(fruit)
TukeyHSD(fruit)

# paso 2 - matriz de dissimilaridade (agrupamento) - distancia euclidiana:
biomasseu<-vegdist(biomass.ed, method='eu', binary=F, diag = F, upper = F) #apenas dei um nome, indice de similaridade (distância euclidiana pura para
analises quantitativas)
biomasseu
biomasseucluster<-hclust(biomasseu, method = "average") # nomeando e gerando a analise de cluster
plot(biomasseucluster, hang = -1) #gera o grafico de cluster
biomasseucluster

biomasseu.1<-vegdist(biomassstand, method='eu', binary=F, diag = F, upper = F) #apenas dei um nome, indice de similaridade (distância euclidiana pura para
analises quantitativas)
biomasseu.1
biomasseucluster.1<-hclust(biomasseu.1, method = "average") # nomeando e gerando a analise de cluster
plot(biomasseucluster.1, hang = -1) #gera o grafico de cluster
biomasseucluster.1

#PCA - grafico chique que agrupa sem decostand
iris.pca <- PCA(biomass.ed, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

pca.biomass.ed<- princomp(biomass.ed, cor=F) #calcula os pesos de cada variavel dos eixos da PCA, usa F pois ja fiz a padronizacao dos dados
anteriormente, caso contrario usaria TRUE
pca.biomass.ed
loadings(pca.biomass.ed) #verifica os pesos encontrados
summary(pca.biomass.ed)
attributes(pca.biomass.ed)
biplot(pca.biomass.ed, main='Titulo', xlab='Eixo 1', ylab='Eixo 2', bstick=T)

x3 <-pca.biomass.ed$scores
x3
y3 <- pca.biomass.ed$scores~biomass$Samples
y3

fit3 <- envfit(pca.biomass.ed, biomass.ed, perm = 9999)
fit3

fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=biomass$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=biomass$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=biomass$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5,
pointsize = 4) + theme_minimal()

#PCA stand
iris.pca.1 <- PCA(biomassstand, scale.unit = FALSE, graph = FALSE)
iris.pca.1
loadings(iris.pca.1) #verifica os pesos encontrados
summary(iris.pca.1)
attributes(iris.pca.1)

pca.biomass.ed.1<- princomp(biomassstand, cor=F) #calcula os pesos de cada variavel dos eixos da PCA, usa F pois ja fiz a padronizacao dos dados
anteriormente, caso contrario usaria TRUE
pca.biomass.ed.1
loadings(pca.biomass.ed.1) #verifica os pesos encontrados
summary(pca.biomass.ed.1)
attributes(pca.biomass.ed.1)
biplot(pca.biomass.ed.1, main='Titulo', xlab='Eixo 1', ylab='Eixo 2', bstick=T)

x4 <-pca.biomass.ed.1$scores
x4
y4 <- pca.biomass.ed.1$scores~biomass$Samples
y4

fit4 <- envfit(pca.biomass.ed.1, biomassstand, perm = 9999)
fit4

fviz_pca_ind(iris.pca.1, label="none")
fviz_pca_ind(iris.pca.1, label="none", habillage=biomass$Samples)

```

```
fviz_pca_ind(iris.pca.1, label="none", habillage=biomass$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca.1, axes = c(1,2), label="var", habillage=biomass$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()
```

```
# MDS - (distancia euclidiana métrica - abiotico)
biomass.mds.eu <- metaMDS(biomass.ed, distance='eu', k=2) #ficar em somente em 2 dimensoes
biomass.mds.eu
attributes(biomass.mds.eu)
plot(biomass.mds.eu, cex=3, display=c('sites')) #deixa apenas os pontos
text(biomass.mds.eu, display='sites')
points(biomass.mds.eu$points[1:3,], col=1, cex=3)
points(biomass.mds.eu$points[4,], col=1, cex=3)
points(biomass.mds.eu$points[5:20,], col=2, cex=3)
points(biomass.mds.eu$points[21:40,], col=3, cex=3)
points(biomass.mds.eu$points[41:60,], col=4, cex=3)
points(biomass.mds.eu$points[61:75,], col=5, cex=3)
points(biomass.mds.eu$points[76:90,], col=5, cex=3)
points(biomass.mds.eu$points[91:105,], col=5, cex=3)
points(biomass.mds.eu$points[106:120,], col=5, cex=3)
points(biomass.mds.eu$points[121:135,], col=6, cex=3)
points(biomass.mds.eu$points[136:147,], col=6, cex=3)
points(biomass.mds.eu$points[148:159,], col=6, cex=3)
points(biomass.mds.eu$points[160:174,], col=6, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.0001'))
```

```
#MDS stand
biomass.mds.eu.1 <- metaMDS(biomassstand, distance='eu', k=2) #ficar em somente em 2 dimensoes
biomass.mds.eu.1
attributes(biomass.mds.eu.1)
plot(biomass.mds.eu.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(biomass.mds.eu.1, display='sites')
points(biomass.mds.eu.1$points[1:3,], col=1, cex=3)
points(biomass.mds.eu.1$points[4,], col=1, cex=3)
points(biomass.mds.eu.1$points[5:20,], col=2, cex=3)
points(biomass.mds.eu.1$points[21:40,], col=3, cex=3)
points(biomass.mds.eu.1$points[41:60,], col=4, cex=3)
points(biomass.mds.eu.1$points[61:75,], col=5, cex=3)
points(biomass.mds.eu.1$points[76:90,], col=5, cex=3)
points(biomass.mds.eu.1$points[91:105,], col=5, cex=3)
points(biomass.mds.eu.1$points[106:120,], col=5, cex=3)
points(biomass.mds.eu.1$points[121:135,], col=6, cex=3)
points(biomass.mds.eu.1$points[136:147,], col=6, cex=3)
points(biomass.mds.eu.1$points[148:159,], col=6, cex=3)
points(biomass.mds.eu.1$points[160:174,], col=6, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.0001'))
```

```
# paso 3 - Correlação linear - syneco e prochlo
biomasscor <- cor.test(biomass.ed$Prochlorococcus, biomass.ed$Synechococcus)
biomasscor
```

```
biomasscor.stand <- cor.test(biomassstand$Prochlorococcus, biomassstand$Synechococcus)
biomasscor.stand
```

```
##### 3.00µm
```

**Riqueza costa DGGE**

```
Riq3dgge <- read.table("~/Documents/R operacional/Riq 3.00 DGGE.txt", header = T, dec=',')
```

```
# paso 1
#entendendo a disposição dos dados:
View(Riq3dgge) #lendo no RStudio
Riq3dgge #lendo no console
attach(Riq3dgge) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(Riq3dgge) #resumo pra saber as significancias e valores da planilha
attributes(Riq3dgge) #mostra os criterios
```

```
Riq3dgge.ed <- Riq3dgge[,-1]
View(Riq3dgge.ed)
```

```
# paso 2 - informações ecologicas:
```

```
# riqueza observada
A <- specnumber(Riq3dgge.ed)
View(A)
```

```
#riqueza estimada
```

```

#estimadores de riqueza por ponto
estimateR(Riq3dgge.ed)

#estimadores de riqueza total - binario
specpool(Riq3dgge.ed)

#chao1 (binario)
C1 <- chao1(Riq3dgge.ed, taxa.row = FALSE)
C1

#chao2 (abundancia)
#C2 <- chao2(Riq3dgge.ed, taxa.row = FALSE)
#C2

#jacknife 1 e 2 (binario e abundancia)
J1 <- jack1(Riq3dgge.ed, taxa.row = FALSE, abund = FALSE)
J1

J2 <- jack2(Riq3dgge.ed, taxa.row = FALSE, abund = FALSE)
J2

#rarefacao
quantile(rowSums(Riq3dgge.ed)) #numero de amostras (amostargens) necessarias para chegar o mais perto do 100% do numero total de otus

#Chao random
all <- specaccum(Riq3dgge.ed, method = "random", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton (DGGE)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Chao rarefaction
all <- specaccum(Riq3dgge.ed, method = "rarefaction", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton (DGGE)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack random
all <- specaccum(Riq3dgge.ed, method = "random", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton (DGGE)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(Riq3dgge.ed, method = "random", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton (DGGE)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack rarefaction
all <- specaccum(Riq3dgge.ed, method = "rarefaction", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton (DGGE)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(Riq3dgge.ed, method = "rarefaction", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton (DGGE)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#adiv (binario)
#Shannon
H <- diversity(Riq3dgge.ed, index="shannon", base=exp(1))
View(H)

#Simpson
S <- diversity(Riq3dgge.ed, index="simpson", base=exp(1))
View(S)

#Bdiv (binario)
#betadiver(help=TRUE)
#indice Jaccard
Jac <- designdist(Riq3dgge.ed, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac)

```

```

scores(Jac)
quantile(Jac)
mean(Jac)

Jac.cluster <- hclust(Jac, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.cluster, 3)
Jac.cof<- cophenetic(Jac.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Jac.cof
plot(Jac.cof)
cor(Jac, Jac.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.982

#indice Sorensen
Sor <- designdist(Riq3dgge.ed, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor)
scores(Sor)
quantile(Sor)
mean(Sor)

Sor.cluster <- hclust(Sor, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.cluster, 3)
Sor.cof<- cophenetic(Sor.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Sor.cof
plot(Sor.cof)
cor(Sor, Sor.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.964

#decostand
#nao precisou transformar - decostand, dados binarios
Riq3dgge.ed.stan <- decostand(Riq3dgge.ed, method='stand') #funcao para padronizar dados, para deixar tudo em escalas proximas, no caso o satnd padroniza
para atingir a distribuicao normal padrao
Riq3dgge.ed.stan

#Riq3dgge.ed.dist.stan <- vegdist(Riq3dgge.ed.stan, method='bray', binary=TRUE)
#RRiq3dgge.ed.dist.stan
#Riq3dgge.ed.dist.stan.cluster <- hclust(Riq3dgge.ed.dist.stan) #gera a analise de cluster
#plot(Riq3dgge.ed.dist.stan.cluster) #gera o grafico de cluster

#nao precisou transformar - decostand, pois nao deu diferenca na aordenacao

# passo 3 - matrizes de ordenacao
#PCA
iris.pca <- PCA(Riq3dgge.ed, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

plot(iris.pca, display = "sites")
fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=Riq3dgge.$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=Riq3dgge.$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=Riq3dgge.$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(1,3), label="var", habillage=Riq3dgge.$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(2,3), label="var", habillage=Riq3dgge.$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()

#NMDS - (distancia Jaccard - biotico)
Riq3dgge.mds.Jac<- metaMDS(Riq3dgge.ed, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
Riq3dgge.mds.Jac
attributes(Riq3dgge.mds.Jac)
plot(Riq3dgge.mds.Jac, cex=3, display=c('sites')) #deixa apenas os pontos
text(Riq3dgge.mds.Jac, display='sites')
points(Riq3dgge.mds.Jac$points[1:2,], col=1, cex=3)
points(Riq3dgge.mds.Jac$points[3:4,], col=1, cex=3)
points(Riq3dgge.mds.Jac$points[5:7,], col=2, cex=3)
points(Riq3dgge.mds.Jac$points[8:10,], col=3, cex=3)
points(Riq3dgge.mds.Jac$points[11:13,], col=4, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))
legend(locator(1), c('stress = 0.0151'))

#MDS stand

```

```

Riq3dgge.mds.Jac.1 <- metaMDS(Riq3dgge.ed.stan, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
Riq3dgge.mds.Jac.1
attributes(Riq3dgge.mds.Jac.1)
plot(Riq3dgge.mds.Jac.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(Riq3dgge.mds.Jac.1, display='sites')
points(Riq3dgge.mds.Jac.1$points[1:2,], col=1, cex=3)
points(Riq3dgge.mds.Jac.1$points[3:4,], col=1, cex=3)
points(Riq3dgge.mds.Jac.1$points[5:7,], col=2, cex=3)
points(Riq3dgge.mds.Jac.1$points[8:10,], col=3, cex=3)
points(Riq3dgge.mds.Jac.1$points[11:13,], col=4, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))
legend(locator(1), c('stress = 0.0151'))

##### 0.22µm

Riq0.22dgge <- read.table("~/Documents/R operacional/Riq 0.22 DGGE.txt", header = T, dec=',')

# paso 1
#entendendo a disposição dos dados:
View(Riq0.22dgge) #lendo no RStudio
Riq0.22dgge #lendo no console
attach(Riq0.22dgge) #ê o cabeçário, ou seja, as VARIAVEIS da planilha e suas posicoes
summary(Riq0.22dgge) #resumo pra saber as significancias e valores da planilha
attributes(Riq0.22dgge) #mostra os criterios

Riq0.22dgge.ed <- Riq0.22dgge[,-1]
View(Riq0.22dgge.ed)

# paso 2 - informações ecológicas:

# riqueza observada
A <- specnumber(Riq0.22dgge.ed)
View(A)

#riqueza estimada
#estimadores de riqueza por ponto
estimateR(Riq0.22dgge.ed)

#estimadores de riqueza total - binario
specpool(Riq0.22dgge.ed)

#chao1 (abundancia)
C1 <- chao1(Riq0.22dgge.ed, taxa.row = FALSE)
C1

#chao2 (presenca e ausencia)
C2 <- chao2(Riq0.22dgge.ed, taxa.row = FALSE)
C2

#jacknife 1 e 2 (binario e abundancia)
J1 <- jack1(Riq0.22dgge.ed, taxa.row = FALSE, abund = FALSE)
J1

J2 <- jack2(Riq0.22dgge.ed, taxa.row = FALSE, abund = FALSE)
J2

#rarefacao
quantile(rowSums(Riq0.22dgge.ed)) #numero de amostras (amostargens) necessarias para chegar o mais perto do 100% do numero total de otus

#Chao random
all <- specaccum(Riq0.22dgge.ed, method = "random", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Picophytoplankton (DGGE)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Chao rarefaction
all <- specaccum(Riq0.22dgge.ed, method = "rarefaction", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Picophytoplankton (DGGE)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack random
all <- specaccum(Riq0.22dgge.ed, method = "random", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Picophytoplankton (DGGE)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")

```

```

boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(Riq0.22dgge.ed, method = "random", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Picophytoplankton (DGGE)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack rarefaction
all <- specaccum(Riq0.22dgge.ed, method = "rarefaction", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Picophytoplankton (DGGE)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(Riq0.22dgge.ed, method = "rarefaction", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Picophytoplankton (DGGE)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#adiv (binario)
#Shannon
H <- diversity(Riq0.22dgge.ed, index="shannon", base=exp(1))
View(H)

#Simpson
S <- diversity(Riq0.22dgge.ed, index="simpson", base=exp(1))
View(S)

#Bdiv (binario)
#betadiver(help=TRUE)
#indice Jaccard
Jac <- designdist(Riq0.22dgge.ed, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac)
scores(Jac)
quantile(Jac)
mean(Jac)

Jac.cluster <- hclust(Jac, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.cluster, 3)
Jac.cof<- cophenetic(Jac.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Jac.cof
plot(Jac.cof)
cor(Jac, Jac.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.982

#indice Sorensen
Sor <- designdist(Riq0.22dgge.ed, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor)
scores(Sor)
quantile(Sor)
mean(Sor)

Sor.cluster <- hclust(Sor, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.cluster, 3)
Sor.cof<- cophenetic(Sor.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Sor.cof
plot(Sor.cof)
cor(Sor, Sor.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.964

#decostand
#nao precisou transformar - decostand, dados binarios
Riq0.22dgge.ed.stan <- decostand(Riq0.22dgge.ed, method='stand') #funcao para padronizar dados, para deixar tudo em escalas proximas, no caso o satnd
padroniza para atingir a distribuicao normal padrao
Riq0.22dgge.ed.stan

#Riq0.22dgge.ed.dist.stan <- vegdist(Riq0.22dgge.ed.stan, method='bray', binary=TRUE)
#RRiq0.22dgge.ed.dist.stan
#Riq0.22dgge.ed.dist.stan.cluster <- hclust(Riq0.22dgge.ed.dist.stan) #gera a analise de cluster
#plot(Riq0.22dgge.ed.dist.stan.cluster) #gera o grafico de cluster

#nao precisou transformar - decostand, pois nao deu diferenca na aordenacao

```

```

# passo 3 - matrizes de ordenacao
#PCA
iris.pca <- PCA(Riq0.22dgge.ed, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

plot(iris.pca, display = "sites")
fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=Riq0.22dgge$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=Riq0.22dgge$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=Riq0.22dgge$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labels = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(1,3), label="var", habillage=Riq0.22dgge$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labels = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(2,3), label="var", habillage=Riq0.22dgge$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labels = 5,
pointsize = 4) + theme_minimal()

#NMDS - (distancia Jaccard - biotico)
Riq0.22dgge.mds.Jac<- metaMDS(Riq0.22dgge.ed, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
Riq0.22dgge.mds.Jac
attributes(Riq0.22dgge.mds.Jac)
plot(Riq0.22dgge.mds.Jac, cex=3, display=c('sites')) #deixa apenas os pontos
text(Riq0.22dgge.mds.Jac, display='sites')
points(Riq0.22dgge.mds.Jac$points[1:2,], col=1, cex=3)
points(Riq0.22dgge.mds.Jac$points[3:4,], col=1, cex=3)
points(Riq0.22dgge.mds.Jac$points[5:7,], col=2, cex=3)
points(Riq0.22dgge.mds.Jac$points[8:10,], col=3, cex=3)
points(Riq0.22dgge.mds.Jac$points[11:13,], col=4, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))
legend(locator(1), c('stress = 0.0001'))

#MDS stand
Riq0.22dgge.mds.Jac.1<- metaMDS(Riq0.22dgge.ed.stan, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
Riq0.22dgge.mds.Jac.1
attributes(Riq0.22dgge.mds.Jac.1)
plot(Riq0.22dgge.mds.Jac.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(Riq0.22dgge.mds.Jac.1, display='sites')
points(Riq0.22dgge.mds.Jac.1$points[1:2,], col=1, cex=3)
points(Riq0.22dgge.mds.Jac.1$points[3:4,], col=1, cex=3)
points(Riq0.22dgge.mds.Jac.1$points[5:7,], col=2, cex=3)
points(Riq0.22dgge.mds.Jac.1$points[8:10,], col=3, cex=3)
points(Riq0.22dgge.mds.Jac.1$points[11:13,], col=4, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))
legend(locator(1), c('stress = ?'))

##### COAST
##### 3.00µm

Riq3 <- read.table("~/Documents/R operacional/Riq 3.00 illumina coast.txt", header = T, dec=',')

# paso 1
#entendendo a disposiçao dos dados:
View(Riq3) #lendo no RStudio
Riq3 #lendo no console
attach(Riq3) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(Riq3) #resumo pra saber as significancias e valores da planilha
attributes(Riq3) #mostra os criterios

Riq3.ed.c <- Riq3[,-1]
View(Riq3.ed.c)

# paso 2 - informações ecologicas:

# riqueza observada
A <- specnumber(Riq3.ed.c)
View(A)

#riqueza estimada
#estimadores de riqueza por ponto
estimateR(Riq3.ed.c)

#estimadores de riqueza total - binario
specpool(Riq3.ed.c)

#chao1 (abundancia)

```

Riqueza costa Illumina

```

C1 <- chao1(Riq3.ed.c, taxa.row = FALSE)
C1

#chao2 (presenca e ausencia)
C2 <- chao2(Riq3.ed.c, taxa.row = FALSE)
C2

#jacknife 1 e 2 (binario e abundancia)
J1 <- jack1(Riq3.ed.c, taxa.row = FALSE, abund = FALSE)
J1

J2 <- jack2(Riq3.ed.c, taxa.row = FALSE, abund = FALSE)
J2

#rarefacao
quantile(rowSums(Riq3.ed.c)) #numero de amostras (amostargens) necessarias para chegar o mais perto do 100% do numero total de otus

#Chao random
all <- specaccum(Riq3.ed.c, method = "random", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Chao rarefaction
all <- specaccum(Riq3.ed.c, method = "rarefaction", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack random
all <- specaccum(Riq3.ed.c, method = "random", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(Riq3.ed.c, method = "random", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack rarefaction
all <- specaccum(Riq3.ed.c, method = "rarefaction", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(Riq3.ed.c, method = "rarefaction", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#adiv (binario)
#Shannon
H <- diversity(Riq3.ed.c, index="shannon", base=exp(1))
View(H)

#Simpson
S <- diversity(Riq3.ed.c, index="simpson", base=exp(1))
View(S)

#Bdiv (binario)
#betadiver(help=TRUE)
#indice Jaccard
Jac <- designdist(Riq3.ed.c, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac)
scores(Jac)
quantile(Jac)
mean(Jac)

Jac.cluster <- hclust(Jac, method="average")#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.cluster, 4)

```

```

Jac.cof<- cophenetic(Jac.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Jac.cof
plot(Jac.cof)
cor(Jac, Jac.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.738

#indice Sorensen
Sor <- designdist(Riq3.ed.c, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor)
scores(Sor)
quantile(Sor)
mean(Sor)

Sor.cluster <- hclust(Sor, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.cluster, 4)
Sor.cof<- cophenetic(Sor.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Sor.cof
plot(Sor.cof)
cor(Sor, Sor.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.704

#decostand
Riq3.ed.c.stan <- decostand(Riq3.ed.c, method='stand') #funcao para padronizar dados, para deixar tudo em escalas proximas, no caso o satnd padroniza para
atingir a distribuicao normal padrao
Riq3.ed.c.stan

#Riq3.ed.c.dist.stan <- designdist(Riq3.ed.c.stan, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
#Riq3.ed.c.dist.stan
#Riq3.ed.c.dist.stan.cluster <- hclust(Riq3.ed.c.dist.stan, method='average') #gera a analise de cluster
#plot(Riq3.ed.c.dist.stan.cluster) #gera o grafico de cluster
#Riq3.ed.c.dist.stan.cof<- cophenetic(Riq3.ed.c.dist.stan.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves
do metodo do vizinho completo ou mais distante
#Riq3.ed.c.dist.stan.cof
#plot(Riq3.ed.c.dist.stan.cof)
#cor(Riq3.ed.c.dist.stan, Riq3.ed.c.dist.stan.cof)

#nao precisou transformar - decostand, pois nao deu diferenca na aordenacao

# passo 3 - matrizes de ordenacao
#PCA
iris.pca <- PCA(Riq3.ed.c, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

plot(iris.pca, display = "sites")
fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=Riq3$$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=Riq3$$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=Riq3$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5, pointsize
= 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(1,3), label="var", habillage=Riq3$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5, pointsize
= 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(2,3), label="var", habillage=Riq3$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5, pointsize
= 4) + theme_minimal()

#NMDS - (distancia Jaccard - biotico)
Riq3.ed.c.mds.Jac<- metaMDS(Riq3.ed.c, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
Riq3.ed.c.mds.Jac
attributes(Riq3.ed.c.mds.Jac)
plot(Riq3.ed.c.mds.Jac, cex=3, display=c('sites')) #deixa apenas os pontos
text(Riq3.ed.c.mds.Jac, display='sites')
points(Riq3.ed.c.mds.Jac$points[1:2,], col=1, cex=3)
points(Riq3.ed.c.mds.Jac$points[3:4,], col=1, cex=3)
points(Riq3.ed.c.mds.Jac$points[5:7,], col=2, cex=3)
points(Riq3.ed.c.mds.Jac$points[8:10,], col=3, cex=3)
points(Riq3.ed.c.mds.Jac$points[11:13,], col=4, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))
legend(locator(1), c('stress = 0.0900'))

#MDS stand
Riq3.ed.c.mds.Jac.1<- metaMDS(Riq3.ed.c.stan, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
Riq3.ed.c.mds.Jac.1
attributes(Riq3.ed.c.mds.Jac.1)

```

```

plot(Riq3.ed.c.mds.Jac.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(Riq3.ed.c.mds.Jac.1, display='sites')
points(Riq3.ed.c.mds.Jac.1$points[1:2,], col=1, cex=3)
points(Riq3.ed.c.mds.Jac.1$points[3:4,], col=1, cex=3)
points(Riq3.ed.c.mds.Jac.1$points[5:7,], col=2, cex=3)
points(Riq3.ed.c.mds.Jac.1$points[8:10,], col=3, cex=3)
points(Riq3.ed.c.mds.Jac.1$points[11:13,], col=4, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))
legend(locator(1), c('stress = ?'))

##### COAST
##### 0.22µm

Riq0.22 <- read.table("~/Documents/R operacional/Riq 0.22 illumina coast.txt", header = T, dec=',')

# paso 1
#entendendo a disposição dos dados:
View(Riq0.22) #lendo no RStudio
Riq0.22 #lendo no console
attach(Riq0.22) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(Riq0.22) #resumo pra saber as significancias e valores da planilha
attributes(Riq0.22) #mostra os criterios

Riq0.22.ed.c <- Riq0.22[,-1]
View(Riq0.22.ed.c)

# paso 2 - informações ecológicas:

# riqueza observada
A <- specnumber(Riq0.22.ed.c)
View(A)

#riqueza estimada
#estimadores de riqueza por ponto
estimateR(Riq0.22.ed.c)

#estimadores de riqueza total - binario
specpool(Riq0.22.ed.c)

#chao1 (abundancia)
C1 <- chao1(Riq0.22.ed.c, taxa.row = FALSE)
C1

#chao2 (presenca e ausencia)
C2 <- chao2(Riq0.22.ed.c, taxa.row = FALSE)
C2

#jacknife 1 e 2 (binario e abundancia)
J1 <- jack1(Riq0.22.ed.c, taxa.row = FALSE, abund = FALSE)
J1

J2 <- jack2(Riq0.22.ed.c, taxa.row = FALSE, abund = FALSE)
J2

#rarefacao
quantile(rowSums(Riq0.22.ed.c)) #numero de amostras (amostargens) necessarias para chegar o mais perto do 100% do numero total de otus

#Chao random
all <- specaccum(Riq0.22.ed.c, method = "random", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Picophytoplankton (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Chao rarefaction
all <- specaccum(Riq0.22.ed.c, method = "rarefaction", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Picophytoplankton (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack random
all <- specaccum(Riq0.22.ed.c, method = "random", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Picophytoplankton (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

```

```

all <- specaccum(Riq0.22.ed.c, method = "random", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Picophytoplankton (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack rarefaction
all <- specaccum(Riq0.22.ed.c, method = "rarefaction", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Picophytoplankton (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(Riq0.22.ed.c, method = "rarefaction", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Picophytoplankton (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Cyano OTUs")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#adiv (binario)
#Shannon
H <- diversity(Riq0.22.ed.c, index="shannon", base=exp(1))
View(H)

#Simpson
S <- diversity(Riq0.22.ed.c, index="simpson", base=exp(1))
View(S)

#Bdiv (binario)
#betadiver(help=TRUE)
#indice Jaccard
Jac <- designdist(Riq0.22.ed.c, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac)
scores(Jac)
quantile(Jac)
mean(Jac)

Jac.cluster <- hclust(Jac, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.cluster, 3)
Jac.cof<- cophenetic(Jac.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Jac.cof
plot(Jac.cof)
cor(Jac, Jac.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.883

#indice Sorensen
Sor <- designdist(Riq0.22.ed.c, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor)
scores(Sor)
quantile(Sor)
mean(Sor)

Sor.cluster <- hclust(Sor, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.cluster, 3)
Sor.cof<- cophenetic(Sor.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Sor.cof
plot(Sor.cof)
cor(Sor, Sor.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.964

#decostand
Riq0.22.ed.c.stan <- decostand(Riq0.22.ed.c, method='stand') #funcao para padronizar dados, para deixar tudo em escalas proximas, no caso o satnd padroniza
para atingir a distribuicao normal padrao
Riq0.22.ed.c.stan

#Riq0.22.ed.c.dist.stan <- designdist(Riq0.22.ed.c.stan, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
#Riq0.22.ed.c.dist.stan
#Riq0.22.ed.c.dist.stan.cluster <- hclust(Riq0.22.ed.c.dist.stan, method='average') #gera a analise de cluster
#plot(Riq0.22.ed.c.dist.stan.cluster) #gera o grafico de cluster
#Riq0.22.ed.c.dist.stan.cof<- cophenetic(Riq0.22.ed.c.dist.stan.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada
atraves do metodo do vizinho completo ou mais distante
#Riq0.22.ed.c.dist.stan.cof
#plot(Riq0.22.ed.c.dist.stan.cof)
#cor(Riq0.22.ed.c.dist.stan, Riq0.22.ed.c.dist.stan.cof)

```

```

#nao precisou transformar - decostand, pois nao deu diferenca na aordenacao

# passo 3 - matrizes de ordenacao
#PCA
iris.pca <- PCA(Riq0.22.ed.c, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

plot(iris.pca, display = "sites")
fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=Riq0.22$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=Riq0.22$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=Riq0.22$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(1,3), label="var", habillage=Riq0.22$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(2,3), label="var", habillage=Riq0.22$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()

#NMDS - (distancia Jaccard - biotico)
Riq0.22.ed.c.mds.Jac<- metaMDS(Riq0.22.ed.c, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
Riq0.22.ed.c.mds.Jac
attributes(Riq0.22.ed.c.mds.Jac)
plot(Riq0.22.ed.c.mds.Jac, cex=3, display=c('sites')) #deixa apenas os pontos
text(Riq0.22.ed.c.mds.Jac, display='sites')
points(Riq0.22.ed.c.mds.Jac$points[1:2,], col=1, cex=3)
points(Riq0.22.ed.c.mds.Jac$points[3:4,], col=1, cex=3)
points(Riq0.22.ed.c.mds.Jac$points[5:7,], col=2, cex=3)
points(Riq0.22.ed.c.mds.Jac$points[8:10,], col=3, cex=3)
points(Riq0.22.ed.c.mds.Jac$points[11:13,], col=4, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))
legend(locator(1), c('stress = 0.1041'))

#MDS stand
Riq0.22.ed.c.mds.Jac.1<- metaMDS(Riq0.22.ed.c.stan, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
Riq0.22.ed.c.mds.Jac.1
attributes(Riq0.22.ed.c.mds.Jac.1)
plot(Riq0.22.ed.c.mds.Jac.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(Riq0.22.ed.c.mds.Jac.1, display='sites')
points(Riq0.22.ed.c.mds.Jac.1$points[1:2,], col=1, cex=3)
points(Riq0.22.ed.c.mds.Jac.1$points[3:4,], col=1, cex=3)
points(Riq0.22.ed.c.mds.Jac.1$points[5:7,], col=2, cex=3)
points(Riq0.22.ed.c.mds.Jac.1$points[8:10,], col=3, cex=3)
points(Riq0.22.ed.c.mds.Jac.1$points[11:13,], col=4, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))
legend(locator(1), c('stress = ?'))

##### 3.00µm
##### All - seq

seq3 <- read.table("~/Documents/R operacional/seq 3.00 illumina.txt", header = T, dec=',')

##### paso 1 - entendendo a disposiçao dos dados All 3.00µm
View(seq3) #lendo no RStudio
seq3 #lendo no console
attach(seq3) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(seq3) #resumo pra saber as significancias e valores da planilha
attributes(seq3) #mostra os criterios

#Syneco

#graficos de media com desvio padrao, funcao BOXPLOT
par(mar = c(5,5,2,2), cex.lab = 1.5)
boxplot(seq3$Synecococcus) #media e desvios do tamanho apenas
boxplot(seq3$Synecococcus~seq3$Samples) #da um ~ com um espaco, eh um comando que da a media e o desvio de uma variavel por outra variavel, o ~
separa os dados

shapiro.test(seq3$Synecococcus) #teste de normalidade p<0,05 para ser normal
#deu normal (tem de transformar em log, e agora usar essa planilha transformada pra tudo)
#seq3stan.syn <- decostand(seq3$Synecococcus, "standardize")
#View(seq3stan.syn)
#shapiro.test(seq3stan.syn)

#ANOVA unifactorial
ANOVA.syn<-aov(seq3$Synecococcus~seq3$Samples, seq3) #fit e o nome que eu dei para aov (analysis of variance) do que eu quero

```

```

ANOVA.syn
anova(ANOVA.syn) #Análise de variancia, probabilidades
TukeyHSD(ANOVA.syn) #Teste Tukey
plot(ANOVA.syn)
plot(TukeyHSD(ANOVA.syn))

#Prochloro

#graficos de media com desvio padrao, funcao BOXPLOT
par(mar = c(5,5,2,2), cex.lab = 1.5)
boxplot(seq3$Prochlorococcus) #media e desvios do tamanho apenas
boxplot(seq3$Prochlorococcus~seq3$Samples) #da um ~ com um espaco, eh um comando que da a media e o desvio de uma variavel por outra variavel, o ~
separa os dados

shapiro.test(seq3$Prochlorococcus) #teste de normalidade p<0,05 para ser normal
#deu normal - tem de transformar e usar essa planilha transformada pra tudo
#seq3stan.pro <- decostand(seq3$Prochlorococcus, "standardize")
#View(seq3stan.pro)
#shapiro.test(seq3stan.pro)

#ANOVA unifactorial
ANOVA.pro<-aov(seq3$Prochlorococcus~seq3$Samples, seq3) #fit e o nome que eu dei para aov (analysis of variance) do que eu quero
ANOVA.pro
anova(ANOVA.pro) #Análise de variancia, probabilidades
TukeyHSD(ANOVA.pro) #Teste Tukey
plot(ANOVA.pro)
plot(TukeyHSD(ANOVA.pro))

#All
#ajustes necessarios p/ estatisticas
seq3ed <- seq3[,-1]
View(seq3ed)

seq3edstand <- decostand(seq3ed, "standardize")
View(seq3edstand)
shapiro.test(seq3edstand$Synechococcus)
shapiro.test(seq3edstand$Prochlorococcus)

#ANOVA multifatorial
fait<-aov(seq3ed$Prochlorococcus~seq3ed$Synechococcus*seq3ed$Prochlorococcus, seq3ed)
anova(fait)
TukeyHSD(fait)

frait<-aov(seq3edstand$Prochlorococcus~seq3edstand$Synechococcus*seq3edstand$Prochlorococcus, seq3edstand)
anova(frait)
TukeyHSD(frait)

# paso 2 - matriz de dissimilaridade (agrupamento) - distancia euclidiana com dados bioticos:
#SEM STAND
seq3ed.eu.1<-vegdist(seq3ed, method='eu', binary=TRUE, diag = F, upper = F) #apenas dei um nome, indice de similaridade (distância euclidiana pura para
analises quantitativas)
seq3ed.eu.1
seq3ed.eu.cluster.1<-hclust(seq3ed.eu.1, method = "average") # nomeando e gerando a analise de cluster
seq3ed.eu.cluster.1
plot(seq3ed.eu.cluster.1, hang = -1) #gera o grafico de cluster
rect.hclust(seq3ed.eu.cluster.1, 4)
seq3ed.eu.cluster.1.cof<- cophenetic(seq3ed.eu.cluster.1) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do
metodo do vizinho completo ou mais distante
seq3ed.eu.cluster.1.cof
plot(seq3ed.eu.cluster.1.cof)
cor(seq3ed.eu.1, seq3ed.eu.cluster.1.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh
o que tiver maior correlacao, no caso 0.704

#STAND
seq3edstand.eu.1<-vegdist(seq3edstand, method='eu', binary=TRUE, diag = F, upper = F) #apenas dei um nome, indice de similaridade (distância euclidiana
pura para analises quantitativas)
seq3edstand.eu.1
seq3edstand.eu.cluster.1<-hclust(seq3edstand.eu.1, method = "average") # nomeando e gerando a analise de cluster
seq3edstand.eu.cluster.1
plot(seq3edstand.eu.cluster.1, hang = -1) #gera o grafico de cluster
rect.hclust(seq3edstand.eu.cluster.1, 3)
seq3edstand.eu.cluster.1.cof<- cophenetic(seq3edstand.eu.cluster.1) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada
atraves do metodo do vizinho completo ou mais distante
seq3edstand.eu.cluster.1.cof
plot(seq3edstand.eu.cluster.1.cof)
cor(seq3edstand.eu.1, seq3edstand.eu.cluster.1.cof) #cor

# paso 3

```

```

#PCA SEM STAND - Usar esse:
iris.pca <- PCA(seq3ed, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=seq3$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=seq3$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=seq3$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz
e = 5, pointsize = 4) + theme_minimal()

#PCA stand
iris.pca.1 <- PCA(seq3edstand, scale.unit = FALSE, graph = FALSE)
iris.pca.1
loadings(iris.pca.1) #verifica os pesos encontrados
summary(iris.pca.1)
attributes(iris.pca.1)

fviz_pca_ind(iris.pca.1, label="none")
fviz_pca_ind(iris.pca.1, label="none", habillage=seq3$Samples)
fviz_pca_ind(iris.pca.1, label="none", habillage=seq3$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca.1, axes = c(1,2), label="var", habillage=seq3$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz
e = 5, pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(1,3), label="var", habillage=seq3$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz
e = 5, pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(2,3), label="var", habillage=seq3$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz
e = 5, pointsize = 4) + theme_minimal()

#MDS - (distancia euclidiana metrica - abiotico)
seq3.mds.eu <- metaMDS(seq3ed, distance='eu', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq3.mds.eu
attributes(seq3.mds.eu)
plot(seq3.mds.eu, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq3.mds.eu, display='sites')
points(seq3.mds.eu$points[1:2,], col=1, cex=3)
points(seq3.mds.eu$points[3:4,], col=1, cex=3)
points(seq3.mds.eu$points[5:7,], col=2, cex=3)
points(seq3.mds.eu$points[8:10,], col=3, cex=3)
points(seq3.mds.eu$points[11:13,], col=4, cex=3)
points(seq3.mds.eu$points[14:16,], col=5, cex=3)
points(seq3.mds.eu$points[17:19,], col=5, cex=3)
points(seq3.mds.eu$points[20:22,], col=5, cex=3)
points(seq3.mds.eu$points[23:25,], col=5, cex=3)
points(seq3.mds.eu$points[26:28,], col=6, cex=3)
points(seq3.mds.eu$points[29:31,], col=6, cex=3)
points(seq3.mds.eu$points[32:34,], col=6, cex=3)
points(seq3.mds.eu$points[35:37,], col=6, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.0862'))

#MDS stand
seq3.mds.eu.1 <- metaMDS(seq3edstand, distance='eu', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq3.mds.eu.1
attributes(seq3.mds.eu.1)
plot(seq3.mds.eu.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq3.mds.eu.1, display='sites')
points(seq3.mds.eu.1$points[1:2,], col=1, cex=3)
points(seq3.mds.eu.1$points[3:4,], col=1, cex=3)
points(seq3.mds.eu.1$points[5:7,], col=2, cex=3)
points(seq3.mds.eu.1$points[8:10,], col=3, cex=3)
points(seq3.mds.eu.1$points[11:13,], col=4, cex=3)
points(seq3.mds.eu.1$points[14:16,], col=5, cex=3)
points(seq3.mds.eu.1$points[17:19,], col=5, cex=3)
points(seq3.mds.eu.1$points[20:22,], col=5, cex=3)
points(seq3.mds.eu.1$points[23:25,], col=5, cex=3)
points(seq3.mds.eu.1$points[26:28,], col=6, cex=3)
points(seq3.mds.eu.1$points[29:31,], col=6, cex=3)
points(seq3.mds.eu.1$points[32:34,], col=6, cex=3)
points(seq3.mds.eu.1$points[35:37,], col=6, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.0905'))

##### 3.00µm
##### All - eco

#restart R

```

```

# paso 1 - informacoes preliminares
seq3 <- read.table("~/Documents/R operacional/seq 3.00 illumina.txt", header = T, dec=',')
seq3ed <- seq3[,-1]

View(seq3)
View(seq3ed)

# paso 2 - informações ecologicas All - eco 3.00µm
# riqueza total
Stotal<- diversityresult (seq3ed, index='richness', method='pooled') #abundancia total por todos os pontos, utiliza a tabela onde tem os dados crus da
quantidade de especies por ponto
Stotal
summary(Stotal)

# riqueza media
Slocal<- diversityresult (seq3ed, index='richness', method='mean') #abundancia por ponto
Slocal
summary(Slocal)

# riqueza observada por ponto
A <- specnumber(seq3ed)
View(A)

#riqueza estimada
#estimadores de riqueza por ponto
estimateR(seq3ed)

#estimadores de riqueza total
specpool(seq3ed)

#chao1 (abundancia)
C1 <- chao1(seq3ed, taxa.row = FALSE)
C1

#chao2 (presenca e ausencia)
C2 <- chao2(seq3ed, taxa.row = FALSE)
C2

#jacknife 1 e 2 (binario e abundancia)
J1 <- jack1(seq3ed, taxa.row = FALSE, abund = FALSE)
J1

J2 <- jack2(seq3ed, taxa.row = FALSE, abund = FALSE)
J2

#rarefacao
quantile(rowSums(seq3ed)) #numero de amostras (amostargens) necessarias para chegar o mais perto do 100% do numero total de otus

#menor abundancia dentre todas as amostras - nocao para igualar as diferentes amostargens
raremax <- min(rowSums(seq3ed))
raremax#590

#numero de especies rarefeitas (que nao estao na amostra, especies raras)
R1 <- rarefy(seq3ed, raremax)#590, nesse caso dessa planilha - menor numero observado de sequencias medias. Comparacao com esforcos diferentes (O
tamanho da subamostra para a comunidade mais pequena, seja um valor único ou um vetor)
View(R1)
summary(R1)
plot(R1)

#curva de rarefacao por ponto - rarefacao
col <- c("black", "darkred", "forestgreen", "orange", "blue", "yellow", "hotpink")
lty <- c("solid", "dashed", "longdash", "dotdash")
pars <- expand.grid(col = col, lty = lty, stringsAsFactors = FALSE)
head(pars)

out <- with(pars[1:37, ],
  rarecurve(seq3ed, step = 20, sample = raremax, display=c('sites'), col=c(1,1,2,3,4,5,5,5,5,6,6,6,6),
    lty = lty, label = FALSE))
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,6,6,6,6))

#curva de rarefacao total - Chao
all <- specaccum(seq3ed, method = "random", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

```

```

#Chao rarefaction
all <- specaccum(seq3ed, method = "rarefaction", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack random
all <- specaccum(seq3ed, method = "random", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(seq3ed, method = "random", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack rarefaction
all <- specaccum(seq3ed, method = "rarefaction", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(seq3ed, method = "rarefaction", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#adiv (binario)
#Shannon
#diversidade geral
HShannon <- diversityresult(seq3ed, index='Shannon', method='pooled') #diversidade total por todos os pontos
HShannon

#diversidade media a cada ponto
h1Shannon <- diversityresult(seq3ed, index='Shannon', method='mean') #diversidade por ponto
h1Shannon

#diversidade por ponto
H <- diversity(seq3ed, index="shannon", base=exp(1))
View(H)
summary(H)

#Simpson
#diversidade geral
SSimpson <- diversityresult(seq3ed, index='Simpson', method='pooled') #diversidade total por todos os pontos
SSimpson

#diversidade media a cada ponto
s1Simpson <- diversityresult(seq3ed, index='Simpson', method='mean') #diversidade por ponto
s1Simpson

#diversidade por ponto
S <- diversity(seq3ed, index="simpson", base=exp(1))
View(S)
summary(S)

#Bdiv (binario)
#betadiver(help=TRUE)
#indice Jaccard
Jac <- designdist(seq3ed, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac)
scores(Jac)
quantile(Jac)
mean(Jac)

Jac.cluster <- hclust(Jac, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.cluster, 3)
Jac.cof<- cophenetic(Jac.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Jac.cof
plot(Jac.cof)

```

```
cor(Jac, Jac.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior correlacao, no caso 0.738
```

```
#decostand  
seq3ed.stan <- decostand(seq3ed, method='stand') #funcao para padronizar dados, para deixar tudo em escalas proximas, no caso o satnd padroniza para atingir a distribuicao normal padrao  
seq3ed.stan
```

```
#indice Jaccard STAND  
Jac.stand <- designdist(seq3ed.stan, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)  
plot(Jac.stand)  
scores(Jac.stand)  
quantile(Jac.stand)  
mean(Jac.stand)
```

```
Jac.stand.cluster <- hclust(Jac.stand, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis  
plot(Jac.stand.cluster, hang = -1) #gera o grafico de cluster  
rect.hclust(Jac.stand.cluster, 3)  
Jac.stand.cof<- cophenetic(Jac.stand.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante  
Jac.stand.cof  
plot(Jac.stand.cof)  
cor(Jac.stand, Jac.stand.cof)
```

```
#indice Sorensen  
Sor <- designdist(seq3ed, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)  
plot(Sor)  
scores(Sor)  
quantile(Sor)  
mean(Sor)
```

```
Sor.cluster <- hclust(Sor, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis  
plot(Sor.cluster, hang = -1) #gera o grafico de cluster  
rect.hclust(Sor.cluster, 3)  
Sor.cof<- cophenetic(Sor.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante  
Sor.cof  
plot(Sor.cof)  
cor(Sor, Sor.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior correlacao, no caso 0.704
```

```
#indice Sorensen STAND  
Sor.stand <- designdist(seq3ed.stan, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)  
plot(Sor.stand)  
scores(Sor.stand)  
quantile(Sor.stand)  
mean(Sor.stand)
```

```
Sor.stand.cluster <- hclust(Sor.stand, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis  
plot(Sor.stand.cluster, hang = -1) #gera o grafico de cluster  
rect.hclust(Sor.stand.cluster, 3)  
Sor.stand.cof<- cophenetic(Sor.stand.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante  
Sor.stand.cof  
plot(Sor.stand.cof)  
cor(Sor.stand, Sor.stand.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior correlacao, no caso 0.704
```

```
# passo 3 - matrizes de ordenacao
```

```
#PCA  
iris.pca <- PCA(seq3ed, scale.unit = FALSE, graph = FALSE)  
iris.pca  
loadings(iris.pca) #verifica os pesos encontrados  
summary(iris.pca)  
attributes(iris.pca)
```

```
plot(iris.pca, display = "sites")  
fviz_pca_ind(iris.pca, label="none")  
fviz_pca_ind(iris.pca, label="none", habillage=seq3$$Samples)  
fviz_pca_ind(iris.pca, label="none", habillage=seq3$$Samples, addEllipses=TRUE, ellipse.level=0.95)  
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=seq3$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5, pointsize = 4) + theme_minimal()  
fviz_pca_biplot(iris.pca, axes = c(1,3), label="var", habillage=seq3$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5, pointsize = 4) + theme_minimal()  
fviz_pca_biplot(iris.pca, axes = c(2,3), label="var", habillage=seq3$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5, pointsize = 4) + theme_minimal()
```

```
#PCA stand
```

```

iris.pca.1 <- PCA(seq3ed.stan, scale.unit = FALSE, graph = FALSE)
iris.pca.1
loadings(iris.pca.1) #verifica os pesos encontrados
summary(iris.pca.1)
attributes(iris.pca.1)

plot(iris.pca.1, display = "sites")
fviz_pca_ind(iris.pca.1, label="none")
fviz_pca_ind(iris.pca.1, label="none", habillage=seq3$$Samples)
fviz_pca_ind(iris.pca.1, label="none", habillage=seq3$$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca.1, axes = c(1,2), label="var", habillage=seq3$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(1,3), label="var", habillage=seq3$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(2,3), label="var", habillage=seq3$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()

#NMDS - (distancia Jaccard - biotico)
seq3ed.mds.Jac<- metaMDS(seq3ed, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq3ed.mds.Jac
attributes(seq3ed.mds.Jac)
plot(seq3ed.mds.Jac, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq3ed.mds.Jac, display='sites')
points(seq3ed.mds.Jac$points[1:2,], col=1, cex=3)
points(seq3ed.mds.Jac$points[3:4,], col=1, cex=3)
points(seq3ed.mds.Jac$points[5:7,], col=2, cex=3)
points(seq3ed.mds.Jac$points[8:10,], col=3, cex=3)
points(seq3ed.mds.Jac$points[11:13,], col=4, cex=3)
points(seq3ed.mds.Jac$points[14:16,], col=5, cex=3)
points(seq3ed.mds.Jac$points[17:19,], col=5, cex=3)
points(seq3ed.mds.Jac$points[20:22,], col=5, cex=3)
points(seq3ed.mds.Jac$points[23:25,], col=5, cex=3)
points(seq3ed.mds.Jac$points[26:28,], col=6, cex=3)
points(seq3ed.mds.Jac$points[29:31,], col=6, cex=3)
points(seq3ed.mds.Jac$points[32:34,], col=6, cex=3)
points(seq3ed.mds.Jac$points[35:37,], col=6, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.1515'))

#NMDS stand
seq3ed.mds.Jac.1<- metaMDS(seq3ed.stan, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq3ed.mds.Jac.1
attributes(seq3ed.mds.Jac.1)
plot(seq3ed.mds.Jac.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq3ed.mds.Jac.1, display='sites')
points(seq3ed.mds.Jac.1$points[1:2,], col=1, cex=3)
points(seq3ed.mds.Jac.1$points[3:4,], col=1, cex=3)
points(seq3ed.mds.Jac.1$points[5:7,], col=2, cex=3)
points(seq3ed.mds.Jac.1$points[8:10,], col=3, cex=3)
points(seq3ed.mds.Jac.1$points[11:13,], col=4, cex=3)
points(seq3ed.mds.Jac.1$points[14:16,], col=5, cex=3)
points(seq3ed.mds.Jac.1$points[17:19,], col=5, cex=3)
points(seq3ed.mds.Jac.1$points[20:22,], col=5, cex=3)
points(seq3ed.mds.Jac.1$points[23:25,], col=5, cex=3)
points(seq3ed.mds.Jac.1$points[26:28,], col=6, cex=3)
points(seq3ed.mds.Jac.1$points[29:31,], col=6, cex=3)
points(seq3ed.mds.Jac.1$points[32:34,], col=6, cex=3)
points(seq3ed.mds.Jac.1$points[35:37,], col=6, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.0001'))

##### 3.00µm
##### COAST

#restart R

seq3.c <- read.table("~/Documents/R operacional/seq 3.00 illumina coast.txt", header = T, dec=',')

# paso 1
#entendendo a disposição dos dados:
View(seq3.c) #lendo no RStudio
seq3.c #lendo no console
attach(seq3.c) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(seq3.c) #resumo pra saber as significancias e valores da planilha
attributes(seq3.c) #mostra os criterios

seq3ed.c.1 <- seq3.c[,-1]
seq3ed.c.1

```

```

taxa.off <- colSums(seq3ed.c.1)
taxa.off

seq3ed.c <- seq3ed.c[, which(colSums(seq3ed.c.1) != 0)]
View(seq3ed.c)

n.taxa.off <- ncol(seq3ed.c.1)-ncol(seq3ed.c)
n.taxa.off

# paso 2 - informações ecologicas
# riqueza total
Stotal<- diversityresult (seq3ed.c, index='richness', method='pooled') #abundancia total por todos os pontos, utiliza a tabela onde tem os dados crus da
quantidade de especies por ponto
Stotal
summary(Stotal)

# riqueza media
Slocal<- diversityresult (seq3ed.c, index='richness', method='mean') #abundancia por ponto
Slocal
summary(Slocal)

# riqueza observada por ponto
A <- specnumber(seq3ed.c)
View(A)

#riqueza estimada
#estimadores de riqueza por ponto
estimateR(seq3ed.c)

#estimadores de riqueza total
specpool(seq3ed.c)

#chao1 (abundancia)
C1 <- chao1(seq3ed.c, taxa.row = FALSE)
C1

#chao2 (presenca e ausencia)
C2 <- chao2(seq3ed.c, taxa.row = FALSE)
C2

#jacknife 1 e 2 (binario e abundancia)
J1 <- jack1(seq3ed.c, taxa.row = FALSE, abund = FALSE)
J1

J2 <- jack2(seq3ed.c, taxa.row = FALSE, abund = FALSE)
J2

#rarefacao
quantile(rowSums(seq3ed.c)) #numero de amostras (amostargens) necessarias para chegar o mais perto do 100% do numero total de otus

#menor abundancia dentre todas as amostras - nocaio para igualar as diferentes amostargens
raremax <- min(rowSums(seq3ed.c))
raremax#3406

#numero de especies rarefeitas (que nao estao na amostra, especies raras)
R1 <- rarefy(seq3ed.c, raremax)#3406, nesse caso dessa planilha - menor numero observado de sequencias medias. Comparacao com esforcos diferentes (O
tamanho da subamostra para a comunidade mais pequena, seja um valor unico ou um vetor)
View(R1)
summary(R1)
plot(R1)

#curva de rarefacao por ponto - rarefacao
col <- c("black", "darkred", "forestgreen", "orange", "blue", "yellow", "hotpink")
lty <- c("solid", "dashed", "longdash", "dotdash")
pars <- expand.grid(col = col, lty = lty, stringsAsFactors = FALSE)
head(pars)

out <- with(pars[1:13, ],
  rarecurve(seq3ed.c, step = 20, sample = raremax, display=c('sites'), col=c(1,1,2,3,4),
    lty = lty, label = FALSE))
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))

#curva de rarefacao total - Chao
all <- specaccum(seq3ed.c, method = "random", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Coastal nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")

```

```

boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Chao rarefaction
all <- specaccum(seq3ed.c, method = "rarefaction", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Coastal nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack random
all <- specaccum(seq3ed.c, method = "random", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Coastal nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(seq3ed.c, method = "random", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Coastal nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack rarefaction
all <- specaccum(seq3ed.c, method = "rarefaction", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Coastal nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(seq3ed.c, method = "rarefaction", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Coastal nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#adiv (binario)
#Shannon
#diversidade geral
HShannon <- diversityresult(seq3ed.c, index='Shannon', method='pooled') #diversidade total por todos os pontos
HShannon

#diversidade media a cada ponto
h1Shannon <- diversityresult(seq3ed.c, index='Shannon', method='mean') #diversidade por ponto
h1Shannon

#diversidade por ponto
H <- diversity(seq3ed.c, index="shannon", base=exp(1))
View(H)
summary(H)

#Simpson
#diversidade geral
SSimpson <- diversityresult(seq3ed.c, index='Simpson', method='pooled') #diversidade total por todos os pontos
SSimpson

#diversidade media a cada ponto
s1Simpson <- diversityresult(seq3ed.c, index='Simpson', method='mean') #diversidade por ponto
s1Simpson

#diversidade por ponto
S <- diversity(seq3ed.c, index="simpson", base=exp(1))
View(S)
summary(S)

#Bdiv (binario)
#betadiver(help=TRUE)
#indice Jaccard
Jac <- designdist(seq3ed.c, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac)
scores(Jac)
quantile(Jac)
mean(Jac)

Jac.cluster <- hclust(Jac, method='average') #upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.cluster, 3)
Jac.cof <- cophenetic(Jac.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante

```

```

Jac.cof
plot(Jac.cof)
cor(Jac, Jac.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.738

#decostand
seq3ed.c.stan <- decostand(seq3ed.c, method='stand') #funcao para padronizar dados, para deixar tudo em escalas proximas, no caso o satnd padroniza para
atingir a distribuicao normal padrao
seq3ed.c.stan

#indice Jaccard STAND
Jac.stand <- designdist(seq3ed.c.stan, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac.stand)
scores(Jac.stand)
quantile(Jac.stand)
mean(Jac.stand)

Jac.stand.cluster <- hclust(Jac.stand, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.stand.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.stand.cluster, 3)
Jac.stand.cof<- cophenetic(Jac.stand.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do
vizinho completo ou mais distante
Jac.stand.cof
plot(Jac.stand.cof)
cor(Jac.stand, Jac.stand.cof)

#indice Sorensen
Sor <- designdist(seq3ed.c, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor)
scores(Sor)
quantile(Sor)
mean(Sor)

Sor.cluster <- hclust(Sor, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.cluster, 3)
Sor.cof<- cophenetic(Sor.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Sor.cof
plot(Sor.cof)
cor(Sor, Sor.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.704

#indice Sorensen STAND
Sor.stand <- designdist(seq3ed.c.stan, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor.stand)
scores(Sor.stand)
quantile(Sor.stand)
mean(Sor.stand)

Sor.stand.cluster <- hclust(Sor.stand, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.stand.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.stand.cluster, 3)
Sor.stand.cof<- cophenetic(Sor.stand.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do
vizinho completo ou mais distante
Sor.stand.cof
plot(Sor.stand.cof)
cor(Sor.stand, Sor.stand.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver
maior correlacao, no caso 0.704

# passo 3 - matrizes de ordenacao
#PCA
iris.pca <- PCA(seq3ed.c, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

plot(iris.pca, display = "sites")
fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=seq3.c$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=seq3.c$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=seq3.c$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(1,3), label="var", habillage=seq3.c$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(2,3), label="var", habillage=seq3.c$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()

```

```

#PCA stand
iris.pca.1 <- PCA(seq3ed.c.stan, scale.unit = FALSE, graph = FALSE)
iris.pca.1
loadings(iris.pca.1) #verifica os pesos encontrados
summary(iris.pca.1)
attributes(iris.pca.1)

plot(iris.pca.1, display = "sites")
fviz_pca_ind(iris.pca.1, label="none")
fviz_pca_ind(iris.pca.1, label="none", habillage=seq3.c$Samples)
fviz_pca_ind(iris.pca.1, label="none", habillage=seq3.c$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca.1, axes = c(1,2), label="var", habillage=seq3.c$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(1,3), label="var", habillage=seq3.c$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(2,3), label="var", habillage=seq3.c$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5,
pointsize = 4) + theme_minimal()

#NMDS - (distancia Jaccard - biotico)
seq3ed.c.mds.Jac<- metaMDS(seq3ed.c, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq3ed.c.mds.Jac
attributes(seq3ed.c.mds.Jac)
plot(seq3ed.c.mds.Jac, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq3ed.c.mds.Jac, display='sites')
points(seq3ed.c.mds.Jac$points[1:2,], col=1, cex=3)
points(seq3ed.c.mds.Jac$points[3:4,], col=1, cex=3)
points(seq3ed.c.mds.Jac$points[5:7,], col=2, cex=3)
points(seq3ed.c.mds.Jac$points[8:10,], col=3, cex=3)
points(seq3ed.c.mds.Jac$points[11:13,], col=4, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))
legend(locator(1), c('stress = 0.0900'))

#NMDS stand
seq3ed.c.mds.Jac.1<- metaMDS(seq3ed.c.stan, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq3ed.c.mds.Jac.1
attributes(seq3ed.c.mds.Jac.1)
plot(seq3ed.c.mds.Jac.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq3ed.c.mds.Jac.1, display='sites')
points(seq3ed.c.mds.Jac.1$points[1:2,], col=1, cex=3)
points(seq3ed.c.mds.Jac.1$points[3:4,], col=1, cex=3)
points(seq3ed.c.mds.Jac.1$points[5:7,], col=2, cex=3)
points(seq3ed.c.mds.Jac.1$points[8:10,], col=3, cex=3)
points(seq3ed.c.mds.Jac.1$points[11:13,], col=4, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))
legend(locator(1), c('stress = 0.0954'))

##### 3.00µm
##### ARCHIPELAGO

#restart R

seq3.a <- read.table("~/Documents/R operacional/seq 3.00 illumina archipelago.txt", header = T, dec=',')

# paso 1
#entendendo a disposição dos dados:
View(seq3.a) #lendo no RStudio
seq3.a #lendo no console
attach(seq3.a) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(seq3.a) #resumo pra saber as significancias e valores da planilha
attributes(seq3.a) #mostra os criterios

seq3ed.a.1 <- seq3.a[,-1]
seq3ed.a.1

taxa.off <- colSums(seq3ed.a.1)
taxa.off

seq3ed.a <- seq3ed.a.1[, which(colSums(seq3ed.a.1) != 0)]
View(seq3ed.a)

n.taxa.off <- ncol(seq3ed.a.1)-ncol(seq3ed.a)
n.taxa.off

# paso 2 - informações ecologicas
# riqueza total
Stotal<- diversityresult (seq3ed.a, index='richness', method='pooled') #abundancia total por todos os pontos, utiliza a tabela onde tem os dados crus da
quantidade de especies por ponto

```

```

Stotal
summary(Stotal)

# riqueza media
Slocal<- diversityresult(seq3ed.a, index='richness', method='mean') #abundancia por ponto
Slocal
summary(Slocal)

# riqueza observada por ponto
A <- specnumber(seq3ed.a)
View(A)

#riqueza estimada
#estimadores de riqueza por ponto
estimateR(seq3ed.a)

#estimadores de riqueza total
specpool(seq3ed.a)

#chao1 (abundancia)
C1 <- chao1(seq3ed.a, taxa.row = FALSE)
C1

#chao2 (presenca e ausencia)
C2 <- chao2(seq3ed.a, taxa.row = FALSE)
C2

#jacknife 1 e 2 (binario e abundancia)
J1 <- jack1(seq3ed.a, taxa.row = FALSE, abund = FALSE)
J1

J2 <- jack2(seq3ed.a, taxa.row = FALSE, abund = FALSE)
J2

#rarefacao
quantile(rowSums(seq3ed.a)) #numero de amostras (amostargens) necessarias para chegar o mais perto do 100% do numero total de otus

#menor abundancia dentre todas as amostras - nocao para igualar as diferentes amsotragens
raremax <- min(rowSums(seq3ed.a))
raremax#590

#numero de especies rarefeitas (que nao estao na amostra, especies raras)
R1 <- rarefy(seq3ed.a, raremax)#590, nesse caso dessa planilha - menor numero observado de sequencias medias. Comparacao com esforcos diferentes (O tamanho da subamostra para a comunidade mais pequena, seja um valor unico ou um vetor)
View(R1)
summary(R1)
plot(R1)

#curva de rarefacao por ponto - rarefacao
col <- c("black", "darkred", "forestgreen", "orange", "blue", "yellow", "hotpink")
lty <- c("solid", "dashed", "longdash", "dotdash")
pars <- expand.grid(col = col, lty = lty, stringsAsFactors = FALSE)
head(pars)

out <- with(pars[1:24, ],
  rarecurve(seq3ed.a, step = 20, sample = raremax, display=c('sites'), col=c(5,5,5,5,6,6,6,6),
    lty = lty, label = FALSE))
legend(locator(1), c('AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(5,5,5,5,6,6,6,6))

#curva de rarefacao total - Chao
all <- specaccum(seq3ed.a, method = "random", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Insular nanophotoplankton - 3,00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Chao rarefaction
all <- specaccum(seq3ed.a, method = "rarefaction", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Insular nanophotoplankton - 3,00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack random
all <- specaccum(seq3ed.a, method = "random", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Insular nanophotoplankton - 3,00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")

```

```

boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(seq3ed.a, method = "random", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Insular nanophotoplankton - 3,00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack rarefaction
all <- specaccum(seq3ed.a, method = "rarefaction", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Insular nanophotoplankton - 3,00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(seq3ed.a, method = "rarefaction", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Insular nanophotoplankton - 3,00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#adiv (binario)
#Shannon
#diversidade geral
HShannon <- diversityresult(seq3ed.a, index='Shannon', method='pooled') #diversidade total por todos os pontos
HShannon

#diversidade media a cada ponto
h1Shannon <- diversityresult(seq3ed.a, index='Shannon', method='mean') #diversidade por ponto
h1Shannon

#diversidade por ponto
H <- diversity(seq3ed.a, index="shannon", base=exp(1))
View(H)
summary(H)

#Simpson
#diversidade geral
SSimpson <- diversityresult(seq3ed.a, index='Simpson', method='pooled') #diversidade total por todos os pontos
SSimpson

#diversidade media a cada ponto
s1Simpson <- diversityresult(seq3ed.a, index='Simpson', method='mean') #diversidade por ponto
s1Simpson

#diversidade por ponto
S <- diversity(seq3ed.a, index="simpson", base=exp(1))
View(S)
summary(S)

#Bdiv (binario)
#betadiver(help=TRUE)
#indice Jaccard
Jac <- designdist(seq3ed.a, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac)
scores(Jac)
quantile(Jac)
mean(Jac)

Jac.cluster <- hclust(Jac, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.cluster, 3)
Jac.cof<- cophenetic(Jac.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Jac.cof
plot(Jac.cof)
cor(Jac, Jac.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.738

#decostand
seq3ed.a.stan <- decostand(seq3ed.a, method='stand') #funcao para padronizar dados, para deixar tudo em escalas proximas, no caso o satnd padroniza para
atingir a distribuicao normal padrao
seq3ed.a.stan

#indice Jaccard STAND
Jac.stand <- designdist(seq3ed.a.stan, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac.stand)
scores(Jac.stand)

```

```
quantile(Jac.stand)
mean(Jac.stand)
```

```
Jac.stand.cluster <- hclust(Jac.stand, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.stand.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.stand.cluster, 3)
Jac.stand.cof<- cophenetic(Jac.stand.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante
Jac.stand.cof
plot(Jac.stand.cof)
cor(Jac.stand, Jac.stand.cof)
```

```
#indice Sorensen
Sor <- designdist(seq3ed.a, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor)
scores(Sor)
quantile(Sor)
mean(Sor)
```

```
Sor.cluster <- hclust(Sor, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.cluster, 3)
Sor.cof<- cophenetic(Sor.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante
Sor.cof
plot(Sor.cof)
cor(Sor, Sor.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior correlacao, no caso 0.704
```

```
#indice Sorensen STAND
Sor.stand <- designdist(seq3ed.a.stan, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor.stand)
scores(Sor.stand)
quantile(Sor.stand)
mean(Sor.stand)
```

```
Sor.stand.cluster <- hclust(Sor.stand, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.stand.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.stand.cluster, 3)
Sor.stand.cof<- cophenetic(Sor.stand.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante
Sor.stand.cof
plot(Sor.stand.cof)
cor(Sor.stand, Sor.stand.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior correlacao, no caso 0.704
```

```
# passo 3 - matrizes de ordenacao
#PCA
iris.pca <- PCA(seq3ed.a, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)
```

```
plot(iris.pca, display = "sites")
fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=seq3.a$$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=seq3.a$$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=seq3.a$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(1,3), label="var", habillage=seq3.a$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(2,3), label="var", habillage=seq3.a$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5,
pointsize = 4) + theme_minimal()
```

```
#PCA stand
iris.pca.1 <- PCA(seq3ed.a.stan, scale.unit = FALSE, graph = FALSE)
iris.pca.1
loadings(iris.pca.1) #verifica os pesos encontrados
summary(iris.pca.1)
attributes(iris.pca.1)
```

```
plot(iris.pca.1, display = "sites")
fviz_pca_ind(iris.pca.1, label="none")
fviz_pca_ind(iris.pca.1, label="none", habillage=seq3.a$$Samples)
fviz_pca_ind(iris.pca.1, label="none", habillage=seq3.a$$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca.1, axes = c(1,2), label="var", habillage=seq3.a$$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsize = 5,
pointsize = 4) + theme_minimal()
```

```
fvis_pca_biplot(iris.pca.1, axes = c(1,3), label="var", habillage=seq3.a$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
fvis_pca_biplot(iris.pca.1, axes = c(2,3), label="var", habillage=seq3.a$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
```

```
#NMDS - (distancia Jaccard - biotico)
seq3ed.a.mds.Jac<- metaMDS(seq3ed.a, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq3ed.a.mds.Jac
attributes(seq3ed.a.mds.Jac)
plot(seq3ed.a.mds.Jac, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq3ed.a.mds.Jac, display='sites')
points(seq3ed.a.mds.Jac$points[1:3,], col=5, cex=3)
points(seq3ed.a.mds.Jac$points[4:6,], col=5, cex=3)
points(seq3ed.a.mds.Jac$points[7:9,], col=5, cex=3)
points(seq3ed.a.mds.Jac$points[10:12,], col=5, cex=3)
points(seq3ed.a.mds.Jac$points[13:15,], col=6, cex=3)
points(seq3ed.a.mds.Jac$points[16:18,], col=6, cex=3)
points(seq3ed.a.mds.Jac$points[19:21,], col=6, cex=3)
points(seq3ed.a.mds.Jac$points[22:24,], col=6, cex=3)
legend(locator(1), c('AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(5,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.1168'))
```

```
#NMDS stand
seq3ed.a.mds.Jac.1<- metaMDS(seq3ed.a.stan, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq3ed.a.mds.Jac.1
attributes(seq3ed.a.mds.Jac.1)
plot(seq3ed.a.mds.Jac.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq3ed.a.mds.Jac.1, display='sites')
points(seq3ed.a.mds.Jac.1$points[1:3,], col=5, cex=3)
points(seq3ed.a.mds.Jac.1$points[4:6,], col=5, cex=3)
points(seq3ed.a.mds.Jac.1$points[7:9,], col=5, cex=3)
points(seq3ed.a.mds.Jac.1$points[10:12,], col=5, cex=3)
points(seq3ed.a.mds.Jac.1$points[13:15,], col=6, cex=3)
points(seq3ed.a.mds.Jac.1$points[16:18,], col=6, cex=3)
points(seq3ed.a.mds.Jac.1$points[19:21,], col=6, cex=3)
points(seq3ed.a.mds.Jac.1$points[22:24,], col=6, cex=3)
legend(locator(1), c('AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(5,5,5,5,6,6,6,6))
legend(locator(1), c('stress = ?'))
```

## Sequencias e índices ecológicos 0,22µm Illumina

```
##### 0.22µm
##### All - seq
```

```
seq0.22 <- read.table("~/Documents/R operacional/seq 0.22 illumina.txt", header = T, dec=',')
```

```
##### paso 1 - entendendo a disposição dos dados All 0.22µm
View(seq0.22) #lendo no RStudio
seq0.22 #lendo no console
attach(seq0.22) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(seq0.22) #resumo pra saber as significancias e valores da planilha
attributes(seq0.22) #mostra os criterios
```

```
#Syneco
```

```
#graficos de media com desvio padrao, funcao BOXPLOT
par(mar = c(5,5,2,2), cex.lab = 1.5)
boxplot(seq0.22$Synecococcus) #media e desvios do tamanho apenas
boxplot(seq0.22$Synecococcus~seq0.22$Samples) #da um ~ com um espaco, eh um comando que da a media e o desvio de uma variavel por outra variavel,
o ~ separa os dados
```

```
shapiro.test(seq0.22$Synecococcus) #teste de normalidade p<0.05 para ser normal
#deu normal (tem de transformar em log, e agora usar essa planilha transformada pra tudo)
#seq0.22stan.syn <- decostand(seq0.22$Synecococcus, "standardize")
#View(seq0.22stan.syn)
#shapiro.test(seq0.22stan.syn)
```

```
#ANOVA unifactorial
ANOVA.syn<-aov(seq0.22$Synecococcus~seq0.22$Samples, seq0.22) #fit e o nome que eu dei para aov (analysis of variance) do que eu quero
ANOVA.syn
anova(ANOVA.syn) #Análise de variancia, probabilidades
TukeyHSD(ANOVA.syn) #Teste Tukey
plot(ANOVA.syn)
plot(TukeyHSD(ANOVA.syn))
```

```
#Prochloro
```

```
#graficos de media com desvio padrao, funcao BOXPLOT
par(mar = c(5,5,2,2), cex.lab = 1.5)
```

```

boxplot(seq0.22$Prochlorococcus) #media e desvios do tamanho apenas
boxplot(seq0.22$Prochlorococcus~seq0.22$Samples) #da um ~ com um espaco, eh um comando que da a media e o desvio de uma variavel por outra variavel,
o ~ separa os dados

shapiro.test(seq0.22$Prochlorococcus) #teste de normalidade p<0,05 para ser normal
#deu normal - tem de transformar e usar essa planilha transformada pra tudo
#seq0.22stan.pro <- decostand(seq0.22$Prochlorococcus, "standardize")
#View(seq0.22stan.pro)
#shapiro.test(seq0.22stan.pro)

#ANOVA unifactorial
ANOVA.pro<-aov(seq0.22$Prochlorococcus~seq0.22$Samples, seq0.22) #fit e o nome que eu dei para aov (analysis of variance) do que eu quero
ANOVA.pro
anova(ANOVA.pro) #An?lise de variancia, probabilidades
TukeyHSD(ANOVA.pro) #Teste Tukey
plot(ANOVA.pro)
plot(TukeyHSD(ANOVA.pro))

#All
#ajustes necessarios p/ estatisticas
seq0.22ed <- seq0.22[, -1]
View(seq0.22ed)

seq0.22edstand <- decostand(seq0.22ed, "standardize")
View(seq0.22edstand)
shapiro.test(seq0.22edstand$Synecococcus)
shapiro.test(seq0.22edstand$Prochlorococcus)

#ANOVA multifatorial
fait<-aov(seq0.22ed$Prochlorococcus~seq0.22ed$Synecococcus*seq0.22ed$Prochlorococcus, seq0.22ed)
anova(fait)
TukeyHSD(fait)

frait<-aov(seq0.22edstand$Prochlorococcus~seq0.22edstand$Synecococcus*seq0.22edstand$Prochlorococcus, seq0.22edstand)
anova(frait)
TukeyHSD(frait)

# paso 2 - matriz de dissimilaridade (agrupamento) - distancia euclidiana com dados bioticos:
#SEM STAND
seq0.22ed.eu.1<-vegdist(seq0.22ed, method='eu', binary=TRUE, diag = F, upper = F) #apenas dei um nome, indice de similaridade (distância euclidiana pura
para analises quantitativas)
seq0.22ed.eu.1
seq0.22ed.eu.cluster.1<-hclust(seq0.22ed.eu.1, method = "average") # nomeando e gerando a analise de cluster
seq0.22ed.eu.cluster.1
plot(seq0.22ed.eu.cluster.1, hang = -1) #gera o grafico de cluster
rect.hclust(seq0.22ed.eu.cluster.1, 4)
seq0.22ed.eu.cluster.1.cof<- cophenetic(seq0.22ed.eu.cluster.1) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves
do metodo do vizinho completo ou mais distante
seq0.22ed.eu.cluster.1.cof
plot(seq0.22ed.eu.cluster.1.cof)
cor(seq0.22ed.eu.1, seq0.22ed.eu.cluster.1.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais
adequado, eh o que tiver maior correlacao, no caso 0.704

#STAND
seq0.22edstand.eu.1<-vegdist(seq0.22edstand, method='eu', binary=TRUE, diag = F, upper = F) #apenas dei um nome, indice de similaridade (distância
euclidiana pura para analises quantitativas)
seq0.22edstand.eu.1
seq0.22edstand.eu.cluster.1<-hclust(seq0.22edstand.eu.1, method = "average") # nomeando e gerando a analise de cluster
seq0.22edstand.eu.cluster.1
plot(seq0.22edstand.eu.cluster.1, hang = -1) #gera o grafico de cluster
rect.hclust(seq0.22edstand.eu.cluster.1, 3)
seq0.22edstand.eu.cluster.1.cof<- cophenetic(seq0.22edstand.eu.cluster.1) #gera uma matriz ficticia com base na analise grafica da matriz de distancia
observada atraves do metodo do vizinho completo ou mais distante
seq0.22edstand.eu.cluster.1.cof
plot(seq0.22edstand.eu.cluster.1.cof)
cor(seq0.22edstand.eu.1, seq0.22edstand.eu.cluster.1.cof) #cor

# paso 3
#PCA SEM STAND - Usar esse:
iris.pca <- PCA(seq0.22ed, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=seq0.22$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=seq0.22$Samples, addEllipses=TRUE, ellipse.level=0.95)

```

```
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=seq0.22$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
```

```
#PCA stand
```

```
iris.pca.1 <- PCA(seq0.22edstand, scale.unit = FALSE, graph = FALSE)
iris.pca.1
loadings(iris.pca.1) #verifica os pesos encontrados
summary(iris.pca.1)
attributes(iris.pca.1)
```

```
fviz_pca_ind(iris.pca.1, label="none")
```

```
fviz_pca_ind(iris.pca.1, label="none", habillage=seq0.22$Samples)
```

```
fviz_pca_ind(iris.pca.1, label="none", habillage=seq0.22$Samples, addEllipses=TRUE, ellipse.level=0.95)
```

```
fviz_pca_biplot(iris.pca.1, axes = c(1,2), label="var", habillage=seq0.22$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
```

```
fviz_pca_biplot(iris.pca.1, axes = c(1,3), label="var", habillage=seq0.22$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
```

```
fviz_pca_biplot(iris.pca.1, axes = c(2,3), label="var", habillage=seq0.22$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
```

```
#MDS - (distancia euclidiana metrica - abiotico)
```

```
seq0.22.mds.eu <- metaMDS(seq0.22ed, distance='eu', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
```

```
seq0.22.mds.eu
```

```
attributes(seq0.22.mds.eu)
```

```
plot(seq0.22.mds.eu, cex=3, display=c('sites')) #deixa apenas os pontos
```

```
text(seq0.22.mds.eu, display='sites')
```

```
points(seq0.22.mds.eu$points[1:2,], col=1, cex=3)
```

```
points(seq0.22.mds.eu$points[3:4,], col=1, cex=3)
```

```
points(seq0.22.mds.eu$points[5:7,], col=2, cex=3)
```

```
points(seq0.22.mds.eu$points[8:10,], col=3, cex=3)
```

```
points(seq0.22.mds.eu$points[11:13,], col=4, cex=3)
```

```
points(seq0.22.mds.eu$points[14:16,], col=5, cex=3)
```

```
points(seq0.22.mds.eu$points[17:19,], col=5, cex=3)
```

```
points(seq0.22.mds.eu$points[20:22,], col=5, cex=3)
```

```
points(seq0.22.mds.eu$points[23:25,], col=5, cex=3)
```

```
points(seq0.22.mds.eu$points[26:28,], col=6, cex=3)
```

```
points(seq0.22.mds.eu$points[29:31,], col=6, cex=3)
```

```
points(seq0.22.mds.eu$points[32:34,], col=6, cex=3)
```

```
points(seq0.22.mds.eu$points[35:37,], col=6, cex=3)
```

```
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,5,6,6,6,6))
```

```
legend(locator(1), c('stress = 0.1017'))
```

```
#MDS stand
```

```
seq0.22.mds.eu.1 <- metaMDS(seq0.22edstand, distance='eu', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
```

```
seq0.22.mds.eu.1
```

```
attributes(seq0.22.mds.eu.1)
```

```
plot(seq0.22.mds.eu.1, cex=3, display=c('sites')) #deixa apenas os pontos
```

```
text(seq0.22.mds.eu.1, display='sites')
```

```
points(seq0.22.mds.eu.1$points[1:2,], col=1, cex=3)
```

```
points(seq0.22.mds.eu.1$points[3:4,], col=1, cex=3)
```

```
points(seq0.22.mds.eu.1$points[5:7,], col=2, cex=3)
```

```
points(seq0.22.mds.eu.1$points[8:10,], col=3, cex=3)
```

```
points(seq0.22.mds.eu.1$points[11:13,], col=4, cex=3)
```

```
points(seq0.22.mds.eu.1$points[14:16,], col=5, cex=3)
```

```
points(seq0.22.mds.eu.1$points[17:19,], col=5, cex=3)
```

```
points(seq0.22.mds.eu.1$points[20:22,], col=5, cex=3)
```

```
points(seq0.22.mds.eu.1$points[23:25,], col=5, cex=3)
```

```
points(seq0.22.mds.eu.1$points[26:28,], col=6, cex=3)
```

```
points(seq0.22.mds.eu.1$points[29:31,], col=6, cex=3)
```

```
points(seq0.22.mds.eu.1$points[32:34,], col=6, cex=3)
```

```
points(seq0.22.mds.eu.1$points[35:37,], col=6, cex=3)
```

```
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,5,6,6,6,6))
```

```
legend(locator(1), c('stress = 0.0689'))
```

```
##### 0.22µm
```

```
##### All - eco
```

```
#restart R
```

```
# paso 1 - informacoes preliminares
```

```
seq0.22 <- read.table("~/Documents/R operacional/seq 0.22 illumina.txt", header = T, dec=',')
```

```
seq0.22ed <- seq0.22[,-1]
```

```
View(seq0.22)
```

```
View(seq0.22ed)
```

```
# paso 2 - informações ecologicas All - eco 0.22µm
```

```
# riqueza total
```

```

Stotal<- diversityresult (seq0.22ed, index='richness', method='pooled') #abundancia total por todos os pontos, utiliza a tabela onde tem os dados crus da
quantidade de especies por ponto
Stotal
summary(Stotal)

# riqueza media
Slocal<- diversityresult (seq0.22ed, index='richness', method='mean') #abundancia por ponto
Slocal
summary(Slocal)

# riqueza observada por ponto
A <- specnumber(seq0.22ed)
View(A)

#riqueza estimada
#estimadores de riqueza por ponto
estimateR(seq0.22ed)

#estimadores de riqueza total
specpool(seq0.22ed)

#chao1 (abundancia)
C1 <- chao1(seq0.22ed, taxa.row = FALSE)
C1

#chao2 (presenca e ausencia)
C2 <- chao2(seq0.22ed, taxa.row = FALSE)
C2

#jacknife 1 e 2 (binario e abundancia)
J1 <- jack1(seq0.22ed, taxa.row = FALSE, abund = FALSE)
J1

J2 <- jack2(seq0.22ed, taxa.row = FALSE, abund = FALSE)
J2

#rarefacao
quantile(rowSums(seq0.22ed)) #numero de amostras (amostargens) necessarias para chegar o mais perto do 100% do numero total de otus

#menor abundancia dentre todas as amostras - noca para igualar as diferentes amsotragens
raremax <- min(rowSums(seq0.22ed))
raremax#3309

#numero de especies rarefeitas (que nao estao na amostra, especies raras)
R1 <- rarefy(seq0.22ed, raremax)#3309, nesse caso dessa planilha - menor numero observado de sequencias medias. Comparacao com esforcos diferentes (O
tamanho da subamostra para a comunidade mais pequena, seja um valor unico ou um vetor)
View(R1)
summary(R1)
plot(R1)

#curva de rarefacao por ponto - rarefacao
col <- c("black", "darkred", "forestgreen", "orange", "blue", "yellow", "hotpink")
lty <- c("solid", "dashed", "longdash", "dotdash")
pars <- expand.grid(col = col, lty = lty, stringsAsFactors = FALSE)
head(pars)

out <- with(pars[1:37, ],
  rarecurve(seq0.22ed, step = 20, sample = raremax, display=c('sites'), col=c(1,1,2,3,4,5,5,5,5,6,6,6,6),
    lty = lty, label = FALSE))
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,6,6,6,6))

#curva de rarefacao total - Chao
all <- specaccum(seq0.22ed, method = "random", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Chao rarefaction
all <- specaccum(seq0.22ed, method = "rarefaction", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack random
all <- specaccum(seq0.22ed, method = "random", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),

```

```

ci.col = "lightblue", main = "Nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(seq0.22ed, method = "random", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
ci.col = "lightblue", main = "Nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack rarefaction
all <- specaccum(seq0.22ed, method = "rarefaction", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
ci.col = "lightblue", main = "Nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(seq0.22ed, method = "rarefaction", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
ci.col = "lightblue", main = "Nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#adiv (binario)
#Shannon
#diversidade geral
HShannon <- diversityresult(seq0.22ed, index='Shannon', method='pooled') #diversidade total por todos os pontos
HShannon

#diversidade media a cada ponto
h1Shannon <- diversityresult(seq0.22ed, index='Shannon', method='mean') #diversidade por ponto
h1Shannon

#diversidade por ponto
H <- diversity(seq0.22ed, index="shannon", base=exp(1))
View(H)
summary(H)

#Simpson
#diversidade geral
SSimpson <- diversityresult(seq0.22ed, index='Simpson', method='pooled') #diversidade total por todos os pontos
SSimpson

#diversidade media a cada ponto
s1Simpson <- diversityresult(seq0.22ed, index='Simpson', method='mean') #diversidade por ponto
s1Simpson

#diversidade por ponto
S <- diversity(seq0.22ed, index="simpson", base=exp(1))
View(S)
summary(S)

#Bdiv (binario)
#betadiver(help=TRUE)
#indice Jaccard
Jac <- designdist(seq0.22ed, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac)
scores(Jac)
quantile(Jac)
mean(Jac)

Jac.cluster <- hclust(Jac, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.cluster, 3)
Jac.cof<- cophenetic(Jac.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Jac.cof
plot(Jac.cof)
cor(Jac, Jac.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.738

#decostand
seq0.22ed.stan <- decostand(seq0.22ed, method='stand') #funcao para padronizar dados, para deixar tudo em escalas proximas, no caso o satnd padroniza para
atingir a distribuicao normal padrao
seq0.22ed.stan

#indice Jaccard STAND
Jac.stand <- designdist(seq0.22ed.stan, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)

```

```

plot(Jac.stand)
scores(Jac.stand)
quantile(Jac.stand)
mean(Jac.stand)

```

```

Jac.stand.cluster <- hclust(Jac.stand, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.stand.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.stand.cluster, 3)
Jac.stand.cof<- cophenetic(Jac.stand.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante
Jac.stand.cof
plot(Jac.stand.cof)
cor(Jac.stand, Jac.stand.cof)

```

```

#indice Sorensen
Sor <- designdist(seq0.22ed, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor)
scores(Sor)
quantile(Sor)
mean(Sor)

```

```

Sor.cluster <- hclust(Sor, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.cluster, 3)
Sor.cof<- cophenetic(Sor.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante
Sor.cof
plot(Sor.cof)
cor(Sor, Sor.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior correlacao, no caso 0.704

```

```

#indice Sorensen STAND
Sor.stand <- designdist(seq0.22ed.stan, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor.stand)
scores(Sor.stand)
quantile(Sor.stand)
mean(Sor.stand)

```

```

Sor.stand.cluster <- hclust(Sor.stand, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.stand.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.stand.cluster, 3)
Sor.stand.cof<- cophenetic(Sor.stand.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante
Sor.stand.cof
plot(Sor.stand.cof)
cor(Sor.stand, Sor.stand.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior correlacao, no caso 0.704

```

# passo 3 - matrizes de ordenacao

```

#PCA
iris.pca <- PCA(seq0.22ed, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

```

```

plot(iris.pca, display = "sites")
fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=seq0.22$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=seq0.22$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=seq0.22$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(1,3), label="var", habillage=seq0.22$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(2,3), label="var", habillage=seq0.22$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()

```

```

#PCA stand
iris.pca.1 <- PCA(seq0.22ed.stan, scale.unit = FALSE, graph = FALSE)
iris.pca.1
loadings(iris.pca.1) #verifica os pesos encontrados
summary(iris.pca.1)
attributes(iris.pca.1)

```

```

plot(iris.pca.1, display = "sites")
fviz_pca_ind(iris.pca.1, label="none")
fviz_pca_ind(iris.pca.1, label="none", habillage=seq0.22$Samples)
fviz_pca_ind(iris.pca.1, label="none", habillage=seq0.22$Samples, addEllipses=TRUE, ellipse.level=0.95)

```

```
fviz_pca_biplot(iris.pca.1, axes = c(1,2), label="var", habillage=seq0.22$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(1,3), label="var", habillage=seq0.22$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(2,3), label="var", habillage=seq0.22$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
```

```
#NMDS - (distancia Jaccard - biotico)
seq0.22ed.mds.Jac<- metaMDS(seq0.22ed, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq0.22ed.mds.Jac
attributes(seq0.22ed.mds.Jac)
plot(seq0.22ed.mds.Jac, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq0.22ed.mds.Jac, display='sites')
points(seq0.22ed.mds.Jac$points[1:2,], col=1, cex=3)
points(seq0.22ed.mds.Jac$points[3:4,], col=1, cex=3)
points(seq0.22ed.mds.Jac$points[5:7,], col=2, cex=3)
points(seq0.22ed.mds.Jac$points[8:10,], col=3, cex=3)
points(seq0.22ed.mds.Jac$points[11:13,], col=4, cex=3)
points(seq0.22ed.mds.Jac$points[14:16,], col=5, cex=3)
points(seq0.22ed.mds.Jac$points[17:19,], col=5, cex=3)
points(seq0.22ed.mds.Jac$points[20:22,], col=5, cex=3)
points(seq0.22ed.mds.Jac$points[23:25,], col=5, cex=3)
points(seq0.22ed.mds.Jac$points[26:28,], col=6, cex=3)
points(seq0.22ed.mds.Jac$points[29:31,], col=6, cex=3)
points(seq0.22ed.mds.Jac$points[32:34,], col=6, cex=3)
points(seq0.22ed.mds.Jac$points[35:37,], col=6, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.1812'))
```

```
#NMDS stand
seq0.22ed.mds.Jac.1<- metaMDS(seq0.22ed.stan, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq0.22ed.mds.Jac.1
attributes(seq0.22ed.mds.Jac.1)
plot(seq0.22ed.mds.Jac.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq0.22ed.mds.Jac.1, display='sites')
points(seq0.22ed.mds.Jac.1$points[1:2,], col=1, cex=3)
points(seq0.22ed.mds.Jac.1$points[3:4,], col=1, cex=3)
points(seq0.22ed.mds.Jac.1$points[5:7,], col=2, cex=3)
points(seq0.22ed.mds.Jac.1$points[8:10,], col=3, cex=3)
points(seq0.22ed.mds.Jac.1$points[11:13,], col=4, cex=3)
points(seq0.22ed.mds.Jac.1$points[14:16,], col=5, cex=3)
points(seq0.22ed.mds.Jac.1$points[17:19,], col=5, cex=3)
points(seq0.22ed.mds.Jac.1$points[20:22,], col=5, cex=3)
points(seq0.22ed.mds.Jac.1$points[23:25,], col=5, cex=3)
points(seq0.22ed.mds.Jac.1$points[26:28,], col=6, cex=3)
points(seq0.22ed.mds.Jac.1$points[29:31,], col=6, cex=3)
points(seq0.22ed.mds.Jac.1$points[32:34,], col=6, cex=3)
points(seq0.22ed.mds.Jac.1$points[35:37,], col=6, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,5,6,6,6,6))
legend(locator(1), c('stress = ?'))
```

```
##### 0.22µm
##### COAST
```

```
#restart R
```

```
seq0.22.c <- read.table("~/Documents/R operacional/seq 0.22 ilumina coast.txt", header = T, dec=',')
```

```
# passo 1
#entendendo a disposição dos dados:
View(seq0.22.c) #lendo no RStudio
seq0.22.c #lendo no console
attach(seq0.22.c) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(seq0.22.c) #resumo pra saber as significancias e valores da planilha
attributes(seq0.22.c) #mostra os criterios
```

```
seq0.22ed.c.1 <- seq0.22.c[,-1]
seq0.22ed.c.1
```

```
taxa.off <- colSums(seq0.22ed.c.1)
taxa.off
```

```
seq0.22ed.c <- seq0.22ed.c.1[, which(colSums(seq0.22ed.c.1) != 0)]
View(seq0.22ed.c)
```

```
n.taxa.off <- ncol(seq0.22ed.c.1)-ncol(seq0.22ed.c)
n.taxa.off
```

```

# paso 2 - informações ecologicas
# riqueza total
Stotal<- diversityresult (seq0.22ed.c, index='richness', method='pooled') #abundancia total por todos os pontos, utiliza a tabela onde tem os dados crus da
quantidade de especies por ponto
Stotal
summary(Stotal)

# riqueza media
Slocal<- diversityresult (seq0.22ed.c, index='richness', method='mean') #abundancia por ponto
Slocal
summary(Slocal)

# riqueza observada por ponto
A <- specnumber(seq0.22ed.c)
View(A)

#riqueza estimada
#estimadores de riqueza por ponto
estimateR(seq0.22ed.c)

#estimadores de riqueza total
specpool(seq0.22ed.c)

#chao1 (abundancia)
C1 <- chao1(seq0.22ed.c, taxa.row = FALSE)
C1

#chao2 (presenca e ausencia)
C2 <- chao2(seq0.22ed.c, taxa.row = FALSE)
C2

#jacknife 1 e 2 (binario e abundancia)
J1 <- jack1(seq0.22ed.c, taxa.row = FALSE, abund = FALSE)
J1

J2 <- jack2(seq0.22ed.c, taxa.row = FALSE, abund = FALSE)
J2

#rarefacao
quantile(rowSums(seq0.22ed.c)) #numero de amostras (amostargens) necessarias para chegar o mais perto do 100% do numero total de otus

#menor abundancia dentre todas as amostras - nocao para igualar as diferentes amostras
raremax <- min(rowSums(seq0.22ed.c))
raremax#3309

#numero de especies rarefeitas (que nao estao na amostra, especies raras)
R1 <- rarefy(seq0.22ed.c, raremax)#3309, nesse caso dessa planilha - menor numero observado de sequencias medias. Comparacao com esforcos diferentes (O
tamanho da subamostra para a comunidade mais pequena, seja um valor único ou um vetor)
View(R1)
summary(R1)
plot(R1)

#curva de rarefacao por ponto - rarefacao
col <- c("black", "darkred", "forestgreen", "orange", "blue", "yellow", "hotpink")
lty <- c("solid", "dashed", "longdash", "dotdash")
pars <- expand.grid(col = col, lty = lty, stringsAsFactors = FALSE)
head(pars)

out <- with(pars[1:13, ],
  rarecurve(seq0.22ed.c, step = 20, sample = raremax, display=c('sites'), col=c(1,1,2,3,4),
    lty = lty, label = FALSE))
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))

#curva de rarefacao total - Chao
all <- specaccum(seq0.22ed.c, method = "random", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Coastal nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Chao rarefaction
all <- specaccum(seq0.22ed.c, method = "rarefaction", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Coastal nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack random

```

```

all <- specaccum(seq0.22ed.c, method = "random", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Coastal nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(seq0.22ed.c, method = "random", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Coastal nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack rarefaction
all <- specaccum(seq0.22ed.c, method = "rarefaction", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Coastal nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(seq0.22ed.c, method = "rarefaction", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
      ci.col = "lightblue", main = "Coastal nanophytoplankton - 3.00µm (Library 16s rRNA)", xlab = "Number of Samples",
      ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#adiv (binario)
#Shannon
#diversidade geral
HShannon <- diversityresult(seq0.22ed.c, index='Shannon', method='pooled') #diversidade total por todos os pontos
HShannon

#diversidade media a cada ponto
h1Shannon <- diversityresult(seq0.22ed.c, index='Shannon', method='mean') #diversidade por ponto
h1Shannon

#diversidade por ponto
H <- diversity(seq0.22ed.c, index="shannon", base=exp(1))
View(H)
summary(H)

#Simpson
#diversidade geral
SSimpson <- diversityresult(seq0.22ed.c, index='Simpson', method='pooled') #diversidade total por todos os pontos
SSimpson

#diversidade media a cada ponto
s1Simpson <- diversityresult(seq0.22ed.c, index='Simpson', method='mean') #diversidade por ponto
s1Simpson

#diversidade por ponto
S <- diversity(seq0.22ed.c, index="simpson", base=exp(1))
View(S)
summary(S)

#Bdiv (binario)
#betadiver(help=TRUE)
#indice Jaccard
Jac <- designdist(seq0.22ed.c, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac)
scores(Jac)
quantile(Jac)
mean(Jac)

Jac.cluster <- hclust(Jac, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.cluster, 3)
Jac.cof<- cophenetic(Jac.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Jac.cof
plot(Jac.cof)
cor(Jac, Jac.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.738

#decostand
seq0.22ed.c.stan <- decostand(seq0.22ed.c, method='stand') #funcao para padronizar dados, para deixar tudo em escalas proximas, no caso o satnd padroniza
para atingir a distribuicao normal padrao
seq0.22ed.c.stan

```

```

#indice Jaccard STAND
Jac.stand <- designdist(seq0.22ed.c.stan, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac.stand)
scores(Jac.stand)
quantile(Jac.stand)
mean(Jac.stand)

Jac.stand.cluster <- hclust(Jac.stand, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.stand.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.stand.cluster, 3)
Jac.stand.cof<- cophenetic(Jac.stand.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante
Jac.stand.cof
plot(Jac.stand.cof)
cor(Jac.stand, Jac.stand.cof)

#indice Sorensen
Sor <- designdist(seq0.22ed.c, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor)
scores(Sor)
quantile(Sor)
mean(Sor)

Sor.cluster <- hclust(Sor, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.cluster, 3)
Sor.cof<- cophenetic(Sor.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante
Sor.cof
plot(Sor.cof)
cor(Sor, Sor.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior correlacao, no caso 0.704

#indice Sorensen STAND
Sor.stand <- designdist(seq0.22ed.c.stan, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor.stand)
scores(Sor.stand)
quantile(Sor.stand)
mean(Sor.stand)

Sor.stand.cluster <- hclust(Sor.stand, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.stand.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.stand.cluster, 3)
Sor.stand.cof<- cophenetic(Sor.stand.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante
Sor.stand.cof
plot(Sor.stand.cof)
cor(Sor.stand, Sor.stand.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior correlacao, no caso 0.704

# passo 3 - matrizes de ordenacao
#PCA
iris.pca <- PCA(seq0.22ed.c, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

plot(iris.pca, display = "sites")
fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=seq0.22.c$$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=seq0.22.c$$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=seq0.22.c$$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(1,3), label="var", habillage=seq0.22.c$$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(2,3), label="var", habillage=seq0.22.c$$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz = 5,
pointsize = 4) + theme_minimal()

#PCA stand
iris.pca.1 <- PCA(seq0.22ed.c.stan, scale.unit = FALSE, graph = FALSE)
iris.pca.1
loadings(iris.pca.1) #verifica os pesos encontrados
summary(iris.pca.1)
attributes(iris.pca.1)

plot(iris.pca.1, display = "sites")
fviz_pca_ind(iris.pca.1, label="none")

```

```

fviz_pca_ind(iris.pca.1, label="none", habillage=seq0.22.c$Samples)
fviz_pca_ind(iris.pca.1, label="none", habillage=seq0.22.c$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca.1, axes = c(1,2), label="var", habillage=seq0.22.c$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(1,3), label="var", habillage=seq0.22.c$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(2,3), label="var", habillage=seq0.22.c$Samples, alpha.var = "contrib", col.var = "black", invisible = "ind", labelszize = 5,
pointsize = 4) + theme_minimal()

#NMDS - (distancia Jaccard - biotico)
seq0.22ed.c.mds.Jac<- metaMDS(seq0.22ed.c, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq0.22ed.c.mds.Jac
attributes(seq0.22ed.c.mds.Jac)
plot(seq0.22ed.c.mds.Jac, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq0.22ed.c.mds.Jac, display='sites')
points(seq0.22ed.c.mds.Jac$points[1:2,], col=1, cex=3)
points(seq0.22ed.c.mds.Jac$points[3:4,], col=1, cex=3)
points(seq0.22ed.c.mds.Jac$points[5:7,], col=2, cex=3)
points(seq0.22ed.c.mds.Jac$points[8:10,], col=3, cex=3)
points(seq0.22ed.c.mds.Jac$points[11:13,], col=4, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))
legend(locator(1), c('stress = 0.1041'))

#NMDS stand
seq0.22ed.c.mds.Jac.1<- metaMDS(seq0.22ed.c.stan, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq0.22ed.c.mds.Jac.1
attributes(seq0.22ed.c.mds.Jac.1)
plot(seq0.22ed.c.mds.Jac.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq0.22ed.c.mds.Jac.1, display='sites')
points(seq0.22ed.c.mds.Jac.1$points[1:2,], col=1, cex=3)
points(seq0.22ed.c.mds.Jac.1$points[3:4,], col=1, cex=3)
points(seq0.22ed.c.mds.Jac.1$points[5:7,], col=2, cex=3)
points(seq0.22ed.c.mds.Jac.1$points[8:10,], col=3, cex=3)
points(seq0.22ed.c.mds.Jac.1$points[11:13,], col=4, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1'), pch=16, col=c(1,1,2,3,4))
legend(locator(1), c('stress = 0.0001'))

##### 0.22µm
##### ARCHIPELAGO

#restart R

seq0.22.a <- read.table ("~/Documents/R operacional/seq 0.22 illumina archipelago.txt", header = T, dec=',')

# paso 1
#entendendo a disposição dos dados:
View(seq0.22.a) #lendo no RStudio
seq0.22.a #lendo no console
attach(seq0.22.a) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
summary(seq0.22.a) #resumo pra saber as significancias e valores da planilha
attributes(seq0.22.a) #mostra os criterios

seq0.22ed.a.1 <- seq0.22.a[,-1]
seq0.22ed.a.1

taxa.off <- colSums(seq0.22ed.a.1)
taxa.off

seq0.22ed.a <- seq0.22ed.a.1[, which(colSums(seq0.22ed.a.1) != 0)]
View(seq0.22ed.a)

n.taxa.off <- ncol(seq0.22ed.a.1)-ncol(seq0.22ed.a)
n.taxa.off

# paso 2 - informações ecologicas
# riqueza total
Stotal<- diversityresult (seq0.22ed.a, index='richness', method='pooled') #abundancia total por todos os pontos, utiliza a tabela onde tem os dados crus da
quantidade de especies por ponto
Stotal
summary(Stotal)

# riqueza media
Slocal<- diversityresult (seq0.22ed.a, index='richness', method='mean') #abundancia por ponto
Slocal
summary(Slocal)

# riqueza observada por ponto
A <- specnumber(seq0.22ed.a)

```

```

View(A)

#riqueza estimada
#estimadores de riqueza por ponto
estimateR(seq0.22ed.a)

#estimadores de riqueza total
specpool(seq0.22ed.a)

#chao1 (abundancia)
C1 <- chao1(seq0.22ed.a, taxa.row = FALSE)
C1

#chao2 (presenca e ausencia)
C2 <- chao2(seq0.22ed.a, taxa.row = FALSE)
C2

#jacknife 1 e 2 (binario e abundancia)
J1 <- jack1(seq0.22ed.a, taxa.row = FALSE, abund = FALSE)
J1

J2 <- jack2(seq0.22ed.a, taxa.row = FALSE, abund = FALSE)
J2

#rarefacao
quantile(rowSums(seq0.22ed.a) #numero de amostras (amostargens) necessarias para chegar o mais perto do 100% do numero total de otus

#menor abundancia dentre todas as amostras - nocao para igualar as diferentes amostragens
raremax <- min(rowSums(seq0.22ed.a))
raremax#7613

#numero de especies rarefeitas (que nao estao na amostra, especies raras)
R1 <- rarefy(seq0.22ed.a, raremax)#7613, nesse caso dessa planilha - menor numero observado de sequencias medias. Comparacao com esforcos diferentes (O tamanho da subamostra para a comunidade mais pequena, seja um valor unico ou um vetor)
View(R1)
summary(R1)
plot(R1)

#curva de rarefacao por ponto - rarefacao
col <- c("black", "darkred", "forestgreen", "orange", "blue", "yellow", "hotpink")
lty <- c("solid", "dashed", "longdash", "dotdash")
pars <- expand.grid(col = col, lty = lty, stringsAsFactors = FALSE)
head(pars)

out <- with(pars[1:24, ],
  rarecurve(seq0.22ed.a, step = 20, sample = raremax, display=c('sites'), col=c(5,5,5,5,6,6,6,6),
    lty = lty, label = FALSE))
legend(locator(1), c('AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(5,5,5,5,6,6,6,6))

#curva de rarefacao total - Chao
all <- specaccum(seq0.22ed.a, method = "random", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Insular nanophotoplankton - 3,00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Chao rarefaction
all <- specaccum(seq0.22ed.a, method = "rarefaction", index = "chao")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Insular nanophotoplankton - 3,00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack random
all <- specaccum(seq0.22ed.a, method = "random", index = "jack1")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Insular nanophotoplankton - 3,00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(seq0.22ed.a, method = "random", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
  ci.col = "lightblue", main = "Insular nanophotoplankton - 3,00µm (Library 16s rRNA)", xlab = "Number of Samples",
  ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#Jack rarefaction
all <- specaccum(seq0.22ed.a, method = "rarefaction", index = "jack1")

```

```

plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
     ci.col = "lightblue", main = "Insular nanophotoplankton - 3,00µm (Library 16s rRNA)", xlab = "Number of Samples",
     ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

all <- specaccum(seq0.22ed.a, method = "rarefaction", index = "jack2")
plot(all, ci.type = "poly", col = "blue", lwd = 2, ci.lty = 0, ylim = c(0, 50),
     ci.col = "lightblue", main = "Insular nanophotoplankton - 3,00µm (Library 16s rRNA)", xlab = "Number of Samples",
     ylab = "Number of Taxa")
boxplot(all, col = "yellow", add = TRUE, pch = "+")

#adiv (binario)
#Shannon
#diversidade geral
HShannon <- diversityresult(seq0.22ed.a, index='Shannon', method='pooled') #diversidade total por todos os pontos
HShannon

#diversidade media a cada ponto
h1Shannon <- diversityresult(seq0.22ed.a, index='Shannon', method='mean') #diversidade por ponto
h1Shannon

#diversidade por ponto
H <- diversity(seq0.22ed.a, index="shannon", base=exp(1))
View(H)
summary(H)

#Simpson
#diversidade geral
SSimpson <- diversityresult(seq0.22ed.a, index='Simpson', method='pooled') #diversidade total por todos os pontos
SSimpson

#diversidade media a cada ponto
s1Simpson <- diversityresult(seq0.22ed.a, index='Simpson', method='mean') #diversidade por ponto
s1Simpson

#diversidade por ponto
S <- diversity(seq0.22ed.a, index="simpson", base=exp(1))
View(S)
summary(S)

#Bdiv (binario)
#betadiver(help=TRUE)
#indice Jaccard
Jac <- designdist(seq0.22ed.a, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac)
scores(Jac)
quantile(Jac)
mean(Jac)

Jac.cluster <- hclust(Jac, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.cluster, 3)
Jac.cof<- cophenetic(Jac.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Jac.cof
plot(Jac.cof)
cor(Jac, Jac.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.738

#decostand
seq0.22ed.a.stan <- decostand(seq0.22ed.a, method='stand') #funcao para padronizar dados, para deixar tudo em escalas proximas, no caso o satnd padroniza
para atingir a distribuicao normal padrao
seq0.22ed.a.stan

#indice Jaccard STAND
Jac.stand <- designdist(seq0.22ed.a.stan, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Jac.stand)
scores(Jac.stand)
quantile(Jac.stand)
mean(Jac.stand)

Jac.stand.cluster <- hclust(Jac.stand, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.stand.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.stand.cluster, 3)
Jac.stand.cof<- cophenetic(Jac.stand.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do
vizinho completo ou mais distante
Jac.stand.cof
plot(Jac.stand.cof)

```

```

cor(Jac.stand, Jac.stand.cof)

#indice Sorensen
Sor <- designdist(seq0.22ed.a, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor)
scores(Sor)
quantile(Sor)
mean(Sor)

Sor.cluster <- hclust(Sor, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.cluster, 3)
Sor.cof<- cophenetic(Sor.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Sor.cof
plot(Sor.cof)
cor(Sor, Sor.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.704

#indice Sorensen STAND
Sor.stand <- designdist(seq0.22ed.a.stan, method = "(A+B-2*J)/(A+B)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
plot(Sor.stand)
scores(Sor.stand)
quantile(Sor.stand)
mean(Sor.stand)

Sor.stand.cluster <- hclust(Sor.stand, method='average')#upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Sor.stand.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Sor.stand.cluster, 3)
Sor.stand.cof<- cophenetic(Sor.stand.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do
vizinho completo ou mais distante
Sor.stand.cof
plot(Sor.stand.cof)
cor(Sor.stand, Sor.stand.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver
maior correlacao, no caso 0.704

# passo 3 - matrizes de ordenacao
#PCA
iris.pca <- PCA(seq0.22ed.a, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

plot(iris.pca, display = "sites")
fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=seq0.22.a$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=seq0.22.a$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=seq0.22.a$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(1,3), label="var", habillage=seq0.22.a$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca, axes = c(2,3), label="var", habillage=seq0.22.a$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()

#PCA stand
iris.pca.1 <- PCA(seq0.22ed.a.stan, scale.unit = FALSE, graph = FALSE)
iris.pca.1
loadings(iris.pca.1) #verifica os pesos encontrados
summary(iris.pca.1)
attributes(iris.pca.1)

plot(iris.pca.1, display = "sites")
fviz_pca_ind(iris.pca.1, label="none")
fviz_pca_ind(iris.pca.1, label="none", habillage=seq0.22.a$Samples)
fviz_pca_ind(iris.pca.1, label="none", habillage=seq0.22.a$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca.1, axes = c(1,2), label="var", habillage=seq0.22.a$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(1,3), label="var", habillage=seq0.22.a$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(2,3), label="var", habillage=seq0.22.a$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsiz
e = 5,
pointsize = 4) + theme_minimal()

#NMDS - (distancia Jaccard - biotico)
seq0.22ed.a.mds.Jac<- metaMDS(seq0.22ed.a, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq0.22ed.a.mds.Jac
attributes(seq0.22ed.a.mds.Jac)
plot(seq0.22ed.a.mds.Jac, cex=3, display=c('sites')) #deixa apenas os pontos

```

```

text(seq0.22ed.a.mds.Jac, display='sites')
points(seq0.22ed.a.mds.Jac$points[1:3,], col=5, cex=3)
points(seq0.22ed.a.mds.Jac$points[4:6,], col=5, cex=3)
points(seq0.22ed.a.mds.Jac$points[7:9,], col=5, cex=3)
points(seq0.22ed.a.mds.Jac$points[10:12,], col=5, cex=3)
points(seq0.22ed.a.mds.Jac$points[13:15,], col=6, cex=3)
points(seq0.22ed.a.mds.Jac$points[16:18,], col=6, cex=3)
points(seq0.22ed.a.mds.Jac$points[19:21,], col=6, cex=3)
points(seq0.22ed.a.mds.Jac$points[22:24,], col=6, cex=3)
legend(locator(1), c('AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(5,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.1602'))

```

#NMDS stand

```

seq0.22ed.a.mds.Jac.1 <- metaMDS(seq0.22ed.a.stan, distance='jaccard', binary=TRUE, k=2) #ficar em somente em 2 dimensoes
seq0.22ed.a.mds.Jac.1
attributes(seq0.22ed.a.mds.Jac.1)
plot(seq0.22ed.a.mds.Jac.1, cex=3, display=c('sites')) #deixa apenas os pontos
text(seq0.22ed.a.mds.Jac.1, display='sites')
points(seq0.22ed.a.mds.Jac.1$points[1:3,], col=5, cex=3)
points(seq0.22ed.a.mds.Jac.1$points[4:6,], col=5, cex=3)
points(seq0.22ed.a.mds.Jac.1$points[7:9,], col=5, cex=3)
points(seq0.22ed.a.mds.Jac.1$points[10:12,], col=5, cex=3)
points(seq0.22ed.a.mds.Jac.1$points[13:15,], col=6, cex=3)
points(seq0.22ed.a.mds.Jac.1$points[16:18,], col=6, cex=3)
points(seq0.22ed.a.mds.Jac.1$points[19:21,], col=6, cex=3)
points(seq0.22ed.a.mds.Jac.1$points[22:24,], col=6, cex=3)
legend(locator(1), c('AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(5,5,5,5,6,6,6,6))
legend(locator(1), c('stress = ?'))

```

## Parâmetros da água x índices ecológicos Illumina

##### All - ENV

```
env <- read.delim2("~/Documents/R operacional/env.txt", header = T, dec=',')
```

# paso 1 - entendendo a disposição dos dados env

```
View(env) #lendo no RStudio
```

```
env #lendo no console
```

```
attach(env) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes
```

```
summary(env) #resumo pra saber as significancias e valores da planilha
```

```
attributes(env) #mostra os criterios
```

```
env.ed <- env[,4:7]
```

```
View(env.ed)
```

```
shapiro.test(env.ed$pH)
```

```
shapiro.test(env.ed$T)
```

```
shapiro.test(env.ed$Salinity)
```

```
shapiro.test(env.ed$DO)
```

```
env.ed.stand <- decostand(env.ed, "standardize")
```

```
View(env.ed.stand)
```

```
shapiro.test(env.ed.stand$pH)
```

```
shapiro.test(env.ed.stand$T)
```

```
shapiro.test(env.ed.stand$Salinity)
```

```
shapiro.test(env.ed.stand$DO)
```

#ANOVA multifatorial

```
fait <- aov(env.ed$pH~env.ed$pH*env.ed$T*env.ed$Salinity*env.ed$DO, env.ed)
```

```
anova(fait)
```

```
TukeyHSD(fait)
```

```
frait <- aov(env.ed.stand$pH~env.ed.stand$pH*env.ed.stand$T*env.ed.stand$Salinity*env.ed.stand$DO, env.ed.stand)
```

```
anova(frait)
```

```
TukeyHSD(frait)
```

# paso 2 - matriz de dissimilaridade (agrupamento) - distancia euclidiana com dados abioticos:

#SEM STAND

```
env.ed.eu.1 <- vegdist(env.ed, method='eu', diag = F, upper = F) #apenas dei um nome, indice de similaridade (distância euclidiana pura para analises quantitativas)
```

```
env.ed.eu.1
```

```
env.ed.eu.cluster.1 <- hclust(env.ed.eu.1, method = "average") # nomeando e gerando a analise de cluster
```

```
env.ed.eu.cluster.1
```

```
plot(env.ed.eu.cluster.1, hang = -1) #gera o grafico de cluster
```

```
rect.hclust(env.ed.eu.cluster.1, 3)
```

```
env.ed.eu.cluster.1.cof <- cophenetic(env.ed.eu.cluster.1) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho completo ou mais distante
```

```
env.ed.eu.cluster.1.cof
```

```

plot(env.ed.eu.cluster.1.cof)
cor(env.ed.eu.1, env.ed.eu.cluster.1.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o
que tiver maior correlacao, no caso 0.704

#STAND
env.ed.stand.eu.1<-vegdist(env.ed.stand, method='eu', diag = F, upper = F) #apenas dei um nome, indice de similaridade (distância euclidiana pura para
analises quantitativas)
env.ed.stand.eu.1
env.ed.stand.eu.cluster.1<-hclust(env.ed.stand.eu.1, method = "average") # nomeando e gerando a analise de cluster
env.ed.stand.eu.cluster.1
plot(env.ed.stand.eu.cluster.1, hang = -1) #gera o grafico de cluster
rect.hclust(env.ed.stand.eu.cluster.1, 3)
env.ed.stand.eu.cluster.1.cof<- cophenetic(env.ed.stand.eu.cluster.1) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada
atraves do metodo do vizinho completo ou mais distante
env.ed.stand.eu.cluster.1.cof
plot(env.ed.stand.eu.cluster.1.cof)
cor(env.ed.stand.eu.1, env.ed.stand.eu.cluster.1.cof) #cor

# paso 3 - PCA
#SEM STAND
iris.pca <- PCA(env.ed, scale.unit = FALSE, graph = FALSE)
iris.pca
loadings(iris.pca) #verifica os pesos encontrados
summary(iris.pca)
attributes(iris.pca)

fviz_pca_ind(iris.pca, label="none")
fviz_pca_ind(iris.pca, label="none", habillage=env$Samples)
fviz_pca_ind(iris.pca, label="none", habillage=env$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca, axes = c(1,2), label="var", habillage=env$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsize = 5, pointsize
= 4) + theme_minimal()

#PCA stand
iris.pca.1 <- PCA(env.ed.stand, scale.unit = FALSE, graph = FALSE)
iris.pca.1
loadings(iris.pca.1) #verifica os pesos encontrados
summary(iris.pca.1)
attributes(iris.pca.1)

fviz_pca_ind(iris.pca.1, label="none")
fviz_pca_ind(iris.pca.1, label="none", habillage=env$Samples)
fviz_pca_ind(iris.pca.1, label="none", habillage=env$Samples, addEllipses=TRUE, ellipse.level=0.95)
fviz_pca_biplot(iris.pca.1, axes = c(1,2), label="var", habillage=env$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(1,3), label="var", habillage=env$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsize = 5,
pointsize = 4) + theme_minimal()
fviz_pca_biplot(iris.pca.1, axes = c(2,3), label="var", habillage=env$Samples, alpha.var ="contrib", col.var = "black", invisible = "ind", labelsize = 5,
pointsize = 4) + theme_minimal()

#MDS - (distancia euclidiana metrica - abiotico)
env.ed.mds.eu<- metaMDS(env.ed, distance='eu', k=2) #ficar em somente em 2 dimensoes
env.ed.mds.eu
attributes(env.ed.mds.eu)
plot(env.ed.mds.eu, cex=3, display=c('sites')) #deixa apenas os pontos
text(env.ed.mds.eu, display='sites')
points(env.ed.mds.eu$points[1:2,], col=1, cex=3)
points(env.ed.mds.eu$points[3:4,], col=1, cex=3)
points(env.ed.mds.eu$points[5:7,], col=2, cex=3)
points(env.ed.mds.eu$points[8:10,], col=3, cex=3)
points(env.ed.mds.eu$points[11:13,], col=4, cex=3)
points(env.ed.mds.eu$points[14:16,], col=5, cex=3)
points(env.ed.mds.eu$points[17:19,], col=5, cex=3)
points(env.ed.mds.eu$points[20:22,], col=5, cex=3)
points(env.ed.mds.eu$points[23:25,], col=5, cex=3)
points(env.ed.mds.eu$points[26:28,], col=6, cex=3)
points(env.ed.mds.eu$points[29:31,], col=6, cex=3)
points(env.ed.mds.eu$points[32:34,], col=6, cex=3)
points(env.ed.mds.eu$points[35:37,], col=6, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.0283'))

#MDS stand
env.ed.stand.mds.eu<- metaMDS(env.ed.stand, distance='eu', k=2) #ficar em somente em 2 dimensoes
env.ed.stand.mds.eu
attributes(env.ed.stand.mds.eu)
plot(env.ed.stand.mds.eu, cex=3, display=c('sites')) #deixa apenas os pontos
text(env.ed.stand.mds.eu, display='sites')
points(env.ed.stand.mds.eu$points[1:2,], col=1, cex=3)

```

```

points(env.ed.stand.mds.eu$points[3:4,], col=1, cex=3)
points(env.ed.stand.mds.eu$points[5:7,], col=2, cex=3)
points(env.ed.stand.mds.eu$points[8:10,], col=3, cex=3)
points(env.ed.stand.mds.eu$points[11:13,], col=4, cex=3)
points(env.ed.stand.mds.eu$points[14:16,], col=5, cex=3)
points(env.ed.stand.mds.eu$points[17:19,], col=5, cex=3)
points(env.ed.stand.mds.eu$points[20:22,], col=5, cex=3)
points(env.ed.stand.mds.eu$points[23:25,], col=5, cex=3)
points(env.ed.stand.mds.eu$points[26:28,], col=6, cex=3)
points(env.ed.stand.mds.eu$points[29:31,], col=6, cex=3)
points(env.ed.stand.mds.eu$points[32:34,], col=6, cex=3)
points(env.ed.stand.mds.eu$points[35:37,], col=6, cex=3)
legend(locator(1), c('AC1', 'AC2', 'PP1', 'PA1', 'IC1', 'AS1', 'AS2', 'AS3', 'AS4', 'FN1', 'FN2', 'FN3', 'FN4'), pch=16, col=c(1,1,2,3,4,5,5,5,6,6,6,6))
legend(locator(1), c('stress = 0.0984'))

# passo 4 - correlacao env
View(env.ed)

env.cor <- cor(env.ed, method = "spearman")
View(env.cor)
plot(env.cor)

corrplot(env.cor, type = "upper", hclust.method = "average", order="hclust",col=brewer.pal(n=8, name="RdBu"), tl.col="black", tl.srt=45, insig = "label_sig")

# mat : is a matrix of data
# ... : further arguments to pass to the native R cor.test function
# Specialized the insignificant value according to the significant level
cor.mtest <- function(mat, ...) {
  mat <- as.matrix(mat)
  n <- ncol(mat)
  p.mat<- matrix(NA, n, n)
  diag(p.mat) <- 0
  for (i in 1:(n - 1)) {
    for (j in (i + 1):n) {
      tmp <- cor.test(mat[, i], mat[, j], ...)
      p.mat[i, j] <- p.mat[j, i] <- tmp$p.value
    }
  }
  colnames(p.mat) <- rownames(p.mat) <- colnames(mat)
  p.mat
}

# matrix of the p-value of the correlation
p.mat <- cor.mtest(env.ed)
View(p.mat)

corrplot(env.cor, type = "upper", hclust.method = "average", order="hclust",col=brewer.pal(n=8, name="RdBu"), tl.col="black", tl.srt=45, insig = "label_sig",
p.mat = p.mat, sig.level = c(.001, .01, .05), pch.cex = .9, diag=FALSE)

# STAND correlacao env
View(env.ed.stand)

env.stand.cor <- cor(env.ed.stand, method = "spearman")
View(env.stand.cor)
plot(env.stand.cor)

corrplot(env.stand.cor, type = "upper", hclust.method = "average", order="hclust",col=brewer.pal(n=8, name="RdBu"), tl.col="black", tl.srt=45, insig =
"label_sig")

# mat : is a matrix of data
# ... : further arguments to pass to the native R cor.test function
# Specialized the insignificant value according to the significant level
cor.mtest <- function(mat, ...) {
  mat <- as.matrix(mat)
  n <- ncol(mat)
  p.mat<- matrix(NA, n, n)
  diag(p.mat) <- 0
  for (i in 1:(n - 1)) {
    for (j in (i + 1):n) {
      tmp <- cor.test(mat[, i], mat[, j], ...)
      p.mat[i, j] <- p.mat[j, i] <- tmp$p.value
    }
  }
  colnames(p.mat) <- rownames(p.mat) <- colnames(mat)
  p.mat
}

# matrix of the p-value of the correlation

```

```

p.mat <- cor.mtest(env.stand.cor)
View(p.mat)

corrplot(env.stand.cor, type = "upper", hclust.method = "average", order="hclust",col=brewer.pal(n=8, name="RdBu"), tl.col="black", tl.srt=45, insig =
"label_sig", p.mat = p.mat, sig.level = c(.001, .01, .05), pch.cex = .9, diag=FALSE)

##### 3.00µm
# ECO 3.00 com ENV

#restart R

# paso 1a - informacoes preliminares env

env <- read.delim2("~/Documents/R operacional/env.txt", header = T, dec=',')

View(env)
attach(env)

# paso 1b - entendendo a disposiçao dos dados 3.00µm

seq3 <- read.table("~/Documents/R operacional/seq 3.00 illumina.txt", header = T, dec=',')

View(seq3) #lendo no RStudio
attach(seq3) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes

seq3ed <- seq3[,-1]
View(seq3ed)

seq3edstand <- decostand(seq3ed, "standardize")
View(seq3edstand)

# RDA - (abiotico e biotico)
# RDA preliminar alpha
View(seq3ed)
seq3.rda <- rda(seq3ed~DO+Dynamics+pH+Salinity+Substrate+T) #o conditioned e o efeito que vc retirou (no caso envtopo) mais os efeitos de interacao entre
os dois, e constrained tem a explicao do que vc comparou (envchem no caso)
seq3.rda
summary(seq3.rda)
plot(seq3.rda)
permutest(seq3.rda, permutations=999) # para saber se a RDA e significativa
anova.cca(seq3.rda, by='axis', step=999) # para saber se eixos da RDA sao significativos
ordistep(seq3.rda) #da a significancia de cada variavel da RDA, ai volta e refaz toda a analize denovo com as variaveis mais significantes, e esta sim sera sua
RDA preliminar beta
vif.cca(seq3.rda) #verifica a colinearidade, valores acima de 10 tem de ser retirados, pois esta inflando a massa de dados, ai recorre se tiver alguma com
acima de dez e roda denovo a RDA, e ai sim essa sera sua definitiva
RsquareAdj(seq3.rda)

# RDA definitiva
seq3ed.rda.on <- rda(seq3ed~T+Salinity+DO+Substrate)
seq3ed.rda.on
summary(seq3ed.rda.on)
permutest(seq3ed.rda.on, permutations=999)
anova.cca(seq3ed.rda.on, by='axis', step=999)
vif.cca(seq3ed.rda.on)
RsquareAdj(seq3ed.rda.on)
plot(seq3ed.rda.on, scalling=1, xlab='RDA 1 - 93.45%', ylab='RDA 2 - 6.55%') #Importance of components do summary
text(locator(1), 'pseudo-F = 9.45') #model permutest
text(locator(1), 'r.square = 0.60') #[1] do RsquareAdj

# RDA outras variáveis
seq3ed.rda.off <- rda(seq3ed~Dynamics+pH)
summary(seq3ed.rda.off)
permutest(seq3ed.rda.off, permutations=999)
anova.cca(seq3ed.rda.off, by='axis', step=999)
RsquareAdj(seq3ed.rda.off)
vif.cca(seq3ed.rda.off)
plot(seq3ed.rda.off, scalling=1, xlab='Eixo 1 - 98.32%', ylab='Eixo 2 - 1.68%') #Importance of components do summary
text(locator(1), 'pseudo-F = 5.75') #model permutest
text(locator(1), 'r.square = 0.25') #[1] do RsquareAdj

#riqueza e diversidade das diferentes variaveis categoricas
#Substrate
S.Apool <- with(env, specpool(seq3ed, Substrate))
S.Apool

boxplot(specnumber(seq3ed) ~ Substrate, data = env, border = "black")

A1 <- diversitycomp(seq3ed, y=env, factor1="Substrate", index="Shannon", method="mean", digits=4)

```

```

A1

A2 <- diversitycomp(seq3ed, y=env, factor1="Substrate", index="Simpson", method="mean", digits=4)
A2

#Dynamics
D.Apool <- with(env, specpool(seq3ed, Dynamics))
D.Apool

boxplot(specnumber(seq3ed) ~ Dynamics, data = env, border = "black")

B1 <- diversitycomp(seq3ed, y=env, factor1="Dynamics", index="Shannon", method="mean", digits=4)
B1

B2 <- diversitycomp(seq3ed, y=env, factor1="Dynamics", index="Simpson", method="mean", digits=4)
B2

#Substrate x Dynamics
C1 <- diversitycomp(seq3ed, y=env, factor1="Substrate", factor2="Dynamics", index="Shannon", method="mean", digits=4)
C1

C2 <- diversitycomp(seq3ed, y=env, factor1="Substrate", factor2="Dynamics", index="Simpson", method="mean", digits=4)
C2

#Mantel test
#cluster - diversidade beta varia dependendo de das variaveis categoricas
#env stand - indice euclidiano

env.ed <- env[,4:7]
View(env.ed)

env.ed.stand <- decostand(env.ed, "standardize")
View(env.ed.stand)

env.ed.stand.eu.1 <- vegdist(env.ed.stand, method='eu', diag = F, upper = F) #apenas dei um nome, indice de similaridade (distância euclidiana pura para
analises quantitativas)
env.ed.stand.eu.1
env.ed.stand.eu.cluster.1 <- hclust(env.ed.stand.eu.1, method = "average") # nomeando e gerando a analise de cluster
env.ed.stand.eu.cluster.1
plot(env.ed.stand.eu.cluster.1, hang = -1) #gera o grafico de cluster
rect.hclust(env.ed.stand.eu.cluster.1, 3)
env.ed.stand.eu.cluster.1.cof <- cophenetic(env.ed.stand.eu.cluster.1) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada
atraves do metodo do vizinho completo ou mais distante
env.ed.stand.eu.cluster.1.cof
plot(env.ed.stand.eu.cluster.1.cof)
cor(env.ed.stand.eu.1, env.ed.stand.eu.cluster.1.cof)

#Bdiv - indice Jaccard
Jac <- designdist(seq3ed, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
Jac
plot(Jac)
scores(Jac)
quantile(Jac)
mean(Jac)

Jac.cluster <- hclust(Jac, method='average') #upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.cluster, 3)
Jac.cof <- cophenetic(Jac.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Jac.cof
plot(Jac.cof)
cor(Jac, Jac.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.738

#Mantel
mantel(Jac, env.ed.stand.eu.1, method = "spearman")
plot(Jac, env.ed.stand.eu.1)

#p
Man <- mantel.rtest(Jac, env.ed.stand.eu.1, nrepet = 999)
Man
plot(Man, main = "Mantel's test")

#betadisper
#Substrate
S <- with(env, betadisper(Jac, Substrate, type = "centroid"))
S

```

```

# Perform test
anova(S) #se for significativo, dis um grupo foi distribuido mais homogeneamente e o outro menos homogeneamente, ve que as variabilidades sao diferentes
TukeyHSD(S, which = "group", ordered = FALSE, conf.level = 0.95)
plot(S)

#Dynamics
D <- with(env, betadisper(Jac, Dynamics, type = "centroid"))
D
# Perform test
anova(D) #se for significativo, dis um grupo foi distribuido mais homogeneamente e o outro menos homogeneamente, ve que as variabilidades sao diferentes
TukeyHSD(D, which = "group", ordered = FALSE, conf.level = 0.95)
plot(D)

#headmap
data.prop <- seq3ed/rowSums(seq3ed)
View(data.prop)

# determine the maximum relative abundance for each column
maxab <- apply(data.prop, 2, max)
maxab
head(maxab)
View(maxab)

# remove the genera with less than 1% as their maximum relative abundance
n1 <- names(which(maxab < 0.01))
data.prop.1 <- data.prop[, -which(names(data.prop) %in% n1)]
View(data.prop.1)
attach(data.prop.1)

# calculate the dissimilarity matrix on the full dataset
#jaccard
data.dist.jac <- designdist(data.prop, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
x <- as.matrix(data.dist.jac)
View(x)
row.clus.jac <- hclust(data.dist.jac, method='average')
plot(row.clus.jac, hang = -1)

row.clus.jac.cof <- cophenetic(row.clus.jac) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do
vizinho completo ou mais distante
row.clus.jac.cof
plot(row.clus.jac.cof)
cor(data.dist.jac, row.clus.jac.cof)

# calculate the dissimilarity matrix on the clear genera dataset you have to transpose the dataset to get the genera as rows
#jaccard
data.dist.jac.1 <- designdist(t(data.prop.1), method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
y <- as.matrix(data.dist.jac.1)
View(y)
col.clus.jac.1 <- hclust(data.dist.jac.1, method='average')
plot(col.clus.jac.1, hang = -1)

col.clus.jac.1.cof <- cophenetic(col.clus.jac.1) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do
vizinho completo ou mais distante
col.clus.jac.1.cof
plot(col.clus.jac.1.cof)
cor(data.dist.jac.1, col.clus.jac.1.cof)

# make the heatmap with Rowv and Col
heatmap(as.matrix(data.prop.1), Rowv = as.dendrogram(row.clus.jac), Colv = as.dendrogram(col.clus.jac.1), col=brewer.pal(n=8, name="RdBu"), margins =
c(10, 3))
heatmap.2(as.matrix(data.prop.1), Rowv = as.dendrogram(row.clus.jac), Colv = as.dendrogram(col.clus.jac.1), col=brewer.pal(n=8, name="RdBu"), margins =
c(10, 3), scale="none", density.info="none", trace="none")

#correlacao genero
View(data.prop.1)

data.prop.1.cor <- cor(data.prop.1, method = "spearman")
View(data.prop.1.cor)
plot(data.prop.1.cor)

corrplot(data.prop.1.cor, type = "upper", hclust.method = "average", order="hclust", col=brewer.pal(n=8, name="RdBu"), tl.col="black", tl.srt=45, insig =
"label_sig")

# mat : is a matrix of data
# ... : further arguments to pass to the native R cor.test function
# Specialized the insignificant value according to the significant level
cor.mtest <- function(mat, ...) {
  mat <- as.matrix(mat)

```

```

n <- ncol(mat)
p.mat<- matrix(NA, n, n)
diag(p.mat) <- 0
for (i in 1:(n - 1)) {
  for (j in (i + 1):n) {
    tmp <- cor.test(mat[, i], mat[, j], ...)
    p.mat[i, j] <- p.mat[j, i] <- tmp$p.value
  }
}
colnames(p.mat) <- rownames(p.mat) <- colnames(mat)
p.mat
}

# matrix of the p-value of the correlation
p.mat <- cor.mtest(data.prop.1)
View(p.mat)

corrplot(data.prop.1.cor, type = "upper", hclust.method = "average", order="hclust",col=brewer.pal(n=8, name="RdBu"), tl.col="black", tl.srt=45, insig =
"label_sig", p.mat = p.mat, sig.level = c(.001, .01, .05), pch.cex = .9, diag=FALSE)

##### 0.22µm
# ECO 0.22 com ENV

#restart R

##### paso 1a - informacoes preliminares env

env <- read.delim2("~/Documents/R operacional/env.txt", header = T, dec=',')

View(env)
attach(env)

##### paso 1b - entendendo a disposição dos dados 0.22µm

seq0.22 <- read.table("~/Documents/R operacional/seq 0.22 ilumina.txt", header = T, dec=',')

View(seq0.22) #lendo no RStudio
attach(seq0.22) #lê o cabeçário, ou seja, as VARIÁVEIS da planilha e suas posicoes

seq0.22ed <- seq0.22[, -1]
View(seq0.22ed)

seq0.22edstand <- decostand(seq0.22ed, "standardize")
View(seq0.22edstand)

# RDA - (abiotico e biotico)
# RDA preliminar alpha
View(seq0.22ed)
seq0.22.rda<- rda(seq0.22ed~DO+Dynamics+pH+Salinity+Substrate+T) #o conditioned e o efeito que vc retirou (no caso envtopo) mais os efeitos de
interacao entre os dois, e constrained tem a explicao do que vc comparou (envchem no caso)
seq0.22.rda
summary(seq0.22.rda)
plot(seq0.22.rda)
permutest(seq0.22.rda, permutations=999) # para saber se a RDA e significativa
anova.cca(seq0.22.rda, by='axis', step=999) # para saber se eixos da RDA sao significativos
ordisurf(seq0.22.rda) #da a significancia de cada variavel da RDA, ai volta e refaz toda a analize denovo com as variaveis mais significantes, e esta sim sera
sua RDA preliminar beta
vif.cca(seq0.22.rda) #verifica a colinearidade, valores acima de 10 tem de ser retirados, pois esta inflando a massa de dados, ai recorre se tiver alguma com
acima de dez e roda denovo a RDA, e ai sim essa sera sua definitiva
RsquareAdj(seq0.22.rda)

# RDA definitiva
seq0.22ed.rda.on<-rda(seq0.22ed~Salinity+DO+Substrate+T)
seq0.22ed.rda.on
summary(seq0.22ed.rda.on)
permutest(seq0.22ed.rda.on, permutations=999)
anova.cca(seq0.22ed.rda.on, by='axis', step=999)
vif.cca(seq0.22ed.rda.on)
RsquareAdj(seq0.22ed.rda.on)
plot(seq0.22ed.rda.on, scalling=1, xlab='RDA 1 - 91.67%', ylab='RDA 2 - 8.32%') #Importance of components do summary
text(locator(1), 'pseudo-F = 11.22') #model permutest
text(locator(1), 'r.square = 0.64') #[1] do RsquareAdj

# RDA outras variáveis
seq0.22ed.rda.off<- rda(seq0.22ed~Dynamics+pH)
summary(seq0.22ed.rda.off)
permutest(seq0.22ed.rda.off, permutations=999)
anova.cca(seq0.22ed.rda.off, by='axis', step=999)

```

```

RsquareAdj(seq0.22ed.rda.off)
vif.cca(seq0.22ed.rda.off)
plot(seq0.22ed.rda.off, scalling=1, xlab='Eixo 1 - 99.89%', ylab='Eixo 2 - 0.11%') #Importance of components do summary
text(locator(1), 'pseudo-F = 4.84') #model permutest
text(locator(1), 'r.square = 0.22') #[1] do RsquareAdj

#riqueza e diversidade das diferentes variaveis categoricas
#Substrate
S.Apool <- with(env, specpool(seq0.22ed, Substrate))
S.Apool

boxplot(specnumber(seq0.22ed) ~ Substrate, data = env, border = "black")

A1 <- diversitycomp(seq0.22ed, y=env, factor1="Substrate", index="Shannon", method="mean", digits=4)
A1

A2 <- diversitycomp(seq0.22ed, y=env, factor1="Substrate", index="Simpson", method="mean", digits=4)
A2

#Dynamics
D.Apool <- with(env, specpool(seq0.22ed, Dynamics))
D.Apool

boxplot(specnumber(seq0.22ed) ~ Dynamics, data = env, border = "black")

B1 <- diversitycomp(seq0.22ed, y=env, factor1="Dynamics", index="Shannon", method="mean", digits=4)
B1

B2 <- diversitycomp(seq0.22ed, y=env, factor1="Dynamics", index="Simpson", method="mean", digits=4)
B2

#Substrate x Dynamics
C1 <- diversitycomp(seq0.22ed, y=env, factor1="Substrate", factor2="Dynamics", index="Shannon", method="mean", digits=4)
C1

C2 <- diversitycomp(seq0.22ed, y=env, factor1="Substrate", factor2="Dynamics", index="Simpson", method="mean", digits=4)
C2

#Mantel test
#cluster - diversidade beta varia dependendo de das variaveis categoricas
#env stand - indice euclidiano

env.ed <- env[,4:7]
View(env.ed)

env.ed.stand <- decostand(env.ed, "standardize")
View(env.ed.stand)

env.ed.stand.eu.1 <- vegdist(env.ed.stand, method='eu', diag = F, upper = F) #apenas dei um nome, indice de similaridade (distância euclidiana pura para
analises quantitativas)
env.ed.stand.eu.1
env.ed.stand.eu.cluster.1 <- hclust(env.ed.stand.eu.1, method = "average") # nomeando e gerando a analise de cluster
env.ed.stand.eu.cluster.1
plot(env.ed.stand.eu.cluster.1, hang = -1) #gera o grafico de cluster
rect.hclust(env.ed.stand.eu.cluster.1, 3)
env.ed.stand.eu.cluster.1.cof <- cophenetic(env.ed.stand.eu.cluster.1) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada
atraves do metodo do vizinho completo ou mais distante
env.ed.stand.eu.cluster.1.cof
plot(env.ed.stand.eu.cluster.1.cof)
cor(env.ed.stand.eu.1, env.ed.stand.eu.cluster.1.cof)

#Bdiv - indice Jaccard
Jac <- designdist(seq0.22ed, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
Jac
plot(Jac)
scores(Jac)
quantile(Jac)
mean(Jac)

Jac.cluster <- hclust(Jac, method='average') #upgma - matriz de distancia observada atraves da media entre as distancias das variaveis
plot(Jac.cluster, hang = -1) #gera o grafico de cluster
rect.hclust(Jac.cluster, 3)
Jac.cof <- cophenetic(Jac.cluster) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do vizinho
completo ou mais distante
Jac.cof
plot(Jac.cof)
cor(Jac, Jac.cof) #correlacao entre a matriz observada e a ficticia, ai vc ve se o metodo qde cluster que vc usou e o mais adequado, eh o que tiver maior
correlacao, no caso 0.738

```

```

#Mantel
mantel(Jac, env.ed.stand.eu.1, method = "spearman")
plot(Jac, env.ed.stand.eu.1)

#p
Man <- mantel.rtest(Jac, env.ed.stand.eu.1, nrepet = 999)
Man
plot(Man, main = "Mantel's test")

#betadisper
#Substrate
S <- with(env, betadisper(Jac, Substrate, type = "centroid"))
S
# Perform test
anova(S) #se for significativo, dis um grupo foi distribuido mais homogeneamente e o outro menos homogeneamente, ve que as variabilidades sao diferentes
TukeyHSD(S, which = "group", ordered = FALSE, conf.level = 0.95)
plot(S)

#Dynamics
D <- with(env, betadisper(Jac, Dynamics, type = "centroid"))
D
# Perform test
anova(D) #se for significativo, dis um grupo foi distribuido mais homogeneamente e o outro menos homogeneamente, ve que as variabilidades sao diferentes
TukeyHSD(D, which = "group", ordered = FALSE, conf.level = 0.95)
plot(D)

#headmap
data.prop <- seq0.22ed/rowSums(seq0.22ed)
View(data.prop)

# determine the maximum relative abundance for each column
maxab <- apply(data.prop, 2, max)
maxab
head(maxab)
View(maxab)

# remove the genera with less than 1% as their maximum relative abundance
n1 <- names(which(maxab < 0.01))
data.prop.1 <- data.prop[, -which(names(data.prop) %in% n1)]
View(data.prop.1)
attach(data.prop.1)

# calculate the dissimilarity matrix on the full dataset
#jaccard
data.dist.jac <- designdist(data.prop, method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
x <- as.matrix(data.dist.jac)
View(x)
row.clus.jac <- hclust(data.dist.jac, method='average')
plot(row.clus.jac, hang = -1)

row.clus.jac.cof <- cophenetic(row.clus.jac) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do
vizinho completo ou mais distante
row.clus.jac.cof
plot(row.clus.jac.cof)
cor(data.dist.jac, row.clus.jac.cof)

# calculate the dissimilarity matrix on the clear genera dataset you have to transpose the dataset to get the genera as rows
#jaccard
data.dist.jac.1 <- designdist(t(data.prop.1), method = "(A+B-2*J)/(A+B-J)", terms = "binary", abcd = FALSE, alphagamma = FALSE)
y <- as.matrix(data.dist.jac.1)
View(y)
col.clus.jac.1 <- hclust(data.dist.jac.1, method='average')
plot(col.clus.jac.1, hang = -1)

col.clus.jac.1.cof <- cophenetic(col.clus.jac.1) #gera uma matriz ficticia com base na analise grafica da matriz de distancia observada atraves do metodo do
vizinho completo ou mais distante
col.clus.jac.1.cof
plot(col.clus.jac.1.cof)
cor(data.dist.jac.1, col.clus.jac.1.cof)

# make the heatmap with Rowv and Col
heatmap(as.matrix(data.prop.1), Rowv = as.dendrogram(row.clus.jac), Colv = as.dendrogram(col.clus.jac.1), col=brewer.pal(n=8, name="RdBu"), margins =
c(10, 3))
heatmap.2(as.matrix(data.prop.1), Rowv = as.dendrogram(row.clus.jac), Colv = as.dendrogram(col.clus.jac.1), col=brewer.pal(n=8, name="RdBu"), margins =
c(10, 3), scale="none", density.info="none", trace="none")

#correlacao genero

```

```

View(data.prop.1)

data.prop.1.cor <- cor(data.prop.1, method = "spearman")
View(data.prop.1.cor)
plot(data.prop.1.cor)

corrplot(data.prop.1.cor, type = "upper", hclust.method = "average", order="hclust",col=brewer.pal(n=8, name="RdBu"), tl.col="black", tl.srt=45, insig =
"label_sig")

# mat : is a matrix of data
# ... : further arguments to pass to the native R cor.test function
# Specialized the insignificant value according to the significant level
cor.mtest <- function(mat, ...) {
  mat <- as.matrix(mat)
  n <- ncol(mat)
  p.mat<- matrix(NA, n, n)
  diag(p.mat) <- 0
  for (i in 1:(n - 1)) {
    for (j in (i + 1):n) {
      tmp <- cor.test(mat[, i], mat[, j], ...)
      p.mat[i, j] <- p.mat[j, i] <- tmp$p.value
    }
  }
  colnames(p.mat) <- rownames(p.mat) <- colnames(mat)
  p.mat
}

# matrix of the p-value of the correlation
p.mat <- cor.mtest(data.prop.1)
View(p.mat)

corrplot(data.prop.1.cor, type = "upper", hclust.method = "average", order="hclust",col=brewer.pal(n=8, name="RdBu"), tl.col="black", tl.srt=45, insig =
"label_sig", p.mat = p.mat, sig.level = c(.001, .01, .05), pch.cex = .9, diag=FALSE)

```