Universidad de Oviedo

Departamento de Biología de Organismos y Sistemas

Programa de Doctorado en Recursos Biológicos y Biodiversidad

“Predicting the metabolic balance of the oceans”

***

“Predicción del balance metabólico de los océanos”

TESIS DOCTORAL

Francisca del Carmen García García

Oviedo, Mayo 2015

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: PREDICCIÓN DEL BALANCE METABÓLICO DE LOS OCÉANOS	Inglés: PREDICTING THE METABOLIC BALANCE OF THE OCEANS

2.- Autor	
Nombre: FRANCISCA DEL CARMEN GARCÍA GARCÍA	DNI/Pasap
Programa de Doctorado: RECURSOS BIOLÓGICOS Y BIODIVERSIDAD	
Órgano responsable: DEPARTAMENTO DE BIOLOGÍA DE ORGANISMOS Y SISTEMAS	

RESUMEN (en español)

El bacterioplancton marino desempeña un papel fundamental en los ciclos biogeoquímicos, principalmente en el ciclo del carbono, siendo responsable de una fracción importante de la respiración heterotrófica. En el contexto de calentamiento global actual el poder cuantificar y predecir el papel del bacterioplancton en la respiración y comparar esos valores con las estimas de producción primaria existentes se ha convertido en un desafío científico y constituye uno de los principales objetivos de esta tesis. Para ello, es necesario profundizar en el conocimiento de cuáles son los principales factores y qué efecto tienen sobre la distribución de la abundancia y el metabolismo del bacterioplancton marino.

Para lograr estos objetivos se ha recopilado una extensa base de datos donde se incorporan datos de abundancia bacteriana y otras variables ambientales (temperatura y clorofila principalmente). En la primera parte de este manuscrito (Capítulo 1 y 2) se han desarrollado metodologías de análisis de archivos obtenidos mediante la técnica de citometría de flujo, herramienta rutinaria para el análisis automatizado de poblaciones de bacterioplancton en sistemas acuáticos. En el Capítulo 1 se ha desarrollado una metodología para detectar poblaciones de bacterioplancton marino de forma automática. Esta automatización y estandarización del proceso de detección de poblaciones de bacterias heterotróficas marinas nos permite hacer comparables muestras de distintas fuentes lo que es de gran utilidad cuando se trabaja con bases de datos extensas. En el Capítulo 2 utilizamos también archivos procedentes de análisis citométricos para desarrollar una metodología que nos permita estimar la 'diversidad citométrica' de cada muestra de citometría de flujo. Este análisis se aplicó a una serie temporal de datos en el mar Cantábrico y los resultados fueron validados con datos de diversidad genética. Este análisis nos permitió comprobar que se pueden estimar cambios en las comunidades de bacterioplancton marino a gran escala utilizando la citometría de flujo, lo que representa un gran avance en ecología microbiana por la aplicabilidad, rapidez y bajo coste de esta técnica frente al uso de técnicas moleculares.

En el Capítulo 3 se presenta un estudio metabólico de alometría de bacterioplancton marino realizado con comunidades muestreadas in situ y separadas por su tamaño a lo largo de un gradiente latitudinal en el Océano Atlántico. Este estudio nos ha permitido demostrar por primera vez en el medio natural que las bacterias tienen una respiración superlineal en función de su tamaño, el cual, a pesar de ser tan pequeño, influye en su metabolismo. Esto es debido a la estrecha relación entre tamaño y complejidad metabólica de estos organismos. Esta relación del metabolismo con el tamaño de las bacterias ha sido encontrada al fraccionar una comunidad dónde todas sus células están bajo las mismas condiciones ambientales, en tamaños y medir la respiración de cada clase de tamaño por separado. Sin embargo al comparar el metabolismo de comunidades de distintas zonas con su tamaño medio este efecto no se observa debido al estrecho

rango de tamaños que poseen estos organismos.

La clorofila ha sido señalada en artículos centrados en aguas superficiales, como una de las principales variables que controlan la distribución de bacterioplancton, pero no se ha cuantificado esta relación en la dimensión vertical.
En el Capítulo 4 se utiliza la base de datos global recopilada para estudiar la distribución de abundancia del bacterioplancton marino. En este capítulo se han caracterizado perfiles verticales de clorofila y de abundancia bacteriana ajustándolos a una distribución Gaussiana. Una vez caracterizados se ha estudiado el acoplamiento de varias variables en la vertical utilizando redes neuronales artificiales. De esta forma, hemos podido comprobar que ambas variables están relacionadas para así poder utilizar un modelo no paramétrico en el que a partir de variables superficiales se pueden predecir la distribución de abundancia bacteriana.
En trabajos recientes se ha demostrado que la respiración planctónica se puede estimar multiplicando la abundancia bacteriana por la actividad metabólica celular utilizando la teoría metabólica de la ecología. En este capítulo hemos utilizado la abundancia predicha y la actividad metabólica celular para realizar un primer intento de predecir la respiración bacteriana global en la capa fótica utilizando imágenes de satélite y comparar esta respiración con las predicciones de producción primaria disponibles. Nuestra base de datos revela un océano netamente heterotrófico con grandes consecuencias para el calentamiento global.

RESUMEN (en Inglés)

Marine bacterioplankton play a key role in the biogeochemical cycles, mainly in the global carbon cycle. It is responsible for a large fraction of the heterotrophic respiration in the ocean. In the current context of global warming, quantifying and predicting the role of bacterioplankton in the respiration and the comparison of these values with the existent estimates of primary production is a scientific challenge and represents the main objective of this study. To achieve this, a deeper study of which are the main factors and the effect that they have on the distribution of abundance and metabolism of marine bacterioplankton is required.

We have collected a global database compiling data of bacterial abundance and other environmental variables (mainly temperature and chlorophyll). In the first part of this manuscript (Chapter 1 and 2) methodologies for analysing flow-cytometry processed files have been developed. Flow-cytometry is a routine tool for the automated processing of bacterioplankton groups in aquatic systems. In Chapter 1, we developed a methodology for automated clustering of marine bacterioplankton groups. The automation and, thus, the standardization of flow-cytometry heterotrophic bacterioplankton data processing allows us to compare samples from different sources which represents an advantage for processing large-scale dataset. In Chapter 2, we also used the same flow-cytometry files to develop a methodology for estimating the 'cytometric diversity' of every single flow-cytometry sample. The methodology was applied to a time-series dataset from the Cantabric Sea and the estimates were validated with a genetic diversity dataset. As a main result we obtained that it is possible to detect changes in marine bacterioplankton communities at large scale using flow-cytometry. This represents a breakthrough in marine microbial ecology due to the applicability, short time of analysis required and low cost of flow cytometry over molecular tools.

In Chapter 3, we present an allometric study of metabolism of natural marine bacterioplankton size-fractionated samples along a latitudinal gradient in the Atlantic Ocean. This study reveals, for the first time in nature, that bacterioplankton respiration rate scales superlinearly with cell size which, despite being so small, affects their metabolism. This is due to the close relationship between size and metabolic complexity of these organisms. The relationship between metabolism and cell size has been found by size-fractionating the community where all cells were under the same external conditions.

We were thus able to measure the respiration of each size fraction separately. However, when we compared the metabolism of communities from different areas with their mean size, the effect was not clearly observed due to the narrow size range of these organisms.

In several studies mainly focused on surface waters, chlorophyll has been revealed as a key factor controlling bacterioplankton distribution. However, less progress have been made in the vertical dimension.
In Chapter 4, we used the global database for studying the marine bacterioplankton distribution. In this Chapter we have characterized chlorophyll and bacterial abundance profiles fitting the data to a Gaussian curve. Also, we studied the coupling of both variables with depth using an artificial neural network. After the validation of the coupling of both variables we were able to apply a non-parametric model for predicting the depth-integrated bacterial abundance distribution.

Recent studies have reported that cell-specific metabolic activity can be estimated by multiplying the bacterial abundance and cell-specific metabolic activity using the Metabolic Theory of the Ecology. In this Chapter we used the predicted abundance and the metabolic activity in an attempt to predict the bacterioplankton respiration in the photic layer at global scale. We also compared our approach with the current primary production estimates. Our database reveals a net heterotrophic ocean with important consequences for global warming.

SR. DIRECTOR DE DEPARTAMENTO DE BIOLOGIA DE ORGANISMOS Y SISTEMAS/
SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN RECURSOS BIOLOGICOS Y BIODIVERSIDAD

Contents

Resumen

El bacterioplancton marino desempeña un papel fundamental en los ciclos biogeoquímicos, principalmente en el ciclo del carbono, siendo responsable de una fracción importante de la respiración heterotrófica. En el contexto de calentamiento global actual el poder cuantificar y predecir el papel del bacterioplancton en la respiración y comparar esos valores con las estimas de producción primaria existentes se ha convertido en un desafío científico y constituye uno de los principales objetivos de esta tesis. Para ello, es necesario profundizar en el conocimiento de cuáles son los principales factores y qué efecto tienen sobre la distribución de la abundancia y el metabolismo del bacterioplancton marino.

Para lograr estos objetivos se ha recopilado una extensa base de datos donde se incorporan datos de abundancia bacteriana y otras variables ambientales (temperatura y clorofila principalmente). En la primera parte de este manuscrito (Capítulo 1 y 2) se han desarrollado metodologías de análisis de archivos obtenidos mediante la técnica de citometría de flujo, herramienta rutinaria para el análisis automatizado de poblaciones de bacterioplancton en sistemas acuáticos. En el Capítulo 1 se ha desarrollado una metodología para detectar poblaciones de bacterioplancton marino de forma automática. Esta automatización y, por lo tanto, estandarización del proceso de detección de poblaciones de bacterias heterotróficas marinas nos permite hacer comparables muestras de distintas fuentes lo que es de gran utilidad cuando se trabaja con bases de datos extensas. En el Capítulo 2 se han utilizado también archivos procedentes de análisis citométricos para desarrollar una metodología que nos permita estimar la "diversidad citométrica" de cada muestra de citometría de flujo. Este análisis se aplicó a una serie temporal de datos en el

mar Cantábrico y los resultados fueron validados con datos de diversidad genética estimados con técnicas moleculares. Este análisis nos permitió comprobar que se pueden estimar cambios en las comunidades de bacterioplancton marino a gran escala utilizando la citometría de flujo, lo que representa un gran avance en ecología microbiana por la aplicabilidad, rapidez y bajo coste de esta técnica frente al uso de técnicas moleculares.

En el Capítulo 3 se ha realizado un estudio metabólico de alometría de bacterioplancton marino utilizando comunidades muestreadas in situ y separadas por su tamaño a lo largo de un gradiente latitudinal en el Océano Atlántico. Este estudio nos ha permitido demostrar por primera vez en el medio natural que las bacterias tienen una respiración superlineal en función de su tamaño, el cual, a pesar de ser tan pequeño, influye en su metabolismo. Esto es debido a la estrecha relación entre tamaño y complejidad metabólica de estos organismos. Esta relación del metabolismo con el tamaño de las bacterias ha sido encontrada al fraccionar una comunidad dónde todas sus células están bajo las mismas condiciones ambientales, en tamaños y medir la respiración de cada clase de tamaño por separado. Sin embargo, al comparar el metabolismo de comunidades de distintas zonas con su tamaño medio este efecto no se observa debido al estrecho rango de tamaños en el que se mueven estos organismos.

La clorofila ha sido la variable señalada en varios artículos centrados principalmente en aguas superficiales, como una de las principales variables que controlan la distribución de bacterioplancton marino. Sin embargo, esta relación no se ha cuantificado en la dimensión vertical. En el Capítulo 4 se utiliza la extensa base de datos global recopilada para estudiar la distribución de abundancia del bacterioplancton marino. En este capítulo se han caracterizado perfiles verticales de clorofila y de abundancia bacteriana ajustándolos a una distribución Gaussiana. Una vez caracterizados se ha estudiado el acoplamiento de varias variables en la vertical utilizando redes neuronales artificiales. De esta forma, hemos podido comprobar que ambas variables están relacionadas para así poder utilizar un modelo no paramétrico en el que a partir de variables superficiales se pueden predecir la distribución de abundancia bacteriana.

Resumen

En trabajos recientes, la temperatura ha sido revelada como el parámetro con mayor influencia en la respiración de comunidades de microorganismos heterotróficos. Estos trabajos han demostrado que la respiración planctónica se puede estimar multiplicando la abundancia bacteriana por la actividad metabólica celular utilizando la teoría metabólica de la ecología. En este capítulo hemos utilizado la abundancia predicha y la actividad metabólica celular para realizar un primer intento de predecir la respiración bacteriana global en la capa fótica utilizando imágenes de satélite y comparar esta respiración con las predicciones de producción primaria disponibles. Nuestra base de datos revela un océano netamente heterotrófico con grandes consecuencias para el calentamiento global.

General Introduction

Marine microbes are ubiquitous in the oceans (Fenchel and Finlay, 2004; Giovannoni and Stingl, 2005). They thrive in all regions (Polar, equatorial, tropical, subtropical, estuarine, coastal, oceanic) and throughout the water column from the surface to the dark deep ocean (Aristegui et al., 2009; Gasol et al., 1997). Thanks to their metabolic versatility marine bacterioplankton are involved in all biogeochemical cycles (Ducklow et al., 2001; Falkowski et al., 2008; Kirchman, 2000) playing a key role in the carbon cycle through the Microbial Carbon Pump (Figure 1) (Jiao et al., 2010). Of all organisms forming the microbial carbon pump, heterotrophic bacteria are the most important in the assimilation of organic dissolved matter (DOM) (Azam et al., 1983; Jiao et al., 2010).

Heterotrophic bacteria (bacteria and Archaea) are the most abundant organisms in the ocean (Azam et al., 1983; Giovannoni and Rappe, 2000). Despite their tiny size, bacterioplankton cells may contain on average around 12 fg C in open ocean (Fukuda et al., 1998). The total heterotrophic bacterioplankton biomass in the open ocean is quite large and becomes as important as phytoplankton total biomass in some regions such as oligotrophic areas (Kirchman, 2000). Furthermore, bacteria have the largest relationship between surface area and volume of all microorganisms, which is an advantage for the incorporation of the dissolved compounds allowing them to out-compete other microbes (Kirchman, 2000).

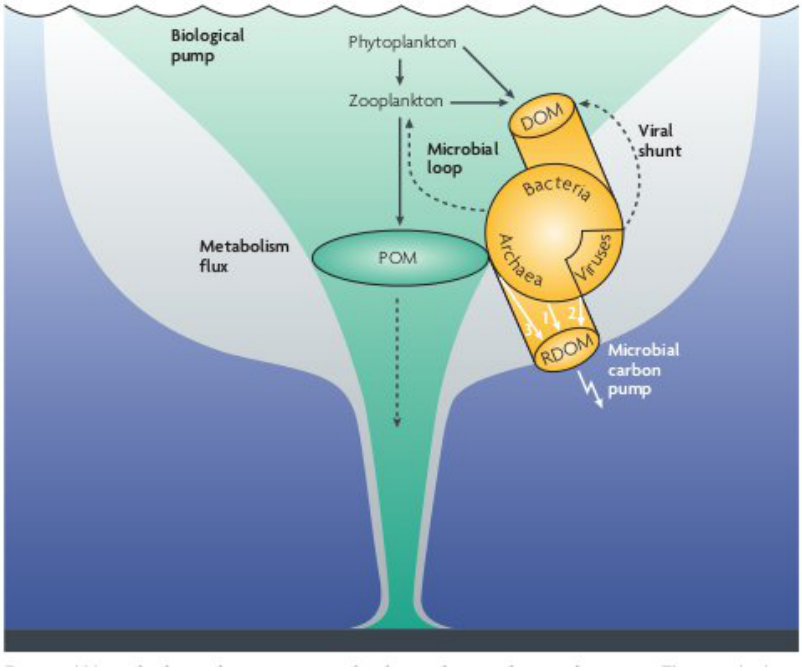

Figura 1: The Microbial Carbon Pump. (Adapted from Jiao et al. (2010))

The role of flow cytometry in the study of bacterioplankton macroecology

In the last decades, flow cytometry has become a fundamental tool in aquatic microbial ecology. Flow cytometry allows the automated analysis of a large number of microscopic cells per second thanks to their properties of light scatter and fluorescence. This tool allows the processing of a large number of samples per day. Thus, flow cytometry is specially used for large-scale studies and has greatly contributed to the comprehension of the macroecological patterns of microbial communities (Li, 2009). Flow cytometry is not only used for counting a great number of cells, but it also allows a quite accurate study of the individual cell size. This has provided a global view about how bacteria are distributed in the ocean in terms of their abundance and biomass (Buitenhuis et al., 2012).

A further step in the study of large-scale patterns of marine microbes was made by Li (1997) with the introduction of the term “cytometric diversity”. Using

the cytometric properties of the cells, Li (1997) classified phytoplankton groups into cytometric categories to achieve an estimate of diversity for these groups. Therefore, flow cytometry has been an essential tool in the macroecological study of microbial communities.

Macroecological patterns of heterotrophic bacterioplankton

A spatio-temporal correlation between bacterial abundance with phytoplankton biomass and temperature has been reported (Gasol and Del Giorgio, 2000; Gasol et al., 1997; Li, 1998; Li et al., 2006, 2004). However, this relationship may vary depending on nutrients conditions. When nutrients are scarce, phytoplankton abundance decrease causing a higher resource limitation for heterotrophic bacterioplankton communities (bottom-up control) (Ducklow and Carlson, 1992). In areas with high resources levels, bacterioplankton abundance decreases associated to the influence of graziers (top-down control) (Gasol, 1994; Li et al., 2004). The ratio heterotrophic/autotrophic biomass varies with the trophic conditions becoming larger under oligotrophic conditions and decreasing with increased eutrophic conditions due to losses by bacterivory and viral lysis (top-down control) (Li et al., 2004).

Bacterioplankton abundance is also strongly correlated with temperature, especially for temperatures below 15 °C. For temperatures above 15 °C the relationship has been reported to be weaker mainly due to resource limitation (Li, 1998). Using a large-scale oceanic database from samples analysed by microscopy and flow cytometry, Li et al. (2004) found a significant relationship by combining both log chlorophyll concentration and temperature with log bacterial abundance. Li et al. (2004) found a general pattern in the relationship between the ratio of bacterial abundance to chlorophyll concentration with temperature. They found maximum bacterial abundance in waters of intermediate levels of chlorophyll concentration and temperature. Fuhrman et al. (2008) found that temperature also explains the latitudinal diversity pattern in marine bacterioplankton.

Heterotrophic bacteriplankton metabolism

Bacterioplankton metabolism is key in the oceanic carbon budget. A large proportion of the marine microbial respiration is carried out by heterotrophic bacterioplankton (5-59%) (del Giorgio and Duarte, 2002; García-Martín et al., 2014; Robinson and Williams, 2005). Heterotrophic prokaryotes are able to efficiently sequester the dissolved organic matter (DOM) generated by planktonic primary producers and other allochthonous or autochthonous inputs.

Both bacterial production and respiration have a great influence in the flux of dissolved organic matter (DOM) in the ocean controlling the dynamic of the microbial loop (Ducklow, 2000). Several studies have evidenced that bacterioplankton metabolism may vary spatio-temporally in the ocean, mainly through the effects of temperature on physiological processes, especially at cellular level (Brown et al., 2004; Kirchman et al., 2005; López-Urrutia and Morán, 2007; Mazuecos et al., 2015; Pomeroy and Wiebe, 2001; Robinson and Williams, 1993; White et al., 1991). Hitherto, temperature has been the main factor associated with bacterial respiration while their production is strongly linked to resource supply (Kirchman et al., 2005; López-Urrutia and Morán, 2007; Regaudie-de Gioux and Duarte, 2013; Scofield et al., 2015; Yvon-Durocher et al., 2010).

Bacterioplankton and climate change

Predicting the metabolic balance of the ocean has been a major goal for marine biogeochemists. An increase in temperature may have dramatic effects on planktonic communities. A reduction in primary producers and, consequently an expansion of the oligotrophic areas in the ocean are expected (Boyce et al., 2010; Sarmiento et al., 2004). The feedback between warming and the capability of the ecosystems to incorpore CO_2 would lead to a decline in the carbon uptakes by planktonic communities (del Giorgio and Duarte, 2002; López-Urrutia et al., 2006).

A half of primary production occurs in the oceans and it is well known the main role of microbes as a keystone of the carbon cycle due to their capability of incorporate DOM and their contribution to the respiration of the ocean. An increase

in the temperature thus may affect the dynamic and the structure of the organisms inhabit the oceans (Figure 2).

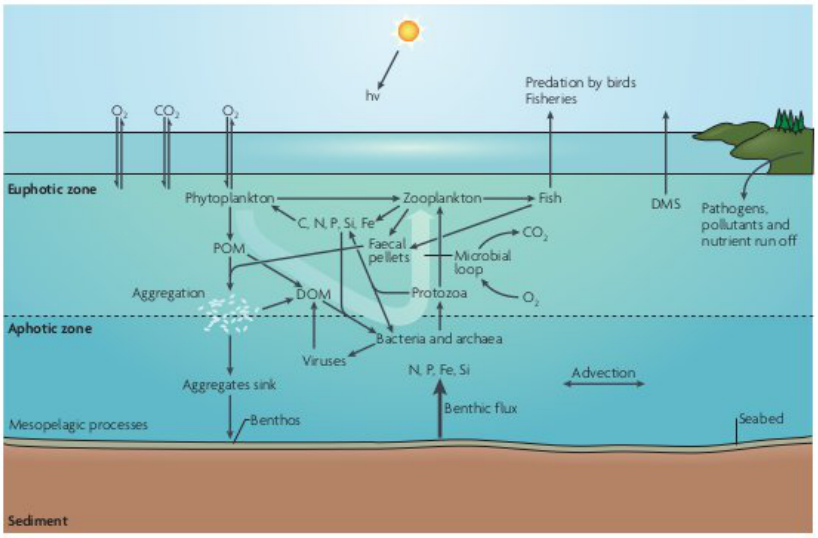

Figura 2: Microbial structure in a marine ecosystem. (Adapted from Azam and Malfatti (2007))

Therefore the understanding of the distribution, diversity and metabolic activity patterns of bacterioplankton for predicting the response of these organisms to increases in temperature is thus required for improving our knowledge about the effect of global warming in the oceans.

Objectives

The main objective of this thesis is to improve our knowledge about the major factors driving the distribution of bacterioplankton abundance. We also aim to use the main patterns found to model bacterioplankton abundance distribution and predict heterotrophic bacteria respiration. To achieve this goal we have first compiled a global database with a large spatio-temporal coverage.

The specific objectives of each chapter are:

Chapter 1

- To developed a methodology for automated clustering bacterioplankton flow cytometry data.
- To test the validity of the methodology for identifying bacterioplankton groups using a long-term database from a coastal environment manually analyzed.

Chapter 2

- To test the applicability of cytometric diversity for detecting changes in the bacterioplankton communities by comparing changes in bacterial community composition analysed by flow cytometry and using a molecular approach (454-pyrosequencing).
- To find ecological patterns of bacterioplankton communities analysed by flow cytometry.

Chapter 3

- To determine the effect of size on the bacterioplankton cell-specific respiration rate in marine natural samples.
- To estimate the scaling exponent between cell size and metabolic rate for bacterioplankton size-fractionated cells along a latitudinal gradient in the Atlantic Ocean.

- To test the universality of the obtained scaling exponent.
- To identify the factors that could affect the allometric relationship between cell size and respiration rate for prokaryotes.

Chapter 4

- To compile a dataset of marine heterotrophic bacterioplankton abundance and several ancillary variables with a large spatio-temporal coverage for analysing the macroecological patterns in marine microbial communities. To this end, more specific objectives were established:
  - To find variables highly correlated with bacterioplankton abundance, especially that can be estimated by remote sensing.
  - To characterize chlorophyll and bacterioplankton abundance profiles and to study how coupled are both variables in the photic layer.
  - To applied an artificial neural network to the characterized vertical profiles to identify the different types of profiles and how are they distributed.
  - To predict bacterioplankton abundance at a global scale using a model based on remotely sensed variables.
  - To calculate bacterial community respiration using the Metabolic Theory of the Ecology and to compare heterotrophic picoplanktonic respiration with global primary production estimations by remote sensing.

Chapter 1

Automated clustering of heterotrophic bacterioplankton in flow cytometry data

Francisca C. García, Ángel López-Urrutia & Xosé Anxelu G. Morán
Published in Aquatic Microbial Ecology, Vol. 72, pp.175–185, (2014)

1.1 Introduction

Heterotrophic prokaryotes make up the largest living biomass of aquatic ecosystems, playing a key role in the carbon cycle by incorporating dissolved organic matter and recycling nutrients in the oceans (Azam et al., 1983; Hansell and Carlson, 1998). The estimation of total bacterioplankton abundance has evolved from the tedious estimation by microscopy techniques to automatic counting using flow cytometry. Flow cytometry allows counting at a rate of 200 to 2000 cells per second. Fast sample processing by flow cytometry has proven very useful for large-scale studies of plankton communities (Gasol and Del Giorgio, 2000). In addition to abundance estimates, interest in automated flow cytometric properties arises from the increasing use of side scatter (SSC) signals as surrogates of cell size (Calvo-Díaz and Morán, 2006; Felip et al., 2007), thus allowing for precise estimation of the biomass contributed by each subgroup.

The analysis of bacterial population abundance by flow cytometry consists of two separate steps. The first step, which is truly automated, is the processing of the sample under the flow cytometer. The second, not automated step, is the posterior analysis of the flow cytometer output. Flow cytometry uses cell properties of light scattering and fluorescence, and records the information of each analyzed cell in a flow cytometry standard (FCS) file. The analysis of the samples by flow cytometry is automatic but the clustering of the data stored in the FCS files is usually done manually. Two widespread groups of heterotrophic bacteria separated by their relative nucleic acid (NA) content, after appropriate NA staining, and commonly referred to as low (LNA) and high (HNA) are found in almost any aqueous sample (Bouvier et al., 2007; Gasol et al., 1999; Li et al., 1995), from freshwater, estuarine, to open-ocean waters. After some debate, a consensus is emerging that LNA and HNA subgroups comprise fundamentally different phylotypes (Schattenhofer et al., 2011; Vila-Costa et al., 2012). These and other groups usually have to be identified by means of drawing the limits between populations in scatterplots, a process referred to as gating. Manual gating is done in 2-dimensional graphical representations of flow cytometry variables of SSC and fluorescence (most frequently red and green). However the limit between

populations is not always clear and significant subjectivity is introduced based on the analyst‘s criterion. In addition, it is often hard to visually discriminate subgroups in flow cytometry samples (Andreatta et al., 2004; Finak et al., 2009), introducing an important error in the bacterial population estimates. The magnitude of this error is even more important in large-scale studies where flow cytometry files are processed by different analysts.

In recent years, many automatic techniques for gating populations have been developed to minimize the time-consuming step of manual processing. Different clustering methods have been proposed (Aghaeepour et al., 2013; Bashashati and Brinkman, 2009; Lahesmaa-Korpinen et al., 2011; Rajwa et al., 2008; Scheuermann et al., 2009). A clustering technique that has given good results in medical and health sciences is model-based clustering. Model-based clustering is an unsupervised clustering technique, meaning that it attempts to classify the data in a given number of homogeneous groups without the need for training by the user with a set of classified examples. This technique tries to find the best model that describes the structure in the data. Lo et al. (2009) proposed a model-based clustering approach based on t mixture models with a Box-Cox transformation. Contrary to previous methods, their algorithm can detect groups with elliptical shape and unbiased to the presence of outliers. One additional advantage of automatic clustering techniques is that the number of groups detected is variable so for example it is possible to easily detect subgroups within the HNA and LNA categories usually detected by manual gating.

Our objective is to evaluate the performance of t-mixture model-based clustering for HNA and LNA bacterioplankton from a coastal environment. We develop a methodology for automated clustering of bacterioplankton flow cytometry data and tested it with a long-term database that has also been analyzed manually. Our method is an adaptation of the flowClust function (Lo et al., 2009) for the identification of bacterioplankton groups. In addition we provide an evaluation of the errors introduced by the subjectivity in the manual gating of bacterioplankton by comparing the analyses of the same flow cytometry plots by different analysts.

1.2 Material and Methods

1.2.1 Sample collection and analysis using the traditional manual gating

A monthly 9 yr time series of flow cytometry bacterioplankton samples collected under the RADIALES programme (Spanish Institute of Oceanography, IEO) at 3 stations off Gijón/Xixón in the southern Bay of Biscay was used to validate the methodology developed. Samples were collected at 8 depths (from the surface to 150 m depth). Bacterioplankton samples were preserved with 1% paraformaldehyde + 0.05% glutaraldehyde and frozen at -80 °C. To analyze heterotrophic bacterioplankton, an aliquot of 0.4 ml was stained with 2.5 μmol L^{-1} SYTO-13 NA fluorochrome (Molecular Probes) and analyzed using a FACSCalibur flow cytometer (BD/Becton, Dickinson and Company) equipped with a laser emitting at 488 nm.

We used a total of 2050 files in this intercomparison. Flow cytometry scatterplots are routinely analyzed manually as part of the ongoing time-series programme. Manual gating has been performed by different analysts (although they all receiving the same training), so a certain degree of variability can be expected in the manually gated data. Fluorescent latex beads (1.0 um diameter, Molecular Probes) were added as an internal standard to relate the measured SSC and fluorescence signals, which might change slightly from sample to sample due to laser drift, to the constant SSC and fluorescence of beads. LNA and HNA bacteria, were easily distinguished within bacterioplankton based on their relative green fluorescence (FL1, a variable related to NA content, (Marie et al., 1997)). LNA cells were almost always smaller (lower SSC values) than the HNA counterparts (Calvo-Díaz and Morán, 2006).

1.2.2 Automatic gating

1.2.2.1 Detection of beads

Because beads are used as the standard they need to be identified very precisely. In many samples, the beads population is a mixture of single beads, doublets, triplets and higher associations. Beads usually lay in an area of the cytogram where no other population exists, making it easier to separate these subpopulations. Hence, we have developed an additional methodology to analyze the beads first, so that they could be used later to correct bacterioplankton data. For automatic analysis of beads, we first selected a window of SSC and FL1 containing the beads population for all samples. A window wide enough is needed to ensure that the beads do not lay outside of the selected window if slight changes in the laser occur.

In this window, the population of single beads is as a bi-normal distribution (i.e. the beads have normal SSC and FL1 distributions). This bi-normal distribution of single beads sometimes has 2 or 3 adjacent subpopulations (doublets, triplets, etc). To automatically gate the subpopulation of single beads within this wide window, and since the single bead population is the most abundant, we first located the SSC and FL1 corresponding to the most abundant histogram class. Then, for the specific brand of beads used in our samples, a range of ± 0.2 log SSC and FL1 units around the mode encompassed all the single beads but removed doublets and triplets and was therefore used to select a reduced window containing all single beads.

Once this reduced window was automatically selected we fitted a normal distribution to the SSC data. For a large number of samples the beads population was superimposed over background noise. Therefore, the beads population was defined as those particles falling within the 99 % confidence intervals of the fitted normal distribution. The same analysis was performed automatically for all selected FCS files. If different beads were used, the ranges of SSC and FL1 for the window selection would need to be readjusted.

1.2.2.2 Detection of bacterioplankton groups

Although bacterioplankton tend to show marked bimodal distributions of FL1, distinct groups may be occasionally hard to detect because they are composed of different subgroups with similar SSC and FL1. In addition, not all groups of bacterioplankton are distributed following a normal distribution. *t*-mixture model-based clustering is supposedly able to cope with these peculiarities. We use the R package flowClust (Clustering for Flow Cytometry) (Lo et al., 2009) to evaluate the ability of this technique to cluster natural bacterial populations.

1.2.2.3 Software

FlowClust uses a model-based clustering approach based on *t* mixture models with a Box-Cox transformation. In *t*-mixture models the number of subpopulations to detect needs to be specified a priori. FlowClust package uses a Expectation-Maximization algorithm (EM) to calculate the parameters when fitting t-distributions to the *n*-subpopulations (Lo et al., 2008), selecting the best fit. To test the goodness of the fitted *t*-distributions, flowClust compares the error and tries to minimize it. A modified Box-Cox transformation is needed to minimize the effects of outliers present in the flow cytometry files. The resulting classification allows each cell to be assigned to one of the *n*-subpopulations. Lo et al. (2008) provides a more detailed description of the flowClust algorithm.

1.2.2.4 File analysis

FCS files were read into R using the Bioconductor package flowCore (Hahne et al., 2009) providing flow cytometric signals for each cell. The variables used in the flowClust function are SSC, FL1 and red fluorescence (FL3), characteristic of autotrophic cells. Once the flow cytometry file is read, we have a multiparameter matrix with all values of each of the cells of that file. Data were log-transformed for subsequent analyses.

As we have explained above, to apply a classification using *t*-mixture models we need to specify the number of clusters (K) to be fitted. But the number of

bacterioplankton populations in a cytogram changes, there can be several groups of noise (eg. viruses, sample contamination). To select the best number of clusters in a sample, and given that the number of subpopulations was rarely greater than 10, we fitted 10 *t*-mixture models to each cytogram, with K from 1 to 10. For each K we calculated the Bayesian Information Criterion (BIC) (Lo et al., 2009) relative to the maximum value of BIC for the 10 fitted models. The most appropriate value of K for each cytogram was selected as the lower K with BIC > 95 %.

We filtered the flow cytometry data before the fit of the 10 *t*-distribution models. Filtering data prior to analysis is a common practice in automatic clustering methodologies (Andreatta et al., 2001; Ribalet et al., 2011; Zare et al., 2010) that reduces the effect of noise and speeds up the analysis. However, this pre-treatment has to be carefully applied to not affect the biologically important information. The filtering procedure is based on a binning of the data, so an image representation of the data is produced. Each pixel of the image represents the number of cells present in the given combination of SSC and FL1. To filter the data, if less than 2 cells fall in a pixel, they are removed. Three sequential filters were performed on each FCS file. First, a 3-dimensional filter (SSC, FL1, FL3 with 40 bins in each axis), then 2-dimensional filters (first SSC vs FL1, then FL1 vs FL3 with 80 bins in each axis). This pre-filtering is an optional step prior to the clustering function, therefore, it could be used or not depending on the expert criterion.

Once we have the optimal number of clusters for each file, the flowClust function returns each of the K groups and each cell was assigned to a given group. The flowClust is performed on the filtered data and hence, even if the threshold to filter the data is carefully selected, the number of cells belonging to each group could be underestimated. To correct this underestimation, we first calculated the boundaries of the cloud of points in the 3-dimensional (SSC, FL1 and FL3) space for each of the K groups. Then, with the unfiltered data we used these boundaries to assign each cell to a group. If a cell is not within any group it, was labelled as an outlier. To calculate the boundaries of a group we used a convolution algorithm that calculates the smallest polygon that includes the input points (using function convhulln of the package “geometry” available in R).

Figure 1.1 shows an example summary graphic of the flowClust results for one file. FlowClust returns a numbered list of groups, but we do not know which cluster represents each subgroup of bacteria (HNA and LNA in this case) or whether the cluster is in fact a group of noise. To assign each cluster to a category, we designed a method to label each group once all FCS files were analysed. We first calculated the mean value of the flow cytometry parameters for each cluster. Then we corrected these data using mean values for the beads for the corresponding FCS file. We plot the corrected mean value for each group using SSC, FL1 and FL3 signals. We obtained a plot of all the centres of the groups of all the FCS files analyzed. We gated this plot manually to select which groups were assigned as HNA or LNA (Figure 1.2). The purpose of this step is to label the groups with names instead of numbers.

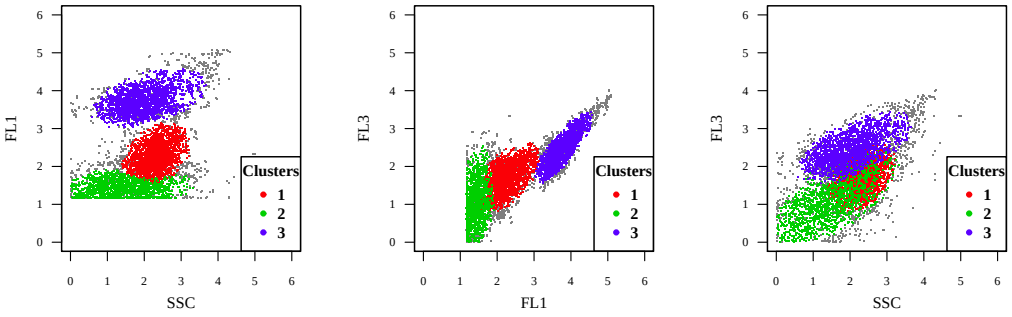

Figura 1.1: Clustering output obtained by the flowClust function for an example file. The axes show the flow cytometer variables used for clustering: Green fluorescence (FL1), Red fluorescence (FL3) and side scatter (SSC)

The R code to automatically analyze bacterioplankton FCS files is provided as annex.

1.2.2.5 **Validation**

We compared our automatic clustering with the manual analysis of 2050 files. The automated analysis of the beads was also compared to the manual analysis. Bacterioplankton groups were divided into two categories, HNA and LNA, to match the groups detected in the manual-gated database processing. To determine

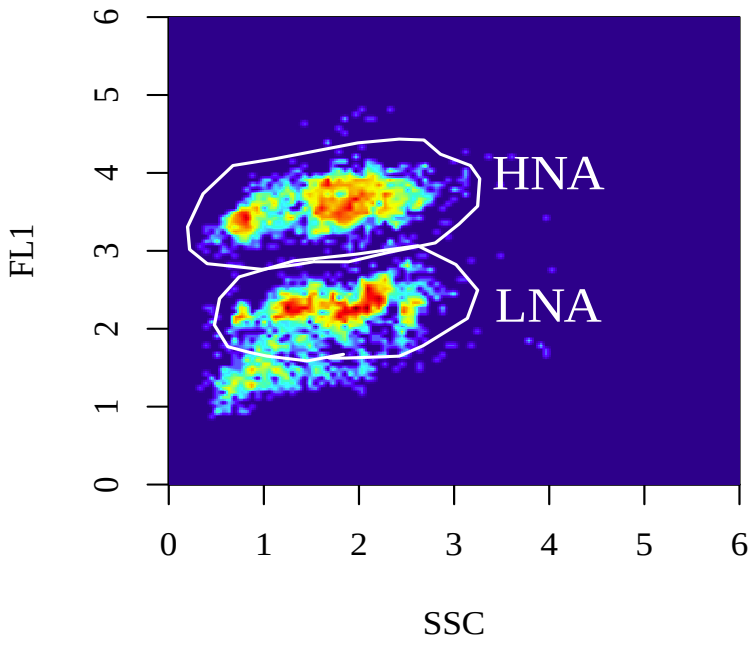

Figura 1.2: Manual gating in the labelling step of flowClust output groups using the parameters side scatter (SSC, x-axis) and green fluorescence (FL1, y-axis). HNA:high nucleic acid content; LNA: low nucleic acid content.

the inherent variability in the manual analysis, 10 FCS files randomly selected from Radiales database were analyzed by 6 different experts. For the 2050 files analysed by the ongoing Radiales monitoring program we have information on the cytometric properties for each group, and mean size values calculated using the arithmetic mean. Since the distribution of flow cytometric properties is frequently not normal, other measures of central tendency like the median might be more appropriate. To evaluate which measure of central tendency is more adequate, we manually analysed 50 files selected to have maximum variability in the data and including a wide range of depths and dates using the R-package. Files manually were analyzed by the same person. We also provide error estimations between different methods and experts.

1.3 Results

The correlation between the automatic clustering and the manual gating data was high for the counts (Figure 1.3A-1.3B-1.3C, $r = 1$ for beads, $r = 0.96$ for LNA, and $r = 0.97$ for HNA groups) but rather low for the average SSC (Figure 1.3E-1.3F, $r = 0.37$ for LNA bacteria, and $r = 0.56$ for HNA bacteria). These estimates of average SSC represent the arithmetic mean of the SSC of the particles within each group, but many bacterioplankton groups have side-scatter distributions which are not normally or even log-normally distributed. Hence, the median of the distribution is a more reliable estimate of the average side-scatter as it is less sensitive to data distribution. Because the Radiales data set was analyzed using the arithmetic mean and it was not feasible to manually reanalyze the 2050 FCS files, we selected 50 files and manually gated the data. We calculated the median instead of the arithmetic mean for both automatic and manual methods and a better correlation between the median SSC estimates was found (Figure S1.1C-S1.1D, $r = 0.76$ for LNA bacteria; $r = 0.99$ for HNA bacteria).

However, a good correlation is not a measure of the correspondence between two methodologies. Usually, when comparing two analyses, a test on whether the slope and intercept of the fit are significantly different than 1 and 0 is provided. For the data shown in Figure 1.3 all slopes and intercepts are significantly different than 1 and 0 respectively. Even for the beads where the differences between the two methodologies are irrelevant (i.e. when manual counts were 1000 beads the automatic method counted 1011) the ANOVA reveals significant differences (test for slope different to 1: $F_{1,2036} = 85, p < 0{,}001$). This suggest that in our database of 2050 FCS files, the ANOVA had so much power that even the slightest differences are detected as statistically significant and, therefore, we can not rely just on the p-value to study the goodness-of-fit of our method.

Despite this problem with exceedingly high power of the ANOVA, the significance of the tests points out to dissimilarities between the methodologies. However, it is also required to study the magnitude and whether these differences are biologically important (Peters, 1991).

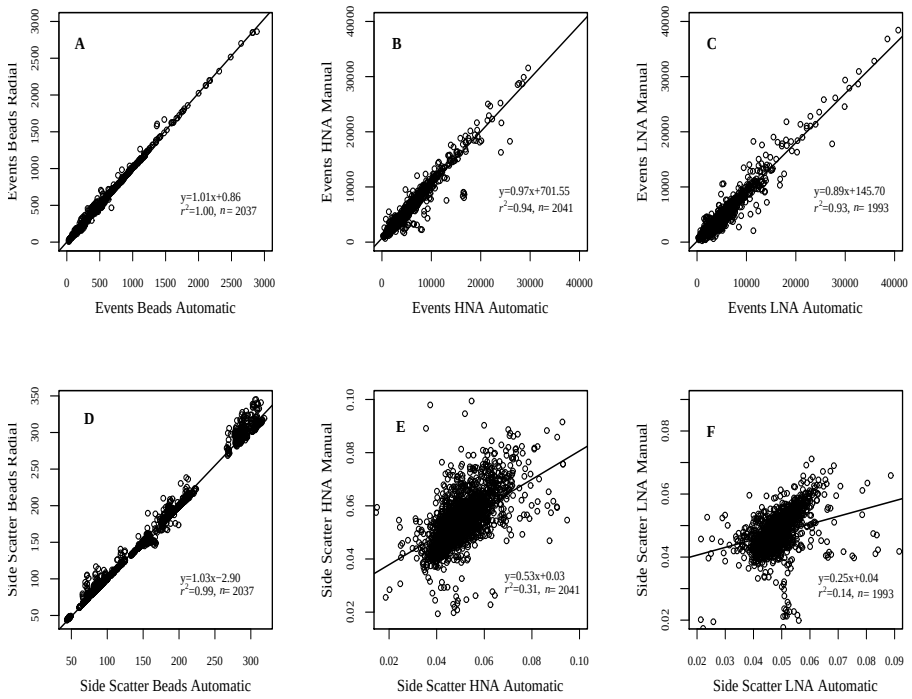

Figura 1.3: Correlation between the output of manual (y-axis) and automatic (x-axis) methods. Upper panels (A, B, C) show the counts and lower panels the side scatter for (A, D) beads, (B & E) high nucleic acid content (HNA), and (C, F) low nucleic acid content (LNA) groups.

To detect the disagreement or bias between the both methodologies, we estimated the error, which represents the ratio between the automatic and the manual methods (Figure 1.4A-1.4B). A positive error hence indicates an overestimation by the automatic method while a negative error implies an underestimation by the automatic method. When this error was calculated we found important differences between the manual and the automatic methods both in the HNA and the LNA bacteria groups, especially at the lowest counts (Figure 1.4A-1.4B).

To understand the origin of the error, we evaluated the influence of the error introduced by the analyst who manually gated the data. We therefore asked an external analyst who was not involved in the processing of the Radiales database

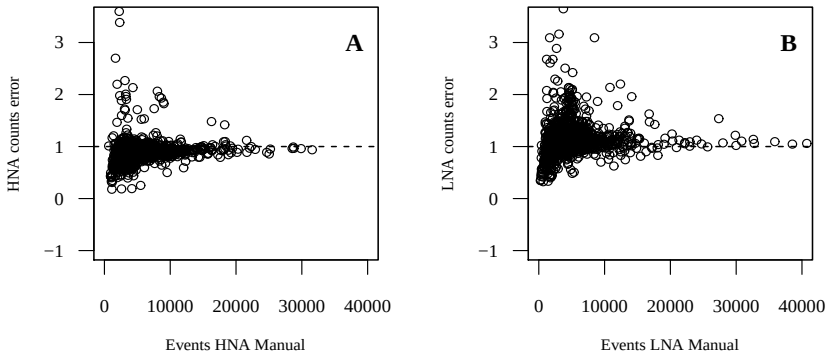

Figura 1.4: Relationship between error (defined as the ratio between the automatic and manual counts) and the number of events analyzed for (A) high nucleic acid content (HNA) and (B) low nucleic acid content (LNA) bacteria using the whole Radiales database.

to manually gate the same 50 files (Figure 1.5C-1.5D). Although the error between the 2 manual gating analysis (Figure 1.5C-1.5D) was lower than the comparison between automatic and manual (Figure 1.5A-1.5B), the differences observed suggest that even when the gating was done manually by 2 different experts important errors were found.

Interestingly, the errors of the latest comparison, and in particular the errors in Figure 1.5C, suggest that the external analyst underestimates the abundance at the lowest counts, very similarly to the automatic method. We hypothesized that this underestimation is due to the fact that when an analyst is asked to gate 50 files for an intercomparison study, he/she pays special attention to adjust the gates as much as possible trying to avoid the inclusion of adjacent noise or outliers. To test this hypothesis we asked an analyst involved in the routine processing of the Radiales database to gate these 50 files (we refer to this analysis as re-analysis data). The counts were quite similar to the external analyst (Figure 1.5E-1.5F) and different to the routine database counts (data not shown). The error was higher for HNA at the lowest counts. We tested for significant differences in the counts for each intercomparison in figure 1.5 using paired *t*-tests. For HNA bacteria, counts obtained by the external analyst and the re-analysis counts were not significantly

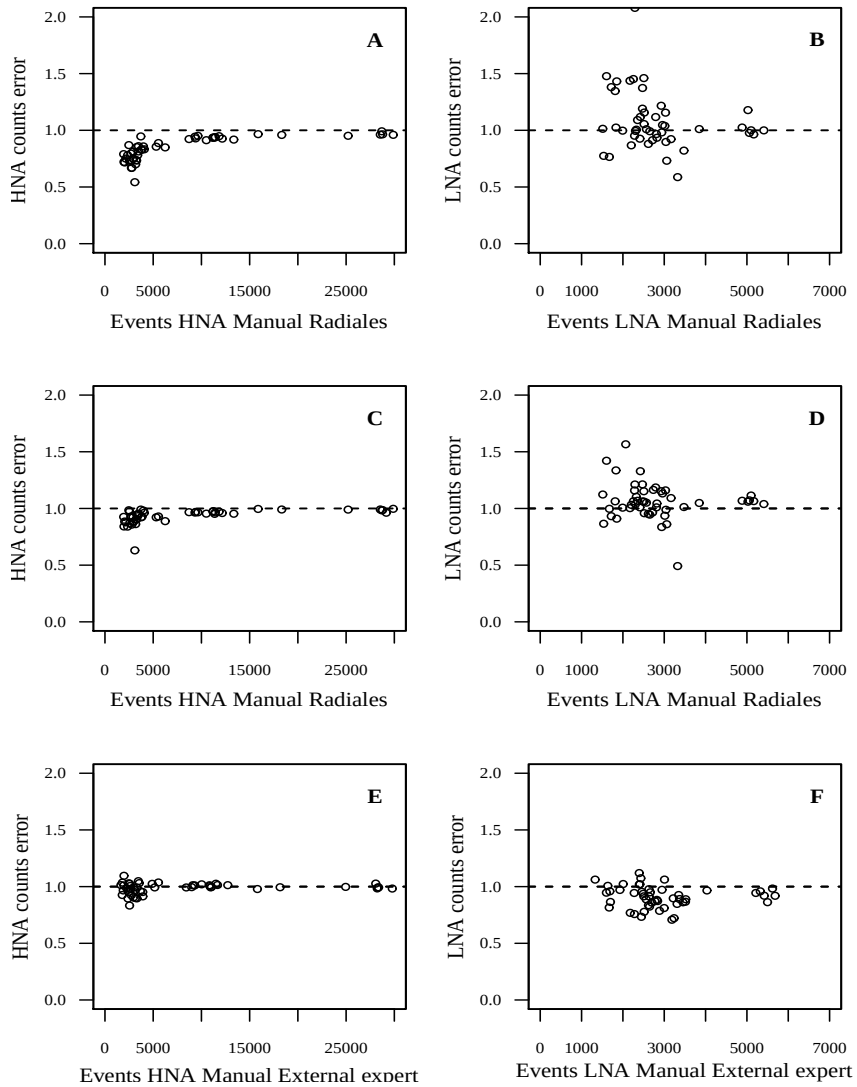

Figura 1.5: Error between different analysts and the automatic and manual methods for bacteria counts using 50 files. (A, B) Ratio between automatic and manual gating (y-axis) against the manual counts (x-axis). (C, D) Ratio between the counts manually obtained by an external expert and the manual counts of the Radiales database. (E, F) Ratio between the gating of an external expert and another expert who collaborated in the processing of the Radiales database. Dashed line is the 1:1 relationship.

different (Figure 1.5E: *t*-test, t = -1.47, df = 49, p = 0.15). However, for the other two cases significant differences were found (Figure 1.5A: *t*-test, t = -20.51, df= 48,

$p < 0.001$; Figure 1.5C: t-test, t = -10.59, df = 49, $p < 0.001$). A different tendency was found for LNA sub-group with significant differences between counts by the external analyst and the re-analysis (Figure 1.5D: t-test, t = 2.7, df = 49, p = 0.009; Figure 1.5F: t-test, t = -7.43, df = 49, $p < 0.001$), and no differences between automatic and manual gating (Figure 1.5B: t-test, t = 1.88, df = 47, p = 0.07).

To obtain a quantitative estimate of the influence of the analyst, we tested the variation in the counts obtained by 6 experts. When the same flow cytometry file was analyzed by different experts the coefficient of variation between HNA and LNA bacterial groups counts was relatively high, with values around 10-20 %. For beads counts the variability dropped to around 5 % (Figure 1.6).

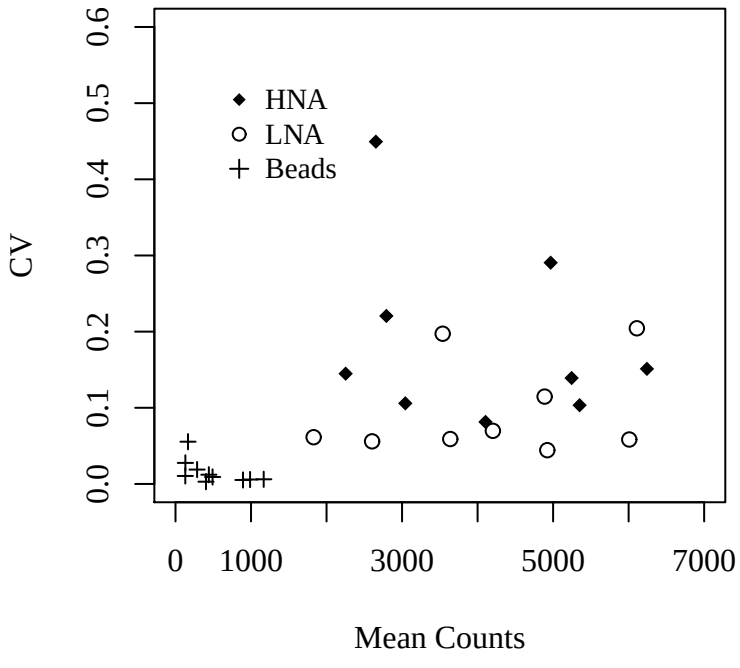

Figura 1.6: Relationship between the coefficient of variation (CV) for 10 flow cytometry standard (FCS) files analyzed by 6 different experts and the mean counts. HNA: high nucleic acid content, LNA: low nucleic acid content.

Similarly to Figure 1.5, the variation between experts as higher for samples with lower abundance of cells for beads and HNA groups, however we did not observe this effect for the LNA subgroup.

Finally, to study if the differences between methods are ecologically significant, we calculated a mean annual cycle using all available surface data at station 3. Seasonal cycles of total bacterial abundance obtained by manual and automatic methods were similar (Figure 1.7A).

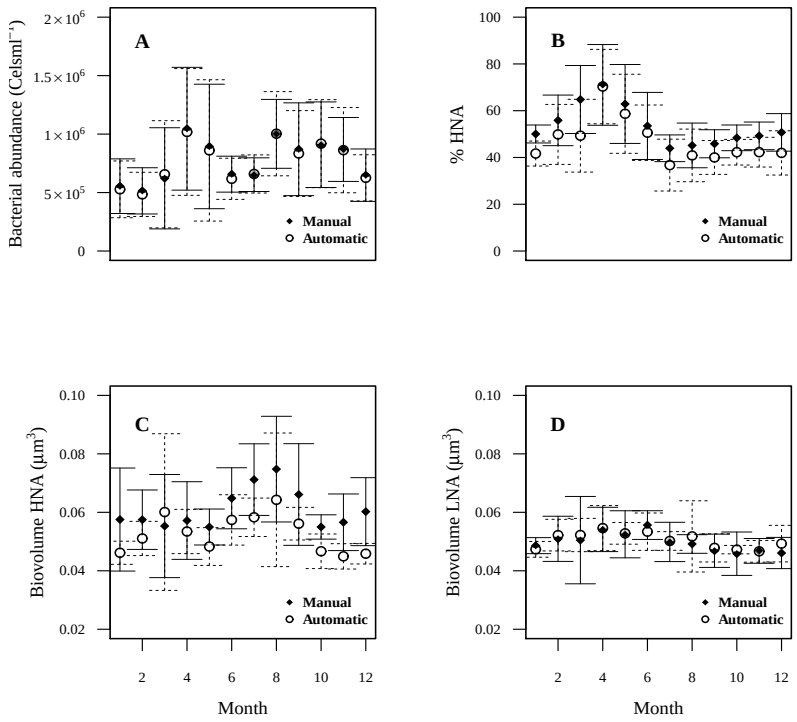

Figura 1.7: Mean annual cycle at the surface of the RADIALES station 3 off Gijón, Spain, for the 2002-2010 period obtained by the manual and automatic methods. (A) monthly mean bacterial abundance counts, (B) monthly mean contribution of high nucleic acd content (HNA) cells, (C) mean biovolume for HNA and (D) mean biovolumes of low nucleic acid content (LNA) cells. Solid error bar: manual SD; dashed error bar: automatic SD.

However, when the contribution of HNA cells (% HNA) was examined, some differences were found (Figure 1.7B) yet showing a similar pattern. Figure 1.7C-

1.7D show the seasonal cycles of mean LNA-HNA cell biovolume. Differences were higher for the HNA group (13%) than for LNA (1.4%) but the pattern was quite similar.

1.4 Discussion

Recently, several automatic techniques for flow cytometry data processing have been widely applied in medical fields (Aghaeepour et al., 2013; Le Meur, 2013; Robinson et al., 2012). However, in microbial ecology, experts still rely largely on manual gating of the FCS files. One peculiarity of flow cytometry samples of bacterioplankton when we process samples is the high variability between samples hampering automated analysis. Although some of these automated methods have been applied to analyze planktonic groups (Andreatta et al., 2004; Ribalet et al., 2011) these methodologies are not able to cope with such dynamic features, due to the presence of noise or variability of the community composition, usually encountered in large scale studies. For example, Ribalet et al. (2011) used a set of pre-defined windows where picoplankton populations should lay and the method detected the spatial dynamics of each group of bacterioplankton within each window. Therefore, their method is not readily applicable whenever the position in the cytogram of each group changes. This is frequently the case in large datasets. Andreatta et al. (2001) used image analysis to identify subgroups of bacterial populations in FCS files but again, manual gates were required to distinguish each bacterial population from the background noise.

Another special feature of bacterioplankton FCS files is that the number of subgroups and the boundaries between them are not always clearly defined. Frequently it is even hard to identify them manually. Usually, it is easy to differentiate between the 2 widespread groups of bacterioplankton assessed here (HNA and LNA), but often other subgroups are also present. In addition, the presence of outliers in the cytogram is quite common and not all automatic methods respond adequately to the presence of a significant contribution of them (Luta, 2011). The model-based clustering used in the flowClust algorithm is robust to the presence of outliers (Finak et al., 2010; Lo et al., 2009). Another interesting

feature of this method is the ability to detect non-elliptical population shapes, such as HNA (e.g. cluster 3 in Figure 1.1) cell distribution in contrast to the more circular of the LNA bacteria (e.g. cluster 1 in Figure 1.1), as it is usually the case in aquatic samples (Bouvier et al., 2007). Although other automated techniques are unbiased for population shape, they are not able to detect overlapping groups (Ge and Sealfon, 2012; Naumann et al., 2010; Naumann and Wand, 2009; Sugar and Sealfon, 2010). Finally, other methods cannot be applied to large data sets due to computational efficiency (Zare et al., 2010). In addition, the flowClust algorithm is provided as open software and it is thus free to use and modify.

The main aim of this work was to apply the flowClust algorithm to develop an automatic and standardized method for processing flow cytometry analysis of heterotrophic bacterioplankton groups that could be routinely adopted as an alternative of manual processing. We tested it under a real case scenario using a large dataset and comparing the results to the traditional, manual gating technique. We used a database consisting of 9 yr of monthly sampling continental shelf bacteria from the surface down to 150 m, characterized by a wide range of natural variability at the seasonal, and spatial (inshore-offshore and vertical gradients) scales (Calvo-Díaz and Morán, 2006; Morán and Calvo-Díaz, 2009).

As we have explained in the results section, bacterioplankton automatically clustered groups were aggregated into two categories, HNA and LNA, to be able to compare them with the manual method. Manual gating is more limited to identify bacterioplankton subgroups or even cyanobacteria, specially when groups such as *Prochcloroccus* overlap with HNA group in natural bacterioplankton samples. Nevertheless, the methodology we propose is able to detect a higher number of bacterioplankton groups and subgroups as there is no a priori restriction on the number of groups that can be detected. However, we recommend to visually inspect flowClust output to understand the results of the automatic clustering. For example, in open ocean surface samples where *Prochcloroccous* populations frequently partially fall within the HNA cell cluster, our methods detected them quite accurately (data not shown). Nevertheless in these type of conditions both in-vivo and stained samples are usually analysed so it would be easier to automatically

gate both samples in a similar way as it is done with manual gating to differentiate autotrophic and heterotrophic cells.

We have shown that the subjectivity of the analyst can introduce around 10-20 % of variability in the manual gating of bacterioplankton samples (Figure 1.6). The consequences of this variability are important especially in large-scale studies where data from different analyst are combined. An automatic approach such as the one we have developed certainly does better at comparing different data sets producing more consistent counts. Despite its importance the subjectivity of manual gating has not been quantified previously. We have shown that not only different analysts reach different counts due to this subjectivity error but that also that the same expert when analysing a large amount of samples can introduced some bias due to fatigue.

The quality of the samples analyzed by flow cytometry is important for the results of the automated analysis. Similarly, when gating is done manually, samples are analyzed more effectively if all the groups are centered with the adequate number of cells per sample and without much noise. In testing the method, major differences between automated and manual counts were due to the presence of these problematic files. It is therefore important to follow recommendations (Gasol and Del Giorgio, 2000; Marie et al., 1997) on using bacterioplankton optimal flow rates and number of cells per sample. At higher rates, populations begin to overlap and it is more difficult to set limits between them, even automatically.

Despite previous filtering was applied to the data, the method is relatively fast (< 5 min. per sample). The processing time for our method is similar to the time spent to process FCS files manually, but the clustering is unsupervised and the time required is only computing time and not analyst time. However, we strongly recommend a visual checking of the analysis output. This method has the advantage to be more objective and reproducible. It is possible to reduce time of analysis using a server with multiple processors as we did. Moreover, the fact that the FCS files are read using the flowCore package allows to have access to the information of each cell in the file. This information allows to know the distribution of any group and thus improve the subsequent statistical treatments. We propose that the beads

should be analyzed before the clustering technique using another method because beads measurements are required to correct for the deviations that cytometer lasers can experience with time (Shapiro, 1995).

FlowClust provides an effective tool for gating the populations, mainly for counts. Significant differences were still found in biovolume estimates (with an empirically determined calibration between SSC and cell diameter (Calvo-Diaz2006)), specially for the HNA group (Figure 1.7B), which was 13 % higher. While the LNA group has a more or less spherical population shape, the HNA group is usually more irregular and with a non-spherical shape. Consequently, the measure of central tendency that is to calculate the mean size introduces more variability for the HNA case (Figure 1.7B-1.7D). % HNA values obtained by the automatic method (Figure 1.7C) is usually lower than the % HNA obtained by manual gating although the distinct seasonal pattern of maxima in April and minima in July (Calvo-Díaz and Morán, 2006; Morán and Calvo-Díaz, 2009) was well reproduced. This is due to both an underestimation of the abundance of HNA cells and an overestimation of LNA cells by the automatic method (Figure 1.4). These methodological deviations between manual and automatic analysis, can result into different ecological patterns in cell size and biomass estimates. These differences become more important in large-scale studies or when we compare different databases. In this type of studies the importance of using objective and standardized methods of clustering in microbial ecology becomes critical.

In summary, our methodology is a powerful tool to analyze groups and subgroups of heterotrophic bacterioplankton, allowing the processing of thousands of files, quickly and with reduced error. The computer-based processing of the FCS files results in the full automation of sample analysis by flow cytometry. It increases the efficiency and quality of the results and makes them comparable with other data from large-scale studies. The technique could be easily adapted to the analysis of phytoplankton samples or even extended to the analysis of viruses.

1.5 Supplementary figure

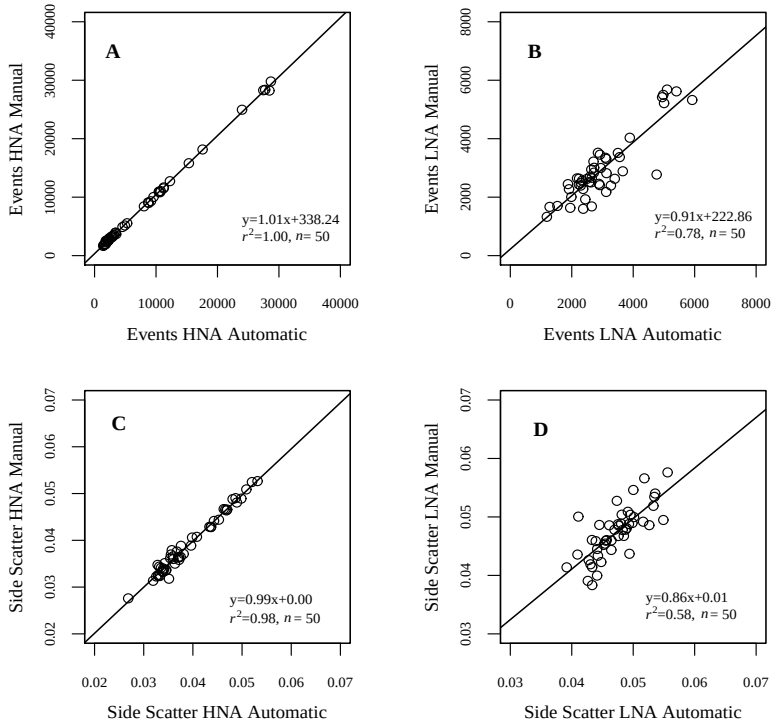

Figura S1.1: Correlation between the output of manual (y-axis) and automatic (x-axis) methods using 50 files from the Radiales database. The upper panels show the relationship for counts of (A) high nucleic acid content (HNA) bacteria and (B) low nucleic acid content (LNA) bacteria, while the lower panels show the comparison between median side scatter of (C) HNA bacteria and (D) LNA bacteria.

Chapter 2

Seasonality in molecular and cytometric diversity of marine bacterioplankton: the reshuffling of bacterial taxa by vertical mixing

2.1 Introduction

Marine bacteria are the most diverse component of planktonic communities and constitute one of the largest biomass stocks in the ocean (Buitenhuis et al., 2012; Giovannoni and Rappe, 2000; Venter et al., 2004). Identifying microbial taxa present in the environment and understanding the composition of microbial communities over spatio-temporal gradients has been a challenge in microbial ecology studies (Giovannoni and Stingl, 2005; Green et al., 2008; Pielou, 1975; Zinger et al., 2011). The first approach to identify bacterial taxa involved their isolation from the environment, and subsequent study in culture. However, the culturability of environmental bacteria is extremely low (Amann et al., 1995), which hampered a reliable characterization of in situ bacterial communities until the application of molecular methods (Giovannoni et al., 1990; Olsen et al., 1986).

During the last two decades, the use of molecular tools has completely changed the way ecologists study the distribution and diversity of marine bacteria taxa (Zinger et al., 2012). Different seasonal, spatial and latitudinal patterns in bacterioplankton phylogenetic composition have been reported using fingerprinting methods such as Denaturing Gradient Gel Electrophoresis (DGGE), Automated Ribosomal intergenic Spacer analysis (ARISA) or Terminal Fragment Length Polymorphysm (TRFLP) (Chow et al., 2013; Fuhrman et al., 2008; Morris et al., 2005; Schauer et al., 2000). While these methods enable a fast comparison of samples, other approaches such as Fluorescence In Situ Hybridization (FISH), have permitted the determination of the abundance of specific target populations or quantitative changes in the phylogenetic composition with a reasonable accuracy (Amann and Fuchs, 2008; Cottrell and Kirchman, 2000; Pernthaler et al., 2002; Schattenhofer et al., 2009). Recently, the development of high-throughput sequencing technologies has represented a breakthrough in this field by increasing several orders of magnitude the number of sequences typically obtained per sample (Salter et al., 2014; Sogin et al., 2006; Vergin et al., 2013; Zinger et al., 2011).

While all molecular approaches have their own advantages and disadvantages (e.g. Alonso-Sáez et al. (2007)), a common limitation is that they are time-consuming and/or expensive for large-scale studies requiring the analysis of

hundreds to thousands of samples. In early ecological studies, other classification criteria such as the photosynthetic pigment concentration were proposed to study the structure and composition of phytoplankton communities (Margalef, 1968). Li (1997) introduced the concept of 'cytometric diversity'to characterize marine phytoplankton communities, using individual cell properties measured by flow cytometry. In natural and stained samples, cytometric properties such as light scatter and fluorescence reflect changes in cell size, complexity, pigment, and nucleic acid content, and therefore, they can be used to classify planktonic cells. While the power of this method in differentiating bacterial taxa is limited to a few (i.e. < 10) groups (Schiaffino et al., 2013) as compared to molecular approaches, its potential applicability to detect changes in community composition would represent an important advance. As compared to molecular analyses, flow cytometry is technically simple, requires small volumes and allows a relatively fast analysis of hundreds to thousands of samples, being a routine method in aquatic microbiology laboratories.

Here, we analysed the seasonal dynamics of heterotrophic bacterioplankton in a decadal time-series in the Southern Bay of Biscay using the cytometric diversity approach. While seasonal changes in the composition of coastal marine bacterioplankton have been widely reported (Alonso-Sáez et al., 2007; Alonso-Sáez et al., 2015; Campbell et al., 2009; Chow et al., 2013; Gilbert et al., 2012), the factors controlling this seasonality are less understood. Using the power of the cytometric method, we performed a large-scale analysis at three continental shelf stations in the Cantabrian Sea (2586 samples in total), at a spatio-temporal resolution not easily achievable with molecular techniques. Our objectives were: (i) to test whether ecological meaningful patterns can be derived from cytometric diversity analysis on marine bacterioplankton, and (ii) to understand the factors responsible for seasonal changes in the diversity and composition of coastal bacterial communities.

2.2 Material and Methods

2.2.1 Sampling sites

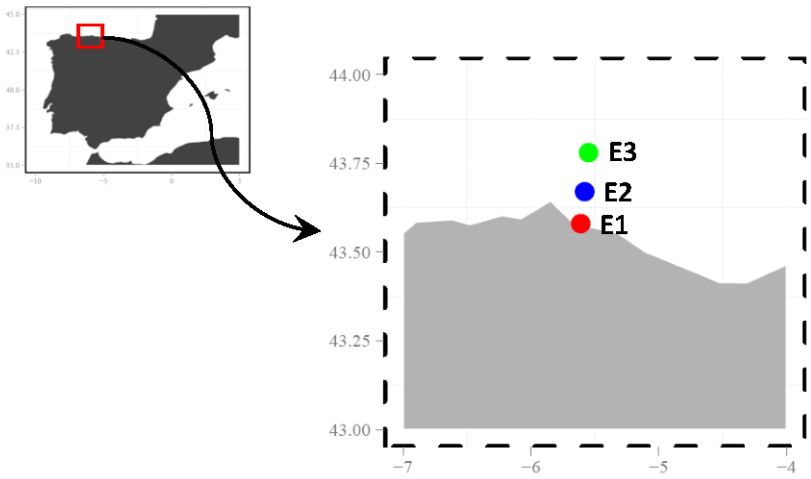

Figura 2.1: Sampling stations. E1, E2 and E3 samples were used for cytometric diversity estimates. E2 surface samples were used for genetic diversity.

Samples were collected monthly off Gijón, from April 2002 to December 2012 as part of the programme RADIALES of the Spanish Institute of Oceanography (IEO). This 10-year time-series study includes 3 stations sampled along a transect perpendicular to the coast located in the continental shelf at 43.58°N, 5.61°W (Station 1, E1), 43.67°N, 5.58°W (Station 2, E2), and 43.78°N, 5.55°W (Station 3, E3)(Figure 2.1). Samples for flow cytometry analysis were collected over the 10 year period at 4 depths at E1 (from the surface to 20 m depth), 8 depths at E2 (from the surface to 100 m depth) and 9 depths at E3 (from the surface to 150 m depth). DNA samples for 454 pyrosequencing were obtained over 3.5 years (July 2009 to December 2012) at the surface (0 m) at E2.

A SeaBird 25 CTD was used to collect temperature and salinity data. The density (σ_t) gradient was used to estimate the mixed layer depth (MLD). The MLD was defined as the depth where a change in density of $\geq$ 0.05 kg m^{-3} for a 5 m interval was first found. Samples for nutrients analyses (NO_3^-, NO_2^- and PO_4^-)

were collected concomitantly with flow-cytometry samples and were frozen until analysis with a Technicon autoanalyzer following the methodology described by Grasshoff (1976).

2.2.2 Flow cytometry analysis and cytometric diversity estimates

Samples for flow cytometry were preserved with 1 % para-formaldehyde + 0.05 % glutaraldehyde and stored at -80 °C until processing, usually within one month of collection. For flow cytometry analysis, 0.4 ml of sample was stained with 2.5 μmol L^{-1} DMSO-diluted SYTO-13 nucleic acid fluorochrome (Molecular Probes) and analyzed with a FACSCalibur flow cytometer (BD/Becton, Dickinson and Company) with a laser emitting at 488 nm. 10 μl volume of 1.0 μm-diameter fluorescent latex beads (Molecular Probes) was added to each sample as an internal standard. Flow cytometry standard (FCS) files were gated automatically following the methodology described in García et al. (2014). This methodology uses a model-based clustering for automatically detect the different groups of cells that appears in a sample and allows us to separate heterotrophic cells from noise or phototrophic cells. Side scatter (SSC), green fluorescence (FL1) and red fluorescence (FL3) are common flow cytometric variables used to identify bacterioplankton groups (Gasol and Del Giorgio, 2000). The settings used for detecting these cytometric parameters for bacterioplankton cells were 400 nm for SSC, 511 nm for FL1 and 590 nm for FL3. Hence, these variables were used in the flowClust function for the automatic clustering of the heterotrophic bacteria populations.

The methodology used by Li (1997) to estimate phytoplankton diversity was followed and adapted to bacterioplankton analysis. For each bacterioplankton FCS file, we selected the clusters corresponding to heterotrophic prokaryotes.

We focused our study on heterotrophic bacterioplankton. Although both Synechococcus and Prochclorococcus appeared in our samples, their abundance was much lower, thus data obtained from cytograms soecifically targeting heterotrophic prokaryotes are less reliable (Calvo-Díaz and Morán, 2006). Therefore, we did not include cyanobacterial clusters in the analyses. The flow cytometer detection channels allow differenciate the cells according to their optical properties. SSC and

FL1 are more important for differentiation of heterotrophic bacterioplankton communities than FL3. Hence, in our analysis we gave more weight to SSC and FL1 by clustering the detection channels into 16 bins each, while FL3 channels were grouped into just 4 bins. We used FL3 simply to obtain a 3-dimensional view of the community to separate heterotrophic bacteria from noise or cyanobacteria groups more efficiently (Gasol and Del Giorgio, 2000). We then estimated the diversity of each sample using the number of cells falling into each of the 1024 (16x16x4) resulting categories. Figure 2.2 shows an example with two FCS files where SSC and FL1 parameters were divided into 16x16 bins obtained from a sample with relatively low and high cytometric diversity (Figure 2.2-A and 2.2-B, respectively).

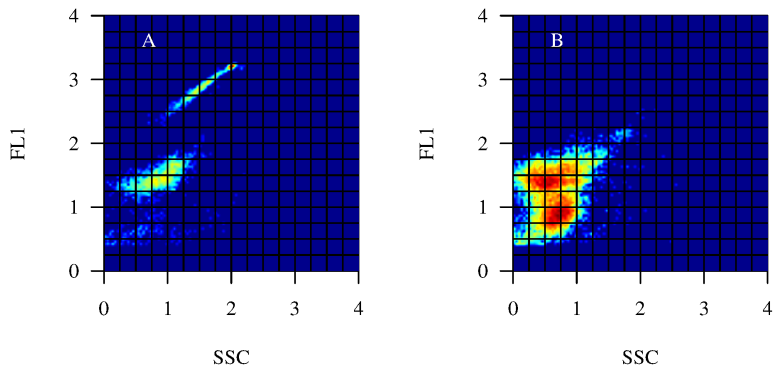

Figura 2.2: Comparison between two flow cytometry standard (FCS) files corresponding to May 2012, when a low cytometric diversity was found (A) and January 2012, when a high cytometric diversity was found (B). The vertical axis is green fluorescence (FL1) and the horizontal axis is side scatter (SSC).

2.2.3 Genetic diversity estimates by 454 pyrosequencing

Samples (0.5 to 2 L) for pyrosequencing were collected monthly from surface waters at E2 using Niskin bottles. Samples were pre-filtered through GF/A filters (Whatman) and filtered onto 0.2 μm diameter polycarbonate filters (Millipore). DNA extraction was performed using Power Water DNA Isolation kit (Mobio).

DNA was re-suspended in 50 μl MilliQ water and quantified using Nanodrop (Thermo). Bacterial sequences were amplified by the use of domain-specific bacterial primers (341F and 805R, Herlemann et al. (2011)) complemented with a 10 bp sample-specific bar-code sequence on the reverse primer. PCR reactions (20 μL) contained a final concentration of 1 μM of each primer, 0.2 mM of each dNTP and 0.02 nM of polymerase Phusion High Fidelity Taq Polymerase. The template DNA concentration varied between 1 and 10 ng per reaction. The PCR cycles started with a 5 min initial denaturation at 98 °C, followed by 25 cycles of 98 °C for 40 sec, 53 °C for 40 sec, and 72 °C for 60 sec and a final 7 min elongation at 72 °C. Triplicate PCR reactions were pooled for each sample. The purified amplicons were subject to pyrosequencing from the reverse primer using a 454 FLX+ system. The raw reads were quality trimmed (with a minimum Phred average quality score of 35 over a 50 bp window) and aligned against a reference SILVA alignment, keeping only those positions that start and stop in the same alignment space. From 488 593 reads initially screened, 248 474 reads were considered of good quality and with a minimum alignment length of 140 bp for further analyses. The resulting sequences were denoised by a preclustering method allowing 1 mismatch (Huse et al., 2010) as recommended (Schloss et al., 2009). Chimeras were removed using Uchime. For OTU construction, reads were clustered at genetic distance cutoffs 0.01 and 0.03 substitution per nucleotide, using the average linkage method. Taxonomic assignment of the OTUs was obtained by classification with SILVA taxonomy using the Wang approach (Wang et al., 2007) as implemented in Mothur. The average confidence in the taxonomic assignment was 99.6 % ± 3.3 % and 98.9 ± 5.6 % for OTUs at the 0.01 and 0.03 cutoff levels. Chloroplasts and cyanobacteria identified by this method were removed from the dataset, and samples were downsized to 4174 reads by random resampling to equal the depth of sequencing of all samples before calculating Shannon diversity indices. Sequences of this study have been deposited in the European Nucleotide Archive (accession number PRJEB6399).

2.2.4 Determination of diversity indices

We calculated Shannon-Weaver diversity indices for the molecular and flow-cytometry dataset to compare the diversity patterns of bacterioplankton commu-

nities. In the case of cytometric diversity, this calculation was based on the total number of cytometric categories detected and the relative abundance of cells in each category. In the case of the molecular dataset, the Shannon diversity index was calculated based on the total number of OTUs detected (at 97 % similarity level) and their evenness in the samples.

The similarity between samples by each method over a 3.5 y period was calculated using Bray-Curtis similarity index, and a Mantel test was performed to compare the two similarity matrices. Bacterioplankton flow-cytometry data and molecular data were standardized before similarity analysis. We also calculated pairwise Bray-Curtis similarity matrices between flow cytometry samples across the depth profile for the three sampling stations. For that purpose, we assessed the vertical variability in species bacterial community composition by calculating, for each station and date, the similarity index between the cytometric composition at the surface and each of the other sampled depths. This vertical similarity analysis was applied to every monthly profile along the seasonal cycle during the 10-year time-series.

All the analyses were performed with R (R Team (2013)) and statistical tests were performed using the R package "vegan" (Oksanen et al., 2012).

2.3 Results

2.3.1 Comparison of genetic and cytometric diversities

In the molecular dataset from E2 (obtained by 16S rRNA tag pyrosequencing over 3.5 years), a cyclical pattern of diversity with maximum Shannon-Weaver index values in winter and minima in early summer was found (using a OTU genetic distance cutoff of 0.03, Figure 2.3). This seasonality is clear both in the monthly time-series (Figure 2.3A) and in the average annual cycle for the 3.5-years period (Figure 2.3B). As expected, the Shannon Weaver diversity indices were higher when using a more stringent distance cutoff for OTU construction (0.01), but the same monthly and average annual patterns of diversity were observed (Figure S2.1, S2.2).

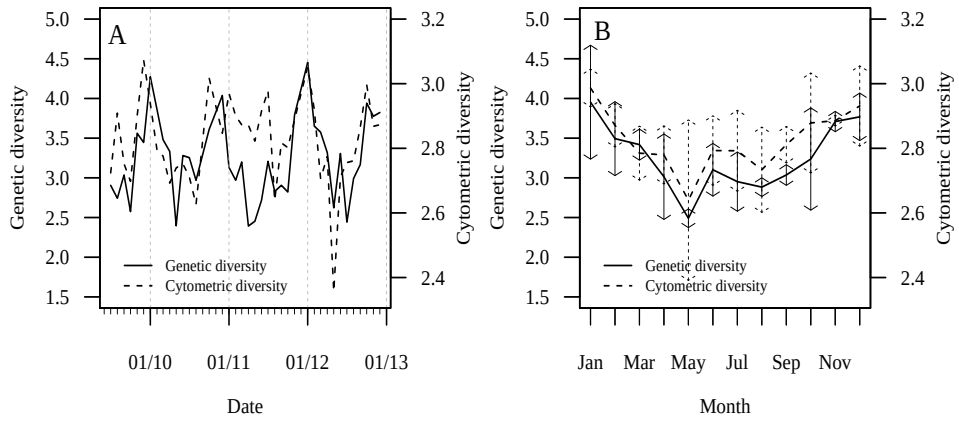

Figura 2.3: Trends of genetic (solid line) and cytometric (dashed line) diversity (Shannon-Weaver index) of surface bacterial communities at E2 during a 3.5-year monthly time series (A) and as a mean annual cycle (B).

The molecular approach has a finer resolution at detecting different bacterial groups as compared to the cytometric approach and, thus, absolute Shannon diversity values obtained by flow cytometry were generally lower (Figure 2.3). However, the cytometric approach provided a similar general trend in the diversity of bacterial communities along the 3.5 years, although with a weaker cyclical pattern (Figure 2.3A). The largest differences between genetic and cytometric diversity were found during 2011, when relatively high cytometric diversity values were observed, especially in spring and early summer compared with anual minima in genetic diversity (Figure 2.3A).

Despite this discrepancy, the average annual cycles had the same pattern with both methodologies, with highest Shannon-Weaver index values during winter and lowest in late spring (Figure 2.3B). While a significant, although a weak correlation was found between individual estimates of Shannon diversity obtained by both methods (Figure 2.4A, $r = 0.46$, $p < 0.01$), monthly averages were highly

correlated (Figure 2.4B, r = 0.93, p < 0.01). These results indicate that despite some individual discrepancies, cytometric diversity can reliably capture general trends in the bacterioplankton community composition.

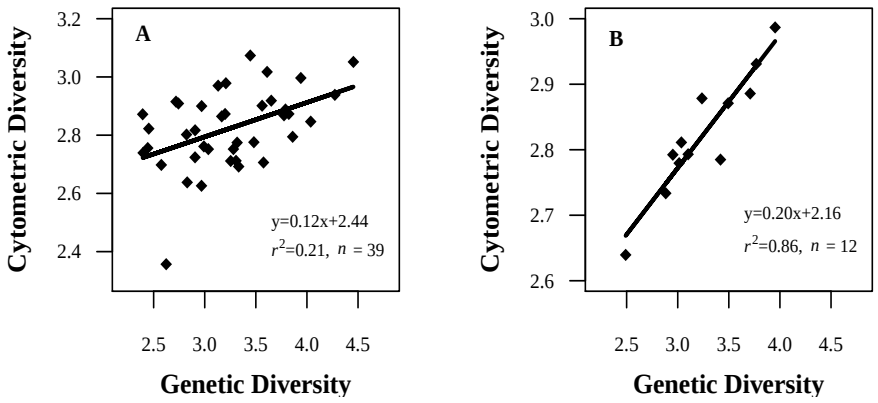

Figura 2.4: Correlation analysis between genetic and cytometric diversity (Shannon-Weaver index) using the 3.5-year monthly sampled database and mean annual cycle.

The similarity of the bacterial community composition based on the molecular and cytometric approaches was also calculated using the Bray-Curtis metric. A Mantel test confirmed a significant relationship between the similarity matrices obtained by both techniques (Figure S2.3, Mantel statistic r = 0.59, p = 0.001).

2.3.2 Cytometric diversity patterns in the Cantabrian Sea

Overall, the decadal seasonal dynamics for the 3 stations analyzed were similar to the pattern found for the E2 using the 3.5-year period (figure 2.3), with higher cytometric diversity in early winter and late autumn than in summer (Figure 2.5). In order to understand the factors leading to temporal changes in surface bacterial cytometric diversity, we tested correlations with different ancillary environmental variables (chlorophyll, temperature, salinity, inorganic nutrients, day length and mixed-layer depth, MLD). Within the variables considered using the 10 year time series, three parameters showed significant correlations of similar strength: day length, inorganic nutrients, (particularly NO_2-NO_3) and MLD. The three variables showed similar correlations with the surface decadal seasonal cytometric diversity

at the three stations studied (Day length: r = - 0.69, - 0.61 and - 0.54; NO_2-NO_3: r = 0.66, 0.62 and 0.23; MLD: r = 0.61, 0.64 and 0.44, for E1, E2 and E3 respectively).

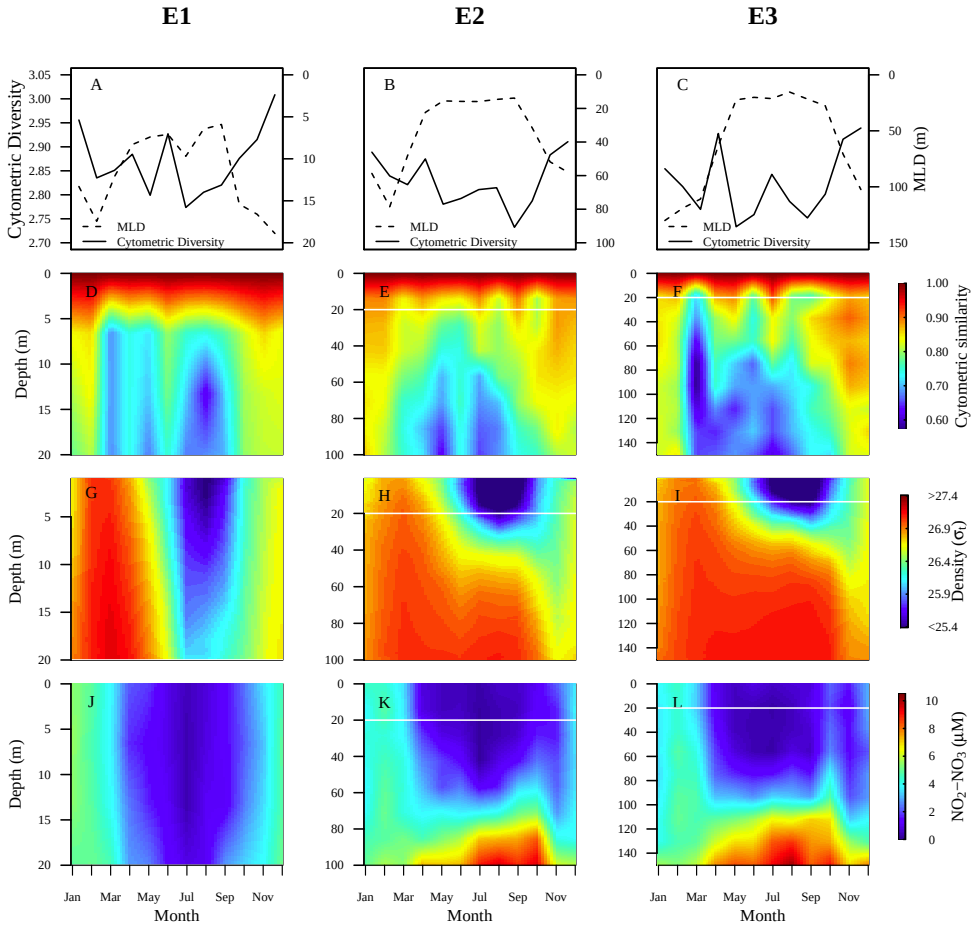

Figura 2.5: Seasonal dynamics along the spatial and vertical gradient of the cytometric diversity and MLD (A, B, C), community similarity (D, E, F), potential density (G, H, I) and nutrient concentration (J, K, L) data estimated from a 10-year monthly time series database collected in the Cantabrian Sea. Community similarity was estimated as the Bray-Curtis similarity between surface data and the different depths. The white line (E, F, H, I, K, L) represents the 20 m depth. Warm colours represent high similarity with the surface sample and cool colours represent low similarity with the surface sample. Surface similarity value is 1 because it represents similarity with itself.

As these three explanatory variables (day length, nutrients and MLD) were

correlated, it is hard to determine the specific influence of each variable on cytometric diversity patterns. Therefore, we explored the influence of these factors on the bacterioplankton composition across another spatio-temporal dimension, the vertical gradient. In particular, we compared the profiles of nutrient concentration and potential density along depth with the similarity in the mean cytometric diversity of the surface sample and each of the remaining sample depths for each month and station during the 10-year time-series. We found that the similarity in cytometric diversity always decreased with depth (Figure 2.5D-2.5E-2.5F).

Nevertheless, during the spring-summer months this decrease in similarity with depth was more pronounced than during late autumn - early winter. This seasonal pattern was found at the three stations but, interestingly, the decay in similarity with depth was sharper at E1 than at E2 or E3 (Figure 2.5D-2.5E-2.5F). During the summer months at E1 the similarity between surface and 20 m is around 0.6 while at E2 and E3, the similarity between surface and 20 m (see Figure 2.5-D and white line in Figure 2.5E-2.5F) was higher, ca. 0.8. In summary, at E1 during summer there is a sharp change in the cytometric similarity which occurs at around 5 m while this drastic change occurs at around 20 meters at E2 and E3.

Vertical nutrient profiles varied between stations and seasonally, and the shallowing of the nitracline was coincident in time with the decrease in cytometric similarity between depth and surface samples. Yet, the nitracline was deeper than the depth where the change in bacterial cytometric similarity occurred. The nutrient profiles (NO_2-NO_3) revealed a sharp gradient at ca. 80 and 120 m at stations E2 and E3 respectively (Figure 2.5J-2.5K-2.5L), while the gradient in density was much shallower (Figure 2.5G-2.5H-2.5I). From April to September no vertical gradient in nutrients was found at E1.

In contrast to nutrients, the MLD was coincident both in time and vertically with changes in bacterial cytometric composition, suggesting this variable was likely the best to explain the cytometric similarity patterns observed at the stations (Figure 2.5A-2.5B-2.5C). During late fall and winter, MLD frequently reached the bottom while during spring-summer conditions it shoaled at all stations resembling the pattern in cytometric similarity. At E1, the MLD during the summer was at around

5-10 meters while at E2 and E3 it shoaled up to only 20 m. This fact was paralleled by a more pronounced decay in cytometric diversity similarity with depth at E1 than at E2 and E3 (Figure 2.5D-2.5E-2.5F). The agreement between the dynamics of the MLD and cytometric diversity was also supported by the potential density profiles (Figure 2.5G-2.5H-2.5I).

We used the Bray-Curtis similarity metric for all the analyses explained above. Thus, we compared similarity in terms of bacterial community composition based on cytometric categories and their relative abundance (i.e. evenness). In order to clarify whether the spatio-temporal variation found is mainly caused by changes in the cytometric categories present or by their evenness, we re-estimated the similarity matrix by using a metric based only on presence/absence data (Jaccard's index). Hence, we replaced the abundance value in each cytometric category by 1 (when there were cells present) or 0 (when there were no cells in that category). The tendency found using Jaccard's index was similar to the pattern found by using the Bray-Curtis index (Figure S2.4), suggesting that the pattern is due to a change in the presence/absence of particular cytometric categories rather than in their relative abundance. In summary, although we found a high correlation between surface bacterial community composition and day length, nutrients and MLD, the analysis of vertical patterns indicate that the dynamics of the MLD is the factor (of those tested) most likely affecting the composition of bacterioplankton communities.

2.4 Discussion

Ribosomal phylogenetic markers are the *de facto* standard to analyze compositional and diversity changes in bacterial assemblages (Giovannoni and Rappe, 2000). However, high-throughput molecular approaches are still expensive and time consuming, posing limitations of spatial and temporal patterns resolution in oceanic microbial communities. A fast and easy processing of high numbers of microbial samples can be achieved by flow cytometry, a method that usually records bio-optical information of thousands of cells per minute. While the single-cell properties of cells obtained by flow cytometry have been used to study the diversity

of phytoplankton communities (Li (1997), Li (2002)), its potential for the smallest-sized microorganisms (bacteria and archaea) remains unexplored and, in general, the power of this method to detect ecological patterns in microbial diversity is unclear. Schiaffino et al. (2013) estimated cytometric diversity of freshwater picoplankton by analysing cytometric populations (clusters) which shared similar properties, instead of the properties of each individual cell, as Li (1997) proposed. In this study we have adapted the methodlology proposed by Li (1997) to study bacterioplankton using 3D to separate the cells falling on each cytometric category instead the 2D originally proposed by Li (1997) for phytoplankton communities. The advantage of incorporate this third parameter is that we have a higher spatial vision to separate the cells falling in each cytometric category, increasing thus, the accuracy of the methodology (see Methods).

Our results show a significant agreement in the temporal pattern of diversity of bacterial communities using their cytometric properties and a molecular standard approach, opening the door to a promising application of the routine flow cytometry method in aquatic systems. At the compositional level, important changes in the cytometric properties of bacterial communities were also found seasonally and vertically, as commonly found by molecular analyses (Fuhrman et al., 2006; Gilbert et al., 2012; Giovannoni and Vergin, 2012). The significant correlation in bacterial community composition based on cytometric and genetic criteria (Mantel test, $r = 0.59$, $p < 0.001$) confirms that, to a certain extent, changes in the cytometric properties of cells in a community reflect changes in microbial community composition.

It should be acknowledged, though, that the resolution of both approaches is clearly different. Indeed, the cytometric diversity was not a strong predictor of individual estimates of molecular diversity (Figure 2.3A). Different taxa probably fall in the same cytometric category and the same taxa may occupy different categories due to variations in morphology and cell-size over their growth cycle. The use of regularly-spaced divisions provided by the flow cytometer detection channels is also likely to split cells belonging to the same taxa into multiple bins and/or to create multi-taxa bins. The evidence presented is thus that the cytometric method can pick out large signals in community variability at coarse scales (both

over time and across depths, Figure 2.5). Thus, the utility of the method is liable to be restricted to detecting large signals over spatial and/or temporal scales, but this limitation is offset by its potential to provide data on changes in diversity at a low cost and effort and with a large spatio-temporal coverage, unfeasible by molecular techniques.

Cytometric diversity estimates are based on bio-optical properties of each individual cell. In the case of heterotrophic bacteria, these features are related mainly to their morphology (such as cell complexity and size, SSC) or nucleic acid content based on the green fluorescence emission (FL1) of the stained cells. These flow cytometry variables have been widely used to separate microbial communities into two main groups, HNA and LNA according to their nucleic acid content and SSC (Bouvier et al., 2007; Calvo-Díaz and Morán, 2006; Gasol and Del Giorgio, 2000). From the initial explanations of differential activity (Bouvier et al., 2007; Gasol et al., 1999; Li et al., 1995), it has been demonstrated that different cytometric groups of bacterioplankton have a distinct genetic composition. Indeed, several studies using flow cytometry cell sorting and subsequent molecular analysis of the sorted cells by CARD-FISH or pyrosequencing have revealed a distinct taxomomic composition of HNA and LNA cells (Mary et al., 2006; Vila-Costa et al., 2012; Zubkov et al., 2001) with little taxonomic overlap between them (Schattenhofer et al., 2011). For example, the group SAR11 has been associated with LNA cells (Mary et al., 2006; Schattenhofer et al., 2011). In agreement with these results, we found a drastic decrease in cytometric diversity in a sample collected in May 2012, when almost no LNA cells were recorded in the samples (Figure 2.2-A) and SAR11 populations were depleted (Alonso-Sáez et al., 2015). This data point with the lowest cytometric diversity increased the correlation found with the Shannon index between both methods (Figure 2.4) and also in the Mantel test results (Figura S1). Nonetheless, if we remove this value, the relationship between both approaches remains significant (Mantel test, $r = 0.44$, $p = 0.001$).

In agreement with previous multiyear time-series of marine bacterioplankton based on molecular data (Caporaso et al., 2012; Fuhrman et al., 2006; Gilbert et al., 2012; Smyth et al., 2014), we also found a pattern of increasing diversity during winter at coastal stations. Day length and, recently, net heat flux have been shown

to be the factors best correlated with patterns of bacterial diversity in the English Channel L4 station, but the mechanistic reason for this correlation remains unclear (Gilbert et al., 2010; Smyth et al., 2014). In our dataset, three parameters, i.e., day length, nitrogen concentration and the mixed layer depth, were significantly correlated with cytometric diversity in the samples of the three coastal stations with similar strength.

Taking advantage of the high spatio-temporal coverage provided by the flow cytometry sampling, we analysed samples along the vertical profiles to help elucidating which factor was responsible for the winter increase in diversity of surface coastal bacterial communities.

Important changes in cytometric diversity were found along the depth profile in most samples, and changes in day length (or associated incoming irradiance at the surface) could not explain compositional changes with depth. We found that other variables such as inorganic nutrients concentrations and, remarkably, MLD had a similarly high explanatory power of bacterial diversity as day length. Furthermore, the vertical profiles of potential density and nitrogen concentrations were different at each station.

A higher similarity between the potential density and cytometric diversity similarity profiles was found, as compared with nutrients profiles (Figure 2.5). In particular, the differentiation in bacterial cytometric composition (Figure 2.5D-2.5E-2.5F) followed changes in the MLD, matching the water-mass differentiation along the water column during stratification (Figure 2.5G-2.5H-2.5I). Conversely, nutrient gradient differed markedly between the three stations. At the coastal station E1 nutrient concentrations were vertically homogeneous, while at the more oceanic stations the nutrient gradient was deeper (Figure 2.5J-2.5K-2.5L) than the depth where cytometric diversity changes occur. In summary, the nitrocline was deeper than the pycnocline, and it was the pycnocline which matched with the cytometric similarity gradients, suggesting that mixing has a more important role than nutrients on bacterioplankton community composition.

The changes in bacterial cytometric composition occurring in surface waters when the stratification breaks were coincident with increases in diversity at all

stations (Figure 2.5). Therefore, we hypothesize that the higher values of bacterial diversity in early winter were due to a vertical reshuffling of bacterial taxa during the mixing period.The effect of mixing in patterns of bacterial diversity have been already highlighted in recent studies (Fuhrman et al., 2015; Vergin et al., 2013). During stratification, taxa adapted to thrive in oligotrophic, warmer waters outcompete other taxa from surface communities resulting in an overall decrease in diversity. In winter months, the MLD is deeper and, therefore, the whole water column is well-mixed and replenished of nutrients causing an increase in the number of bacterial taxa found at the surface. These bacterial taxa rising to the surface may grow under similar conditions than at depth. The finding of rare taxa that are typically abundant in the deep ocean such as SAR324 and SAR406, in surface samples from E2 during winter (Alonso-Sáez et al., 2015) also supports our hypothesis. However, based on our data, we cannot rule out the possibility that the increase in diversity was due to taxa that were originally in surface waters but under detection, and grew favored by other drivers associated with water column mixing not measured in this study (e.g. micronutrients or vitamins).

Changes in community composition following mixing by overturning have been described in the Bermuda Atlantic Time-Series Study (BATS) (Morris et al., 2005). Salter et al. (2015) have recently reported seasonal increases in diversity of SAR11 communities strongly associated with mixing using a 7-year time series in NW Mediterranean. However, a mechanistic explanation for the increase of bacterial diversity has been hindered by the lack of high vertical resolution in bacterial composition. The use of flow cytometry to complement molecular techniques has allowed us to increase the spatio-temporal resolution of the time-series analysed and to propose the species reshuffling by vertical mixing hypothesis to explain the seasonal pattern in bacterioplankton diversity and the late autumn-early winter maxima in bacterial diversity in coastal systems.

In summary, as in the seminal application of Li (1997) to small phytoplankton, our study reveals the utility and the potential of the cytometric diversity to assess heterotrophic bacterioplankton community structure and dynamics in aquatic systems, especially for large-scale studies where the application of molecular techniques is more limited. We took advantage of the high throughput offered by

flow cytometry to analyse changes in diversity over spatial, vertical and temporal scales and found that the reshuffling of species by vertical mixing may be an important factor for the recurrent early winter maxima in bacterial diversity. The use of flow cytometry for studying the diversity of different components of microbial communities can be a useful complementary tool to molecular approaches to provide new insights in the understanding of global microbial community spatio-temporal distribution.

2.5 Supplementary figures

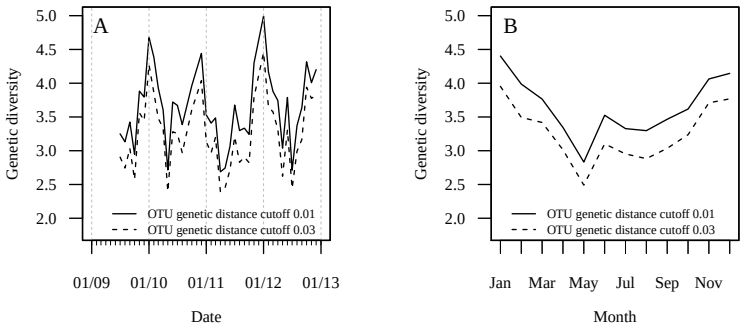

Figura S2.1: Trends of phylogenetic diversity (Shannon-Weaver index) of surface bacterial communities at E2 during a 3.5-year monthly time series (A) and as a mean annual cycle (B) when two different percentages of similarity are used: 99 % (solid line) and 97 % (dashed line).

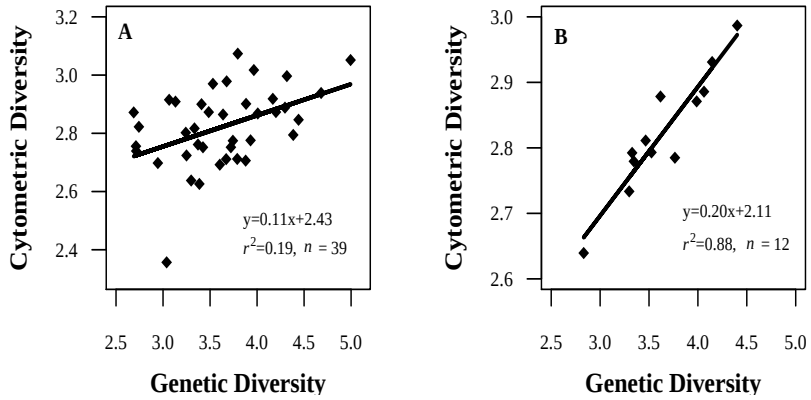

Figura S2.2: Correlation analysis between genetic (OTU 99% similarity) and cytometric diversity (Shannon-Weaver index) using the 3.5-year monthly sampled database and mean annual cycle.

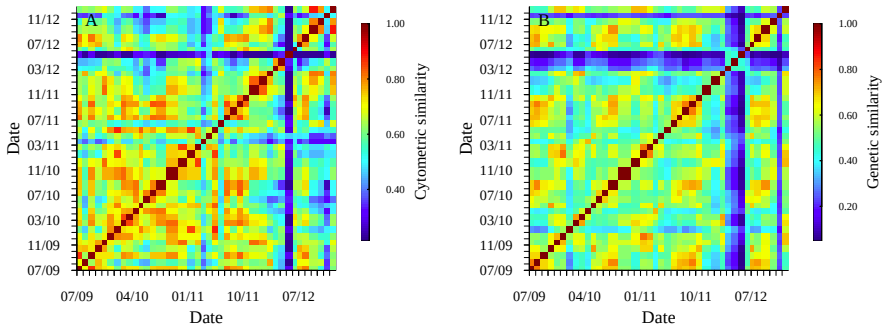

Figura S2.3: Bacterioplankton similarity matrix estimated using Bray-Curtis algorithm with 3.5-year monthly surface samples analyzed by flow-cytometry (left panel) and by pyrosequencing techniques (right panel).

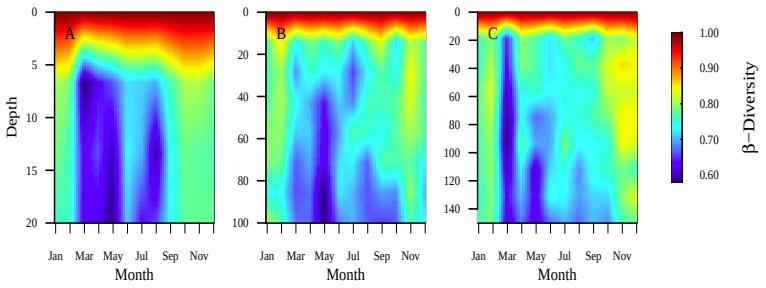

Figura S2.4: Community similarity data estimated from a 10-year monthly time series database collected in the Cantabrian Sea . Community similarity was estimated as the Jaccard similarity between surface data and the different depths accross the water column. Warm colours represent high similarity with the surface sample and cool colours represent low similarity with the surface sample. Surface similarity value is 1 because represent similarity with itself.

Chapter 3

The allometry of the smallest: Superlinear scaling of microbial metabolic rates in the Atlantic Ocean

3.1 Introduction

Heterotrophic bacterioplankton, one of the simplest organisms in the ocean, are responsible for a large fraction of the respiration in marine ecosystems, controlling the carbon cycle mainly in the vast oligotrophic areas (del Giorgio and Duarte, 2002; Robinson and Williams, 2005; Williams, 1981). Marine bacteria and archaea form an extremely diverse and specialised assemblage (Giovannoni and Stingl, 2005). They hold high capability to acquire genetic diversity mainly by lateral gene transfer (LGT) and other mechanisms for gene expansion favouring their functional diversity and, therefore, a widespread diversification of metabolic pathways (Doolittle and Papke, 2006; Whitman et al., 1998). This implies a competitive advantage allowing them to be adapted to the surrounding conditions (Giovannoni and Stingl, 2005; Massana and Logares, 2013) which largely explains the changes in bacterial composition recorded along different spatio-temporal gradients (Gilbert et al., 2009; Giovannoni et al., 1996; Morris et al., 2005; Pommier et al., 2007). Therefore, bacterial taxonomic and functional versatility have been the main variables used to explain changes in bacterioplankton metabolism.

Size and functional diversity are considered key factors controlling species metabolism (Edwards et al., 2012; Finkel et al., 2010; Litchman et al., 2007; Marañón et al., 2013). Although much progress has been made in allometric studies of other small organisms such as phytoplankton, the effects of cell size on marine heterotrophic bacterioplankton respiration has been hardly considered and continue essentially unanswered (Makarieva et al., 2005; Massana and Logares, 2013). The effects of cell size on metabolism are usually characterized using a power law (Brown et al., 2004; DeLong et al., 2010; West et al., 1997),

$$Y = Y_0 M^b \tag{3.1}$$

where Y is the metabolic rate, Y_0 is a normalised constant, M is the body mass of an individual and b is the scaling exponent.

The debate on the exact value of the allometric exponent b is extensive (Brown et al., 2004; Enquist et al., 2003; Hemmingsen, 1960; Huete-Ortega et al., 2012;

Kleiber, 1932; West et al., 1997). Recent empirical studies have suggested values for the scaling exponent in small-size organisms significantly different from the traditional $\frac{3}{4}$ scaling suggesting the need to further review the value of this exponent (Dodds et al., 2001; Makarieva et al., 2005; Zubkov, 2014). In this sense, DeLong et al. (2010) have found that the scaling of metabolism is not universal for all forms of life and is sublinear (b<1) for metazoan, linear (b=1) for protists and superlinear (b>1) for prokaryotes. Following DeLong et al. (2010) hypothesis, the superlinear scaling for prokaryotes is the result of the correlation between genome size and cell size within this functional group. Larger cells contain larger genomes which allows for an increasing diversity of substrates that can be metabolized leading to higher metabolic power.

The allometric studies of prokaryotes have generally used data compilations mostly from laboratory cultures (DeLong et al., 2010; Makarieva et al., 2005; Marañón et al., 2013). Most of the information on the allometry of microorganisms in the ocean is for phytoplankton groups (Finkel et al., 2010; López-Urrutia et al., 2006; Marañón et al., 2007; Zubkov, 2014), while the allometric scaling of natural bacterial communities remains mostly unexplored. Therefore, our main objective is to test whether the metabolic rates of prokaryotic picoplanktonic groups in nature follow the superlinear allometric scaling. We carried out *in vivo* 2-para (iodo- phenyl)-3(nitrophenyl)-5(phenyl)tetrazolium chloride reduction capacity (ivINT) measurements of size-fractionated natural marine microbial planktonic communities. Our aim is to determine how individual bacterioplankton respiration rate scales with cell size and to identify the biotic and abiotic factors that could affect this relationship.

3.2 Material and Methods

3.2.1 Sampling strategy

Seawater from 30 stations along a latitudinal transect in the Atlantic Ocean was taken during the 2011 Atlantic Meridional cruise (AMT21) on board the RRS Discovery. The latitudinal transect covered from 50°N (Avonmouth, United

Kingdom) to 50°S (Punta Arenas, Chile). One station was sampled daily between the 29th of September and the 14th of November of 2011. Seawater samples were collected in Niskin bottles at predawn, from one or two different depths: near surface (2-5 m depth) and the depth of the chlorophyll maimum (DCM). Seawater was transferred from the Niskin bottle to a 3 L bottle for subsequent subsampling and analysis of size-calibration experiments, in vivo INT reduction and flow cytometric determination of the size structure, as outlined below. All bottles were previously rinsed with fresh water, HCl and rinsed twice with milliQ water.

3.2.2 In vivo INT reduction capacity method

In vivo INT reduction capacity measurements were performed using 2-para (iodo- phenyl)-3(nitrophenyl)-5(phenyl)tetrazolium chloride salt (INT) as electron acceptor. This method has the advantage of estimating the respiration of the natural population allowing the filtration of the community into size classes after the incubation, which is not possible to perform by the dark bottle incubation (Winkler) method. Despite the criticism regarding the validity of the method (Maldonado et al., 2012), there is a good correlation between the INT reduced by planktonic organism and their respiration measured by dark bottle incubation (Winkler method) (Martínez-García et al., 2009) therefore the INT reduction capacity method could be considered as a proxy of respiration rate. However, the ratio between the oxygen consumption and the INT reduction is variable (García-Martín et al, personal communication). Hence, reduced INT moles have not been converted to O_2 units.

Four replicates of 500 ml were collected in plastic bottles. One replicate was immediately fixed by adding formaldehyde (2% w/v final concentration) and used as a killed control. After 15-20 minutes, all replicates were inoculated with a sterile solution of 8 mM iodonitrotetrazolium salt (INT) to give a final concentration of 0.8 mM. All replicates were incubated between 2 and 4 hours. Samples were fixed by adding formaldehyde, as for the killed control. Samples were sequentially filtered after 15 minutes through 0.8, 0.6, 0.4 and 0.2 μm pore size polycarbonate filters,

air-dried, and stored frozen in 1.5 ml cryovials at -20 °C until further processing. The 0.8 μm filter was used to remove the larger planktonic cells from the analysis. So then, respiration was estimated for three different size classes: 0.8-0.6, 0.6-0.4, 0.4-0.2 μm. The reduction of the INT was determined following (Martínez-García et al., 2009).

3.2.3 Size-structure

A water sample of 500 ml was collected for the flow cytometric determination of the size structure and filtered without any reactive through the same sequence of 0.8, 0.6, 0.4, 0.2 μm pore size Nuclepore filters. A 500 μL aliquot from each filtrate was fixed with paraformaldehyde 2% final concentration and used for flow cytometry analyses of heterotrophic bacteria. The samples were stained with SYBR Green I (1:10000 final dilution of initial stock) and potassium citrate (0.1% w/v) and kept 1 hour in the dark before analysis. Then, to calibrate the flow cytometric signal, 25 μL of a mixture of yellow-green 0.5, 1.0 μm beads (Fluoresbrite Microparticles, Polysciences) in pre-filtered seawater was added to each sample. Samples were analysed for 60 seconds at a low flow rate (15 μlmin^{-1} approximately) with a FACSCalibur flow cytometer.

3.2.4 Bacterioplankton abundance, respiration per cell and cytometric properties estimates

Flow cytometry analyses during the size-fractionated respiration experiments were used to estimate abundance and bio-volume of the community. The ivINT method estimates the respiration of the plankton fraction retained between the two consecutive filters (0.8-0.6, 0.6-0.4, 0.4-0.2, respectively). However, flow cytometry measurements were performed on the filtrates of each filter. The cell abundance retained by one filter was calculated by subtracting the number of cells in the filtrate from the number of cells in the suspension before passing through that filter.

To estimate the mean size of the cells retained on each filter we used a flow

cytometer parameter related to cell size, the side scatter (SSC) (Burkill et al., 1993; Calvo-Díaz and Morán, 2006; Zubkov et al., 1998). SSC units were transformed into size units by using a bacterioplankton size-calibration performed during the AMT21 cruise (See Supplementary information). The flow cytometer returns, for each sample, a FCS file with the recorded flow cytometric signals for each cell. Each flow cytometry standard (FCS) file was processed using the Bioconductor package flowCore (Hahne et al., 2009), hence, we were able to access to the cell size-distribution of the community. To estimate the size distribution of the cells retained on each filter we subtracted the size distribution of the cells in the filtrate from the size distribution in the suspension before passing through that filter. Knowledge on the abundance and mean size of the cells retained on each filter allowed us to estimate the average respiration per cell and the size scaling of respiration. The respiration per cell was calculated as the respiration of the different fractions divided by the cell abundance at that fraction.

3.2.5 Statistical analysis

The theoretical relationship between bacterioplankton respiration per cell and cell bio-volume (Eq. 3.1) can be linearized by logarithmic transformation. The equation has two unknowns; the normalizing constant α and the scaling exponent β, that were estimated from in situ measurements using a mixed effects model (Gelman and Hill, 2007) due to the hierarchical structuring of our experimental design. This allowed us to account for unbalanced observations and to quantify variability associated to changes between replicates within the same experiment, and among different experiments along the transect;

$$\ln R_i = \alpha_{jk[i]} + \beta_{jk[i]} \ln V_i + \varepsilon_i \tag{3.2}$$

where R and V represent the respiration and the bio-volume measurements. The subscript $jk[i]$ indicate that respiration was measured for the replicate k within experiment j. Intercepts and slopes were modelled as normal random deviates with mean μ_α and μ_β, respectively. Deviations from this mean vector for each replicate were characterized by a variance-covariance matrix partitioning total

variance among changes between replicates and between experiments (*i.e.* σ_k^2 and σ_j^2). Remaining errors were assumed independent and normally distributed ($\varepsilon \sim N(0, \sigma_{\ln R_i})$. The model was fitted using restricted maximum likelihood (Team, 2014) with the package lme4 (Bates et al., 2014). Data were centered before the analysis to avoid spurious inflation of parameter correlations. Model selection based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) discarded simplified versions of the model presented above (*e.g.* constant intercepts or slopes) and models considering differences between replicates within experiments (*i.e.* the best model assumed $\sigma_k = 0$ for both α and β, highlighting the robustness of our measurements). We assumed that the magnitude of errors in respiration per cell was much larger than the magnitude of errors in cell size measurements, and thus that the later can be safely ignored. Although this assumption is common in allometric studies (Gillooly et al., 2002; Makarieva et al., 2005), it might lead to the underestimation of scaling exponents (DeLong et al., 2010; Warton et al., 2006). Therefore, estimates performed using ordinary least squares (OLS) method were compared with estimates derived from reduced major axis (RMA) regression in order to explore the effect of the fitting method in the superlinearity found, the scaling exponent and intercept for each station (Supplementary information). A comparison with estimates derived using reduced major axis regression for each experiment resulted in around a 20% increase of scaling exponents, so we consider our approach conservative to reject the hypothesis of linear or sublinear scaling.

3.3 Results

The size-fractionation experiments along the 30 stations sampled revealed a strong correlation between cell size and metabolic activity along a latitudinal transect across the Atlantic Ocean (Figure 3.1). We parameterized a mixed-effect model to analyse the relationship between respiration per cell and cell size for bacterioplankton groups along the Atlantic Ocean (See Methods). The best model was a lineal mixed effects model with random intercept and slope (Eq. 3.2), Akaike Information Criterion (AIC) = 549.98, Bayesian Information Criterion (BIC) =

571.23, p < 0.0001). This implies significant differences in both slope and intercept along the Atlantic latitudinal transect experiments.

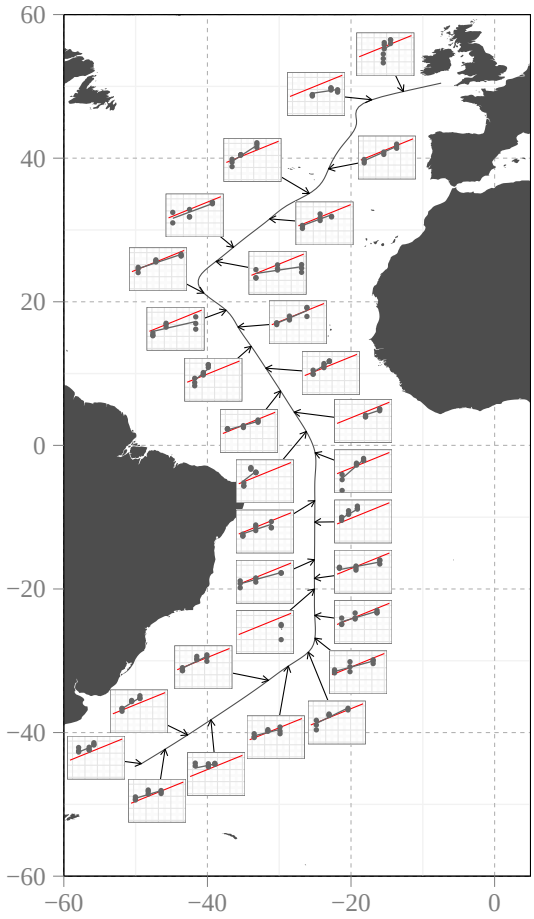

Figura 3.1: The relationship between respiration per cell (y-axis) and bio-volume (x-axis) for each experiment along the Atlantic latitudinal transect and the general fit provided by the model. Grey dots represent the values for each size-fraction and grey dashed line the linear fit provided by the mixed model for each experiment. The red line corresponds to the population level, mean scaling relationship for all experiments.

The overall slope in our model suggests that respiration rate scales with bio-volume superlinearly, with a scaling exponent significantly larger than one (Figure 3.2, $\beta_1 = 1.67 \pm 0.19$, $p < 0.0001$). For 80% of the experiments performed along the latitudinal transect the slope was significantly higher than the unity.

Despite the model concluded that there are significant differences between stations, the comparison of the fitted slope for each experiment with the general slope (Figure 3.1) revealed that the superlinearly detected for the overall slope was also followed by the majority of the experiments. The comparison between the scaling exponents obtained by ordinary least squares (OLS) and reduced major axis (RMA) (Supplementary information, Table S3.1) further demonstrate the superlinear scaling recorded by the mixed effect models used in our analyses.

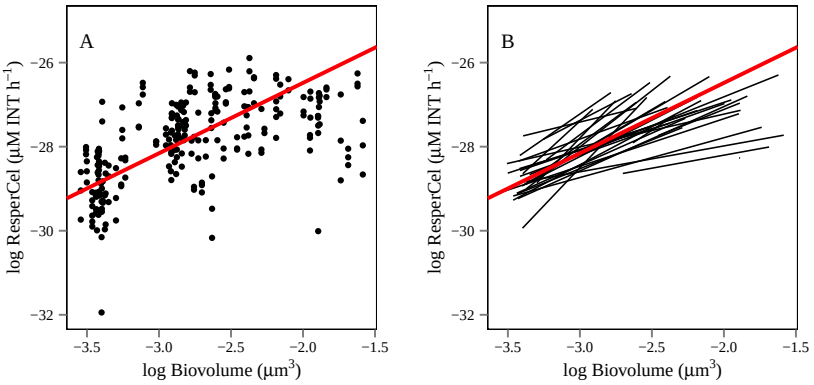

Figura 3.2: Relationship between respiration per cell (y-axis) and bio-volume (x-axis) for each experiment size-fraction (A) and the linear fit for each experiment (B). The red line in each panel represent the mean scaling relationship derived from the mixed effects model.

A Caterpillar plot for the slope residuals (Figure 3.3) revealed that for most of the experiments (> 75 %) the 95 % confidence intervals overlapped with the general slope. Nevertheless, some experiments deviate significantly from the overall trend (red lines in Figure 3.3). Exclusion of these outliers do not affect the non-universality in the scaling exponent along the transect.

In an effort to understand the differences found for the slope, we tested the relationship between the scaling exponents provided by the lineal mixed effects model and different external factors: latitude, temperature, depth and chlorophyll. However, none of the factors tested could be clearly associated to these deviations.

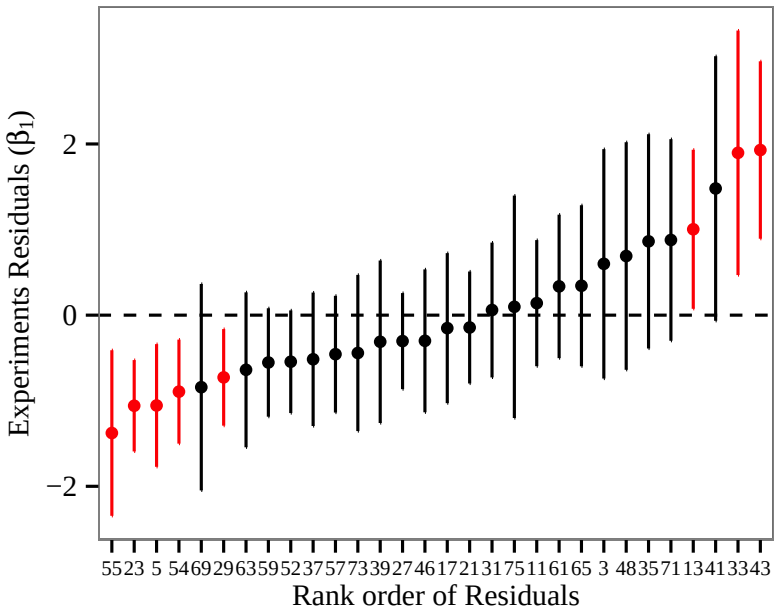

Figura 3.3: Caterpillar plot comparing the ranked experiment level residuals for the 30 size-fraccionated experiments along the transect with the 95 % confidence intervals for the scaling exponent parameter (β_1). Those experiments whose β_1 confidence intervals do not overlap with the mean scaling exponent provided by the model are represented in red. The red lines represents the stations where the confidence intervals of β_1 do not overlap the mean escaling exponent.

Those bacterioplankton communities composed of larger cells evidence a decrease in the scaling exponent (Figure 3.2 A, B). Indeed, there is a strong inverse correlation between the scaling exponent and the mean size of the cells that integrated the community in each experiment ($r = -0{,}57, p < 0{,}01$, Figure 3.4 A). This decrease in the scaling exponent when the community is composed of larger cells implies that changes in community composition may be causing the differences in the slope term. We explored this hypothesis through the use of the average bio-volume of the cells retained by each of the different pore-size filters in an experiment. While no remarkable changes were found in the mean bio-volume for the two smallest size fractions, there is a considerable range of average cell-

sizes of the largest size fraction along the experiments. A significant relationship was found between the slope term and the mean bio-volume of the cells retained on the 0.6 μm filter ($r = -0{,}68, p < 0{,}001$, Fig. 3.4 B). Nonetheless, no remarkable changes were found in the mean bio-volume for the other size fractions along the experiments. This suggests that changes in the larger-size populations of the microbial community may be responsible for the drop-off observed in the scaling exponent.

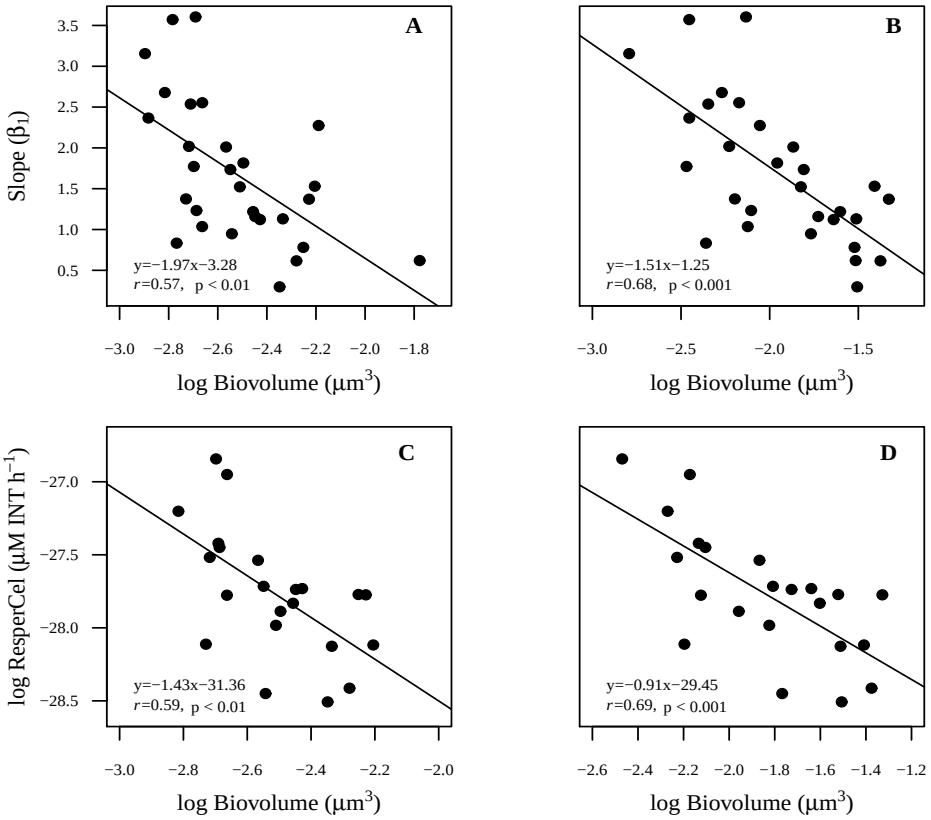

Figura 3.4: Relationship between the β_1 (A, B, y-axis) and size-normalized respiration per cell (C, D, y-axis) obtained for each single experiment (A,B) and for experiments performed at the DCM (C, D), respectively, with the mean bio-volume of the community (A, C, x-axis) and the mean bio-volume of the cells retained on the 0.6 μm filter (B, D, x-axis). All parameters are in log scale except β_1 .

In addition to a change in slope, Figure 3.2 also denoted a decrease in the height

of the scaling relationship with increasing mean bio-volume of the community. Because the slopes differ, comparison of the intercepts provided by the model is meaningless. To study these changes in the height of the scaling relationship we selected an intermediate cell volume (0.068 μm^3) and we calculated a size-normalized respiration rate per cell at this fixed size for all the experiments. To minimise the variability, only experiments performed at the DCM were taken into account. Similarly to the scaling exponent, the size-normalized respiration was highly correlated with average bio-volume, especially for 0.6 μm filter size, indicating that communities composed of larger, more complex cells have lower metabolic rates per cell for a given cell size (Figure 3.4 C, D). These effects of community composition are also evidenced by a decreasing size-normalized metabolism with an increasing percentage of *Prochlorococcus* abundance in relation to heterotrophic bacteria ($r = -0{,}60, p < 0{,}01$) and a strong, positive relation with heterotrophic bacteria abundance ($r = 0{,}71, p < 0{,}001$).

These compositional changes were paralleled by changes in several ancillary environmental parameters. When the DCM is located deeper, size-normalized metabolism decreased ($r = 0{,}68, p < 0{,}001$). Although the DCM was deeper mainly in oligotrophic areas, a nutrient limitation index (calculated as the difference between the MLD (mixed layer depth) and nitracline depth (Behrenfeld et al., 2002)) showed no clear relationship with size-normalized respiration ($r = 0{,}35, p = 0{,}12$). This implies that albeit significant correlations were found, no environmental parameter could explain alone the differences found in the heights along the study.

3.4 Discussion

Despite the effect of cell size on prokaryotic picoplankton metabolism has been hardly considered to be important, our experiments with natural oceanic communities demonstrate a strong effect of cell bio-volume on metabolic processes for tiny organisms like bacteria. Our results reveal a superlinear scaling (> 1) for the smallest (< 0.8 μm) planktonic organisms along a latitudinal transect in the Atlantic Ocean. The model providing a best fit to our data estimated an overall, station-average scaling exponent value larger than one (b = 1.67) and in the range

of the value of 1.7-2.0 reported by DeLong et al. (2010) from a compilation of data for terrestrial and aquatic prokaryotes mostly from cultures. DeLong et al. (2010) hypothesized that the superlinear scaling is due to the correlation between cell size and genome length in prokaryotes (Doolittle, 2002; Shuter et al., 1983; Tanaka et al., 2003). Larger cells have longer genomes resulting in an increase in the number of metabolic pathways and, hence, in their functional diversity (Islas et al., 2004; Price et al., 2004; Stepkowski and Legocki, 2001; Whitman et al., 1998). Accordingly, the metabolism of prokaryotes would not be constrained by surface area as in protists but by the number of metabolic pathways that they can use to acquire energy. When the size of prokaryotes is large enough so that they have a complete complement of most metabolic pathways, surface area constraints their metabolism (DeLong et al., 2010)

Our model also concludes that this general scaling exponent is not universal with significant deviations along the transect (Figure 3.3). Our analyses also revealed that the heterogeneity around the universal scaling relationship may be explained in terms of changes in community composition and size distribution. Communities composed of smaller cells have higher scaling exponents than those composed of larger cells (Figure 3.4 A, B). Along the Atlantic transect there are compositional changes in the relative importance of the phylogenetic groups within prokaryotic picoplankton (bacteria, archaea and cyanobacteria) (Fuhrman et al., 1993; Giovannoni and Stingl, 2005) and in the cell sizes within each group. Although our experiments separated the community into three size fractions bellow 0.8 μm, these compositional changes result in shifts in the size range covered in our experiments and in the mean cell sizes of the whole community. We were able to detect these changes thanks to the measurement of the size frequency distributions within each size fraction using flow cytometry. If we use the average size of the community as an indicator of cell complexity, taking to account the relationship between cell size and the number of genes of a prokaryotic cell reported by DeLong et al. (2010), a similar pattern can be observed in our experiments with communities composed of large and thus, more complex cells having an scaling exponent closer to the unity (Figure 3.4 A, B).

We hypothesize that these changes are the result of shifts in the allometric scaling between functional groups within prokaryotes. Experiments where heterotrophic prokaryotic groups (bacteria and archaea) predominate have a metabolism which is not constrained by surface area and scale superlinearly as suggested by DeLong et al. (2010). There is a drop-off in the scaling exponent (Figure 3.4 A, B) when larger, more complex cyanobacteria, especially *Prochlorococcus*, dominate the largest size fractions of the community as in the Atlantic oligotrophic areas (Zubkov et al., 2000*a*, 1998). We suggest that these larger, more complex prokaryotes fall close to the evolutionary transition between prokaryotes and protists, in a range where surface area starts to constraint metabolism (DeLong et al., 2010) and, hence, are expected to follow a scaling closer to linearity.

The change in the size scaling slope was accompanied by a marked decreasing in the heights of the scaling relationship for communities dominated by larger, more complex cells (Figure 3.2). The size-normalized respiration at the DCM was inversely correlated with mean cell size of each community, especially for the largest size fraction (Figure 3.4 C, D). *Prochlorococcus* predominates at the DCM depth in both Atlantic oligotrophic gyres (Zubkov et al., 1998), a pattern that we also found in the 0.6 μm filter in our experiments. The resulting increasing ratio of *Prochlorococcus* abundance to heterotrophic bacteria abundance in oligotrophic areas is reflected in a decrease in the size-normalized respiration (Figure 3.2).

The relevance of cell size in the metabolism of phytoplankton groups has been fully recognized (Edwards et al., 2012; Finkel et al., 2010; Litchman et al., 2007; López-Urrutia et al., 2006; Zubkov, 2014). The importance of these groups on the productivity of the oceans together with their wide size range, spanning several orders of magnitude, has favoured that phytoplankton has been largely studied compared with other groups much smaller as bacteria, whose allometry has remained essentially unknown. Here, despite the narrow size range covered (0.2 - 0.8 μm), we have demonstrated that cell size is a key factor controlling the respiration of marine bacterioplankton. The strong compositional changes along the transect and its correlation with the size scaling parameters highlights the influence of cell complexity on the allometry of the microbial community. We have shown that cell size is a key functional trait in bacterioplankton communities.

3. The allometry of the smallest: Superlinear scaling of microbial metabolic rates in the Atlantic Ocean

Understanding the deviations from the general allometric scaling from knowledge of bacterial taxonomic and functional diversity might allow a better comprehension of their contribution to the biogeochemical cycles, especially the marine carbon cycle.

3.5 Supplementary information

Calibration of the side scatter flow-cytometric signal for size estimates

The size calibration experiments consisted on the filtration of the water sample through different pore-size filters and the analyses of the filtrate by flow-cytometry. The filtration system consisted of a 47 mm of diameter filter holder that allowed the collection of the filtered sample directly into a flow-cytometry falcon tube. Samples were filtered by gravity through 0.2, 0.4, 0.6, 0.8 and 1.0 μm Nuclepore filters. A 4 mL subsample from the unfiltered water was filtered through each filter. In addition, 4 mL of unfiltered water were taken as control. Filters were previously moistened with 0.1 μm filtered seawater. A 500 μL aliquot from each 4 mL filtrate was used for flow-cytometry analyses of heterotrophic bacterioplankton and fixed with paraformaldehyde 2% final concentration. Samples were stained and processed by flow cytometry as we have detailed in the experimental setup section.

In order to establish a relationship between SSC and cell size, we used the nominal pore size of each filter as an estimate of the cell size of particles that are retained with 50% efficiency. Thus, for each filtrate we attempt to find the SSC of the cells that are retained with 50% efficiency. For each population, we counted the number of cells falling in each SSC flow-cytometer channel, and compared the number of cells in the filtrate to the number of cells in the original samples. This yields, for each channel, the percentage of cells that pass through the filter. Hence, for each filter we obtained a calibration point as the SSC for which 50% of the particles were retained. Then, for each filter and population we obtain a curve that shows the decrease in the percentage of particles retained as SSC increases. These curves were smoothed using a lowess algorithm and the SSC of the bin at which 50% the original population was retained using these smoothed curves for each filter. The SSC values were corrected using the SSC of the 1.0 μm beads. The nominal filter pore size and the SSC data obtained above were related using a reduced major axis (RMA) regression to obtain the calibration equations (Figure S3.1).

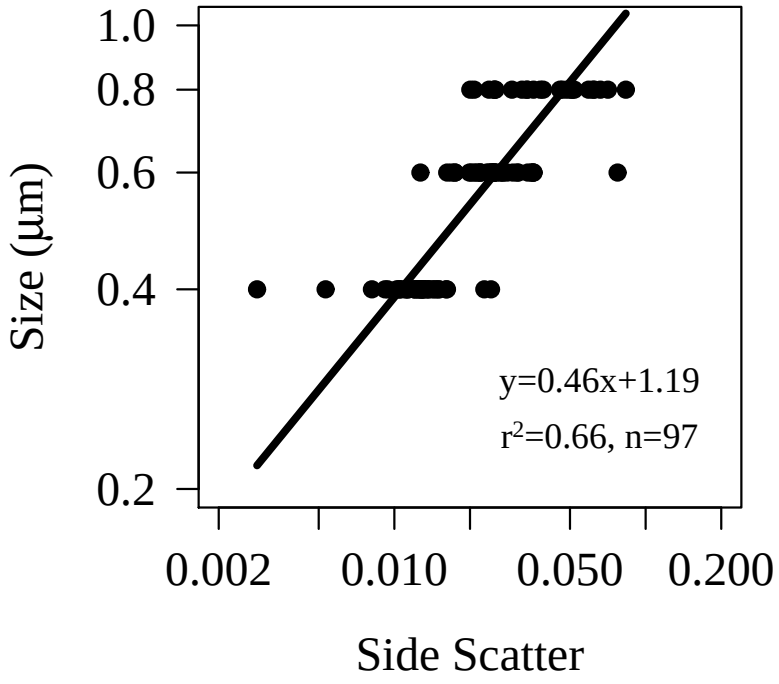

Figura S3.1: Calibration curve between the logarithm of nominal pore size (y-axis) and mean side scatter (x-axis) of bacterioplankton cells in the corresponding filtrate (0.8, 0.6, 0.4 μm.

Table S3.1. In-vivo INT reduction capacity experiments allometric scaling. Data shown are the parameter estimates α and β for each in-vivo INTreduction capacity experiment along the AMT21 fitting the data by using ordinary least squares (OLS) regression and reduced major axis (RMA) regression. Model fit were: ln(r) = α+β*ln(w) where r is the respiration per cell in μM INT h^{-1} and w is the biovolume in μm^3. r^2 and p are the square correlation coefficient and the significance level, respectively and n is the number of data in each regression.

	OLS Regression		RMA Regression				
Station	α	β	α	β	r^2	p	n
3	-9.32 (4.19)	7.31 (1.76)	-2.21 (4.92)	10.13 (1.97)	0.52	0.028	9
5	-26.54 (0.43)	0.86 (0.17)	-26.04 (0.44)	1.10 (0.17)	0.61	0.012	9
11	-22.84 (0.48)	1.89 (0.16)	-22.68 (0.47)	1.94 (0.16)	0.94	10^{-5}	8
13	-19.34 (0.93)	2.89 (0.33)	-18.95 (0.97)	3.03 (0.34)	0.91	10^{-5}	9
17	-23.85 (0.65)	1.54 (0.21)	-23.26 (0.91)	1.75 (0.28)	0.78	0.0037	8
21	-23.85 (0.75)	1.61 (0.31)	-23.16 (0.49)	1.88 (0.26)	0.74	0.0063	8
23	-26.79 (0.72)	0.60 (0.28)	-25.70 (0.58)	1.01 (0.18)	0.35	0.094	9
27	-23.99 (0.33)	1.35 (0.14)	-23.79 (0.40)	1.42 (0.16)	0.90	0.0001	9
29	-25.95 (1.04)	0.96 (0.34)	-24.86 (0.78)	1.36 (0.22)	0.49	0.035	9
31	-22.97 (1.10)	1.76 (0.34)	-22.41 (0.83)	1.96 (0.25)	0.81	0.001	9
33	-13.01 (1.81)	4.92 (0.62)	-12.23 (1.58)	5.18 (0.55)	0.90	10^{-5}	9
35	-19.19 (0.81)	2.92 (0.30)	-18.81 (0.67)	3.05 (0.25)	0.92	10^{-5}	9
37	-24.98 (0.28)	1.06 (0.09)	-24.74 (0.28)	1.10 (0.096)	0.91	0.0002	8
39	-23.23 (0.43)	1.92 (0.19)	-23.11 (0.39)	1.98 (0.18)	0.94	0.001	6
41	-13.50 (2.86)	4.59 (0.90)	-9.50 (3.05)	5.89 (0.91)	0.61	0.013	9
43	-15.78 (2.15)	4.31 (0.81)	-14.62 (2.37)	4.71 (0.88)	0.84	0.0005	9
46	-24.34 (0.93)	1.38 (0.29)	-23.43 (0.86)	1.69 (0.25)	0.67	0.007	9
48	-18.98 (0.88)	2.74 (0.29)	-18.49 (0.93)	2.90 (0.30)	0.89	0.0001	9
52	-25.17 (0.32)	1.11 (0.15)	-24.90 (0.36)	1.21 (0.16)	0.85	0.0004	9
54	-25.89 (0.42)	0.67 (0.14)	-25.45 (0.39)	0.83 (0.15)	0.65	0.0084	9
57	-24.67 (0.49)	1.17 (0.20)	-24.13 (0.52)	1.36 (0.19)	0.74	0.003	9
59	-24.91 (0.37)	1.05 (0.13)	-24.35 (0.47)	1.25 (0.16)	0.70	0.005	9
61	-21.86 (0.90)	2.07 (0.35)	-21.16 (1.15)	2.32 (0.41)	0.80	0.003	8
63	-25.42 (0.96)	0.87 (0.31)	-24.37 (0.67)	1.24 (0.21)	0.50	0.033	9

65	-21.78 (1.06)	2.10 (0.34)	-20.95 (0.89)	2.39 (0.27)	0.77	0.002	9
69	-26.89 (0.67)	0.10 (0.24)	-25.19 (0.43)	0.69 (0.17)	0.02	0.727	8
71	-19.40 (0.75)	2.78 (0.26)	-19.09 (0.77)	2.88 (0.27)	0.93	10^{-5}	9
73	-24.55 (0.65)	1.07 (0.22)	-23.71 (0.72)	1.36 (0.22)	0.62	0.012	9
75	-21.45 (0.79)	1.94 (0.29)	-23.31 (0.76)	1.34 (0.23)	0.55	0.022	9

Chapter 4

Bacterioplankton distribution in the euphotic layer: an attemp to predict the metabolic balance of the oceans

4.1 Introduction

Marine planktonic communities constitute the base of the marine food chain and have a central role in the oceanic carbon cycle. Marine phytoplankton are responsible for most of the assimilation of CO_2 through photosynthesis while heterotrophic bacterioplankton respiration represents a substantial contribution to the overall production of CO_2 in the oceans, mainly in the least productive areas (del Giorgio and Duarte, 2002; Duarte and Agusti, 1998; Ducklow, 1995; Morel and Antoine, 2002; Robinson and Williams, 2005; Suess, 1980; Williams, 1981). Quantification of the photosynthesis and respiration terms is crucial for determining the efficiency of the microbial food web, and therefore the metabolic balance of the oceans.

Significantly more progress has been reached both in the sampling and modelling of phytoplankton carbon assimilation compared to our still very limited knowledge of global community respiration. This fact is exacerbated by the development of primary production models based on remotedly sensed data. Remote sensing provides, routinely, a map of near-surface satellite estimates of chlorophyll biomass (Behrenfeld and Falkowski, 1997; Eppley et al., 1985; Smith et al., 1982). The large spatio-temporal coverage provided is an unvaluable source of information on the primary production dynamics in the ocean. Two major problems have been reported related to this tool. Firstly, the poor agreement found when satellite-derived chlorophyll measurements are compared with local in-situ measurements due to the need to incorporate light scattering in the applied predictive models (Behrenfeld et al., 2002; Campbell et al., 2002). Secondly, satellite-derived measurements have the inconvenient that they only cover the ocean surface while the primary producers are distributed along all the photic layer.

Current primary production models either use vertically generalized models (e.g. the carbon-based productivity model (CbPM) and Vertically Generalized Production Model (VGPM) or estimate the vertical distribution of phytoplankton biomass and productivity (Behrenfeld and Falkowski, 1997; Behrenfeld et al., 2006; Platt and Sathyendranah, 1995; Platt et al., 1988; Westberry et al., 2008). In this regard, several studies have characterised chlorophyll profiles by fitting the

vertical data to a Gaussian curve (Platt and Sathyendranah, 1995; Platt et al., 1988; Sathyendranath et al., 1995, 1991). Richardson et al. (2002, 2003) developed a methodology for predicting the shape of chlorophyll profiles from surface variables by applying an artificial neural network, the Self-Organizing maps (SOM). Depth-integrated primary production estimates on a global scale can be thus obtained thanks to the development of multiregression models linking primary production with satellite-derived chlorophyll (Behrenfeld et al., 2006; Richardson et al., 2003; Westberry et al., 2008). Far less progress has been made in the estimation of global heterotrophic respiration rate (del Giorgio and Cole, 1998; del Giorgio and Duarte, 2002; Gasol et al., 2008).

During the last decades, the increasing number of marine bacterioplankton samples, analysed mainly thanks to the introduction of flow cytometry, has improved our knowledge about the macroecological patterns of bacterioplankton communities (Li, 2009). It should be also emphasized the contribution of international open ocean sampling programs at large-scale. Most of them started at 90s and still continue today being the major source of bacterial data currently available. The most outstanding programs in aquatic microbial ecology are the global Joint Global Ocean Flux Study (JGOFS), the Atlantic Meridional Transect (AMT) program centred in the latitudinal study of the Atlantic ocean with a particular research interest in the dynamic of the oligotrophic gyres and the MAREDAT dataset with a remarkable spatio-temporal coverage of the Artic ocean (Aiken et al., 2000; Buitenhuis et al., 2012; Ducklow and Harris, 1993; Fasham et al., 1999; Hanson et al., 2000; Li, 2009; Robinson et al., 2006; Zubkov et al., 2000*b*).

These global databases have evidenced a significant relationship between bacterioplankton abundance and environmental variables, such as temperature and phytoplankton abundance (Gasol and Duarte, 2000; Li, 1998; Li and Harrison, 2001). It is widely accepted the coupling between phytoplankton as source of organic matter and bacterioplankton that incorporates those substrates for growth (Gasol and Duarte, 2000; Li et al., 2004; Moran et al., 2002; Nagata, 2000). However, this relationship is non-universal and may vary depending on the trophic conditions (Gasol and Duarte, 2000; Li et al., 2004). Li et al. (2004), using an

extensive database of bacterioplankton and chlorophyll biomass reported a strong coupling between bacterial abundance and chlorophyll concentration exhibiting a power relationship that is modified under different temperature and nutrients availability conditions. The decrease in the power slope showed under high nutrients conditions has been explained through an increase in the top-down control (Ducklow et al., 1999; Li et al., 2004). Albeit chlorophyll is not a robust proxy of organic matter, the dependence of bacteria abundance on the resources supply by phytoplankton has been widely demonstrated (Ducklow and Carlson, 1992; Gasol and Duarte, 2000; Moran et al., 2002).

Temperature has great influence on many metabolic processes and several studies have reported that bacterial production and respiration rate are highly affected by this parameter (Kirchman et al., 2005; López-Urrutia et al., 2006; Pomeroy and Wiebe, 2001). In this sense, López-Urrutia and Morán (2007) developed an equation derived from the Metabolic Theory of the Ecology (MTE) for estimating bacterioplankton cell-specific respiration rate (BR_i) using temperature as independent variable.

$$BR_i = 3{,}21 * 10^{11} e^{\frac{-0{,}589}{kT}} \tag{4.1}$$

where T is the absolute temperature and k is Boltzmann′s constant ($8{,}62 * 10^{-5} eV/K$).

Combining both bacterial abundance and cell-specific metabolism can be a powerful tool for predicting the heterotrophic respiration at global scale. In this study, we have compiled a global oceanographic database of heterotrophic bacterioplankton abundance and ancillary environmental information to improve our knowledge about bacterioplankton distribution in the ocean. We have characterized the coupling in the vertical distribution patterns of chlorophyll and bacterial abundance using unsupervised pattern recognition techniques (Richardson et al., 2002, 2003). We then use these relationships to model bacterioplankton depth-integrated global distribution by using satellite-derived variables. The final goal of our study is to attempt to estimate bacterial global heterotrophic respiration and to compare these

predictions with the current estimates of primary production obtained by a vertically generalized primary production model (Behrenfeld and Falkowski, 1997).

4.2 Material and Methods

4.2.1 Data compilation

Data were compiled from different oceanographic cruises including a total of 58779 bacterioplankton samples. Data include a wide range of regions from polar to equatorial latitudes and from coastal to oceanic waters, up to 5000 metres (Figure 4.1).

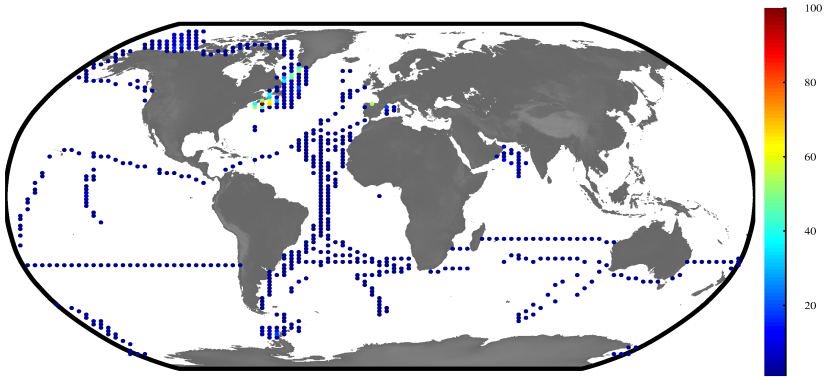

Figura 4.1: Geographic distribution of the sampling points compiled in this study. Cold colours represent areas less sampled and warm colours represent areas where more samples were collected.

Table S4.1 summarises the sources of data compiled for this study. The dataset includes data from public repositories as JGOFS and MAREDAT (85%) and primary sources. We only considered cruises with measures of bacterial abundance, temperature and chlorophyll, but we also included other biological and

environmental variables when available (primary production, bacterial production, biomass, salinity, PAR, nutrients). We considered bacteria and archaea as bacteria group along the study. Most of the bacterial abundance measurements were analysed by flow cytometry. However when microscope data were available we also included them.

4.2.2 Satellite-derived data

We extracted chlorophyll concentration (Chl), sea surface temperature (SST), particulate organic carbon concentration (POC), daily mean Photosynthetically Available Radiation (PAR), diffuse attenuation coefficient at 490 nm (Kd490) variables from satellite data for each location of the dataset. Chl, PAR, POC and Kd490 data were extracted from SeaWIFS L3 with 9 km resolution (1 pixel = 9 km, http:// oceandata. sci. gsfc. nasa. gov/ SeaWiFS/ Mapped) and SST data were extracted from the AVHRR sensor. We downloaded daily SST datasets with 4 km (1 pixel= 4 km) resolution from the NASA site (ftp:// podaac-ftp. jpl. nasa. gov/ allData/ avhrr/ L3/ pathfinder$_ $v5/ daily/ night/ 04km/). The compiled dataset ranged from April 1989 to November 2011, so for data after 2009, satellite data were extracted from the MODIS Aqua L3 with a resolution of 9 km (1 pixel = 9 km, http:// oceandata. sci. gsfc. nasa. gov/ MODISA/ Mapped/).

SeaWiFS Data Analysis System program (SeaDAS) was used to extract the data. We first extracted satellite data from the same day and location of each field data collected. When no satellite data was available we attempted to obtain data by considering also surrounding pixels (up to a maximum distance of 20 pixels). When we obtained a missing value for this point we used an increasing number of search radius starting at radius of pixel 1 to pixel 20. We used the lower pixel radius with available data. When satellite data from the daily image was not available we used a monthly image or, ultimately, a monthly climatology image. A quality flag (QF) was used for controlling the satellite-derived data. A value of 0 was used for data extracted from the same day and pixel and an increasing value was used as precision decreased using the highest value for monthly climatology with a pixel area of 20x20 pixels.

Climatological images were used for making global predictions of bacterial abundance and respiration. POC, Chl and Kd490 were extracted from monthly average 2002-2012 climatology (Modis Aqua 0.083° x 0.083°). The water column depth was estimated using the GEBCO global database. The depth of the euphotic layer (Z_{eu}) was calculated with the Kd490 images ($ln(0{,}01)/Kd490$).

We compiled also temperature data at different depths from the annual climatology from the Word Ocean Atlas 13 (WOA13) database. Data were then stored in maps with a resolution of one degree at the 29 depths levels available from 0 to 300 metres in the Ocean Atlas. Primary production data were extracted from the Ocean Productivity site (`http:// orca. science. oregonstate. edu/ data/ 1x2/ monthly/ vgpm. r2013. m. chl. m. sst4/ hdf/`). The primary production was estimated with the Vertically Generalized Production Model (VGPM) algorithm (Behrenfeld and Falkowski, 1997) and stored in images with a resolution of 1080x2160.

4.2.3 Parameterization of vertical profiles

A quantitative study was performed to characterize the vertical profiles of chlorophyll and bacterial abundance using the shifted Gaussian model described by Platt et al. (1988); Sathyendranath et al. (1995). The Gaussian equation has four parameters for describing the shape of the profile. B_o represents the background chlorophyll concentration (mg m^{-3}), h is the total chlorophyll concentrations beneath the curve (mg m^{-3}), σ is the width of the peak (mg m^{-3}) and z_m represents the depth of the chlorophyll maximum (m).

$$B_{(z)} = B_0 + \frac{h}{\sigma\sqrt{2\pi}} e^{-\frac{(z-z_m)^2}{2\sigma^2}} \tag{4.2}$$

$B_{(z)}$ is the chlorophyll concentration as a function of depth (mg m^{-3}). When we characterized bacterioplankton abundance profiles, $B_{(z)}$ and h where the background bacterioplankton abundance and the abundance beneath the curve respectively (cell m^{-3}). This model was fitted applying a maximum likelihood estimation (MLE) using the bbmle R package (Bolker, 2012). The quasi-Newton

algorithm was used for fitting the model. We constrained the parameters B_0, h and σ to be positive as recommended Platt and Sathyendranah (1995). Figure 4.2 shows an example of a chlorophyll profile. We can see how the profile shape changes depending on the value of each parameter. These parameters thus provide a valuable information about how a variable is distributed with depth and, therefore allow us to study the agreement between both variables (bacterial counts and chlorophyll biomass) with depth.

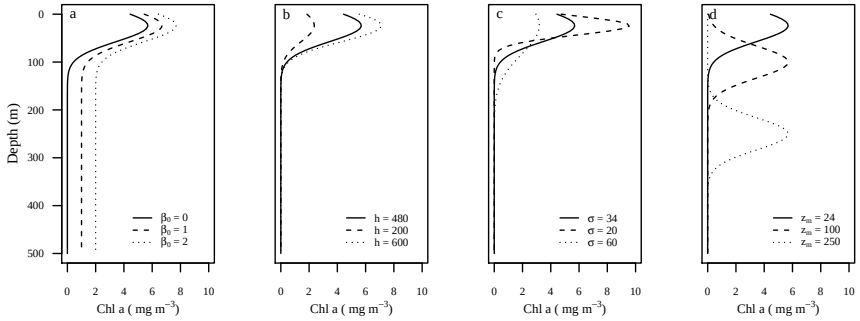

Figura 4.2: Changes of a typical chlorophyll profile when the parameter B_0 (a), σ (b), h (c), z_m (d) acquire different values.

Only profiles containing paired measurements of bacterial counts and chlorophyll concentration with more than 4 values for each single profile were characterized and thus considered for the analysis. The correlation coefficient between the measured and the estimated values was calculated for each profile as an estimate of the goodness-of-fit. We then validated how the MLE model performed by comparing the predicted values estimated using the parameters extracted from the MLE model with the in situ measured values.

4.2.4 Application of the SOM to the characterized profiles

The output parameters obtained from the MLE model for chlorophyll and bacterial abundance were used to characterize the shapes of the vertical profiles and how both variables were coupled along different regions. For this analysis, we applied

an unsupervised common artificial neural network (ANN), the self-organizing map (SOM) (Wehrens and Buydens, 2007) using the eight parameters of the Gaussian model (4 parameters for the bacterial and 4 parameters for the chlorophyll vertical profile) previously estimated for each profile as input data.

All the characterized profiles were included in the analyses, but when the z_m predicted for a profile was higher than the maximum sampled depth we removed it from the SOM analysis. The main reason was that in those cases the Gaussian curve described by the data was non-completely recorded and we can get non-realistic predictions. The final matrix with the selected profiles parameters for each variable (chlorophyll and bacterial counts), previously standardized, was introduced in the SOM model. The model output is a dimension-reduced matrix where the profiles sharing similar patterns are clustered in the same category. The SOM was performed using the kohonen R package (Wehrens and Buydens, 2007). The SOM analysis is an useful artificial neural network for visualizing high-dimensional data. Similarly to other ANN, SOM operates in two main steps: the training and mapping. The training allows to build a map using the input data while the mapping step automatically classifies and, thus simplifies the input data in a low-dimensional data vector. The self-organizing map is composed of nodes associated to a weight and a position vectors that provide a lower-dimensional map space (for further details see Kohonen (1997)).

A 5x5 dimension hexagonal map was selected obtaining 25 different categories for classifying the 2420 profiles. We then compared and separated the profiles for each category taking into account the similarity between chlorophyll and bacterioplankton abundance profiles. We divided the world's ocean into the Longhurst areas (Longhurst et al., 1995) and, for each region, we studied the percentage of profiles where both variables shared a similar vertical distribution and those where chlorophyll and bacterial abundance vertical profiles were uncoupled.

4.2.5 Predicting the shape of bacterioplankton profiles

We applied a Generalized Additive Model (GAM) for predicting the four parameters that characterize bacterial abundance profiles. The non-parametric

regression models are used for fitting non-parametric curves. The GAM is a non-parametric model without a priori constraint on the shape of the relationship. In these models the linear predictor depends on a sum of smooth functions. The model was fitted with the mgcv R package (Wood, 2015) using as input a set of variables (SST, surface chlorophyll concentration, depth of the water column, month). The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were used to select the best model and thus the optimal combination of input variables to estimate each parameter. A Leave-One-Out Cross-Validation (LOOCV) was used to estimate the goodness-of-the fit for each GAM model. To further study the GAM model performance, 50% of the dataset was used as input to fit the model and the rest of the data were used for predicting purpose.

4.2.6 **Predictions of global bacterial abundance and metabolic distribution**

The GAM models fitted using the entire database were then used to predict the Gaussian parameters at a global scale. Models with different variables were fitted and compared using the AIC and BIC criterion. The variables selected as the most likely to explain the cell abundance distribution were then used as input parameters.

The global maps of the four parameters in the Gaussian equation were used to calculate global vertical profiles of bacterial abundance using Eq. 4.2 to estimate the global distribution. The global abundance was calculated for the 29 depths levels available from 0 to 300 metres in the WOA13 Temperature Atlas. The bacterial abundance was then integrated from surface to the Z_{eu} to obtain the global bacterial abundance in the euphotic layer.

In order to obtain the respiration per cell in each depth level, we used the expression described by López-Urrutia and Morán (2007) as a function of temperature (Eq. 4.1). We then multiplied the image of predicted global bacterial, in each depth level, and the respiration per cell in order to calculate the global bacterial respiration. The mean-annual bacterioplankton depth-integrated respiration was then compared with mean annual global primary production estimates (see Satellite-derived data section). In order to make all the satellite images obtained

from different sources (and resolutions) comparable, we down-sized all the images to a 1°x1° resolution.

All data analyses were performed using R program (R Team (2013)).

4.3 Results

4.3.1 General description

The majority of the 58779 bacterioplankton samples comprising this study were from the Atlantic Ocean (69%). Artic and antartic waters were represented by the 16 and 7% of the data respectively. The remaining 8% was distributed between Paficic ocean, Indian ocean and the Mediterranean sea (Table S4.1). The global ocean was highly represented, ranging from 83°N to 78°S and containing data from surface to 5441 metres. Important time-series were included in the database as AZMP (North Atlantic), AZOMP (Labrador sea), Radiales (Cantabrian Sea) and BBMO (Mediterranean Sea) providing to our analysis also a high temporal coverage.

An overview of the dataset is provided by the density distribution of bacterioplankton abundance with temperature and chlorophyll (Figure 4.3). The data covered nearly the full environmental range of temperature (Figure 4.3 a) and chlorophyll (Figure 4.3 b). Both chlorophyll and bacterial abundance appeared lognormally distributed. Chlorophyll geometric mean value was of 0.35 mg m^{-3} with a maximum value of 39.85 mg m^{-3} recorded in the Cantabrian Sea and bacterial abundance geometric mean was $2.79*10^5$ cells ml^{-1} with a maximum value of $9.67*10^6$ cells ml^{-1} recorded in the Mediterranean Sea.

4.3.2 Characterization of vertical profiles

A total of 2420 chlorophyll and bacterioplankton abundance profiles were characterized. Figure S4.1 reveals the goodness-of-the fit, for most of the analysed profiles, the correlation obtained between measured and estimated values was higher than 0.75 for a 85 and 81% of chlorophyll and bacterial abundance profiles

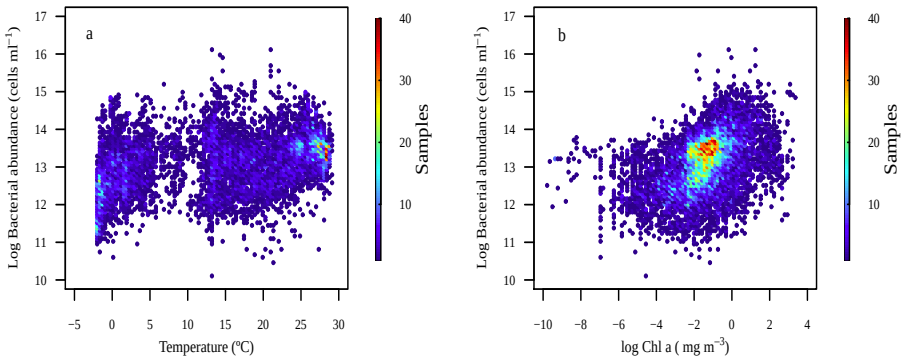

Figura 4.3: Observations of bacterial abundance as a function of temperature (a) and chlorophyll concentration (b) stored in our dataset.

respectively. The accuracy of the MLE fitting procedure for characterizing vertical profiles of bacterial counts and chlorophyll concentration was further evidenced by comparing the complete vertical profile measured and estimated using the equation 4.2 (Figure 4.4 a: r^2 = 0.98 for chlorophyll; Figure 4.4 b: r^2 = 0.95 for bacterioplankton).

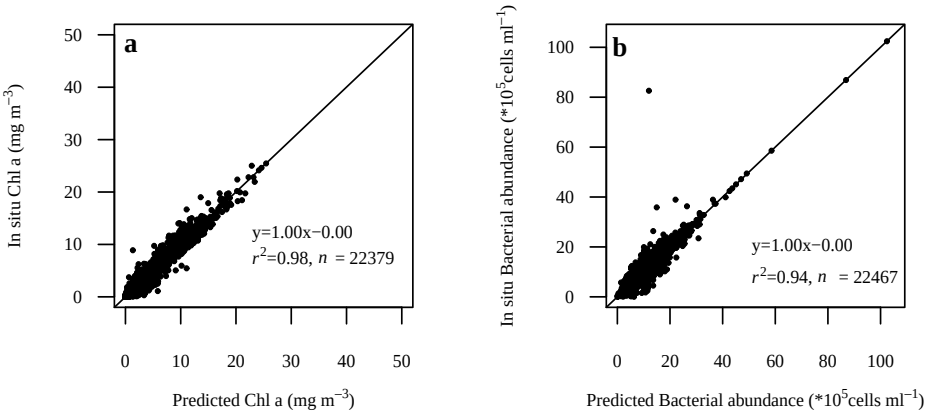

Figura 4.4: Comparison between in-situ chlorophyll concentration (a) and bacterial abundance (b) and predicted chlorophyll and bacterial abundance using MLE models for all data used in the profile's characterization.

4.3.3 Coupling of vertical profiles

The SOM model divided all the chlorophyll and bacterioplankton abundance profiles into the 25 selected categories as a function of the eight characterized parameters and therefore, of the shape described for each profile along the dataset (Figure 4.5). For most of the categories, the shape described for both variables was quite similar. However, the shape of profiles included into some categories (e.g. profiles a, d, e, f, h) evidenced important differences between the two variables. Table S4.2 summarizes the parameter values of each SOM category for bacterioplankton and chlorophyll and the number of profiles within each category. Despite the discrepancies found, those profiles where chlorophyll and bacterioplankton were uncoupled were less represented (259 profiles) than those sharing similar shape (1802 profiles).

Based on the differences between chlorophyll and bacterioplankton characterized parameters in each category, we divided the profiles into two categories: coupled and uncoupled profiles. The 87 % of the profiles (20 categories) were considered coupled ($r > 0.7$, p-value $< 10^{-16}$) while 13 % of the profiles (5 categories) were considered uncoupled.

To further study the concordances and discrepancies between the vertical profiles of the two variables and how they were distributed we used the 57 provinces defined by Longhurst et al. (1995) (Figure 4.6). The mapping of the percentage of profiles coupled showed that for almost the totality of the provinces covered by our dataset, more than the 50 % of the profiles were coupled (Figure 4.6 a), especially the regions located in latitudes over the 40°. In mid-latitudes we found some discrepancies (Figure 4.6 a-4.6 b). Remarkably, we found no discrepancies in all the provinces sampled near to coastal regions.

4.3.4 Selecting parameters to predict the shape of bacterioplankton profiles using non-parametric models

The best set of parameters to estimate the bacterioplankton B_0 parameter were the surface chlorophyll concentration, surface temperature, the month of the year and

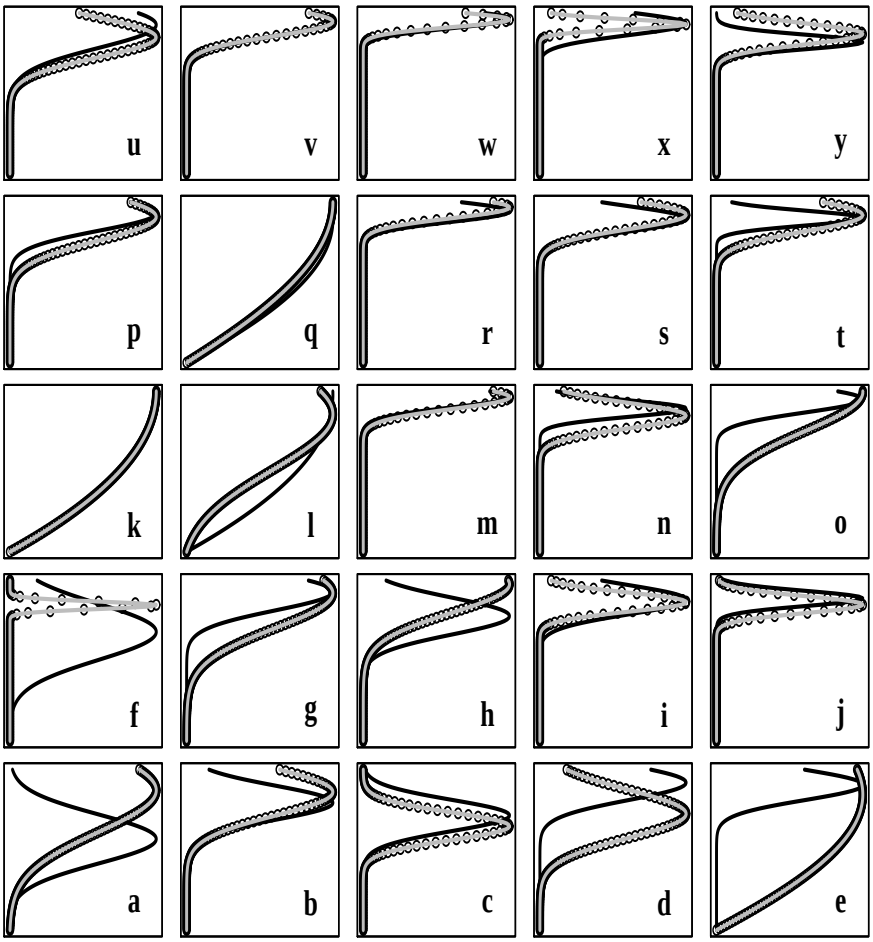

Figura 4.5: SOM profiles categories. Black line represents chlorophyll profiles and grey line represents bacterial abundance profiles. y-axis is the depth from 0 to 300 metres.

the depth of the water column ($p < 0.001$). We fitted the GAM model using the whole database and the Gaussian Cross Validation error (GCV) value obtained for this model was 0.32. A sensitivity analysis was also done for validating the fit (LOOCV), which revealed that the data included in the GAM were quite robust for the analysis. A 28 % of the variance was explained (0.27-0.3) and the Gaussian Cross Validation error (GCV) value was 0.32 (0.3-0.33). This analysis includes only the data corresponding to the stations where vertical profiles were available.

The rest of the parameters describing the shape of the profile (σ, h and z_m) were

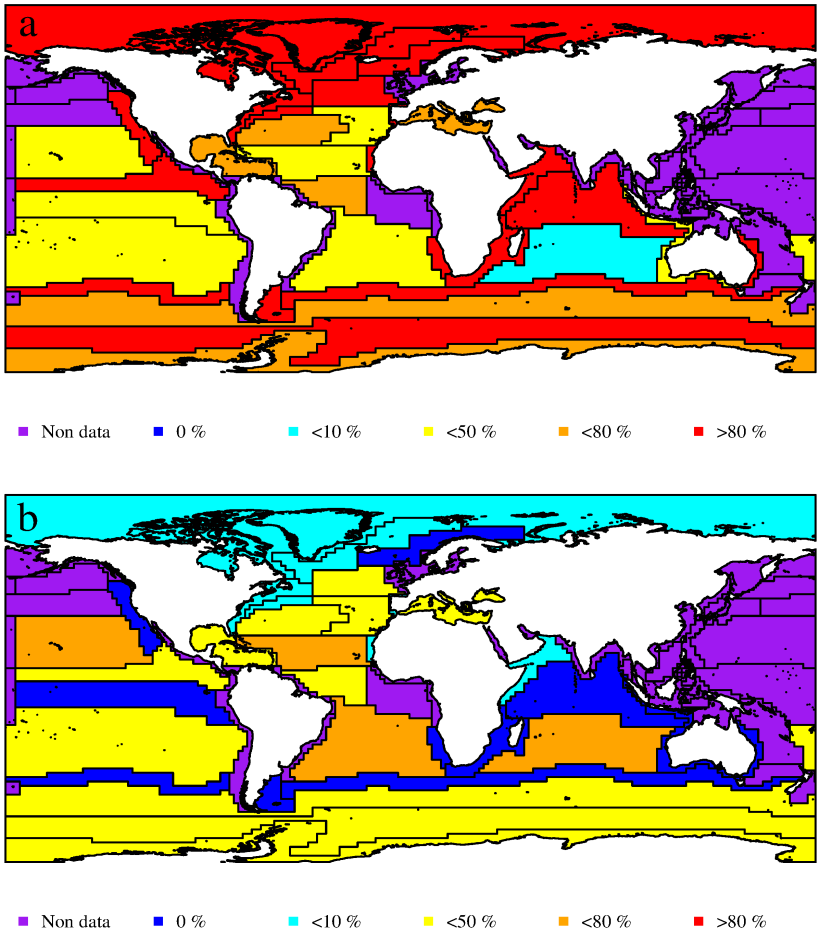

Figura 4.6: Distribution of the characterized profiles in the Longhurst areas. Panel a shows the distribution of the percentage of profiles coupled and panel b the distribution of the uncoupled profiles in the Longhurst areas.

also estimated from surface chlorophyll concentration, surface temperature, month and the depth of the water column as input variables. The variance explained was 18.3%, 26.1% and 16.4% and the GCV error was 0.4, 0.9 and 0.67 for σ, h and z_m respectively. Although all the data were used to fit the final GAM models, an additional fit using 50% of the dataset for fitting and the other 50% for predicting purpose was performed. This analysis allowed to understand how each parameter was predicted and detect possible under or overestimations (Figure 4.7).

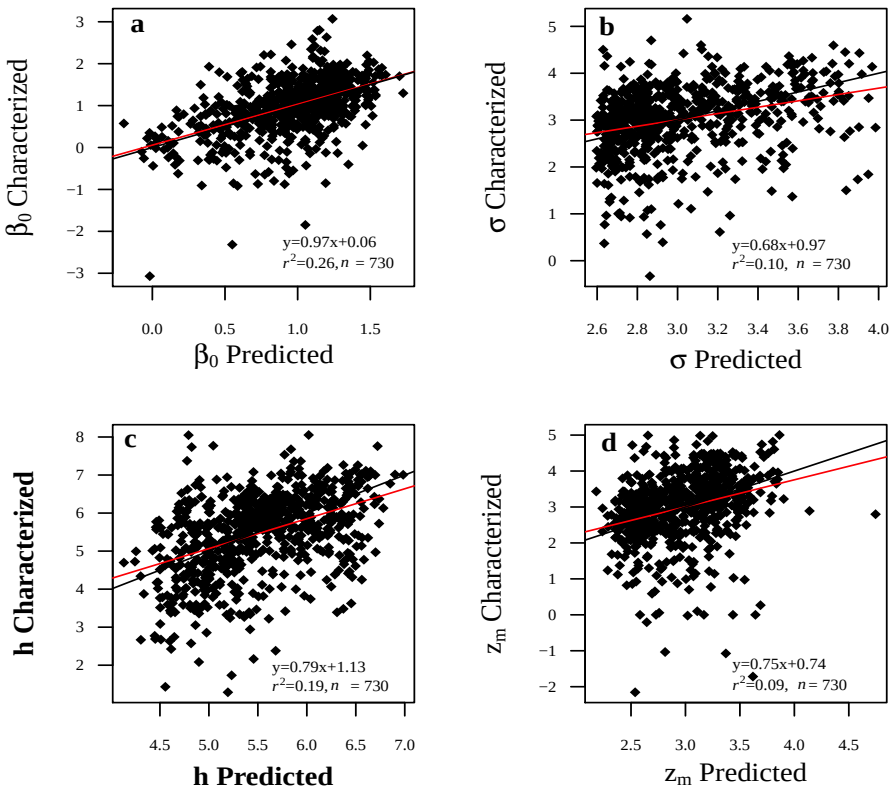

Figura 4.7: Comparison of the different parameters describing the shape of a Gaussian curve characterized by the MLE model (y-axis) and predicted using the GAM model (x-axis) in log scale. Black line represents the model fit and red line is the line (0, 1).

We observed that better predictions were done for the parameter B_0, with a slope close to 1, while for the other parameters we found an overall trend to overestimate low values and underestimate high values when comparing the characterized and predicted parameters (Figure 4.7). The four predicted parameters were then introduced in the Eq. 4.2 to calculate the bacterial counts predicted at the different sampled depths.

The comparison of these counts with the measured counts also revealed a slight overestimation of high values and underestimation of low values (Figure 4.8).

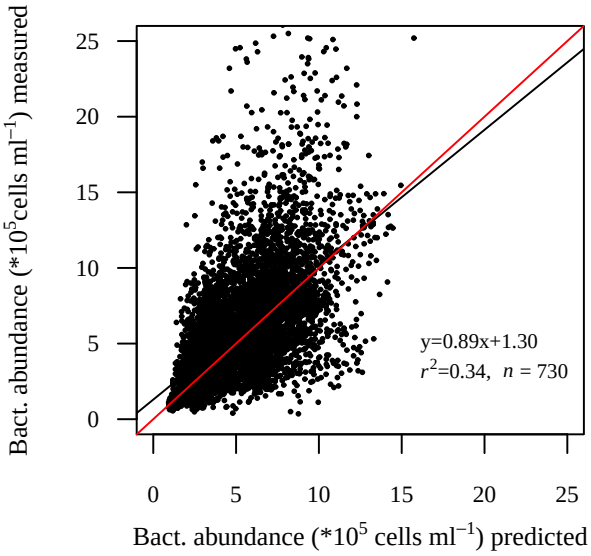

Figura 4.8: Comparison between bacterial abundance measured (y-axis) and predicted (x-axis) in cells ml^{-1}. Black line represents the model fit and red line is the line (0, 1).

4.3.5 Predictions of global bacterial abundance and metabolism using the GAM models

The non-parametric models were then used to calculate the mean annual global heterotrophic abundance distribution in the photic layer from satellite images. The maximum depth of the Z_{eu} was 244 metres corresponding to the clear oligotrophic oceanic waters. On average, the areas with the highest estimated bacterial abundance were the oligotrophyc areas, with a maximum value of $1*10^{14}$ cells m^{-2} estimated in the Southern Pacific ocean. The area where lower vertically integrated abundance was estimated was the Artic ocean, with $7.19*10^{11}$ cell m^{-2}. The mean annual abundance was $5.13 \pm 5*10^{13}$ cell m^{-2} (Figure 4.9 b).

The global bacterioplankton abundance estimated for the photic ocean was $1.69*10^{28}$ cells. However, the distribution pattern found at surface (5 metres, Figure

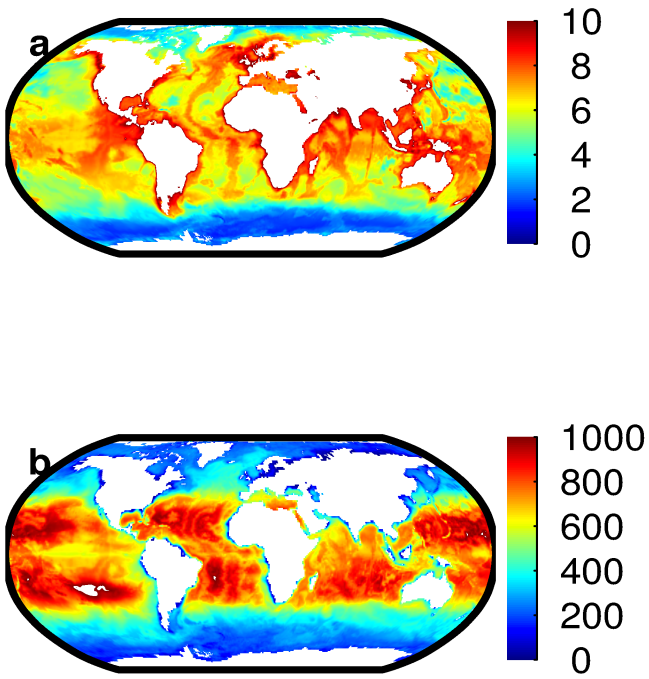

Figura 4.9: Bacterial abundance distribution. Panel a represents the global distribution of bacterial abundance at 5 metres depth (10^{11} cells m^{-3}). Panel b represents the depth-integrated bacterial abundance global distribution (10^{11} cells m^{-2}).

4.9 a) was quite different to the depth-integrated (Figure 4.9 b). The areas where a higher surface abundance was predicted were coastal regions with values around $1*10^{12}$cell m^{-3} while polar areas showed a lower abundance.

The mean annual global abundance estimated and the satellite-derived SST images were used to predict the mean annual global heterotrophic picoplankton community respiration (Figure 4.10 a).

The mean respiration obtained was 1.36 ± $1.2*10^{15}$ fg C day^{-1}. In lower latitudinal waters 30-40° (both N and S) higher respiration values were predicted. The maximum value predicted was $6.8*10^{15}$ fg C day^{-1} in the Caribbean sea. The minimum respiration estimated was $9*10^{12}$ fg C day^{-1} from the Artic ocean.

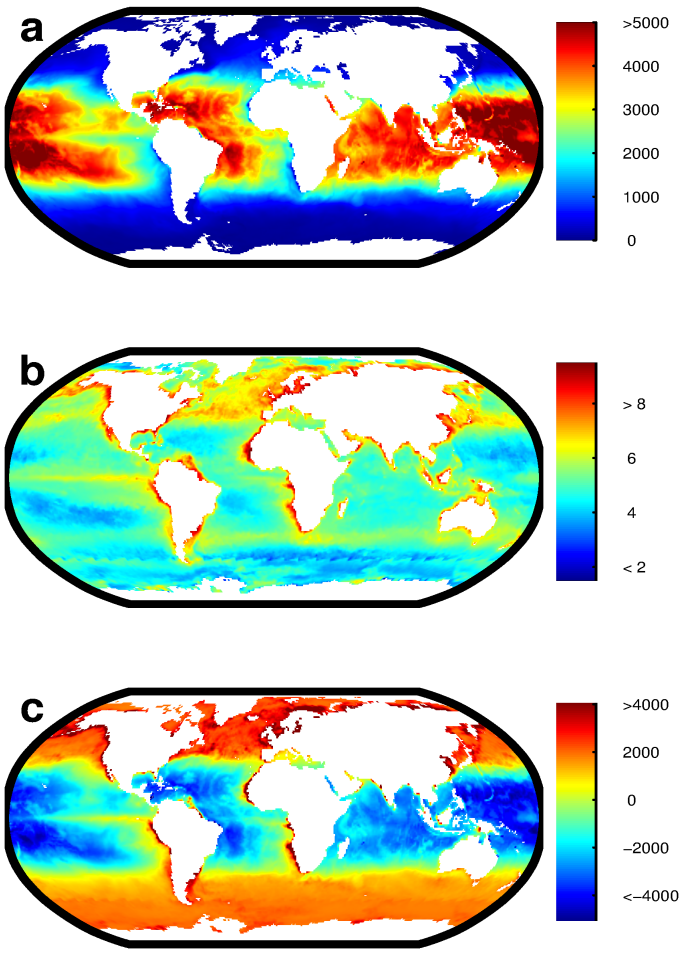

Figura 4.10: Metabolic predictions. Depth-integrated bacterial respiration (a) (10^{12} fg C m^{-2} day^{-1}). Depth-integrated primary production (b) (10^{12} fg C m^{-2} day^{-1}). Balance between bacterial respiration and primary productions (c) (10^{12} fg C m^{-2} day^{-1}).

Overall, in all polar and coastal waters where the depth of the euphotic layer was shallower, we predicted lower depth-integrated respiration while the oligotrophic areas where the euphotic layer was deeper, higher values were predicted. We then estimated the global respiration value for the photic layer that amounted to 146 Gt C yr^{-1}.

Regarding the primary production, the mean annual value estimated was 456

± 622 mg C m^2 day^{-1}. The areas more productive were coastal regions and the northern North Atlantic. The maximum value recorded was 11027 mg C m^{-2} day^{-1} from the upwelling of the eastern subtropical North Atlantic Ocean (Figure 4.10 b). The less productive areas were mainly the subtropical gyres and polar waters with a minimum value of 18 mg C m^{-2} day^{-1} predicted in the Southern ocean (Figure 4.10 b). The mean global primary production estimated was 51.83 Gt yr^{-1}.

Our results revealed a net heterotrophic ocean where respiration obtained was around 3 times higher than the production in the euphotic layer. Combining both images, an extensive area with negative values was found in low latitudes (40°) while the coastal and latitudes over 40° were net autotrophic (Figure 4.10 c). Despite we found a similar coverage of autotrophic and heterotrophic areas the magnitude of the respiration in mid-latitudes was higher than those found for primary production with a resulting negative balance of 94 Gt C yr^{-1}.

4.4 Discussion

Our results point out a strong vertical agreement between bacterioplankton abundance and chlorophyll concentration for 87% of the analysed profiles. Previous studies have demonstrated that both variables are highly correlated in surface waters (Ducklow and Carlson, 1992; Gasol and Duarte, 2000; Li, 1998; Li and Harrison, 2001; Zubkov et al., 1998). Concordances were also found in the vertical gradient along different ecological provinces in the North Atlantic (Li and Harrison, 2001). In other regions, similar vertical agreement have been reported (Church et al., 2006; Cochlan et al., 1993)

However, to our knowledge, this is the first time that single profiles of both variables were compared and characterized to study this relationship at a global scale. The MLE of the Gaussian shape of all vertical profiles was very accurate, most of the vertical profiles estimated and measured were strongly correlated (r^2 > 0.75, Figure S4.1, Figure 4.4). It means that the majority of vertical profiles of bacterial abundance and, also chlorophyll biomass, follow a Gaussian distribution pattern in the oceans similarly to the results obtained in previous studies using chlorophyll profiles (Richardson et al., 2002).

Combining the SOM approach with the characterized profiles was crucial to determine the different types of profiles that appear in the ocean. Therefore, our results further demonstrated the link between bacterial abundance and phytoplankton resources also in the vertical gradient. Nevertheless, some discrepancies were found, especially in the subtropical gyres. The origin of these differences was not clear. It might come from methodological problems or it can really exist a decoupling of both variables in those areas. In that case, the decoupling might come from a bottom-up control in areas where nutrients become limiting. It is further known that in oligotrophic areas phytoplankton abundance decreases due to the scarcity of nutrients while the bacterioplankton abundance remains constant or suffers slight decreases (Li and Harrison, 2001).

Despite the high flexibility of the GAM models and the wide range covered by the dataset, the variance explained by our models was weak (Figure 4.8). This means that, on the one hand, including all the characterized profiles may cause the deviance of the slope recorded for some parameters from 1 which is translated into slight over and underestimates in the bacterial counts (Figure 4.8). On the other hand, other factors with a considerable influence in the bacterial abundance distribution (e.g. mineral nutrients and organic carbon) that we are not included in the analysis may cause the low predictive capacity shown by our models (Apple et al., 2006). In our approach we have not included the effect of a possible top-down control by bacterivores and viruses which has been reported to influence the relationship between these variables (Li et al., 2004). It is also required to note that when we compiled data from different sources it is necessary to assume that some bias might be introduced from the different methodologies used. The methodological noise introduced in an extensive database is hard to determine.

Our results point out the relevance of heterotrophic picoplankton in the least productive areas. This is because Z_{eu} is deeper in the clear oligotrophyc gyres where a pronounced decrease of resources and, therefore, of primary producers occurs while the abundance of heterotrophic bacteria remains stable. Hence, a change in the ratio primary production/heterotrophic respiration towards the dominance of the heterotrophs is recorded (Figure 4.10). Our approach is a first attempt to predict and quantify the distribution of heterotrophic bacteria in the

photic layer of the ocean providing a global view that allows to compare the roles of heterotrophs and phototrophs with the advantage of including the vertical dimension in the approach. It should be emphasized that even if a coarse estimation is obtained in this study, the progress made allows to simplify what is happening in the nature providing valuable information for the marine ecology community.

The estimates of cell-specific respiration were obtained with the equations developed by López-Urrutia and Morán (2007) based on the MTE. They demonstrated that bacterioplankton cell-specific respiration rate is regulated by temperature but not by the availability of substrates. Additionally in Chapter 2 we have shown that the respiration rate responds to the changing cell size using a size-fractionated community whose cells were under the same external conditions. However, when the respiration rates and cell size of the sampled communities from different locations where studied the effect of size in the overall community metabolism was not evidenced. Hence, we have considered that temperature is the main factor controlling the respiration rate for bacterioplankton using thus the equations reported by López-Urrutia and Morán (2007) and the predicted abundance for estimating the bacterial community respiration of the ocean. Bacterial respiration has been reported to contribute to about 50 % of the total community respiration (del Giorgio and Duarte, 2002; Rivkin and Legendre, 2001; Robinson and Williams, 2005), albeit this contribution might be quite variable (García-Martín et al., 2014).

Our coarse estimate of the mean global respiration in the euphotic layer (around 1.36 g C m^{-2} day^{-1}) is comparable to the value reported by Williams (1998) and Duarte and Agusti (1998) in surface waters (around 1.2 g C m^{-2} day^{-1}). This value has been reported to be quite high due to an overestimation in eutrophic areas (del Giorgio and Duarte, 2002). However, the bacterial abundance predicted by our approach ($1.69*10^{28}$ cells) is lower than previous estimates of heterotrophic prokaryotes in the upper 200 metres of the oceans ($3.6*10^{28}$ cells) (Whitman et al., 1998).

In our study, where the respiration estimates were based on heterotrophic abundance and temperature measurements, we know that 1) bacterioplankton

counts were highly coupled with chlorophyll concentration especially in polar and coastal regions with some discrepancies in oligotrophic areas, 2) our GAM model, overestimates bacterial counts in areas with higher abundance, and subestimates in areas with less bacterial cells, 3) We did not considered organic matter in our analysis which is a main source of bacteria (Apple et al., 2006). Therefore, the high value of bacterial abundance and respiration presented in this study has to be taken into account as a coarse estimation that simplifies the picoplankton metabolic complexity of the oceans providing a valuable information. However, at the local scale some disagreements are evident and further model developments might be required.

This study reveals the importance of improving the methodologies and increasing the sampling effort focusing on the heterotrophic picoplanktonic organisms and their substrates. In this sense, the contribution of international programs and cruises such as JGOFS, AMT, MAREDAT and the circumnavigation Malaspina has been crucial. The main advantage of an increase in the accuracy of bacterioplankton distribution models is to be able to compare them with the current production models outputs. Despite the improvements required, our approach in terms of abundance and respiration has been addressed to be in good agreement with previous estimates.

In summary, our study provides a global view of how bacterioplankton communities are distributed in the photic area of the ocean and their coupling with chlorophyll biomass by characterising the profiles of both variables in this layer at global scale. These analyses have revealed the possibility of using these relationships to quantify bacterial abundance from remote sensing. Our model provides a quantitative basis for predicting and understanding shifts in bacterioplankton distributions.

4.5 Supplementary figure and tables

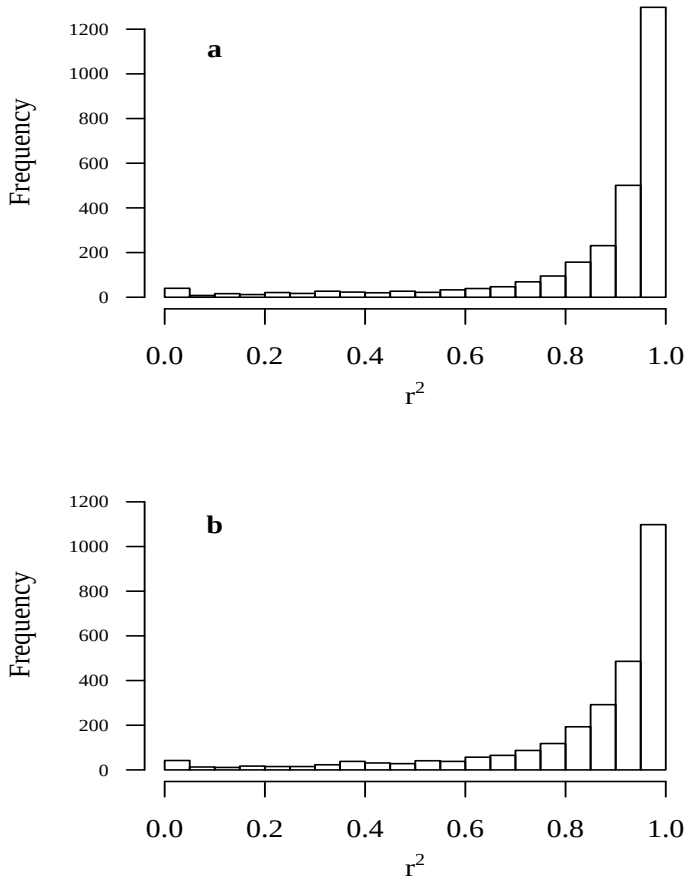

Figura S4.1: Frecuency distribution of correlation between profiles measured and characterized for chlorophyll concentration (a) and bacterial abundance (b)

4. Bacterioplankton distribution in the euphotic layer: an attemp to predict the metabolic balance of the oceans

Table S4.1. Data sources.

Cruise	Date	Area	Reference	n
Artic	2002−present	Artic	(Li, 2009)	9272
AZMP	1997−present	North Atlantic	(Li, 2009; Li et al., 2004)	35376
AZOMP	1994−present	Labrador Sea	(Li, 2009; Li et al., 2004)	
Beagle	Aug2003−Feb2004	Antartica	(Li, 2009)	232
ICEFISH	2004	Antartica	(Li, 2009)	112
CONAFRICA	Mar−Apr2006	NE Atlantic	(Baltar et al., 2007)	45
Dharma	Dec1998	Antartic Ocean	(Díez et al., 2004)	313
Latitud−1	Mar−Apr1995	Atlantic Ocean	(Vázquez-Domínguez et al., 2008)	173
Latitud−2	Oct−Nov1995	Atlantic Ocean	(Vázquez-Domínguez et al., 2008)	328
Coca−1	Sep2002	NE Atlantic	(Alonso-Sáez et al., 2007)	151
Coca−2	May−Jun2003	NE Atlantic	(Alonso-Sáez et al., 2007)	109
Inco−I	Apr−May1997	NE Atlantic	Gasol(Unpub)	59
Inco−II	Aug−Sept1998	Atlantic	Gasol(Unpub)	90
FRUELA−1	Dec1995−Jan1996	Antartica	(Pedrós-Alió et al., 2002)	335
FRUELA−2	Jan−Feb1996	Antartica	(Pedrós-Alió et al., 2002)	291
WECOMA	May2002	NE Pacific	(del Giorgio et al., 2011)	21
BBMO	1997−2010	Mediterranean Sea	Gasol	185
Efluvio04	Sep2004	Mediterranean Sea	(Latasa et al., 2010)	64
Efluvio05	Mar-Apr2005	Mediterranean Sea	(Latasa et al., 2010)	104
Microdeep	Mar2010	Mediterranean Sea	Gasol	257
Varimed95	Jun1995	Mediterranean Sea	Gasol	372
Varimed96	Jun1996	Mediterranean Sea	Gasol(Unpub)	257
Radiales	Apr2002−Sep2010	Cantabrian Sea	Moran(in prep)	2054
Malaspina	Dec2010−Jul2011	Atlantic/Indian/Pacific	Gasol	1111
AMT17	2007	Atlantic Ocean	Zubkov(Unpub)	711
AMT21	Sep−Nov2011	Atlantic Ocean	Garcia(Unpub)	38
AtlantisII	Apr−Jun1989	North Atlantic	JGOFSProgram	368
Endeavor	Jun−Jul1989	North Atlantic	JGOFSProgram	36
TTN−043	Jan1995	Arabian Sea	JGOFSProgram	262
TTN−045	Mar−Apr1995	Arabian Sea	JGOFSProgram	339
TTN−049	Jul−Aug1995	Arabian Sea	JGOFSProgram	354
TTN−050	Aug−Sept1995	Arabian Sea	JGOFSProgram	588

TTN−054	Nov−Dec1995	Arabian Sea	JGOFSProgram	748
TT−007	Feb−Mar1992	Equatorial Pacific	JGOFSProgram	582
TT−008	Mar−April1992	Equatorial Pacific	JGOFSProgram	75
TT−011	Aug−Sept1992	Equatorial Pacific	JGOFSProgram	366
TT−012	Oct1992	Equatorial Pacific	JGOFSProgram	110
NBP−96	Oct−Nov1996	Southern Ocean	JGOFSProgram	158
NBP−97	Jan−Feb1997	Southern Ocean	JGOFSProgram	153
NBP−97	Apr1997	Southern Ocean	JGOFSProgram	110
NBP−97	Nov−Dec1997	Southern Ocean	JGOFSProgram	198
RR−6	Oct−Nov1997	Southern Ocean	JGOFSProgram	36
RR−7	Dec1997	Southern Ocean	JGOFSProgram	212
RR−8	Jan1998	Southern Ocean	JGOFSProgram	50
RR−9	Feb−Mar1998	Southern Ocean	JGOFSProgram	299
BATS90−97	Jan1992−Jul1997	Bermudas	Montes−Hugo	15
AESOPS	Jan−Apr1997	RossSea	Ducklow, Smith	84
Palter03	Jan−Feb2003	Antartica	Montes−Hugo	38
lterjan96	Jan1996	Antartica	Montes−Hugo	964

Table S4.2. Chlorophyll and bacterioplankton abundance parameters estimated in each SOM category (C).

			Chl				Bact. Abu.		
C	B_0	σ	h	z_m	B_0	σ	h	z_m	n
a	0.068	45.81	25.3	129.66	0.98	78.13	1365.22	38.64	68
b	0.08	31.90	30.28	61.89	1.12	44.05	697.23	41.89	136
c	0.23	29.16	35.32	85.76	4.04	25.78	373.97	105.81	20
d	0.19	32.60	74.98	24.30	1.77	44.89	438.25	82.65	97
e	0.32	28.26	38.98	28.55	0.60	337.38	2457.06	46.62	7
f	0.20	52.36	12.48	96.03	3.57	6.93	76068.38	46.40	2
g	0.11	34.79	110.32	21.12	0.21	65.19	1132.86	23.79	182
h	0.07	34.48	28.34	66.55	1.92	53.38	756.46	6.91	85
i	0.17	27.40	63.39	35.99	4.71	17.88	281.56	40.50	133
j	2.88	14.24	235.97	33.57	4.73	15.85	565.79	45.14	20
k	0	1483.94	592.81	6.63	0	1517.23	6978.46	0	2
l	0	1483.94	592.81	6.63	0	1517.23	6978.46	0	6
m	0.07	19.37	49.83	10.59	2.22	22.16	390.11	10.80	423
n	5.12	15.42	222.03	32.25	3.98	23.79	256.26	45.30	14
o	10.63	24.45	982.74	15.02	1.78	72.65	1059.34	1.58	2

p	0.19	32.75	739.61	20.57	2.02	44.27	772.60	27.43	76
q	0	461.20	143.87	33.14	0.22	424.26	2676.55	0	3
r	0.18	19.33	70.66	17.31	4.22	20.09	421.05	9.73	274
s	0.07	23.68	40.89	31.10	2.17	27.12	415.40	23.31	246
t	1.68	13.93	144.30	29.70	2.87	30.32	482.83	24.08	43
u	0.23	42.98	1339.30	22.08	3.13	37.12	271.81	45.41	17
v	0.25	27.86	364.98	14.87	3.36	26.70	466.80	16.02	98
w	0.24	17.40	56.32	10.72	7.05	14.22	243.24	12.02	84
x	0.25	21.35	63.91	19.87	12.91	9.25	116.31	20.91	22
y	0.16	13.64	23.59	53.97	42.67	19.67	3454.05	38.89	1

General Discussion

The goal of this thesis was to disentangle and understand the main factors controlling marine bacterioplankton distribution and metabolism for predicting their role in the metabolism of the ocean. The main difficulty to achieve this goal was the scarcity and methodological inconsistency within datasets which introduce an important variability in the observed mathematical results.

On the other hand it is further known that all the factors affecting planktonic communities are highly interrelated, what makes difficult to identify all the factors affecting bacterioplankton and, thus to quantify their effects on microbial communities. The availability of empirical datasets has been crucial to solve our main goals and, therefore, to improve the understanding of the marine bacterioplankton ecology. However, there have been several results in this thesis that should be further discussed.

Flow cytometry: a promising tool for assessing the dynamic of microbial communities in the ocean

A relevant point in our results has been to show the potential of flow cytometry for assessing changes in bacterioplankton communities. Flow cytometry is a routine automatic tool in aquatic microbial ecology. However, the potential of this tool, especially regarding the information stored in the flow cytometry standard (FCS) files is not completely harnessed in this area of the ecology.

The FCS files contain valuable information of each single cell about light scattering and fluorescences. However, the information generally used in aquatic microbial studies are the mean counts of each manual analysed group, the mean value of light scatter (as measure related to size) and the fluorescences. Here, we have proposed two methodologies for applying to the FCS files that allow, firstly the automated clustering of the different groups present in each sample objectively. This routine allows the access to the information about the distribution of side scatter and/or fluorescences for each group which offers a wide range of possibilities for detecting morphological differences between the groups present in the different samples. Secondly, the applicability of the concept of "cytometric diversity" proposed by Li (1997) for assessing changes in bacterioplankton community validated with molecular analyses presented in Chapter 2 supposes a further step for obtaining more information of the flow cytometric samples and, therefore of the microbial populations analysed.

The capability of combining both methodologies and incorporating them to the routine flow cytometry analyses may increase the potential of this valuable technique, especially in macroecological studies of microbial communities. Obviously these tools provide an overall view of the dynamic of the community being required additional analyses, using more specific tools (e.g. molecular analysis) to obtain a profound and more accurate assessing of these communities.

The non-universality of the scaling exponent for microbial groups

One of the most outstanding results presented in this thesis is the evidence of the effect of cell size in the metabolism of marine bacterioplankton analysed in Chapter 3. Our results revealed that bacterioplakton respiration rate scales superlinearly with cell size with a scaling exponent > 1 which varies depending on the bacterioplankton community composition. This value differs of the traditional $\frac{3}{4}$ established as universal by the Metabolic Theory of the Ecology (MTE) and was found not universal along the study. The universality of this exponent is a focus of debate and some studies have shown discrepancies mainly for the smallest organisms (DeLong et al., 2010; Makarieva et al., 2005).

Our results suggest, in agreement with the cited studies, that the effect of size in the metabolism of microbes need to be further explored to improve our knowledge of how the composition of the community is affecting to their metabolism. This fact is particularly relevant when we attempt to estimate the respiration of bacterioplankton in Chapter 4. However, our results do not evidence an effect of size in the respiration rate when we considered communities from different areas of the ocean. It seems that the effect of size is meaningless in the respiration rate of these organisms compared with other variables as temperature which has been further demonstrated to be a key factor controlling the respiration rate of picoheterotrophs (Apple et al., 2006; del Giorgio and Williams, 2009; López-Urrutia and Morán, 2007)

Main factors controlling the bacterial abundance distribution

In an effort to understand how related are bacterioplankton abundance and chlorophyll concentration, we have compared both variables in the vertical gradient. This study has revealed that it is possible to predict bacterial abundance from chlorophyll concentration measurements quite accurately. However in the oligotrophic areas we have found some discrepancies between both variables. It seems to suggest that other parameters highly related with bacterial abundance are required to be incorporated in routine samples as the dissolved organic matter.

Bacteria incorporates directly from the ocean the dissolved organic matter (DOM) (Apple et al., 2006; Jiao et al., 2010; Kirchman et al., 2005). The pattern of distribution for dissolved organic carbon (DOC) in surface waters (Hansell et al., 2009), is quite similar to the pattern found for bacterioplankton communities (Chapter 4). A decrease of the DOC concentrations have been detected in high latitudes, similar to the pattern that we found for bacterial counts while in the subtropical gyres the DOC concentration were quite significant (Hansell et al., 2009). This suggest that in some areas of the ocean organic carbon measurements could be a better proxy of bacterial abundance than chlorophyll concentration. Hence, increasing the number of paired measurements of bacterial counts and

DOM might be a next step for improving our predictions and thus, to confirm whether the ocean is really as heterotrophic as our primary approach reveals.

Conclusiones

1. La identificación manual de los distintos grupos de bacterioplancton procedentes de archivos analizados por citometría de flujo introduce un porcentaje de error importante en los contajes debido a la subjetividad introducida por el analista.

2. La automatización de la detección de grupos de células procedentes de archivos citométricos utilizando algoritmos matemáticos permite analizar un gran número de muestras en poco tiempo y con apenas supervisión. Esto hace que muestras procendentes de distintas fuentes puedan ser comparables reduciendo el error introducido por la subjetividad del analista.

3. La citometría de flujo nos permite detectar cambios en la composición de la comunidad de bacterioplancton heterotrófico gracias al concepto de "diversidad citométrica". Dividiendo las células analizadas en distintas categorias citométricas, hemos podido obtener una estima de diversidad basada en las características morfológicas de las células. La comparación de estimas de diversidad citométrica con estimas de diversidad realizadas con técnicas moleculares nos ha permitido validar este concepto para la detección de cambios composicionales a gran escala.

4. La aplicación del concepto de diversidad citométrica a la serie temporal Radiales de análisis citométricos en el Cantábrico, nos ha permitido detectar cambios estacionales en la diversidad con máximos superficiales en invierno y mínimos en verano. El estudio de la dinámica temporal en 3 estaciones costeras a lo largo de la serie temporal junto con el estudio de otras variables como nutrientes, capa de mezcla y duración del día entre otras, nos ha permitido establecer la mezcla

como principal variable responsable de los aumentos en diversidad detectados en invierno. Nuestros resultados revelan que estos aumentos son debidos a un incremento en el número de especies y que especies típicas de aguas más profundas aparecen en superficie durante el periodo de mezcla. Esto nos ha permitido establecer una relación entre la mezcla y el afloramiento de especies desde las capas mas profundas hacia la superficie en los meses de invierno.

5. El tamaño celular de los microorganismos, contrariamente a lo que se pensaba, debido fundamentalmente a su pequeño tamaño, influye en el metabolismo de estos organismos. Este metabolismo aumenta considerablemente con el tamaño celular gracias a la relación entre tamaño celular y número de genes codificantes de proteinas que existe en estos organismos. A medida que aumenta el número de genes y por lo tanto su complejidad metabólica el exponente alometrico aumenta a valores por encima de 1. Al subdividir poblaciones de bacterioplancton marino (< 0.8 μm de diametro) en distintos rangos de tamaño, hemos podido observar estas diferencias ya que cuando comunidades de distintas zonas son estudiadas, esta relación no puede ser claramente establecida debido a la influencia de otros factores externos altamente relacionados con el metabolismo celular de estos organismos.

6. La respiracion celular del bacterioplancton marino escala de manera superlineal con su tamaño celular pero este exponente, contrariamente a lo expuesto por la Teoria Metabolica de la Ecologia. no es universal para organismos unicelulares como las bacterias heterotroficas. Diferencias significativas en el exponente alometrico fueron encontradas a lo largo del gradiente latitudinal estudiado. Este exponente desciende hacia 1 a medida que organismos mas evolucionados (cianobacterias) adquieren relevancia en la composicion de las comunidades.

7. La base de datos recopilada en el capítulo 4 aporta una gran cobertura espacio-temporal que nos ha permitido profundizar en la macroecología del bacterioplancton marino y establecer su relación con otras variables utilizando datos empíricos. Esta base de datos, una de las mas completas en microbiologia marina, recopila datos de zonas polares en su mayoria (oceano Artico y Antartico) y un gran numero de datos recogidos en zonas costeras y oceanicas incluyendo

los grandes giros oligotróficos. La alta resolución espacial y la incorporación de todo tipo de sistemas gracias al amplio gradiente ambiental incorporado, le proporcionan una gran solidez haciéndola única y fundamental para el estudio y parametrización de la relación del bacterioplancton marino con otras variables.

8. El bacterioplancton marino esta altamente ligado a los productores primarios estimados a partir de la concentracion de clorofila tanto en superficie como en profundidad. La caracterización de perfiles verticales a escala global para ambas variables, demuestra la alta relación entre abundancia bacteriana y clorofila en la vertical.

9. Relacionar variables como la abundancia de microorganismos con variables medibles a partir de satélite, nos ha permitido modelar estas relaciones y predecir los patrones de distribución a nivel global.

10. Las discrepancias encontradas entre abundancia bacteriana y clorofila en los giros oligotróficos, así como la baja varianza explicada por las variables predictivas en nuestra aproximación, revelan que variables que actúan como limitantes de recursos para las bacterias como por ejemplo, la materia orgánica, deben ser incorporadas a los análisis rutinarios. Esto pone de manifiesto la necesidad de vincular materia orgánica real con variables ópticas para poder obtener estas estimas de sensores remotos. El avance en las cuestiones expuestas aquí proporcionaría grandes avances en la predicción de microorganimos heterotróficos en el océano, proporcionando a estas estimas una mayor precisión que las acerque y haga mas comparables con las obtenidas para productores primarios.

11. La respiración heterotrófica llevada a cabo por las bacterias en el océano podria tener valores mucho mayores de los revelados por otros estudios. Esto supondría que el océano es netamente heterotrófico teniendo graves consecuencias para el calentamiento global y suponiendo un riesgo elevado de cambiar la dinámica metabólica de las comunidades planctónicas debido a la elevada relación que hay entre temperatura y metabolismo planctónico.

Annexes

Annex 1: R code for automated clustering bacterioplankton cells

Internal functions

CleanMatrix < − function(FullMatrix, Allx, Ally,
BreaksX = 75,
BreaksY = 75,
xlims = range(Allx, na.rm = TRUE),
ylims = range(Ally, na.rm = TRUE),
Threshold = 2){ Threshold to delete cells with lower number of counts
Assigns each point to a XYCells
Xbrks < − seq(xlims[1], xlims[2], length = BreaksX)
Ybrks < − seq(ylims[1], ylims[2], length = BreaksY)
Xfactor < − cut(Allx, breaks = Xbrks, include.lowest = TRUE)
Yfactor < − cut(Ally, breaks = Ybrks, include.lowest = TRUE)
XYfactor < − factor(paste(Xfactor, Yfactor, sep = "_"))
This is the factor that specifies for each point to what cells it belongs
Unos < − rep(1, length = length(Allx))
Dummy vector to count number of points in each cell
Counts < − tapply(Unos, XYfactor, sum)
Sums the number of points in each level of XYfactor
Celdas < − levels(XYfactor) All the levels in XYfactor
CellsToDelete < − Celdas[which(Counts > Threshold)]
Levels that have counts lower than Threshold
ParticlesToDelete < − XYfactor%in%
CellsToDelete Rows of particles to delete
CleanFullMatrix < − FullMatrix[ParticlesToDelete,]
return(CleanFullMatrix) }

CleanMatrix3D < − function(FullMatrix, Allx, Ally, Allz,
BreaksX = 40,
BreaksY = 40,
BreaksZ = 40,

xlims = range(Allx, na.rm = TRUE),
ylims = range(Ally, na.rm = TRUE),
zlims = range(Allz, na.rm = TRUE),
Threshold = 2){ Threshold to delete cells with lower number of counts
Assigns each point to a XYCells
Xbrks < − seq(xlims[1], xlims[2], length = BreaksX)
Ybrks < − seq(ylims[1], ylims[2], length = BreaksY)
Zbrks < − seq(zlims[1], zlims[2], length = BreaksZ)
Xfactor < − cut(Allx, breaks = Xbrks, include.lowest = TRUE)
Yfactor < − cut(Ally, breaks = Ybrks, include.lowest = TRUE)
Zfactor < − cut(Allz, breaks = Zbrks, include.lowest = TRUE)
XYfactor < − factor(paste(Xfactor, Yfactor, Zfactor, sep = "_"))
This is the factor that specifies for each point to what cells it belongs
Unos < − rep(1, length = length(Allx)) Dummy vector to count number of points in
each cell
Counts < − tapply(Unos, XYfactor, sum) Sums the number of points in each level
of XYfactor
Celdas < − levels(XYfactor) All the levels in XYfactor
CellsToDelete < − Celdas[which(Counts > Threshold)]
Levels that have counts lower than Threshold ParticlesToDelete< −XYfactor%in%CellsToDelete
Rows of particles to delete
CleanFullMatrix< −FullMatrix[ParticlesToDelete,]
return(CleanFullMatrix)}

Reformat.FCS < − function(File.FCS,
The complete path to the FCS file to be analysed
column.names.fcs, Column names for the FCS file
CleanBeads = TRUE){

FlowData < − read.FCS(File.FCS)
Log transfrom the raw Flowcytometer data
tf < − transformList(from = colnames(FlowData), tfun = log)

```
FlowData < − tf%on% FlowData


   if(CleanBeads){NoBolas < − which(exprs(FlowData)[, 3] <= 6.0)
exprs(FlowData) < − exprs(FlowData)[NoBolas, ]}
Convert FCS object into a Matrix
return(FlowData)}


   Extracts the best information criterio for each K in a flowClust object
BICf < − function (x) as.numeric(unlist(lapply(x, function(q) try(q@BIC, silent =
TRUE))))


   changepoint < − function (BIC, NTestGroups) {
K < − 1: NTestGroups Sometimes BIC contains NaN
BICfinite < − is.finite(BIC)
K < − K[BICfinite]
BIC < − BIC[BICfinite]
MaxBIC < − BIC[which(BIC == max(BIC))]
if(MaxBIC < 0){ Proportion < − MaxBIC / BIC
} else if (MaxBIC > 0) Proportion < − BIC / MaxBIC
Group < − c(1: NTestGroups)
change.point < − min(Group[which(Proportion >= 0.95)])
return (change.point)}


   recover.cell < − function(MatFlow,Groups,OriginalRows){
MatFlow < − MatFlow[, c("SSC", "FL1", "FL3")]
Full.labels.groups < − rep(NA, nrow(MatFlow))
labels.groups < − Map(Groups)
New.labels.groups < − matrix(NA, nrow(MatFlow), ncol = max(labels.groups,
na.rm = TRUE))
labels.groups[is.na(labels.groups)] < − 0
Full.labels.groups[OriginalRows] < − labels.groups
```

```
DetectInCloud< −function(MatFlow, Full.labels.groups, GroupCloud){
ps < − MatFlow[which(Full.labels.groups == GroupCloud), ]
xs < − MatFlow[which(is.na(Full.labels.groups)), ]
phull < − convhulln(ps, options = "Pp")
phull2 < − convhulln(rbind(ps, xs), options = "Pp")
nrp < − nrow(ps)
nrx < − nrow(xs)
outside < − unique(phull2[phull2 > nrp])-nrp
done < − FALSE
while(done){
phull3 < − convhulln(rbind(ps, xs[-(outside), ]), options = "Pp")
also.outside < − (1: nrx)[-outside][unique(phull3[phull3 > nrp])-nrp]
outside < − c(outside, also.outside)
done < − length(also.outside) == 0 }


   Rows.particles.inside < − (1: nrow(xs))[ -outside]
INFULL < − rep(NA,length = length(Full.labels.groups))
INFULL[which(is.na(Full.labels.groups))][Rows.particles.inside] < − GroupCloud
return(INFULL)}


   mysample < − function(x){
NoNAS < − which(is.na(x))
Take < − NA
if(length(NoNAS) > 1) {
x < − x[NoNAS]
Take < − sample(x, 1)
} else if (length(NoNAS) == 1) Take < − x[NoNAS]
return(Take)}


   for (i in 1: max(labels.groups)) New.labels.groups[, i] < − DetectInCloud(MatFlow,
Full.labels.groups, GroupCloud = i)
```

```
New.particles < − apply(New.labels.groups , 1, mysample)
Full.labels.groups[which(is.na(Full.labels.groups))] < −
New.particles[which(is.na(Full.labels.groups))]
return(Full.labels.groups) }


   Beads.function < − function (MatFlow, x, y, RegionBeads){
BeadsWindow < − which(RegionBeads[1]> y & RegionBeads[2] < y & Region-
Beads[3] > x & RegionBeads[4] < x)
x < − x[BeadsWindow]
y < − y[BeadsWindow]
BreaksX = seq(min(range(x)) - 0.005, max(range(x)) + 0.005, by = 0.005) BreaksY
= seq(min(range(y)) - 0.005, max(range(y)) + 0.005, by = 0.005) HistBeadsX < −
hist(x, breaks = BreaksX, plot = FALSE)
HistBeadsY < − hist(y, breaks = BreaksY, plot = FALSE)
Distribution maximum
MAXIMOX< − BreaksX[which(HistBeadsX$counts == max(HistBeadsX$counts))][1]
MAXIMOY < − BreaksY[which(HistBeadsY$counts == max(HistBeadsY$counts))][1]
Values 0.2 higher and lower of the maximum are taken
PositionX < − which(MAXIMOX - 0.2 < x & x < MAXIMOX + 0.2)
PositionY < − which(MAXIMOY - 0.2 < y & y < MAXIMOY + 0.2)
PositionBeads < − which(is.na(match(PositionX,PositionY)))
X < − x[PositionBeads]
Y < − y[PositionBeads]
Fit the distribution to a normal
AA < − fitdistr(X, "normal")
MeanSize < − coef(AA)[1]
StdError < − coef(AA)[2]
CI 95 %
LowerStdSize < − MeanSize - 1.96*StdError
UpperStdSize < − MeanSize + 1.96*StdError
BeadsCounts < − length(which(LowerStdSize < X & X < UpperStdSize))
BeadsX < − mean(exp(X[which(LowerStdSize < X & X < UpperStdSize)]))
```

BeadsY < − mean(exp(Y[which(LowerStdSize < X & X < UpperStdSize)]))
return(c(BeadsCounts,BeadsX,BeadsY))}

Makecluster function

Detection and clustering of different groups of particles within FCS files from Flow-cytometer

Detection and clustering of different groups of particles within FCS files from Flow-cytometer, returning a matrix with all cells labelled with the group number which they belong and the counts and main properties (counts, fluorescences) of the group used to correct the data.
FunctionMakecluster(FileName, NTestGroups = 10, column.names.fcs = c("FSC", "SSC", "FL1", "FL2", "FL3", "FL3-A", "FL3-W","Time"), RestricNGroups = FALSE, ScalingF1 = FALSE, Filtered = TRUE, MinimumToClean = 5000, CleanBeads = TRUE, AnalizeBeads = TRUE, AnalizeBeads = TRUE, RegionBeads = c(8.5, 6, 6.5, 3))

FileName: A character vector indicating the name of the fcs file to analyze

NTestGroups: Maximum Number of groups to try to make. Default is 10

column.names.fcs: A character vector with the column names of the FCS files. Default is c("FSC", "SSC", "FL1", "FL2", "FL3", "FL3-A", "FL3-W","Time")

Filtered: Whether to filter the FCS file to clean it. Default is TRUE

MinimumToClean: Minimum number of particles to clean the FCS. Default is 50000

CleanBeads: Whether to remove beads, changing different FL1 levels. Default is TRUE

AnalizeBeads: Whether to remove all particles except beads. Default is TRUE

RegionBeads: SSC and FL1 limits to define beads population x = SSC y = FL1 c(xmax,xmin,ymax,ymin). Default is c(8.5, 6, 6.5, 3)

This function is based on the following steps: Reads the extended FCS files (Reformat.fcs) and creates a matrix with a column with particle position in the original FCS file (MatFlow).
Filters the data to clean the FCS file. The filter is applied in three steps: first is in 3D (CleanMatrix3D) using 3 parameters (SSC, FL1, FL3) and in 2D (CleanMatrix) using 2 parameters (SSC, FL1; FL1, FL3). Makes the clustering on the filtered FCS file for the selected number of groups (flowClust) Extracts the best information criterio for each K in a flowClust object (BICf) Uses the information on the best clustering (Groups) and the original FCS file to recover the particles removed in the filtered step (recover.cell). Reads the FCS file again (Reformat.fcs) Analizes the beads (Beads.function) using the properties SSC and FL1.

FunctionMakecluster returns a list including:

- matrix with all the data stored in the original FCS file (FlowData)
- matrix filtered to be used as input of floClust function (FlowData.filtered)
- matrix with the fcs file filtered (flowCluts.res)
- flowClust output (Groups)
- a numeric vector with the group number which every cell belong (Full.labels.groups)
- number of groups detected (change.point)
- a numeric vector which includes the counts of beads, mean FL1 and SSC of beads(Beads)
- Name of the FCS file analyzed (FileName)

Required packages
library("flowClust")
library("geometry")
library("MASS")

Makecluster < − function(FileName, NTestGroups = 10, Maximum Number of groups to try to make
column.names.fcs = c("FSC", "SSC", "FL1", "FL2", "FL3", "FL3-A", "FL3-W","Time"), Filtered = TRUE, Whether to filter the FCS file to clean it
MinimumToClean = 5000, Minimum number of particles to clean the FCS
CleanBeads = TRUE, Whether to remove beads, TODO: change different FL1 levels
AnalizeBeads = TRUE,
RegionBeads = c(8.5, 6, 6.5, 3)), SSC and FL1 limits to define beads population
x = SSC y = FL1 c(xmax,xmin,ymax,ymin){
source(“Internals.R”) Load internal functions
Reads FCS file and restructures the matrix
FlowData < − Reformat.FCS(FileName, column.names.fcs)
MatFlow < − exprs(FlowData)
MatFlow < − as.data.frame(MatFlow)
MatFlow < − cbind(MatFlow, 1: nrow(MatFlow)) Create a column with particle positions in original FCS files
colnames(MatFlow) < − c(column.names.fcs, “OriginalRow”)
Applies the filtered if required if(Filtered & nrow(MatFlow) > MinimumToClean){
CleanMatFlow < − CleanMatrix3D(MatFlow, MatFlow$FL1, MatFlow$FL3, MatFlow$SSC, Threshold = 1, BreaksX = 40, BreaksY = 40, BreaksZ = 40)
CleanMatFlow < − CleanMatrix(CleanMatFlow, CleanMatFlow$SSC, CleanMatFlow$FL1, Threshold = 1, BreaksX = 80, BreaksY = 80)
CleanMatFlow < − CleanMatrix(CleanMatFlow, CleanMatFlow$FL1, CleanMatFlow$FL3, Threshold = 1, BreaksX = 80, BreaksY = 80)
OriginalRows < − CleanMatFlow$OriginalRow
CleanMatFlow < − CleanMatFlow[, -ncol(CleanMatFlow)]

colnames(CleanMatFlow) < − colnames(FlowData)
FlowData.filtered < − FlowData
exprs(FlowData.filtered) < − as.matrix(CleanMatFlow)
Makes N-clusterings on the filtered FCS file from 1 groups to N selected number of groups
flowClust.res < − flowClust(FlowData.filtered, varNames = c("SSC-H", "FL1-H", "FL3-H"), K = 1: NTestGroups, B = 1000, B.init = 100, tol = 1e-05, tol.init = 0.01, randomStart = 50)} else { FlowData.filtered < − NA

flowClust.res < − flowClust(FlowData, varNames = c("SSC-H", "FL1-H", "FL3-H"), K = 1: NTestGroups, B = 1000, B.init = 100, tol = 1e-05, tol.init = 0.01, randomStart = 50)}
Extracts the best information criterio for each K in a flowClust object
BIC < − BICf(flowClust.res)
Finds the best number of clusters based on BIC
change.point < − changepoint(BIC, NTestGroups)
Selects from the N flowclusts the one determined in the previous step
Groups < − flowClust.res[[change.point]]
returns results from the best clustering
If the FCS files was filtered recover the particles filtered and assign them to a group
if(Filtered & nrow(MatFlow) > MinimumToClean) Full.labels.groups < − recover.cell(MatFlow = MatFlow, Groups = Groups, OriginalRows = OriginalRows) else Full.labels.groups < − NA
analyze beads
if(AnalizeBeads){
FlowData2 < − Reformat.FCS(FileName, column.names.fcs, CleanBeads = FALSE)
MatFlow < − exprs(FlowData2)
MatFlow < − as.data.frame(MatFlow)
colnames(MatFlow) < − c(column.names.fcs)
Beads < − Beads.function(MatFlow = MatFlow, x = MatFlow$SSC, y = MatFlow$FL1,

RegionBeads = RegionBeads)}
else Beads< −NA
return(list(FlowData, FlowData.filtered, flowClust.res,
Groups, Full.labels.groups, change.point, Beads, FileName))}

Bibliografía

Aghaeepour, N., G. Finak, H. Hoos, T. R. Mosmann, R. Brinkman, R. Gottardo, and R. H. Scheuermann. 2013. Critical assessment of automated flow cytometry data analysis techniques. Nature Methods **10**: 228–238. 16, 29

Aiken, J., N. Rees, S. Hooker, P. Holligan, A. Bale, D. Robins, G. Moore, R. Harris, and D. Pilgrim. 2000. The Atlantic Meridional Transect: overview and synthesis of data. Progress In Oceanography **45**: 257–312. 80

Alonso-Sáez, L., V. Balague, E. L. Sa, O. Sanchez, J. M. González, J. Pinhassi, R. Massana, J. Pernthaler, C. Pedrós-Alió, and J. M. Gasol. 2007. Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH. Fems Microbiology Ecology **60**: 98–112. 37, 38, 102

Alonso-Sáez, L., L. Díaz-Pérez, and X. A. G. Morán. 2015. The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. Environ Microbiology . 38, 50, 52

Amann, R., and B. M. Fuchs. 2008. Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. Nature Reviews Microbiology **6**: 339–348. 37

Amann, R. I., W. Ludwig, and K. H. Schleifer. 1995. Phylogenetic Identification and In-situ Detection of Individual Microbial-cells Without Cultivation. Microbiological Reviews **59**: 143–169. 37

Andreatta, S., M. M. Wallinger, J. Piera, J. Catalan, R. Psenner, J. S. Hofer, and R. Sommaruga. 2004. Tools for discrimination and analysis of lake bacterioplankton subgroups measured by flow cytometry in a high-resolution depth profile. Aquatic Microbial Ecology **36**: 107–115. 16, 29

Andreatta, S., M. M. Wallinger, T. Posch, and R. Psenner. 2001. Detection of subgroups from flow cytometry measurements of heterotrophic bacterioplankton by image analysis. Cytometry **44**: 218–225. 20, 29

Apple, J. K., P. A. del Giorgi, and W. M. Kemp. 2006. Temperature regulation of bacterial production, respiration, and growth efficiency in a temperate salt-marsh estuary. Aquatic Microbial Ecology **43**: 243–254. 98, 100, 109

Aristegui, J., J. M. Gasol, C. M. Duarte, and G. J. Herndl. 2009. Microbial oceanography of the dark ocean's pelagic realm. Limnology and Oceanography **54**: 1501–1529. 5

Azam, F., T. Fenchel, J. G. Field, J. S. Gray, L. A. Meyerreil, and F. Thingstad. 1983. The Ecological Role of Water-column Microbes In the Sea. Marine Ecology Progress Series **10**: 257–263. 5, 15

Azam, F., and F. Malfatti. 2007. Microbial structuring of marine ecosystems (vol 5, pg 782-791, 2007). Nature Reviews Microbiology **5**: 966–U23. 9

Baltar, F., J. Aristegui, J. M. Gasol, S. Hernandez-Leon, and G. J. Herndl. 2007. Strong coast-ocean and surface-depth gradients in prokaryotic assemblage structure and activity in a coastal transition zone region. Aquatic Microbial Ecology **50**: 63–74. 102

Bashashati, A., and R. R. Brinkman. 2009. A survey of flow cytometry data analysis methods. Advances in bioinformatics pages 584–603. 16

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2014. Linear mixed-effects models using Eigen and S4. R package version 1.1-7, http://CRAN.R-project.org/package=lme4 . 64

Behrenfeld, M. J., and P. G. Falkowski. 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. Limnology And Oceanography **42**: 1–20. 79, 82, 84

Behrenfeld, M. J., E. Maranon, D. A. Siegel, and S. B. Hooker. 2002. Photoacclimation and nutrient-based model of light-saturated photosynthesis for quantifying oceanic primary production. Marine Ecology-Progress Series **228**: 103–117. 69, 79

Behrenfeld, M. J., R. T. O'Malley, D. A. Siegel, C. R. McClain, J. L. Sarmiento, G. C. Feldman, A. J. Milligan, P. G. Falkowski, R. M. Letelier, and E. S. Boss. 2006. Climate-driven trends in contemporary ocean productivity. Nature **444**: 752–755. 79, 80

Bolker, B. 2012. bbmle: Tools for general maximum likelihood estimation. R package version 1.0.4.1: http://CRAN.R-project.org/package=bbmle . 84

Bouvier, T., P. A. del Giorgio, and J. M. Gasol. 2007. A comparative study of the cytometric characteristics of High and Low nucleic-acid bacterioplankton cells from different aquatic ecosystems. Environmental Microbiology **9**: 2050–2066. 15, 30, 50

Boyce, D. G., M. R. Lewis, and B. Worm. 2010. Global phytoplankton decline over the past century. Nature **466**: 591–596. 8

Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West. 2004. Toward a metabolic theory of ecology. Ecology **85**: 1771–1789. 8, 59

Buitenhuis, E. T., W. K. W. Li, M. W. Lomas, D. M. Karl, M. R. Landry, and S. Jacquet. 2012. Picoheterotroph (Bacteria and Archaea) biomass distribution in the global ocean. Earth System Science Data **4**: 101–106. 6, 37, 80

Burkill, P. H., R. J. G. Leakey, N. J. P. Owens, and R. F. C. Mantoura. 1993. Synechococcus and Its Importance To the Microbial Foodweb of the Northwestern Indian-ocean. Deep-sea Research Part Ii-topical Studies In Oceanography **40**: 773–782. 63

Calvo-Díaz, A., and X. A. G. Morán. 2006. Seasonal dynamics of picoplankton in shelf waters of the southern Bay of Biscay. Aquatic Microbial Ecology **42**: 159–174. 15, 17, 30, 32, 40, 50, 63

Campbell, B. J., L. Yu, T. R. A. Straza, and D. L. Kirchman. 2009. Temporal changes in bacterial rRNA and rRNA genies in Delaware (USA) coastal waters. Aquatic Microbial Ecology **57**: 123–135. 38

Campbell, J., D. Antoine, R. Armstrong, K. Arrigo, W. Balch, R. Barber, M. Behrenfeld, R. Bidigare, J. Bishop, M. E. Carr, W. Esaias, P. Falkowski, N. Hoepffner, R. Iverson, D. Kiefer, S. Lohrenz, J. Marra, A. Morel, J. Ryan, V. Vedernikov, K. Waters, C. Yentsch, and J. Yoder. 2002. Comparison of algorithms for estimating ocean primary production from surface chlorophyll, temperature, and irradiance. Global Biogeochemical Cycles **16**: 1035. 79

Caporaso, J. G., K. Paszkiewicz, D. Field, R. Knight, and J. A. Gilbert. 2012. The Western English Channel contains a persistent microbial seed bank. Isme Journal **6**: 1089–1093. 50

Chow, C.-E. T., R. Sachdeva, J. A. Cram, J. A. Steele, D. M. Needham, A. Patel, A. E. Parada, and J. A. Fuhrman. 2013. Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. Isme Journal **7**: 2259–2273. 37, 38

Church, M. J., H. W. Ducklow, R. M. Letelier, and D. M. Karl. 2006. Temporal and vertical dynamics in picoplankton photoheterotrophic production in the subtropical North Pacific Ocean. Aquatic Microbial Ecology **45**: 41–53. 97

Cochlan, W. P., J. Wikner, G. F. Steward, D. C. Smith, and F. Azam. 1993. Spatial-distribution of Viruses, Bacteria and Chlorophyll-a In Neritic, Oceanic and Estuarine Environments. Marine Ecology Progress Series **92**: 77–87. 97

Cottrell, M. T., and D. L. Kirchman. 2000. Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. Applied and Environmental Microbiology **66**: 5116–5122. 37

del Giorgio, P., and P. J. Williams. 2009. Respiration in aquatic ecosystems. Oxford University Press, New York. 109

del Giorgio, P. A., and J. J. Cole. 1998. Bacterial growth efficiency in natural aquatic systems. Annual Review of Ecology and Systematics **29**: 503–541. 80

del Giorgio, P. A., R. Condon, T. Bouvier, K. Longnecker, C. Bouvier, E. Sherr, and J. M. Gasol. 2011. Coherent patterns in bacterial growth, growth efficiency, and leucine metabolism along a northeastern Pacific inshore-offshore transect. Limnology and Oceanography **56**: 1–16. 102

del Giorgio, P. A., and C. M. Duarte. 2002. Respiration in the open ocean. Nature **420**: 379–384. 8, 59, 79, 80, 99

DeLong, J. P., J. G. Okie, M. E. Moses, R. M. Sibly, and J. H. Brown. 2010. Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. Proceedings of the National Academy of Sciences of the United States of America **107**: 12941–12945. 59, 60, 64, 70, 71, 108

Díez, B., R. Massana, M. Estrada, and C. Pedrós-Alió. 2004. Distribution of eukaryotic picoplankton assemblages across hydrographic fronts in the Southern Ocean, studied by denaturing gradient gel electrophoresis. Limnology and Oceanography **49**: 1022–1034. 102

Dodds, P. S., D. H. Rothman, and J. S. Weitz. 2001. Re-examination of the '"3/4-law.ºf metabolism. Journal of Theoretical Biology **209**: 9–27. 60

Doolittle, R. F. 2002. Biodiversity: Microbial genomes multiply. Nature **416**: 697–700. 70

Doolittle, W. F., and R. T. Papke. 2006. Genomics and the bacterial species problem. Genome Biology **7**: 116. 59

Duarte, C. M., and S. Agusti. 1998. The CO2 balance of unproductive aquatic ecosystems. Science **281**: 234–236. 79, 99

Ducklow, H. 2000. Bacterial production and biomass in the oceans. In D. L. Kirchman (ed.), Microbial Ecology of the Oceans. 1st edn. pp. 85-120, Wiley-Liss, Inc., New York, N.Y. 8

Ducklow, H., and C. Carlson, 1992. Oceanic Bacterial Production. Pages 113–181 *in* K. Marshall, editor. Advances in Microbial Ecology, volume 12. Springer US. 7, 81, 97

Ducklow, H., C. Carlson, and W. Smith. 1999. Bacterial growth in experimental plankton assemblages and seawater cultures from the Phaeocystis antarctica bloom in the Ross Sea, Antarctica. Aquatic Microbial Ecology **19**: 215–227. 81

Ducklow, H. W. 1995. Ocean Biogeochemical Fluxes - New Production and Export of Organic-matter From the Upper Ocean. Reviews of Geophysics **33**: 1271–1276. 79

Ducklow, H. W., and R. P. Harris. 1993. Introduction to the {JGOFS} North Atlantic bloom experiment. Deep Sea Research Part II: Topical Studies in Oceanography **40**: 1 – 8. 80

Ducklow, H. W., D. K. Steinberg, and K. O. Buesseler. 2001. Upper ocean carbon export and the biological pump. Special Issue-JGOFS. Oceanography . 5

Edwards, K. F., M. K. Thomas, C. A. Klausmeier, and E. Litchman. 2012. Allometric scaling and taxonomic variation in nutrient utilization traits and maximum growth rate of phytoplankton. Limnology and Oceanography **57**: 554–566. 59, 71

Enquist, B. J., E. P. Economo, T. E. Huxman, A. P. Allen, D. D. Ignace, and J. F. Gillooly. 2003. Scaling metabolism from organisms to ecosystems. Nature **423**: 639–642. 59

Eppley, R. W., E. Steward, M. R. Abbott, and U. Heyman. 1985. Estimating Ocean Primary Production From Satellite Chlorophyll - Introduction To Regional Differences and Statistics For the Southern-california Bight. Journal of Plankton Research **7**: 57–70. 79

Falkowski, P. G., T. Fenchel, and E. F. Delong. 2008. The microbial engines that drive Earth's biogeochemical cycles. Science **320**: 1034–1039. 5

Fasham, M. J. R., P. W. Boyd, and G. Savidge. 1999. Modeling the relative contributions of autotrophs and heterotrophs to carbon flow at a Lagrangian JGOFS station in the Northeast Atlantic: The importance of DOC. Limnology and Oceanography **44**: 80–94. 80

Felip, M., S. Andreatta, R. Sommaruga, V. Straskrabova, and J. Catalan. 2007. Suitability of flow cytometry for estimating bacterial biovolume in natural plankton samples: Comparison with microscopy data. Applied and Environmental Microbiology **73**: 4508–4514. 15

Fenchel, T., and B. J. Finlay. 2004. The ubiquity of small species: Patterns of local and global diversity. Bioscience **54**: 777–784. 5

Finak, G., A. Bashashati, R. Brinkman, and R. Gottardo. 2009. Merging mixture components for cell population identification in flow cytometry. Advances in bioinformatics pages 247646–247646. 16

Finak, G., J. M. Perez, A. Weng, and R. Gottardo. 2010. Optimizing transformations for automated, high throughput analysis of flow cytometry data. Bmc Bioinformatics **11**: 546. 29

Finkel, Z. V., J. Beardall, K. J. Flynn, A. Quigg, T. A. V. Rees, and J. A. Raven. 2010. Phytoplankton in a changing world: cell size and elemental stoichiometry. Journal of Plankton Research **32**: 119–137. 59, 60, 71

Fuhrman, J. A., J. A. Cram, and D. M. Needham. 2015. Marine microbial community dynamics and their ecological interpretation. Nature reviews. Microbiology **13**: 133–46. 52

Fuhrman, J. A., I. Hewson, M. S. Schwalbach, J. A. Steele, M. V. Brown, and S. Naeem. 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proceedings of the National Academy of Sciences of the United States of America **103**: 13104–13109. 49, 50

Fuhrman, J. A., K. Mccallum, and A. A. Davis. 1993. Phylogenetic Diversity of Subsurface Marine Microbial Communities From the Atlantic and Pacific Oceans. Applied and Environmental Microbiology **59**: 1294–1302. 70

Fuhrman, J. A., J. A. Steele, I. Hewson, M. S. Schwalbach, M. V. Brown, J. L. Green, and J. H. Brown. 2008. A latitudinal diversity gradient in planktonic marine bacteria. Proceedings of the National Academy of Sciences of the United States of America **105**: 7774–7778. 7, 37

Fukuda, R., H. Ogawa, T. Nagata, and I. Koike. 1998. Direct determination of carbon and nitrogen contents of natural bacterial assemblages in marine environments. Applied and Environmental Microbiology **64**: 3352–3358. 5

García, F. C., A. López-Urrutia, and X. A. Morán. 2014. Automated clustering of heterotrophic bacterioplankton in flow-cytometry data. Aquatic Microbial Ecology **72**: 175–185. 40

García-Martín, E. E., S. McNeill, P. Serret, and R. J. G. Leakey. 2014. Plankton metabolism and bacterial growth efficiency in offshore waters along a latitudinal transect between the UK and Svalbard. Deep-sea Research Part I-oceanographic Research Papers **92**: 141–151. 8, 99

Gasol, J. M. 1994. A Framework For the Assessment of Top-down Vs Bottom-up Control of Heterotrophic Nanoflagellate Abundance. Marine Ecology-progress Series **113**: 291–300. 7

Gasol, J. M., and P. A. Del Giorgio. 2000. Using flow cytometry for counting natural planktonic bacteria and understanding the structure of planktonic bacterial communities. Scientia Marina **64**: 197–224. 7, 15, 31, 40, 41, 50

Gasol, J. M., P. A. del Giorgio, and C. M. Duarte. 1997. Biomass distribution in marine planktonic communities. Limnology and Oceanography **42**: 1353–1363. 5, 7

Gasol, J. M., and C. M. Duarte. 2000. Comparative analyses in aquatic microbial ecology: how far do they go? Fems Microbiology Ecology **31**: 99–106. 80, 81, 97

Gasol, J. M., J. Pinhassi, L. Alonso-Saez, H. Ducklow, G. J. Herndl, M. Koblizek, M. Labrenz, Y. Luo, X. A. G. Moran, T. Reinthaler, and M. Simon. 2008. Towards a better understanding of microbial carbon flux in the sea. Aquatic Microbial Ecology **53**: 21–38. 80

Gasol, J. M., U. L. Zweifel, F. Peters, J. A. Fuhrman, and A. Hagstrom. 1999. Significance of size and nucleic acid content heterogeneity as measured by

flow cytometry in natural planktonic bacteria. Applied and Environmental Microbiology **65**: 4475–4483. 15, 50

Ge, Y., and S. C. Sealfon. 2012. flowPeaks: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding. Bioinformatics **28**: 2052–2058. 30

Gelman, A., and J. Hill. 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. 1 edition. Cambridge University Press, New York, USA. 63

Gilbert, J., P. Somerfield, B. Temperton, S. Huse, and I. Joint. 2010. Day-length is central to maintaining consistent seasonal diversity in marine bacterio-plankton. Available from Nature Precedings, http://hdl.handle.net/10101/npre. 2010.4406.1. . 51

Gilbert, J. A., D. Field, P. Swift, L. Newbold, A. Oliver, T. Smyth, P. J. Somerfield, S. Huse, and I. Joint. 2009. The seasonal structure of microbial communities in the Western English Channel. Environmental Microbiology **11**: 3132–3139. 59

Gilbert, J. A., J. A. Steele, J. G. Caporaso, L. Steinbrueck, J. Reeder, B. Temperton, S. Huse, A. C. McHardy, R. Knight, I. Joint, P. Somerfield, J. A. Fuhrman, and D. Field. 2012. Defining seasonal marine microbial community dynamics. Isme Journal **6**: 298–308. 38, 49, 50

Gillooly, J. F., E. L. Charnov, G. B. West, V. M. Savage, and J. H. Brown. 2002. Effects of size and temperature on developmental time. Nature **417**: 70–73. 64

Giovannoni, S. J., T. B. Britschgi, C. L. Moyer, and K. G. Field. 1990. Genetic Diversity In Sargasso Sea Bacterioplankton. Nature **345**: 60–63. 37

Giovannoni, S. J., and M. S. Rappe. 2000. Evolution, diversity and molecular ecology of marine prokaryotes. pp. 47-84, Wiley-Liss, Inc., New York, N.Y. 5, 37, 48

Giovannoni, S. J., M. S. Rappe, K. L. Vergin, and N. L. Adair. 1996. 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the Green Non-Sulfur bacteria. Proceedings of the National Academy of Sciences of the United States of America **93**: 7979–7984. 59

Giovannoni, S. J., and U. Stingl. 2005. Molecular diversity and ecology of microbial plankton. Nature **437**: 343–348. 5, 37, 59, 70

Giovannoni, S. J., and K. L. Vergin. 2012. Seasonality in Ocean Microbial Communities. Science **335**: 671–676. 49

Grasshoff, K. 1976. Methods of sea-water analysis. Verlag Chemie. 40

Green, J. L., B. J. M. Bohannan, and R. J. Whitaker. 2008. Microbial biogeography: From taxonomy to traits. Science **320**: 1039–1043. 37

Hahne, F., N. LeMeur, R. R. Brinkman, B. Ellis, P. Haaland, D. Sarkar, J. Spidlen, E. Strain, and R. Gentleman. 2009. flowCore: a Bioconductor package for high throughput flow cytometry. Bmc Bioinformatics **10**: 106. 19, 63

Hansell, D. A., and C. A. Carlson. 1998. Deep-ocean gradients in the concentration of dissolved organic carbon. Nature **395**: 263–266. 15

Hansell, D. A., C. A. Carlson, D. J. Repeta, and R. Schlitzer. 2009. Dissolved Organic Matter In the Ocean A Controversy Stimulates New Insights. Oceanography **22**: 202–211. 109

Hanson, R., H. Ducklow, and J. Field. 2000. The changing ocean carbon cycle, a midterm synthesis of Joint Global Ocean Flux Study. Cambridge University Press, Cambridge. 80

Hemmingsen, A. 1960. Energy metabolism as related to body size and respiratory surfaces, and its evolution. Rep Steno Mem Hosp Nord Insulinlab **9**: 1–110. 59

Herlemann, D. P. R., M. Labrenz, K. Juergens, S. Bertilsson, J. J. Waniek, and A. F. Andersson. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. Isme Journal **5**: 1571–1579. 42

Huete-Ortega, M., P. Cermeno, A. Calvo-Diaz, and E. Maranon. 2012. Isometric size-scaling of metabolic rate and the size abundance distribution of phytoplankton. Proceedings of the Royal Society B-biological Sciences **279**: 1815–1823. 59

Huse, S. M., D. M. Welch, H. G. Morrison, and M. L. Sogin. 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environmental Microbiology **12**: 1889–1898. 42

Islas, S., A. Becerra, P. L. Luisi, and A. Lazcano. 2004. Comparative genomics and the gene complement of a minimal cell. Origins of Life and Evolution of the Biosphere **34**: 243–256. 70

Jiao, N., G. J. Herndl, D. A. Hansell, R. Benner, G. Kattner, S. W. Wilhelm, D. L. Kirchman, M. G. Weinbauer, T. Luo, F. Chen, and F. Azam. 2010. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. Nature Reviews Microbiology **8**: 593–599. 5, 6, 109

Kirchman, D. 2000. Microbial ecology of the oceans 2nd edition. 620 pp. 2nd edition. Wiley-Liss, Inc., New York, N.Y. 5

Kirchman, D. L., R. R. Malmstrom, and M. T. Cottrell. 2005. Control of bacterial growth by temperature and organic matter in the Western Arctic. Deep-sea Research Part Ii-topical Studies In Oceanography **52**: 3386–3395. 8, 81, 109

Kleiber, M. 1932. Body size and metabolism. Hilgardia **6**: 315–353. 60

Kohonen, T. 1997. Self-Organizing Maps. Springer, Berlin. 86

Lahesmaa-Korpinen, A. M., E. Jalkanen, S, P. Chen, E. Valo, J. Núñez-Fontarnau, V. Rantanen, A. Oghabian, J. Vakkila, K. Porkka, S. Mustjoki, and S. Hautaniemi. 2011. FlowAnd: Comprehensive Computational Framework for Flow Cytometry Data Analysis. Proteomics Bioinformatics **4**: 11. 16

Latasa, M., R. Scharek, M. Vidal, G. Vila-Reixach, A. Gutiérrez-Rodriguez, M. Emelianov, and J. M. Gasol. 2010. Preferences of phytoplankton groups for waters of different trophic status in the northwestern Mediterranean Sea. Marine Ecology Progress Series **407**: 27–42. 102

Le Meur, N. 2013. Computational methods for evaluation of cell-based data assessment - Bioconductor. Curr Opin Biotechnol **24**: 105–111. 29

Li, W. K. W. 1997. Cytometric diversity in marine ultraphytoplankton. Limnology and Oceanography **42**: 874–880. 6, 7, 38, 40, 49, 52, 108

Li, W. K. W. 1998. Annual average abundance of heterotrophic bacteria and Synechococcus in surface ocean waters. Limnology and Oceanography **43**: 1746–1753. 7, 80, 97

Li, W. K. W. 2002. Macroecological patterns of phytoplankton in the northwestern North Atlantic Ocean. Nature **419**: 154–157. 49

Li, W. K. W. 2009. From cytometry to macroecology: a quarter century quest in microbial oceanography. Aquatic Microbial Ecology **57**: 239–251. 6, 80, 102

Li, W. K. W., and W. G. Harrison. 2001. Chlorophyll, bacteria and picophytoplankton in ecological provinces of the North Atlantic. Deep-sea Research Part Ii-topical Studies In Oceanography **48**: 2271–2293. 80, 97, 98

Li, W. K. W., W. G. Harrison, and E. J. H. Head. 2006. Coherent sign switching in multiyear trends of microbial plankton. Science **311**: 1157–1160. 7

Li, W. K. W., E. J. H. Head, and W. G. Harrison. 2004. Macroecological limits of heterotrophic bacterial abundance in the ocean. Deep-sea Research Part I-oceanographic Research Papers **51**: 1529–1540. 7, 80, 81, 98, 102

Li, W. K. W., J. F. Jellett, and P. M. Dickie. 1995. DNA distributions in planktonic bacteria stained with TOTO or TO-PRO. Limnology and Oceanography **40**: 1485–1495. 15, 50

Litchman, E., C. A. Klausmeier, O. M. Schofield, and P. G. Falkowski. 2007. The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level. Ecology Letters **10**: 1170–1181. 59, 71

Lo, K., R. R. Brinkman, and R. Gottardo. 2008. Automated Gating of flow cytometry data via robust model-based clustering. Cytometry Part A **73A**: Int Soc Analyt Cytol. 19

Lo, K., F. Hahne, R. R. Brinkman, and R. Gottardo. 2009. flowClust: a Bioconductor package for automated gating of flow cytometry data. Bmc Bioinformatics **10**: 145. 16, 19, 20, 29

Longhurst, A., S. Sathyendranath, T. Platt, and C. Caverhill. 1995. An Estimate of Global Primary Production In the Ocean From Satellite Radiometer Data. Journal of Plankton Research **17**: 1245–1271. 86, 90

López-Urrutia, A., and X. Morán. 2007. Resource limitation of bacterial production distorts the temperature dependence of oceanic carbon cycling. Ecology **88**: 817–822. 8, 81, 87, 99, 109

López-Urrutia, A., E. San Martín, R. P. Harris, and X. Irigoien. 2006. Scaling the metabolic balance of the oceans. Proceedings Of The National Academy Of Sciences Of The United States Of America **103**: 8739–8744. 8, 60, 71, 81

Luta, G. 2011. On Extensions of k-Means Clustering for Automated Gating of Flow Cytometry Data. Cytometry Part A **79A**: 3–5. 29

Makarieva, A. M., V. G. Gorshkov, and B. L. Li. 2005. Energetics of the smallest: do bacteria breathe at the same rate as whales? Proceedings of the Royal Society B-biological Sciences **272**: 2219–2224. 59, 60, 64, 108

Maldonado, F., T. T. Packard, and M. Gómez. 2012. Understanding tetrazolium reduction and the importance of substrates in measuring respiratory electron transport activity. Journal of Experimental Marine Biology and Ecology **434**: 110–118. 61

Marañón, E., P. Cermeno, D. C. López-Sandoval, T. Rodríguez-Ramos, C. Sobrino, M. Huete-Ortega, J. M. Blanco, and J. Rodríguez. 2013. Unimodal size scaling of phytoplankton growth and the size dependence of nutrient uptake and use. Ecology Letters **16**: 371–379. 59, 60

Marañón, E., P. Cermeno, J. Rodríguez, M. V. Zubkov, and R. P. Harris. 2007. Scaling of phytoplankton photosynthesis and cell size in the ocean. Limnology and Oceanography **52**: 2190–2198. 60

Margalef, R. 1968. Perspectives in ecological theory. Univ. Chicago. Press, Chicago IL. 111p. 38

Marie, D., F. Partensky, S. Jacquet, and D. Vaulot. 1997. Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. Applied and Environmental Microbiology **63**: 186–193. 17, 31

Martínez-García, S., E. Fernández, M. Aranguren-Gassis, and E. Teira. 2009. In vivo electron transport system activity: a method to estimate respiration in natural marine microbial planktonic communities. Limnology and Oceanography-methods **7**: 459–469. 61, 62

Mary, I., J. L. Heywood, B. M. Fuchs, R. Amann, G. A. Tarran, P. H. Burkill, and M. V. Zubkov. 2006. SAR11 dominance among metabolically active low nucleic acid bacterioplankton in surface waters along an Atlantic meridional transect. Aquatic Microbial Ecology **45**: 107–113. 50

Massana, R., and R. Logares. 2013. Eukaryotic versus prokaryotic marine picoplankton ecology. Environmental Microbiology **15**: 1254–1261. 59

Mazuecos, I. P., J. Aristegui, E. Vazquez-Dominguez, E. Ortega-Retuerta, J. M. Gasol, and I. Reche. 2015. Temperature control of microbial respiration and growth efficiency in the mesopelagic zone of the South Atlantic and Indian Oceans. Deep-sea Research Part I-oceanographic Research Papers **95**: 131–138. 8

Morán, X. A. G., and A. Calvo-Díaz. 2009. Single-cell vs. bulk activity properties of coastal bacterioplankton over an annual cycle in a temperate ecosystem. Fems Microbiology Ecology **67**: 43–56. 30, 32

Moran, X. A. G., M. Estrada, J. M. Gasol, and C. Pedros-Alio. 2002. Dissolved primary production and the strength of phytoplankton bacterioplankton coupling in contrasting marine regions. Microbial Ecology **44**: 217–223. 80, 81

Morel, A., and D. Antoine. 2002. Oceanography - Small critters - Big effects. Science **296**: 1980–1982. 79

Morris, R. M., K. L. Vergin, J. C. Cho, M. S. Rappe, C. A. Carlson, and S. J. Giovannoni. 2005. Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. Limnology and Oceanography **50**: 1687–1696. 37, 52, 59

Nagata, T. 2000. Production mechanisms of dissolved organic matter. In: Kirchman, D.L. (Ed.), Microbial Ecology of the Oceans. pp. 121–152., Wiley-Liss, NewYork. 80

Naumann, U., G. Luta, and M. P. Wand. 2010. The curvHDR method for gating flow cytometry samples. Bmc Bioinformatics **11**: 44. 30

Naumann, U., and M. P. Wand. 2009. Automation in High-Content Flow Cytometry Screening. Cytometry Part A **75A**: 789–797. 30

Oksanen, J., F. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2012. vegan: Community Ecology Package. R package version 2.0-3. R Foundation for Statistical Computing, Vienna, Austria. http://CRAN.R-project.org/package=vegan . 43

Olsen, G. J., D. J. Lane, S. J. Giovannoni, N. R. Pace, and D. A. Stahl. 1986. Microbial Ecology and Evolution - A Ribosomal-rna Approach. Annual Review of Microbiology **40**: 337–365. 37

Pedrós-Alió, C., D. Vaqué, N. Guixa-Boixereu, and J. M. Gasol. 2002. Prokaryotic plankton biomass and heterotrophic production in western Antarctic waters during the 1995-1996 Austral summer. Deep-sea Research Part Ii-topical Studies In Oceanography **49**: 805–825. 102

Pernthaler, A., J. Pernthaler, and R. Amann. 2002. Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. Applied and Environmental Microbiology **68**: 3094–3101. 37

Peters, R. H. 1991. A critique for ecology. Cambridge Univ. Press. 23

Pielou, E. 1975. Ecological diversity. Wiley-Interscience. 37

Platt, T., and S. Sathyendranah. 1995. Software for Use in Calculation of Primary Production in the Oceanic Water Column. Available at http://ioccg/software/ocean$_p$*roduction*/*index.html*. 79, 80, 85

Platt, T., S. Sathyendranath, C. M. Caverhill, and M. Lewis. 1988. Ocean Primary Production and Available Light - Further Algorithms For Remote-sensing. Deep-sea Research Part A-oceanographic Research Papers **35**: 855–879. 79, 80, 84

Pomeroy, L. R., and W. J. Wiebe. 2001. Temperature and substrates as interactive limiting factors for marine heterotrophic bacteria. Aquatic Microbial Ecology **23**: 187–204. 8, 81

Pommier, T., B. Canback, L. Riemann, K. H. Bostrom, K. Simu, P. Lundberg, A. Tunlid, and A. Hagstrom. 2007. Global patterns of diversity and community structure in marine bacterioplankton. Molecular Ecology **16**: 867–880. 59

Price, N. D., J. L. Reed, and B. O. Palsson. 2004. Genome-scale models of microbial cells: Evaluating the consequences of constraints. Nature Reviews Microbiology **2**: 886–897. 70

Rajwa, B., M. Venkatapathi, K. Ragheb, P. P. Banada, E. D. Hirleman, T. Lary, and J. P. Robinson. 2008. Automated classification of bacterial particles in flow by multiangle scatter measurement and support vector machine classifier. Cytometry Part A **73A**: 369–379. 16

Regaudie-de Gioux, A., and C. M. Duarte. 2013. Global patterns in oceanic planktonic metabolism. Limnology and Oceanography **58**: 977–986. 8

Ribalet, F., D. M. Schruth, and E. V. Armbrust. 2011. flowPhyto: enabling automated analysis of microscopic algae from continuous flow cytometric data. Bioinformatics **27**: 732–733. 20, 29

Richardson, A. J., M. C. Pfaff, J. G. Field, N. F. Silulwane, and F. A. Shillington. 2002. Identifying characteristic chlorophyll a profiles in the coastal domain using an artificial neural network. Journal of Plankton Research **24**: 1289–1303. 80, 81, 97

Richardson, A. J., N. F. Silulwane, B. A. Mitchell-Innes, and F. A. Shillington. 2003. A dynamic quantitative approach for predicting the shape of phytoplankton profiles in the ocean. Progress In Oceanography **59**: 301–319. 80, 81

Rivkin, R. B., and L. Legendre. 2001. Biogenic carbon cycling in the upper ocean: Effects of microbial respiration. Science **291**: 2398–2400. 99

Robinson, C., A. J. Poulton, P. M. Holligan, A. R. Baker, G. Forster, N. Gist, T. D. Jickells, G. Malin, R. Upstill-Goddard, R. G. Williams, E. M. S. Woodward, and M. V. Zubkov. 2006. The atlantic Meridional Transect (CMT) Programme: A contextual view 1995-2005. Deep-sea Research Part Ii-topical Studies In Oceanography **53**: 1485–1515. 80

Robinson, C., and P. J. Williams. 2005. Respiration and its measurement in surface marine waters, in: Respiration in aquatic ecosystems. Oxford Univ. Press. 8, 59, 79, 99

Robinson, C., and P. J. L. Williams. 1993. Temperature and Antarctic Plankton Community Respiration. Journal of Plankton Research **15**: 1035–1051. 8

Robinson, J. P., B. Rajwa, V. Patsekin, and V. Davisson. 2012. Computational analyses of highthroughput flow cytometry data . 29

Salter, I., P. E. Galand, S. K. Fagervold, P. Lebaron, I. Obernosterer, M. J. Oliver, M. T. Suzuki, and C. Tricoire. 2014. Seasonal dynamics of active SAR11 ecotypes in the oligotrophic Northwest Mediterranean Sea. The ISME Journal pages 1–14. 37

Salter, I., P. E. Galand, S. K. Fagervold, P. Lebaron, I. Obernosterer, M. J. Oliver, M. T. Suzuki, and C. Tricoire. 2015. Seasonal dynamics of active SAR11 ecotypes in the oligotrophic Northwest Mediterranean Sea. The ISME journal **9**: 347–60. 52

Sarmiento, J. L., R. Slater, R. Barber, L. Bopp, S. C. Doney, A. C. Hirst, J. Kleypas, R. Matear, U. Mikolajewicz, P. Monfray, V. Soldatov, S. A. Spall, and R. Stouffer. 2004. Response of ocean ecosystems to climate warming. Global Biogeochemical Cycles **18**: GB3003. 8

Sathyendranath, S., A. Longhurst, C. M. Caverhill, and T. Platt. 1995. Regionally and seasonally differentiated primary production in the North Atlantic. Deep-sea Research Part I-oceanographic Research Papers **42**: 1773–1802. 80, 84

Sathyendranath, S., T. Platt, E. P. W. Horne, W. G. Harrison, O. Ulloa, R. Outerbridge, and N. Hoepffner. 1991. Estimation of New Production In the Ocean By Compound Remote-sensing. Nature **353**: 129–133. 80

Schattenhofer, M., B. M. Fuchs, R. Amann, M. V. Zubkov, G. A. Tarran, and J. Pernthaler. 2009. Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. Environmental Microbiology **11**: 2078–2093. 37

Schattenhofer, M., J. Wulf, I. Kostadinov, F. O. Gloeckner, M. V. Zubkov, and B. M. Fuchs. 2011. Phylogenetic characterisation of picoplanktonic populations with high and low nucleic acid content in the North Atlantic Ocean. Systematic and Applied Microbiology **34**: 470–475. 15, 50

Schauer, M., R. Massana, and C. Pedros-Alio. 2000. Spatial differences in bacterioplankton composition along the Catalan coast (NW Mediterranean) assessed by molecular fingerprinting. Fems Microbiology Ecology **33**: 51–59. 37

Scheuermann, R., Y. Quian, C. Wei, and I. Sanz. 2009. ImmPort FLOCK: Automated cell population identification in high dimensional flow cytometry data. The Journal of Immunology **182(Meeting Abstracts 1)**: 42–17. 16

Schiaffino, R. M., J. M. Gasol, I. Izaguirre, and F. Unrein. 2013. Picoplankton abundance and cytometric group diversity along a trophic and latitudinal lake gradient. Aquatic Microbial Ecology **68**: 231–250. 38, 49

Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, and C. F. Weber. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology **75**: 7537–7541. 42

Scofield, V., S. M. S. Jacques, J. R. D. Guimaraes, and V. F. Farjalla. 2015. Potential changes in bacterial metabolism associated with increased water temperature and nutrient inputs in tropical humic lagoons. Frontiers In Microbiology **6**: 310. 8

Shapiro, H. 1995. Practical flow cytometry. Third Edition. Wiley-Liss. 32

Shuter, B. J., J. E. Thomas, W. D. Taylor, and A. M. Zimmerman. 1983. Phenotypic Correlates of Genomic Dna Content In Unicellular Eukaryotes and Other Cells. American Naturalist **122**: 26–44. 70

Smith, R. C., R. W. Eppley, and K. S. Baker. 1982. Correlation of Primary Production As Measured Aboard Ship In Southern-california Coastal Waters and As Estimated From Satellite Chlorophyll Images. Marine Biology **66**: 281–288. 79

Smyth, T. J., I. Allen, A. Atkinson, J. T. Bruun, R. A. Harmer, R. D. Pingree, C. E. Widdicombe, and P. J. Somerfield. 2014. Ocean Net Heat Flux Influences Seasonal to Interannual Patterns of Plankton Abundance. PLoS ONE **9**: e98709. 50, 51

Sogin, M. L., H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl. 2006. Microbial diversity in the deep sea and the underexplored rare biosphere". Proceedings of the National Academy of Sciences of the United States of America **103**: 12115–12120. 37

Stepkowski, T., and A. B. Legocki. 2001. Reduction of bacterial genome size and expansion resulting from obligate intracellular lifestyle and adaptation to soil habitat. Acta Biochimica Polonica **48**: 367–381. 70

Suess, E. 1980. Particulate Organic-carbon Flux In the Oceans - Surface Productivity and Oxygen Utilization. Nature **288**: 260–263. 79

Sugar, I. P., and S. C. Sealfon. 2010. Misty Mountain clustering: application to fast unsupervised flow cytometry gating. Bmc Bioinformatics **11**: 502. 30

Tanaka, N., M. Hirahata, S. Miyazaki, and H. Sugawara. 2003. The status quo of microbial genomic data available in the public domain: archaea and bacteria. World Fed Cult Collect Lett **36**: 13–20. 70

Team, R. C. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://CRAN.R-project.org/package=vegan . 43, 88

Team, R. C., 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 64

Vázquez-Domínguez, E., C. M. Duarte, S. Agustí, K. Juergens, D. Vaqué, and J. M. Gasol. 2008. Microbial plankton abundance and heterotrophic activity across the Central Atlantic Ocean. Progress In Oceanography **79**: 83–94. 102

Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Y. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304**: 66–74. 37

Vergin, K. L., B. Done, C. A. Carlson, and S. J. Giovannoni. 2013. Spatiotemporal distributions of rare bacterioplankton populations indicate adaptive strategies in the oligotrophic ocean. Aquatic Microbial Ecology **71**: 1–U129. 37, 52

Vila-Costa, M., J. M. Gasol, S. Sharma, and M. A. Moran. 2012. Community analysis of high- and low-nucleic acid-containing bacteria in NW Mediterranean coastal waters using 16S rDNA pyrosequencing. Environmental Microbiology **14**: 1390–1402. 15, 50

Wang, K., M. Li, and M. Bucan. 2007. Pathway-based approaches for analysis of genomewide association studies. American Journal of Human Genetics **81**: 1278–1283. 42

Warton, D. I., I. J. Wright, D. S. Falster, and M. Westoby. 2006. Bivariate line-fitting methods for allometry. Biological Reviews **81**: 259–291. 64

Wehrens, R., and L. M. C. Buydens. 2007. Self- and super-organizing maps in R: The kohonen package. Journal of Statistical Software **21**: 1–19. 86

West, G. B., J. H. Brown, and B. J. Enquist. 1997. A general model for the origin of allometric scaling laws in biology. Science **276**: 122–126. 59, 60

Westberry, T., M. J. Behrenfeld, D. A. Siegel, and E. Boss. 2008. Carbon-based primary productivity modeling with vertically resolved photoacclimation. Global Biogeochemical Cycles **22**: GB2024. 79, 80

White, P. A., J. Kalff, J. B. Rasmussen, and J. M. Gasol. 1991. The Effect of Temperature and Algal Biomass On Bacterial Production and Specific Growth-rate In Fresh-water and Marine Habitats. Microbial Ecology **21**: 99–118. 8

Whitman, W. B., D. C. Coleman, and W. J. Wiebe. 1998. Prokaryotes: The unseen majority. Proceedings of the National Academy of Sciences of the United States of America **95**: 6578–6583. 59, 70, 99

Williams, P. J. L. 1981. Microbial Contribution To Overall Marine Plankton Metabolism - Direct Measurements of Respiration. Oceanologica Acta **4**: 359–364. 59, 79

Williams, P. J. L. 1998. The balance of plankton respiration and photosynthesis in the open oceans. Nature **394**: 55–57. 99

Wood, S. 2015. Package'mgcv'. Available: http://cran.r-project.org/web/packages/mgcv/mgcv.pdf . 87

Yvon-Durocher, G., J. I. Jones, M. Trimmer, G. Woodward, and J. M. Montoya. 2010. Warming alters the metabolic balance of ecosystems. Philosophical Transactions of the Royal Society B-biological Sciences **365**: 2117–2126. 8

Zare, H., P. Shooshtari, A. Gupta, and R. R. Brinkman. 2010. Data reduction for spectral clustering to analyze high troughput flow cytometry data. BMC Bioinformatics **11**: 403. 20, 30

Zinger, L., L. A. Amaral-Zettler, J. A. Fuhrman, M. C. Horner-Devine, S. M. Huse, D. B. M. Welch, J. B. H. Martiny, M. Sogin, A. Boetius, and A. Ramette. 2011. Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems. Plos One **6**: e24570. 37

Zinger, L., A. Gobet, and T. Pommier. 2012. Two decades of describing the unseen majority of aquatic microbial diversity. Molecular Ecology **21**: 1878–1896. 37

Zubkov, M. V. 2014. Faster growth of the major prokaryotic versus eukaryotic CO2 fixers in the oligotrophic ocean. Nature Communications **5**: 3776. 60, 71

Zubkov, M. V., B. M. Fuchs, P. H. Burkill, and R. Amann. 2001. Comparison of cellular and biomass specific activities of dominant bacterioplankton groups in stratified waters of the Celtic Sea. Applied and Environmental Microbiology **67**: 5210–5218. 50

Zubkov, M. V., M. A. Sleigh, and P. H. Burkill. 2000*a*. Assaying picoplankton distribution by flow cytometry of underway samples collected along a meridional transect across the Atlantic Ocean. Aquatic Microbial Ecology **21**: 13–20. 71

Zubkov, M. V., M. A. Sleigh, P. H. Burkill, and R. J. G. Leakey. 2000*b*. Picoplankton community structure on the Atlantic Meridional Transect: a comparison between seasons. Progress In Oceanography **45**: 369–386. 80

Zubkov, M. V., M. A. Sleigh, G. A. Tarran, P. H. Burkill, and R. J. G. Leakey. 1998. Picoplanktonic community structure on an Atlantic transect from 50 degrees N to 50 degrees S. Deep-sea Research Part I-oceanographic Research Papers **45**: 1339–1355. 63, 71, 97

Esta tesis ha sido financiada por el Ministerio de Economía y competitividad, a través de una beca de Formación de Personal Investigador (FPI).

La mayor parte del trabajo realizado se llevó a cabo en el Centro Oceanográfico de Gijón, perteneciente al Instituto Español de Oceanografía - Ministerio de Economía y Competitividad.

Esta tesis ha estado en el marco del proyecto METabolic OCean Analysis (METOCA) financiado por el Plan Nacional Español de Investigación+Desarrollo+Innovación (I+D+I). Ademas esta tesis se ha beneficiado del proyecto CONSOLIDER Malaspina 2010 y del programa Atlantic Meridional Transect (AMT)

Predicting the metabolic balance of the oceans

FRANCISCA GARCIA GARCIA

Oviedo, 2015