DNA metabarcoding derived biotic indices for marine monitoring and assessment

EVA AYLAGAS MARTÍNEZ

PhD Thesis, 2017

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

PhD Thesis

DNA metabarcoding derived biotic indices for marine monitoring and assessment

Presented by
Eva Aylagas Martínez

Thesis Supervisors
Dr. Naiara Rodríguez-Ezpeleta
Dr. Ángel Borja Yerro

Department of Zoology and Cellular Biology
Year 2017

“Whatever you do will be insignificant,

but it is very important that you do it.”

Mahatma Gandhi

A Marcos y Hugo

ACKNOWLEDGEMENTS

No puedo empezar de otra manera sino dando las gracias a mis supervisores, Ángel y Naiara, por haber confiado en mí para llevar este trabajo adelante. Esta tesis no hubiera sido posible sin vuestra ayuda, paciencia y apoyo durante estos cuatro años. Ángel, gracias por transmitirme tu confianza, por ayudarme a crecer como investigadora y por tu inmensa paciencia. Naiara, sin tu constancia y perseverancia no estaría donde estoy. Gracias por todo el tiempo que le has dedicado a esta tesis y por la ilusión que has puedo en ella (¡casi tanta como yo!).

Me gustaría agradecer a las instituciones, organizaciones y proyectos que han financiado esta tesis: AZTI, Fundación Centros Tecnológicos – Iñaki Goenaga, DEVOTES (DEVelopment Of innovative Tools for understanding marine biodiversity and assessing good Environmental Status - http://www.devotes-project.eu) financiado por la Union Europea (7th Framework Program 'The Ocean of Tomorrow' Theme, grant agreement no. 308392), Agencia Vasca del Agua (URA) y Convenio con la Dirección de Pesca y Acuicultura (Convenio Marco). También a Iñaki Saiz Salinas, del departamento de Zoología y Biología Celular Animal de la UPV, gracias por aceptar la tutorización de esta tesis y por tu enorme ayuda con todos los trámites administrativos.

Quiero agradecer a toda la gente de AZTI que en algún momento me ha echado una mano durante este tiempo, en especial a Xabier, Javi, María Uyarra, Laura Alonso, Guillem, Iñigo Muxika, Germán, Ibon y Luis. Y las palabras de ánimo constantes de Iratxe, Mikel, Imanol e Irati. De una manera muy especial a Iñaki Mendibil. Millones de gracias por tratar con tanta delicadeza mis muestras y por las horas compartidas en el laboratorio; aquellos tiempos de contaminaciones y risas han marcado el rumbo de esta tesis.

A mis compañeros de guerra que empezaron conmigo a darlo todo: Jon, Ernesto, Maddi, Iñigo y Jorge: nadie nos dijo que esto iba a ser fácil! Y a todos los demás doctorandos que me han acompañado durante los primeros años (Natalia, Mireia, Jon López, Jon Uranga, Nagore Cuevas, Nagore Zaldua eta Aizokorri) o los últimos (Itxaso, Oihane, Natalia, Ane, Anaïs, Blanca, Sarai, Isa, Kemal, Maite y Unai) de todo este camino. Y de una manera muy especial a Nerea y Lohitzune por todo el apoyo que me habéis dado hasta el ultimísimo momento, gracias, gracias, gracias!!

I would like also to thank to all the colleges I have met during the different travels I made during these four years. Matt and Laetitia, thank you for the great time we spent in Bocas del Toro (Panama) and all the advices you gave me when I was just starting this trip. Also, Susana and John, thanks for showing me the nicest part of KAUST, and especially for all the help you have provided me with this thesis. Finally, special thanks to Danovaro for your support during my time in Ancona and to Cinzia, Mike and Antonio. It was a pleasure working with you.

A vosotros, Marta, Juan, César, Victor, Carlos, Marina y Carmen, que me hacéis sentir el calor del Mediterráneo de una manera tan especial. A vosotras, Donostiarras, que sin saber muy bien de qué va todo esto no dejáis de decirme lo mucho que valgo. Y a mis compañeras de viajes y aventuras Oihane, Lucía y Tuca, la distancia sigue sin poder con nuestra amistad. A todos vosotros gracias por estar siempre ahí.

No puedo olvidar a la gente que me he encontrado por el camino y ha llegado a convertirse en mi familia. Esti y Unai, habéis sido un gran apoyo en mis idas y venidas a Sukarrieta. Gracias por ser simplemente como sois! Y como no, a mi a mis "forasteras", Andrea y Ángela. Encontrar almas gemelas no es fácil y menos con tanta paciencia como vosotras. Eneko, no te escapas de esta! Eskerrikasko benetan.

Gracias, muy especialmente, a mi familia, por vuestro apoyo incondicional, vuestra cercanía a pesar de la distancia, vuestra comprensión y vuestra generosidad absoluta. No tengo palabras para agradeceros todo lo que habéis hecho por mí. Diego, esta tesis es en parte tuya; gracias, una vez más, por creer en mí. Gracias Vanessa, Marcos y Hugo por los momentos de felicidad que hemos compartido juntos (algunos desde hace muy muy poquito tiempo) y los que nos quedan. Eskerrak eman baita ere Urnietako nire familiari, Maddi eta Patxi, gracias por hacerme sentir una más en casa, sabéis que sois muy especiales para mí.

Gracias, a ti, Xabi, porque te mereces todo y porque cada día a tu lado es una caja de sorpresas que no cambiaría por nada del mundo. Gracias por tu cariño, paciencia y positividad. Ya hemos terminado la tesis!

CONTENTS

Summary

The Water Framework and Marine Strategy Framework Directives have been developed within European legislation in order to protect and restore aquatic ecosystems. For this purposes, performing environmental status assessments that allow an integrated ecosystem management is required. Environmental status assessment is based on monitoring indicators; however, there is a demand of developing innovative monitoring tools and including new indicators within the assessments in order to increase speed, accuracy and cost-efficiency of monitoring programs. Among the promising innovative tools that can ease monitoring and potentially lead to new indicators are those based on genomic techniques, and in particular, on metabarcoding, a method that allows the accurate and cost-efficient taxonomic identification of multiple environmental samples even when including early developmental stages, cryptic species or degraded specimens. As such, this technique has been applied as a powerful alternative to traditional methodologies to detect toxic species, understand trophic interactions by analysing faeces or stomach contents or monitor early introduction of non-indigenous species, among others. However, before being included in current European legislation, metabarcoding needs to be benchmarked against morphology-based species identification, and proved capable of improving current monitoring by being more cost-effective and/or of producing new indicators that can be included in the mentioned directives. This Thesis explores the application of metabarcoding for responding to the need of new monitoring tools and indicators that allow more cost-effective and integrated environmental status assessments required for application in European directives. The specific aims of this Thesis are presented in six different Chapters.

Chapter 1 sets the ground for the implementation of a metabarcoding-based macroinvertebrate biotic index. The AZTI's Marine Biotic Index (AMBI) is one of the most successful and worldwide applied biotic indices for assessing seafloor integrity and derive health status of marine environments. Yet, this index is based on the benthic macroinvertebrate diversity, which is traditionally inferred by morphological taxonomic identification of the specimens present in each sample, which is time

consuming and requires not always available specialized taxonomical expertise. This chapter assesses the potential of metabarcoding as a cost-effective alternative to visual taxonomy by evaluating the requirements for the implementation of a genetics-based AMBI (gAMBI). The *in silico* analyses performed using available DNA sequence data showed that information about presence/absence of the most frequently occurring species provides accurate AMBI values and revealed the most suitable genetic fragment for metabarcoding marine macroinvertebrates. The results obtained set the basics for the implementation of the gAMBI, which has direct implications for a faster and cheaper marine monitoring and health status assessment.

Chapter 2 presents optimized protocols for macroinvertebrate metabarcoding data generation. The absence of protocols for metabarcoding marine benthic macroinvertebrates is one of the limitations that are preventing the potential inclusion of this technique in routine monitoring programs. A standardized protocol describing all steps regarding processing and manipulation of environmental samples for macroinvertebrate community characterization is presented. Detailed procedures for benthic environmental sample collection, processing, enrichment for macroinvertebrates, homogenization, and subsequent DNA extraction are provided.

Chapter 3 provides guidelines for the analysis of macroinvertebrate metabarcoding sequence data. Once the procedures presented in Chapter 2 have been followed, obtained DNA is used for the construction of amplicon libraries, which are sequenced using high-throughput sequencing. This technology produces a high amount of sequence data that need to be properly analysed in order to obtain accurate biodiversity assessments; yet, to date, no standardized pipelines are available for the analysis of macroinvertebrate metabarcoding data. This chapter presents detailed procedures for analysis of high-throughput sequence data derived from metabarcoding marine benthic macroinvertebrate samples based on two barcodes of the mitochondrial cytochrome oxidase I (COI) gene. In addition, this chapter shows how sequence data can be used for the calculation of benthic indices for environmental monitoring.

Chapter 4 validates metabarcoding for characterizing macroinvertebrate communities. For this purpose, a comprehensive study benchmarking metabarcoding against morphology for environmental monitoring based on benthic indices was performed. For that aim, benthic macroinvertebrate samples of known composition were analysed using alternative metabarcoding protocols and results compared to those obtained based on morphology. The comparisons highlighted the influence of the metabarcoding protocol in the obtained taxonomic composition and suggested that using inappropriate metabarcoding conditions can lead to erroneous biodiversity assessments. Additionally, a biotic index inferred from the list of macroinvertebrate taxa obtained using DNA-based taxonomic assignments (gAMBI) showed to be comparable to that inferred using morphological identification (AMBI). Thus, the analyses proved metabarcoding valid for ecological status assessment and will contribute to accelerating the implementation of this technique in regular monitoring programs.

Chapter 5 describes the application of metabarcoding in an ongoing monitoring program to confirm the suitability, in a real context, of this technique to provide accurate biotic indices. In this attempt, different variants of AMBI and gAMBI were inferred from paired samples collected from the same locations over multiple estuarine and coastal locations and compared. The results revealed that metabarcoding-based accurate inferences of marine ecological status are possible and that gAMBI succeeded in discriminating ecological status classes. Furthermore, compared to morphology based inferences, metabarcoding is both more time and cost effective. These results highlight that metabarcoding will contribute in a significant manner to improve large scale monitoring programs.

Chapter 6 responds to the necessity of including new indicators within current European directives by the development of a new biotic index based on bacterial communities. Biotic indices for used to assess seafloor integrity are mostly based on the analysis of benthic macroinvertebrate communities. Due to their high sensitivity to pollution and fast response to environmental changes, bacterial assemblages could complement the information provided by benthic metazoan communities as indicators of human-induced impacts, but so far, this biological component has not been well explored for this purpose. In this chapter, metabarcoding was applied for characterizing the bacterial assemblage composition of 51 estuarine and coastal stations characterized by different environmental conditions and human-derived pressures. Using the relative abundance of putative indicator bacterial taxa, a biotic index that was significantly correlated with a sediment quality index calculated on the basis of organic and inorganic compound concentrations was calculated. This new index based on bacterial assemblage composition can be a sensitive tool for providing a fast environmental assessment and allow a more comprehensive integrative ecosystem approach for environmental management.

The results obtained in the different Chapters are analyzed from an integrative point of view in the ***General discussion*** section. Overall, the results of the present work support the inclusion of metabarcoding as an appropriate approach for evaluating the health status of marine environments that can contribute to increase speed in providing monitoring results, which will greatly benefit implementation of current European directives.

RESUMEN

La Directiva Marco del Agua y la Directiva Marco de la Estrategia Marina tienen como principal objetivo proteger y restaurar los ecosistemas acuáticos, para lo cual es necesario evaluar el estado ecológico de una manera que permita realizar una gestión de los ecosistemas de forma integrada. Para ello, dicha evaluación se basa en el seguimiento de una serie de indicadores. No obstante, en los últimos años ha aumentado la necesidad tanto de desarrollar nuevas técnicas de seguimiento de los indicadores como de incluir indicadores nuevos que permitan realizar una evaluación ambiental más rápida, precisa y eficaz en términos económicos. Entre las herramientas innovadoras más prometedoras que permiten facilitar el seguimiento del estado y dar lugar a nuevos indicadores se encuentran las técnicas basadas en genómica, de las cuales cabe destacar el metabarcoding. Esta técnica permite realizar de forma simultánea la identificación taxonómica en numerosas muestras ambientales, de manera precisa y a bajo coste, lo que facilita la caracterización de comunidades biológicas, incluyendo cualquier estado de desarrollo, especies crípticas e incluso especímenes degradados. Así, el metabarcoding se propone como una técnica alternativa a los métodos tradicionales para detectar especies tóxicas, entender interacciones tróficas mediante el análisis de heces o contenidos estomacales, o efectuar el seguimiento de la introducción temprana de especies invasoras, entre otras aplicaciones. No obstante, antes de incluir el metabarcoding como herramienta de evaluación ambiental en directivas europeas, es necesario comparar su capacidad para generar los mismos resultados que los obtenidos con las herramientas tradicionales para la identificación de especies (basadas en caracteres morfológicos) y, por tanto, examinar su potencial para incluir nuevos indicadores de evaluación del estado. Esta Tesis investiga la aplicación del metabarcoding para responder a la necesidad de desarrollar nuevas herramientas de seguimiento que disminuyan costes y permitan llevar a cabo una evaluación ambiental integrada. Los objetivos específicos de esta Tesis se presentan en seis capítulos.

El ***Capítulo 1*** establece las bases para la implementación de un índice biótico de macroinvertebrados basado en metabarcoding. Uno de los

índices bióticos más utilizados a nivel mundial para evaluar la integridad de las comunidades bentónicas, y determinar así el estado de los ambientes marinos es el “AZTI’s Marine Biotic Index” (AMBI). Este índice se basa en la diversidad de los macroinvertebrados y se calcula asignando cada especie identificada a un grupo ecológico, que es función de la respuesta de la especie al estrés. Dichas especies son identificadas mediante técnicas basadas en morfología. Este proceso tiene una importante limitación en el tiempo necesario para llevarlo a cabo así como una dependencia, no siempre disponible para todos los phyla, de expertos taxónomos. En este capítulo evaluamos el potencial del metabarcoding como una técnica alternativa para disminuir el tiempo y los costes necesarios para calcular un índice biótico basado en genómica (gAMBI). Los análisis realizados *in silico* usando la información pública de secuencias de ADN de las especies bajo estudio mostraron que la información de presencia/ausencia de las especies más frecuentes en los muestreos proporciona valores suficientemente precisos de AMBI. Además, los resultados obtenidos permitieron determinar el marcador genético y los cebadores más apropiados para caracterizar macroinvertebrados usando metabarcoding. Los resultados obtenidos asientan las bases para la implementación de gAMBI, lo cual tiene implicaciones directas para un seguimiento marino y una evaluación del estado ecológico más rápidos y económicos.

El ***Capítulo 2*** presenta protocolos optimizados para la generación de datos de metabarcoding en macroinvertebrados bentónicos. La falta de protocolos estandarizados de metabarcoding para estas comunidades es una de las limitaciones que están impidiendo la inclusión de esta técnica en programas de seguimiento rutinarios. En este capítulo se define un protocolo que describe todos los pasos necesarios para el procesado y manipulación de muestras ambientales destinadas a caracterizar la comunidad de macroinvertebrados de sustrato blando. Así, se proporcionan en detalle indicaciones para recolectar la muestra, procesarla, llevar a cabo su homogeneización y posterior extracción de ADN.

En el ***Capítulo 3*** de esta Tesis, se presentan procedimientos estandarizados para el análisis de secuencias obtenidas a partir de metabarcoding aplicado a la caracterización de macroinvertebrados. Una vez realizados los primeros pasos para el procesamiento de la muestra y la consecuente extracción de ADN (presentados en el Capítulo 2), se procede a la construcción de librerías genéticas, las cuales son secuenciadas en plataformas de alto rendimiento. Estas tecnologías producen una gran cantidad de datos de secuencias que requieren un adecuado análisis para llevar a cabo una correcta evaluación de la biodiversidad. Así, este capítulo presenta un procedimiento detallado para analizar este tipo de datos basados en dos marcadores del gen citocromo oxidasa I (COI). Además, haciendo uso de dicha información, se detalla la manera en la que datos de secuenciación se pueden emplear para el cálculo de índices bióticos para la evaluación del estado ecológico.

El ***Capítulo 4*** muestra la validación de metabarcoding para caracterizar las comunidades de macroinvertebrados. Para ello, se ha realizado una evaluación comparativa entre el metabarcoding y la identificación morfológica utilizada en monitoreo ambiental, basada en índices bentónicos. Con este objetivo, se analizaron muestras de macroinvertebrados bentónicos de composición conocida, usando protocolos de metabarcoding alternativos, y los resultados se compararon con los obtenidos mediante identificación morfológica. Las comparaciones demostraron la influencia del protocolo de metabarcoding en la composición taxonómica obtenida, sugiriendo que el uso inapropiado de condiciones de metabarcoding puede producir evaluaciones de biodiversidad erróneas. Además, el cálculo de un índice biótico mediante el uso de una lista de especies de macroinvertebrados obtenida usando asignaciones basadas en ADN (gAMBI), mostró que los resultados eran comparables a los calculados usando la identificación morfológica (AMBI). En conclusión, los análisis han probado que el metabarcoding es válido para la evaluación del estado ecológico y que puede contribuir a acelerar la implementación de esta técnica en programas rutinarios de monitoreo.

En el ***Capítulo 5*** se describe la aplicación de metabarcoding en un caso concreto de monitoreo con el fin de confirmar la capacidad de la técnica para obtener índices bióticos que den respuesta a situaciones reales. Así, diferentes versiones de AMBI y gAMBI se han obtenido de muestras recolectadas en localidades de estuario y costeras. Los resultados mostraron que la técnica de metabarcoding es válida para evaluar el estado ecológico marino y que gAMBI es capaz de discriminar diferentes clases de estado ecológico en las muestras analizadas. Además, comparado con técnicas tradicionales basadas en morfología, el metabarcoding no solamente permite un monitoreo más rápido sino de menor coste. Estos resultados indican de manera notable que esta técnica contribuirá significativamente a la mejora de las evaluaciones ambientales a gran escala.

El ***Capítulo 6***, por su parte, responde a la necesidad de incluir nuevos indicadores en las directivas europeas mediante el desarrollo de un nuevo índice biótico basado en comunidades bacterianas. La mayoría de los índices bióticos utilizados para evaluar la integridad de los ecosistemas bentónicos están basados en el análisis de comunidades de macroinvertebrados. Debido a la gran sensibilidad que presentan las bacterias a la presencia de contaminantes, y su rápida respuesta a cambios ambientales, estos organismos pueden complementar la información proporcionada por las comunidades de macroinvertebrados como indicadores de impactos antropogénicos. No obstante, este componente biológico no ha sido evaluado, hasta la fecha, para este objetivo. En este capítulo, se ha caracterizado la comunidad bacteriana correspondiente a muestras de sedimento recogidas en 51 localidades de estuario y costa con diferentes condiciones ambientales y presiones antropogénicas utilizando la técnica de metabarcoding. Usando la abundancia relativa de bacterias potencialmente indicadoras de alteraciones ambientales, se ha calculado un índice biótico que se correlacionó significativamente con un índice de calidad del sedimento, basado en concentraciones de compuestos orgánicos e inorgánicos. Este nuevo índice biótico basado en la comunidad bacteriana puede ser utilizado como una herramienta para proporcionar una evaluación ambiental de una forma rápida y permitir así llevar a cabo una gestión ecosistémica de manera integral, utilizándolo junto a otros índices y componentes ecosistémicos.

Los resultados obtenidos en los diferentes capítulos de esta Tesis se han analizado de una forma integrada y se presentan en la sección de ***Discusión general***. En general, los resultados demuestran la posibilidad de la inclusión del metabarcoding como una técnica apropiada para identificar muestras de biodiversidad y contribuir a evaluar el estado del medio ambiente marino, proporcionando resultados de una manera más rápida y eficiente, contribuyendo a la implementación de las directivas europeas relacionadas con el medio marino.

INTRODUCTION

1. The importance of assessing marine systems

Marine environments, both coastal and offshore, are being severely impacted by traditional and emerging human activities (Borja et al., 2016b) such as shipping, fishing, wastewater discharging, recreation and renewable energy production (OSPAR, 2009). This is translated into habitat loss, overexploitation of resources, eutrophication, pollution by hazardous substances and introduction of non-indigenous species, all among the main threats of marine biodiversity (Diaz and Rosenberg, 2008; Halpern et al., 2008) that compromise the sustainability of marine ecosystems and their services. As a response to the fast environmental degradation (Lotze et al., 2006), the United Nations Convention on the Law of the Sea (UNCLOS, 1982), the international basic legal framework that governs the use of the oceans and seas, establishes an international obligation to protect and sustainably use marine resources. Among the initiatives developed for protecting and restoring the aquatic environment within recent European legislation are the Water Framework Directive (WFD, 2000/60/EC) and the Marine Strategy Framework Directive (MSFD, 2008/56/EC).

The WFD applies to all surface waters (including freshwater, transitional and coastal waters) of the European Union (EU), while the MSFD covers the waters from the costal baseline to 200 nautical miles, which is known as Exclusive Economic Zone (EEZ). The WFD requires “Good Ecological Status” to be achieved based upon the assessment of a variety of biological (i.e. phytoplankton, macroalgae, seagrasses, macroinvertebrates and fishes), physico-chemical (e.g. salinity, transparency, oxygenation, pollutants and nutrient status) and hydro-morphological (e.g. depth variation, wave exposure and tidal regime) quality elements (Heiskanen et al., 2004). The MSFD aims at protecting the marine environment by achieving and/or maintaining “Good Environmental Status (GES)” by 2020, and constitutes one of the major legal frameworks for the protection of marine biodiversity along with the EU Biodiversity Strategy 2020 (COM/2011/0244) and the Convention on Biological Diversity (CBD, 2000). In the MSFD, environmental status assessment takes into account marine ecosystem structure, function and processes, encompassing physical, chemical, physiographic, geographic and climatic factors, and integrates these conditions with anthropogenic impacts and activities (European Commission, 2008) (Borja et al., 2015a). Some of the indicators for biological quality elements under the WFD can be applied to the MSFD (Borja et al., 2010), which may provide an easier and reliable way to implement this complex directive.

The MSFD relies upon an “integrated ecosystem-based management approach” (CBD, 2000) to assess environmental status. This approach requires that the evaluation of human activities and their pressures is performed by simultaneously measuring a variety of ecosystem components together with physico-chemical parameters and elements of pollution (Borja et al., 2010). Ecosystem-based management has been used as a way to consider the complex interactions of the biological, physical, chemical and human components of an ecosystem instead of managing ecosystem elements individually (Borja et al., 2013b). The MSFD evaluates the status of the marine environment based on eleven qualitative descriptors (Table I.1). Within each descriptor, the European Commission Decision 2010/477/EU provides a set of 29 criteria that have 56 associated indicators (Berg et al., 2015), whose evaluation require adequate and rigorous spatiotemporal monitoring (Borja et al., 2011).

2. Improved marine biomonitoring within European directives

Descriptor 1, “Biodiversity”, is one of the cornerstones of the MSFD (Heiskanen et al., 2016), but additional descriptors, such as “Non-indigenous species”, “Marine food webs”, “Human-induced eutrophication” or “Sea floor integrity” are also related to biodiversity. As a central MSFD descriptor, it is suggested that biodiversity should act as one of the key elements for attaining GES at the rest of the descriptors (Figure I.1). Thus, following the scheme of the integrated ecosystem approach, the pressures and their effects on biodiversity should be simultaneously analysed and complemented for determining GES in the different descriptors. In order to fulfil these requirements, monitoring methodologies must be able to determine the effects of human pressures over large geographical and temporal scales and to provide rapid and comparable results across different regions. As such, there is a need for developing new monitoring approaches that allow evaluating the biodiversity related descriptors in an integrative way (Borja and Elliott, 2013).

Table I.1. Marine Strategy Framework Directive (MSFD) qualitative descriptors describing the environment condition required to assume Good Environmental Status (GES).

MSFD descriptor name	Qualitative descriptors which describe what the environment will look like when GES has been achieved
D1: Biological diversity	Biological diversity is maintained. The quality and occurrence of habitats and the distribution and abundance of species are in line with prevailing physiographic, geographic and climatic conditions
D2: Non-indigenous species	Non-indigenous species introduced by human activities are at levels that do not adversely alter the ecosystems
D3: Commercially exploited fish and shellfish	Populations of all commercially exploited fish and shellfish are within safe biological limits, exhibiting a population age and size distribution that is indicative of a healthy stock
D4: Marine food webs	All elements of the marine food webs, to the extent that they are known, occur at normal abundance and diversity and levels capable of ensuring the long-term abundance of the species and the retention of their full reproductive capacity
D5: Human-induced eutrophication	Human-induced eutrophication is minimised, especially adverse effects thereof, such as losses in biodiversity, ecosystem degradation, harmful algae blooms and oxygen deficiency in bottom waters
D6: Sea floor integrity	Sea-floor integrity is at a level that ensures that the structure and functions of the ecosystems are safeguarded and benthic ecosystems, in particular, are not adversely affected
D7: Hydrographical conditions	Permanent alteration of hydrographical conditions does not adversely affect the ecosystem
D8: Concentrations of contaminants	Concentrations of contaminants are at levels not giving rise to pollution effects
D9: Contaminants in fish and other seafood	Contaminants in fish and other seafood for human consumption do not exceed levels established by Community legislation or other relevant standards
D10: Marine litter	Properties and quantities of marine litter do not cause harm to the coastal and marine environment
D11: Energy, including underwater noise	Introduction of energy, including underwater noise, is at levels that do not adversely affect the marine environment

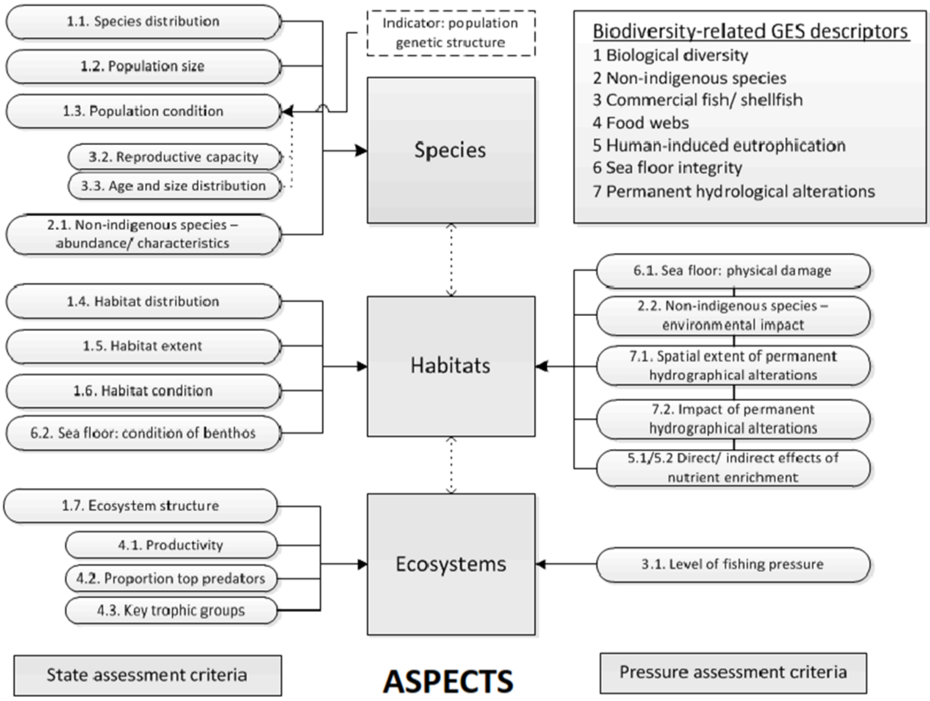

Figure I.1. Schematic representation of the biodiversity-related descriptors defined in the Marine Strategy Framework Directive for assessing Good Environmental Status (GES). The pressures and state assessment criteria are associated to the different ecological scales to which they are linked (ecosystems, habitats or species). Source: DEVOTES project (http://www.devotes-project.eu).

Traditional methodologies used to assess biodiversity have limitations in providing rapid, cost-effective, taxonomically comprehensive and spatiotemporally wide-range measurements (Danovaro et al., 2016). For example (i) *in situ* observational surveys (e.g. for seabed ecosystems) are often not comprehensive enough and have significant spatiotemporal gaps (Blondeau-Patissier et al., 2004), (ii) traditional sampling methods, such as grabs, are ineffective at some areas dominated by hard substrata, (iii) morphology-based biological community characterization is time consuming and taxonomic expertise dependent (Dafforn et al., 2014), and (iv) evaluation of eutrophication based on optical metrics of phytoplankton is subjected to natural variability of pigmentation or cell size (Goela et al., 2015). These limitations inherent to traditional methodologies for biodiversity assessment impede both, making comparisons over time and space and providing rapid results for monitoring (De Jonge et al., 2006). As a consequence, integrated ecosystem management cannot be performed. Recently developed technologies present advantages for improving and easing marine monitoring including ability to provide higher taxonomic resolution and faster outcomes, and to cover wider geographic areas and larger temporal scales (Danovaro et al., 2016). Yet, although promising, new technologies still need to be tested prior to their application in routine marine monitoring. The main innovative tools being developed can be placed into four main categories: systems for *in situ* analysis, remote sensing, modelling and genomics (Table I.2).

Table I.2. Innovative technologies for marine monitoring and biodiversity related descriptors to which each tool can be applied. MSFD: Marine Strategy Framework Directive. D: descriptors (for equivalence, see Table I.1)

Technology	MSFD descriptors
Instruments for *in situ* analysis	D1, D2, D3, D4, D6
Remote sensing	D1, D5
Modelling	D1, D3
Genomics	D1, D2, D3, D4, D5, D6

3. The promise of genomic tools for marine monitoring

Among the innovative tools proposed to improve and ease marine monitoring, those based on genomics are considered particularly promising (Bourlat et al., 2013) and can be applied to a variety of Descriptors (Table I.3). Genomic tools allow measuring variables that were not possible to measure before and provide alternatives to ease measurements otherwise performed with traditional methodologies. Thus, by analysing nucleotide sequences, a wide variety of genomic techniques can be potentially applied to taxonomically and metabolically characterize biological communities, rapidly detect toxic or invasive species, determine connectivity among populations or assign individuals to populations, among others (Bourlat et al., 2013). Furthermore, the recent development of high-throughput sequencing technologies has produced large amounts of genetic data from a variety of organisms that can be used to rapid and cost-effectively measure various indicators.

Table I.3. Genomic tools for marine monitoring and descriptors in which they can be applied. Barcoding and metabarcoding consist on sequencing a portion of the genome of an individual or of the whole community, respectively. Metagenomics and metatranscriptomics consist on sequencing the genome or the transcriptome of the whole community, respectively. Microarrays and qPCR (quantitative PCR) quantify a gene or transcript in high or low numbers of samples respectively. SNP genotyping allows assessing population structure and assigning individuals to populations based on DNA sequence variations. Table adapted from Borja et al. (2016b) and Bourlat et al. (2013)

Genomic tool	Application to monitoring	MSFD descriptors
Barcoding and Metabarcoding	Community taxonomic characterization	D1, D2, D4, D5, D6
Metagenomics	Community metabolic potential characterization	D1, D2, D4, D5, D6
Metatranscriptomics	Community metabolic activity characterization	D1, D2, D4, D5, D6
Microarrays	Metabolic activity characterization and high-throughput species detection and quantification and gene expression quantification	D2, D5
qPCR	Low-throughput species detection dn quantification and gene expression quantification	D2, D5
SNP genotyping	Connectivity assessment and assignment of individuals to populations	D1, D3

Indicators of the biodiversity related descriptors are often monitored by characterizing biological communities (e.g. *composition of ecosystem components (i.e. species)* – D1; *occurrence and spatial distribution of non-indigenous species* – D2; *distribution of key trophic groups/species* – D4; *presence of particularly sensitive and/or tolerant species* – D6). This characterization currently relies on morphological analyses that imply the knowledge of taxonomic experts, who are generally specialized on some specific groups of organisms (Bacher, 2012). Moreover, morphological identification can introduce biases due to erroneous species classification, especially in the presence of cryptic species, damaged specimens or larval/juvenile stages (Kochzius et al., 2008), and is often limited to large organisms (Pawlowski et al., 2012). These limitations make biodiversity assessments costly and time-consuming and impede a comprehensive characterization of the biological community in large scale monitoring programs (Bourlat et al., 2013). In this context, genomic tools represent an opportunity for improving biodiversity assessments.

3.1. DNA barcoding and metabarcoding

The most promising genetic techniques for improving biodiversity assessments are barcoding and metabarcoding (Figure I.2), which consist respectively on taxonomically assigning a specimen or a mixture of specimens contained in a 'bulk' sample by means of a standardized short DNA fragment (barcode) that is compared against a reference database containing the correspondence between barcodes and taxonomy (Hebert et al., 2003a; Taberlet et al., 2012a). Barcoding has been applied in biodiversity conservation, environmental management and the study of trophic interactions (Valentini et al., 2009; Taylor and Harris, 2012). However, the process is quite laborious because it requires each species be processed individually (Cameron et al., 2006; Stein et al., 2014). In contrast, metabarcoding allows analysing whole samples without needing to isolate individual organisms (Creer et al., 2010), which, on top of overcoming dependence on taxonomic expertise, allows rapid analyses of several samples and, consequently reduces monitoring costs and allows performing large-scale surveys (Yu et al., 2012; Kelly et al., 2014b).

The successful application of metabarcoding in biodiversity assessments of marine ecosystems relies on a series of premises that need to be fulfilled. First, a barcode that is present in all target species must be selected. The barcode should have enough sequence variability to allow distinction among related species (Hebert et al., 2003a) and must be flanked by conserved regions so that universal primers for amplification of all target organisms during the amplification through the polymerase chain reaction (PCR) could be designed (Leray et al., 2013; Lobo et al., 2013). Second, a database containing the correspondence between barcodes and taxonomy should exist so that classification of the maximum number of unknown barcodes into species can be performed (Zepeda Mendoza et al., 2015). Indeed, it has been stated that species identification by metabarcoding is as good and reliable as complete and accurate the reference database is (Wangensteen and Turon, 2016). Third, standardized protocols both for sample processing and data analysis must ensure the reliable characterization of the target community and allow generating reproducible and comparable results (Creer et al., 2016).

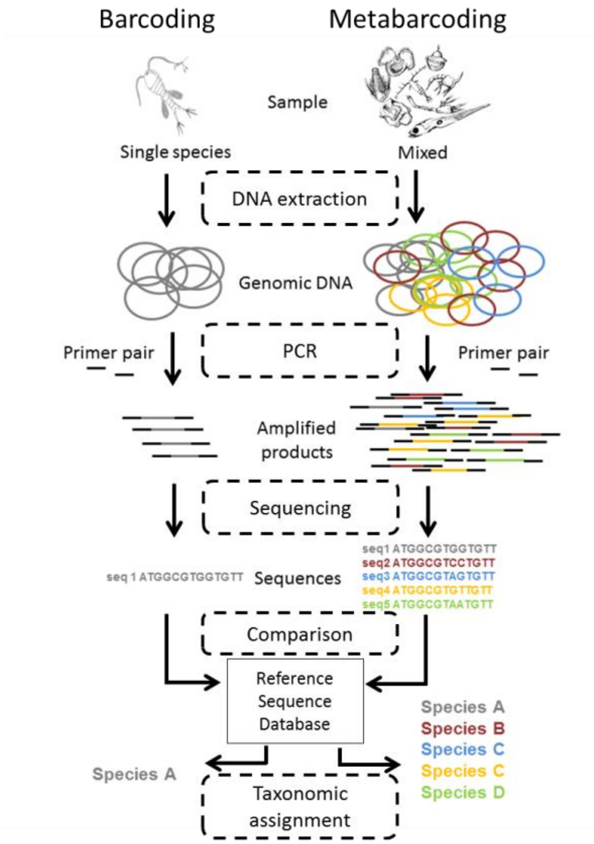

Figure I.2. Schematic representation of the steps involved in barcoding and metabarcoding. Selected barcodes are amplified from genomic DNA. Amplified products are identical in barcoding, whereas a mixture of amplified products is obtained in metabarcoding. Once the amplified products have been sequenced, taxonomic assignment is preformed based on comparison of the obtained sequences to a reference sequence database. Modified from Corell and Rodriguez-Ezpeleta (2014)

So far, metabarcoding has been mostly applied to bacteria (Sogin et al., 2006; Bartram et al., 2011; Zinger et al., 2011; Caporaso et al., 2012; Sun et al., 2013; Ferrera and Sanchez, 2016) and microbial eukaryotes (Stoeck et al., 2009; Chariton et al., 2010; Edgcomb et al., 2011; Pawlowski et al., 2014b; Laroche et al., 2016), but recently, an increasing number of studies characterizing marine metazoans through metabarcoding have been performed, and have targeted meiofauna (Creer et al., 2010; Fonseca et al., 2014; Dell'Anno et al., 2015; Guardiola et al., 2015; Guardiola et al., 2016), zooplankton (Lindeque et al., 2013; Pearman et al., 2014; Hirai et al., 2015; Pearman and Irigoien, 2015; Abad et al., 2016; Bucklin et al., 2016), phytoplankton (Yoon et al., 2016), fishes (Thomsen et al., 2012a; Kelly et al., 2014a; Turner et al., 2015), and benthic macroinvertebrates (e.g. Leray and Knowlton, 2015; Pearman et al., 2016a). This wide range of applications anticipates that metabarcoding can be potentially used for the assessment of various MSFD descriptors such as “Biodiversity”, “Non-indigenous species”, “Marine food webs”, “Human-induced eutrophication” and “Seafloor integrity”.

4. Improving the assessment of Seafloor Integrity

The MSFD states that seafloor good environmental status will be achieved when “*it is at a level that ensures that the structure and functions of the ecosystems are safeguarded and benthic ecosystems, in particular, are not adversely affected*”. Benthic habitats play a significant role in ecological processes and are one of the most important sources of ecosystem services (Harley et al., 2006). Thus, comprehensively monitoring benthic ecosystems is essential to ensure the sustainability of marine resources. Due to the large variety of seafloor types (soft substrata – sand, mud, gravel and mixed sediments; hard substrata – bedrock and boulders; and, biogenic substrata – mussel beds and cold-water coral reefs), it is necessary to define indicators and standardized methods that provide accurate information about the status of benthic ecosystem as a whole and of their alteration by human induced pressures (Fisher et al., 2001). These indicators can be based on the presence of particularly sensitive or tolerant species or can also be indices calculated from several parameters such as species diversity, number of species and proportion of different types of species in benthic samples (Rees et al., 2008).

4.1. Benthic macroinvertebrate community as indicator of seafloor integrity

Macroinvertebrate communities are frequently monitored in benthic systems and used as indicators for a variety of reasons: (i) they live in sediments, where the exposure to contaminants and oxygen stress is most evident (Engle, 2000), (ii) they are relatively sedentary, reflecting the quality of the immediate environment (Dauer, 1993; Weisberg et al., 1997), (iii) they present long life cycles, allowing integration of water and sediment quality changes over time (Dauer, 1993), (iv) they respond rapidly to both anthropogenic and natural pressures (Marques et al., 1993; Lerberg et al., 2000), (v) they include a wide range of species with different tolerance levels to pollution, allowing their inclusion into different functional response groups (Grall and Glémarec, 1997), and (vi) they represent the link with higher trophic levels and some species are, or are prey of, commercially important species (McLusky and Elliot, 2004).

Benthic macroinvertebrates are monitored within the MSFD to assess seafloor integrity and used to calculate biotic indices that integrate the information obtained from the macroinvertebrate community into a single number to ease the interpretation of the ecological status (Diaz et al., 2004; Pinto et al., 2009). A variety of benthic macroinvertebrate community-based biotic indices have been developed, such as the AZTI´s Marine Biotic Index (AMBI; Borja et al., 2000), the BENTIX (Simboura and Zenetos, 2002) or the Benthic quality index (BQI; Rosenberg et al., 2004). One of the most applied benthic community-based biotic indices is the AMBI (see review in Borja et al., 2015b), which is based on a prior assignment of macroinvertebrate species to five ecological groups according to their sensitivity to an increasing stress gradient (Figure I.3), and calculated applying the formula AMBI = (0 × %GI) + (1.5 × %GII) + (3 × %GIII) + (4.5 × %GIV) + (6 × %GV)/100, where percentages of individuals of each ecological group are multiplied by a factor so that lower and higher AMBI values indicate less and more disturbed status respectively.

- *Ecological Group I:* Species very sensitive to organic enrichment and present under undisturbed conditions (initial or climatic state)
- *Ecological Group II:* Species indifferent to enrichment, always present in low densities with non-significant variations with time (from initial state to slight unbalanced)
- *Ecological Group III:* Species tolerant to excess organic matter enrichment. These species may occur under normal conditions, but their populations are stimulated by organic enrichment (slight unbalanced situations)

- *Ecological Group IV:* Second-order opportunistic species (slight to pronounced unbalanced situations)
- *Ecological Group V:* First-order opportunistic species (pronounced unbalanced situations)

Currently, the number of macroinvertebrate species for which ecological group has been assigned, and are therefore used for the calculation of AMBI, is of over 7,000, belonging to 19 different phyla, being the most abundant Annelida, Mollusca, Arthropoda, Echinodermata and Cnidaria (http://ambi.azti.es).

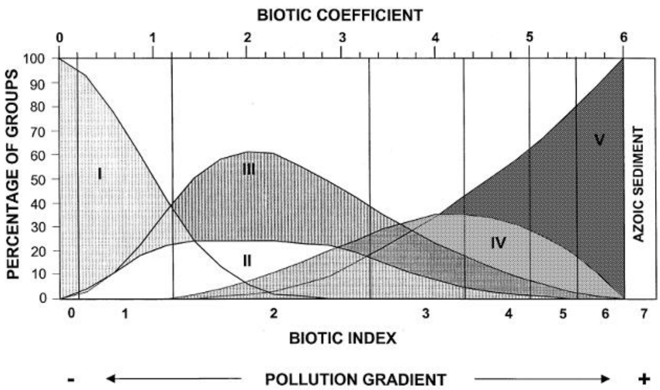

Figure I.3. Theoretical model from which AMBI is constructed. The model provides the ordination of benthic macroinvertebrate species into five ecological groups according to their sensitivity to an increasing pollution gradient. The relative proportion of abundance of each group in a sample provides a continuous value (biotic coefficient) and an equivalent discreet biotic index to discriminate disturbance classes: undisturbed [0 - 1.2], slightly disturbed [1.3 - 3.3], moderately disturbed [3.4 - 5], heavily disturbed [5.1 - 6] and extremely disturbed [6.1 – 7]. Source: Borja et al. (2000)

AMBI allows the detection of anthropogenic pressures in the environment as it can be used to measure the evolution of the ecological status of a particular region (Muxika et al., 2005). It was initially developed to evaluate the status of different locations in the Basque coast (northern Spain), but has been successfully applied to different geographical areas worldwide (Borja et al., 2009b) covering a range of different impact sources with increasing number of users in European marine waters, such as the Baltic (Zettler et al., 2007) and Mediterranean (Ponti et al., 2008) seas, and the Atlantic ocean (Salas et al., 2004), as well as in the South American Atlantic region (Muniz et al., 2005; Valenca and Santos, 2012).

Since AMBI was developed, several refinements of the index to include other metrics describing the benthic community integrity have been performed in response to the WFD and MSFD requirements. For example, the Multivariate-AMBI (M-AMBI) adds species richness and diversity (Muxika et al., 2007) and the biomass(B)-AMBI (BAMBI) adds species biomass (Warwick et al., 2010). The calculation of the different versions of AMBI is usually done by isolation and taxonomic identification based on morphology of each macroinvertebrate specimen. The abovementioned potential erroneous classification of the species is translated into an incorrect assignment of ecological groups (Ranasinghe et al., 2012) and, as a result, inaccurate disturbance classification. The limitations of the morphology-based taxonomic assignment are especially evident when analyzing several samples in real monitoring programs. In such surveys, large number of sites are monitored so that a high number of benthic samples are analyzed from which rigorous species identification needs to be performed for calculating AMBI (Borja et al., 2013a). As a consequence, obtaining accurate and rapid AMBI data in order to respond in a timely manner to environmental management directives is sometimes unfeasible. Thus, alternative monitoring methodologies are essential for providing a more accurate and rapid characterization of the macroinvertebrate community.

Metabarcoding represents a potential alternative to overcome the issues related with morphology-based macroinvertebrate community assessments. To date, few studies have been performed to assess the capability of metabarcoding for characterizing macroinvertebrates (Carew et al., 2013; Dafforn et al., 2014; Gibson et al., 2015; Lejzerowicz et al., 2015). In general, they prove the potential of the technique for environmental assessment purposes. However, the challenges associated to the different steps of metabarcoding analyses have not been exhaustively analysed for this community, which is crucial to establish the best procedures for accurately and reliably performing biodiversity assessments in the future. In this sense, before using metabarcoding for macroinvertebrate community characterization, robust studies benchmarking the technique against morphology are needed. In view of the indicative potential of macroinvertebrates in the context of the MSDF and the potential capacity of metabarcoding for characterizing this community, developing a genomic version of AMBI could ease and increase the speed in the assessment of seafloor integrity.

4.2. Microbial assemblages as indicator of seafloor integrity

Bacteria and microbial eukaryotes are crucial in the functioning of marine ecosystems (Azam and Malfatti, 2007; Gasol et al., 2008) and essential for the maintenance of marine food webs (Cotner and Biddanda, 2002). An understanding of their composition and dynamics is critical for studying ecosystem functions and services. Marine microbial community composition and metabolic activity are highly sensitive to environmental changes, such as in temperature, pH or oxygen (Hoppe et al., 2008; Burns et al., 2013). In particular, bacterial assemblages present the capacity to rapidly respond to natural or anthropogenic pressures (Zhang et al., 2008b; Chiellini et al., 2013; Zhang et al., 2014). Further, they play an important role in benthic systems as they are essential in recycling organic matter (Pusceddu et al., 2009). Several studies have demonstrated the potential use of bacteria as indicators of human-impacted environments (Nogales et al., 2011; Lozada et al., 2014). For example, certain bacterial taxa have been identified as indicators of organic enriched sediments from locations influenced by fish farming activities (Aranda et al., 2015) in eutrophic estuaries (Sun et al., 2013) or harbour areas (Zhang et al., 2008a; Ziegler et al., 2016).

In contrast with the widely recognized relevance of bacterial processes in marine ecosystem functioning and their response to human induced pressures, the MSFD so far does not include the evaluation of bacterial communities as indicator along the different descriptors (Caruso et al., 2015). The high complexity of these communities in terms of diversity and functioning in natural ecosystems (Nogales et al., 2011) and the difficulties in their taxonomic identification have limited the use of this community as indicators. However, ignoring the evaluation of bacterial communities within the MSFD impedes the application of an integrated ecosystem approach-based management, which should include the evaluation of all ecosystem components, from microorganisms to mammals (Borja et al., 2008). As a response, it has been recently proposed that including this biological component within the MSFD would be of great benefit (Caruso et al., 2015).

In recent years, the advent of high-throughput sequencing techniques has allowed the characterization of bacterial assemblages from different marine environments (Wang et al., 2012; Zhou et al., 2013; Ye et al., 2016; Zhang et al., 2016) to an extent that was inconceivable only few years ago (Barberan et al., 2014). Remarkably, it has been possible to identify key microorganisms involved in important ecosystem processes (Gilbride et al., 2006; Tan et al., 2015a) as well as to rapidly characterize bacterial assemblages from several samples simultaneously at low cost using metabarcoding (Ferrera and Sanchez, 2016). In this context, the capacity of bacterial communities for rapidly respond to environmental changes and the capability for characterizing this biological component using metabarcoding can be combined to develop a new biotic index using bacteria, which can potentially increase the confidence level in the classification of the ecological status, as it can complement the information provided by other biological communities (e.g. macroinvertebrates) as an early warning sign to assess impacts.

5. Development of DNA metabarcoding-based biotic indices

When designing a metabarcoding study for characterizing whole communities, there are many decisions to make. The process is linear (Figure I.4), and the steps usually consist on sample collection, sample processing, DNA extraction, barcode PCR amplification, High-Throughput Sequencing (HTS), bioinformatic data analysis, and data interpretation. Hence, if the objective is developing a biotic index, validation is required. Nevertheless, the protocols used for each step can vary widely based on the question, the environment and the target community.

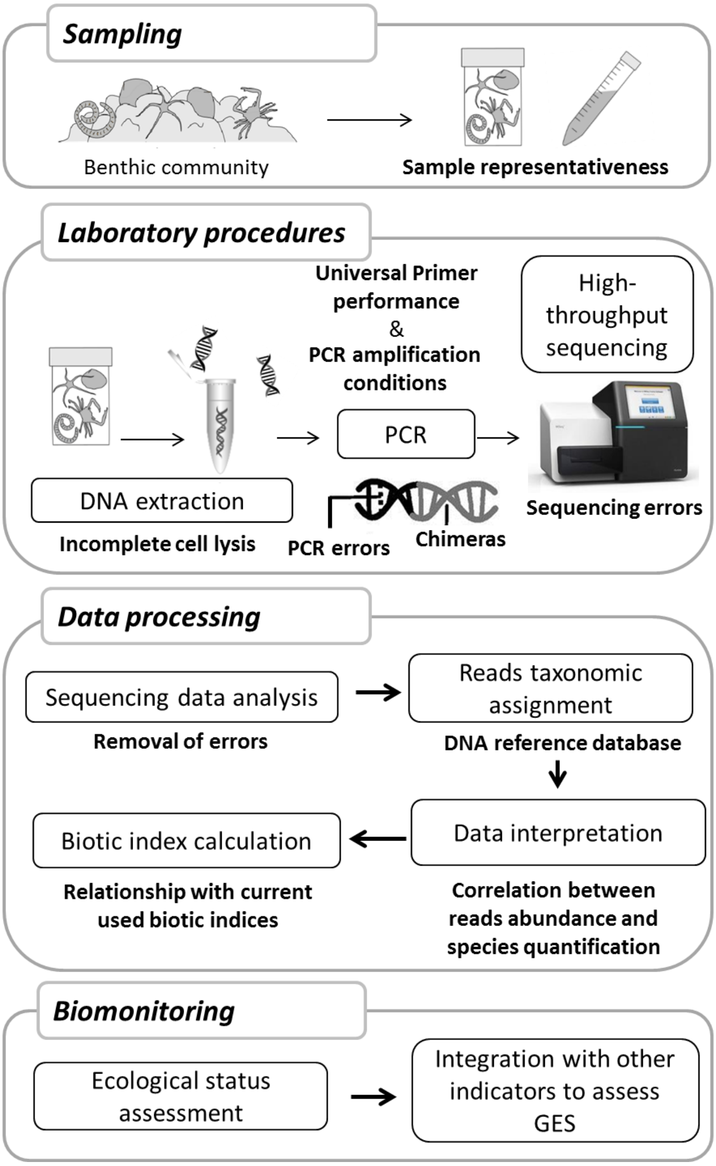

Figure I.3. General view of the metabarcoding-based biomonitoring workflow. Critical issues of the approach are indicated in bold and addressed in this PhD Thesis. Modified from Pawlowski et al. (2014b)

5.1. Sample collection, processing and DNA extraction

The size range of the target organisms and patchiness in their distribution typically determine how much sample is processed for DNA extraction (Creer et al., 2016). For microbial communities, a small volume of sample material (2.5 gr of sediment) is usually enough for performing downstream analysis (Pawlowski et al., 2014a) because they are small size and usually homogeneously distributed. Furthermore, commercially available kits have been developed for the extraction of microbial DNA from small sediment samples. In contrast, when the target species are larger (e.g. macroinvertebrate community), the presence of a wide size range (from 1 mm to several cm) and heterogeneously distributed specimens makes it crucial to ensure a subsample for analysis that is representative of the whole community (Wangensteen and Turon, 2016). Importantly, correct processing of the sample is required to ensure that DNA is effectively extracted from all species present. If incorrect procedures are undertaken in this steps, these can be passed on downstream and affect the inferred community composition. For example incomplete cell lysis or uneven degradation during DNA extraction phase could affect the inferred community composition especially as taxa will be affected differently. For metabarcoding macroinvertebrates, there are not established procedures that ensure the reliability of these practices. Thus, developing and testing protocols to process the samples and to effectively extract DNA from all taxa is essential to ensure accurate and reproducible results (Creer et al., 2016).

5.2. Barcode amplification and sequencing

Protocols for bacterial community composition analyses using metabarcoding are already developed and have been tested and validated in a wide range of environments (Caporaso et al., 2011; Sinclair et al., 2015). Indeed, universal primers to target a fragment of the 16S rRNA gene have being defined (Caporaso et al., 2012; Klindworth et al., 2012). In contrast, these aspects are not well established for the macroinvertebrate community. Different barcodes such as portions of the small and large subunits of the nuclear ribosomal RNA (18S and 28S rRNA) genes (Machida and Knowlton, 2012) and of the mitochondrial cytochrome oxidase I (COI) (Meusnier et al., 2008) and small subunit of the ribosomal RNA (16S rRNA) genes (Sarri et al., 2014) have been proposed for metabarcoding. To date, there are no standardized universal primer sets to reliably apply metabarcoding for characterizing this community. As such, testing the performance of different universal primers designed to amplify a variety of marine metazoans taxa (Leray et al., 2013; Lobo et al., 2013) and selecting the most suitable to retrieve the highest number of macroinvertebrate species present in a sample is required. At the same time, establishing the PCR conditions that most accurately characterize the macroinvertebrate community is crucial for a successful application of metabarcoding to biomonitoring (Deagle et al., 2014). In addition, the PCR amplification step is a source of errors that needs to be carefully addressed (Pawlowski et al., 2014b). For example increasing the PCR cycle number, increases detection of some taxa, but also the potential for PCR errors, such as insertion of erroneous bases or formation of chimeric sequences (i.e. DNA artefacts generated during the PCR step which consists of DNA molecules with two or more fragments from two or more original DNA sequences).

Once the PCR products are obtained, they are sequenced on a HTS platform. HTS platforms are capable of sequencing multiple DNA molecules in parallel (each nucleotide sequence is called a 'read'), enabling hundreds of millions of DNA molecules to be sequenced simultaneously. The currently available sequencing platforms differ in the total number of reads obtained, the length of these reads and the average error rate per read (Loman et al., 2012). The selection of an appropriate sequencing platform is an important consideration that requires understanding the specificities of each technology. The MiSeq (Illumina) sequencing platform provides high yield (up to 25 million reads per run), can sequence relatively long overlapping fragments (300 bp paired-end reads), and is fast (results are available in less than a week), which makes it the most commonly used sequencing platform for metabarcoding-based biodiversity assessment studies nowadays (Caporaso et al., 2012; Wangensteen and Turon, 2016).

5.3. *Data processing*

After the sequencing process, an essential part of the metabarcoding workflow, is the bioinformatic analysis of the large amount of sequence data generated. The sequencing analysis procedure can be summarized in removal of reads or read fragments that contain potential errors (inserted during the PCR amplification process or during the sequencing) (Caporaso et al., 2011) and taxonomic assignment of the retained reads using curated databases. For the former step, appropriate pipelines for data processing that ensure the correct sequence analysis must be utilized, and for the second, the reliability depends on the correct taxonomic assignment of reference sequences and on the number of taxonomically assigned sequences in the reference database. Regarding bacteria, curated reference databases have been gathered for the 16S rRNA, such as SILVA (Quast et al., 2013), and for metazoans, BOLD (Ratnasingham and Hebert, 2007) is the best curated database containing thousands of reference COI sequences.

Whereas in bacteria, a positive association between sequencing read number and abundance can be obtained (Turnbaugh et al., 2010), this association in macroinvertebrates is constrained mostly due to biological factors such as multicellularity, variation in tissue cell density, and in inter and intra specific variations in gene copy number (Bik et al., 2012; Pompanon et al., 2012). Additionally, technical factors associated to metabarcoding can introduce errors along the different steps (i.e. biases during DNA extraction, PCR, sequencing and bioinformatics analysis) that are likely to impede making estimation of species biomass or abundance from sequence data (Porazinska et al., 2010). Attempts to quantify the relationships between abundance or biomass and read number in metazoans using metabarcoding have yielded low correlations (Carew et al., 2013; Zhou et al., 2013; Hirai et al., 2015), but have not been based on carefully controlled experimental conditions. Studies performed in controlled experiments show better associations between biomass and metabarcoding read number, but this relations vary across taxa (Elbrecht and Leese, 2015). This limitation is likely to prevent the use of metabarcoding for macroinvertebrate community-based biomonitoring relying on abundance metrics (Yu et al., 2012). Consequently, attempting a genomic version of AMBI requires deeper studies understanding the effect of read abundance in species quantification and estimation of biotic indices.

5.4. Biotic index calculation

Information on the ecological groups of macroinvertebrate species is well stablished and available, but metabarcoding protocols for analyzing this community are scarce and/or not well evaluated. In this context, the development of a genomics-based macroinvertebrate biotic index would be centered in the comparison of the taxonomic assignments obtained with morphology and metabarcoding and in assessing the effect of the potential differences in taxonomic inferences in biotic index calculations.

On the other side, metabarcoding protocols for bacterial community analyses are well established, but information on assignment of bacteria to ecological groups according to their response to a pollution gradient is lacking. In this context, the development of a bacterial community-based index will be centered in a prior classification of bacterial taxa into ecological groups son that the increasing interest in including the evaluation of bacterial communities within the MSFD (Caruso et al., 2015) can be taken into account. For that aim, the characterization of this component from estuarine and coastal sediments using metabarcoding and the evaluation of its response to a gradient of pollution is required so that a biotic index that can be routinely integrated within monitoring programs can be developed.

The abovementioned biases and unresolved issues for developing macroinvertebrate and bacterial community metabarcoding-based biotic indices need to be solved before applying this approach in regular biomonitoring surveys. Thus, developing and testing protocols and robust methods that provide solutions to these shortcomings is essential to allow policy questions to be answered rapidly and reliably and will improve the knowledge for performing integrative assessments of marine waters under an ecosystem approach in the context of European directives.

HYPOTHESIS & OBJECTIVES

1. Hypothesis

Considering the gaps for the implementation of metabarcoding in marine environmental monitoring and assessment, the following hypothesis is posed as a basis of this Thesis:

“Metabarcoding-based biomonitoring represents a rapid, accurate and cost-effective alternative to traditional methodologies for environmental monitoring by characterizing macroinvertebrate and bacterial communities, which are or could be used as indicators of ecosystem health to assess marine ecological status,”

2. Objectives

In order to test the hypothesis, the Thesis addresses the following general objective:

“To validate metabarcoding for biomonitoring by characterizing marine benthic macroinvertebrate and bacterial communities by means of developed procedures to implement this approach in the context of current environmental management directives”.

The general objective has been subdivided in a series of operational objectives that are shown below and are addressed in the different Chapters of this Thesis:

1. To set the basics for the calculation of a metabarcoding-based biotic index by analyzing all the genetic information available for the macroinvertebrate species included in AMBI. The following sub-objectives were defined:
   a. To determine the performance of a new biotic index calculated using only the species for which genetic information is available
   b. To determine the minimum reference database size and content required to calculate an accurate biotic index

c. To identify the best primer set to retrieve the most complete representation of the macroinvertebrate taxonomic diversity
d. To increase the size of the reference database by including new DNA sequences of species for which no genetic information is available

2. To develop standardized laboratory procedures for macroinvertebrate sample processing and for bioinformatic analyses of sequence data so that reliability and reproducibility of the approach is ensured. The following sub-objectives were defined:

a. To define laboratory guidelines and detailed steps for extracting good quality and integrity DNA representative of the whole community from benthic sediment samples collected for macroinvertebrate metabarcoding-based biomonitoring
b. To define a bioinformatic pipeline for obtaining the taxonomic composition of benthic macroinvertebrate samples from sequence data generated using HTS platforms
c. To establish guidelines for the use of HTS data-derived taxonomic information to calculate benthic macroinvertebrate-based biotic indices

3. To benchmark metabarcoding-based macroinvertebrate community assessments using samples of known taxonomic composition in order to test the technique for environmental biomonitoring. The following sub-objectives were defined:

a. To compare the accuracy of taxonomic assignments and biotic indices obtained from two different DNA extraction strategies
b. To compare the accuracy of taxonomic assignments and biotic indices obtained using two different DNA barcodes
c. To compare the accuracy of taxonomic assignments and biotic indices obtained from different PCR amplification conditions
d. To compare the accuracy of taxonomic assignments and biotic indices obtained using two different DNA sources

4. To compare metabarcoding and morphology-based biotic indices derived from benthic macroinvertebrate samples of unknown taxonomic composition. The following sub-objectives were defined:

a. To apply the developed protocols for sample processing and data analysis to gather macroinvertebrate taxonomic compositions of several benthic samples simultaneously
b. To compare metabarcoding and morphology derived taxonomic compositions of paired samples collected from the same locations
c. To compare metabarcoding and morphology-based biotic indices of paired samples collected from the same locations

5. To compare the cost-effectiveness of metabarcoding and morphology-based biomonitoring. The following sub-objectives were defined:
a. To compare the time required from sample collection to calculation of the biotic index of metabarcoding and morphology
b. To compare the costs involved from sample collection to calculation of the biotic index of metabarcoding and morphology

6. To explore the potential of bacterial assemblages as indicators of ecosystem health and to develop a biotic index based on the response of this community to a gradient of pollution. The following sub-objectives were defined:
a. To characterize the bacterial community from different sediment samples using metabarcoding
b. To gather the documented response of the different bacterial taxa to pollution
c. To assign ecological groups to the different bacterial taxa in order to associate them with their tolerance to stress
d. To develop a biotic index using the information of the newly classified taxa
e. To test and validate the newly developed biotic index by evaluating it in samples subjected to different pollution pressures

THESIS STRUCTURE

To achieve the objectives defined, this Thesis has been divided into six Chapters, from which the first five focus on the validation of metabarcoding for characterizing macroinvertebrates and the last one focuses on the development of a biotic index using bacterial communities. The work presented in this Thesis has been included in six manuscripts (five of them published in international peer-reviewed journals and one in preparation); each chapter consists on one publication.

Chapter 1 presents the requirements for the implementation of a metabarcoding-based biotic index and shows, using available sequence data, that information about presence/absence of the most frequently occurring species provides accurate biotic index values.

Chapter 2 details standardized procedures for benthic environmental sample collection, processing and homogenizing, and for extracting and preparing DNA for metabarcoding.

Chapter 3 describes the analysis of high-throughput sequence data derived from marine benthic macroinvertebrate metabarcoding and provides guidelines on how sequencing reads should be used for the calculation of benthic indices for environmental monitoring.

Chapter 4 comprehensively benchmarks metabarcoding and morphology -based taxonomic identification and describes how the limitations of metabarcoding should be addressed for the development of a metabarcoding-based biotic index.

Chapter 5 describes the application of the newly developed laboratory and bioinformatics protocols to bulk environmental samples and shows that metabarcoding provides biomonitoring conclusions comparable to those obtained using traditional methodologies, while being more cost-effective.

Chapter 6 classes bacterial taxa according to their tolerance to pollution and makes use of this information to develop and validate a new biotic index that is significantly correlated with a sediment quality index.

The outcomes of this Thesis are integrated in the "General Discussion" section, with conclusions drawn at the "Conclusions" section. "Further recommendations" regarding the use of metabarcoding in marine environmental policy are provided. The "References" and the "Supplementary Material" (figures and tables) sections are provided at the end of this document.

Chapter 1

ECOLOGICAL STATUS ASSESSMENT USING DNA METABARCODING: TOWARDS A GENETICS-BASED MARINE BIOTIC INDEX (gAMBI)

Published as: **Aylagas, E.**, Borja, A., and Rodríguez-Ezpeleta, N. (2014). Environmental status assessment using DNA metabarcoding: towards a genetics-based Marine Biotic Index (gAMBI). PLoS ONE 9(3), e90529. doi: 10.1371/journal.pone.0090529.

1. Introduction

Increasing human activities in seas and oceans are likely to produce impacts on marine ecosystems (Claudet and Fraschetti, 2010; Halpern et al., 2012). Yet, the United Nations Convention on the Law of the Sea (UNCLOS, 1982), further supported by the 1992 Convention on Biological Diversity (CBD, 2000), establishes an international obligation to sustainably use marine resources. Additionally, several national or regional initiatives (e.g. the Australian Oceans Policy, the Canadian Oceans Act and Oceans Strategy, the USA Oceans Act, and the European Water and Marine Strategy Framework Directives (WFD, 2000/60/EC and MSFD, 2008/56/EC)) have been developed to protect, conserve or enhance marine ecosystems. These initiatives rely on the assessment of ecological integrity and marine health status (Borja and Dauer, 2008), which requires adequate and rigorous spatiotemporal monitoring of multiple ecosystem components (De Jonge et al., 2006; Borja et al., 2009a; Borja et al., 2011).

Among the components to be monitored, marine benthic macroinvertebrates are frequently used as indicators of ecosystem health. Benthic indices summarize complex biological information such as community composition in a single number that ranks sites on a scale from good to bad status (Ranasinghe et al., 2012). Numerous different benthic indices have been developed in recent times (Diaz et al., 2004; Pinto et al., 2009), allowing managers to identify impacted sites and decide on habitat restoration measures. One of the most successful indices used worldwide is the AZTI's Marine Biotic Index (AMBI; Borja et al., 2000), which is officially used in many European countries and has been tested in America, Africa, Asia and Oceania (Borja et al., 2009b), where examples of its application can be found (Ranasinghe et al., 2012; Valenca and Santos, 2012).

AMBI is based on abundance-weighted pollution tolerances of the species present in a sample, with tolerance being expressed categorically as one of five ecological groups (sensitive to pressure, indifferent, tolerant, opportunist of second order and opportunist of first order). Currently a list of about 6,000 worldwide species with ecological group assigned is available (http://ambi.azti.es). In addition, (Warwick et al., 2010) and Muxika et al. (2012) have proposed the use of this index based upon presence/absence and biomass of species (i.e. (pa)AMBI and (B)AMBI, respectively). All forms of AMBI require each species to be sorted and identified under a stereomicroscope. This is a time and resource consuming process that has limitations in some cases, as for example when damaged specimens o immature life stages are present (Ranasinghe et al., 2012).

Despite the importance of monitoring and assessment, the current economic crisis is leading some countries to pay attention on their monitoring budgets (Borja and Elliott, 2013). This fact has led researchers to investigate new and cost-effective methods to monitor and assess marine waters (Frolov et al., 2013). Genomic methods are a promising avenue to analyze biological systems, especially due to the recent advent of high-throughput sequencing technologies (Bourlat et al., 2013). Among these methods, DNA barcoding and metabarcoding have the potential to increase speed, accuracy and resolution of species identification, while decreasing its cost in biodiversity monitoring (Ji et al., 2013).

Barcoding consists of taxonomically assigning a specimen based on sequencing a short standardized DNA fragment (barcode). In the metabarcoding approach, the analysis is extended to a community of individuals (of different species) rather to a single individual (Taberlet et al., 2012a; Ji et al., 2013). In both cases, sequences need to be compared to a reference library that contains the correspondence between the barcodes and taxonomical classification. Several studies have used "metabarcoding" to study marine and tropical rainforest meiofauna (Creer et al., 2010), soil fauna (Yang et al., 2013), arthropods (Yu et al., 2012; Ji et al., 2013), zooplankton (Machida and Tsuda, 2010) and fish gut contents (Leray et al., 2013).

The efficiency and accuracy in taxonomic identification using metabarcoding largely depend on the targeted barcode, which should be taxonomically informative (Liu et al., 2008), and primer set used for amplification, which should be adequate for the target species (Leray et al., 2013). Primers can therefore be group specific, if the goal is to describe the diversity of species of a specific taxonomic (i.e. nematodes in sediments; see Creer et al., 2010), or wide range, if the goal is to obtain a comprehensive analysis of samples containing species from numerous phyla (Leray et al., 2013). If required, a cocktail of wide range and group specific primers can be used to cover the comprehensive biodiversity of the samples under study (Prosser et al., 2013).

For animals, the most commonly used barcode is a 658 bp section of the mitochondrial cytochrome c oxidase subunit I gene (COI) (Hebert et al., 2003a). This gene has a faster substitution rate, compared to nuclear rRNA genes, which makes it suitable for species discrimination (Hebert et al., 2003a). Yet, alternatives have been developed for cases when COI sequences are insufficient to distinguish recognized species (Hebert et al., 2003b) or when amplification is challenging (Creer et al., 2010) . Among the alternatives, the nuclear 18S small subunit rRNA (18S rRNA) is the most widely used (Markmann and Tautz, 2005), although other markers such as the nuclear 28S rRNA and the mitochondrial 12S rRNA have also been suggested (Machida and Tsuda, 2010; Machida and Knowlton, 2012).

Attempting a (meta)barcoding approach for the AMBI calculation is challenging as the species that compose the index belong to different taxonomic groups. Searching the appropriate genetic markers and primers for the target organisms is mandatory to cover the maximum spectrum of species within a sample and therefore avoid underestimations. Furthermore, a large enough barcode reference library is needed to comprehensively determine the biodiversity in the samples. In this chapter, we evaluate the potential of an AMBI based on taxonomic identification by (meta)barcoding. For that purpose, we analyze the genetic resources available for the AMBI species, and determine the minimum reference library size and content required to calculate an accurate index. Additionally, we identify the best primers to retrieve the most complete representation of the AMBI taxonomic diversity and provide sequences for 22 species for which no genetic resources were available.

2. Methods

2.1. Datasets: species, sequences and case studies

Species list and assignment into one of the five ecological groups defined by the index were retrieved from the AMBI 5.0 software (http://ambi.azti.es). Taxonomic classification of the 5,977 retrieved soft-bottom macroinvertebrate species was done through the World Register of Marine Species (WoRMS) (www.marinespecies.org) and verified in the European Register of Marine Species (ERMS) (www.marbef.org). Sequences of the mitochondrial cytochrome oxidase I (*COI*) and nuclear 18S ribosomal RNA (*18S rRNA*) genes of the 5,977 species were searched in GenBank database (accession: July 2013) and retrieved when available. The case studies used for subsequent analyses consisted on a subset of

734 samples of soft-bottom macroinvertebrates collected during annual surveys conducted by the Littoral Water Quality Monitoring and Control Network of the Basque Country, northern Spain (Borja et al., 2009c), in 32 and 51 coastal and estuarine stations between 1995 and 2001 and between 2002 and 2011, respectively. From the samples collected, 694 contain at least one individual and are the ones used in further analyses, being the remainder azoic.

2.2. AMBI and (pa)AMBI calculation and agreement measures

AMBI (calculated using the number of individuals of each species) and (pa)AMBI (calculated using presence /absence (pa) of each species ignoring number of individuals) values were calculated based on the proportional occurrences of benthic macrofaunal species among five ecological groups according to the pollution gradient. This gradient ranges from Ecological Group I – species very sensitive to organic enrichment and present under unpolluted conditions, to Ecological Group V – first-order opportunistic species present in pronounced unbalanced situations, and is calculated using the formula: AMBI = (0 × % GI) + (1.5 × % GII) + (3 × % GIII) + (4.5 × % GIV)+ (6 × % GV) / 100, where percentages represent number of individuals (AMBI) or species ((pa)AMBI) of each ecological group (Borja et al., 2000). AMBI and (pa)AMBI values are grouped in categorical pollution levels (i.e. quality classes): “unpolluted” from 0 to 1.2, “slightly polluted” from 1.3 to 3.3, “moderately polluted” from 3.4 to 5, “heavily polluted” from 5.1 to 6 and “extremely polluted” from 6.1 to 7. AMBI 5.0 software and an in-house R script were used for automated (pa)AMBI value calculations.

Cohen´s Kappa (Cohen, 1960) was used to determine the agreement between pollution levels obtained for the same stations but using different species sets. The level of agreement is described using the ranges suggested by Monserud and Leemans for each value of Kappa (Monserud and Leemans, 1992): < 0.05, no agreement; 0.05-0.20, very poor; 0.20-0.40, poor; 0.40-0.55, fair; 0.55-0.70, good; 0.70-0.85, very good; 0.85-0.99, excellent and 1, perfect. In order to determine if the Kappa value obtained with the x most frequent species (x being 10, 25 and 50%) is significantly better than that obtained with the same number of species selected randomly, we subsampled 100 times x species and calculated the (pa)AMBI of each station considering this subset of species. The Kappa of each of the 100 subsets was calculated with respect to the original species list and the confidence interval of the obtained distribution was used to assign a p value to the Kappa obtained with the most frequent species.

2.3. Primer pair analysis

Primers designed to amplify *COI* and *18S rRNA* gene fragments across representative species of marine macroinvertebrates were retrieved from the bibliography (Table S1.1 and Figure S1.1). From the total sequences for *COI* and *18S rRNA* genes retrieved from GenBank, multiple sequences from the same species were removed by applying cd-hit (Niu et al., 2010) separately for each taxa. This program groups sequences according to a similarity threshold (which was set to 0.9 in this case) and selects the longest one as representative of the group.

Predicting the performance of a primer pair against a target sequence requires the putative annealing region of the primer to be present in the sequence. Because some of the retrieved sequences are partial and/or do not include the primer region, not all primer pairs can be tested against all sequences. Therefore, in order to avoid false negatives, we tested each primer pair only on the sequences that contain the putative annealing region. For that purpose we used the *COI* region of the complete mitochondrial gene from *Mytilus galloprovincialis* (Accession number DQ399833) and the *18S rRNA* gene from *Aplysia punctata* (Accession number AJ224919) as reference to determine the most external nucleotide position of each primer for *COI* and *18S rRNA* respectively. Then, each sequence was compared with the reference using BLAST (Altschul et al., 1990) and, for each primer pair, only those included within the primer pair external positions were selected (See Figure S1.1 for regions tested for each primer primer). Additionally, due to the low number of sequences to be tested for *COI*, we retrieved a total of 3,687 complete metazoan mitochondrial genome sequences (all those available) from the NCBI Organelle Genome Resources database (November 2013), from where 84 sequences were selected for the analysis as belonged to species of the AMBI. Each primer pair was evaluated against its correspondent sequence set using PrimerProspector (Walters et al., 2011) with default parameters. For species that contained more than one sequence, if at least one of them amplifies, the species is considered positive for this primer.

2.4. Animal samples, DNA extraction, PCR and sequencing

The stations that, according to the data series, contain the most frequent species were selected for DNA barcoding. For this purpose, specimens were manually separated, *visu* identified and preserved separately in ethanol until DNA extraction. Taxonomic identification was done by experts from the Cultural Society INSUB following the identification protocols accepted and applied by the scientific community. Total genomic DNA from 115 species belonging to 9 phyla (Annelida, Arthropoda, Cnidaria, Echinodermata, Mollusca, Nematoda, Nemertea, Plathyhelminthes and Sipuncula) was extracted from 1 mm^3 of tissue (which in some cases, came from more than one individual) using the Wizard SV 96 Genomic DNA Purification System (Promega) following manufacturer's instructions. The 658 bp region of the *COI* gene was amplified using the forward dgLCO-1490 and the reverse dgHCO-2198 degenerated primer pair (Meyer, 2003). All PCRs were performed in a 20 µl volume containing 1 X PCR buffer with 1.8 mM MgCl2, 3% DMSO, 0.2 mM dNTP, 1.25U TAQ polymerase (ROCHE), 0.4 µM of each primer, and 80–100 ng of DNA template. The thermal cycling conditions were based on (Meyer, 2003) and consisted of 95 °C for 2 minutes; 35 cycles of 95 °C for 40 seconds, 45 °C for 40 seconds, and 72 °C for 60 seconds, followed by a final extension of 72 °C for 7 minutes and a final cooling at 4 °C. PCR products were purified with ExoSAP-IT (AFFYMETRIX) and Sanger sequenced.

3. Results and discussion

3.1. Species-level taxonomic identification but not species abundance is required for a reliable index calculation

AMBI calculation requires that each identified species be assigned to an ecological group based on its taxonomic identification (Borja et al., 2000). Because ecological groups are associated to species names, this taxonomic identification has to be as precise as to determine the species to which the individual belongs. In order to determine if taxonomic identification to higher taxonomic levels (genus, family, class or phylum) would suffice for ecological group assignment and therefore AMBI calculation, we analyzed the distribution of the AMBI species into taxonomic levels and ecological groups (Figure 1.1). Unfortunately, even within the same genus, there exist species belonging to different ecological groups, meaning that the identification to the species level is required for a reliable AMBI calculation.

The calculation of the currently implemented AMBI is based on the number of individuals of each species found in each sample (Borja et al., 2000). Although this information, including species abundance, could be achieved through DNA barcoding of single individuals, this method is much more time consuming and much less cost effective than metabarcoding, which consists on sequencing all individuals present in a sample at once . Yet, the suitability of metabarcoding for gAMBI calculation requires further studies. Ji et al. (Ji et al., 2013) have recently shown that metabarcoding data leads to similar alpha- and beta-diversity estimates than individual taxonomic identification and, therefore, to similar policy conclusions; however, the identification of all species present in a sample with their abundances, required for the implementation of AMBI, from sequence read data is not yet possible (Yu et al., 2012).

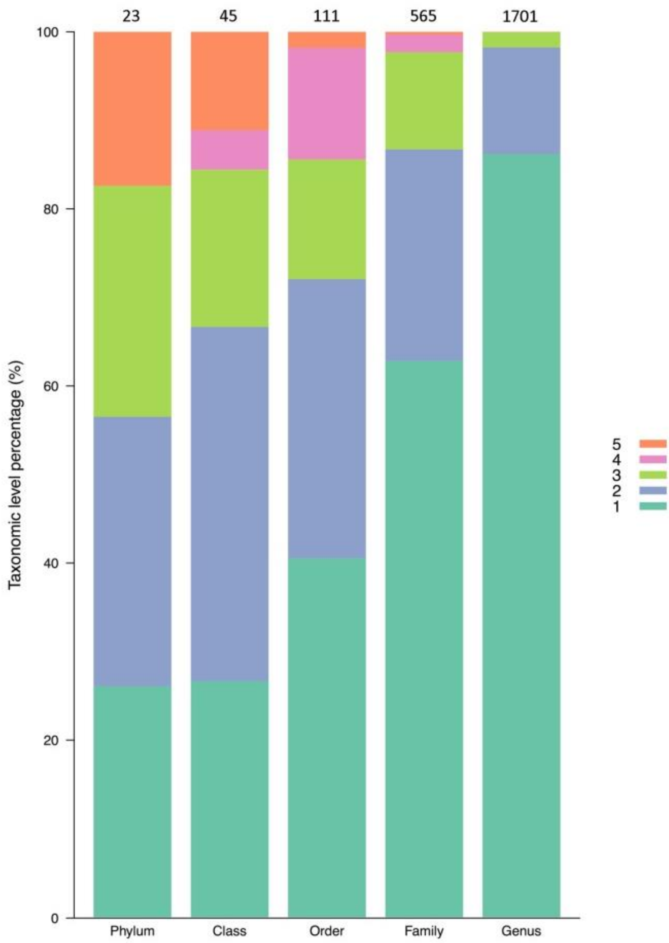

Figure 1.1. Relationship between taxonomic levels and ecological groups. Proportion of taxonomic levels composed by species belonging to the same (1) or different (2,3,4,5) ecological groups. Numbers above bars indicate the total phyla, orders, classes, families and genera and different colors indicate number of different ecological groups.

Biological factors such as multicellularity, variation in tissue cell density, and inter and intra specific variations in gene copy number will lead to different DNA per gram of tissue extracted (Pompanon et al., 2012), making estimation of number of individuals from sequence data impossible. Alternatively, biomass estimations could be used to calculate BAMBI, a version of AMBI based on biomass. Though, several technical factors such as biases during DNA extraction, PCR, pooling, sequencing and bioinformatics sorting (Amend et al., 2010; Porazinska et al., 2010) make estimation of biomass from sequence reads also a difficult challenge. Therefore, it seems that for now genetic data could only provide relevant information to an index that does not rely on species abundance. Fortunately, the (pa)AMBI, based on presence/absence of each occurring species, provides biotic index values that are strongly related to the AMBI values (Muxika et al., 2012). This is also confirmed by our dataset from where we obtain a very good agreement (Kappa k=0.77) between AMBI and (pa)AMBI values (Figure 1.2). Thus, obtaining presence/absence data from genetic analyses is enough for a reliable biotic index calculation.

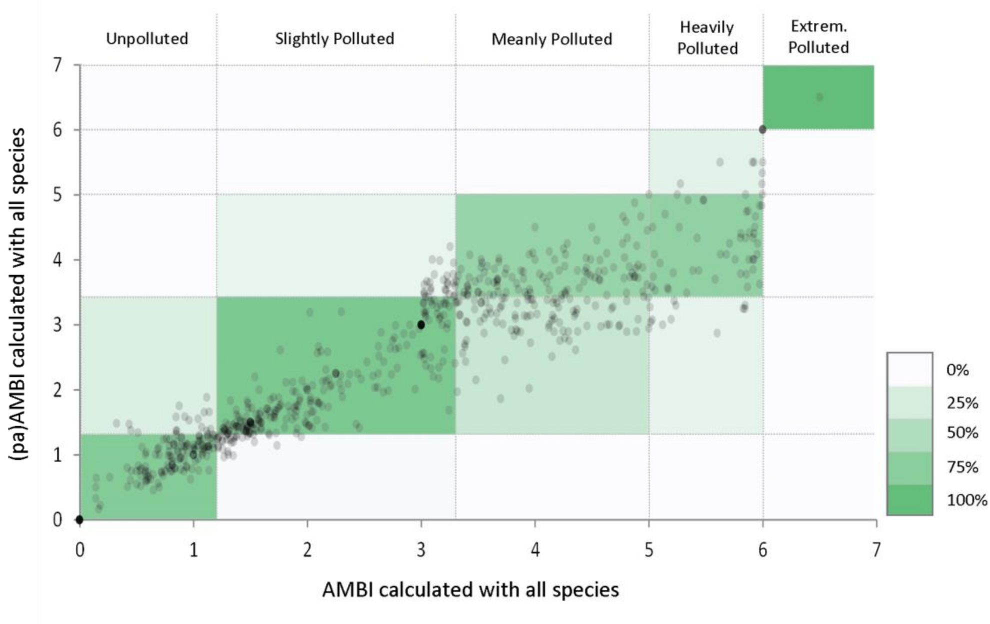

Figure 1.2. Correspondence between AMBI and (pa)AMBI values. Relationship between AMBI and (pa)AMBI values calculated for 694 cases. Vertical and horizontal lines indicate pollution level assessment thresholds. Color scale indicates percentage of agreement for each pollution level, meaning the number of samples that fall in the same category. Dark green color located in the diagonal reflects the best agreement between samples.

3.2. AMBI species classification and available genetic data

From the 5,977 taxa included in the AMBI species list, 90% fall into five phyla: Annelida (2,148 species), Mollusca (1,506 species), Arthropoda (1,448 species), Echinodermata (188 species) and Cnidaria (133 species). The remaining 10% fall into 19 phyla that contain each less than 100 taxa (Figure 1.3). We explored the sequences available in the GenBank database for these species for the most widely used genetic markers for animal barcoding: *COI* and *18S rRNA* (Hebert et al., 2003a; Hebert et al., 2003b; Creer et al., 2010; Hajibabaei et al., 2011). For the former, 15,619 sequences belonging to 855 species were retrieved, whilst for the later, 2,295 sequences belonging to 940 species were retrieved. Among them, 471 species have sequences for both markers. Although the number of species for which *COI* and *18S rRNA* sequences are available is virtually the same, more sequences for the former are available.

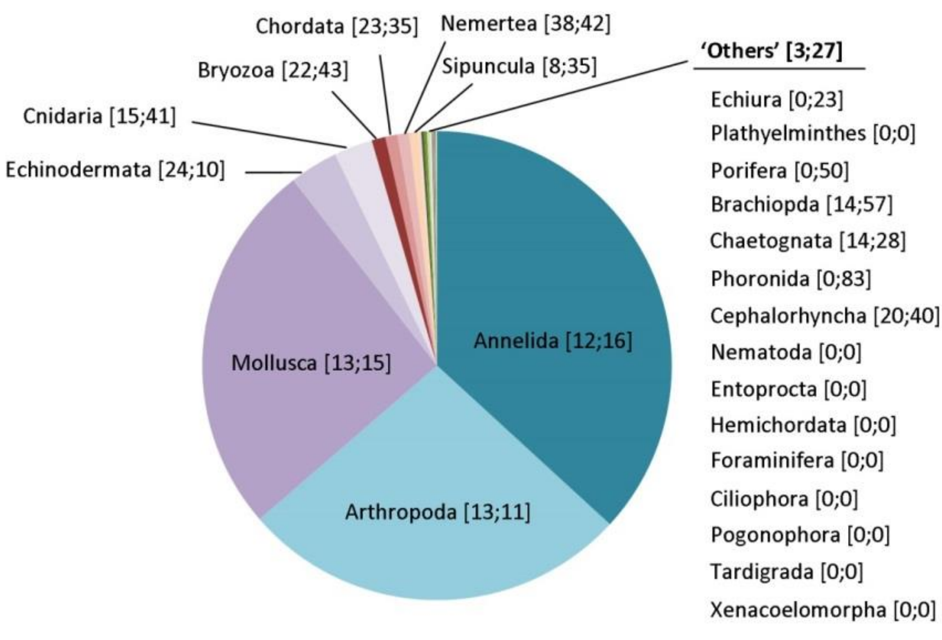

Figure 1.3. AMBI list phyla and available genetic data. Numbers in brackets indicate proportion of sequences for *COI* or *18S rRNA* available for each phylum.

This is due to the popularity of the *COI* marker in barcoding studies (Hebert et al., 2003a) and to the extended used of this gene in molecular systematic studies leading to submission of sequences from the same species spanning different geographical areas (Meyer, 2003; Hardy et al., 2010; Matzen da Silva et al., 2011). Notably, only about 15% of the species included in the AMBI list have *COI* and/or

18S rRNA genes sequenced, which may be insufficient for the implementation of a biotic index based on barcoding or metabarcoding for taxonomic identification.

3.3. Available sequence data is not sufficient to calculate reliable AMBI values

In order to determine if data from only 15% of the species in the AMBI list is sufficient to provide reliable (pa)AMBI values, we gathered data from 694 cases studies (see Methods). The total number of different species found along the total serial data is 924, of which only 143 (15%) and 185 (20%) have *COI* or/and *18S rRNA* sequenced, respectively (note that some species may have sequences for both genes). For each case study, we calculated the (pa)AMBI considering all species and the (pa)AMBI considering only the species with *COI* or *18S rRNA* sequence available (Figure 1.4). The level of agreement between samples is fair (Kappa value of 0.502) for *COI* and poor (Kappa value of 0.244) for *18S rRNA*, meaning that the available genetic data is not sufficient or does not fulfill the requirements for a reliable AMBI calculation.

Ranasinghe et al. (2012) suggested that an even distribution of taxa across the disturbance gradient is needed for a reliable index calculation, condition that is not met by neither the *COI* nor *18S rRNA* datasets. Notably, the distribution of species into ecological groups of the *18S rRNA* dataset is considerably different from that of the whole dataset, being ecological group III predominant (Figure 1.5). This may explain the large number of cases where this dataset yields (pa)AMBI of 3 regardless of the (pa)AMBI values obtained with the whole dataset. Also, the slightly higher agreement obtained with the *COI* dataset, despite being composed by less species may be explained by a more even distribution of the species into ecological groups. Thus, not only the number of species, but their distribution along the different ecological groups affects the reliability in (pa)AMBI values calculation.

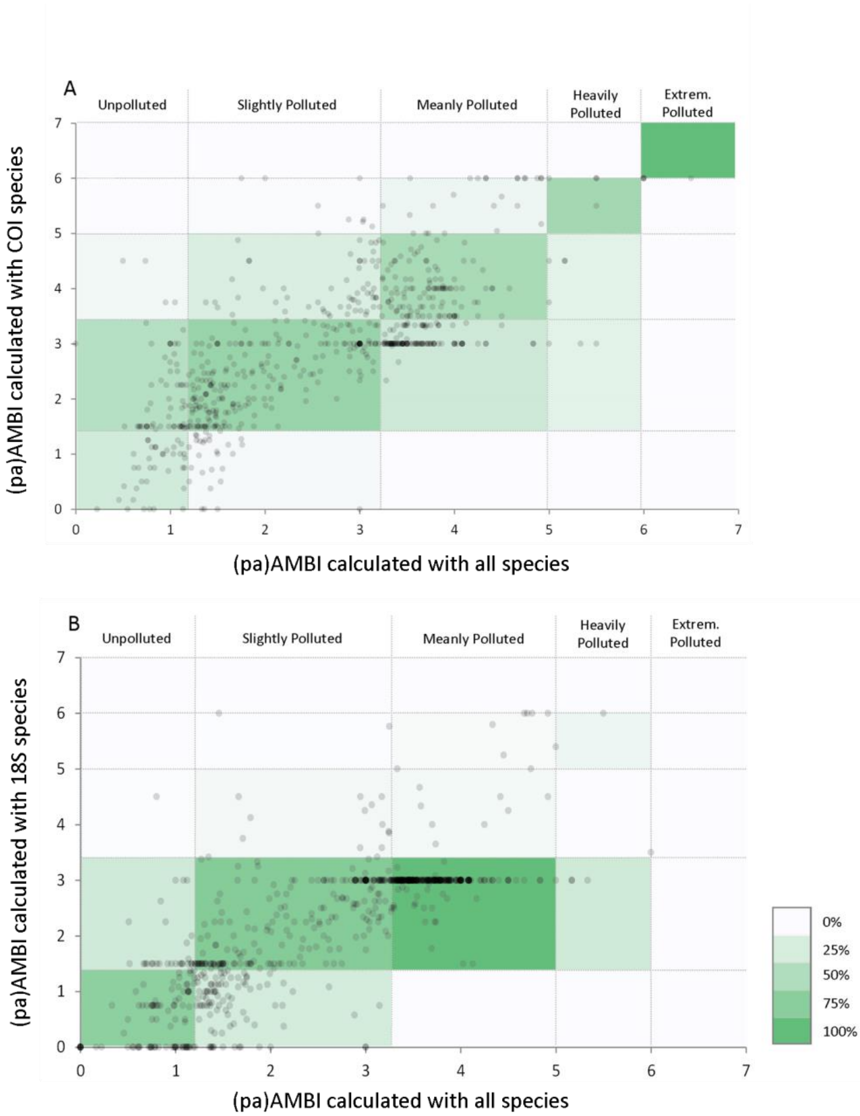

Figure 1.4. (pa)AMBI values calculated with all or with only sequenced species. Relationship between (pa)AMBI calculated with all species and (pa)AMBI calculated with the current (A) *COI* and (B) *18S rRNA* sequenced species. Vertical and horizontal lines indicate assessment thresholds pollution levels. Color scale as in Figure 1.2.

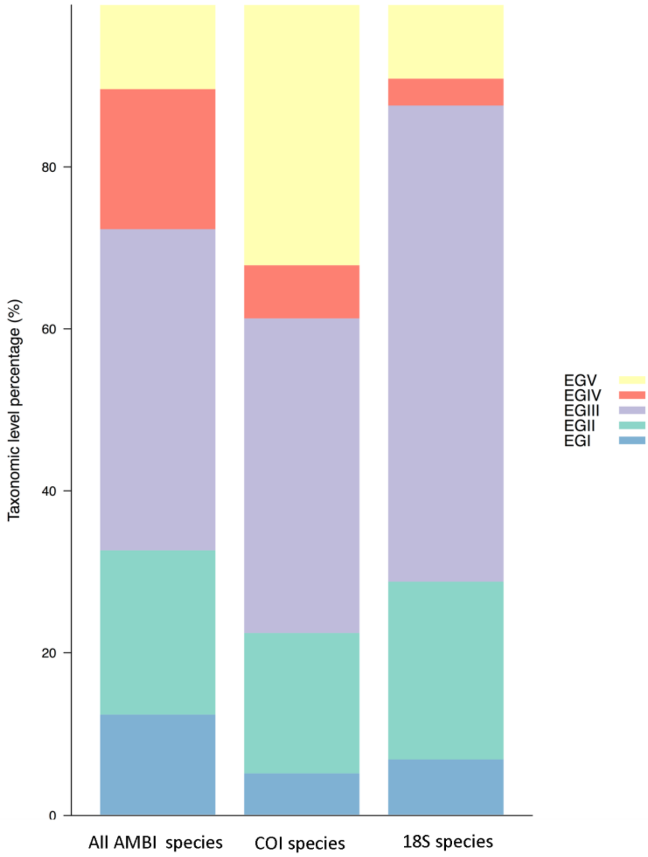

Figure 1.5. Distribution of sequenced taxa along the pollution gradient. Proportion of species, based on frequency, of each ecological group in each dataset (all species, *COI* sequenced species and *18S rRNA* sequenced species).

3.4. How many species are necessary for an accurate AMBI calculation?

In order to determine the minimum number of species required to calculate accurate AMBI values, agreement tests between (pa)AMBI values obtained with the full set of species and (pa)AMBI values calculated with increasing percentages of the most frequent species were performed (Figure 1.6). Obtained Kappa values are very good (0.85 for 10% of the most frequent species) and excellent (0.93 for 25% and 0.98 for 50%). Importantly, the observed agreement is not due to the number of species selected, but to the fact that they are the most frequent ones. That is, the Kappa values obtained when using the same number of randomly selected species are significantly lower than the ones obtained using the most frequent species (p values of 1.44×10^{-5}, 2.03×10^{-5} and 0.0035 for 10%, 25% and 50% respectively). Notably, the distribution of the most frequent species in ecological groups is, in all cases, similar to that of the whole species list (Figure S1.2). Therefore, in order to increase DNA reference library the effort must be focused on barcoding the most frequent species, which can in low number be sufficient to provide reliable (pa)AMBI values.

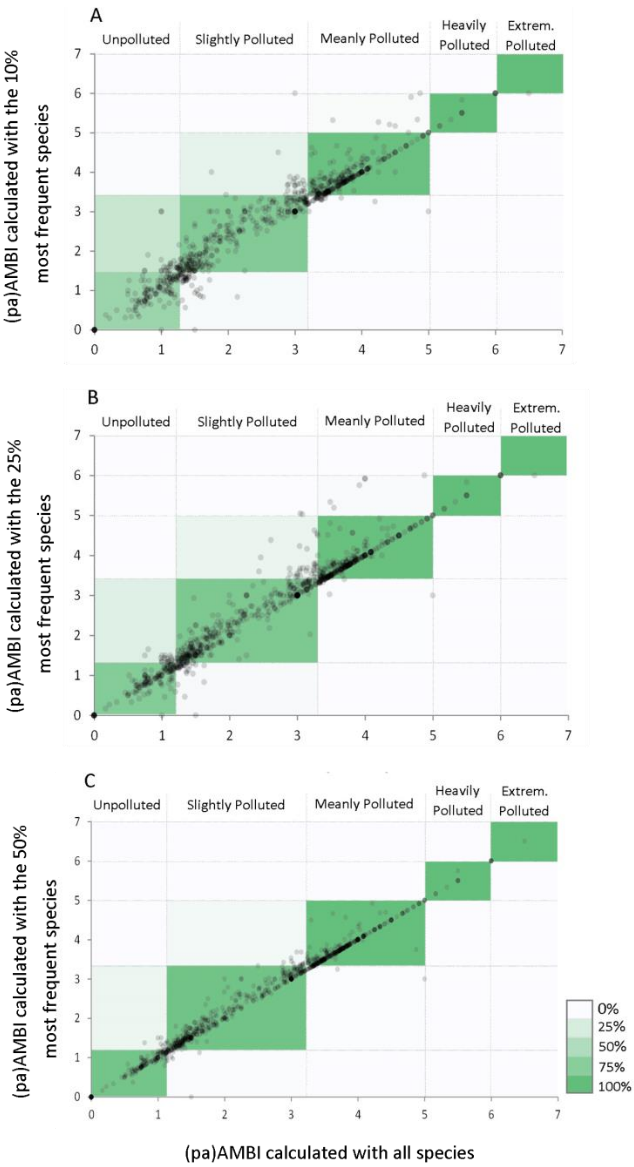

Figure 1.6. (pa)AMBI calculated with all or with the most frequent species. Relationship between (pa)AMBI calculated with all species and (pa)AMBI calculated with the 10% (A) 25% (B) and 50% (C) most frequent species. Vertical and horizontal lines indicate assessment thresholds for pollution levels. Color scale as in Figure 1.2.

3.5. Evaluation of primer pairs: taxonomic coverage

Suitable genetic markers and primers that amplify the largest number of species are necessary to efficiently increase the AMBI species list reference library. We assessed the performance of primer pairs designed to amplify the most used genetic markers for Metazoa, *COI* and *18S rRNA*, in the available sequences from these genes for the species of interest.

Despite the large number of *COI* sequences available, very few include the complete gene sequence (Figure S1.1), limiting primer analysis. Thus, in order to increase the number of sequences tested in the analysis, 84 complete mitochondrial sequences - belonging to 84 species of the AMBI list - were included. Fifteen primer pairs that are included within the 658 bp 'Folmer region' (Folmer et al., 1994; Meyer, 2003) were tested for 15 phyla, from which only Mollusca, Arthropoda, Echinodermata and Annelida had more than 10 sequences (Figure 1.7). For the remaining phyla, less than 10 sequences could be tested. Only one sequence of Hemichordata and Chaetognata was tested for each, from which no amplification was obtained with any of the primer pair (data not shown). Among the primer pairs, jgLCO1490 × jgHCO2198 potentially amplify 80% of the 101 sequences tested; only Mollusca had less than 90% (50%) potentially amplifying species. Primers designed to target a shorter region (319 bp), could be tested for a higher number of species. Among them, mlCOIintF × HCO2198, mlCOIintF × dgHCO2198 and mlCOIintF × jgHCO2198 potentially amplify 9, 12 and 35%, respectively, of the 118 sequences tested.

The difference in performance of these primers could be explained by the presence of more number of degenerated bases in the last one. This could also improve the performance of the dgLCO1490 × dgHCO2198 (Meyer, 2003) pair versus the "traditional" Folmer pair, LCO1490 × HCO2198, although this could not be confirmed with available sequences. Although the lack of complete sequences for *CO1* gene that include the potential primer binding sites limit our analysis, our results confirm that the degenerated primers that cover the complete Folmer region and a shorter region (319 bp) are the best performing ones (Meyer, 2003; Geller et al., 2013; Ji et al., 2013; Leray et al., 2013).

More species could be tested for *18S rRNA* data, although the reduced number of sequences available for some phyla (*e.g*. Cephalorhyncha, Chaetognata, Echinodermata, Echiura, Phoronida and Porifera) limits inferences related to these groups. The highest taxa coverage is shown for the primer pair 18eF × 18lR (Figure 1.8), with 98% of the 118 species tested potentially amplifying; only Echinodermata and Mollusca had less than 100% (75 and 96% respectively) potentially amplifying species. Although apparently less successful in terms of percentage of species potentially amplifying among the tested ones (ranging from 97.1 to 94.2%), the remaining universal primers could be tested in all phyla. In particular, primer pair #3Fx#5_RC has an amplification success of 97.1% and all phyla and almost all species could be tested. Thus, according to our results, primer pair #3Fx#5_RC is the best performing for *18S rRNA* macroinvertebrate amplification. The primer pair selected by other authors as best performing (Machida and Knowlton, 2012) also provides successful amplification rates although slightly lower (94%).

3.6. DNA barcoding of AMBI species

In order to start increasing the reference library for a future gAMBI, we attempted to sequence the *COI* gene fragment amplified with the dgLCO1490 × dgHCO2198 primer pair from the most frequent species. From 115 individuals selected, 56 amplified and 22 gave a sequencing product. The specimens have been submitted to BOLD (http://www.boldsystems.org) with BINs BOLD:AAJ1248, ACJ4563, ACJ4767, ACH4094, ACJ2906, ACG2010, ACJ4318, ACJ2494, ABU8508, ACJ4125, ACJ4592, ACJ4543, ABA9346, ACJ2932, ACJ2637, ACJ2931, ACJ4785, ACJ4313, ACJ2499, ACJ2492, ACJ2498 and ACJ4512; and the sequences deposited in GenBank with accession numbers KF808157 - KF808178. The 22 new sequenced species have been included in the list of sequenced *COI* species for (pa)AMBI calculations. Among them, 8 taxa (*Magelona johnstoni, Urothoe pulchella, Protodorvillea kefersteini, Polygordius appendiculatus, Glycera unicornis, Diogenes pugilator*, *Scolaricia* sp. and *Glycinde nordmanni*) are within the 10% most frequent, 6 (*Ampelisca sarsi*, *Chamelea striatula*, *Phyllodoce lineata*, *Pseudomystides limbata*, *Necallianassa truncata* and *Haplostylus normani*), within the 25% most frequent and 4 (*Hyala vitrea, Sabellaria spinulosa, Bathyporeia tenuipes* and *Paradoneis ilvana*), within the 50% most frequent taxa, whilst 4 taxa (*Thracia phaseolina, Paradoneis sp., Magelona minuta* and *Sthenelais limicola*) are not part of the most frequent species. The level of agreement between (pa)AMBI calculated with all species and (pa)AMBI calculated with *COI* species (included the abovementioned) is good (Kappa value of 0.617), improving the one obtained with the previously available resources for this gene.

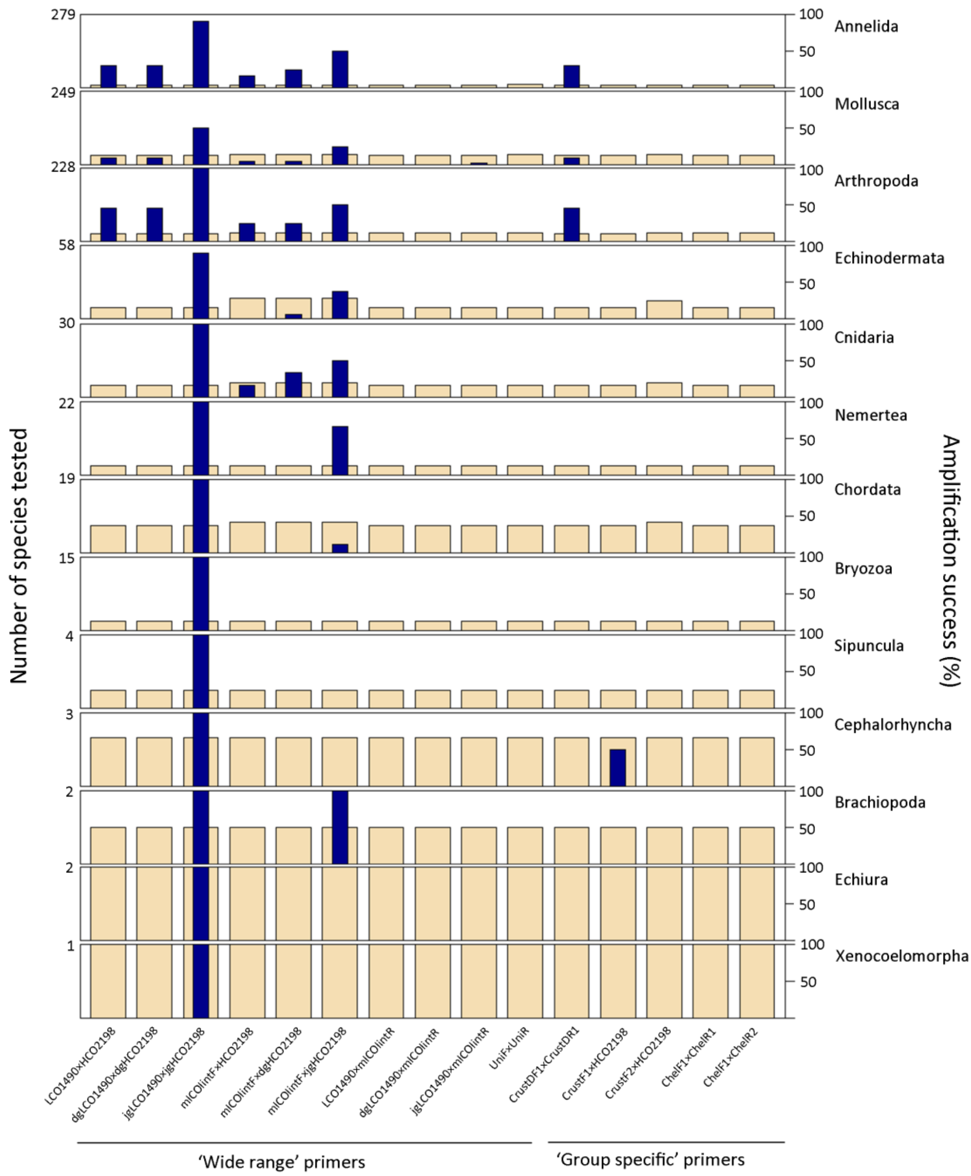

Figure 1.7. Taxa coverage for *COI* primer pairs. Percentage of species potentially amplified for each combination of primer pair and phylum. Wheat color bars represent number of species tested per primer and dark blue color bars percentage of species (within the tested ones) potentially amplified for each primer pair. The maximum value on the left Y axis indicates the total number of species for which *COI* sequence is available per phylum.

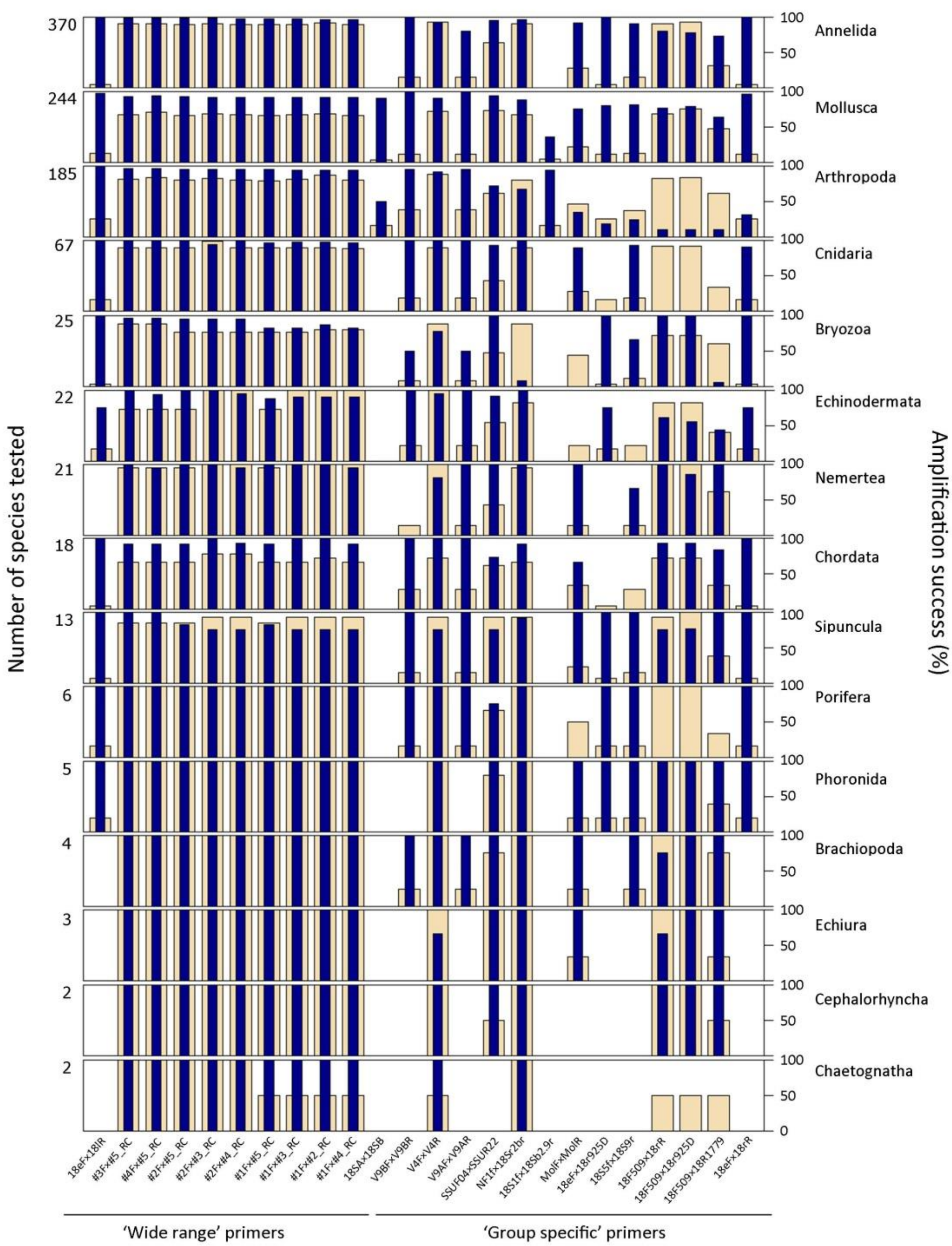

Figure 1.8. Taxa coverage for *18S rRNA* primer pairs. Percentage of species potentially amplified for each combination of primer pair and phylum. Wheat color bars represent number of species tested per primer and dark blue color bars percentage of species (within the tested ones) potentially amplified for each primer pair. The maximum value on the left Y axis indicates the total number of species for which *18S rRNA* sequence is available per phylum.

4. Outlook

Overall, our results place DNA barcoding as a viable alternative to visual species identification in the context of taxonomic assignment for gAMBI calculation; though, this viability is subject to increasing the number of sequences in the reference library. According to our results, this increase should be performed focusing on the most frequently occurring species, as their presence in the reference library, even in a small percentage, is enough for an accurate gAMBI calculation.

Here, we have focused on the use of (meta)barcoding techniques to ease the first step for the calculation of AMBI: taxonomic identification. However, it could be possible to think about a new version of gAMBI based on total biodiversity metabarcoding profile that would not require finding a particular set of species previously defined. Therefore, besides working on increasing the gAMBI reference library, we are also focusing on comparing samples analyzed by visual taxonomy and by metabarcoding in order to explore more practical genetics-based alternatives to AMBI.

Regardless of whether we pursue species or higher taxonomic level identification, increasing the reference library of sequences is mandatory, and even if the cost of doing so depends on many factors, there is no doubt that it will remain significant (Bourlat et al., 2013). Yet, once the initial investment for building the library is made, each individual in a sample can be identified by DNA barcoding per about $5 (Cameron et al., 2006), and a whole sample per about $50 if it is bulk processed by metabarcoding (rough calculation assuming multiplexing 100 samples on the Illumina MiSeq platform and without considering the bioinformatics processing of the data). Needing still optimization of several analytical steps, the optimal cost-efficiency of DNA techniques for taxonomic identification has not yet been achieved, but has already overtaken that of visual identification (Tautz et al., 2003).

Our ultimate goal is to develop genetics-based tools for a cheaper and faster assessment of the marine quality, which is nowadays suffering from methodological and budget limitations (Borja and Elliott, 2013). Besides their cost-efficiency, genomics-based methods allow a rapid and reliable identification of specimens, irrespective of the taxonomic group or available taxonomic expertise.

Showing that a genomics-based AMBI is a viable alternative to a morphological identification-based AMBI, we foresee the use of this index for monitoring regions where no taxonomic expertise and/or sufficient monitoring budget is available.

Chapter 2

MARINE SEDIMENT SAMPLE PRE-PROCESSING FOR MACROINVERTEBRATES METABARCODING: MECHANICAL ENRICHMENT AND HOMOGENIZATION

Published as: **Aylagas, E.**, Mendibil, I., Borja, Á., and Rodríguez-Ezpeleta, N. (2016). Marine Sediment Sample Pre-processing for Macroinvertebrates Metabarcoding: Mechanical Enrichment and Homogenization. Frontiers in Marine Science 3. doi: 10.3389/fmars.2016.00203.

1. Introduction

Biomonitoring has become essential to address changes in the quality of the environment as a response to the several pressures that are threatening marine ecosystems (Halpern et al., 2008). The rapid response of benthic organisms to a range of natural and anthropogenic pressures makes this community a suitable ecological component for marine biomonitoring (Johnston and Roberts, 2009). Above all, macroinvertebrates are widely used to assess environmental quality through the calculation of benthic indices (Diaz et al., 2004; Borja et al., 2015b). Yet, the fast environmental degradation and the necessity of cost-effective methods for biodiversity assessment urge the need of new tools that allow species identification in a much faster way compared to morphological methodologies (Bourlat et al., 2013). The advent of HTS technologies has favored the application of DNA-based biodiversity assessment methods (Creer et al., 2016) and, in particular, DNA metabarcoding has become a promising technique for rapid, accurate and cost-effective taxonomic identification of the benthic macroinvertebrate community in environmental samples (Elbrecht and Leese, 2015; Aylagas et al., 2016a).

DNA metabarcoding involves the amplification of a particular DNA region (barcode) to resolve the total genomic DNA extracted from an environmental sample into distinct taxa, typically species, by using universal primers (Taberlet et al., 2012a). Coupled with HTS, the technique enables the simultaneous identification of the taxonomic composition of several independent samples by matching the unknown amplified DNA barcode to a DNA reference database (ideally, every organism within a sample can be detected). Metabarcoding has been proven useful in the identification of metazoan community composition from a wide variety of aquatic environments (Chariton et al., 2010; Cowart et al., 2015; Dowle et al., 2015; Elbrecht and Leese, 2015; Lejzerowicz et al., 2015; Leray and Knowlton, 2015; Zaiko et al., 2015), and recent studies have proved that the ecological ecosystem condition addressed through the calculation of DNA-based biotic indices is comparable to that inferred using morphological identification (Dowle et al., 2015; Lejzerowicz et al., 2015; Aylagas et al., 2016a). However, metabarcoding is not a fully established methodology for marine monitoring. Therefore standardization of procedures is necessary, which requires of optimized protocols that allow the reliability and reproducibility of the approach. In this sense, significant efforts have been made to standardize different steps of the metabarcoding workflow by addressing the issues regarding to PCR amplification (Aylagas et al., 2016a), barcode region (Carew et al., 2013), primer selection (Leray et al., 2013), library preparation (Bourlat et al., 2016) and bioinformatics analysis for data interpretation (Aylagas and Rodríguez-Ezpeleta, 2016).

A major limitation for environmental DNA metabarcoding studies of benthic macroinvertebrate communities that has not been properly addressed is the manipulation of the sample to be analyzed. Usually, sediment and organic matter carried over using marine benthic community sampling methods result in large sample volume, which needs to be correctly processed so that DNA representing the whole community can be extracted. However, the amount of collected material, the nature of the sample (e.g. mud sediments require different processing than coarse sands) and the size of the target organisms make, in some cases, DNA extraction of the entire sample unfeasible. The requisite of an adequate metabarcoding study is that the sample must be representative of the whole community. Thus, because each sample is different, the pre-processing strategy must be carefully considered in order to retrieve a reliable representation of the macroinvertebrate community. Additionally, routine application of metabarcoding for biomonitoring requires each step of sample collection, handling, pre-processing, DNA extraction and DNA library preparation and sequencing be standardized so that results from different laboratories can be compared and combined (Deiner et al., 2015).

Different approaches can be used to recover DNA from sediment samples. Generally, the size range of the target organisms determines the amount of sediment to be processed and the protocol used (Creer et al., 2016). For studies targeting small size metazoans (e.g. meiofauna), the procedures can rely on extracting DNA from small sediment samples (i.e. 5 gr of sediment) without any pre-processing step (Lejzerowicz et al., 2015), targeting extracellular DNA (Guardiola et al., 2015; Pearman et al., 2016b) or performing some separation via decantation/flotation (Creer et al., 2010). However, when the fraction to be investigated is larger (e.g. macroinvertebrates) samples need first be processed via decantation protocols so that the macroinvertebrate community is separated from the sediment. Recently, Aylagas et al. (2016a) showed that following protocols to target the extracellular DNA from sediment samples, only a small proportion of the macroinvertebrate taxa are retrieved, whilst the isolation of organisms followed by homogenization and DNA extraction reliably characterized the macroinvertebrate community through DNA metabarcoding.

The objective of the present protocol is to extract good quality and integrity DNA from complex environmental samples which is representative of the whole macroinvertebrate community. For that purpose, we present guidelines for the processing of benthic sediment samples collected for metabarcoding-based biomonitoring. We detail the steps necessary to: (i) preserve the benthic sample to ensure DNA integrity, (ii) isolate organic fraction from the sediment by decantation, (iii) homogenize the sample in order to achieve a good community representation, and (iv) extract DNA of good quality and integrity. The efficiency of sediment decantation and homogenization steps detailed in this protocol have previously shown to help providing accurate metabarcoding taxonomic inferences that are comparable to those inferred from morphology (Leray and Knowlton, 2015). Thus, followed by the well-established metabarcoding procedures for library preparation (Bourlat et al., 2016) and bioinformatics analysis (Aylagas and Rodríguez-Ezpeleta, 2016) this protocol represents the first steps of the procedure to gather the taxonomic list of several benthic samples simultaneously. This information can be ultimately used for a variety of applications that rely on the macroinvertebrate community characterization of the samples such as the calculation of benthic indices for ecological status assessment (Aylagas and Rodríguez-Ezpeleta, 2016), the detection of non-indigenous species (Zaiko et al., 2015) or large-scale spatio-temporal biodiversity assessments (Leray and Knowlton, 2015; Chain et al., 2016). Finally, a Notes section is dedicated to discuss various artefacts and pitfalls to consider throughout the description of the protocol.

2. Materials and Equipment

2.1. Sample collection and preservation

1. Gloves
2. 0.5 m^2 sampling squares
3. Van Veen grab (0.07 – 0.1 m^2)
4. 1 mm mesh size sieve (45 cm diameter)
5. Ethanol 96%
6. 1 L storing flasks
7. Spatula

2.2. Sample processing

Decantation

8. Graduated cylinder with stopper (500 ml, 1 L, 2 L)
9. Deionized water
10. 1 mm mesh size sieve (20 cm diameter)
11. Tweezers
12. Stereomicroscope
13. Milli-Q water
14. Ethanol 96%

Homogenization and DNA extraction

15. Blender (PHILIPS hr2095 700W 2 L glass jar) for large volume samples or porcelain mortar (Thermo Scientific) for small volume samples
16. 50 ml falcon tubes
17. Ethanol 96%
18. 20 µm mesh size filter
19. Spatula
20. Mo Bio PowerMax® Soil DNA Isolation Kit (for large volume samples) or Mo Bio PowerSoil® DNA Isolation Kit (for small volume samples)
21. Proteinase K (20 mg/ml)
22. Shaking incubator
23. Water bath

2.3. DNA overall quality assessment, purification and normalization

24. Agarose
25. SYBR® Safe DNA Gel Stain (Thermo Scientific)
26. HyperLadder™ 1 kbp (BIOLINE)
27. Electrophoresis equipment
28. Nanodrop® ND-1000 (Thermo Scientific)
29. Qubit dsDNA HS Assay Kit (Thermo Scientific)
30. 1.5 ml Eppendorf tubes
31. Mo Bio PowerClean Pro DNA Clean-Up Kit
32. MilliQ water

3. Procedures

3.1. Sample collection and preservation

DNA-free materials thoroughly cleaned between locations must be used to avoid cross-contamination (*see Note 1*), and samples should be preserved under appropriate conditions to guarantee DNA integrity.

1. Collect soft benthic samples using 0.5 m^2 sampling squares in intertidal locations concurring with the low tide or using a van Veen grab from a boat on sublittoral stations.
2. Pass through a 1 mm mesh size sieve.
3. Preserve the retained material in 96% ethanol (*see Note 2*) in a 5:1 volumetric ratio using 1 L flask and store at 4 ºC until further analysis (*see Note 3a: Safe stopping point*).

3.2. Sample processing

Decantation (*0.5 h*)

Humic substances, co-extracted with DNA, inhibit enzymes such as the Taq Polymerase used in PCR reactions to amplify DNA, representing the primary inhibitory compound associated with sediment samples (Matheson et al., 2010). This inhibition represents a potential bias for DNA metabarcoding studies performed on sediment samples and, if not properly addressed, can lead to generation of false negative results (Thomsen and Willerslev, 2015). At the same time, the heterogenic composition of the benthic macroinvertebrate community would require extracting all DNA within a sample in order to detect all species present. As this step is logistically unfeasible, the homogenization of the sample is required, so that a subsample is representative of the whole community. The volume of sediment processed may significantly vary among samples, which could imply a great impact on the sample representativeness. In this sense, low amounts of sediment in the sample allow for more representative homogenized subsamples. For these reasons, it is recommended to separate the organic fraction from the sediment before proceeding with DNA extraction. Depending on sediment type (Figure 2.1), this separation can be totally or partially performed through a decantation process. Medium to coarse grain sediments can often be completely removed through decantation but muddy or fine sediments may decant with the organic matter and impede the complete sediment removal. The sample processing workflow is shown in Figure 2.2.

1. Transfer each sample into a graduated cylinder up to ¼. For 50 to 200 ml volume samples use the 500 ml cylinder; for 200 to 500 ml, the 1 L; and for 500 to 2 L the 2 L graduate cylinder.
2. Fill up with deionized water, cover the cylinder and shake vigorously to resuspend animals and other organic matter.
3. After 5 seconds or when the sediment has been deposited on the bottom of the flask, gently pour the water with the suspended matter onto a 1 mm mesh size sieve so that resuspended organic material decants onto the sieve and the sediment is retained in the cylinder.
4. Repeat steps 2 and 3 five times or until no organic particles can be observed after shaking.
5. Collect the organic material into the corner of the sieve and pour into a blender-jar containing ethanol 96% or into a mortar (Figure 2.2). Large amounts of recovered material (i.e. organisms together with a fraction of organic matter) require sampling homogenization using a blender unit that allows big volume sample processing. In contrast, samples from sediments with low amount of organic matter allow the successful isolation of organisms which can be easily homogenized using a mortar.
6. Check sieve under a stereomicroscope for attached animals and examine sediment for remaining shelled organisms that are not separated through decantation (e.g. bivalves, gastropods); recover with the help of tweezers and add to the previously decanted material (*see Note 3b: Safe stopping point*).

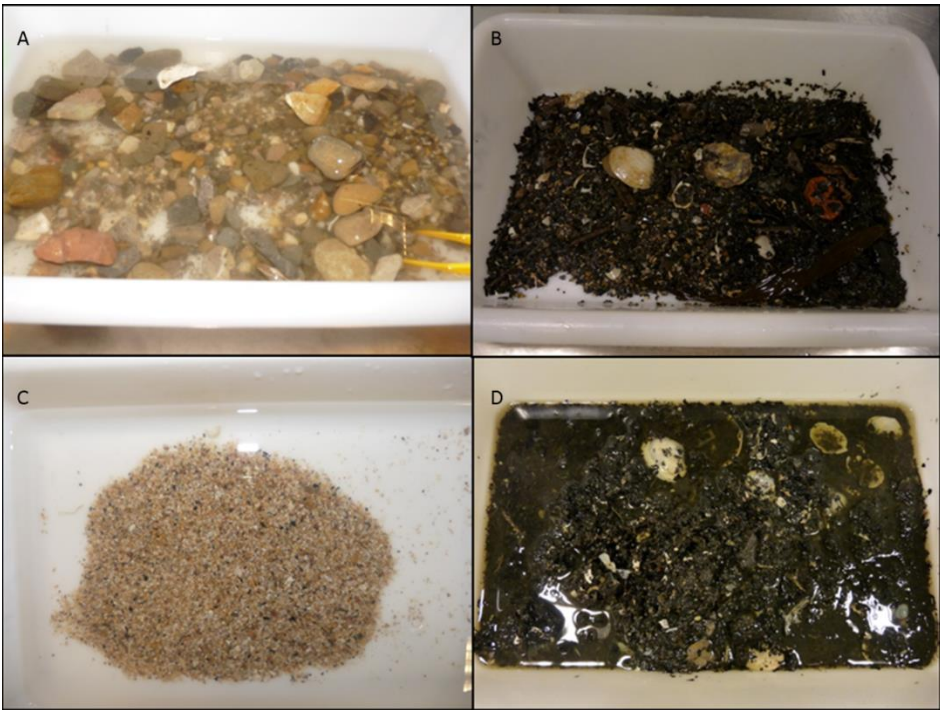

Figure 2.1. Different types of sediment samples collected from intertidal and sub-littoral benthic environments. A) Coarse Sands, B) Medium Sands, C) Fine sands and D) Mud.

Homogenization and DNA extraction *(2 h, Overnight and 3 h)*

The biomass of the decanted organic material may greatly differ among samples, which predetermines subsequent sample pre-processing and DNA extraction procedures. Large amounts of organic material recovered (i.e. the recovered material contains macroinvertebrates and lots of organic matter or big-sized organisms) are followed by Blender homogenization and DNA extraction using the PowerMax Soil DNA Isolation kit; conversely, samples with a range of recovered biomass from 10 – 200 mg (i.e. the recovered material contains animals for the most part) are processed using Mortar homogenization followed by DNA extraction using the PowerSoil DNA Isolation Kit (see Figure 2.2 for schematic representation of the workflow).

Blender homogenization

1. Homogenize the sample until no fragments of animals and other organic material can be observed in the final homogenate.
2. Pour the material through a 20 µm sieve to remove the ethanol and mix the blended material using a spatula. Rinse using ethanol until no material remains in the blender jar.
3. Take two subsamples of 10 gr from the homogenized sample and preserve the remaining material in a flask with ethanol 96% in a 5:1 volumetric ratio using 50 ml falcon tube and store at -20 ºC (*see Note 3c: Safe stopping point*).
4. Extract DNA from each of the two subsamples (*see Note 4*) using the PowerMax Soil DNA Isolation kit following manufacturer´s instructions but replacing the initial bead-beating step by adding proteinase K (0.4 mg/ml) to the power bead solution and incubating samples in a shaking incubator overnight at 56 ºC (Leray and Knowlton, 2015).

Mortar homogenization

1. Pour isolated organisms through a 20 µm sieve to remove the ethanol if sample has been stored before homogenization and place in a mortar.
2. Homogenize animals for 5 minutes or until a mixture has been formed and collect homogenized material in 2 ml Eppendorf tubes (*see Note 3c: Safe stopping point*).
3. Extract DNA from whole homogenate or from a subsample of up to 25 mg using the PowerSoil DNA Isolation Kit following manufacturer´s instructions but replacing the initial bead-beating step, by adding proteinase K (0.4 mg/ml) to the power bead solution and incubating samples in a shaking incubator overnight at 56 °C (Leray and Knowlton, 2015).

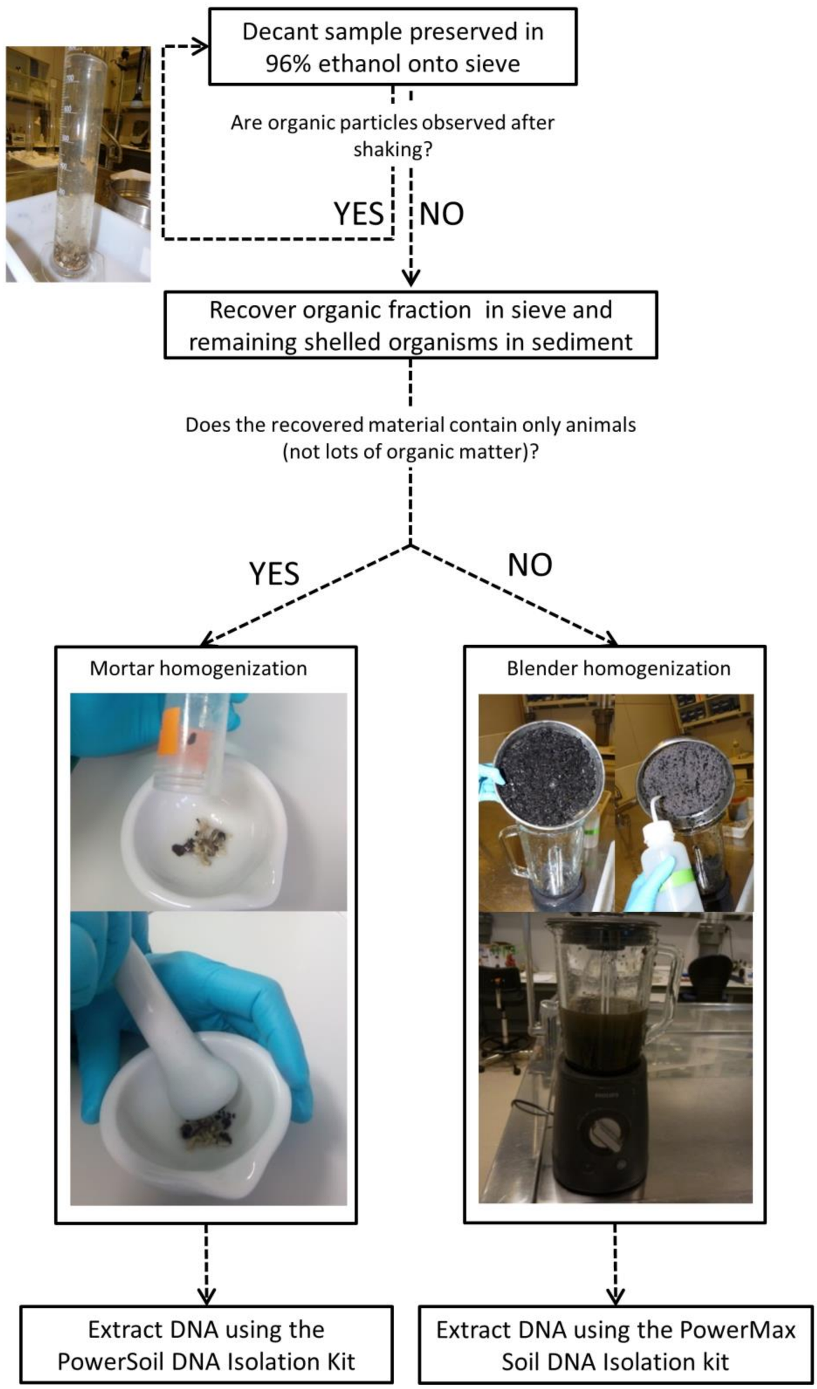

Figure 2.2. Illustration of workflow for bulk sample processing

3.3. DNA overall quality assessment, purification and normalization *(3 h)*

1. Assess DNA integrity migrating about 100 ng of DNA on an agarose 1.0% gel stained with SYBR® Safe (Figure 2.3), purity using the Nanodrop® ND-100 system, and quantity using a Qubit® 2.0 Fluorometer with the Qubit® dsDNA HS Assay Kit.
2. Pool the same amount of DNA derived from each extraction replicate in a single tube.
3. Purify DNA using PowerClean Pro DNA Clean-Up Kit following manufacturer´s instructions (*see Note 5*).
4. Normalize DNA at 5 ng/µl using milliQ water (*see Note 3d: Safe stopping point*)
5. Use DNA as a template for downstream analysis.

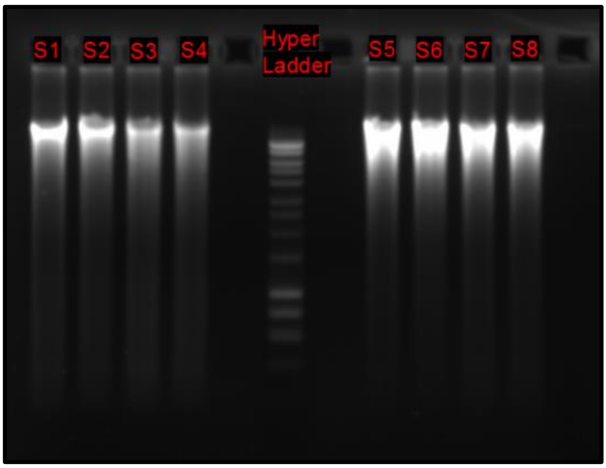

Figure 2.3. DNA integrity of 8 environmental samples processed as described in the present protocol. DNA extraction was performed using the PowerMax Soil DNA Isolation Kit. HyperLadder™ 1 kbp

4. Anticipated Results

The protocol described here provides guidelines to resolve the first steps needed for metabarcoding-based benthic macroinvertebrate community assessment: sample collection, preservation and processing, and extraction of representative DNA of good quality and integrity. The standardization of these three steps is crucial to further obtaining accurate taxonomic inferences from metabarcoding data.

Macroinvertebrate samples used for benthic monitoring can occur in different types of sediment (coarse, medium and fine sands, and muds), and contain organisms of heterogeneous size (from 1 mm to several cm) and nature (soft or containing hard, shell or spiny calcium carbonate exoskeleton, gelatinous, etc.), which implies that DNA extraction may not be equally effective for all types of sediment or organismal types. Our protocol is based on large sediment volumes (> 100 ml) to ensure that all organisms are present, preserved in appropriate conditions to prevent DNA degradation, that are mortar or blender beaten to ensure breaking of hard exoskeletons.

DNA extracted from complex environmental samples need to be representative and of good quality and integrity. The steps presented here ensure both (i) macroinvertebrate community representation by homogenizing samples from which subsamples are taken before DNA extraction, and (ii) good quality and integrity DNA by utilizing kits-based extraction protocols specifically designed for isolating high-quality environmental DNA from soil or sediment. The procedures described in the present protocol for decantation, homogenization and DNA extraction have been recently applied to sediment samples from estuarine and coastal locations with different level of anthropogenic pressures. The DNA extracted from each environmental sample was amplified following the protocol for amplicon library preparation and sequencing (Bourlat et al., 2016) and the resulting reads analyzed using the pipeline for bioinformatics analysis of metabarcoding data (Aylagas and Rodríguez-Ezpeleta, 2016). Using the retrieved macroinvertebrate taxonomic list from each sample, the marine biotic index AMBI (Borja et al., 2000) was calculated, showing comparable results to that inferred using morphological species identification from samples of the same locations (Chapter 5). Thus, the promising results obtained using the present protocol for environmental biomonitoring contributes to accelerating the implementation of metabarcoding for ecological status assessment.

Finally, in response to the necessity of more cost-effective approaches than the traditional morphological species identification, the present protocol followed by DNA amplification coupled with HTS proves to be a suitable cheaper alternative for biodiversity assessment. Although several procedures involving less sample manipulation prior DNA extraction are well-established for small metazoans metabarcoding studies (Guardiola et al., 2015; Lejzerowicz et al., 2015; Pearman et al., 2016b), these approaches cannot be accommodated for macroinvertebrates. In this context, the standardization of the sample pre-processing through mechanical enrichment and homogenization before DNA extraction will ensure the reproducibility of the results and may help to the establishment of macroinvertebrates metabarcoding for environmental biomonitoring.

5. NOTES

Note 1. Recommendations to prevent cross-contamination

DNA-based approach to characterize metazoan communities is very sensitive to contamination. Avoiding cross-contamination is essential to ensure the success of DNA metabarcoding-based biodiversity studies. During sample collection, decantation and homogenization steps, material (sieves, graduated cylinders, blender jar, mortar and tweezers) must be cleaned between samples by soaking in 10% bleach for a minimum of 5 min and gently rinsing with deionized water. Finally, these recommendations must be followed:

- The working area must be cleared and previously cleaned using 10% bleach
- Gloves and lab coat must be worn during manipulation of samples
- Pre and post-amplification laboratory areas should be differentiated
- Sterile filter pipette tips must be used and changed between samples

Note 2. Environmental sample preservation for DNA-based studies

DNA degradation is critical for metabarcoding marine benthic community assessment. In this sense, the detection of some of the species present in an environmental sample may be reduced if DNA integrity has been altered. The process of DNA degradation starts at the moment an organism dies, when cell membranes break and allow entrance of bacteria and other threats with the subsequent release of DNAses that degrade DNA. Thus, avoiding DNA degradation requires storing the sample as soon as collected in appropriate preserving agents (ethanol or other reagents such as RNA later) that prevent DNAse activity (Rodriguez-Ezpeleta et al., 2013). Although formalin has traditionally been used to store marine benthic organism samples, as it preserves morphological structure and allows visual identification, it is toxic and degrades DNA (Serth et al., 2000); thus, ethanol 96% is recommended to preserve samples for molecular studies (Stein et al., 2013).

Note 3. Safe stopping points

a. If sample processing is not immediately performed, bulk benthic sample must be preserved in ethanol at 4 °C until further use (Stein et al., 2013).

b. If homogenization is not immediately performed, pour decanted material into a 2 ml Eppendorf tube, a 50 ml falcon tube or a 1 L flask (depending on the amount of recovered material) containing ethanol 96% and store at -20 °C until homogenization.

c. If DNA extraction is not immediately performed, store homogenized sample in a falcon tube containing ethanol 96% at -20 °C until DNA extraction.

d. Preserve DNA at -20 °C for downstream analysis.

Note 4. Subsample representativeness

Homogenization is performed in order to solve the problem of representativeness issues in large volume samples from which the whole macroinvertebrate community is aimed to be characterized. The best community characterization using DNA-based approaches would require the DNA extraction of the total sample; yet, this cannot be achieved in a reasonable time and commercial kits are not designed for samples up to 10 g. Therefore, a good homogenization step is crucial to ensure the representativeness of the whole community in a subsample. However, we recommend performing two DNA extractions on two subsamples from the homogenized sample to further guarantee a reliable representation of the whole community. In order to ease following steps of the protocol, the DNA replicates are pooled and purified prior amplicon library preparation. Finally, one of the issues related with metabarcoding of different size organisms (from 1 mm to several cm) is the homogenization of exceptionally large specimens with the remaining sample. The

DNA of large organisms may mask the presence of other biota in the sample, which may lead to false negative results. In this case, body parts from large specimens can be subsampled or set aside for standard DNA barcoding.

Note 5. Recommendation to avoid inhibition issues related to humic substances

Even though DNA extraction kits used in this protocol are appropriate to remove humic substances, applying cleaning columns further removes other potential PCR inhibitors such as calcium carbonates, silicates, proteins and algal polysaccharides.

Chapter 3

ANALYSIS OF ILLUMINA MISEQ METABARCODING DATA: APPLICATION TO BENTHIC INDICES FOR ENVIRONMENTAL MONITORING

Published as: **Aylagas, E.**, and Rodríguez-Ezpeleta, N. (2016). "Analysis of Illumina MiSeq amplicon reads: application to benthic indices for environmental monitoring" *in Marine Genomics Methods and Protocols,* Methods in Molecular Biology, ed. Bourlat, S.J. (Springer, New York), vol. 1452. doi: 10.1007/978-1-4939-3774-5

1. Introduction

Metabarcoding, the simultaneous amplification of a standardized DNA fragment specific for a species from the total DNA extracted from an environmental sample, allows the rapid, accurate and cost-effective identification of the entire taxonomic composition of thousands of samples simultaneously (Zepeda Mendoza et al., 2015). This is particularly relevant for monitoring programs relying on the application of benthic indices, which are based on indicator species or ecological groups of species classified according to their sensitivity to stress (Aylagas et al., 2014). Implementation of metabarcoding in regular monitoring programs requires both standardized laboratory and data analysis procedures so that results across studies can be compared (Tedersoo et al., 2015). Here, we describe the data analysis procedures developed to derive the benthic macroinvertebrate taxonomic composition of an environmental sample from MiSeq amplicon reads such as the ones generated using the protocols described in Fonseca and Lallias (2016), Bourlat et al. (2016) and Leray et al. (2016). We will focus on barcodes based on two regions of the most commonly used gene for Metazoa, the mitochondrial cytochrome oxidase I (COI) (Hebert et al., 2003b): a "long region" of 658 bp amplified using the LCO1490 - HCO2198 (Folmer et al., 1994), dgLCO1490 - dgHCO2198 (Meyer, 2003) or jgLCO1490 - jgHCO2198 (Geller et al., 2013) primer pairs, and a "short region" of 313 bp amplified using the mlCOIintF (Leray et al., 2013) forward primer with the HCO2198, dgHCO2198 or jgHCO2198 reverse primers. The analysis for the long region is especially challenging as, unlike in the short region, the reads do not overlap, which requires additional read and database preparation steps.

Additionally, we describe the application of Illumina MiSeq amplicon analysis to environmental monitoring based on benthic macroinvertebrate indices. One of the most successful biotic indices used worldwide is AMBI, which uses marine benthic macroinvertebrates as indicators of ecosystem health (Borja et al., 2000). Calculation of the currently implemented AMBI is based on abundance-weighted pollution tolerances of the species present in a sample (tolerance expressed categorically as one of five ecological groups - sensitive to pressure, indifferent, tolerant, opportunist of second order and opportunist of first order). However, estimating abundances from sequence data is difficult due to biological factors such as multicellularity, variation in tissue cell density, inter and intra specific variations in gene copy number, and technical limitations such as PCR biases (some sequences are amplified more than others) and PCR and sequencing errors (Yu et al., 2012). Thus, biodiversity estimation of the species present in a sample using sequencing data should rely on presence-absence metrics (Elbrecht and Leese, 2015), such as the (pa)AMBI, based on presence/absence of each species and providing biotic index values that are strongly related to the AMBI values (Aylagas et al., 2014).

2. Materials

2.1. Sequencing reads

We assume that 300 bp long forward and reverse sequence reads are provided by the sequencing facility, and demultiplexed based on the barcodes assigned to each sample as described in Bourlat et al. (2016) and Leray et al. (2016). There should be two files per sample in compressed fastq format, usually with extension ".fastq.gz".

2.2. Software

All analyses described in the methods sections 3.1 to 3.4 are performed on a Unix-based environment. The programs listed below need to be previously installed in the system:

1. FastQC (Andrews, 2010): http://www.bioinformatics.babraham.ac.uk/projects/fastqc
2. Trimmomatic (Bolger et al., 2014): http://www.usadellab.org/cms/?page=trimmomatic
3. FLASH (Magoč and Salzberg, 2011): http://ccb.jhu.edu/software/FLASH
4. mothur (Schloss, 2009): http://www.mothur.org
5. Cd-hit (Li and Godzik, 2006): http://weizhong-lab.ucsd.edu/cd-hit/

For section 3.5, the AMBI software needs to be installed on a Windows environment.

6. AMBI (Borja et al., 2000): http://ambi.azti.es

2.3. Databases

A database that contains the correspondence between each taxon and its barcode is needed for taxonomic assignment. Here, we will use the most complete and curated database for the CO1 marker, the BOLD database. Generating a formatted database with all CO1 barcodes requires the retrieval of aligned sequences and taxonomy files from BOLD (http://www.boldsystems.org) using an existing account (*see Note 1*).

1. Aligned sequences are retrieved by searching "Public records" from the "Workbench" section using the option "Let BOLD align my sequences" (*see Note 2*).
2. Taxonomy files (in TSV format) are retrieved by using the "Access Published & Released Data" from the taxonomy browser.

3. Methods

Taxonomic assignment of reads is described in section 3.4 and is based on the MiSeq SOP tutorial (Kozich et al., 2013) of mothur. This tutorial starts with the raw reads; however, due to the nature of our data (*i.e.* non-overlapping forward and reverse reads), the need for a custom database and the fact that this tutorial does not consider quality scores, we have introduced a preprocessing step of the raw data described in sections 3.1 and 3.2 for the CO1 short and long regions respectively (see *Note 3*), and a database preparation step described in section 3.3. In section 3.5, we describe the calculation of benthic indices based on the taxonomic assignment of amplicon reads.

Throughout the methods section, "$" indicates Unix commands run in the terminal window, whereas "mothur>" indicates commands run inside mothur (*see Note 4*).

3.1. Preparation of reads for analysis of the COI "short region"

The "short region" amplicons are 313 bp long, meaning that, with 300 bp long MiSeq forward and reverse reads, an overlap of 237 bp is expected.

1. Check the quality of the reads using FastQC:

```
$ fastqc S1_R1.fastq.gz S1_R2.fastq.gz
```

This will generate a .fastqc.html file for each forward and reverse file that can be visualized in any web browser. The plots generated contain relevant information on the library preparation process and sequence quality (see the FastQC documentation for more information). If everything looks as expected, continue to the next step (*see Note 5*).

2. Remove primer sequences (the first 26 bases of the forward and reverse reads, see Note 6) using Trimmomatic:

```
$ trimmomatic SE -phred33 -trimlog S1_R1.logfile
S1_R1.fastq.gz S1_R1_crop.fastq.gz HEADCROP:26
$ trimmomatic SE -phred33 -trimlog S1_R2.logfile
S1_R2.fastq.gz S1_R2_crop.fastq.gz HEADCROP:26
```

This will result in two output files, S1_R1_crop.fastq and S1_R2_crop.fastq, that contain the forward and reverse reads without the primer sequence.

3. Merge the forward and reverse reads with a minimum and maximum required overlap length between two reads of 217 and 257 bp, respectively (see Note 7):

```
$ flash S1_R1_crop.fastq.gz S1_R2_crop.fastq.gz
-M 257 -m 217 -o S1 -z
```

This will generate five output files: S1.hist and S1.histogram that contain numeric and visual histograms of merged read lengths, S1.extendedFrags.fastq.gz that contains the merged reads, and S1.notCombined_1.fastq.gz and S1.notCombined_2.fastq.gz that contain the forward and reverse reads that were not merged respectively.

4. Remove reads that have an average quality (Phred score) below 25 using the SLIDINGWINDOW option in Trimmomatic and choosing as window length the total length of the amplicon (*see Note 8*):

```
$ trimmomatic SE -phred33 -trimlog
S1_extendedFrags_trimmed.logfile
S1.extendedFrags.fastq.gz S1_ready.fastq.gz
SLIDINGWINDOW:313:25
```

This will generate an output file (S1_ready.fastq.gz) that contains only the reads with an average Phred score above 25.

5. Uncompress the S1_ready.fastq.gz file and transform it into a fasta file using mothur:

```
$ gunzip S1_ready.fastq.gz
$ mothur "#fastq.info(fastq=S1_ready.fastq)"
```

This will generate the S1_ready.fasta file that will be used as the input for section 3.4.

3.2. Preparation of reads for analysis of the COI "long region"

The "long region" amplicons are 658 bp long, meaning that with 300 bp long MiSeq forward and reverse reads, a non-sequenced gap of 109 bp is expected.

1. Check the quality of the reads using FastQC:

```
$ fastqc S1_R1.fastq.gz S1_R2.fastq.gz
```

This will generate a .fastqc.html file for each forward and reverse file that can be visualized in any web browser. The plots generated contain relevant information on the library preparation process and sequence quality (see the FastQC documentation for more information). If everything looks as expected, continue to the next steps, but note at which position the reads have an average quality below 25 (*see Note 9*).

2. Trim the forward and reverse reads at the position where the average quality is below 25 (see Note 8) (260 and 200 for the forward and reverse reads in this example):

```
$ trimmomatic SE -phred33 -trimlog S1_R1.logfile
S1_R1.fastq.gz S1_R1_cut.fastq.gz CROP:260
$ trimmomatic SE -phred33 -trimlog S1_R2.logfile
S1_R2.fastq.gz S1_R2_cut.fastq.gz CROP:200
```

Note that, after this trimming step, the non-sequenced gap gets longer (249 bp in this example)

3. Remove primer sequences (the first 25 and 26 bases of the forward and reverse reads respectively; see Note 6) using Trimmomatic:

```
$ trimmomatic SE -phred33 -trimlog
S1_R1_cut.logfile S1_R1_cut.fastq.gz
S1_R1_crop.fastq.gz HEADCROP:25
$ trimmomatic SE -phred33 -trimlog
S1_R2_cut.logfile S1_R2_cut.fastq.gz
S1_R2_crop.fastq.gz HEADCROP:26
```

This will result in two output files: S1_R1_crop.fastq and S1_R2_crop.fastq.

4. Uncompress the S1_R1_crop.fastq.gz and S1_R2_crop.fastq.gz files:

```
$ gunzip *_crop.fastq.gz
```

This will generate S1_R1_crop.fastq and S1_R2_crop.fastq files.

5. Transform the S1_R1_crop.fastq and S1_R2_crop.fastq files into fasta files and reverse-complement the reverse reads:

```
$ mothur "#fastq.info(fastq=S1_R1_crop.fastq)"
$ mothur "#fastq.info(fastq=S1_R2_crop.fastq)"
$ mothur "#reverse.seqs(fasta=S1_R2_crop.fasta)"
```

6. Paste the forward (S1_R1_crop.fasta) and reverse-complemented reverse reads (S1_R2_crop.rc.fasta) generated in the previous step to create an artificial barcode consisting of the trimmed forward and reverse reads. Because the forward and reverse files are in the same order, a simple paste command can be used.

```
$ paste -d '\0' S1_R1_crop.fasta
S1_R2_crop.rc.fasta | cut -d '>' -f1,2 >
S1_ready.fasta
```

This will generate the S1_ready.fasta file that will be the input for step 3.4. In this example, the barcode is 409 bp read long, which corresponds to the “long region” that lacks a 249 bp long internal fragment.

3.3. Database preparation

We start with the files described in section 2.3 that are required to generate the database: the aligned sequences (with .fasta extension) and the taxonomy (with .txt extension).

1. Remove identical sequences from the alignment file and keep one as a representative sequence in order to reduce the size of the database:

```
$ cd-hit –i BOLDdb.fasta -o BOLDdb_clean.fasta –
c 1 M2000
```

2. Trim the sequences down to the 658 bp Folmer CO1 fragment (retain the sequence between positions 38 and 714) using a sequence alignment editor (e.g. Biodedit; Hall, 1999).

3. Keep the header with the sequence identifier preceded by ">":

```
$ cut -d '|' -f1 BOLDdb_clean.fasta >
BOLDrefdb.fasta
```

4. From the taxonomy file, keep only the columns and lines needed and convert to mothur file format (Figure 3.1):

```
$ grep -v 'processid' BOLDtaxonomy.txt | cut -
f1,9,11,13,15,19,21 | sed 's/\t/;/g' | cut -d
';' -f1 > BOLDtaxonomy1.txt
$ grep -v 'processid' BOLDtaxonomy.txt | cut -
f1,9,11,13,15,19,21 | sed 's/\t/;/g' | cut -d
';' -f2- | sed 's/ /_/g' | sed 's/$/;/g'>
BOLDtaxonomy2.txt
$ paste BOLDtaxonomy1.txt BOLDtaxonomy2.txt >
BOLDtax.txt
```

5. Retain only the identifiers contained in the reference CO1 alignment (*see Note 10*):

```
$ grep '>' BOLDrefdb.fasta | cut -d '>' -f2 >
identifiers.txt
$ fgrep –f identifiers.txt BOLDtax.txt >
BOLDreftax.txt
```

If using the COI "long region", continue with this step:

6. Remove the 249 bp gap fragment from the BOLDrefdb.fasta file (from positions 246 to 498) using a sequence alignment editor to construct the BOLDgaprefdb.fasta database (*see Note 11*).

```
GBAN1430-06     Metazoa;Annelida;Polychaeta;Terebellida;Pectinariidae;Pectinaria;Pectinaria_koreni;
GBAN2075-09     Metazoa;Annelida;Polychaeta;Phyllodocida;Nereididae;Hediste;Hediste_diversicolor;
BCAS090-14      Metazoa;Annelida;Polychaeta;Terebellida;Ampharetidae;Auchenoplax;Auchenoplax_crinita;
GBCMD2886-09    Metazoa;Arthropoda;Malacostraca;Decapoda;Portunidae;Carcinus;Carcinus_maenas;
GBML0020-06     Metazoa;Mollusca;Scaphopoda;Dentaliida;Dentaliidae;Antalis;Antalis_entalis;
NBMOL004-11     Metazoa;Mollusca;Gastropoda;Littorinimorpha;Littorinidae;Littorina;Littorina_littorea;
BCAS104-15      Metazoa;Echinodermata;Ophiuroidea;Ophiurida;Ophiuridae;Ophiura;Ophiura_texturata;
```

Figure 3.1. An extract of the BOLDreftax.txt file used for the taxonomic assignment of reads. The taxonomy file is a two column text file where the first column is the sequence identifier and the second a string of taxonomic information separated by semi-colons.

3.4. Taxonomic assignment of amplicon reads

We assume that we start with quality trimmed and merged reads for the COI “short region” (section 3.1) or CO1 “long region” (section 3.2) and that we have an appropriately formatted database (section 3.3). Usually, steps 3.1 and 3.2 have generated files for more than one sample (probably hundreds), which need to be merged into a single file (let´s assume here we only have three samples: S1, S2 and S3). The commands used in this section, and their input and output file requirements are carefully explained in the mothur manual.

1. Merge the .fasta files generated in steps 3.1 or 3.2 for each sample and create a group file to assign sequences to a specific sample; for simplicity, rename the group file to a shorter name:
```
$ cat S1_ready.fasta S2_ready.fasta
S3_ready.fasta > all.fasta
$ mothur “#make.group(fasta=S1_ready.fasta-
S2_ready.fasta-S3_ready.fasta, groups=S1-S2-S3)”
$ mv
S1_ready.S2_ready.S3_ready.groups.all.groups
```

2. Discard sequences with at least one ambiguous base (*see Note 12*), retain only unique reads (*see Note 13*) and count the number of sequences per group:
```
mothur> screen.seqs(fasta=all.fasta,
group=all.groups, maxambig=0, processors=8)
mothur> unique.seqs(fasta=all.good.fasta)
```

```
mothur> count.seqs(name=all.good.names,
group=all.good.groups)
```

3. Align the sequences (here, the COI "short region" is used as an example) to the corresponding CO1 reference database using the Needleman-Wunsch global alignment algorithm. Retain sequences that align inside the barcode region (*see Note 14*) and are longer than a given threshold (*see Note 15*). In order to obtain a cleaner alignment, regions of the alignment with no data and resulting redundancies are removed.

```
mothur> align.seqs(fasta=all.good.unique.fasta,
reference=BOLDrefdb.fasta, processors=8, flip=T)
mothur> screen.seqs(fasta=all.good.unique.align,
count=all.good.count_table, minlength=200,
start=420, end=550, processors=8)
mothur>
filter.seqs(fasta=all.good.unique.good.align,
processors=8)
mothur>
unique.seqs(fasta=all.good.unique.good.filter.fa
sta, count=all.good.good.count_table)
```

4. Remove sequences that occur only once among all samples (singletons) (*see Note 16*).

```
mothur>
split.abund(fasta=all.good.unique.good.filter.un
ique.fasta,
count=all.good.unique.good.filter.count_table,
cutoff=1)
```

5. Remove potential chimeric sequences using UCHIME (Edgar et al., 2011) *De novo* mode:

```
mothur>
chimera.uchime(fasta=all.good.unique.good.filter
.unique.abund.fasta,
```

```
count=all.good.unique.good.filter.abund.count_ta
ble, processors=8)
mothur>
remove.seqs(accnos=all.good.unique.good.filter.u
nique.abund.uchime.accnos,
fasta=all.good.unique.good.filter.unique.abund.f
asta,
count=all.good.unique.good.filter.abund.count_ta
ble)
```

6. Assign taxonomy to the sequences using the Wang approach (Wang et al., 2007). Taxonomic assignments are done using the aligned reference database and the reference taxonomy file created in section 3.3.

```
mothur>
classify.seqs(fasta=all.good.unique.good.filter.
unique.abund.pick.fasta, count=
all.good.unique.good.filter.abund.pick.count_tab
le, template=BOLDrefdb.fasta,
taxonomy=BOLDreftax.txt, cutoff=90, method=wang,
processors=8)
```

7. Cluster sequences into "Operational Taxonomic Units" (OTUs) based on the previous taxonomic classification. Count the number of times an OTU is observed in order to have information about the incidence of the OTUs in the different samples.

```
mothur>
phylotype(taxonomy=all.good.unique.good.filter.u
nique.abund.pick.BOLDreftax.wang.taxonomy)
mothur>
make.shared(list=all.good.unique.good.filter.uni
que.abund.pick.BOLDreftax.wang.tx.list,
count=all.good.unique.good.filter.abund.pick.cou
nt_table)
```

This will create a file with the count of the number of reads in each OTU, for each sample.

8. Assign taxonomy to each OTU.

```
mothur>
classify.otu(list=all.good.unique.good.filter.un
ique.abund.pick.BOLDreftax.wang.tx.list,
count=all.good.unique.good.filter.abund.pick.cou
nt_table,
taxonomy=all.good.unique.good.filter.unique.abun
d.pick.BOLDreftax.wang.taxonomy)
```

9. Combine the files obtained in steps 7 and 8 into a single table to generate the OTU table that contains the count of the number of sequences in each OTU, for each sample, and the taxonomy of that OTU.

The OTU table is the final output obtained using this protocol, which can be used as an input file for diversity metrics estimations (i.e. alpha and beta diversity). See also chapter 1 by Lehmann and chapter 15 by Leray and Knowlton (Bourlat 2016) on the calculation of diversity indices using the OTU table. The OTU table can also be used for the calculation of biotic indices, biodiversity monitoring programs and other biodiversity studies that are based on sample taxonomic composition.

3.5. Calculation of benthic indices from sequence data

The biotic index calculation procedure described here is based on presence/absence data obtained from the taxonomic analysis of amplicon reads performed in section 3.4 and carried out according to the "Instructions for the use of the AMBI index" protocol (Borja et al., 2012). Detailed information about each step can be found in the manual.

We assume that we start with the OTU table, for which taxonomic assignment of the reads has been performed.

1. Import the OTU table into a spreadsheet and open it in R, Excel, or any other program to manipulate data.
2. Transform relative abundance of the retained taxa into presence/absence data. Simply change the number of reads to 1 if they represent more than 0.01% of the total taxa and keep the rest of the cells blank.

3. Open the AMBI software, import the spreadsheet and calculate the AMBI index. The result will show the ecological quality of the stations under study (Figure 3.2), allowing the monitoring of a site after an impact or the detection of gradients from the source of a certain impact. In addition, detailed information on the percentage of taxa assigned to each ecological group for each station can be displayed (Figure 3.3).

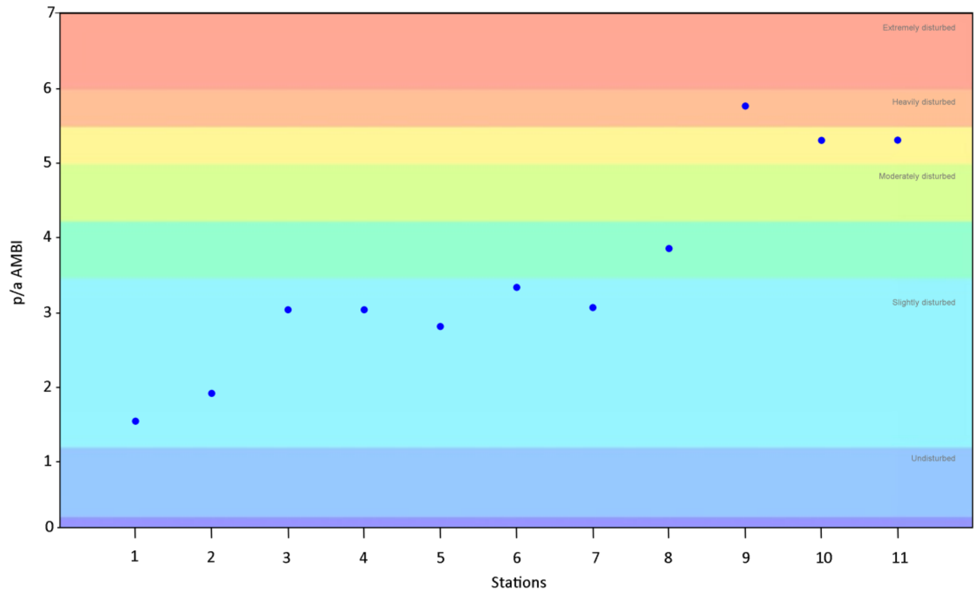

Figure 3.2. Ecological quality for 11 arbitrary stations used as an illustrating example.

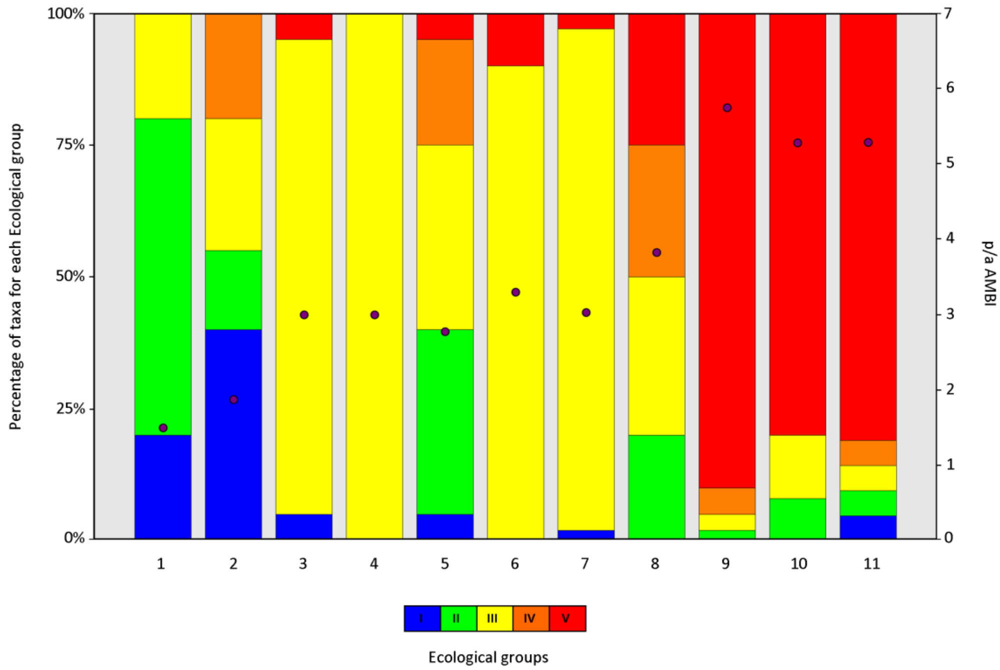

Figure 3.3. Percentage of taxa from each ecological group and derived presence/absence(pa) AMBI values for 11 arbitrary stations used as an illustrating example.

4. Notes

1. A new BOLD account can be created at http://www.boldsystems.org/index.php/MAS_Management_NewUserApp
2. Minimum required fields in the record search are Taxonomy, Marker and select to include public records. Note that record search can be performed by taxonomic level (e.g. phylum), although some groups need to be split into lower levels (e.g. Chordata has to be split into classes) due to download limitations of 50,000 records in a unique search. If that is the case, you will need to concatenate the resulting files to create a single file.
3. Sections 3.1 and 3.2 describe the steps needed to process one sample (named here "S1") for which two raw data files (S1_R1.fastq.gz and S1_R2.fastq.gz), corresponding to 300 bp long forward and reverse reads, have been provided. Processing the hundreds of files usually generated in one MiSeq run would require

the use of scripts including loops, which is not covered in this chapter.

4. mothur can be executed using an interactive mode, batch mode and command line mode; see the mothur webpage for more explanations on how to use each mode.
5. It is expected that quality drops towards the end of the reads. For the analysis of the “short region”, this is not an issue because the large overlapping region allows the poor quality bases at the ends of the forward reads to be compensated by the good quality ones of the beginning of the reverse reads and vice versa.
6. If different primers are used, these values need to be adjusted to the appropriate primer length.
7. We found that using a minimum and a maximum overlap of respectively minus and plus 20 bases from the expected overlap (237 bp in the case of the “short region”) provides good results.
8. Quality score thresholds are somewhat arbitrary. We found that 25 is not too strict, neither too loose, but other values are equally appropriate.
9. It is expected that the quality of the read drops towards the end; for the analysis of the “long region”, this is an issue because there is no overlap.
10. The taxonomy file downloaded from BOLD also includes taxa for which no barcode is available. Because the database must contain the same identifiers in both alignment and taxonomy files, these additional taxa need to be removed.
11. Removing the 249 bp gap fragment in the database facilitates the alignment of the query sequences to the reference alignment. We found that even changing the alignment parameters, mothur is not able to correctly aligning the COI “long region” sequences to the complete reference database.
12. Discarding all reads that contain at least one ambiguous base may lead to too few reads remaining; in such cases, it is possible to exclude only those reads with more than a certain number of ambiguous bases.
13. The unique.seqs commands is applied several times in order to reduce the number of reads analyzed by returning only the unique

sequences found; it has no effect on the output as it is not a filtering step.

14. Before aligning sequences to the reference alignment, verify the start and end positions on the alignment - this will facilitate following filtering steps. For the COI “short region”, we retained sequences that start at or before position 420 and end at or after position 550; for the COI “long region” these positions are 60 and 300 respectively.
15. We used 200 bp for the COI “short region” and 300 bp for the COI “long region”.
16. It is assumed that reads that occur only once (singletons) are most likely to be due to PCR or sequencing errors than to be real data.

Chapter 4

BENCHMARKING DNA METABARCODING FOR BIODIVERSITY-BASED MONITORING AND ASSESSMENT

<u>Published as</u>: **Aylagas, E.**, Borja, A., Irigoien, X., and Rodríguez-Ezpeleta, N. (2016). Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. Frontiers in Marine Science 3 (96). doi: 10.3389/fmars.2016.00096

1. Introduction

Environmental biomonitoring in coastal and marine ecosystems often relies on comprehensively, accurately and repeatedly characterizing the benthic macroinvertebrate community (Yu et al., 2012). These organisms are considered a good indicator of ecosystem health and have demonstrated a rapid response to a range of natural and anthropogenic pressures (Johnston and Roberts, 2009). As a result, the macroinvertebrate community has been largely used to develop biotic indices (Diaz et al., 2004; Pinto et al., 2009; Borja et al., 2015b), such as the AMBI (Borja et al., 2000), used worldwide to assess the marine benthic status (Borja et al., 2015b). Nevertheless, biomonitoring based upon benthic organisms has limitations because species identification requires extensive taxonomic expertise and it is time-consuming, expensive and laborious (Yu et al., 2012; Wood et al., 2013; Aylagas et al., 2014). The rapid development of HTS technologies represents a promising opportunity for easing the implementation of molecular approaches for biomonitoring programs (Bourlat et al., 2013; Dowle et al., 2015). In particular, DNA metabarcoding (Taberlet et al., 2012a) allows the rapid and cost-effective identification of the entire taxonomic composition of thousands of samples simultaneously (Zepeda Mendoza et al., 2015) and the ability to provide a more comprehensive community analysis than traditional assessments (Dafforn et al., 2014), which can enable the calculation of benthic indices in a much faster and accurate way compared to morphological methodologies.

Metabarcoding consists of simultaneously amplifying a standardized DNA fragment specific for a species (*barcode*) from the total DNA extracted from an environmental sample using conserved short DNA sequences flanking the barcode (*primers*) (Hajibabaei, 2012; Cristescu, 2014). The obtained barcodes are then high-throughput sequenced and compared to a previously generated DNA sequence reference database from well-characterized species for taxonomic assignment (Taberlet et al., 2012a). In the case of animals, different barcodes such as portions of the small and large subunits of the nuclear ribosomal RNA (18S and 28S rRNA) genes (Machida and Knowlton, 2012) and of the mitochondrial cytochrome oxidase I (COI) (Meusnier et al., 2008) and 16S rRNA genes (Sarri et al., 2014) have been proposed for metabarcoding. The COI gene is by far the most commonly used marker for metazoan metabarcoding (Ratnasingham and Hebert, 2013), for which thousands of reference sequences are available in public databases (the Barcode of Life Database (BOLD) contains >1,000,000 COI sequences belonging to animal species) and several amplification primers have been designed (more than 400 COI primers are published in the Consortium for the Barcode of Life (CBOL) primer database).

Several studies have used metabarcoding to characterize the metazoan taxonomic composition of aquatic environments (Porazinska et al., 2009; Chariton et al., 2010; Fonseca et al., 2014; Dell'Anno et al., 2015; Leray and Knowlton, 2015; Chain et al., 2016), and an increasing number of studies have directly applied the approach for environmental biomonitoring purposes (Ji et al., 2013; Dafforn et al., 2014; Pawlowski et al., 2014a; Chariton et al., 2015; Gibson et al., 2015; Pochon et al., 2015; Zaiko et al., 2015). Initial studies inferring biotic indices from molecular data show the potential of metabarcoding for evaluating aquatic ecosystem quality (Lejzerowicz et al., 2015; Visco et al., 2015). However, before implementation of metabarcoding in regular biomonitoring programs, this approach needs to be benchmarked against morphological identification so that accurate taxonomic inferences and derived biotic indices can be ensured (Aylagas et al., 2014; Carugati et al., 2015). The accuracy of metabarcoding-based taxonomic inferences relies on the retrieval of a wide range of taxonomic groups from a given environmental sample using the appropriate barcode, primers and amplification conditions (Deagle et al., 2014; Kress et al., 2015), and on the completeness of the reference database (Zepeda Mendoza et al., 2015). Some attempts have been performed to compare morphological versus metabarcoding-based taxonomic inferences; yet, results are inconclusive as some studies do not apply both approaches to the same sample and/or have focused on a particular taxonomic group (Hajibabaei et al., 2012; Carew et al., 2013; Zhou et al., 2013; Gibson et al., 2014; Cowart et al., 2015; Zimmermann et al., 2015). A recent study (Gibson et al., 2015) has performed morphological and metabarcoding-based taxonomic identification on the same freshwater aquatic invertebrate samples, but limited their visual identifications to family level. Only two studies (Dowle et al., 2015; Elbrecht and Leese, 2015) have performed a robust benchmarking of metabarcoding using freshwater invertebrates and showed that this technique can be successfully applied to biodiversity assessment. In marine metazoans, all studies have focused only on plankton samples (Brown et al., 2015; Mohrbeck et al., 2015; Albaina et al., 2016). Thus, an exhaustive evaluation of metabarcoding for marine benthic metazoan taxonomic inferences is still lacking.

The use of extracellular DNA (the DNA released from cell lysis (Taberlet et al., 2012b)) for biodiversity monitoring is increasingly applied to water (e.g.(Ficetola et al., 2008; Foote et al., 2012; Thomsen et al., 2012a; Kelly et al., 2014a; Davy et al., 2015; Valentini et al., 2016), soil (Taberlet et al., 2012b) and sediment samples (Guardiola et al., 2015; Turner et al., 2015; Pearman et al., 2016b). Constituting a significant fraction of the total DNA (Dell'Anno and Danovaro, 2005; Pietramellara et al., 2009; Torti et al., 2015), it is assumed that the taxonomic composition of the free DNA present in the environment reflects the biodiversity of the sample (Ficetola et al., 2008), which would simplify DNA extraction protocols (Pearman et al., 2016b) and allow the detection of organisms that are even larger than the sample itself (Foote et al., 2012; Thomsen et al., 2012a; Kelly et al., 2014a; Davy et al., 2015). Thus, this method appears as a promising cost-effective alternative for macroinvertebrate diversity monitoring, but no robust evidence that the entire macroinvertebrate community can be detected using extracellular DNA exists so far.

The lack of a thorough comparison between morphological and metabarcoding-based taxonomic inferences of marine metazoa and of an evaluation of the use of metabarcoding for marine biotic index estimations prevents the application of metabarcoding in routine biomonitoring programs. Here we benchmark alternative metabarcoding protocols based on a combination of different DNA sources (extracellular DNA and DNA extracted from previously isolated organisms), barcodes (short and long COI regions) and amplification conditions against benthic macroinvertebrate samples of known taxonomic composition. Additionally, we test the effect of the discrepancies between morphological and DNA-based taxonomic inferences in marine biomonitoring through the evaluation of the molecular based taxonomies performance when incorporated for the calculation of the AMBI and prove the suitability of molecular data based biotic indices to assess marine ecological status.

2. Methods

The experimental design followed to compare the performance of molecular and morphological-based taxonomic inferences is summarized in Figure 4.1.

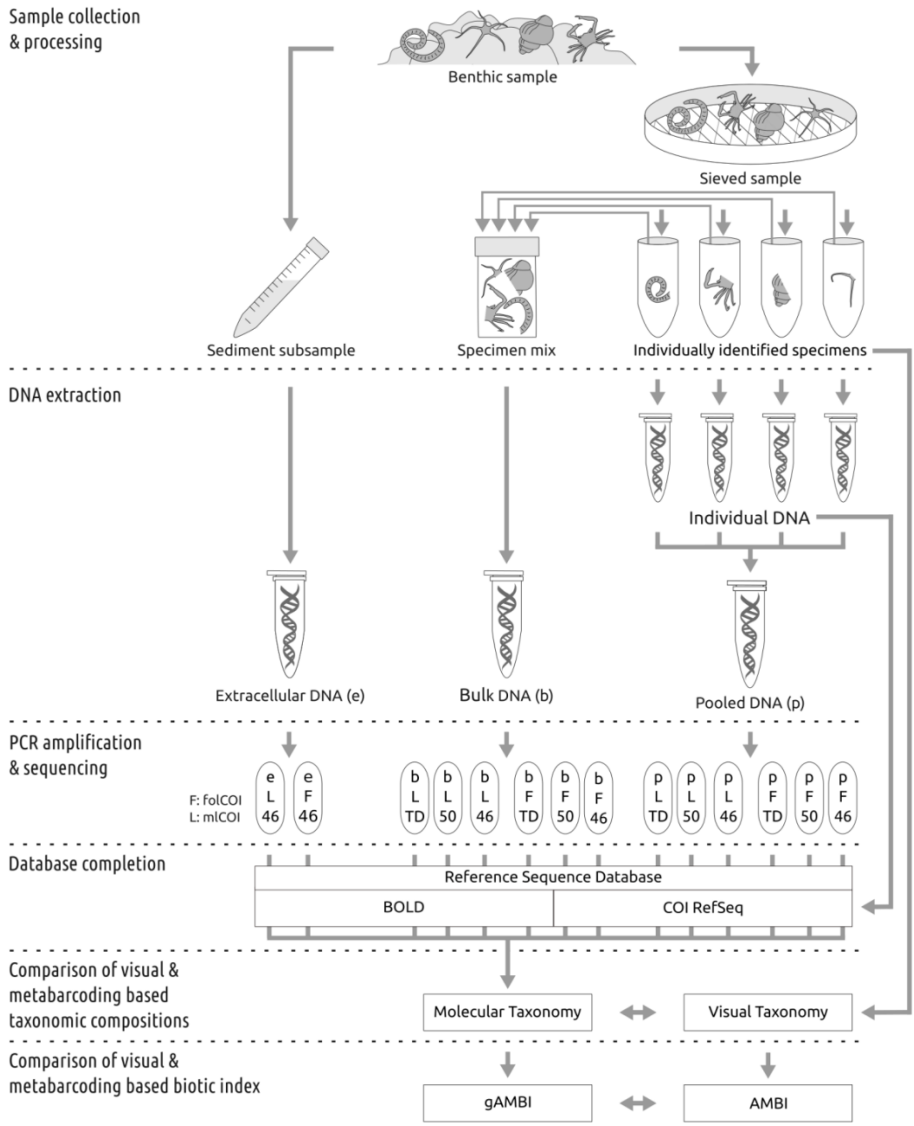

Figure 4.1. Workflow for sample processing. See Methods section for detailed explanations.

2.1. Sample collection and processing

Benthic samples were collected from 11 littoral stations (sampling depth ranging from 100 to 740 m) along the Basque Coast, Bay of Biscay (Figure 4.2), during March 2013, using a van Veen grab (0.07–0.1 m^2). At each location, after sediment homogenization, one subsample of sediment was taken from the surficial layer of the grab and stored in a sterile 15 ml falcon tube at -80 °C until extracellular DNA extraction (see below). In order to collect the benthic macroinvertebrate community (organism size >1 mm) present in each sample, the remaining sediment was sieved on site through a 1 mm size mesh, and the retained material preserved in 96% ethanol at 4 °C until processing (<6 months). Macroinvertebrate specimens were sorted and identified to the lowest possible taxonomic level based on morphology. Following taxonomic classification, each sample was divided into two identical subsamples by taking equal amount of tissue per taxa for each subsample. Tissues from one subsample were pooled and used for bulk DNA extraction. Each tissue of the second subsample was used for individual DNA extraction (see below).

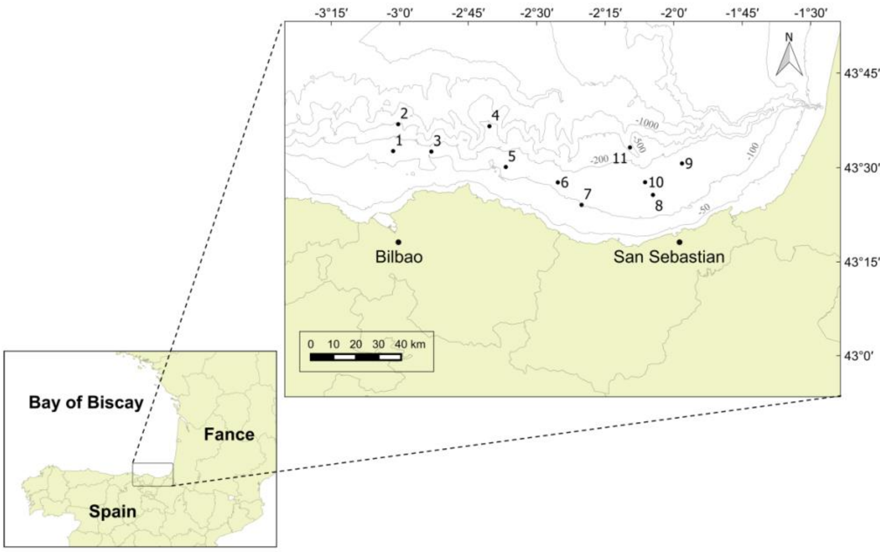

Figure 4.2. Map depicting the 11 sampling stations along the Basque Coast from where samples were collected

2.2. Extracellular, individual and bulk DNA extraction

Extracellular DNA was extracted following an optimized protocol (Taberlet et al., 2012b). Briefly, 5 g of each sediment sample were mixed with 7.5 ml of saturated phosphate buffer and an equal volume of chloroform:isoamyl alcohol (IAA). After centrifugation for 5 min at 4000 g, the aqueous phase was passed through a second round of chloroform:IAA purification and ethanol precipitated before elution of resulting DNA pellet in 100 µl Milli-Q water. For individual and bulk processing, total genomic DNA from each tissue and from the mix of tissues composing each sample, respectively, were extracted using the Wizard® Genomic DNA Purification kit (Promega, WI, USA) in a 125 µl of Milli-Q water final elution. The possible presence of PCR inhibitors in the bulk and extracellular DNA were removed using the Mobio PowerClean® DNA Clean-Up Kit. Genomic DNA integrity was assessed by electrophoresis, migrating about 100 ng of GelRed™-stained DNA on an agarose 1.0% gel, DNA purity was assessed using the Nanodrop® ND-1000 (Thermo Scientific) system and DNA concentration was determined with the Quant-iT dsDNA HS assay kit using a Qubit® 2.0 Fluorometer (Life Technologies). About 20 ng of each individually extracted DNA were used for DNA barcoding of single species (see details below). Subsequently, 5 µl of each individually extracted DNA at original concentration were pooled (hereafter referred as "pooled DNA"). Extracellular, bulk and pooled DNA were used for PCR amplification and sequencing (see below).

2.3. Individual PCR amplification and Sanger sequencing

Individual DNA barcoding was performed for the species for which no COI barcode was available in public databases (see table Supplementary Material). The standard 658 bp COI barcode (*folCOI*) was targeted using the dgLCO1490 × dgHCO2198 primer pair (Meyer, 2003). Each individual DNA sample was amplified in a total volume reaction of 20 µl using 10 µl of Phusion® High-Fidelity PCR Master Mix (Thermo Scientific), 0.2 µl of each primer (10 µM) and 20 ng of genomic DNA. The thermocycling profile consisted of an initial 30 sec denaturation step at 98 °C, followed by up to 35 cycles of 10 sec at 98 °C, 30 sec at 48 °C and 45 sec at 72 °C, and a final 5 min extension step at 72 °C. PCR products were considered positive when a clear single band of expected size was visualized on a 1.7% agarose gel. Samples with negative product were further amplified with the mlCOIintF ×

dgHCO2198 primer pair (Leray et al., 2013) targeting a 313 bp fragment of the COI gene (*mlCOI*). Negative samples were included with each PCR run as external control. PCR products were purified with ExoSAP-IT (Affymetrix) and Sanger sequenced.

2.4. PCR amplification for library preparation and Illumina MiSeq sequencing

Indexed paired-end libraries of pooled amplicons were prepared using two nested PCRs from the extracellular, bulk and pooled (mix of 5 µl of individually extracted DNA at original concentration) DNA obtained from each of the 11 collected samples. In parallel, three of the samples were processed per triplicate and considered independently in downstream analysis. For the first PCR, two universal primer pairs with overhang Illumina adapters were used to amplify two different length COI barcodes (the *mlCOI* and the *folCOI*). Three different PCR profiles were used to amplify each COI barcode from the bulk and pooled DNAs (46 and 50 °C annealing temperatures and a touchdown profile), whilst the extracellular DNA COI barcodes were amplified with 46 °C annealing temperature. PCRs were performed in a total volume of 20 µl using 10 µl of Phusion® High-Fidelity PCR Master Mix (Thermo Scientific), 0.5 µl of each primer (10 µM) and 2 µl of genomic DNA (5 ng/µl). The PCR conditions for the two different annealing temperatures consisted on an initial 30 sec denaturation step at 98 °C, 27 cycles of 10 sec at 98 °C, 30 sec at 46 or 50 °C and 45 sec at 72 °C, and a final 5 min extension at 72 °C. For the touchdown profile the PCR conditions consisted on an initial 30 sec denaturation step at 98 °C, 16 cycles of 10 sec at 98 °C, 30 sec at 62 °C (-1 °C per cycle) and 60 sec at 72 °C, followed by 17 cycles at 46 °C annealing temperature, and a final 5 min extension at 72 °C (Leray et al., 2013). Negative controls were included with each PCR. Generated amplicons were purified with AMPure XP beads (Beckman Coulter), eluted in 50 µL MilliQ water and used as templates for the generation of the dual-indexed amplicons in the second PCR round following the "16S Metagenomic Sequencing Library Preparation" protocol (Illumina). Purified PCR products were quantified using the Quant-iT dsDNA HS assay kit using a Qubit® 2.0 Fluorometer (Life Technologies) and further normalized for all samples. Pools of 96 equal concentration amplicons were sequenced using the 2 × 300 paired-end on a MiSeq (Illumina).

2.5. DNA barcode reference database

Trace files of Sanger sequences obtained from individual PCR amplifications were edited and trimmed to remove low quality bases (Qvalue <30) using SeqTrace 0.9.0 (Stucky, 2012) and checked for frame shifts using EXPASY (Gasteiger et al., 2003). COI sequences are available in 'BCAS project' at BOLD (http://www.boldsystems.org) and in GenBank (accession numbers KT307619 - KT307707). To generate our DNA reference database, we retrieved a total of 1,123,601 public COI aligned sequences from 96,641 different taxa from BOLD (October 2014), including the sequences generated in this study (COI RefSeq). After removing duplicates, a total of 505,033 sequences were kept and trimmed to the 658 bp Folmer COI fragment to generate the "BOLD database". A smaller customized DNA reference database was generated using the 4,231 sequences corresponding to species included in the AMBI list (see below) (available at http://ambi.azti.es) extracted from the "BOLD database" to build the "AMBI database". For the analyses of the *folCOI* reads, the 249 bp not sequenced internal fragment (see below) was removed from these two databases to construct the "BOLD gapped database" and the "AMBI gapped database". The four resulting databases were formatted according to mothur (Schloss, 2009) standards.

2.6. Amplicon sequence analysis

Demultiplexed reads were quality checked using FastQC (Andrews, 2010) and primer sequences removed using Trimmomatic 0.33 (Bolger et al., 2014). Since the *mlCOI* paired-end reads overlap in 237 bp and the *folCOI* paired-end reads do not overlap, different preprocessing steps are needed for each COI fragment. Forward and reverse *mlCOI* reads were merged using FLASH (Magoč and Salzberg, 2011) with a minimum and maximum overlap of respectively 20 bases below and above the expected overlapping region, and the resulting reads were trimmed using Trimmomatic at the first sliding window of 50 bp with an average quality score below 30. The *folCOI* forward and reverse reads were trimmed at 260 and 200 bp respectively based on the quality decrease after these positions observed on FastQC plots. Each pair of forward and reverse-complemented reverse read was pasted to create a 409 bp read that corresponds to the *folCOI* barcode without a 249 bp internal fragment. Further details on this new pipeline developed to analyze the universal 658 bp COI barcode which is too long for most high-throughput sequencing applications such as the Illumina MiSeq are detailed elsewhere (Aylagas and Rodríguez-Ezpeleta, 2016). Preprocessed reads from both barcodes were independently analyzed with mothur following the MiSeq standard operating procedure (Kozich et al., 2013). Briefly, sequences with ambiguous bases were discarded and the rest, aligned to the corresponding BOLD and AMBI reference databases. Only those *mlCOI* and *folCOI* reads aligning inside the barcode region and longer than 200 bp and 300 bp respectively were kept. After chimera removal using the *de novo* mode of UCHIME (Edgar et al., 2011), sequences were grouped into phylotypes according to the taxonomic assignments made based on the Wang method (Wang et al., 2007) using a bootstrap value of 90. The sequences that did not return any taxonomic assignment against the BOLD database were blasted against the NCBI non redundant database. All Sequences have been deposited in the Dryad Digital Repository (doi: dx.doi.org/10.5061/dryad.4t3t2).

2.7. Comparison of morphological and metabarcoding-based taxonomic compositions

Only taxa representing at least 0.01% of the reads in one station were considered present in the taxonomic composition inferred from molecular data. An in-house script (Figure 4.3) was used to calculate the degree of match between the molecular and morphologically inferred taxonomic compositions of each station. The detection success was normalized for each sample and transformed to percentage of matches (100% of matches means all taxa identified based on morphology have been detected using DNA-based approaches). Differences in mean values of the taxa detection percentages between DNA extraction methods, primers and PCR conditions were examined using a t-test at alpha=0.05. Patterns of sample dissimilarity were visualized using non-metric multidimensional scaling (nMDS) based on taxa presence/absence and abundance using the Jaccard and Bray-Curtis indices respectively obtained using molecular approaches.

```
database: AMBI database

FOREACH sample
        species_list: list of taxa identified at the species level in the morphological taxonomy
        taxa_listoflists: list of lists that contains:
                genus_list: list of taxa identified at the genus level in the morphological
                taxonomy
                family_list: list of taxa identified at the family level in the morphological
                taxonomy
                order_list: list of taxa identified at the order level in the morphological
                taxonomy
                class_list: list of taxa identified at the class level in the morphological
                taxonomy
                phylum_list: list of taxa identified at the phylum level in the morphological
                taxonomy
                remaining_list: empty list
        moltax_list: list of taxa found in the molecular taxonomy
        visNoMatch: empty list
        visMatch: empty list
        molNoMatch: moltax_list

        FOREACH species IN species_list
                IF species IN database
                        IF species IN moltax_list
                                add species to visMatch
                                remove species from molNoMatch
                             exact match
                        ELSE
                                add species to visNoMatch
                             no match
                ELSE
                        add species´ genus to genus_list
        FOREACH taxa_list IN taxa_listoflists
                FOREACH taxa IN taxa_list
                        IF taxa IN database
                                IF taxa NOT IN moltax_list
                                        add taxa to visNoMatch
                                        no match
                                ELSEIF taxa IN database only once
                                        add taxa to visMatch
                                        remove taxa from molNoMatch
                                        match with
                                ELSE
                                        IF taxa IN moltax_list only once
                                                IF in visMatch only once
                                                        add taxa to visMatch
                                                        match with already used
                                                ELSE
                                                        add taxa to visMatch
                                                        remove taxa from
              molNoMatch
                                                        match with
                                        ELSE
                                                match with various
                        ELSE:
                                add taxa to next taxa_list
```

Figure 4.3. (previous page) Procedure followed to assign matches between morphological and molecularly inferred taxonomies. *Exact match* indicates that a species identified in the morphological taxonomy is found under the same name in the molecular taxonomy; *no match* indicates that the first taxonomic level of a taxa identified in the morphological taxonomy that has barcode in the database is not found in the molecular taxonomy; the categories *match with*, *match with already used* and *match with various* are considered *match*, but have been created to differentiate them from the cases where the taxonomic identification has not been done at the species level and/or the morphologically identified taxa has no barcode in the database.

2.8. Comparison of morphological and metabarcoding-based biotic indices

In order to compare morphological and metabarcoding-based biotic indices, we used AMBI, which is a status assessment index based on the pollution tolerances of the taxa present in a sample, with tolerance being expressed categorically into ecological groups (EGI, sensitive to pressure; EGII, indifferent; EGIII, tolerant; EGIV, opportunist of second order; and, EGV, opportunist of first order). We calculated the presence/absence(pa) morphology-based AMBI and the presence/absence(pa) genetics-based AMBI (Aylagas et al., 2014) inferred through DNA metabarcoding of each sample, using the AMBI 5.0 software (http://ambi.azti.es). The relationships among (pa)AMBI and (pa)gAMBI values were examined using standardized major axis (SMA) estimation (Warton et al., 2006) using the software SMATR (Falster et al., 2003). In order to evaluate the performance of (pa)gAMBI for each condition, root-mean-square error (RMSE) and bias were calculated (Walther and Moore, 2005).

3. Results

3.1. Morphological and molecular analysis

In total, 138 macroinvertebrate taxa belonging to 9 different phyla were morphologically identified in the 11 stations. Representatives of two main phyla, Annelida and Arthropoda, are present at all stations, with 94 and 21 taxa, respectively, whereas less represented phyla (Mollusca, Chaetognata, Cnidaria, Echinodermata, Nemertea, Nematoda and Sipuncula) are absent from some stations and include less number of taxa (Table S4.1). Individual DNA barcoding was successful on 61 and 24 of the 106 identified species with no COI barcode in public databases, for which new *folCOI* and *mlCOI* barcodes were generated, respectively, and included in the reference database. Despite this effort to increase the reference database, 21 species remain without barcode because amplification of both barcodes failed.

For each station, two condition combinations were tested for the extracellular DNA (two different barcodes) and six for the bulk and pooled DNAs (two different barcodes and three different PCR profiles). From the 238 samples analyzed, including triplicates performed on three of the stations, 14 had no PCR amplification (see Table S4.2 for clarification on the number of samples produced for molecular analysis). The 224 remaining resulted in 16 million reads, from which about 56% passed quality filters and were used for taxonomic analysis. Of the total reads obtained from extracellular DNA, 71.5 and 73.4% could not be assigned to any metazoan phylum using the customized BOLD database and 24.9 and 25.6% were not assigned to Metazoa for *mlCOI* and *folCOI*, respectively. When blasted against NCBI, the reads obtained using *mlCOI* matched with bacteria (0.6%), non-metazoan eukaryotes (84%), metazoans (12.2%) or did not provide any match (3%), and the reads obtained using *folCOI* matched with bacteria (66.6%), non-metazoan eukaryotes (6%), metazoans (4.2%), archaea (0.05%) or did not provide any match (23.2%). The percentages of non-metazoan reads are much lower for bulk (0.03 and 0.04%) and pooled DNA (0.1 and 0.3%), and the proportion of Metazoa reads with no phylum assigned are lower for *mlCOI* (23.2 and 10.6% for bulk and pooled DNA, respectively) than for *folCOI* (29.94 and 31.6% for bulk and pooled DNA, respectively).

3.2. Comparison of morphological and molecular-based taxonomic compositions

From the taxonomic inferences obtained using molecular approaches, only macroinvertebrates were considered for sample comparison (e.g. Chordata records were excluded for downstream analysis). The average percentage of recovered taxa (molecular taxonomy matches visual taxonomy) over all stations using different conditions is shown in Figure 4.4 (see Figure S4.1 for percentage of recovered taxa considering only species level identification). Matches for taxonomic inferences based on metabarcoding of extracellular DNA are very low (3.4% and 3.1% for *folCOI* and *mlCOI* respectively), with only taxa from three phyla (Mollusca, Annelida and Nemertea) retrieved (Table S4.3). Results obtained between replicates from the same sample reveal similar taxonomic inferences. No significant differences were observed between the percentage of matches obtained using bulk and pooled DNA (p value > 0.05). Interestingly, the *mlCOI* barcode outperforms the *folCOI* barcode (p value < 0.05 for bulk and pooled DNA) and, within the *mlCOI*, the 46 and 50 °C annealing temperatures outperform the touchdown profile both for bulk and pooled DNA (p values < 0.05). Overall, the best performing condition is the *mlCOI* barcode amplified using 46 °C annealing temperature, which results in a percentage of recovered taxa of 62.4% for all matches and of 76.3% for only matches at species level.

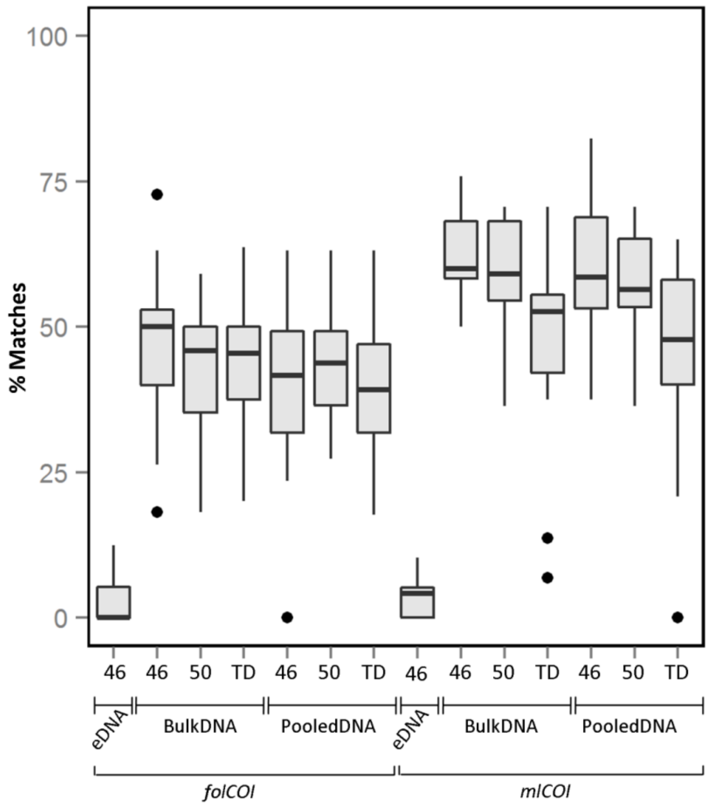

Figure 4.4. Boxplot showing the percentage of matches obtained between morphological and molecularly inferred taxonomic compositions over all stations. All matches using extracellular DNA (eDNA), bulk and pooled DNA approaches using different PCR conditions (46 or 50 °C annealing temperatures or TD: touchdown profile) for *folCOI* and *mlCOI* barcodes.

Using molecular approaches we were able to retrieve taxa that had not been morphologically identified. Representatives of Annelida (e.g. *Tubificoides amplivasatus*, *Chloeia parva* and *Mugga wahrbergi*), Arthropoda (e.g. *Scyllarus arctus* and *Limnoria* sp.), Mollusca (e.g. *Nucula nucleus*, *Galeomma turtoni, Thyasira ferruginea* and *Entalina tetragona*) and Echinodermata (e.g. *Ophiura albida* and *Macrophiothrix* sp.) were solely identified using DNA-based approaches. Moreover, we were able to find taxa belonging to two phyla that were not morphologically identified even at phylum level: two families (Triaenophoridae and Echinobothriidae) and one order (Acoeala) of Platyhelminthes and one family (Hemiasterellidae) of Porifera. As illustrated by the nMDS ordination plot of beta diversity (Figure 4.5), the greatest disparity in macroinvertebrate composition inferred using molecular taxonomy of each station was shown by the extracellular DNA approach.

3.3. Comparison of morphological and metabarcoding-based biotic indices

The correlation between (pa)AMBI and (pa)gAMBI values obtained from the taxonomic composition inferences using the AMBI database is shown in Figure 4.6. The (pa)AMBI values that best correlate with (pa)gAMBI values are those obtained using bulk and pooled DNA approaches at 46 or 50 °C annealing temperatures obtained with *mlCOI* (Table 4.1). Generally, (pa)gAMBI values tend to score lower than (pa)AMBI values (negative bias over all stations). This tendency can be also observed in the variation of the percentage of taxa found belonging to each ecological group obtained using morphological and molecular taxonomic identifications (Figure S4.2). The non-detection of taxa belonging to tolerant and opportunistic ecological groups (III, IV and V) when using *folCOI*, especially for pooled DNA method, leads to poor correlations between (pa)AMBI and (pa)gAMBI values.

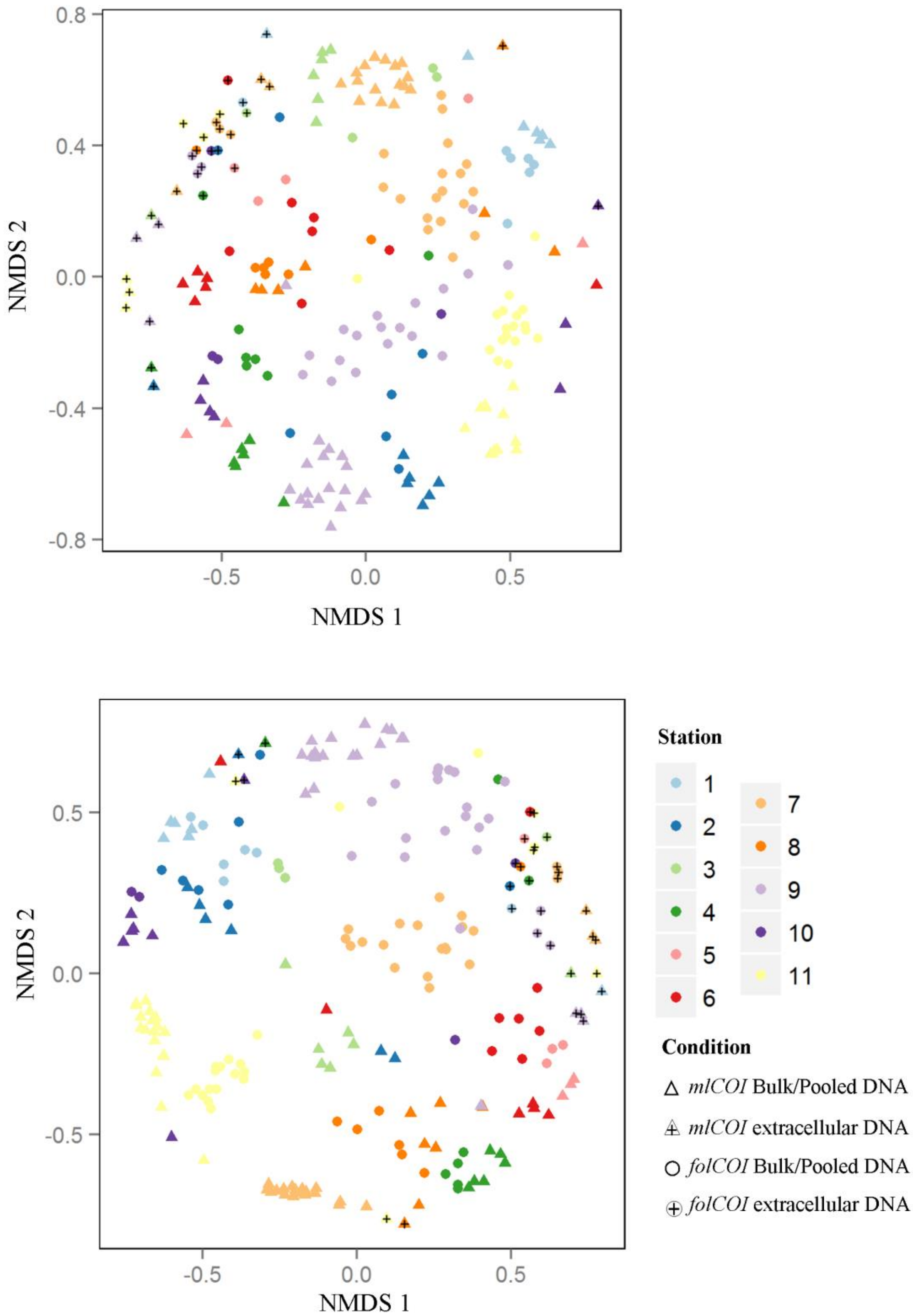

Figure 4.5. Non-metric multidimensional scaling (nMDS) plots. (Top) Jaccard (presence-absence) and (Bottom) Bray-Curtis (abundance) dissimilarities for 32 samples of extracellular DNA and 192 samples of bulk or pooled DNA approaches, from 11 littoral stations for the two barcodes (*mlCOI* and *folCOI*).

4. Discussion

4.1. Effect of PCR-based analysis biases on taxonomic inferences

Finding the primer pair and PCR conditions that most accurately recover the organisms present in an environmental sample is crucial for a successful application of metabarcoding to biomonitoring. Several studies analyzing the same samples with morphological and molecular taxonomy have been performed so far to benchmark COI-based metabarcoding in animals, all focusing exclusively on freshwater or terrestrial macroinvertebrates (Hajibabaei et al., 2012; Carew et al., 2013; Gibson et al., 2014; Dowle et al., 2015; Elbrecht and Leese, 2015) or carried out under morphological identifications limited to high taxonomic levels (Gibson et al., 2015). Thus, studies on marine benthic communities that prove the suitability of DNA-based approaches for environmental biomonitoring are lacking. Using samples of known taxonomic composition, we show that an alternative barcode that targets a shorter region of the COI gene outperforms the 658 bp region that is commonly used for metabarcoding metazoans (Carew et al., 2013; Ji et al., 2013; Dowle et al., 2015; Elbrecht and Leese, 2015; Zaiko et al., 2015). Our data corroborate previous studies unveiling the lack of universality in the COI primers, which is translated to biases during PCR step (Pochon et al., 2013; Deagle et al., 2014). However, the increased performance of the short region, previously demonstrated for individual barcoding on marine metazoans (Leray et al., 2013) and metabarcoding in insects (Brandon-Mong et al., 2015) proves that the *mlCOI* barcode retrieves a high proportion of the morphologically identified taxa. This fact also corroborates the preferred use of small barcodes for metabarcoding, which provide pair-end overlaps on Illumina sequencing and good taxonomic resolution for species identification (Meusnier et al., 2008). Additionally, the *folCOI* barcode returns more reads with no match and metazoan reads not assigned to any specific phylum, which could be attributed to the fact that longer barcodes can accumulate more errors during the PCR and sequencing processes (Schirmer et al., 2015).

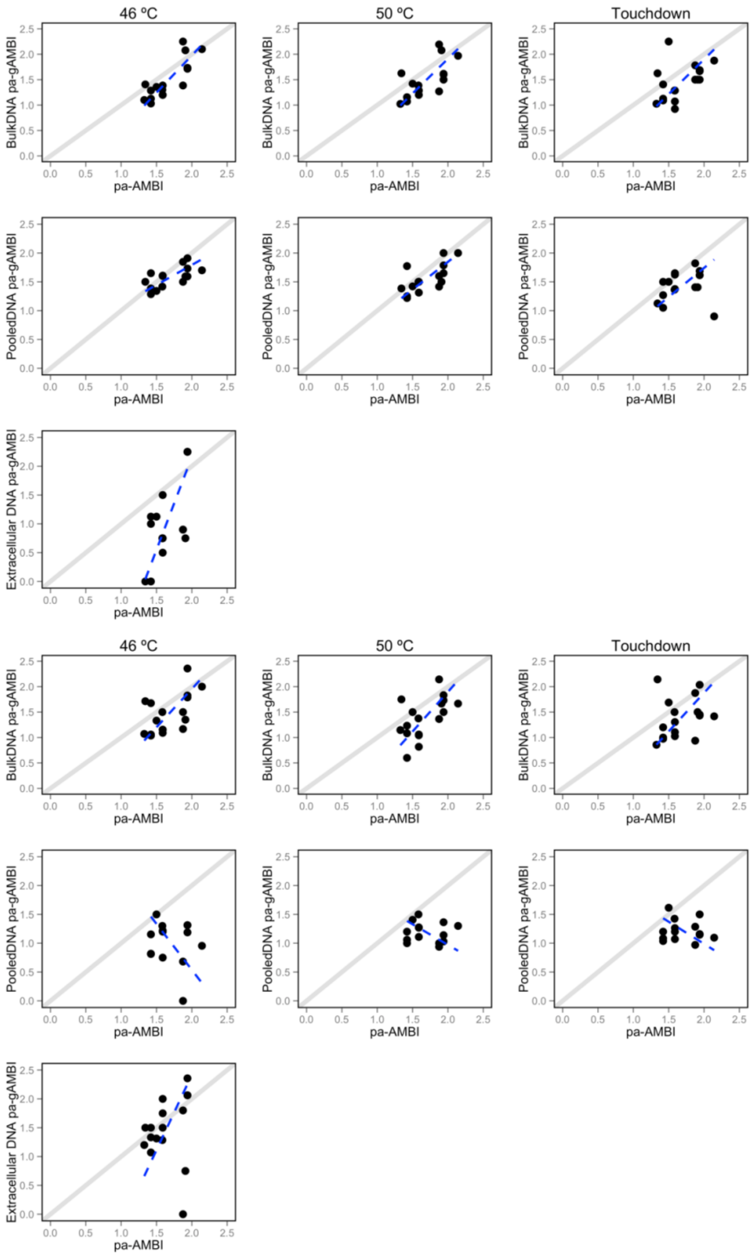

Figure 4.6. Relationship between (pa)AMBI and (pa)gAMBI values. For each DNA-based approach (extracellular, bulk and pooled DNA) and PCR condition (46 or 50 °C annealing temperatures or Touchdown profile) displayed separately for each barcode - *mlCOI* (top) and *folCOI* (bottom). Each dot shows the relationship between the (pa)AMBI (x-axis) and (pa)gAMBI value (y-axis) for each station. The dotted lines represent the results of model II regression and the diagonal showing perfect correlation between the two observations is depicted.

The effect of the PCR annealing temperature has been shown to affect retrieved taxonomic composition in bacterial and archaeal metabarcoding using the 16S rRNA gene (Sipos et al., 2007; Lee et al., 2012; Pinto and Raskin, 2012). Here we show that the use of inappropriate PCR conditions can also affect the final taxonomic assignment in metazoan metabarcoding analyses. Our results show that a constant low annealing temperature (46 or 50 °C) provides more accurate taxonomic inferences compared to the touchdown profile, which contrasts with previous studies (Hansen et al., 1998; Simpson et al., 2000; Leray et al., 2013). Moreover, it is well established that the more PCR cycles, the more spurious sequences and chimera are formed during PCR (Haas et al., 2011), which could explain the lower taxa detection rate when using the touchdown profile (which includes 5 more cycles). Further, the nature of the organisms and their size may bias DNA extraction (i.e. hard shells or chitin exoskeleton can prevent cell lysis and DNA from small organisms can be less effectively extracted). Here, we have ensured that DNA from all organisms is present in the pooled sample by pooling individually extracted DNAs, and show that the results of the pooled DNA and bulk extracted DNA are comparable.

Table 4.1. Results from the regression model between traditional and molecularly inferred (pa)AMBI values. *: Significant correlations ($p < 0.05$). TD: touchdown PCR profile.

Barcode	Condition	R^2	BIAS	RMSE
mlCOI	Bulk DNA 46 °C	0.68*	-0.18	0.28
	Bulk DNA 50 °C	0.49*	-0.21	0.32
	Bulk DNA TD	0.21	-0.22	0.39
	Pooled DNA 46 °C	0.41*	-0.11	0.22
	Pooled DNA 50 °C	0.46*	-0.14	0.23
	Pooled DNA TD	0.03	-0.26	0.40
	Extracellular DNA	0.42*	-0.59	0.83
folCOI	Bulk DNA 46 °C	0.33*	-0.21	0.37
	Bulk DNA 50 °C	0.49*	-0.29	0.43
	Bulk DNA TD	0.07	-0.29	0.49
	Pooled DNA 46 °C	0.02	-0.69	0.83
	Pooled DNA 50 °C	0.01	-0.52	0.59
	Pooled DNA TD	0.01	-0.48	0.57
	Extracellular DNA	0.15	-0.11	0.61

4.2. The use of extracellular DNA for biodiversity estimations

The extracellular DNA-based metabarcoding for biodiversity assessments has the potential of detecting big-size organisms in small samples, which facilitates sampling strategies and could resulting in a more cost-effective approach for environmental biomonitoring (Taberlet et al., 2012b; Thomsen et al., 2012a; Thomsen and Willerslev, 2015). Several studies have used extracellular DNA from the water column to detect vertebrates (Ficetola et al., 2008; Thomsen et al., 2012a; Valentini et al., 2016) freshwater macroinvertebrates (Goldberg et al., 2013; Mächler et al., 2014) and benthic eukaryotes (Guardiola et al., 2015; Pearman et al., 2016b). Yet, so far, this approach has not been proved valid for biodiversity assessment as no comparison with samples of known taxonomic composition has been performed. To our knowledge, only one attempt exists to detect the whole freshwater benthic macroinvertebrate community from extracellular DNA extracted from samples of known composition (Hajibabaei et al., 2012), but the authors used the preservative ethanol as controlled environment containing the free DNA rather than natural scenarios. In our analyses, only a small proportion of the taxa identified using morphological methods are retrieved using extracellular DNA present in the sediment. Indeed, even considering the taxa not identified through morphological taxonomy, the extracellular DNA-based analyses only identify 30 macroinvertebrate taxa over all stations, which is much lower than the total diversity inferred from morphology and from DNA extracted from the isolated organisms. Therefore, the striking differences obtained between morphological and extracellular DNA metabarcoding-based taxonomic inferences suggest that further studies are needed before using sediment extracellular DNA as a suitable source for macroinvertebrate biodiversity assessment; yet, more experiments testing the effect of sediment sample size, DNA degradation scenarios or DNA extraction protocols are required, as it is possible that sampling more deeply in the sediment, or using the water column provides better results, and/or that the optimal DNA extraction procedure has not been employed (Corinaldesi et al., 2005).

4.3. Effect misinterpreting community composition in environmental biomonitoring

Environmental biomonitoring programs rely on the detection of a wide range of taxonomic groups, which are usually amplified using universal primers (Leray et al., 2013). The abovementioned biases inherent to PCR-based analyses can lead to greater recovery of sequences of some species and the exclusion of others (Elbrecht and Leese, 2015; Piñol et al., 2015). Thus, it is important to see whether in samples containing species from numerous phyla, metabarcoding is also able to retrieve a high proportion of taxa that suffices for environmental monitoring. In general, we show a high percentage of recovery using bulk DNA among the nine different phyla identified using morphological approach. However, in our metabarcoding analyses, some taxa identified using morphological methodologies remain undetected using both short and long COI barcodes, whereas others appear only using metabarcoding. The species exclusively detected using metabarcoding represent potential cryptic species (e.g. *Thyasira flexuosa*/*Thyasira ferruginea* and *Ophiura texturata*/*Ophiura albida*) or unable to be classified based on morphological characters. Further, some additional identified taxa (i.e. two phyla detected from extracellular DNA (Platyhelminthes and Porifera)) may either represent organisms which had been missed by taxonomy based on morphology and metabarcoding from previously isolated organisms due to their small size (<1 mm) or detected due to the fact that the free DNA has been transported from other localities (Roussel et al., 2015).

Consequences of the misinterpretation of the taxonomic composition could result in erroneous biodiversity assessment, which may impede the implementation of DNA metabarcoding in regular biomonitoring programs (Cowart et al., 2015; Chariton et al., 2015; Lejzerowicz et al., 2015; Zaiko et al., 2015). In particular, calculation of biotic indices based on pollution tolerances assigned to the taxa retrieved from the sample (Maurer et al., 1999; Borja et al., 2000) may be affected by the approach used for taxonomic assignment. We show that, despite using the metabarcoding conditions that most accurately detect the morphologically identified taxa, some differences between both approaches are observed. Yet, in general, (pa)AMBI values obtained from metabarcoding analyses provide significant presence-absence community estimations and can be used for calculating biotic indices.

5. Conclusions

Metabarcoding represents a promising opportunity to overcome the time-consuming and high cost of morphology-based species identification. Thus, once the technique is proved as appropriate for providing accurate taxonomic identifications, it is anticipated that it will be routinely used in biomonitoring programs in the near future. Here, we demonstrate through an exhaustive benchmarking study design that, using the appropriate conditions, metabarcoding presents a great potential to characterize biodiversity and to provide accurate biotic indices. Thus, our findings will contribute to accelerating the implementation of metabarcoding for ecological status assessment.

Chapter 5

METABARCODING-BASED MARINE BIOMONITORING AND ASSESSMENT: FROM SCIENTIFIC CONCEPTS TO MANAGEMENT APPLICATIONS

In preparation: **Aylagas, E.,** Borja, A., Muxika, I., Irigoien, X. and Rodríguez-Ezpeleta, N. Metabarcoding-based marine biomonitoring and assessment: from scientific concepts to management directives.

1. Introduction

Molecular techniques for the characterization of biological communities are transforming marine ecology and represent a great opportunity for improving the conservation of the marine environment (Bik et al., 2012; Dafforn et al., 2014; Goldberg et al., 2015). Since Taberlet et al. (2012a) introduced the term 'DNA metabarcoding' to designate the high-throughput taxonomic characterization of complex samples (i.e. soil, water, sediment) using an amplified short fragment from the total extracted DNA, the technique is being evaluated for biodiversity assessment with ecosystem conservation purposes due to its advantages over traditional methodologies (Wangensteen and Turon, 2016). For example, metabarcoding has been shown effective for: (i) rapidly characterizing biological communities from environmental samples (i.e. water or sediment) (Dell'Anno et al., 2015), (ii) identifying species at all life cycle stages or degraded specimens (Ardura et al., 2016), (iii) detecting toxic species (e.g. toxic algae) (Penna and Galluzzi, 2013), (iv) understanding trophic interactions by analysing faecal samples or stomach contents (Albaina et al., 2016), (v) early detecting invasive species (Zaiko et al., 2015), (vi) providing high resolution for fish species detection (Thomsen et al., 2012a) and (vii) reliably characterizing indicators for marine ecological status assessment, such as the phytoplankton (Visco et al., 2015) or benthic macroinvertebrate communities (Lejzerowicz et al., 2015; Aylagas et al., 2016a). Yet, despite this evidenced potential of metabarcoding for accurate monitoring, the gap between the scientific literature and management applications suggests that this approach needs to be more effectively translated for policy making.

Recently, the European Water and Marine Strategy Framework Directives (WFD, 2000/60/EC and MSFD, 2008/56/EC), which have been developed for protecting and restoring the aquatic environment within recent European legislation, have highlighted the need to develop faster, more cost-effective and reliable tools for assessing marine environmental status (Heiskanen et al., 2016). Current assessments of biological components are hindered by the time and cost associated to the use of morphological identification and observational survey based monitoring, which require, in addition, often lacking high level of taxonomic expertise (Bacher, 2012; Pochon et al., 2013). Metabarcoding enables the simultaneous taxonomic characterization of hundreds of samples at relatively low cost and in just few weeks, which will allow answering real policy questions in a timely manner, at high spatial and temporal resolution, with relatively low required effort (Ji et al., 2013). Applying metabarcoding within European directives context can provide valuable insights into the status assessment of the marine environment through the evaluation of the different defined indicators (i.e. phytoplankton, macrophytes, zoobenthos and fish communities; see Heiskanen et al., 2016) and will bring greater capacity for efficient monitoring.

During the last few years, significant effort has been devoted to test, validate and review the potential of metabarcoding for accurately monitor biological communities (Danovaro et al., 2016; Goldberg et al., 2016). Several studies have highlighted some limitations of the technique. For example, the sample processing strategy can strongly influence the species detection success (Creer et al., 2016), barcode selection and PCR biases can prevent the detection of some taxa (Deagle et al., 2014), and estimations of individual number and/or biomass from metabarcoding data is not possible (Elbrecht and Leese, 2015)). Despite these downsides, the overall conclusion is the promise of metabarcoding for a rapid and cost-effective environmental management (Ji et al., 2013; Dowle et al., 2015). However, applications in the context of routine monitoring programs are lacking. Yet, the time to put metabarcoding into practice in real policy questions and verify its effectiveness has definitely come. Here, in order to test the potential of metabarcoding in a real management context, we have used the Basque (northern Spain) estuarine and coastal monitoring network program as a case of study. For that purpose, we have compared the macrobenthic community-based traditional and genomic versions of the AZTI´s Marine Biotic index (Borja et al., 2000; Aylagas et al., 2014), which are respectively inferred through morphology and metabarcoding. By comparing their performance for determining the marine ecological status, we evaluated the potential of metabarcoding-based biotic indices for routine monitoring programs.

2. Methods

2.1. Sampling and morphology-based taxonomic assignment

From the 51 locations of the Basque coast monitoring network (Borja et al., 2016a), a total of 11 coastal and 7 estuarine locations were selected for this study (Figure 5.1). This selection was based on previous morphological surveys performed in the 51 locations so that the selected locations present different macroinvertebrate taxonomic compositions, a wide range of AMBI values and come from sediments of different nature (i.e. coarse, medium and fine sands and, mud). From each location, four sediment samples were collected using a van Veen grab (0.07 – 0.1 m^2) in the coastal (32 – 107 m water depth range) and sublittoral estuarine stations (5 – 9 m water depth range), and using 0.5 m^2 sampling squares in the intertidal estuarine locations. Each sediment sample was sieved on site through a 1 mm mesh size; from each site, three samples were stored in formalin at room temperature and one in 96% ethanol (5:1 v/v) at 4 °C. From the formalin stored samples, macroinvertebrate specimens were counted and identified to the lowest possible taxonomic level; biomass of each taxa was determined as ash-free dry weight, obtained by drying at 80 °C for 48 h in an oven and incinerating at 450 °C for 4 h in a muffle furnace. The ethanol stored samples were processed for metabarcoding-based taxonomic assignment as detailed below.

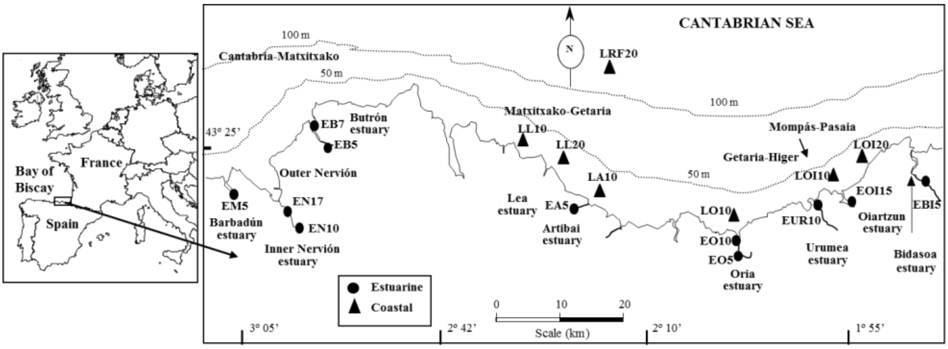

Figure 5.1. Location of the 18 stations, selected for this research, within the monitoring program network samples of the Basque coast (northern Spain).

2.2. Metabarcoding-based taxonomic assignment

The ethanol preserved samples were processed for genomic DNA extraction as described in Aylagas et al. (2016b). From the total extracted DNA, a 313 bp fragment of the mitochondrial cytochrome oxidase I (COI) gene was amplified using the degenerated metazoan universal primers mlCOIintF-dgHCO2198 (Leray et al., 2013) with overhang Illumina adapters as in Bourlat et al. (2016) with the following Index-PCR conditions: an initial 3 min denaturation step at 98 °C; 27 cycles of 10 sec at 98 °C, 30 sec at 46 °C and 45 sec at 72 °C; and a final 5 min extension at 72 °C. Equimolar concentrations of each dual-indexed PCR products were pooled and sequenced on the Illumina MiSeq platform with 2 × 300 bp paired-end v3 chemistry. Sequences were demultiplexed using the MiSeq Reporter version 2.4.60.8. Sequence analysis and taxonomic assignment were performed following the pipeline described in Aylagas and Rodríguez-Ezpeleta (2016).

2.3. Biotic indices calculation using morphology and metabarcoding-based taxonomic assignments

Differences between morphology and metabarcoding-based marine benthic macroinvertebrate taxonomic compositions obtained for each station were visualized using an in-house script (see details in Aylagas et al., 2016a), and tested using the Analysis of Molecular Variance (AMOVA) implemented in mothur (Schloss, 2009) from the distance matrix constructed based on taxa presence/absence and abundance using the Jaccard and Bray-Curtis indices, respectively, obtained using morphology and metabarcoding. Different versions of the AMBI were calculated based on pollution tolerances of the species present in a sample, with tolerance being expressed categorically as one of five ecological groups (I: sensitive to pressure, II: indifferent, III: tolerant, IV: opportunist of second order and V: opportunist of first order) using the AMBI 5.0 software (http://ambi.azti.es). AMBI, (B)AMBI and (pa)AMBI (Muxika et al., 2012) were calculated based on abundance, biomass (B) and presence/absence (pa) of the morphologically identified specimens, respectively. Alternatively, the genetic versions of the index, gAMBI and (pa)gAMBI (Aylagas et al., 2014) were calculated using metabarcoding derived read count and presence/absence for each identified taxa, respectively. The agreement between disturbance classifications obtained from the different versions of AMBI and gAMBI was analyzed using a Kappa analysis (Cohen, 1960).

2.4. Relative cost of metabarcoding vs. morphology-based ecological status assessment

To compare cost-efficiency and wait time between metabarcoding and morphology-based biotic index calculation, we estimated for each approach the time required and the costs involved from sample collection until calculation of the biotic index. The calculation of the costs included reagents, consumables and personnel needed to process samples and analyze data to an endpoint where the AMBI and gAMBI are obtained. In both cases, the personnel cost was considered 40 € hour^{-1}. The sequencing costs were calculated assuming multiplexing 96 samples on the Illumina MiSeq platform (note that Illumina kits for pooling and sequencing together up to 384 samples are available (Illumina, 2014)).

3. Results and Discussion

3.1. Do morphology and metabarcoding-based marine monitoring provide comparable conclusions?

From the total high-quality reads obtained, about 30% were assigned to macroinvertebrates (see Table 5.1) accounting for 114 different taxa, from which 72 were classified at species level being the remaining assigned at genus, family, class, order or phylum level. From the 207 morphologically identified taxa, an average of 20% (range from 0 to 66.6%) were detected using metabarcoding, being the taxonomic composition at the presence/absence and abundance level significantly different between both methodologies ($p < 0.001$). The percentage of common taxa between morphological replicates was about 50% (range from 18 to 81%), with no significant differences in taxonomic composition ($p > 0.05$).

These results reveal discrepancies in the community characterization using morphology and metabarcoding. Although some species were detected using both methodologies, between morphological replicates, in general, the same taxonomic groups are equally represented, whilst the metabarcoding sample differs in community composition (Figure 5.2). Using universal primers in bulk metabarcoding studies entails that some taxa present in the sample are not amplified and consequently undetected (Leray et al., 2013; Deagle et al., 2014; Gibson et al., 2014). Thus, it has been shown that metabarcoding typically recovers about 80% or even less of the taxa present in a mock community (Dowle et al., 2015; Aylagas et al., 2016a). Here, apart from primer biases, the different sample processing used for morphology and metabarcoding-based taxonomic identification (i.e. manual isolation of specimens vs. extracting DNA from a representative subsample) could have intensified these differences (Creer et al., 2016). Thus, using easy sample manipulation protocols for metabarcoding reduces the processing time and allows standardizing the technique, but carries the risk of favoring the detection of big-size specimens and the non-detection of small ones (Elbrecht and Leese, 2015) and the amplification of non-targeted taxa, such as other metazoans, fungi or protists present in the environmental sample (Lejzerowicz et al., 2015).

Table 5.1. Number of quality-filtered reads, percentage of reads corresponding to macroinvertebrate taxa and number of macroinvertebrate taxa identified per station using metabarcoding

Sample	Number of quality filtered reads	Percentage of macroinvertebrate quality filtered reads	Number of macroinvertebrate taxa
EA5	85,627	40.9	37
EB5	74,326	61.5	3
EB7	84,417	50.7	10
EBI5	34,048	5.5	13
EM5	75,409	21.9	14
EN10	72,198	0.6	6
EN17	51,535	55.2	10
EO10	80,335	8.0	13
EO5	57,912	58.7	11
EOI15	60,319	0.8	4
EOI15R	70,053	0.7	5
EOI15R2	50,191	0.7	3
EUR10	74,123	66.4	3
EUR10R	69,737	66.2	3
EUR10R2	68,360	66.0	3
LA10	50,933	1.4	15
LL10	83,842	0	0
LL20	62,595	3.3	8
LO10	85,539	12.2	5
LOI10	59,260	18.1	10
LOI20	73,589	0.2	2
LREF20	88,878	62.5	21

The different taxonomic compositions obtained with morphology and metabarcoding could imply that characterizing the macroinvertebrate community using metabarcoding provides contrasting management conclusions compared to that obtained using morphology. Using a known mixture of different macroinvertebrate species, metabarcoding has recently shown to provide taxonomic compositions that suffice for environmental monitoring, which is explained by the comparable percentage of taxa belonging to each ecological group detected using morphology and metabarcoding (Aylagas et al., 2016a).

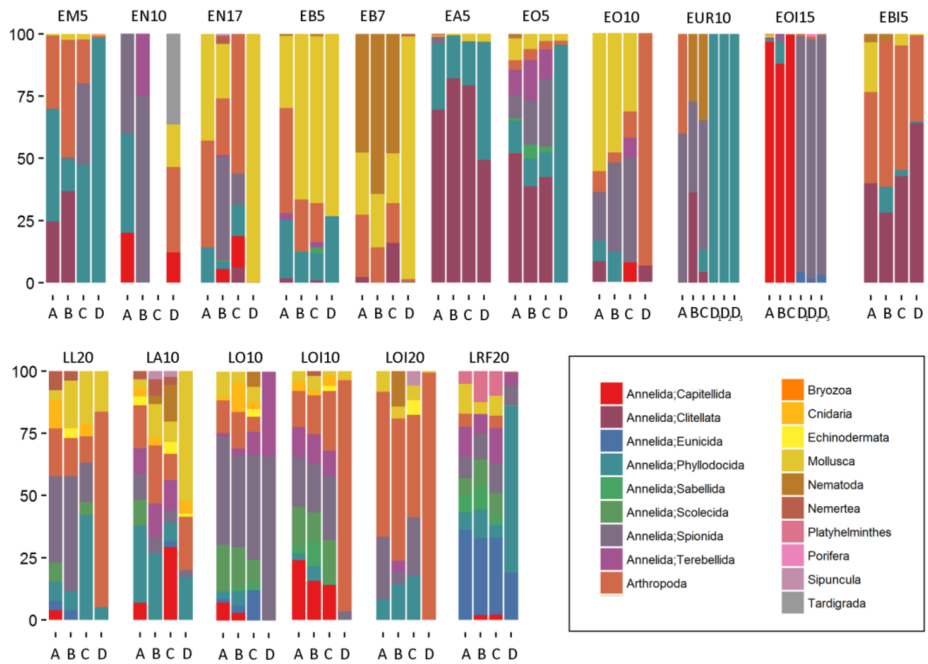

Figure 5.2. Proportions of macroinvertebrate taxa identified using morphology (A, B, C) and metabarcoding (D) for each station. The taxonomy is presented at the order level for Annelida, and at the phylum level for remaining groups. Metabarcoding replicates performed on two of the samples are shown as D_1, D_2 and D_3.

Hence, by calculating gAMBI and (pa)gAMBI we found good correlations ($r^2 > 0.65$) when compared with the different versions of AMBI (Figure 5.3). The agreement was "good" between (pa)AMBI and (pa)gAMBI, and between AMBI and gAMBI, and "excellent" between (B)AMBI and gAMBI, for which 14 out of the 17 stations were classified under the same ecological status category. Same comparisons performed between all versions of AMBI calculated from the three replicate samples resulted in very good correlations ($r^2 > 0.75$) although some discrepancies in the ecological status categories were observed (Figure S5.1). These results indicate that biomonitoring conclusions obtained using metabarcoding are comparable to those obtained using traditional methodologies, especially when using biomass.

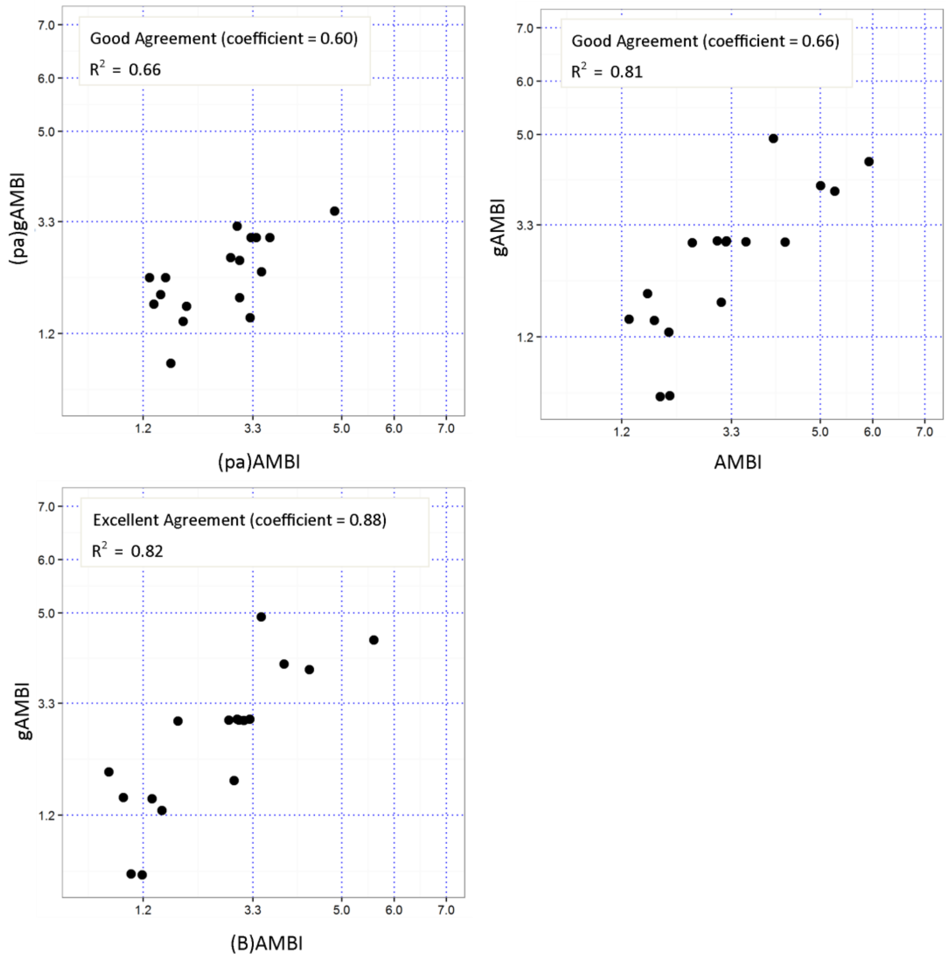

Figure 5.3. Correspondence between (pa)AMBI and (pa)gAMBI, AMBI and gAMBI, and (B)AMBI and gAMBI. Vertical and horizontal lines are depicted using threshold values to discriminate disturbance classes: undisturbed [0 - 1.2], slightly disturbed [1.3 - 3.3], moderately disturbed [3.4 - 5], heavily disturbed [5.1 - 6] and extremely disturbed [6.1 – 7]. The results of the Kappa analysis are shown.

Interestingly, the discrepancies found between the ecological status assigned to the same station using morphology or metabarcoding techniques are not higher than the ones found between the replicates analyzed through morphology. The boundaries of the quality classes used here to define the ecological status are the same for morphological and metabarcoding approaches. We suggest that by adapting these boundaries in metabarcoding, as done in the WFD intercalibration exercise (Borja et al., 2007), the final ecological status obtained by both approaches could fit better (Cai et al., 2014). Overall, the results obtained prove that metabarcoding is able to provide accurate management conclusions for the current European directives.

3.2. Can metabarcoding provide abundance metrics?

Metabarcoding provides information of species occurrence, which could be used for biodiversity assessments (Yu et al., 2012; Ficetola et al., 2015). However, biomonitoring usually relies on abundance metrics. Thus, finding correlation between sequence data and species abundance has focused the attention of a number of studies (Thomsen et al., 2012b; Goldberg et al., 2013; Evans et al., 2016). Some attempts to evaluate the relationships between macroinvertebrate species abundance or biomass and read number, obtained low associations from samples with a mixture of taxa with different abundances and biomass (Dowle et al., 2015; Elbrecht and Leese, 2015). Here, using the number of specimens and biomass of those taxa detected from morphology and metabarcoding methodologies, we found a significant positive correlation with number of reads using metabarcoding (Pearman´s $r = 0.84$, $p < 0.0001$ for abundance and Pearman´s $r = 0.8$, $p < 0.0001$ for biomass, Figure 5.4). Although this correlation could only be tested using just few taxa detected at the species level with both techniques, this finding represents a step forward for the implementation of metabarcoding in management.

Due to the difficulties in estimating species abundances or biomass using PCR-based approaches (Piñol et al., 2015), it has been suggested that biomonitoring using metabarcoding should rely on presence/absence metrics (Yu et al., 2012; Dowle et al., 2015). Hence, since the metabarcoding-based AMBI was developed, only the presence/absence version has been proposed to assess ecological status (Aylagas et al., 2014). Yet, this version of the biotic index reduces importance of dominant taxa to the overall community (Warwick et al., 2010; Muxika et al., 2012) and might produce erroneous assessments. For the first time, we show here that gAMBI, calculated using metabarcoding derived read counts, provides a more comprehensive evaluation of the ecological status than using the presence/absence version, (pa)gAMBI, and generates comparable results to those biotic indices inferred using taxa abundances or biomass. These results should be strongly considered for biomonitoring as they show the potential of metabarcoding for providing species relative proportions, and, therefore implementing DNA metabarcoding for biodiversity quantification in environmental management.

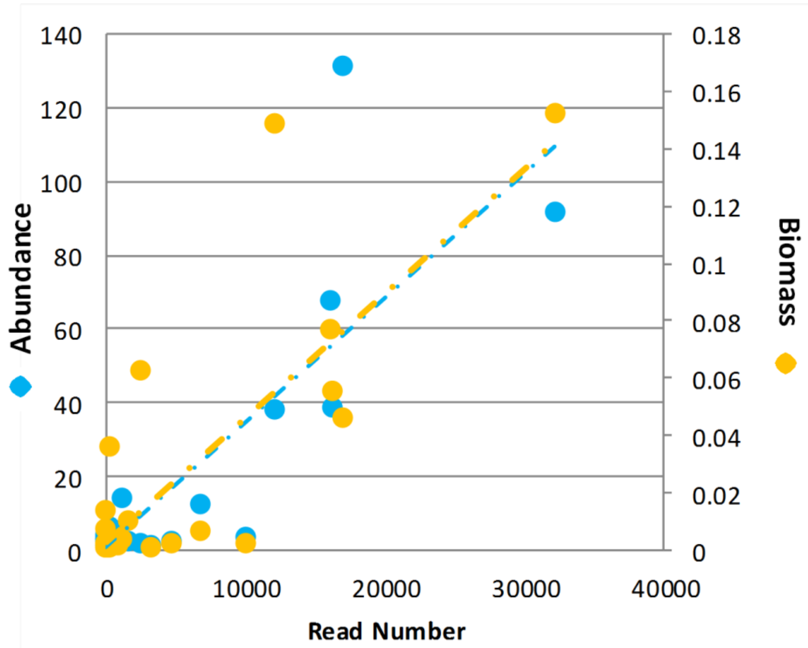

Figure. 5.4. Relationships between the abundance and biomass of each taxon at each site determined using morphology, and the number of reads generated for each taxon by metabarcoding.

3.3. Cost-effectiveness of metabarcoding for environmental biomonitoring

Since both, morphology and metabarcoding-based biotic indices yield similar results, both methodologies are able to detect changes in the ecological status of the community analyzed. The differences come from the costs to process the samples and wait time until results are obtained. Metabarcoding-based biomonitoring reduces costs when several samples are analyzed simultaneously (Figure 5.5). We inferred that above 20 samples analyzed, metabarcoding is more cost-effective. The greater advantage of metabarcoding is the number of samples that can be sequenced simultaneously, which on top of decreasing costs it increases speed of the process. In the case of AMBI, the estimated time to calculate the index for one sample is about 6.5 hours (assuming expertise in the classification of the specimens) independently on the number of samples analyzed. In contrast, metabarcoding allows the calculation of gAMBI from 96 samples in about 190 hours (less than 2 hours per sample).

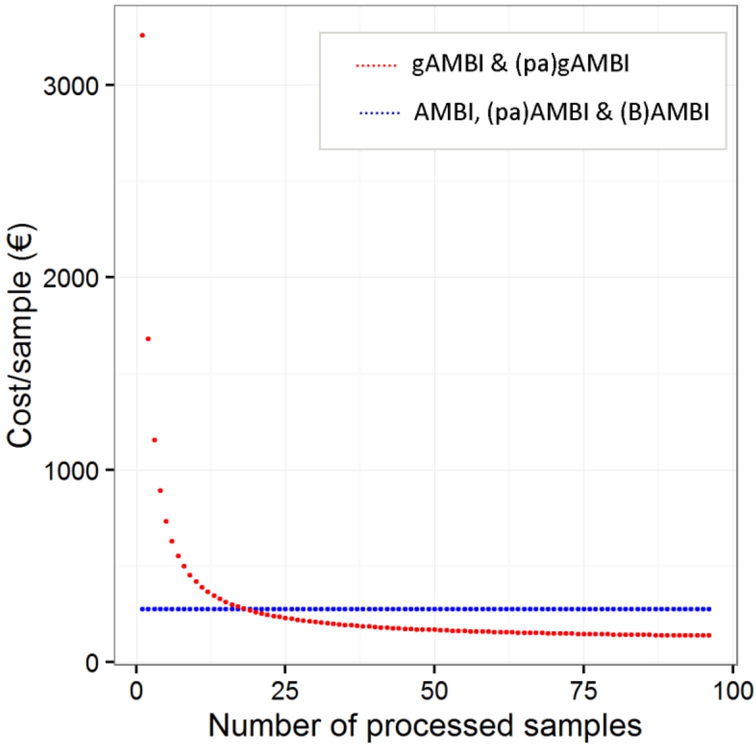

Figure 5.5. Costs associated to the calculation of morphology and metabarcoding-based biotic indices. The figure shows how the cost for calculating the genetic version of the biotic index (gAMBI) for one sample decreases with the number of samples analyzed.

These gains in cost-efficiency made possible by metabarcoding can greatly benefit large-scale biomonitoring programs (Ji et al., 2013). The cheaper alternative of DNA methods for species identification has been previously shown on a single targeted taxa (e.g. Biggs et al., 2015; Sigsgaard et al., 2015; Smart et al., 2016). Here, we have extended the cost-efficiency study to the whole sampled benthic macroinvertebrate community and integrated the costs analysis in a monitoring program. We estimated that the Basque monitoring network (Borja et al., 2016a) spends annually 44,000 € evaluating the macroinvertebrate ecological status of 51 estuarine and coastal locations (analyzing 3 sample replicates per location). Analyzing samples and providing biomonitoring results using morphological identification of the taxa detected requires about 1,000 hours for all locations (excluding field sampling time), which in practice represents 6 months of work. Based on current cost estimates, metabarcoding enables reducing three times the time required to analyze the same number of samples within the monitoring program (gAMBI can be calculated for the complete monitoring program in about 280 hours, which in practice represents less than 2 months of work) and two times the costs involved from sample collection untill calculation of the biotic index (the economical investment required to obtained gAMBI is about 20,000 €). Thus, the cheaper alternative of metabarcoding for providing results in a great number of samples in a relatively short period of time offers the opportunity for implementing the technique in large-scale biomonitoring programs.

3.4. Remarks for routinely applying metabarcoding in large-scale biomonitoring

Due to the potential of metabarcoding for cost effectively and comprehensively assessing biological communities, metabarcoding-based biomonitoring can be reliably used as a complementary tool to the currently established methodologies for ecological status assessments. Here, we provide some suggestions for applying metabarcoding in large-scale biomonitoring programs:

a) The indicator AMBI is officially used in many European countries and has been tested in America, Africa, Asia and Oceania (Borja et al., 2015b), where examples of its application can be found (Ranasinghe et al., 2012; Valenca and Santos, 2012). We suggest that one way to extend in time and space existing monitoring and introducing it in new areas is using the more cost-effective gAMBI, for which establishing and following common protocols from sample collection until calculation of gAMBI is required.

b) The improvement of the reference database will enable metabarcoding to be more reliably used in monitoring surveys. We suggest collaboration between molecular ecologists and taxonomists for accurate characterization of species and deposition of high quality sequences in public databases, starting from the most frequent (Aylagas et al., 2014). This will require some investment as the barcoding costs are estimated in about 5 € per individual (see Stein et al., 2014).

c) As it has been done for macroinvertebrates, we suggest assessing the potential of metabarcoding for characterizing other biological communities. Thus, currently used indicators which are monitored using traditional techniques, could be more cost-effectively characterized using metabarcoding and increase speed and accuracy in monitoring programs within European directives. The technique can be easily integrated by adapting some components of the approach to the target community such as selection of barcode, primer pair and reference library.

4. Conclusion

In the light of the results obtained in this study we confirm the suitability of metabarcoding to reliably, rapidly and cost-effectively respond to environmental management needs. The difficulty in obtaining comparable results from different locations by assessing the ecological status using morphology is aggravated by the subjectivity of the taxonomists identifying the samples, which could be overcome if metabarcoding techniques are applied. Also, conservation budgets are limited, making monitoring programs decrease sampling frequency, which reduces the biomonitoring resolution and impedes a comprehensive assessment of the ecosystem integrity, limiting results for environmental management. Metabarcoding will allow increasing the frequency of monitoring programs and obtaining comparable results for large-scale monitoring in just few weeks. Therefore, since metabarcoding has demonstrated to represent a reliable and cost-effective method for ecological status assessment, its integration in routinely biomonitoring programs will greatly benefit environmental management.

Chapter 6

A BACTERIAL COMMUNITY-BASED INDEX TO ASSESS THE ECOLOGICAL STATUS OF ESTUARINE AND COASTAL ENVIRONMENTS

Published as: **Aylagas, E.,** Á. Borja, M. Tangherlini, A. Dell'Anno, C. Corinaldesi, C. T. Michell, X. Irigoien, R. Danovaro, N. Rodríguez-Ezpeleta (in press), A bacterial community-based index to assess the ecological status of estuarine and coastal environments. *Marine Pollution Bulletin***.** doi: 10.1016/j.marpolbul.2016.10.050.

1. Introduction

The ability to monitor marine ecosystems is crucial to avoid adverse effects of anthropogenic activities (Halpern et al., 2008). The environmental quality of these systems is comprehensively evaluated through the analysis of physico-chemical (e.g. nutrients, contaminants, organic matter content) and biological (phytoplankton, zooplankton, benthic invertebrates, algae, seagrasses and fishes) components (Borja et al., 2008). In particular, biotic indices are extensively used for environmental monitoring estuarine and coastal ecosystems, and most rely on the analysis of benthic macroinvertebrate communities (Diaz et al., 2004; Borja et al., 2015b), which present short- and mid- term responses to a wide variety of anthropogenic impacts.

Despite being sensitive indicators of human-induced impacts, bacterial assemblages have not been adequately considered for the analysis of ecosystem functioning and in biomonitoring (Danovaro and Pusceddu, 2007; Caruso et al., 2015). This biological component can respond rapidly in terms of diversity, physiology and functional characteristics (Doiron et al., 2012; Hajipour et al., 2012; Sun et al., 2013) to environmental changes due to natural or anthropogenic pressures (Zhang et al., 2008b; Chiellini et al., 2013; Zhang et al., 2014). However, studies addressing the use of bacterial assemblages as indicators of marine ecological status are very limited due to the high complexity of microbial communities in terms of diversity and functioning in natural ecosystems (Nogales et al., 2011) and the difficulties in the taxonomic identification of environmental bacteria compared to macro-organisms.

Fortunately, the advent of molecular methods based on HTS technologies has provided new insights into the knowledge of bacterial assemblage composition from different marine environments (Wang et al., 2012; Sun et al., 2013; Tan et al., 2015b; Ziegler et al., 2016). It has also allowed the identification of key microorganisms involved in important ecosystem processes (Gilbride et al., 2006; Tan et al., 2015a) and has increased our capability of characterizing bacterial assemblages in several samples simultaneously and rapidly at low cost (Ferrera and Sanchez, 2016). Therefore, the fast response of bacteria to environmental changes and the easy access of cost-efficient HTS technologies allow the integration of taxonomic composition of bacterial assemblages as indicators of ecological quality (Caruso et al., 2015), complementing the information provided by benthic metazoan communities as an early warning sign to assess impacts.

Currently, there is an increasing concern for performing integrative assessments of marine waters under an ecosystem approach, including all components from microorganisms to mammals (Borja et al., 2008). In particular, two main directives are aimed at safeguarding the integrity of aquatic systems within Europe: (i) the Water Framework Directive (WFD; 2000/60/EC), in freshwater, estuarine and coastal areas; and (ii) the Marine Strategy Framework Directive (MSFD; 2008/56/EC), in marine waters. Several methods have been developed to assess the status within the WFD (Birk et al., 2012) and the MSFD (Borja et al., 2016b) including the complementary use of different indicators to evaluate ecosystem integrity. Current assessment methods present an important gap regarding the use of microbial assemblages (Heiskanen et al., 2016). In this sense, the inclusion of the microbial component in regular monitoring programs may be key for a better understanding of the connection between biodiversity and ecosystem function (Strong et al., 2015), and will provide valuable information for detecting the effects of anthropogenic pressures on marine environments (Caruso et al., 2015).

Here we used HTS of the 16S rRNA gene to analyze the benthic prokaryotic assemblage composition of 51 coastal and estuarine locations of the Basque coast, northern Spain, under different anthropogenic pressures. Due to the high level of human-induced impacts that historically have affected the area, the ecological quality of the Basque coast has been monitored during the past two decades (Borja et al., 2013a); however, bacterial assemblages have not been considered yet within this monitoring program network. The aim of our study is to analyze the bacterial assemblage composition at locations subjected to different anthropogenic impacts in order to: (i) develop a new index based on the bacterial community composition for the ecological status assessment of estuarine and coastal environments, (ii) validate the index by determining its response to different anthropogenic pressures, and (iii) evaluate the performance of the index compared with a biotic index based on marine benthic macroinvertebrates.

2. Material and methods

2.1. Study area and sample collection

Historically, the ecological quality of the Basque coast has been significantly altered by human activities (Cearreta et al., 2000). Most of the estuaries and coastal areas have supported urban and industrial discharges (resulting in an increase of organic matter and consumption of oxygen) and the construction of different artificial structures that alter hydrological features (dykes and port construction, dredging, sediment disposal and land reclamation) (Borja et al., 2009c). In order to determine their impact, the Basque coast and estuaries have been monitored since 1994 through a monitoring program network that includes the analyses of physico-chemical (in water and sediment, such as concentrations of metals and organic compounds) and biological (e.g. macroinvertebrates) components (Borja et al., 2009c; Borja et al., 2016a).

Within the monitoring network, three replicates for benthic macroinvertebrates and one for sediment analyses are undertaken in 32 stations located in 14 estuarine locations, and in 19 stations located in 4 coastal areas not directly influenced by freshwaters (Figure 6.1). A summary of the main significant pressures affecting the area is provided in Table 6.1. For this study, four sediment samples were collected in winter (January-February) 2013 from each of the 51 stations using a van Veen grab in the coastal (30 – 113 m water depth range) and sublittoral estuarine stations (5 – 24 m water depth range), and using 0.5 m^2 sampling squares in the intertidal estuarine locations. Three replicates were sieved on site through a 1 mm mesh size and preserve in formalin for morphological identification of macroinvertebrates. Surface sediment subsamples (top 1 cm) were collected from the fourth sample and stored in sterile 15 mL falcon tubes at -80 °C for the analysis of the composition of microbial assemblages. Additional sediment subsamples were used for the determination of grain size, organic matter content and inorganic and organic contaminant concentrations. In the same samples, the redox potential (Eh) was measured in the top 10 mm of the sediment with a combined Pt-ring electrode (Langmuir, 1971). Salinity was measured in situ using a CTD-Seabird 25 multiprobe. According to the pressures affecting to each site, the outer part of the Lea and Butroe estuaries are considered the most pristine estuarine areas. In contrast, the Nervion and Oiartzun estuarine stations are the most impacted sites along the study area.

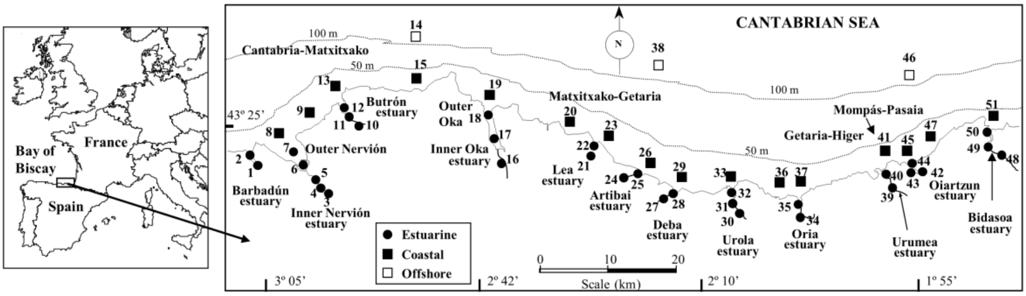

Figure 6.1. Locations of the 51 stations within the Monitoring Program Network sampled within the Basque coast (northern Spain).

2.2. Sediment characteristics

The grain size of sediments was determined by using the dry sieving technique at 60 °C (Folk, 1974). The organic matter (OM) content was determined by loss of weight on ignition at 450 °C during 6 h (Dean, 1974). Metal concentrations (Cd, Cr, Cu, Ni, Hg, Pb and Zn) were analyzed within the fine-sediment content sediment fraction (< 63 µm) (Kersten and Smedes, 2002), obtained by sieving samples previously oven-dried at 60 °C. Analysis were performed using the acid-extractable metal concentration method by means of an acid mixture of HCl/HNO3 (1:2, v/v) (Menchaca et al., 2012). The concentration of polychlorinated biphenyls (PCBs: PCB-28, PCB-52, PCB-101, PCB-118, PCB-138, PCB-153 and PCB-180) and polycyclic aromatic hydrocarbons (PAHs: acenaphthene, acenaphthylene, antracene, fluorene, naphthalene, phenanthrene, benzo(*a*)antracene, benzo(*a*)pyrene, chrysene, dibenzo(*a*,*h*)antracene, fluoranthene, pyrene, benzo(*e*)pyrene, benzo(*b*)fluorantene, indeno(1,2,3)pyrene, and benzo(*g*,*h*,*i*)perylene) was determined following the procedure described by Bartolomé et al. (2005).

Table 6.1. Main significant pressures affecting the ecological quality of the stations selected within the estuarine and coastal areas, and actions taken to mitigate such impacts on the environment, within the Basque Country. The year(s) of the pressure, or mitigating action, are shown in brackets (modified from Borja et al. (2016a). WTP = Water Treatment Plant. (*) stations not included in the analysis (low reads number).

Water body	Station	Pressures	Actions
Barbadun	1	Oil refinery (1968 to present), urban discharges	Oil refinery effluent deviation (1999)
	2	Influence from oil refinery, small urban discharges	WTP (1988)
Inner Nerbioi	3	Changes in morphology, pollutants	WTP (1990, 2002), mining closure (1990s), steel industry closure (1995)
	4	Changes in morphology, steel industry (19th century-1990s), pollutants	
	5	Changes in morphology, steel industry (19th century-1990s)	
Outer Nerbioi	6	Dredging, port enlargement (1992-present)	Sewerage scheme (1993-present)
	7	Port enlargement (1992-present)	
Butroe	10	Small urban discharges	Sewerage scheme (1993–present), WTP (1997, 2006)
	11	Small urban discharges	
	12	Marina construction (1993), dredging (1991, 2008, 2008)	
Inner Oka	16	Urban discharges	WTP (1974)
Outer Oka	17	Shipyard (1943-present), dredging (1995, 1998, 1999, 2003, 2009), urban discharges	Some discharge deviation, WTP (1999, under construction)
	18	Small urban discharges, dredging (1995, 1998, 1999, 2003, 2009)	
Lea	21	Small urban discharges	WTP (1995, 2005), basin water treatment
	22	Small urban discharges, port	Discharge deviation (1993–1995)
Artibai	24	Urban & industrial discharges	Basin water treatment, discharge deviation, WTP (2011)
	25	Dredging (1998, 2003, 2009), port	
Deba	27	Urban and industrial discharges	Basin pollutants removal (since 1998), WTP (1996)
	28	Marina construction (1999)	
Urola	30	Small urban discharges	WTP (2007)
	31	Dredging (2000–2005)	
	32	Dredging (2000–2005), port construction (1997–1998)	
Oria	34	Urban discharges	Basin water treatment, WTP (2000), discharge deviation (2016)
	35	Land-claim (2001), port construction (2005)	

Urumea	39	Urban constructions	Discharge deviation (2001), WTP (2006)
	40	Urban and industrial discharges	
Oiartzun	42	Urban and industrial discharges, pollutants, changes in morphology	Discharge deviation (1996, 2001), WTP (2007)
	43*	Urban and industrial discharges, pollutants, changes in morphology, dredging (decreasing since 1995), port	
	44	Urban and industrial discharges, pollutants, changes in morphology, dredging (decreasing since 1995), port	
Bidasoa	48	Urban discharges	Discharge deviation (since 1999), WTP (2003)
	49	Port/marina construction (1992, 1997, 1998), channelling	
	50	Port/marina construction (1992, 1997, 1998), channelling	
Cantabria–Matxitxako	8	Urban discharges, industry, intermittent dredging disposal	WTP and submarine outfall (2013)
	9	Blast furnace slag disposal (1980–1995)	
	13*	Urban discharges	
	15*	Undisturbed	
Matxitxako –Getaria	19	Small urban discharges, intermittent dredging disposal	WTP (2014), basin water treatment
	20	Undisturbed	
	23	Small urban discharges	
	26*	Small urban discharges, intermittent dredging disposal	
	29	Small urban discharges	
	33	Small urban discharges	
Getaria–Higer	36	Sediment disposal (2001- 2003), urban discharges (2000)	WTP and submarine outfall (2003)
	37	Intermittent dredging disposal, urban discharges (2000)	
	45	Sediment disposal (2001- 2003)	
	47*	Small urban discharges	
	51	Urban discharges	
Mompas	41	Urban and industrial discharges (1970-2001)	WTP (2007) and submarine outfall (2001)
Offshore	**14***	No pressures	No actions
	38	No pressures	
	46	Sinks of suspended particulate matter	

2.3. Anthropogenic pressure

The gradient of anthropogenic pressure was determined using the marine regional Sediment Quality Guidelines (SQG) for metals (Menchaca et al., 2012), PCBs and PAHs (Menchaca et al., 2014) established for the Basque Country, together with redox potential and organic matter content as additional proxies of anthropogenic pressure. Hence, we have derived an Index of Pressure, by normalizing each variable to a value ranging from 0 (total absence of pressure for a pollutant) to 5 (maximum pressure). Concentrations above SQG represent adverse effects, and we considered this limit as 2 (the boundary between acceptable/not acceptable levels) in the scale 0-5.The maximum pressure has been considered as the maximum concentration registered in the area for 1995-2014 time series. For example, to calculate the value: (i) SQG for Pb is 78 mg kg^{-1} (equal to 2 in the Pressure Index scale); in station 1, Pb concentration is 46.2 mg kg^{-1}, meaning that in the scale from 0 to 5, this value is equivalent to 1.18; (ii) for total PCBs the SQG value is 24.6 µg kg^{-1} (equal to 2 in the Pressure Index scale); in station 5, the total PCBs concentration is 197 µg kg^{-1}, which is normalized to 4.78 in the Pressure Index. After normalizing the value for each component, the average of all individual predictors for each station was calculated to define the total Pressure Index, which was used as an independent way to validate the bacterial index.

2.4. Bacterial community analysis

Total DNA was extracted from 1.5 gr of the surface sediment subsamples using the Power Soil® DNA Isolation kit (Mobio) following manufacturer´s instructions. A 250 bp fragment of the V4 region of the 16S rRNA gene was amplified using the bacterial/archaeal universal primers 419F- CAGCMGCCGCGGTAA and 806R-GGACTACHVGGGTWTCTAAT (Klindworth et al., 2012) with overhang Illumina adapters. Each sample was amplified three times in a total volume reaction of 25 µL using 12.5 µL of 2x KAPA HiFi HotStart ReadyMix, 5 µL of each forward and reverse primers and 2.5 µL of normalized DNA (5 ng/µL). PCR reactions were as follows: 95 °C for 2 min, 30 cycles of 98 °C for 20 s, 50 °C for 60 s and 72 °C for 90 s and a final extension of 5 min at 72 °C. The three replicates were pooled, purified using AMPure XP beads (Beckman Coulter) and used as a template for the attachment of dual Illumina indices in a second PCR round following the “16S Metagenomic Sequencing Library Preparation” protocol (Illumina). Negative controls were added to all reactions. Equimolar concentrations of each final dual-indexed PCR product were pooled and sequenced on the Illumina Miseq, 2 × 300 bp paired-end v3 chemistry according to the manufacture´s specifications. Sequences were demultiplexed using the Miseq Reporter version 2.4.60.8. The data are accessible via GenBank under the SRA accession numbers: SRP075964.

Primer and adaptor sequences were removed using cutadapt (Martin, 2011), and paired-end reads were then merged with FLASH (Magoč and Salzberg, 2011) using a maximum overlap of 290 bp and the default 10 bp minimum overlap. At this step, a total of 6 samples produced less than 8,000 reads and were removed; the remaining 45 were kept for downstream analysis. Merged pairs were subsequently quality-trimmed with USEARCH (Edgar, 2010) to remove sequences with a maximum expected error >0.75 (Edgar and Flyvbjerg, 2015) and shorter than 100 bp, and finally trimmed to a common length of 250 bp. The retained high-quality reads were then submitted to the Minimum Entropy Decomposition pipeline (Eren et al., 2015) to identify Operational Taxonomic Units (OTUs), allowing for 3 nucleotides of maximum variation and a minimum of 250 sequences per cluster; outliers were relocated after the first stage of clustering. Taxonomic assignment of OTU node representatives was carried out with the SINA aligner on the SILVA database, release 123 (Quast et al., 2013). Standard OTU richness and Shannon index were calculated from OTUs generated at 97% of similarity by the USEARCH pipeline from high quality reads, after removal of singletons, using the *fasta_diversity* command and the evenness index was calculated following the Pielou's index.

2.5. Macroinvertebrate community analysis and AMBI

From each sieved replicate, macroinvertebrate specimens were separated under a stereomicroscope and identified by expert taxonomists at the lowest possible taxonomic level. From the list of macroinvertebrate taxa and abundances obtained from each station, the AMBI (Borja et al., 2000) was determined using AMBI 5.0 software (http://ambi.azti.es). AMBI is based on the response of macroinvertebrate species to gradients of pressure and requires the classification of each species into one of five ecological groups (EG): EGI, sensitive species to pressure; EGII, indifferent species; EGIII, tolerant species; EGIV, second order opportunistic species; and EGV, first order opportunistic species. The relative proportion of each group in a sample provides a number ranging from 0 (undisturbed sample) to 7 (extremely disturbed), being 7 the azoic situation.

2.6. Developing a bacterial biotic index

To develop the bacterial biotic index we followed the strategy of the AMBI. In the case of bacteria, as there was not sufficient previous information available to classify different taxa into one of the five groups, we classified them into two ecological groups: EGI*,* as taxa not associated with pollution inputs (including sensitive and indifferent taxa) and EGIII, as taxa associated with pollution inputs (including tolerant and opportunistic taxa). A similar approach has been applied with invertebrates in some variations of AMBI (e.g. in Bentix; Simboura and Zenetos (2002)).

Since the index proposed here is based on the microbial assemblage composition (micro) analyze through 16S rRNA gene amplicon sequencing and inspired in AMBI, we have named it microgAMBI. To calculate it, the formula is based on the relative bacterial family abundance of each group, within each sample, to obtain a continuous index, where:

$$\text{microgAMBI} = [(0 \times \%EGI) + (6 \times \%EGIII)] / 100$$

The range of microgAMBI is from 0 to 6, where 0 represents 100% of sequences assigned to EGI, whilst 6 represents 100% of sequences assigned to EGIII. We have not considered an azoic situation from microgAMBI since azoic situation for bacteria in marine environments is extremely rare. We found 39 different families which were assigned to EGIII based on literature records, taking into account their ecological role associated with pollution inputs (see Table S6.1). This potential role includes: (i) dominance in organic matter-enriched sediments; (ii) organic pollution response; (iii) dominant presence in anoxic methane-rich sediments; (iv) identification as nitrite oxidizer and related to nitrogen inputs; (v) presence in sulfide-rich wastewaters; (vi) presence in wastewater treatment plants; (vii) role in methanogenic degradation of alkanes; (viii) role in aromatic compounds biodegradation, including petroleum products pollution, as complex PAHs; and (ix) potential pathogens. From the 226 prokaryotic taxa retrieved from the sequence taxonomic analysis clustered at family level, the 39 potential indicators of pollution families were assigned to EGIII. From the remaining 187 families, 169 were assigned to EGI as they did not include any member from the potential indicator families and 18 were not assigned (i.e., cases where taxonomic assignment gave unknown families that belong to a class or order for which at least one family has been reported as potential indicator of pollution (e.g.

Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales; unknown); see list in Table S6.2).

Using the series of continuous values, an ecological status classification has been proposed, following the criteria of the WFD. This classification is based on the contribution of each ecological group to the final microgAMBI value. We have considered the boundary between good and moderate status when there is 60% of EGI and 40% of EGIII; this means that the boundary is 2.4 (Table 6.2).

Table 6.2. Proposed ecological status quality classes for microgAMBI, and class boundaries, determined from the contribution of each ecological group (EG).

Class boundaries	Contribution of each group	Ecological Status
0 < microgAMBI ≤ 1.2	> 80% EGI	High
1.3 < microgAMBI ≤ 2.4	60% EGI and 40% EGIII	Good
2.5 < microgAMBI ≤ 3.6	40% EGI and 60% EGIII	Moderate
3.7 < microgAMBI ≤ 4.8	20% EGI and 80% EGIII	Poor
4.9 < microgAMBI ≤ 6	> 80% EGIII	Bad

2.7. Statistical analysis

Sequence abundance data per cluster was transformed to relative abundance (as a percentage of the total number of sequences per sample) and used as an input for the statistical analyses carried out with the Primer 6 + statistical package (Plymouth Marine Laboratory, UK). A similarity matrix calculated using the Bray-Curtis distance after square root transformation of the relative abundance data was used to perform hierarchical clustering based on group average (the mean distance apart of two groups, averaging over all between group pairs) and for non-metric multidimensional scaling (nMDS). Similarity percentage (SIMPER) analysis, based on the Bray-Curtis similarity matrix, was carried out to determine the taxa contributing to the top 10% of differences between groups and similarities within each group. The contribution of these taxa to differences between samples (or groups) were analyzed using the two-group Welch´s *t*-test (a variation of the Student´s *t*-test used when two groups cannot be assumed to have equal variance) on the Statistical Analysis of Metagenomic Profiles (STAMP; Parks et al., 2014). An nMDS plot was constructed to explore samples segregation according to abiotic parameters and after redundancies removal, 9 variables were selected (salinity, organic matter (%), redox potential, ∑PCB, ∑PAH and concentration of Zn, Pb, Cd and Hg) to examine the relationships between bacterial assemblage composition and environmental variables using distance-based linear models (DISTLM). The most parsimonious model was re-run using only the variables selected for this model and distance-based redundancy analysis (dbRDA) was performed to visualize the influence of predictor variables identified by the DISTLM. The relationship between microgAMBI and the Pressure Index, microgAMBI and AMBI, and putative indicators taxa and contaminant variables, was calculated with a linear model regression analysis (lm) performed in R (R Development Core Team, 2014). The significant response of the putative indicators taxa to contaminant variables was evaluated using forward-selection.

3. Results

3.1. Environmental characteristics and anthropogenic pressures

A summary of the physico-chemical attributes examined for the sediment samples together with concentration of contaminants are presented in Table S6.3. Salinity of the bottom waters ranged from 0.07 to 35.1 in the estuarine stations and from 35.1 and 35.7 in coastal stations. The organic matter content in the sediment was characterized by a wide variability in the estuarine stations with values ranging from 0.4 to 10.8%, whilst coastal values ranged from 0.9 to 2.9%. The redox potential showed positive values across coastal and estuarine areas except for one station of the Inner Nervion estuary. Metal concentration exceeded the values established by the SQG for 6 out of the 7 analyzed variables in stations 4, 5 (in the Inner Nervion estuary), 42, 44 (in the Oiartzun estuary) and 49 (in the Bidasoa estuary). In particular, Zn concentrations at stations 42 and 50 and Pb concentrations at station 50 were more than one order of magnitude higher than the SQG values. The total concentrations of aromatic compounds exceeded the SQG for PAHs at the estuarine stations 4, 5 (in the inner Nervion estuary) and 21 (in the Lea estuary), and for PCBs at 10 stations belonging to the Nervion, Lea, Deba, Urola, Oria and Oiartzun estuaries.

Overall, the pressure index ranged from 0.7 to 1.5 and from 0.5 to 3.8 at coastal and estuarine stations, respectively (Table S6.3). The highest values were obtained at Inner Nervion, Oiartzun and Bidasoa estuaries and the pressure showed, in general, a decreasing trend from the inner to the outer part of the estuaries.

3.2. Bacterial diversity

A total of 5,651,697 combined read pairs were obtained, which were further reduced to 5,087,534 after quality check and trimming. Across all samples, the number of OTUs generated at 97% of similarity ranged from 670 to 9,703 with mean values of 4,625 and 2,762 OTUs for estuarine and coastal stations, respectively (Table 6.3). The Shannon index indicated a diversity range across all samples from 5.3 to 7.4 with the highest diversity found at estuarine areas (stations 22 and 48). The Pielou's index ranged from 0.63 to 0.86 and showed the lowest evenness at estuarine stations 3, 27, 28 and 34, for which values lower than 0.7 were found.

Table 6.3. Alpha diversity metrics calculated for each station of estuarine (E) and coastal (C) water bodies.

Station	Richness	Shannon index	Pielou's evenness	Station	Richness	Shannon index	Pielou's evenness
1 (E)	1,886	6.3	0.835	28 (E)	6,561	6.1	0.694
2 (E)	2,190	6.2	0.806	29 (C)	2,691	6.4	0.810
3 (E)	3,405	5.3	0.652	30 (E)	4,614	6.2	0.735
4 (E)	670	5.4	0.830	31 (E)	3,745	6.6	0.802
5 (E)	3,948	6.3	0.761	32 (E)	4,732	6.8	0.804
6 (E)	2,845	6.1	0.767	33 (C)	3,516	6.7	0.821
7 (E)	2,187	6.0	0.780	34 (E)	4,500	5.6	0.666
8 (C)	2,230	6.3	0.817	35 (E)	6,969	7.1	0.802
9 (C)	3,977	6.5	0.784	36 (C)	3,121	6.5	0.808
10 (E)	3,046	6.5	0.810	37 (C)	1710	6.1	0.819
11 (E)	2,551	6.1	0.778	38 (C)	4,004	6.4	0.772
12 (E)	2,342	6.3	0.812	39 (E)	9,703	7.2	0.784
16 (E)	2,314	6.2	0.800	40 (E)	4,662	6.0	0.710
17 (E)	1,477	6.0	0.822	41 (C)	2,695	5.9	0.747
18 (E)	955	5.6	0.816	42 (E)	9,469	6.9	0.754
19 (C)	2,863	6.4	0.804	44 (E)	8,096	6.9	0.767
20 (C)	3,329	6.6	0.814	45 (C)	3,766	6.2	0.753
21 (E)	6,929	7.1	0.803	46 (C)	2,031	6.3	0.827
22 (E)	6,762	7.4	0.839	48 (E)	8,051	7.4	0.823
23 (C)	969	5.9	0.858	49 (E)	6,631	6.9	0.784
24 (E)	7,831	7.1	0.792	50 (E)	868	5.4	0.798
25 (E)	7,733	6.9	0.771	51 (C)	1,765	6.5	0.869
27 (E)	5,712	5.5	0.636				

3.3. Prokaryotic assemblage composition

From the 5,087,534 high quality reads, a total of 2,222,256 were included in the Minimum Entropy Decomposition pipeline to form 2045 different clusters (OTUs). More than 98% of the sequences obtained were affiliated with Bacteria. Proteobacteria was the most abundant phylum (57.2 - 83.5%) across samples followed by Bacteroidetes and Atribacteria. Yet, the later was exclusively present at station 4 of the Inner Nervion estuary, and represented 15% of the total bacterial community in this sample.

The assemblage composition at the class level was dominated by γ-, δ- and β-Proteobacteria (Figure 6.2). A cluster encompassing the coastal and six estuarine stations, showed a high contribution of γ-Proteobacteria, and Flavobacteria in two stations in particular. Estuarine stations were clustered in two groups dominated by δ- and β-Proteobacteria, respectively, whilst ε-Proteobacteria was dominant only at one estuarine station. At the order level, Desulfobacterales (within class δ-Proteobacteria, Figure S6.1) was the dominant group within estuarine stations, and was significantly less important (Welch's *t*-test; $p < 0.005$) at coastal stations (Figure S6.2). Representatives of Desulfobacteraceae and Desulfobulbaceae families contributed to the dominance of this order (Figure. S6.3). Stations from coastal areas were characterized by a high contribution of the γ-Proteobacteria order Xanthomonadales (7 – 24%), in particular the JTB255 marine benthic group and an unclassified BD7-8 marine group. Representatives of this order showed a significant lower contribution (0.5 – 17%) across estuarine stations.

Burkholderiales was the second most abundant order found in estuarine stations, but presented a high contribution at sites 27, 28 and 30 (> 20%) mostly by the family Comamonadaceae. Within Flavobacteriales, the family Flavobacteriaceae presented a high relative abundance (10.9 – 27.3%) at certain stations (27, 28, 31, 32, 40 and 50) of four estuaries (Deba, Urola, Urumea and Bidasoa) and at station 46 within coastal areas (see Figure 6.1 for location). Yet, different OTUs within this family showed different contributions across the stations. The ε-Proteobacteria family Helicobacteraceae, within Campylobacterales, was found with a significant high relative abundance (41%) at station 34 of the Oria estuary.

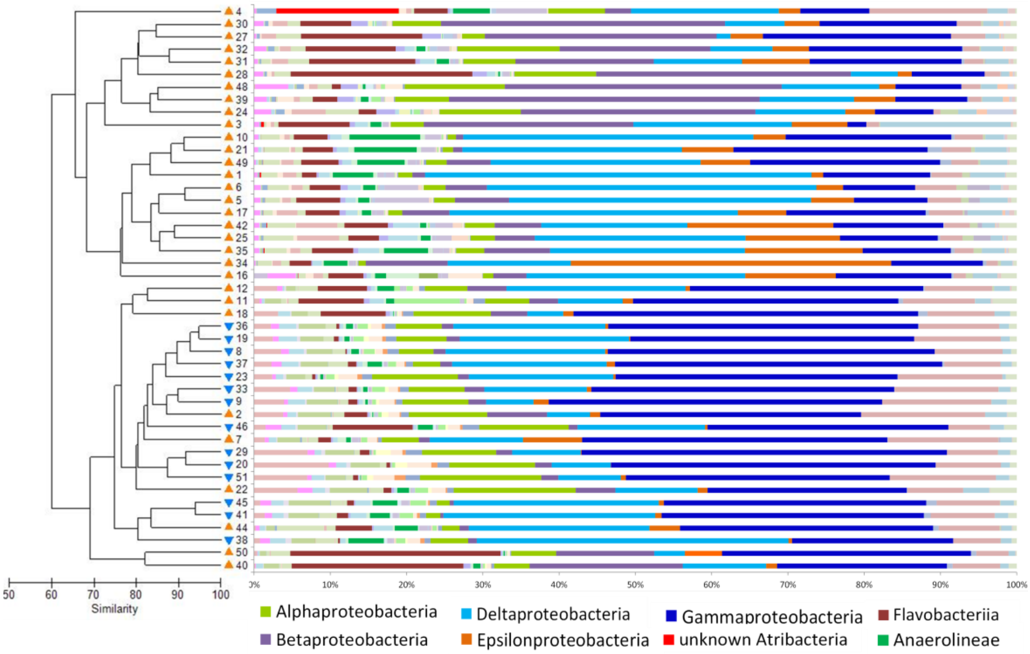

Figure 6.2. Prokaryotic community composition along the Basque Coast. Bray-Curtis dendrogram of 16S rRNA sequences clustered at class level, followed by colored bars depicting the community assemblages for each station. Classes for which the contribution to the overall community composition at least at one sample is > 10% are highlighted. Stations of estuarine (orange) and coastal (blue) areas are accordingly colored.

On the basis of the distance-based linear model, we determined that contaminants and environmental variables that significantly contributed in explaining differences in community composition across all samples were, in decreasing order of influence, salinity, redox potential and PCBs (Table 6.4). Together, these variables explained 35.7% of community variation. The first dbRDA coordinate axis explained 34.3% of the total variation in the community and the second axis 8% (Figure 6.3).

3.4. microgAMBI and ecological status

Putative bacterial taxa at the family level indicators of anthropogenic pressures (Table S6.1) were identified across the study region. Table 6.5 shows the most abundant bacterial taxa found along the study area (contribution of a minimum of 5% to the overall bacterial assemblage in at least one station) that significantly correlated with variables indicators of anthropogenic pressures measured across all stations. In general, impacted sites (high values of Pressure Index) showed a higher contribution of sequences of bacteria here defined as indicators of contamination. For instance, stations across Deba, which exceeded the SQG for organic compounds, and Urola and Oria estuaries, that in addition presented high concentration of organic matter content (> 5%) showed a high total contribution of the Comamonadaceae or Helicobacteraceae families (27.4 and 41.7%, respectively). Taxa significantly correlated with organic matter content (Desulfobacteraceae) and redox potential (Desulfobulbaceae) contributed with higher proportion in the estuarine stations than to the coastal ones. Regarding the family Flavobacteraceae, significant correlation was found with concentration of Chromium.

Table 6.4. DistLM results of relative abundant prokaryotic community data against 9 predictor variables selected for inclusion in the full analysis, *n* = 44 (station 20 was removed due to not availability of environmental data) (9999 permutations).

	Marginal test			Forward selection sequential test			
Predictor variable	Pseudo-F	*P*	Percent variation explained	Pseudo-F	*P*	Percent variation explained	Cumulative variation explained
Salinity	9.0027	**0.0010**	0.1765	9.0027	**0.0010**	0.1765	0.1765
Redox potential	7.8193	**0.0010**	0.1570	7.9959	**0.0010**	0.1344	0.3109
PCBs	3.9334	**0.0020**	0.0856	2.8698	**0.0130**	0.0461	0.3570
OM	7.2591	**0.0010**	0.1474	1.1970	0.2700	0.0191	0.3762
Hg	0.8492	0.4800	0.0198	1.0558	0.3540	0.0169	0.3930
PAHs	1.0725	0.3540	0.0249	0.9487	0.4410	0.0152	0.4082
Cd	2.3170	**0.0370**	0.0523	0.9721	0.4290	0.0156	0.4238
Zn	1.4021	0.1920	0.0323	1.3717	0.1950	0.0217	0.4455
Pb	0.8822	0.5210	0.0206	0.8267	0.5370	0.0132	0.4587

Bold: significantly correlated with prokaryotic community assemblages at $\alpha = 0.05$. PCBs: polychlorinated biphenyls; OM: organic matter (%); PAHs: polycyclic aromatic hydrocarbon

Exploring at the lowest taxonomic level, we observed that the OTUs found across this family, varied along different stations. For stations 28 (from Deba estuary), 40 (Urola) and 3 (Nervion) one single OTU contributed in 5.7, 8.6 and 5%, respectively, to the overall bacterial assemblage of these samples. A different OTU of the same family showed to be dominant (15% overall bacterial assemblage) at station 50, located at the Bidasoa estuary. These OTUs were found at a relative proportion < 1% in the remaining stations.

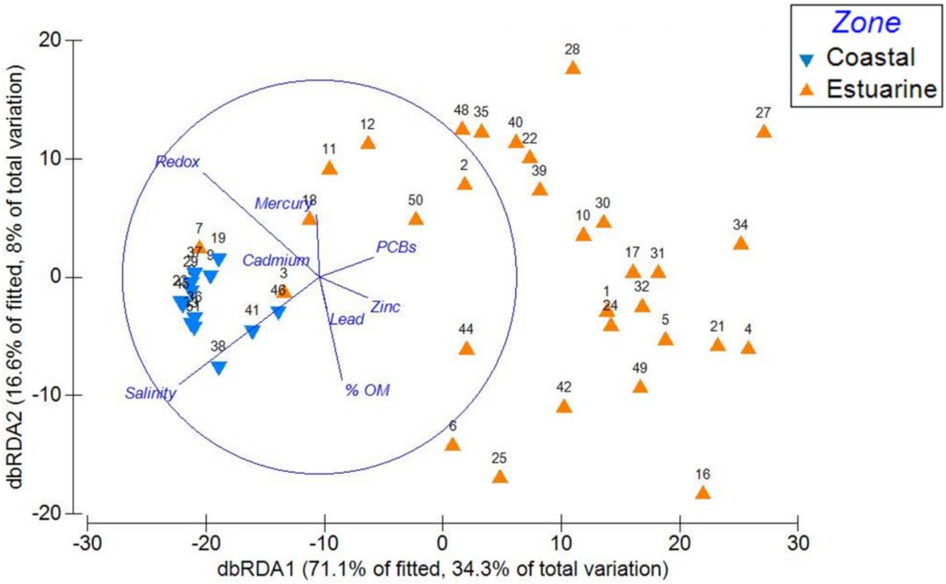

Figure 6.3. dbRDA ordination plot showing the relationships between bacterial assemblage composition and predictor variables determined by forward selection DISTLM. Length of overlaid vectors indicates the relative influence of the fitted predictor variable. OM: organic matter; PCB: polychlorinated biphenyls.

Table 6.5. Summary of significant bacterial taxa at the family level associated with pollution inputs based on forward-selection analysis. OM: organic matter

Taxa	Contribution to the different stations (relative abundance %)	Variable	Std. Error	t value	p value
Desulfobacteraceae	1 (21.6), 4 (8.5), 5 (9.7) 6 (18.8), 16 (7.3), 17 (9.83), 49 (11.5), 21 (10) , 35 (8.8), 41 (8.3)	OM	0.042	3.016	0.004
Desulfobulbaceae	1 (16.2), 3 (11.4), 5 (17.7), 6 (14.3), 10 (10), 12 (7), 17 (13.75), 24 (7.8), 34 (6.3), 35 (7.14), 44 (9), 49 (8)	Redox Potential	0.001	3.667	0.001
Atribacteria	4 (16)	Cadmium	0.046	8.983	< 0.001
Flavobacteriaceae	11 (7.9), 18 (8.4), 27 (15.6), 28 (23.1), 31 (12.4), 32 (11), 40 (22.3), 3 (9.1), 50 (27), 46 (10)	Chromium	0.004	4.215	< 0.001
Helicobacteraceae	16 (11.3), 25 (12.1), 34 (41.7), 42 (18.3),	OM	0.042	3.885	< 0.001
Comamonadaceae	3 (10), 24 (15.5), 27 (23.4), 28 (22), 30 (27.4), 31 (11.4), 32(14), 39 (15.4), 40 (8.1), 48 (14.4),	Chromium	0.006	3.676	0.001
Rhodobacteraceae	32 (6.24)	Chromium	0.002	3.381	0.002

The microgAMBI values ranged from 0.38 to 4.43, so that 12, 9, 15 and 9 stations were classified as sites with a high, good, moderate and poor ecological status, respectively. No stations with bad status were found (Figure 6.4).

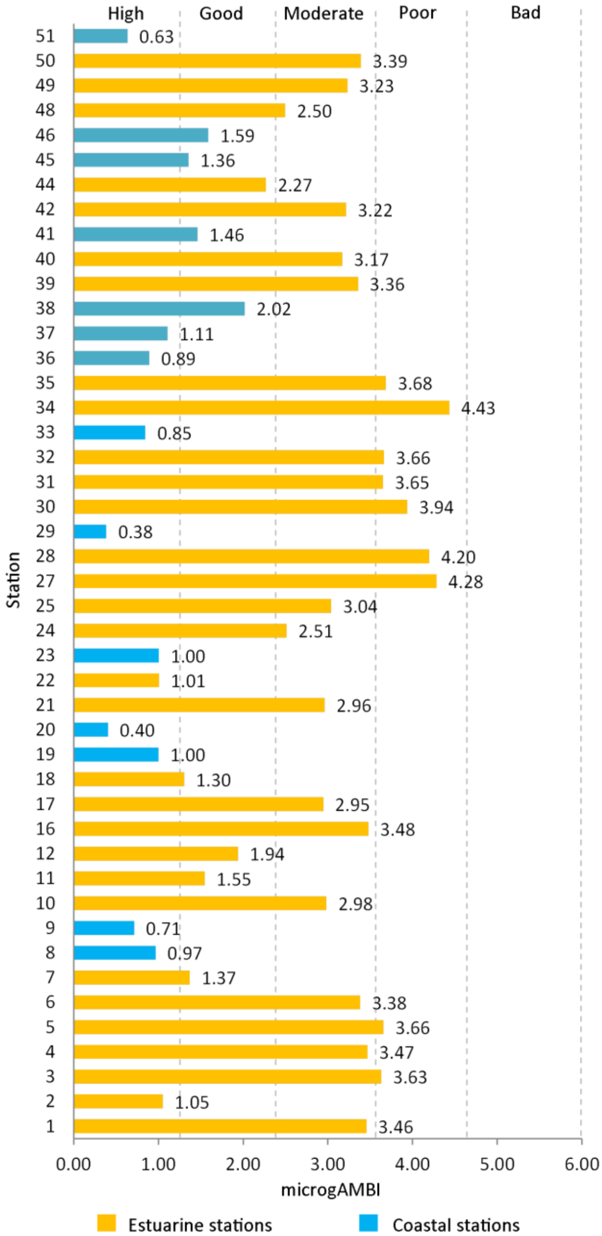

Figure 6.4. microgAMBI values obtained at each station. Thresholds for ecological status assessment are depicted.

Significant differences in taxa proportions among the different ecological status categories identified were observed, which highlighted the contribution of the families Desulfobulbaceae, Comamonadaceae, Flavobacteraceae, Moraxellaceae and Helicobacteraceae in the samples classified as poor or moderate status when considering all samples together (Figure 6.5) and only considering estuarine stations (Figure 6S.4). In addition, based on the SQG criteria, the ecological status obtained for each sample based on bacterial taxa indicators was contrasted. We found a significant correlation between microgAMBI and the pressure index, where low microgAMBI values were associated with the lowest anthropogenic pressures (Figure 6.6) and higher microgAMBI values were obtained with increasing pressure. A high significant positive correlation was found between microgAMBI and AMBI (Figure 6.7). As microgAMBI, in general, AMBI presented higher values in the inner estuarine stations and decreased towards the coastal zone (Figure S6.5), which resulted in 10 and 9 estuarine stations assigned as moderately and heavily disturbed, respectively. Stations that resulted in AMBI values > 5 (Heavily disturbed) were assigned as moderate or poor ecological status using microgAMBI, except for one station of the Barbadun estuary (2), which resulted as high ecological status for microgAMBI (Figure 6.7, Figure S6.5).

4. Discussion

Despite the recognized role of microorganisms in ecosystem functioning (Strong et al., 2015) and their sensitivity to environmental changes (Zhang et al., 2008a; Zhang et al., 2008b; Ager et al., 2010; Chiellini et al., 2013; Aranda et al., 2015), the use of bacterial assemblages as indicator of ecosystem health has been greatly ignored (Nogales et al., 2011; Caruso et al., 2015; Strong et al., 2015; Heiskanen et al., 2016). Here, we have developed a biotic index (microgAMBI) based on the benthic bacterial community analyzed through 16S rRNA gene amplicon sequencing. The purpose of microgAMBI is to summarize information about the relative abundance of putative indicator bacterial taxa in an environmental sample into a single measurement. Our findings show that the analysis of the bacterial community composition can be used to generate an index able to detect gradients of environmental perturbation in marine systems.

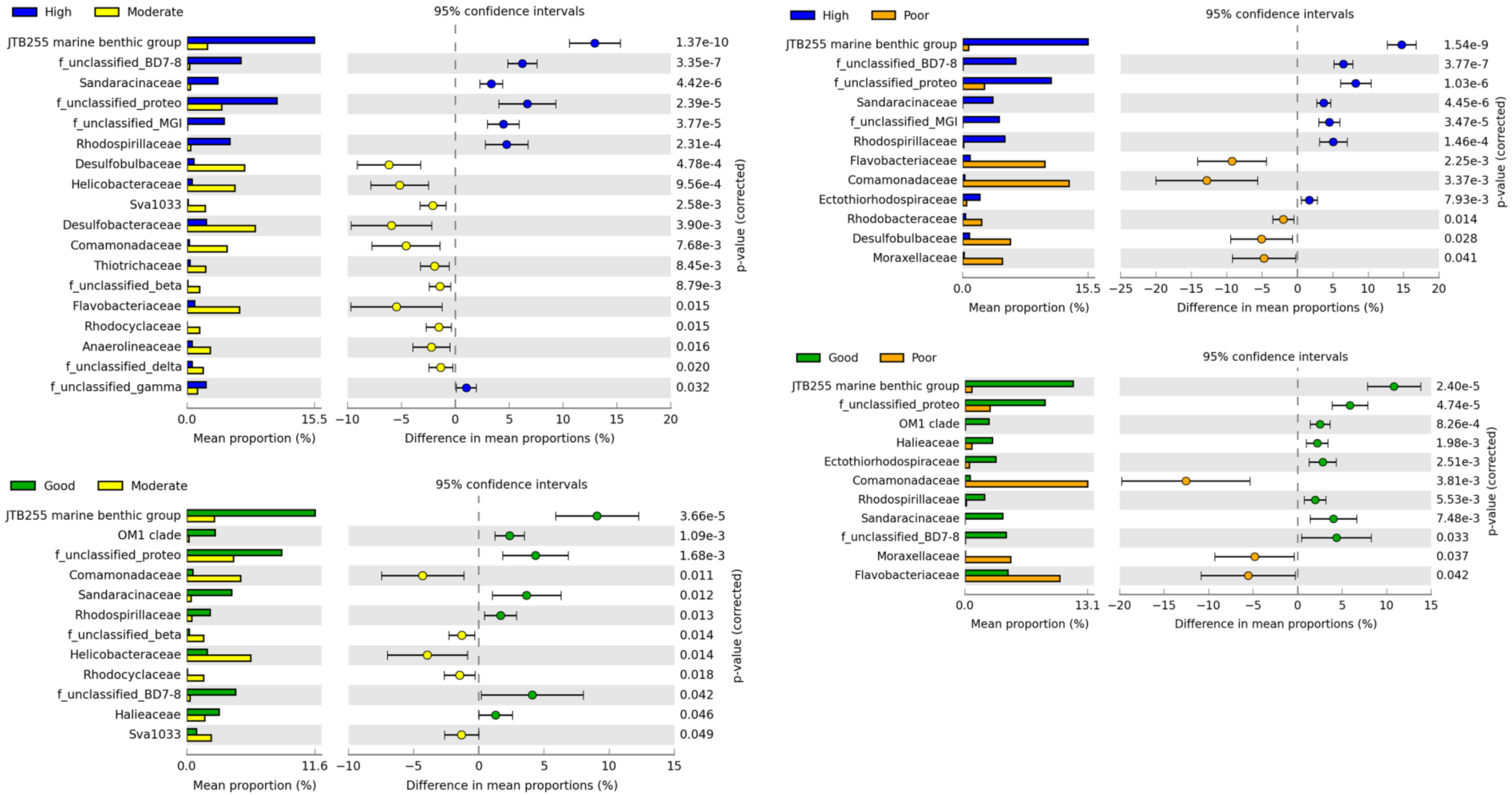

Figure 6.5. Detail of the relative contribution of the different bacterial taxa at the family level contributing to significant differences (Welch's t-test; $p < 0.005$) between the four classes (High, Good, Moderate and Poor) defined from the microgAMBI inferred for each station.

The present study is focused on a program network that monitors the estuarine, coastal and shelf areas (< 200m), which are those more intensively exploited and impacted by land-based activities (Halpern et al., 2008). Due to the natural heterogeneity of these systems is difficult to define links between the observed changes in bacterial community and human-induced perturbations (Nogales et al., 2011). Here we found that the prokaryotic assemblage composition was dependent on salinity, and also responded to increases in the concentration of anthropogenic compounds (PCB and cadmium) and organic matter content.

Among the human perturbations occurring in estuarine and coastal systems, nutrient enrichment can result in considerable changes in composition and function of the microbial component (Crump et al., 2007). The consequences of nutrient enrichment are associated to conditions of hypoxia or anoxia and low redox potential (Nogales et al., 2011), which favor the increase of sulphide-oxidizing bacteria (SOB) and sulphate-reducing bacteria (SRB) (Asami et al., 2005). In the study area we found a relevant fraction of sulphur-oxidizing Gammaprotebacteria and SRB represented by the deltaproteobacteria Desulfobacteraceae and Desulfobulbaceae, which could indicate organic enrichment and oxygen depletion at some stations.

Our results showed a significant correlation between the increase of organic matter content and the relative abundance of Desulfobacteraceae. The presence of this group in eutrophic estuaries and harbor sediments has been previously reported (Zhang et al., 2008b; Sun et al., 2013). Hence, due to their role on a variety of processes regarding organic matter turnover, biodegradation of pollutants, and sulfur and carbon cycles (Zhang et al., 2008b), SOB and SRB have been used as indicator of nutrient inputs to marine sediments (Sun et al., 2013; Aranda et al., 2015). Here we found that the highest contribution of SRB (>10%) occurred at some of the stations located in areas severely affected by different anthropogenic impacts (see Solaun et al., 2013) such as the Barbadun, Nervion, Oka, Oiartzun and Bidasoa estuaries. In particular, the Nervion estuary has been historically affected by organic nutrient inputs from surrounding anthropogenic activities (Cearreta et al., 2000). In this regard, one station of this estuary was characterized by a low redox potential, which could explain the presence of Atribacteria, recently reported in anoxic sediments (Carr et al., 2015).

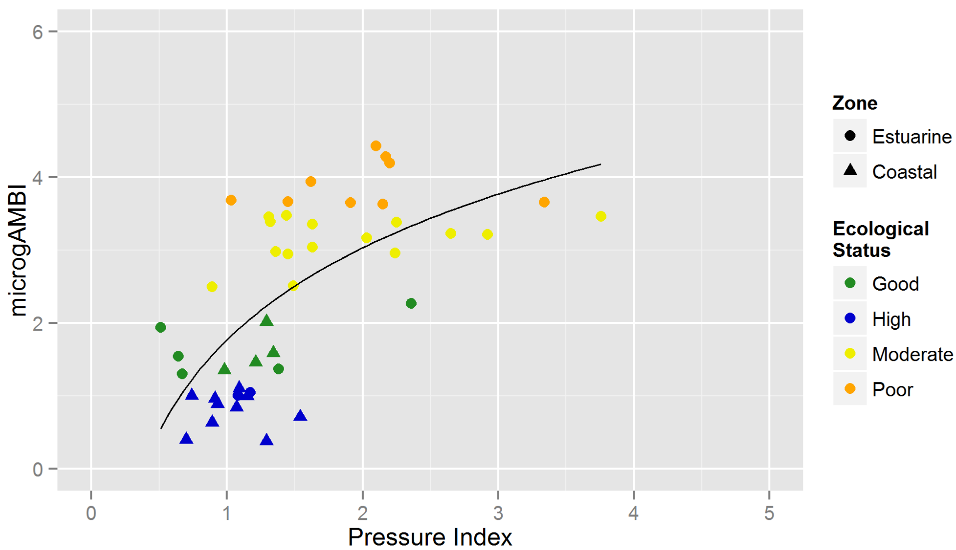

Figure 6.6. Logarithmic regression between the Pressure Index and microgAMBI. Significant correlation was found ($r^2 = 0.43$, $p < 0.01$). Each point shows the relationship between the pressure index and microgAMBI for each station, colored according to the ecological status assigned from the microgAMBI value obtained.

Alterations in bacterial assemblage due to contamination of organic and inorganic pollutants have been reported (Dell'Anno et al., 2003; Sun et al., 2013; Lozada et al., 2014). These components reach the marine environment by direct discharges, runoff from land or river discharges and accumulate in sediments. Here, we detected a high contribution (> 30%) of hydrocarbon degrading bacteria (e.g. *Flavobacterium* sp., *Acinetobacter* sp. and family Commamonadaceae), previously reported from hydrocarbon-polluted marine environments (Lozada et al., 2014), at Deba and Urola estuaries. In particular, one OTU of *Sulfuricurvum* sp. highly contributed (21.1%) to the bacterial assemblage of one station in the Oria estuary. The genus has been reported as a contributor to the oxidation of reduced sulfur compounds in iron seeps (Haaijer et al., 2008) but has been also found in groundwater contaminated with petroleum (Campbell et al., 2006). Historically, Deba, Urola and Oria estuaries have supported industrial activities. Although wastewater treatment plants and improvements in industrial management during the last few years have led to a decrease in the concentration of contaminants (Solaun et al., 2013), still several components exceed the SQG. The presence of a relevant fraction of bacteria here defined as indicator of pollution in these estuaries resulted in high microgAMBI values, ranging from 3.6 to 4.4, resulting in a poor ecological status classification.

The analyses regarding the response of bacterial taxa to different impacts, place Flavobacteriaceae as indicator of an increase in the concentration of chromium. Here, two stations of the outer part of the Urumea and Bidasoa estuaries and one offshore station presented a high contribution of this family (10-20%); interestingly, a high proportion was represented by a single OTU. As microgAMBI places this family in EGIII, the presence of this group in the external stations of the estuaries Urumea and Bidasoa resulted in microgAMBI values that classified these stations as moderate status. This could be related to the actions taken under the sanitation plan through the construction of a water treatment plant and a diversion of discharges via a submarine outfall outside the estuaries (Borja et al., 2009c; Solaun et al., 2013). Flavobacteraceae has shown to increase the relative abundance in samples collected from wastewater treatment systems (Shchegolkova et al., 2016) and marine sites impacted by sedimentation, local sewage and municipal wastewater (Ziegler et al., 2016). However, Flavobacteraceae commonly occurs in marine samples (Giovannoni and Rappé, 2000), so that using this family as indicators might result in some cases as incorrect ecological status assessment.

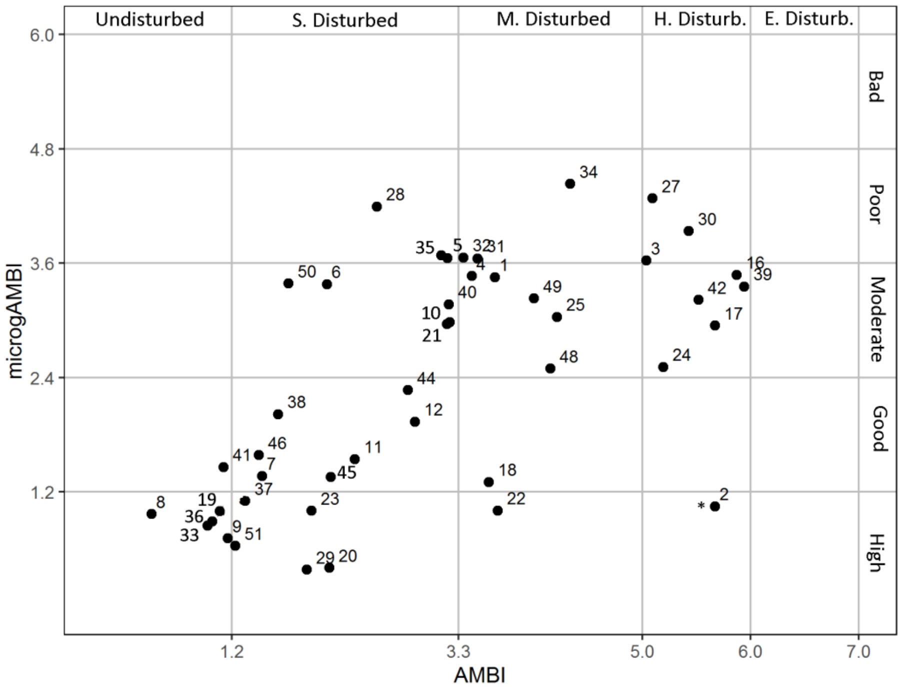

Figure 7. Relationship between AMBI and microgAMBI ($r^2 = 0.46$, $p < 0.01$). For each station, the correlation between AMBI and microgAMBI is represented by each point. Vertical and horizontal lines indicate pollution level assessment thersholds for AMBI and microgAMBI, respectively. S. Disturbed: Slightly disturbed, M. Disturbed: Moderately disturbed, H. Disturb.: Heavily disturbed, E. Disturb.: Extremely disturbed. Station numbers are depicted. * Station 2 was excluded for correlation; correlation considering all stations resulted in $r^2 = 0.36$, $p < 0.01$.

These findings provide useful information in the understanding of the link between the bacterial assemblage composition and the putative source of contamination in polluted marine environments. However, there are still several challenges that using bacterial assemblages as indicators will need to address to become considered in monitoring. Although microgAMBI relies on the relative abundance of bacterial taxa at the family level assigned to one of the two ecological groups, further investigation focusing on the response of particular bacterial taxa, at low levels of taxonomic resolution (genus and OTU level), to certain contaminants is necessary to move beyond these findings.

In order to validate the bacterial biotic index microgAMBI, we developed a pressure index based on organic matter content, redox potential and concentrations of metals, PAHs and PCBs. We found that the microgAMBI was significantly correlated with this sediment quality index and, thus, can provide valuable information to determine the ecological status of certain anthropogenic impacted sites. In addition, microgAMBI provided comparable results to that obtained using AMBI, based on macroinvertebrates. Yet, some discrepancies must be considered between both indices. In particular, one station in the study region resulted in contradictory results when analyzed using each approach. Whilst AMBI assesses Station 2 as heavily disturbed, based upon macroinvertebrates (AMBI = 5.67), the bacteria community resulted in a high ecological status (microgAMBI = 1.05). This station belongs to the outer part of the estuary Barbadun and is affected by strong wave energy. This results in extremely dynamic sand banks, which prevent the settlement of permanent macroinvertebrate communities and normally are dominated by early colonizers. This dominance leads to high AMBI values, even if the area is not polluted, being this disturbance of natural origin (Borja et al., 2013a). In contrast, when calculating microgAMBI, there is a dominance of typical sensitive bacterial marine communities, present in undisturbed locations, being the contribution of bacterial families assigned to EGIII (e.g. Desulfobateraceae and Desulfobulbaceae) very low (< 1 %). Despite these differences, in general, AMBI and microgAMBI provided similar ecological status assessment. However, this contribution does not intend to replace one index by another, but presenting a more comprehensive understanding of an integrated ecosystem approach, in which different ecosystem components must be used to assess the ecological status holistically (Borja et al., 2009c).

5. Conclusions

Overall, our findings indicate that the microgAMBI can provide useful information for the evaluation of anthropogenic impacts occurring in estuarine and coastal environments. In this sense, microgAMBI can be used as a complementary tool to the currently applied biotic indices based on macroinvertebrates for ecological status assessment of marine environments in response to European directives, such as the WFD and the MSFD.

HTS technologies provide the opportunity to monitor marine ecosystems in a robust and cost-effective way. It has been reported that metabarcoding is a valid and accurate technique for environmental biomonitoring, which applied for the calculation of current existing biotic indices based on macroinvertebrates, such as AMBI, can significantly reduce the effort required to obtain final results. Here, we have applied the metabarcoding analysis on bacterial assemblages demonstrating that the results obtained using this approach can be useful for assessing the ecological status of marine environments.

The present index has been developed on samples collected from the Basque coast; however, given the wide array of ecological conditions considered, the methodology can be potentially applied to other coastal areas. Finally, further investigation on the response of specific bacterial taxa to selected pollutants may help to improve the proposed index, as well as to include new potential indicator taxa to be considered for ecological status assessment.

GENERAL DISCUSSION

As a result of the implementation of the WFD and MSFD to promote sustainable use of the seas and conserve marine ecosystems, innovative monitoring methodologies and indicators, allowing improving environmental status assessments for achieving an integrated ecosystem management, are being investigated (Borja et al., 2016b). Among the incipient innovative methodologies (Danovaro et al., 2016), this Thesis focuses on metabarcoding, which is one of the most promising genomic tools for performing faster, cost-effective and more accurate marine biodiversity assessments (Bourlat et al., 2013; Borja et al., 2016b). Regarding new indicators, it focuses on the suitability of bacterial community-based biotic indices for assessing ecological status. In particular, the work presented in this Thesis focuses on assessing seafloor integrity within the MSFD, and through the different Chapters, the main limitations that are preventing the use of metabarcoding for this purpose are addressed and solutions for a future implementation of the technique in current European directives are provided.

1. Requirements on abundance estimation for calculating macroinvertebrate-based biotic indices

Due to biological and technical factors (see Introduction), estimating metazoan species abundance from metabarcoding sequence data is challenging (Porazinska et al., 2010; Thomsen et al., 2012b; Goldberg et al., 2013; Evans et al., 2016). Thus, it has been assumed that DNA-based biodiversity estimates should rely only on presence/absence metrics (Ranasinghe et al., 2012; Yu et al., 2012; Ji et al., 2013; Mächler et al., 2014; Dowle et al., 2015). Thus, we investigated the possibility of developing a biotic index using only presence/absence metabarcoding data and showed that presence/absence(pa)AMBI, provides similar results to AMBI (Chapter 1). Also, we demonstrated that there is good correlation between (pa)gAMBI and (pa)AMBI (Chapter 4). These findings alleviate the need of quantifying species for assessing the marine ecological status using molecular tools. In this context, being able to provide presence/absence species estimations, using incidence read data (Chariton et al., 2015), metabarcoding might represent a great opportunity for developing a genomic version of AMBI using only incidence taxa information (i.e. (pa)gAMBI). But it is well known that biodiversity metrics are more accurate when using abundance of the target communities (Yu et al., 2012); in particular, the most widely used versions of AMBI rely on the abundance (AMBI) or biomass (BAMBI) of the taxa detected (Muxika et al., 2012), and it is known that, although capable of discriminating disturbance classes, (pa)AMBI is not as accurate as their abundance or biomass-based versions (Warwick et al., 2010). Thus, obtaining a reliably read count-based gAMBI would be more appropriate for ecological status assessments than (pa)gAMBI. We found very good correlation between gAMBI and AMBI and gAMBI and BAMBI, derived from a good correlation between the number of metabarcoding reads and number of individuals per specie and between the number of metabarcoding reads and biomass of each species (Chapter 5). We also showed a better performance of gAMBI vs. (pa)gAMBI in assessing ecological status of the sites under study. This fact represents the first insight for the use of abundance metabarcoding data for ecological status assessment purposes beyond presence-absence metrics. This finding represents an advance for the implementation of metabarcoding in environmental management. However, efforts in this direction must be performed to better establish the relationship between species abundance or biomass and read counts in order to further implementing metabarcoding for metazoan taxa abundance-based estimations in monitoring.

2. Marker and primer selection for accurately characterizing biological communities

Selecting the most suitable barcode for accurately discriminating species (Hebert et al., 2003a) is required for a successful metabarcoding study (Wangensteen and Turon, 2016). In the case of animals, the mitochondrial cytochrome c oxidase subunit I gene (COI) and the nuclear 18S small subunit rRNA (18S rRNA) are the most commonly used markers (Leray and Knowlton, 2015; Zaiko et al., 2015; Guardiola et al., 2016; Pearman et al., 2016a). The suitability of COI for metabarcoding purposes has been criticized due the high variability in the primer binding sites (Deagle et al., 2014); however, the higher capability of this marker to distinguish closely related metazoan species compared to 18S rRNA (Hebert et al., 2003b;

Bucklin et al., 2011) has resulted in COI being the preferred marker for metazoan metabarcoding studies (Carew et al., 2013; Leray et al., 2013; Gibson et al., 2015; Leray and Knowlton, 2015; Zaiko et al., 2015). Furthermore, COI is represented by hundreds of thousands of curated sequences in the Barcode of Life Database (http://www.boldsystems.org/).

Suitable primers that amplify the selected marker of the largest number of target species are necessary. Among the available primer sets for amplifying COI barcodes, those that target a 658 bp fragment (Hebert et al., 2003a), and a more recently proposed shorter fragment of 313 bp (Leray et al., 2013) showed to be the most suitable for amplifying a wide range of AMBI species in a *in silico* analysis (Chapter 1). Additionally, analyses based on samples composed by a known variety of macroinvertebrate taxa belonging to different phyla showed that the primer pair targeting the short COI region was able to amplify a higher percentage of the species present in the sample (Chapter 4), which is also the one that entails easier, more straight forward and less error probe bioinformatic analyses (Chapter 3).

It should be noted that the selection of a good primer set should be accompanied by the use of suitable PCR conditions (Sipos et al., 2007). The comparison of different PCR annealing profiles on the amplification success revealed that using inappropriate conditions substantially affects the biological community characterization, which is translated into erroneous ecological status assessments (Chapter 4). The good performance of a low and constant annealing temperature during the PCR contrasted with previous recommended procedures, such as those proving, through barcoding, that a touchdown profile should be applied for amplifying macroinvertebrates (Leray et al., 2013).

In sum, this Thesis permitted determining both the primer pair and PCR conditions that more reliably assess the macroinvertebrate community. These findings should be strongly considered in future applications of metabarcoding for characterizing macroinvertebrates with ecological assessment purposes as they allow performing accurate biomonitoring and, if applied to different areas, results being comparable, thus, improving an integrated ecosystem-based management.

3. Reference database for accurate taxonomic assignment of unknown sequences

The success of metabarcoding for species identification relies on the number of target species with publicly available sequences (Cristescu, 2014). In this respect, only few species in the AMBI list presented sequences in public databases at the beginning of this work, being not enough for providing accurate inferences of gAMBI (Chapter 1). Increasing the reference database requires following rigorous species identification protocols so that the sequences are tied to a specimen that has been formally identified (Bucklin et al., 2011). This is a slow process which, together with the economical investment required for barcoding each species (i.e. each individual in a sample can be identified by DNA barcoding per about $5 (Cameron et al., 2006)), makes the completion of such database for all species included in the AMBI list in a relatively short period of time challenging.

According to *in silico* analyses (Chapter 1), accurate (pa)gAMBI values can be derived from a reference database that contains only 10% of the target species, if they are among the most frequently occurring. Yet, due to failures in amplification of some taxa in real samples (Chapters 4 and 5), this number could be higher. Nonetheless, this finding has a notable implication for a quick implementation of metabarcoding for ecological status assessment since with only few more barcoded species, gAMBI could provide comparable status results to those obtained using AMBI. This Thesis has contributed with new barcodes of 129 macroinvertebrate species, which are available in BOLD (http://www.boldsystems.org) at "BCAS project" and in GenBank using accession numbers KT307619–KT307707 and KF808157 - KF808178. Together with other studies barcoding marine and freshwater macroinvertebrate species (e.g. Carew et al., 2013; Laforest et al., 2013; Dell'Anno et al., 2015; Leray and Knowlton, 2015; Vivien et al., 2015; Miralles et al., 2016; Shackleton and Rees, 2016), this data will shorten time when DNA sequences of a large number of the AMBI species list will be available in public databases.

It is well known that the more complete the database, the more reliable is the biological community characterization (Wangensteen and Turon, 2016); yet, the taxonomic level at which sequences are identified in reference databases is also critical. This is particularly important for the assignment of ecological groups for the calculation biotic indices, which in most cases needs to be done at the species level (Chapter 1). However, a high number of sequences deposited in the public databases do not contain the taxonomic classification at this level and remain as genus, family or even phyla, so that many unknown sequences are taxonomically assigned to one of these levels (Chapter 4) and therefore, ecological groups cannot be assigned. In this context, the improvement of the reference database and thus the ability to assign sequences to known species will enable metabarcoding to be more reliably used in monitoring surveys.

4. Effect of metabarcoding biases in marine monitoring

Biases present in the different steps of metabarcoding, such as environmental samples manipulation (Creer et al., 2016), inefficient DNA extraction for some taxa (Deiner et al., 2015), or uneven amplification by primer pairs (Deagle et al., 2014) or under certain PCR conditions can lead to greater recovery of sequences of some species and the non-detection of others (Elbrecht and Leese, 2015; Piñol et al., 2015). This limitation must be taken into account as results might provide inaccurate ecological status assessments, especially when attempting to calculate biotic indices, where the inference of the ecological status according to the ecological groups assigned to species may be affected by the approach used for taxonomic identification (Maurer et al., 1999; Borja et al., 2000; Simboura and Zenetos, 2002). Our results revealed that extracting DNA from bulk samples (composed of mixed individuals but without sediment) using appropriate DNA extraction kits performed equally well than extracting DNA individually (Chapter 4); yet analyses of real samples collected directly from the sediment and homogenized as described in Chapter 2 suggest that sampling processing approach affects inferred taxonomic composition.

Both, effect of sample processing and primer biases were reflected using samples of known and unknown taxonomic composition in Chapters 4 and 5, respectively, by the fact that some species, despite being present in the reference database, were not detected using metabarcoding. These discrepancies between both assessments have been shown in other studies of similar characteristics (Dowle et al., 2015; Elbrecht and Leese, 2015; Gibson et al., 2015; Lejzerowicz et al., 2015) and must be taken into consideration when interpreting biomonitoring results. In this context, the different taxonomic compositions obtained between morphology and metabarcoding could imply that characterizing the macroinvertebrate community using metabarcoding provides contrasting management conclusions compared to that obtained using morphology. Fortunately, these discrepancies did no negatively affected for a successful assessment of ecological status, since morphology and metabarcoding-based analysis generated comparable AMBI and gAMBI values. This result positions gAMBI as a suitable alternative for assessing ecological status and will contribute to accelerating the implementation of metabarcoding in current European directives.

5. Importance of standardizing procedures for ensuring reproducible and comparable results

As a response to the potential showed by metabarcoding for performing accurate and cost-efficient biodiversity assessments (Yu et al., 2012; Ji et al., 2013), characterizing metazoan communities from aquatic environments using this approach is gaining importance during the last few years (e.g. Guardiola et al., 2015; Leray and Knowlton, 2015; Pearman and Irigoien, 2015), and the interest for its application in ecological status assessments has increased notably (Baird and Hajibabaei, 2012; Dafforn et al., 2014; Chariton et al., 2015; Lejzerowicz et al., 2015; Laroche et al., 2016). The absence of metabarcoding protocols for characterizing macroinvertebrates is one of the limitations that prevents the potential inclusion of the technique in routine monitoring programs within current European directives. By producing standardized procedures for macroinvertebrate sample processing (Chapter 2) and for bioinformatic analyses of sequence data (Chapter 3), we have generated essential information for ensuring reliability and reproducibility of these practices which represents a step forward the implementation of metabarcoding in routine monitoring. As a response to the necessity of developing techniques that allow performing comparable results for improving an integrated ecosystems approach, this Thesis provides all steps necessary for performing reproducible and comparable metabarcoding-based monitoring practices across sites.

6. Extracellular DNA for biodiversity assessments

The sample processing and DNA extraction protocols suggested in this Thesis, including decantation, sieving and some morphological sorting (Chapter 2), are a step forward the application of metabarcoding for macroinvertebrate biodiversity assessments, but are still tedious and time consuming. Some research acknowledge that small sediment samples, despite not containing whole specimens, may contain DNA (included in feces, cell debris, scales,…) that is representative of the species inhabiting in the area (Lejzerowicz et al., 2015); various authors have relied on the species signal preserved in this so called extracellular DNA (Guardiola et al., 2015; Pearman et al., 2016b) for assessing biodiversity of small metazoans, thus, reducing the sample manipulation effort. By targeting macroinvertebrate extracellular DNA through the application of specific protocols developed for this aim (Taberlet et al., 2012b), we showed that only a small proportion of the taxa identified using morphological methods were retrieved (Chapter 4). Furthermore, the characterized macroinvertebrate community using this DNA source did not suffice for accurate gAMBI calculation, suggesting that the suitability of extracellular DNA-based analyses for macroinvertebrates taxonomic assessments and environmental management purposes is still dependent of further research.

7. Improving biomonitoring cost- efficiency

Probably, the greatest benefit of metabarcoding for environmental monitoring is the adequacy of the technique for efficient monitoring (Ji et al., 2013; Biggs et al., 2015; Sigsgaard et al., 2015; Smart et al., 2016). Even considering sample processing and data analysis time issues, and cost in producing HTS data, metabarcoding is more cost-effective than morphology for biotic indices calculation (Chapter 5). Using the Basque monitoring network as a case of study, we have confirmed the potential of metabarcoding for reducing monitoring costs and time compared to traditional biodiversity assessment based on morphological identification of the species. Therefore, this Thesis confidently confirms, despite other advantages such as being independent on taxonomic expertise and allowing identification of damaged specimens and early developmental stages, that metabarcoding is also less expensive, which makes it suitable for large monitoring programs involving several samples and reduces the time required from sample collection to calculation of the biotic index, which makes it suitable to respond in a timely manner within the different directives.

8. Achievement of an integrated ecosystem-based management through the inclusion of new indicators

The gap regarding the use of microbial indicators within the MSFD (Caruso et al., 2015) is impeding comprehensively assessing the environmental status and therefore limiting the achievement of an integrated ecosystem-based management (Heiskanen et al., 2016). In order to contribute in the improvement of the integrated approach performance, we have performed the first attempt to include bacterial assemblages as indicators within MSFD (Chapter 6) and revealed that the newly developed bacterial derived biotic index (microgAMBI) is significantly correlated with a sediment quality index calculated on the basis of organic and inorganic compound concentrations. In this context, microgAMBI was able to discriminate disturbance situations across the analyzed samples.

The microgAMBI can provide additional information for the evaluation of anthropogenic impacts occurring in estuarine and coastal environments. In this sense, microgAMBI can be used as a complementary tool to the currently applied biotic indices based on macroinvertebrates in response to European directives as it represents an early warning signal to assess impacts (e.g. in aquaculture and other human activities). Yet, as this index has been developed using samples from the Basque coast and uniquely validated in stations from this geographic area, further research is needed in order to evaluate the capability of microgAMBI in providing ecological status assessments in other coastal areas. Furthermore, the inclusion of additional bacterial taxa indicator of pollution might help to improving the effectiveness of microgAMBI in terms of assessing ecological status over a wide variety of locations.

FURTHER RECOMMENDATIONS

In the present Thesis we show the potential of metabarcoding for improving comprehensive evaluation of the marine ecological status in a cost-efficient manner. Despite some limitations of the technique that have been highlighted in the different Chapters, metabarcoding-based biomonitoring can be reliably used as a complementary tool to, and with time, replace currently established methodologies for ecological status assessment of marine environments in response to European directives such as the WFD and the MSFD. Thus, we provide some suggestions for implementing and making a productive use of this technique in large-scale biomonitoring programs:

a) Due to its advantages in terms of cost-effectiveness and independence of taxonomic expertise, we favor the implementation of gAMBI for routine ecological assessment. For that aim, we recommend applying the same standardized protocols throughout the sites evaluated so that results can be comparable across laboratories, and time and cost can be reduced by combining hundreds of samples in the same analysis batch. In order to further continue validating the technique, we also recommend analyzing few sporadic samples with morphology-based taxonomic identification and compare biomonitoring results.

b) As gAMBI is comparable to AMBI, it is anticipated that this new index is also able to detect changes in the marine environment. Yet, intercalibration with other indices currently in use (e.g. BENTIX or BQI) and inclusion as part of a multivariate AMBI (gM-AMBI) is desirable to further validate its usefulness within directives such as the WFD or the MSFD.

c) The first bacterial community-based benthic biotic index developed to date, microgAMBI, may contribute to a more comprehensive ecological status assessment as the information it provides is complementary to other biotic indices. Thus, we recommend: (i) adding new putative pollution indicator taxa to the current list established here; (ii) testing performance of microgAMBI in different geographic areas and under different human activities and pressures, using gradients of degradation; and (iv) investigating its dependence on seasonality.

Despite the evidenced potential of metabarcoding for accurate monitoring, this Thesis has highlighted the lack of application of scientific knowledge at the decision making stage. The recommendations made in this Thesis are aimed to move forward this direction; however, more efforts are required to remove the gap between the scientific literature and management applications.

Regarding other potential applications of metabarcoding for ecological status assessments, the results of this Thesis can be extrapolated to other biological communities. In this context, the technique can be easily integrated by adapting some components of the approach to the target community such as barcode, primer set and reference database. For example, following the strategies for the development of gAMBI or microgAMBI, we recommend further investigation on microbial eukaryotic communities with potential to respond to stress for the development of new complementary biotic indices. This will greatly contribute for a better assessment of the ecological status and will allow achieving an integrated ecosystem management.

For future monitoring, more technical development is necessary. Yet, metabarcoding presents a wide variety of advantages that make it suitable to replace present biomonitoring methods over time. Detecting invasive species, accurately identifying species from stomach contents, developing future genomic-based biotic indices, identifying toxic species or developing protocols for reliably detecting big size organisms using extracellular DNA are some of the directions that must be taken for further applying metabarcoding in ecological status assessments.

CONCLUSIONS AND THESIS

Taking into account the objectives of this Thesis we can conclude that:

1. DNA metabarcoding has been validated as a viable tool for marine macroinvertebrate species identification. Furthermore, a genomic-based biotic index (gAMBI) has been developed. The minimum number of barcoded species, necessary to calculate gAMBI, has been established in 10% of the most frequently occurring species from the current list; however, the accuracy of gAMBI will increase by increasing the number of COI sequences in public reference databases. Yet, biases associated to the different metabarcoding steps might imply that not all species present in the database are detected. We have identified two primer pairs as the best primer sets to retrieve the most complete representation of macroinvertebrate diversity. We have increased in 129 the number of species in the AMBI list with publicly available barcodes. The gAMBI can serve as an alternative to the current biotic indices based on morphology as it can provide faster, more accurate and cost-efficient monitoring results.

2. Standardized protocols for macroinvertebrate sample processing and for obtaining the taxonomic composition of a benthic macroinvertebrate sample from sequence data using a curated pipeline have been developed. In addition, guidelines for the use of sequence data derived taxonomic information to calculate gAMBI have been provided. The contribution of such detailed procedures represents a step forward towards the implementation of metabarcoding in routine monitoring as they provide all necessary steps for performing reproducible and comparable metabarcoding-based assessments across sites.

3. The comprehensive benchmarking study has permitted establishing the conditions in which metabarcoding should be performed for accurately assessing benthic macroinvertebrate diversity. We have demonstrated that using the appropriate DNA extraction strategy, primer set, barcode and PCR conditions, metabarcoding presents great potential to provide accurate biotic indices. These findings will contribute to accelerating the implementation of metabarcoding for ecological status assessment.

4. The application of the developed protocols for sample processing and data analysis on different benthic samples allowed us obtaining successful relationships between gAMBI and AMBI, which confirms the potential of metabarcoding to accurately infer the marine ecological status. However, the discrepancies observed between the macroinvertebrate taxonomic compositions inferred using each methodology reveals that the technique is still unable to provide exact biodiversity characterization and this must be taken into consideration when interpreting biomonitoring results.

5. The cost-efficiency of metabarcoding compared to traditional methodologies relying on morphological analysis will allow performing a better environmental management. Metabarcoding technique proved to reduce the dependency on taxonomic expertise and decreased by three and two times, respectively, the time and costs required from sample collection until calculation of the biotic index when using metabarcoding compared to morphology. Therefore, metabarcoding will contribute in a significant manner to improving large monitoring programs.

6. We were successful in characterizing the bacterial community using the metabarcoding technique. Using published data on bacterial responses to pollution, we were able to develop for the first time a new index based on the same concept as AMBI (named microgAMBI). This index, tested and validated using a gradient of pressure, will provide useful information for the evaluation of anthropogenic impacts occurring in estuarine and coastal environments. Thus, this biotic index can be used as a complementary tool to the currently applied biotic indices based on macroinvertebrates for ecological status assessment of marine environments in response to European directives, such as the WFD and the MSFD.

Thesis

Hence, taking into account the achievement of these objectives, we consider that our hypothesis has been proven and the resulting Thesis is:

"Metabarcoding-based biomonitoring represents a rapid and cost-efficient approach to assess the marine ecological status by characterizing biological communities (i.e. macroinvertebrate and bacterial communities) used as indicators of ecosystem health, in a much faster an accurate way than current monitoring methodologies applied in environmental management, and might improve the performance of integrative assessments of marine waters under an ecosystem approach-based management"

References

Abad, D., Albaina, A., Aguirre, M., Laza-Martínez, A., Uriarte, I., Iriarte, A., et al. (2016). Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. Marine Biology 163(149). doi: 10.1007/s00227-016-2920-0.

Ager, D., Evans, S., Li, H., Lilley, A.K., and van der Gast, C.J. (2010). Anthropogenic disturbance affects the structure of bacterial communities. Environ Microbiol 12(3), 670-678. doi: 10.1111/j.1462-2920.2009.02107.x.

Albaina, A., Aguirre, M., Abad, D., Santos, M., and Estonba, A. (2016). 18S rRNA V9 metabarcoding for diet characterization: a critical evaluation with two sympatric zooplanktivorous fish species. Ecol Evol 6(6), 1809–1824 doi: 10.1002/ece3.1986.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

Amend, A.S., Seifert, K.A., and Bruns, T.D. (2010). Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Mol Ecol 19(24), 5555-5565. doi: 10.1111/j.1365-294X.2010.04898.x.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aranda, C.P., Valenzuela, C., Matamala, Y., Godoy, F.A., and Aranda, N. (2015). Sulphur-cycling bacteria and ciliated protozoans in a Beggiatoaceae mat covering organically enriched sediments beneath a salmon farm in a southern Chilean fjord. Mar Pollut Bull 100(1), 270-278. doi: 10.1016/j.marpolbul.2015.08.040.

Ardura, A., Morote, E., Kochzius, M., and Garcia-Vazquez, E. (2016). Diversity of planktonic fish larvae along a latitudinal gradient in the Eastern Atlantic Ocean estimated through DNA barcodes. PeerJ 4, e2438. doi: 10.7717/peerj.2438.

Asami, H., Aida, M., and Watanabe, K. (2005). Accelerated sulfur cycle in coastal marine sediment beneath areas of intensive shellfish aquaculture. Appl Environ Microbiol 71(6), 2925-2933. doi: 10.1128/AEM.71.6.2925-2933.2005.

Aylagas, E., Borja, A., Irigoien, X., and Rodríguez-Ezpeleta, N. (2016a). Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. Front Mar Sci 3(96). doi: 10.3389/fmars.2016.00096

Aylagas, E., Borja, A., and Rodríguez-Ezpeleta, N. (2014). Environmental status assessment using DNA metabarcoding: towards a genetics based Marine Biotic Index (gAMBI). PLoS ONE 9(3), e90529. doi: 10.1371/journal.pone.0090529.

Aylagas, E., Mendibil, I., Borja, Á., and Rodríguez-Ezpeleta, N. (2016b). Marine Sediment Sample Pre-processing for Macroinvertebrates Metabarcoding: Mechanical Enrichment and Homogenization. Frontiers in Marine Science 3. doi: 10.3389/fmars.2016.00203.

Aylagas, E., and Rodríguez-Ezpeleta, N. (2016). "Analysis of Illumina MiSeq amplicon reads: application to benthic indices for environmental monitoring,". in Marine Genomics Methods and Protocols, Methods in Molecular Biology, ed. Bourlat, S.J. (Springer, New York), vol. 1452. doi: 10.1007/978-1-4939-3774-5.

Azam, F., and Malfatti, F. (2007). Microbial structuring of marine ecosystems. Nature Reviews Microbiology 5, 782-791. doi: 10.1038/nrmicro1747.

Bacher, S. (2012). Still not enough taxonomists: reply to Joppa et al. Trends Ecol Evol 27, 65-66. doi: 10.1016/j.tree.2011.11.003.

Baird, D.J., and Hajibabaei, M. (2012). Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Molecular Ecology 21, 2039-2044.

Barberan, A., Casamayor, E.O., and Fierer, N. (2014). The microbial contribution to macroecology. Front Microbiol 5, 203. doi: 10.3389/fmicb.2014.00203.

Bartolomé, L., Cortazar, E., Raposo, J.C., Usobiaga, A., Zuloaga, A., and Etxebarria, N. (2005). Simultaneous microwave-assisted extraction of polycyclic aromatic hydrocarbons, polychlorinated biphenyls, phthalate esters and nonylphenols in sediments. J Chromatogr 1068, 229–236. doi: 10.1016/j.chroma.2005.02.003.

Bartram, A.K., Lynch, M.D., Stearns, J.C., Moreno-Hagelsieb, G., and Neufeld, J.D. (2011). Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. Appl Environ Microbiol 77(11), 3846-3852. doi: 10.1128/AEM.02772-10.

Berg, T., Furhaupter, K., Teixeira, H., Uusitalo, L., and Zampoukas, N. (2015). The Marine Strategy Framework Directive and the ecosystem-based approach - pitfalls and solutions. Mar Pollut Bull 96(1-2), 18-28. doi: 10.1016/j.marpolbul.2015.04.050.

Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Dejean, T., and Griffiths, R.A. (2015). Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (Triturus cristatus). Biological Conservation 183, 19-28. doi: dx.doi.org/10.1016/j.biocon.2014.11.029.

Bik, H.M., Porazinska, D.L., Creer, S., Caporaso, J.G., Knight, R., and Thomas, W.K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. Trends Ecol Evol 27(4), 233-243. doi: 10.1016/j.tree.2011.11.010.

Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., et al. (2012). Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. Ecol Indic 18, 31-41. doi: 10.1016/j.ecolind.2011.10.009.

Blondeau-Patissier, D., Tilstone, G.H., Martinez-Vicente, V., and Moore, G.F. (2004). Comparison of bio-physical marine products from SeaWiFS, MODIS and a bio-optical model with in situ measurements from Northern European waters. Journal of Optics A: Pure and Applied Optics 6(9), 875-889.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics 30, 2114-2120.

Borja, A., and Dauer, D.M. (2008). Assessing the environmental quality status in estuarine and coastal systems: Comparing methodologies and indices. Ecological Indicators 8, 6.

Borja, A., and Elliott, M. (2013). Marine monitoring during an economic crisis: The cure is worse than the disease. Marine Pollution Bulletin 68, 1-3.

Borja, A., Franco, J., and Perez, V. (2000). A Marine Biotic Index to Establish the Ecological Quality of Soft-Bottom Benthos Within European Estuarine and Coastal Environments. Mar Pollut Bull 40, 12. doi: 10.1016/S0025-326X(00)00061-8.

Borja, A., Josefson, A.B., Miles, A., Muxika, I., Olsgard, F., Phillips, G., et al. (2007). An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European Water Framework Directive. Marine Pollution Bulletin 55, 42-52. doi: 10.1016/j.marpolbul.2006.08.018.

Borja, A., Bricker, S.B., Dauer, D.M., Demetriades, N.T., Ferreira, J.G., Forbes, A.T., et al. (2008). Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. Mar Pollut Bull 56(9), 1519-1537. doi: 10.1016/j.marpolbul.2008.07.005.

Borja, A., Bald, J., Franco, J., Larreta, J., Muxika, I., Revilla, M., et al. (2009a). Using multiple ecosystem components, in assessing ecological status in Spanish (Basque Country) Atlantic marine waters. Marine Pollution Bulletin 59, 54-64.

Borja, A., Miles, A., Occhipinti-Ambrogi, A., and Berg, T. (2009b). Current status of macroinvertebrate methods used for assessing the quality of European marine waters: implementing the Water Framework Directive. Hydrobiologia 633, 181-196.

Borja, A., Muxika, I., and Rodríguez, J.G. (2009c). Paradigmatic responses of marine benthic communities to different anthropogenic pressures, using M-AMBI, within the European Water Framework Directive. Marine Ecology 30(2), 214-227. doi: 10.1111/j.1439-0485.2008.00272.x.

09.Borja, A., Elliott, M., Carstensen, J., Heiskanen, A.S., and Van de Bund, W. (2010). Marine management - Towards an integrated implementation of the European Marine Strategy Framework and the Water Framework Directives. Marine Pollution Bulletin 60(2175-2186).

Borja, Á., Galparsoro, I., Irigoien, X., Iriondo, A., Menchaca, I., Muxika, I., et al. (2011). Implementation of the European Marine Strategy Framework Directive: A methodological approach for the assessment of environmental status, from the Basque Country (Bay of Biscay). Marine Pollution Bulletin 62, 889-904.

Borja, A., Muxika, I., and Mader, J. (2012). Instructions for the use of the AMBI index software (Version 5.0). Revista Investigacion Marina 19(3), 11.

Borja, A., Bald, J., Belzunce, M.J., Franco, J., Garmendia, J.M., Larreta, J., et al. (2013a). Red de seguimiento del estado ecológico de las aguas de transición y costeras de la Comunidad Autónoma del País Vasco. Informe de AZTI-Tecnalia para la Agencia Vasca del Agua. 657.

Borja, A., Elliott, M., Andersen, J.H., Cardoso, A.C., Carstensen, J., and Ferreira, J.G. (2013b). Good Environemntal Status of marine ecosystems: what is it and how do we know when we have attained it? Marine Pollution Bulletin 76(16-27). doi: 10.1016/j.marpolbul.2013.08.042.

Borja, A., Elliott, M., Andersen, J.H., Cardoso, A.C., Carstensen, J., Ferreira, J.G., et al. (2015a). Report on potential Definition of Good Environmental Status Deliverable 6.2. DEVOTES project. 62pp.

Borja, A., Marin, S.L., Muxika, I., Pino, L., and Rodriguez, J.G. (2015b). Is there a possibility of ranking benthic quality assessment indices to select the most responsive to different human pressures? Mar Pollut Bull 97(1-2), 85-94. doi: 10.1016/j.marpolbul.2015.06.030.

Borja, A., Chust, G., Rodriguez, J.G., Bald, J., Belzunce-Segarra, M.J., Franco, J., et al. (2016a). 'The past is the future of the present': Learning from long-time series of marine monitoring. Sci Total Environ 566-567, 698-711. doi: 10.1016/j.scitotenv.2016.05.111.

Borja, A., Elliott, M., Andersen, J.H., Berg, T., Carstensen, J., Halpern, B.S., et al. (2016b). Overview of Integrative Assessment of Marine Systems: The Ecosystem Approach in Practice. Frontiers in Marine Science 3. doi: 10.3389/fmars.2016.00020.

Bourlat, S.J., Borja, A., Gilbert, J., Taylor, M.I., Davies, N., Weisberg, S.B., et al. (2013). Genomics in marine monitoring: new opportunities for assessing marine health status. Mar Pollut Bull 74(1), 19-31. doi: 10.1016/j.marpolbul.2013.05.042.

Bourlat, S.J., Haenel, Q., Finnman, J., and Leray, M. (2016). "Preparation of amplicon libraries for metabarcoding of marine eukaryotes using Illumina MiSeq: the dual-PCR method". in Marine Genomics Methods and Protocols, Methods in Molecular Biology, ed. Bourlat, S.J. (Springer, New York), vol. 1452. doi: 10.1007/978-1-4939-3774-5.

Brandon-Mong, G.J., Gan, H.M., Sing, K.W., Lee, P.S., Lim, P.E., and Wilson, J.J. (2015). DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. Bull Entomol Res 105(6), 717-727. doi: 10.1017/S0007485315000681.

Brown, E.A., Chain, F.J., Crease, T.J., MacIsaac, H.J., and Cristescu, M.E. (2015). Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities? Ecol Evol 5(11), 2234-2251. doi: 10.1002/ece3.1485.

Bucklin, A., Steinke, D., and Blanco-Bercial, L. (2011). DNA barcoding of marine metazoa. Ann Rev Mar Sci 3, 471-508. doi: 10.1146/annurev-marine-120308-080950.

Bucklin, A., Lindeque, P.K., Rodriguez-Ezpeleta, N., Albaina, A., and Lehtiniemi, M. (2016). Metabarcoding of marine zooplankton: prospects, progress and pitfalls. Journal of Plankton Research 38(3), 393-400. doi: 10.1093/plankt/fbw023.

Burns, R.J., DeForest, J.L., Marxsen, J., Sinsabaughd, R.L., Strombergere, M.E., Wallensteinf, M.D., et al. (2013). Soil enzymes in a changing environment: Current knowledge and future directions. Soil Biology and Biochemistry 58, 216-234. doi: dx.doi.org/10.1016/j.soilbio.2012.11.009.

Cai, W., Borja, A., Liu, L., Meng, W., Muxika, I., and Rodríguez, J.G. (2014). Assessing benthic health under multiple human pressures in Bohai Bay (China), using density and biomass in calculating AMBI and M-AMBI. Marine Ecology 35, 180-192. doi: 10.1111/maec.12067.

Cameron, S., Rubinoff, D., and Will, K. (2006). Who Will Actually Use DNA Barcoding and What Will It Cost? . Systematic Biology 55(5), 844-847. doi: 10.1080/10635150600960079.

Campbell, B.J., Engel, A.S., Porter, M.L., and Takai, K. (2006). The versatile epsilon-proteobacteria: key players in sulphidic habitats. Nat Rev Microbiol 4(6), 458-468. doi: 10.1038/nrmicro1414.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, A.C., Turnbaugh, P.J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. PNAS 108, 4516-4522.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6(8), 1621-1624. doi: 10.1038/ismej.2012.8.

Carew, M.E., Pettigrove, V.J., Metzeling, L., and Hoffmann, A.A. (2013). Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. Front Zool 10(1), 45. doi: 10.1186/1742-9994-10-45.

Carr, S.A., Orcutt, B.N., Mandernack, K.W., and Spear, J.R. (2015). Abundant Atribacteria in deep marine sediment from the Adelie Basin, Antarctica. Front Microbiol 6, 872. doi: 10.3389/fmicb.2015.00872.

Carugati, L., Corinaldesi, C., Dell'Anno, A., and Danovaro, R. (2015). Metagenetic tools for the census of marine meiofaunal biodiversity: An overview. Mar Genomics 24, 11-20. doi: 10.1016/j.margen.2015.04.010.

Caruso, G., La Ferla, R., Azzaro, M., Zoppini, A., Marino, G., Petochi, T., et al. (2015). Microbial assemblages for environmental quality assessment: Knowledge, gaps and usefulness in the European Marine Strategy Framework Directive. Crit Rev Microbiol, 1-22. doi: 10.3109/1040841X.2015.1087380.

CBD (2000). Ecosystem Approach. Fifth Conference of the Parties to the Convention on Biodiversity. Nairobi, Kenya, May. (www.iisd.ca/biodiv/cop5/).

Cearreta, A., Irabien, M.J., Leorri, E., Yusta, I., Croudace, I.W., and Cundy, A.B. (2000). Recent Anthropogenic Impacts on the Bilbao Estuary, Northern Spain: Geochemical and Microfaunal Evidence. Estuar Coast Shelf S 50(4), 571-592. doi: 10.1006/ecss.1999.0582.

Claudet, J., and Fraschetti, S. (2010). Human-driven impacts on marine habitats: a regional meta-analysis in the Mediterranean Sea. Biological Conservation 143(9), 2195-2206.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37-46.

Corell, J., and Rodriguez-Ezpeleta, N. (2014). Tuning of protocols and marker selection to evaluate the diversity of zooplankton using metabarcoding. Revista Investigacion Marina 21(2), 19-39.

Corinaldesi, C., Danovaro, R., and Dell'Anno, A. (2005). Simultaneous recovery of extracellular and intracellular DNA suitable for molecular studies from marine sediments. Appl Environ Microbiol 71(1), 46-50. doi: 10.1128/AEM.71.1.46-50.2005.

Cotner, J.B., and Biddanda, B.A. (2002). Small Players, Large Role: Microbial Influence on Biogeochemical Processes in Pelagic Aquatic Ecosystems. Ecosystems 5(2), 105-121. doi: 10.1007/s10021-001-0059-3.

Cowart, D.A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Mine, J., et al. (2015). Metabarcoding Is Powerful yet Still Blind: A Comparative Analysis of Morphological and Molecular Surveys of Seagrass Communities. PLoS ONE 10(2), e0117562. doi: 10.1371/journal.pone.0117562.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W.K., et al. (2016). The ecologist's field guide to sequence-based identification of biodiversity. Methods Ecol. Evol. 7(9), 1008-1018. doi: 10.1111/2041-210x.12574.

Creer, S., Fonseca, V.G., Porazinska, D.L., Giblin-Davis, R.M., Sung, W., Power, D.M., et al. (2010). Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. Mol Ecol 19 Suppl 1, 4-20. doi: 10.1111/j.1365-294X.2009.04473.x.

Cristescu, M.E. (2014). From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. Trends Ecol Evol 29(10), 566-571. doi: 10.1016/j.tree.2014.08.001.

Crump, B.C., Peranteau, C., Beckingham, B., and Cornwell, J.C. (2007). Respiratory succession and community succession of bacterioplankton in seasonally anoxic estuarine waters. Appl Environ Microbiol 73(21), 6802-6810. doi: 10.1128/AEM.00648-07.

Chain, F.J.J., Brown, E.A., MacIsaac, H.J., Cristescu, M.E., and Cowie, R. (2016). Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. Divers Distrib 22(5), 493–504 doi: 10.1111/ddi.12427.

Chariton, A.A., Court, L.N., Hartley, D.M., Colloff, M.J., and Hardy, C.M. (2010). Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. Front Ecol Environ 8(5), 233-238. doi: 10.1890/090115.

Chariton, A.A., Stephenson, S., Morgan, M.J., Steven, A.D., Colloff, M.J., Court, L.N., et al. (2015). Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. Environ Pollut 203, 165-174. doi: 10.1016/j.envpol.2015.03.047.

Chiellini, C., Iannelli, R., Verni, F., and Petroni, G. (2013). Bacterial communities in polluted seabed sediments: a molecular biology assay in Leghorn Harbor. ScientificWorldJournal 2013, 165706. doi: 10.1155/2013/165706.

Dafforn, K.A., Baird, D.J., Chariton, A.A., Sun, M.Y., Brown, M.V., Simpson, S.L., et al. (2014). "Faster, Higher and Stronger? The Pros and Cons of Molecular Faunal Data for Assessing Ecosystem Condition". in Advances in Ecological Research 51. doi: 10.1016/b978-0-08-099970-8.00003-8.

Danovaro, R., and Pusceddu, A. (2007). Biodiversity and ecosystem functioning in coastal lagoons: Does microbial diversity play any role? Estuar Coast Shelf S 75(1-2), 4-12. doi: 10.1016/j.ecss.2007.02.030.

Danovaro, R., Carugati, L., Berzano, M., Cahill, A.E., Carvalho, S., Chenuil, A., et al. (2016). Implementing and innovating marine monitoring approaches for assessing marine environmental status. Frontiers in Marine Science 3. doi: 10.3389/fmars.2016.00213.

Dauer, D.M. (1993). Biological criteria, environmental health and estuarine macrobenthic community structure. Marine Pollution Bulletin 26, 249-257. doi: 10.1016/0025-326X(93)90063-P.

Davy, C.M., Kidd, A.G., and Wilson, C.C. (2015). Development and Validation of Environmental DNA (eDNA) Markers for Detection of Freshwater Turtles. PLoS ONE 10(7), e0130965. doi: 10.1371/journal.pone.0130965.

De Jonge, V.N., Elliott, M., and Brauer, V.S. (2006). Marine monitoring: its shortcomings and mismatch with the EU Water Framework Directive's objectives. Marine Pollution Bulletin 53, 5-19.

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F., and Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. Biol Lett 10(9). doi: 10.1098/rsbl.2014.0562.

Dean, W.E.J. (1974). Determination of carbonate and organic matter in calcareous sediments and sedimentary rocks by loss on ignition: comparison with other methods. J Sediment Petrol 44, 242-248. doi: 10.1306/74D729D2-2B21-11D7-8648000102C1865D.

Deiner, K., Walser, J.-C., Mächler, E., and Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. Biol Conserv 183, 53-63. doi: 10.1016/j.biocon.2014.11.018.

Dell'Anno, A., and Danovaro, R. (2005). Extracellular DNA plays a key role in deep-sea ecosystem functioning. Science 309, 2179–2179.

Dell'Anno, A., Mei, M.L., Ianni, C., and Danovaro, R. (2003). Impact of bioavailable heavy metals on bacterial activities in coastal marine sediments. World Journal of Microbiology and Biotechnology 19(1), 93-100. doi: 10.1023/A:1022581632116.

Dell'Anno, A., Carugati, L., Corinaldesi, C., Riccioni, G., and Danovaro, R. (2015). Unveiling the Biodiversity of Deep-Sea Nematodes through Metabarcoding: Are We Ready to Bypass the Classical Taxonomy? PLoS ONE 10(12), e0144928. doi: 10.1371/journal.pone.0144928.

Diaz, R.J., and Rosenberg, R. (2008). Spreading dead zones and consequences for marine ecosystems. Science 321.

Diaz, R.J., Solan, M., and Valente, R.M. (2004). A review of approaches for classifying benthic habitats and evaluating habitat quality. J Environ Manage 73(3), 165-181. doi: 10.1016/j.jenvman.2004.06.004.

Doiron, K., Pelletier, E., and Lemarchand, K. (2012). Impact of polymer-coated silver nanoparticles on marine microbial communities: a microcosm study. Aquat Toxicol 124-125, 22-27. doi: 10.1016/j.aquatox.2012.07.004.

Dowle, E.J., Pochon, X., Banks, J., Shearer, K., and Wood, S.A. (2015). Targeted gene enrichment and high throughput sequencing for environmental biomonitoring: a case study using freshwater macroinvertebrates. Mol Ecol Resour 16, 1240–1254 doi: 10.1111/1755-0998.12488.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19), 2460-2461. doi: 10.1093/bioinformatics/btq461.

Edgar, R.C., and Flyvbjerg, H. (2015). Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31(21), 3476-3482. doi: 10.1093/bioinformatics/btv401.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27(16), 2194-2200. doi: 10.1093/bioinformatics/btr381.

Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C., et al. (2011). Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness ISME J 5(8), 1344-1356. doi: 10.1038/ismej.2011.6.

Elbrecht, V., and Leese, F. (2015). Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass--Sequence Relationships with an Innovative Metabarcoding Protocol. PLoS ONE 10(7), e0130324. doi: 10.1371/journal.pone.0130324.

Engle, V.D. (2000). Application of the indicator evaluation guidelines to an index of benthic condition for Gulf of Mexico estuaries. In: Jackson, L.E., Kurtz, J.C., Fisher, W.S. (Eds.), Evaluation Guidelines for Ecological Indicators. EPA/620/R-99/005. U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC,, 107 pp.

Eren, A.M., Morrison, H.G., Lescault, P.J., Reveillaud, J., Vineis, J.H., and Sogin, M.L. (2015). Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. ISME J 9(4), 968-979. doi: 10.1038/ismej.2014.195.

Evans, N.T., Olds, B.P., Renshaw, M.A., Turner, C.R., Li, Y., Jerde, C.L., et al. (2016). Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Mol Ecol Resour 16(1), 29-41. doi: 10.1111/1755-0998.12433.

Falster, D., Warton, D., and Wright, I. (2003). SMATR: standardised major axis tests and routines. See http://www.bio.mq.edu.au/ecology/SMATR.

Ferrera, I., and Sanchez, O. (2016). Insights into microbial diversity in wastewater treatment systems: How far have we come? Biotechnol Adv 34(5), 790-802. doi: 10.1016/j.biotechadv.2016.04.003.

Ficetola, G.F., Miaud, C., Pompanon, F., and Taberlet, P. (2008). Species detection using environmental DNA from water samples. Biol Lett 4(4), 423-425. doi: 10.1098/rsbl.2008.0118.

Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., et al. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Mol Ecol Resour 15(3), 543-556. doi: 10.1111/1755-0998.12338.

Fisher, W.S., Jackson, L.E., Suter, G.W., and Bertram, P. (2001). Indicators for Human and Ecological Risk Assessment: A U.S. Environmental Protection Agency Perspective. Human and Ecological Risk Assessment: An International Journal 7(5). doi: dx.doi.org/10.1080/20018091094781.

Folk, R.L. (1974). Petrology of sedimentary rocks. Austin, TX: Hemphill Publishing Company.

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3(5), 294-299.

Fonseca, V.G., Carvalho, G.R., Nichols, B., Quince, C., Johnson, H.F., Neill, S.P., et al. (2014). Metagenetic analysis of patterns of distribution and diversity of marine meiobenthic eukaryotes. Global Ecology and Biogeography 23(11), 1293-1302. doi: 10.1111/geb.12223.

Fonseca, V.G., and Lallias, D. (2016). "Metabarcoding Marine Sediments: Preparation of Amplicon Libraries",. in Marine Genomics Methods and Protocols, Methods in Molecular Biology, ed. Bourlat, S.J. (Springer, New York), vol. 1452. doi: 10.1007/978-1-4939-3774-5.

Foote, A.D., Thomsen, P.F., Sveegaard, S., Wahlbert, M., Kielgast, J., Kyhn, L.A., et al. (2012). Investigating the Potential Use of Environmental DNA (eDNA) for Genetic Monitoring of Marine Mammals. PLoS ONE 7(8). doi: 10.1371/journal.pone.0041781.g001.

Frolov, S., Kudela, R.M., and Bellingham, J.G. (2013). Monitoring of harmful algal blooms in the era of diminishing resources: A case study of the U.S. West Coast. Harmful Algae 21-22, 1-12.

Gasol, J.M., Pinhassi, J., Alonso-Sáez, L., Ducklow, H., Herndl, G.J., Koblízek, M., et al. (2008). Towards a better understanding of microbial carbon flux in the sea. Aquat Microb Ecol 53(1), 21-38. doi: doi.org/10.3354/ame01230.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31, 3784-3788.

Geller, J., Meyer, C., Parker, M., and Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochromecoxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. Mol Ecol Resour 13(5), 851-861. doi: 10.1111/1755-0998.12138.

Gibson, J., Shokralla, S., Porter, T.M., King, I., van Konynenburg, S., Janzen, D.H., et al. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. Proc Natl Acad Sci U S A 111(22), 8007-8012. doi: 10.1073/pnas.1406468111.

Gibson, J., Shokralla, S., Curry, C., Baird, D.J., Monk, W.A., King, I., et al. (2015). Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. PLoS ONE. doi: 10.5061/dryad.vm72v10.1371/journal.pone.0138432.

Gilbride, K.A., Lee, D.Y., and Beaudette, L.A. (2006). Molecular techniques in wastewater: Understanding microbial communities, detecting pathogens, and real-time process control. J Microbiol Methods 66(1), 1-20. doi: 10.1016/j.mimet.2006.02.016.

Giovannoni, S.J., and Rappé, M.S. (2000). Evolution, diversity, and molecular ecology of marine prokaryotes. in D. Kirchman (ed.), Microbial ecology of the oceans. Wiley-Liss, Inc., New York, N.Y., 47-84.

Goela, P.C., Icely, J., Cristina, S., Danchenko, S., Angel DelValls, T., and Newton, A. (2015). Using bio-optical parameters as a tool for detecting changes in the phytoplankton community (SW Portugal). Estuarine, Coastal and Shelf Science 167, 125-137. doi: 10.1016/j.ecss.2015.07.037.

Goldberg, C.S., Sepulveda, A., Ray, A., Baumgardt, J., and Waits, L.P. (2013). Environmental DNA as a new method for early detection of New Zealand mudsnails (Potamopyrgus antipodarum). Freshwater Science 32(3), 792-800. doi: 10.1899/13-046.1.

Goldberg, C.S., Strickler, K.M., and Pilliod, D.S. (2015). Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. Biol Conserv 183, 1-3. doi: 10.1016/j.biocon.2014.11.040.

Goldberg, C.S., Turner, C.R., Deiner, K., Klymus, K.E., Thomsen, P.F., Murphy, M.A., et al. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods in Ecology and Evolution 7(11), 1299-1307. doi: 10.1111/2041-210x.12595.

Grall, J., and Glémarec, M. (1997). Using biotic indices to estimate macrobenthic community perturbations in the Bay of Brest. Estuarine, Coastal and Shelf Science 44(Supplement 1), 43-53. doi: 10.1016/S0272-7714(97)80006-6.

Guardiola, M., Uriz, M.J., Taberlet, P., Coissac, E., Wangensteen, O.S., and Turon, X. (2015). Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Marine Canyons. PLoS ONE 10(10). doi: 10.1371/journal.pone.013963310.5061/.

Guardiola, M., Wangensteen, O.S., Taberlet, P., Coissac, E., Uriz, M.J., and Turon, X. (2016). Spatio-temporal monitoring of deep-sea communities using metabarcoding of sediment DNA and RNA. PeerJ 4, e2807. doi: 10.7717/peerj.2807.

Haaijer, S.C., Harhangi, H.R., Meijerink, B.B., Strous, M., Pol, A., Smolders, A.J., et al. (2008). Bacteria associated with iron seeps in a sulfur-rich, neutral pH, freshwater ecosystem. ISME J 2(12), 1231-1242. doi: 10.1038/ismej.2008.75.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., et al. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and

454-pyrosequenced PCR amplicons. Genome Res 21(3), 494-504. doi: 10.1101/gr.112730.110.

Hajibabaei, M. (2012). The golden age of DNA metasystematics. Trends Genetics 28(11), 535–537.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A., and Baird, D.J. (2011). Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE 6(4), 17497. doi: 10.1371/journal.pone.0017497.

Hajibabaei, M., Spall, L.J., Shodralla, S., and Konynenburg, S. (2012). Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. BMC Ecology 12(28).

Hajipour, M.J., Fromm, K.M., Ashkarran, A.A., Jimenez de Aberasturi, D., de Larramendi, I.R., Rojo, T., et al. (2012). Antibacterial properties of nanoparticles. Trends Biotechnol 30(10), 499-511. doi: 10.1016/j.tibtech.2012.06.004.

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., et al. (2008). A Global Map of Human Impact on Marine Ecosystems. Science 319(5865), 948-952. doi: 10.1126/science.1149345.

Halpern, B.S., Longo, C., Hardy, D., McLeod, K.L., Samhouri, J.F., Katona, S.K., et al. (2012). An index to assess the health and benefits of the global ocean. Nature 488(7413), 615-620.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 4.

Hansen, M.C., Tolker-Nielsen, T., Givskov, M., and Molin, S. (1998). Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region. FEMS Microbiol Ecol 26, 141-149.

Hardy, S.M., Carr, C.M., Hardman, M., Steinke, D., Corstorphine, E., and Mah, C. (2010). Biodiversity and phylogeography of Arctic marine fauna: insights from molecular tools. Marine Biodiversity 41(1), 195-210. doi: 10.1007/s12526-010-0056-x.

Harley, C.D., Randall Hughes, A., Hultgren, K.M., Miner, B.G., Sorte, C.J., Thornber, C.S., et al. (2006). The impacts of climate change in coastal marine systems. Ecol Lett 9(2), 228-241. doi: 10.1111/j.1461-0248.2005.00871.x.

Hebert, P.D., Cywinska, A., Ball, S.L., and deWaard, J.R. (2003a). Biological identifications through DNA barcodes. Proc Biol Sci 270(1512), 313-321. doi: 10.1098/rspb.2002.2218.

Hebert, P.D., Ratnasingham, S., and deWaard, J.R. (2003b). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc Biol Sci 270 Suppl 1, S96-99. doi: 10.1098/rsbl.2003.0025.

Heiskanen, A.S., van de Bund, W., Cardoso, A.C., and Nóges, P. (2004). Towards good ecological status of surface waters in Europe – Interpretation and harmonisation of the concept. Water Science and Technology 49(7), 169-177.

Heiskanen, A.-S., Berg, T., Uusitalo, L., Teixeira, H., Bruhn, A., Krause-Jensen, D., et al. (2016). Biodiversity in Marine Ecosystems—European Developments toward Robust Assessments. Frontiers in Marine Science 3. doi: 10.3389/fmars.2016.00184.

Hirai, J., Kuriyama, M., Ichikawa, T., Hidaka, K., and Tsuda, A. (2015). A metagenetic approach for revealing community structure of marine planktonic copepods. Mol Ecol Resour 15(1), 68-80. doi: 10.1111/1755-0998.12294.

Hoppe, H.G., Breithaupt, P., Walther, K., Koppe, R., Bleck, S., Sommer, U., et al. (2008). Climate warming in winter affects the coupling between phytoplankton and bacteria during the spring bloom: a mesocosm study. Aquatic Microbial Ecology 51(2), 105-115. doi: doi.org/10.3354/ame01198.

Illumina (2014). Nextera® XT DNA Library Preparation Kit, Pub. No. 770-2012-011. http://www.illumina.com/documents/products/datasheets/datasheet_nextera_xt_dna_sample_prep.pdf.

Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., et al. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecol Lett 16(10), 1245-1257. doi: 10.1111/ele.12162.

Johnston, E.L., and Roberts, D.A. (2009). Contaminants reduce the richness and evenness of marine communities: a review and meta-analysis. Environ Pollut 157(6), 1745-1752. doi: 10.1016/j.envpol.2009.02.017.

Kelly, R.P., Port, J.A., Yamahara, K.M., and Crowder, L.B. (2014a). Using environmental DNA to census marine fishes in a large mesocosm. PLoS ONE 9(1), e86175. doi: 10.1371/journal.pone.0086175.

Kelly, R.P., Port, J.A., Yamahara, K.M., Martone, R.G., Lowell, N., Thomsen, P.F., et al. (2014b). Harnessing DNA to improve environmental management. Science 344(6191), 1455-1456. doi: 10.1126/science.1251156.

Kersten, M., and Smedes, F. (2002). Normalization procedures for sediment contaminants in spatial and temporal trend monitoring. J Environ Monit 4, 109–115.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2012). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research 41(1), e1-e1. doi: 10.1093/nar/gks808.

Kochzius, M., Nolte, M., Weber, H., Silkenbeumer, N., Hjorleifsdottir, S., Hreggvidsson, G.O., et al. (2008). DNA microarrays for identifying fishes. Mar Biotechnol (NY) 10(2), 207-217. doi: 10.1007/s10126-007-9068-3.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79(17), 5112-5120. doi: 10.1128/AEM.01043-13.

Kress, W.J., Garcia-Robledo, C., Uriarte, M., and Erickson, D.L. (2015). DNA barcodes for ecology, evolution, and conservation. Trends Ecol Evol 30(1), 25-35. doi: 10.1016/j.tree.2014.10.008.

Laforest, B.J., Winegardner, A.K., Zaheer, O.A., Jeffery, N.W., Boyle, E., and Adamowicz, S. (2013). Insights into biodiversity sampling strategies for freshwater microinvertebrate faunas through bioblitz campaigns and DNA barcoding. BMC Ecology 13(13).

Langmuir, D. (1971). Eh-pH determination. In: Carver RE, editor. Sedimentary petrology. John Wiley & Sons, p. 597–634.

Laroche, O., Wood, S.A., Tremblay, L.A., Ellis, J.I., Lejzerowicz, F., Pawlowski, J., et al. (2016). First evaluation of foraminiferal metabarcoding for monitoring environmental impact from an offshore oil drilling site. Mar Environ Res 120, 225-235. doi: 10.1016/j.marenvres.2016.08.009.

Lee, C.K., Herbold, C.W., Polson, S.W., Wommack, K.E., Williamson, S.J., McDonald, I.R., et al. (2012). Groundtruthing next-gen sequencing for microbial ecology-biases and errors in community structure estimates from PCR amplicon pyrosequencing. PLoS ONE 7(9), e44224. doi: 10.1371/journal.pone.0044224.

Lejzerowicz, F., Esling, P., Pillet, L., Wilding, T.A., Black, K.D., and Pawlowski, J. (2015). High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. Sci Rep 5, 13932. doi: 10.1038/srep13932.

Leray, M., and Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. PNAS 112(7), 2076–2081. doi: 10.1073/pnas.142499711210.5883/DS-ARMS.

Leray, M., Yang, Y.J., Meyer, P.C., Mills, C.S., Agudelo, N., Ranwez, V., et al. (2013). New versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front Zool 10(34).

Leray, M., Haenel, Q., and Bourlat, S.J. (2016). "Preparation of Amplicon Libraries for Metabarcoding of Marine Eukaryotes Using Illumina MiSeq: The Adapter Ligation Method,". in Marine Genomics Methods and Protocols, Methods in Molecular Biology, ed. Bourlat, S.J. (Springer, New York), vol. 1452. doi: 10.1007/978-1-4939-3774-5.

Lerberg, S.B., Holland, A.F., and Sanger, D.M. (2000). Responses of tidal creek macrobenthic communities to the effects of watershed development. Estuaries 23(6), 838-853. doi: doi:10.2307/1353001.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22(13), 1658-1659. doi: 10.1093/bioinformatics/btl158.

Lindeque, P.K., Parry, H.E., Harmer, R.A., Somerfield, P.J., and Atkinson, A. (2013). Next generation sequencing reveals the hidden diversity of zooplankton assemblages. PLoS One 8(11), e81327. doi: 10.1371/journal.pone.0081327.

Liu, Z., DeSantis, T.Z., Andersen, G.L., and Knight, R. (2008). Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Res 36(18), 120. doi: 10.1093/nar/gkn491.

Lobo, J., Costa, M.P., Teixeira, M.A., Ferreira, M.S., Costa, M.H., and Costa, F.O. (2013). Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. BMC Ecology 13(34).

Loman, N.J., Misra, R.V., Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J., et al. (2012). Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30(5), 434-439. doi: 10.1038/nbt.2198.

Lotze, H.K., Lenihan, H.S., Bourque, B.J., Bradbury, R., Cooke, R.G., Kay, M.C., et al. (2006). Depletion, degradation, and recovery potential of estuaries and coastal seas. Science 312, 1806-1809.

Lozada, M., Marcos, M.S., Commendatore, M.G., Gil, M.N., and Dionisi, H.M. (2014). The Bacterial Community Structure of Hydrocarbon-Polluted Marine Environments as the Basis for the Definition of an Ecological Index of Hydrocarbon Exposure. Microbes and Environments 29(3), 269-276. doi: 10.1264/jsme2.ME14028.

Machida, J.R., and Tsuda, A. (2010). Dissimilarity of Species and Forms of Planktonic Neocalanus copepods Using Mitochondrial COI, 12S, Nuclear ITS, and 28S Gene Sequences. PLoS ONE 5(4), 6. doi: 10.1371/journal.pone.0010278.t001.

Machida, J.R., and Knowlton, N. (2012). PCR Primers for Metazoan Nuclear 18S and 28S Ribosomal DNA Sequences. PLoS ONE 7(9). doi: 10.1371/journal.pone.0046180.t001.

Mächler, E., Deiner, K., Steinmann, P., and Altermatt, F. (2014). Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. Freshwater Science 33(4), 1174-1183. doi: 10.1086/678128.

Magoč, T., and Salzberg, S. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27(21), 2957-2963. doi: 10.1093/bioinformatics/btr507.

Markmann, M., and Tautz, D. (2005). Reverse taxonomy: an approach towards determining the diversity of meiobenthic organisms based on ribosomal RNA signature sequences. Philos Trans R Soc Lond B Biol Sci 360(1462), 1917-1924. doi: 10.1098/rstb.2005.1723.

Marques, J.C., Maranhão, P., and Pardal, M.A. (1993). Human Impact Assessment on the Subtidal Macrobenthic Community Structure in the Mondego Estuary (Western Portugal). Estuarine, Coastal and Shelf Science 37(4), 403-419. doi: 10.1006/ecss.1993.1064.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17(1), 10-12. doi: 10.14806/ej.17.1.200.

Matheson, C.D., Gurney, C., Esau, N., and Lehto, R. (2010). Assessing PCR Inhibition from Humic Substances. Open Enzym Inhib J 3, 38-45.

Matzen da Silva, J., Creer, S., dos Santos, A., Costa, A.C., Cunha, M.R., Costa, F.O., et al. (2011). Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca). PLoS ONE 6(5), 19449. doi: 10.1371/journal.pone.0019449.

Maurer, D., Nguyen, H., Robertson, G., and Gerlinger, T. (1999). The Infaunal Trophic Index (ITI): its suitability for marine environmental monitoring. Ecol Appl 9, 14.

McLusky, D.S., and Elliot, M. (2004). The estuarine ecosystem – ecology, threats and management. Oxford University Press, 214 pp.

Menchaca, I., Borja, Á., Belzunce-Segarra, M.J., Franco, J., Garmendia, J.M., Larreta, J., et al. (2012). An empirical approach to the determination of metal regional Sediment Quality Guidelines, in marine waters, within the European Water Framework Directive. Chemistry and Ecology 28(3), 205-220. doi: 10.1080/02757540.2011.651129.

Menchaca, I., Rodríguez, J.G., Borja, Á., Belzunce-Segarra, M.J., Franco, J., Garmendia, J.M., et al. (2014). Determination of polychlorinated biphenyl and polycyclic aromatic hydrocarbon marine regional Sediment Quality Guidelines within the European Water Framework Directive. Chemistry and Ecology 30(8), 693-700. doi: 10.1080/02757540.2014.917175.

Meusnier, I., Singer, G.A., Landry, J.F., Hickey, D.A., Hebert, P.D., and Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9, 214. doi: 10.1186/1471-2164-9-214.

Meyer, P.C. (2003). Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. Biol J Linn Soc 79, 401-459.

Miralles, L., Ardura, A., Arias, A., Borrell, Y.J., Clusa, L., Dopico, E., et al. (2016). Barcodes of marine invertebrates from north Iberian ports: Native diversity and resistance to biological invasions. Marine Pollution Bulletin 112, 183-188.

Mohrbeck, I., Raupach, M.J., Martinez Arbizu, P., Knebelsberger, T., and Laakmann, S. (2015). High-Throughput Sequencing-The Key to Rapid Biodiversity Assessment of Marine Metazoa? PLoS ONE 10(10), e0140342. doi: 10.1371/journal.pone.0140342.

Monserud, R.A., and Leemans, R. (1992). Comparing global vegetation maps with the Kappa statistic. Ecol Modell 62, 275-293.

Muniz, P., Venturini, N., Pires-Vanin, A.M., Tommasi, L.R., and Borja, A. (2005). Testing the applicability of a Marine Biotic Index (AMBI) to assessing the ecological quality of soft-bottom benthic communities, in the South America Atlantic region. Mar Pollut Bull 50(6), 624-637. doi: 10.1016/j.marpolbul.2005.01.006.

Muxika, I., Borja, A., and Bonne, W. (2005). The suitability of the marine biotic index (AMBI) to new impact sources along European coasts. Ecological Indicators 5(1), 19-31. doi: dx.doi.org/10.1016/j.ecolind.2004.08.004.

Muxika, I., Borja, A., and Bald, J. (2007). Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive. Marine Pollution Bulletin 55(1-6), 16-29. doi: dx.doi.org/10.1016/j.marpolbul.2006.05.025.

Muxika, I., Somerfield, P.J., Borja, Á., and Warwick, R.M. (2012). Assessing proposed modifications to the AZTI marine biotic index (AMBI), using biomass and production. Ecological Indicators 12(1), 96-104. doi: 10.1016/j.ecolind.2011.04.030.

Niu, B., Fu, L., Sun, S., and Li, W. (2010). Artificial and natural duplicates in pyrosequencing reads of metagenomic data. BMC bioinformatics 11(187).

Nogales, B., Lanfranconi, M.P., Pina-Villalonga, J.M., and Bosch, R. (2011). Anthropogenic perturbations in marine microbial communities. FEMS Microbiol Rev 35(2), 275-298. doi: 10.1111/j.1574-6976.2010.00248.x.

OSPAR (2009). Trend Analysis of Maritime Human Activities and their Collective Impact on the OSPAR Maritime Area, Publication Number: 44./2009. London: OSPAR Commission.

Parks, D.H., Tyson, G.W., Hugenholtz, P., and Beiko, R.G. (2014). STAMP: Statistical analysis of taxonomic and functional profiles. Bioinformatics. doi: 10.1093/bioinformatics/btu494

Pawlowski, J., Audic, S., Adl, S., Bass, D., Berney, C., Bowser, S.S., et al. (2012). CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. PLoS Biol 10(11)(e1001419). doi: 10.1371/journal.pbio.1001419.g001.

Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T., and Wilding, T.A. (2014a). Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. Mol Ecol Resour 14(6), 1129-1140. doi: 10.1111/1755-0998.12261.

Pawlowski, J., Lejzerowicz, F., and Esling, P. (2014b). Next-Generation Environmental Diversity Surveys of Foraminifera: Preparing the Future. Biol. Bull 227, 93-106. doi: 10.1086/BBLv227n2p93

Pearman, J.K., and Irigoien, X. (2015). Assessment of Zooplankton Community Composition along a Depth Profile in the Central Red Sea. PLoS One 10(7), e0133487. doi: 10.1371/journal.pone.0133487.

Pearman, J.K., Anlauf, H., Irigoien, X., and Carvalho, S. (2016a). Please mind the gap - Visual census and cryptic biodiversity assessment at central Red Sea coral reefs. Mar Environ Res 118, 20-30. doi: 10.1016/j.marenvres.2016.04.011.

Pearman, J.K., El-Sherbiny, M.M., LanzÃ©n, A., Al-Aidaroos, A.M., and Irigoien, X. (2014). Zooplankton diversity across three Red Sea reefs using pyrosequencing. Frontiers in Marine Science 1. doi: 10.3389/fmars.2014.00027.

Pearman, J.K., Irigoien, X., and Carvalho, S. (2016b). Extracellular DNA amplicon sequencing reveals high levels of benthic eukaryotic diversity in the central Red Sea. Mar Genomics 26, 29-39. doi: 10.1016/j.margen.2015.10.008.

Penna, A., and Galluzzi, L. (2013). The quantitative real-time PCR applications in the monitoring of marine harmful algal bloom (HAB) species. Environ Sci Pollut Res Int 20(10), 6851-6862. doi: 10.1007/s11356-012-1377-z.

Pietramellara, G., Ascher, J., Borgogni, F., Ceccherini, M.T., Guerri, G., and Nannipieri, P. (2009). Extracellular DNA in soil and sediment: fate and ecological relevance. Biology and Fertility of Soils 45(3), 219-235

Pinto, A.J., and Raskin, L. (2012). PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PLoS ONE 7(8), e43093. doi: 10.1371/journal.pone.0043093.

Pinto, R., Patrício, J., Baeta, A., Fath, B.D., Neto, J.M., and Marques, J.C. (2009). Review and evaluation of estuarine biotic indices to assess benthic condition. Ecol Indic 9(1), 1-25. doi: 10.1016/j.ecolind.2008.01.005.

Piñol, J., Mir, G., Gomez-Polo, P., and Agusti, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Mol Ecol Resour 15(4), 819-830. doi: 10.1111/1755-0998.12355.

Pochon, X., Bott, N.J., Smith, K.F., and Wood, S.A. (2013). Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. PLoS ONE 8(9), e73935. doi: 10.1371/journal.pone.0073935.

Pochon, X., Wood, S.A., Keeley, N.B., Lejzerowicz, F., Esling, P., Drew, J., et al. (2015). Accurate assessment of the impact of salmon farming on benthic sediment enrichment using foraminiferal metabarcoding. Mar Pollut Bull 100(1), 370-382. doi: 10.1016/j.marpolbul.2015.08.022.

Pompanon, F., Deagle, B.E., Symondson, W.O., Brown, D.S., Jarman, S.N., and Taberlet, P. (2012). Who is eating what: diet assessment using next generation sequencing. Mol Ecol 21(8), 1931-1950. doi: 10.1111/j.1365-294X.2011.05403.x.

Ponti, M., Pinna, M., Basset, A., Moncheva, S., Trayanova, A., Georgescu, L.P., et al. (2008). Quality assessment of Mediterranean and Black Sea transitional waters: comparing responses of benthic biotic indices. Aquatic Conservation: Marine and Freshwater Ecosystems 18(S1), S62-S75. doi: 10.1002/aqc.952.

Porazinska, D.L., Giblin-Davis, R.M., Faller, L., Farmerie, W., Kanzaki, N., Morris, K., et al. (2009). Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. Mol Ecol Resour 9(6), 1439-1450. doi: 10.1111/j.1755-0998.2009.02611.x.

Porazinska, D.L., Sung, W., Giblin-Davis, R.M., and Thomas, W.K. (2010). Reproducibility of read numbers in high-throughput sequencing analysis of nematode community composition and structure. Mol Ecol Resour 10(4), 666-676. doi: 10.1111/j.1755-0998.2009.02819.x.

Prosser, S.W.J., Velarde-Aguilar, M.G., León-Règagnon, V., and Hebert, P.D.N. (2013). Advancing nematode barcoding: A primer cocktail for the cytochrome c oxidase subunit I gene from vertebrate parasitic nematodes. Molecular Ecology Resources.

Pusceddu, A., Dell'Anno, A., Fabiano, M., and Danovaro, R. (2009). Quantity and bioavailability of sediment organic matter as signatures of benthic trophic status. Marine Ecology Progress Series 375, 41-52. doi: 10.3354/meps07735.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41(Database issue), D590-596. doi: 10.1093/nar/gks1219.

R Development Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Ranasinghe, J.A., Stein, E.D., Miller, P.E., and Weisberg, S.B. (2012). Performance of Two Southern California Benthic Community Condition Indices Using Species Abundance and Presence-Only Data: Relevance to DNA Barcoding. PLoS ONE 7(8). doi: 10.1371/journal.pone.0040875.g001.

Ratnasingham, S., and Hebert, P.D. (2007). BOLD: The Barcode of Life Data System (www.barcodinglife.org). Molecular Ecology Notes 7, 355-364. doi: 10.1111/j.1471-8286.2006.01678.x.

Ratnasingham, S., and Hebert, P.D. (2013). A DNA-based registry for all animal species: the barcode index number (BIN) system. PLoS ONE 8(7), e66213. doi: 10.1371/journal.pone.0066213.

Rees, H.L., Hyland, J.L., Mercer Clarke, C.S.L., Roff, J.C., and Ware, S. (2008). Environmental Indicators: utility in meeting regulatory needs. An overview. ICES Journal of Marine Science 65, 1381-1386.

Rodriguez-Ezpeleta, N., Mendibil, I., Álvarez, P., and Cotano, U. (2013). Effect of fish sampling and tissue storage conditions in DNA quality: considerations for genomic studies. Revista Investigacion Marina 20(6), 77-87.

Rosenberg, R., Blomquist, M., Nilsson, H.C., Cederwall, H., and Dimming, A. (2004). Marine quality assessment by use of benthic species-abundance distribution: a proposed new protocol within the European Union Water framework Directive. Mar Pollut Bull 49, 728-739. doi: 10.1016/j.marpolbul.2004.05.013.

Roussel, J.-M., Paillisson, J.-M., Tréguier, A., Petit, E., and Cadotte, M. (2015). The downside of eDNA as a survey tool in water bodies. J Appl Ecol 52(4), 823-826. doi: 10.1111/1365-2664.12428.

Salas, F., Neto, J.M., Borja, A., and Marques, J.C. (2004). Evaluation of the applicability of a marine biotic index to characterize the status of estuarine ecosystems: the case of Mondego estuary (Portugal). Ecological Indicators 4(3), 215-225. doi: dx.doi.org/10.1016/j.ecolind.2004.04.003.

Sarri, C., Stamatis, C., Sarafidou, T., Galara, I., Godosopoulos, V., Kolovos, M., et al. (2014). A new set of 16S rRNA universal primers for identification of animal species. Food Control 43, 35-41. doi: 10.1016/j.foodcont.2014.02.036.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T., and Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res 43(6), e37. doi: 10.1093/nar/gku1341.

Schloss, P.D. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75(23), 7537-7541.

Serth, J., Kuczyk, A., Paeslack, U., Lichtinghagen, R., and Jonas, U. (2000). Quantitation of DNA Extracted after Micropreparation of Cells from Frozen and Formalin-Fixed Tissue Sections. Am J Pathol 156(4), 1189-1196.

Shackleton, M., and Rees, G.M. (2016). DNA barcoding Australian macroinvertebrates for monitoring programs: benefits and current short comings. Marine and Freshwater Research 67, 380-390.

Shchegolkova, N.M., Krasnov, G.S., Belova, A.A., Dmitriev, A.A., Kharitonov, S.L., Klimina, K.M., et al. (2016). Microbial Community Structure of Activated Sludge in Treatment Plants with Different Wastewater Compositions. Front Microbiol 7, 90. doi: 10.3389/fmicb.2016.00090.

Sigsgaard, E.E., Carl, H., Møller, P.R., and Thomsen, P.F. (2015). Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. Biol Conserv 183, 46-52. doi: 10.1016/j.biocon.2014.11.023.

Simboura, N., and Zenetos, A. (2002). Benthic indicators to use in ecological quality classification of Mediterranean soft bottom marine ecosystems, including a new biotic index. Mediterranean Marine Science 3(2), 77-111. doi: 10.12681/mms.249.

Simpson, J.M., McCracken, V.J., Gaskins, H.R., and Mackie, R.I. (2000). Denaturing Gradient Gel Electrophoresis Analysis of 16S Ribosomal DNA Amplicons To Monitor Changes in Fecal Bacterial Populations of Weaning Pigs after Introduction of Lactobacillus reuteri Strain MM53. Appl Environ Microbiol 66(11), 4705-4714.

Sinclair, L., Osman, O.A., Bertilsson, S., and Eiler, A. (2015). Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the illumina platform. PLoS One 10(2), e0116955. doi: 10.1371/journal.pone.0116955.

Sipos, R., Szekely, A.J., Palatinszky, M., Revesz, S., Marialigeti, K., and Nikolausz, M. (2007). Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. FEMS Microbiol Ecol 60(2), 341-350. doi: 10.1111/j.1574-6941.2007.00283.x.

Smart, A.S., Weeks, A.R., van Rooyen, A.R., Moore, A., McCarthy, M.A., Tingley, R., et al. (2016). Assessing the cost-efficiency of environmental DNA sampling. Methods in Ecology and Evolution 7(11), 1291-1298. doi: 10.1111/2041-210x.12598.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci U S A 103(32), 12115-12120. doi: 10.1073/pnas.0605127103.

Solaun, O., Rodriguez, J.G., Borja, A., Gonzalez, M., and Saiz-Salinas, J.I. (2013). Biomonitoring of metals under the water framework directive: detecting temporal trends and abrupt changes, in relation to the removal of pollution sources. Mar Pollut Bull 67(1-2), 26-35. doi: 10.1016/j.marpolbul.2012.12.005.

Stein, E.D., White, B.P., Mazor, R.D., Miller, E.P., and Pilgrim, E.M. (2013). Evaluating Ethanol-based Sample Preservation to Facilitate Use of DNA Barcoding in Routine Freshwater Biomonitoring Programs Using Benthic Macroinvertebrates. PLoS ONE 8(1). doi: 10.1371/journal.pone.0051273.t001.

Stein, E.D., Martinez, M.C., Stiles, S., Miller, P.E., and Zakharov, E.V. (2014). Is DNA barcoding actually cheaper and faster than traditional morphological methods: results from a survey of freshwater bioassessment efforts in the United States? PLoS ONE 9(4), e95525. doi: 10.1371/journal.pone.0095525.

Stoeck, T., Behnke, A., Christen, R., Amaral-Zettler, L., Rodriguez-Mora, M.J., Chistoserdov, A., et al. (2009). Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. BMC Biol 7, 72. doi: 10.1186/1741-7007-7-72.

Strong, J.A., Andonegi, E., Bizsel, K.C., Danovaro, R., Elliott, M., Franco, A., et al. (2015). Marine biodiversity and ecosystem function relationships: The potential for practical monitoring applications. Estuar Coast Shelf S 161, 46-64. doi: 10.1016/j.ecss.2015.04.008.

Stucky, B.J. (2012). SeqTrace: A Graphical Tool for Rapidly Processing DNA Sequencing Chromatograms. J Biomol Tech 23(3), 90-93. doi: 10.7171/jbt.12-2303-004.

Sun, M.Y., Dafforn, K.A., Johnston, E.L., and Brown, M.V. (2013). Core sediment bacteria drive community response to anthropogenic contamination over multiple environmental gradients. Environ Microbiol 15(9), 2517-2531. doi: 10.1111/1462-2920.12133.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C.-., and Willerslev, E. (2012a). Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol 21, 2045-2050.

Taberlet, P., Prud'Homme, S.M., Campione, E., Roy, J., Miquel, C., Shehzad, W., et al. (2012b). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Mol Ecol 21(8), 1816-1820. doi: 10.1111/j.1365-294X.2011.05317.x.

Tan, B., Ng, C., Nshimyimana, J.P., Loh, L.L., Gin, K.Y., and Thompson, J.R. (2015a). Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities. Front Microbiol 6, 1027. doi: 10.3389/fmicb.2015.01027.

Tan, E.L., Mayer-Pinto, M., Johnston, E.L., and Dafforn, K.A. (2015b). Differences in Intertidal Microbial Assemblages on Urban Structures and Natural Rocky Reef. Front Microbiol 6, 1276. doi: 10.3389/fmicb.2015.01276.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., and Vogler, A.P. (2003). A plea for DNA taxonomy. Trends in Ecology and Evolution 18(2), 70-74.

Taylor, H.R., and Harris, W.E. (2012). An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. Mol Ecol Resour 12(3), 377-388. doi: 10.1111/j.1755-0998.2012.03119.x.

Tedersoo, L., Ramirez, K.S., Nilsson, R.H., Kaljuvee, A., Koljalg, U., and Abarenkov, K. (2015). Standardizing metadata and taxonomic identification in metabarcoding studies. Gigascience 4, 34. doi: 10.1186/s13742-015-0074-5.

Thomsen, P.F., and Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv 183, 4-18. doi: 10.1016/j.biocon.2014.11.019.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Moller, P.R., Rasmussen, M., and Willerslev, E. (2012a). Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLoS ONE 7(8), e41732. doi: 10.1371/journal.pone.0041732.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T., et al. (2012b). Monitoring endangered freshwater biodiversity using environmental DNA. Mol Ecol 21(11), 2565-2573. doi: 10.1111/j.1365-294X.2011.05418.x.

Torti, A., Lever, M.A., and Jorgensen, B.B. (2015). Origin, dynamics, and implications of extracellular DNA pools in marine sediments. Mar Genomics 24(185-196). doi: 10.1016/j.margen.2015.08.007.

Turnbaugh, P.J., Quince, C., Faith, J.J., McHardy, A.C., Yatsunenko, T., Niazi, F., et al. (2010). Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proc Natl Acad Sci U S A 107(16), 7503-7508. doi: 10.1073/pnas.1002355107.

Turner, C.R., Uy, K.L., and Everhart, R.C. (2015). Fish environmental DNA is more concentrated in aquatic sediments than surface water. Biol Conserv 183, 93-102. doi: 10.1016/j.biocon.2014.11.017.

UNCLOS (1982). United Nations Convention on the Law of the Sea, signed at Montego Bay, Jamaica, on 10 December 1982. 202 www.un.org/Depts/los/index.htm.

Valenca, A.P., and Santos, P.J. (2012). Macrobenthic community for assessment of estuarine health in tropical areas (Northeast, Brazil): review of macrofauna classification in ecological groups and application of AZTI Marine Biotic Index. . Mar Pollut Bull 64(9), 1809-1820. doi: 10.1016/j.marpolbul.2012.06.003.

Valentini, A., Pompanon, F., and Taberlet, P. (2009). DNA barcoding for ecologists. Trends Ecol Evol 24, 110-117. doi: dx.doi.org/10.1016/j.tree.2008.09.011.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., et al. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol Ecol 25(4), 929-942. doi: 10.1111/mec.13428.

Visco, J.A., Apotheloz-Perret-Gentil, L., Cordonier, A., Esling, P., Pillet, L., and Pawlowski, J. (2015). Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data. Environ Sci Technol 49(13), 7597–7605. doi: 10.1021/es506158m.

Vivien, R., Wyler, S., Lafont, M., and Pawlowski, J. (2015). Molecular barcoding of aquatic oligochaetes: implications for biomonitoring. PLoS One 10(4), e0125485. doi: 10.1371/journal.pone.0125485.

Walters, W.A., Caporaso, J.G., Lauber, C.L., Berg-Lyons, D., Fierer, N., and Knight, R. (2011). PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. Bioinformatics 27, 3.

Walther, B.A., and Moore, J.L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography 28(6), 815-829. doi: 10.1111/j.2005.0906-7590.04112.x.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73(16), 5261-5267. doi: 10.1128/AEM.00062-07.

Wang, Y., Sheng, H.F., He, Y., Wu, J.Y., Jiang, Y.X., Tam, N.F., et al. (2012). Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. Appl Environ Microbiol 78(23), 8264-8271. doi: 10.1128/AEM.01821-12.

Wangensteen, O.S., and Turon, X. (2016). Metabarcoding techniques for assessing biodiversity of marine animal forests. Marine Animal Forests. The Ecology of Benthic Biodiversity Hotspots., 1-34.

Warton, D.I., Wright, I.J., Falster, D.S., and Westoby, M. (2006). Bivariate line-fitting methods for allometry. Biol Rev 81(2), 259-291. doi: 10.1017/s1464793106007007.

Warwick, R.M., Robert Clarke, K., and Somerfield, P.J. (2010). Exploring the marine biotic index (AMBI): variations on a theme by Ángel Borja. Marine Pollution Bulletin 60, 554-559. doi: dx.doi.org/10.1016/j.marpolbul.2009.11.009.

Weisberg, S.B., Ranasinghe, J.A., Dauer, D.M., Schaffner, L.C., Díaz, R.J., and Frithsen, J.B. (1997). An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. Estuaries 20, 149-158.

Wood, S.A., Smith, K.F., Banks, J.C., Tremblay, L.A., Rhodes, L., Mountfort, D., et al. (2013). Molecular genetic tools for environmental monitoring of New Zealand's aquatic habitats, past, present and the future. New Zeal J Mar Fresh 47(1), 90-119. doi: 10.1080/00288330.2012.745885.

Yang, C.X., Ji, Y.Q., Wang, X.Y., Yang, C.Y., and Yu, D.W. (2013). Testing three pipelines for 18S rDNA-based metabarcoding of soil faunal diversity. Science China-Life Sciences 56(1), 73-81. doi: 10.1007/s11427-012-4423-7.

Ye, Q., Wu, Y., Zhu, Z., Wang, X., Li, Z., and Zhang, J. (2016). Bacterial diversity in the surface sediments of the hypoxic zone near the Changjiang Estuary and in the East China Sea. Microbiologyopen 5(2), 323-339. doi: 10.1002/mbo3.330.

Yoon, T.H., Kang, H.E., Kang, C.K., Lee, S.H., Ahn, D.H., Park, H., et al. (2016). Development of a cost-effective metabarcoding strategy for analysis of the marine phytoplankton community. PeerJ 4, e2115. doi: 10.7717/peerj.2115.

Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C., et al. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods Ecol Evol 3(4), 613-623. doi: 10.1111/j.2041-210X.2012.00198.x.

Zaiko, A., Samuiloviene, A., Ardura, A., and Garcia-Vazquez, E. (2015). Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. Mar Pollut Bull 100, 53-59. doi: 10.1016/j.marpolbul.2015.09.030.

Zepeda Mendoza, M.L., Sicheritz-Ponten, T., and Gilbert, M.T. (2015). Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses. Brief Bioinform, 1-14. doi: 10.1093/bib/bbv001.

Zettler, M.L., Schiedek, D., and Bobertz, B. (2007). Benthic biodiversity indices versus salinity gradient in the southern Baltic Sea. Mar Pollut Bull 55(1-6), 258-270. doi: 10.1016/j.marpolbul.2006.08.024.

Zhang, L., Kang, M., Xu, J., Shuai, Y., Zhou, X., Yang, Z., et al. (2016). Bacterial and archaeal communities in the deep-sea sediments of inactive hydrothermal vents in the Southwest India Ridge. Sci Rep 6, 25982. doi: 10.1038/srep25982.

Zhang, W., Ki, J.-S., and Qian, P.-Y. (2008a). Microbial diversity in polluted harbor sediments I: Bacterial community assessment based on four clone libraries of 16S rDNA. Estuar Coast Shelf S 76(3), 668-681. doi: 10.1016/j.ecss.2007.07.040.

Zhang, W., Song, L., Ki, J., Lau, C., Li, X., and Qian, P. (2008b). Microbial diversity in polluted harbor sediments II: Sulfate-reducing bacterial community assessment using terminal restriction fragment length polymorphism and clone library of dsrAB gene. Estuar Coast Shelf S 76(3), 682-691. doi: 10.1016/j.ecss.2007.07.039.

Zhang, Y., Chen, L., Sun, R., Dai, T., Tian, J., Liu, R., et al. (2014). Effect of wastewater disposal on the bacterial and archaeal community of sea sediment in an industrial area in China. FEMS Microbiol Ecol 88(2), 320-332. doi: 10.1111/1574-6941.12298.

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., et al. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. Gigascience 2(1), 4. doi: 10.1186/2047-217X-2-4.

Ziegler, M., Roik, A., Porter, A., Zubier, K., Mudarris, M.S., Ormond, R., et al. (2016). Coral microbial community dynamics in response to anthropogenic impacts near a major city in the central Red Sea. Mar Pollut Bull 105(2), 629-640. doi: 10.1016/j.marpolbul.2015.12.045.

Zimmermann, J., Glockner, G., Jahn, R., Enke, N., and Gemeinholzer, B. (2015). Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. Mol Ecol Resour 15(3), 526-542. doi: 10.1111/1755-0998.12336.

Zinger, L., Amaral-Zettler, L.A., Fuhrman, J.A., Horner-Devine, M.C., Huse, S.M., Welch, D.B., et al. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. PLoS One 6(9), e24570. doi: 10.1371/journal.pone.0024570.

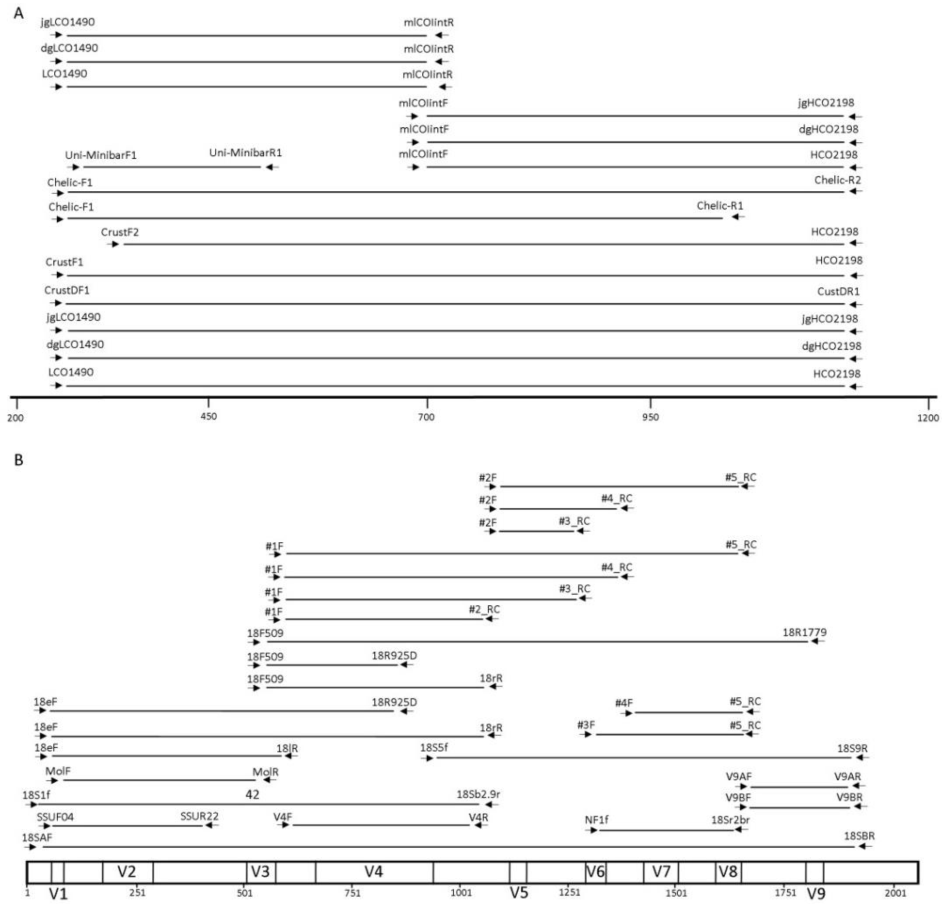

Figure S1.1. Primer pair positions. Position of the primer pairs tested for *COI* (A) on the *COI* region of the complete mitochondrial gene of *Mytilus galloprovincialis* (Accession number DQ399833) and for *18S rRNA* (B) on the *18S rRNA* sequence of *Aplysia punctate* (Accession number AJ224919).

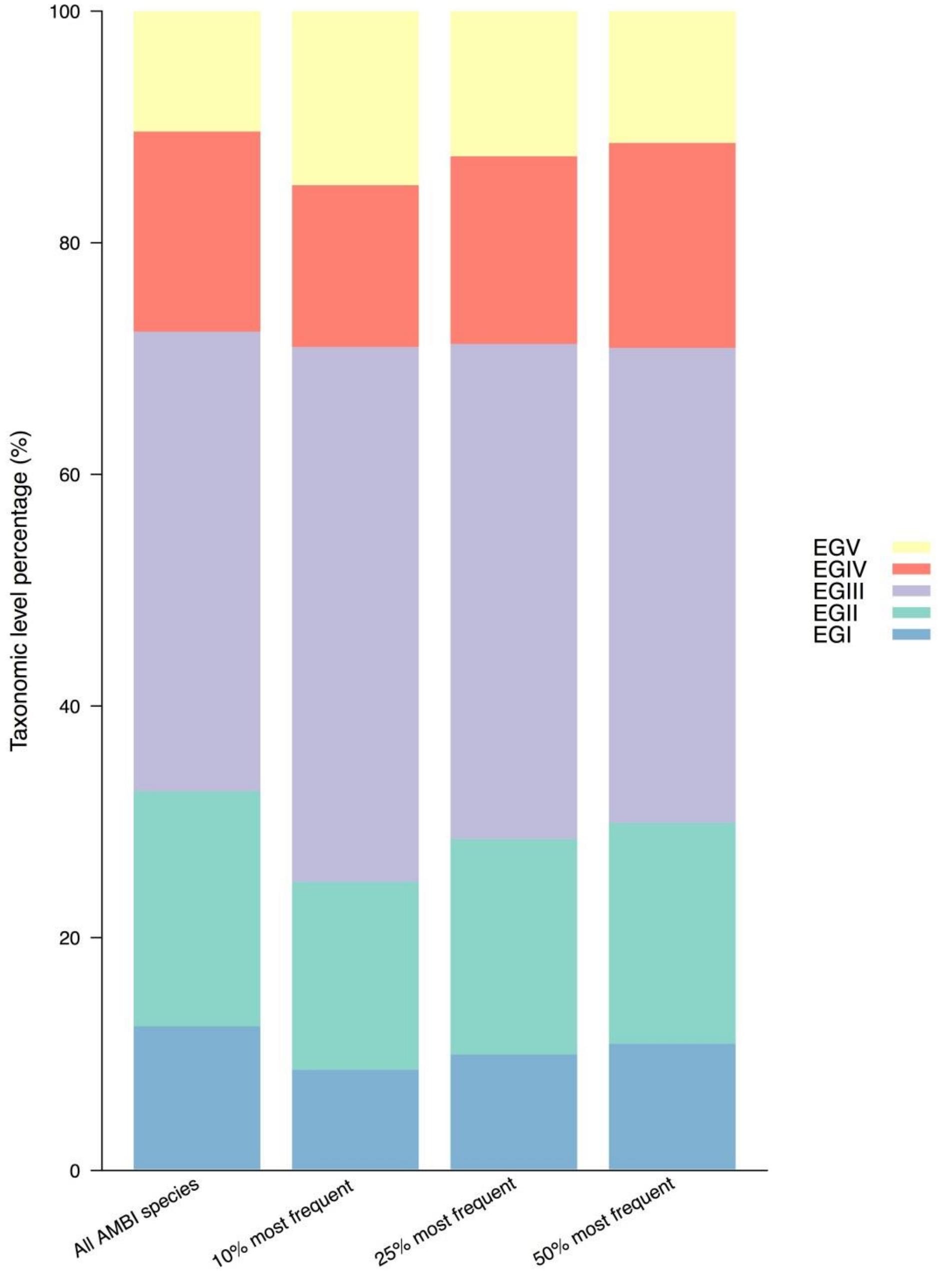

Figure S1.2. Distribution of most frequent taxa along the pollution gradient. Proportion of species, based on frequency, of each ecological group in each dataset (all species, 10% most frequent, 25% most frequent and 50% most frequent).

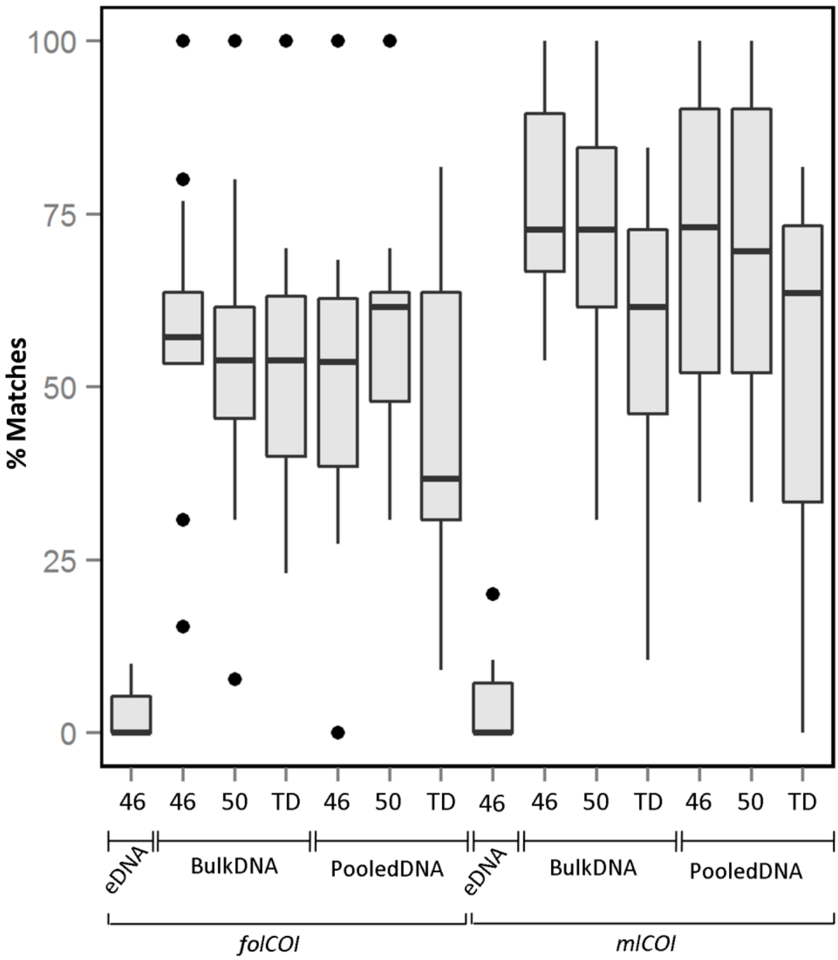

Figure S4.1. Percentage of *exact matches* (only taxa identified to species level using morphological methods) obtained between morphological and molecularly inferred taxonomic compositions over all stations for extracellular DNA (eDNA), Bulk and Pooled DNA approaches using different PCR conditions (46 or 50 ºC annealing temperatures or TD: touchdown profile) for *folCOI* and *mlCOI* barcodes. Median and error bars are depicted.

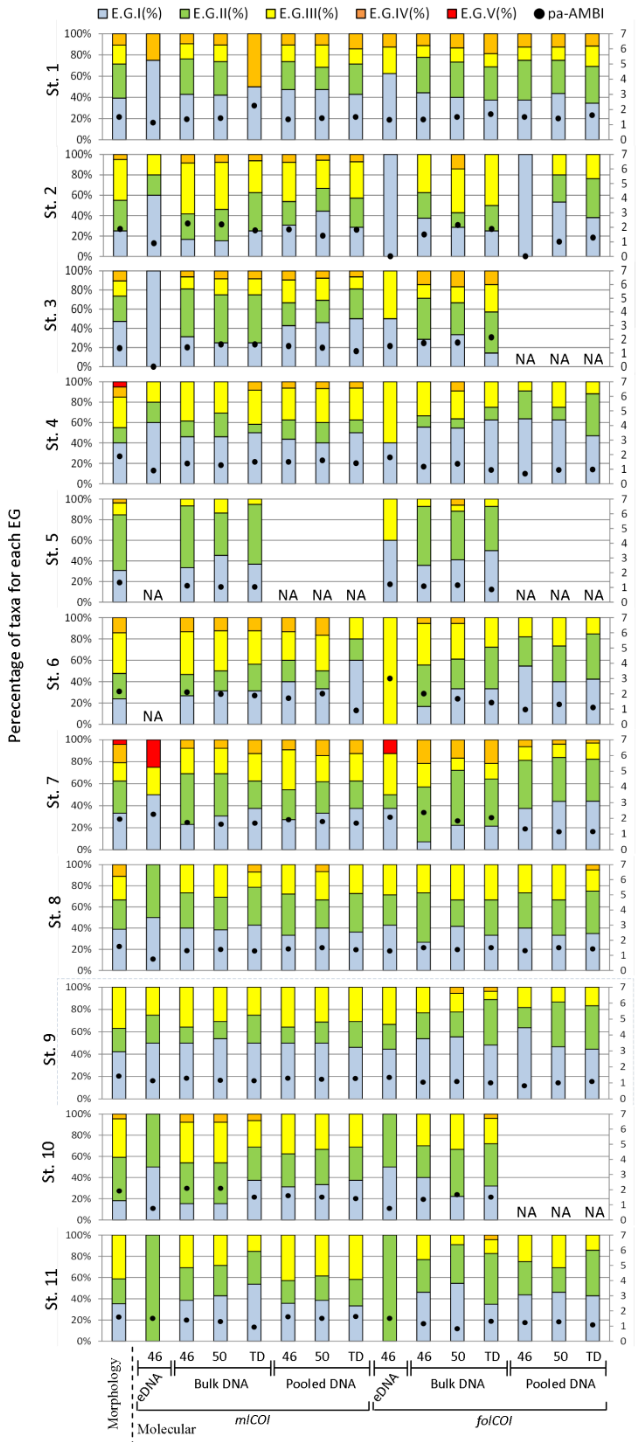

Figure S4.2. Percentage of taxa found for each Ecological Group (E.G.) and (pa)AMBI values (dot) inferred for each condition according to the taxa detected among samples using morphological methodologies and molecularly inferred taxonomies from extracellular DNA (eDNA), Bulk and Pooled DNA approaches using different PCR conditions (46 or 50 °C annealing temperatures or TD: touchdown profile) for *folCOI* and *mlCOI* barcodes. NA indicates samples that were

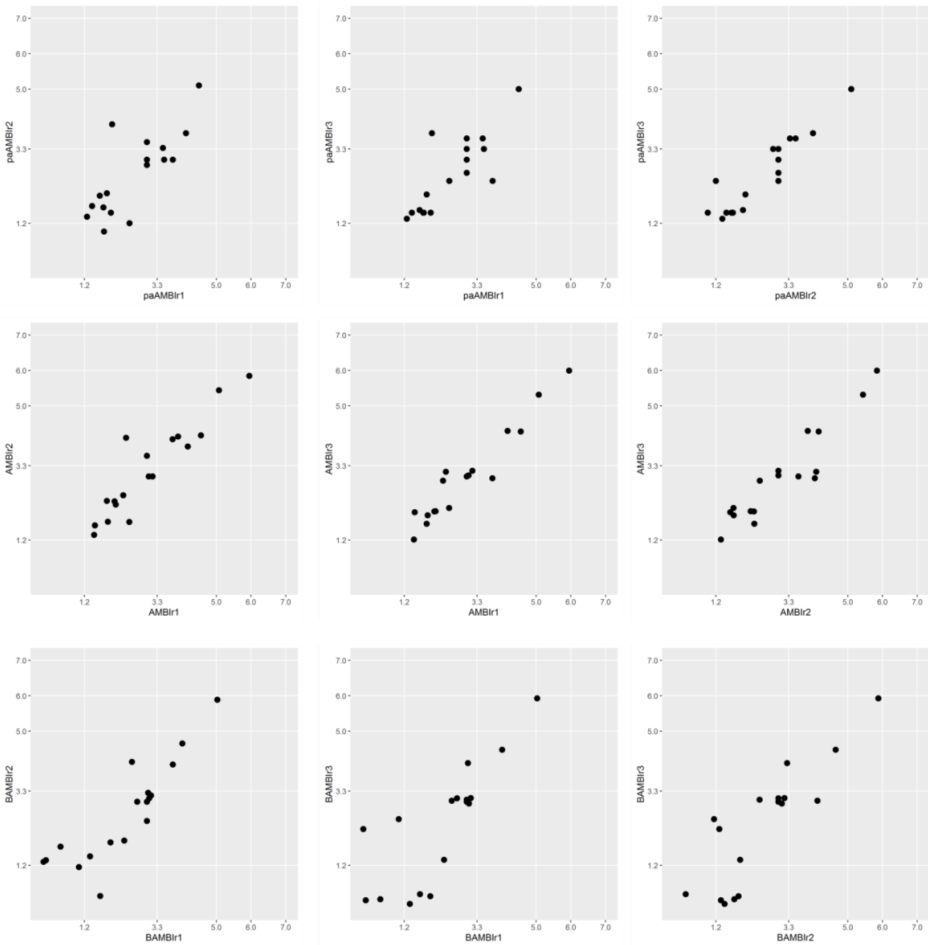

Figure S5.1. Comparisons between each form of AMBI calculated from each sample replicate. From each station, three samples were analyzed thorough morphology and AMBI, (B)AMBI and (pa)AMBI were calculated. Comparisons between replicates are shown. Vertical and horizontal lines are depicted using threshold values to discriminate disturbance classes: undisturbed [0 - 1.2], slightly disturbed [1.3 - 3.3], moderately disturbed [3.4 - 5], heavily disturbed [5.1 - 6] and extremely disturbed [6.1 – 7].

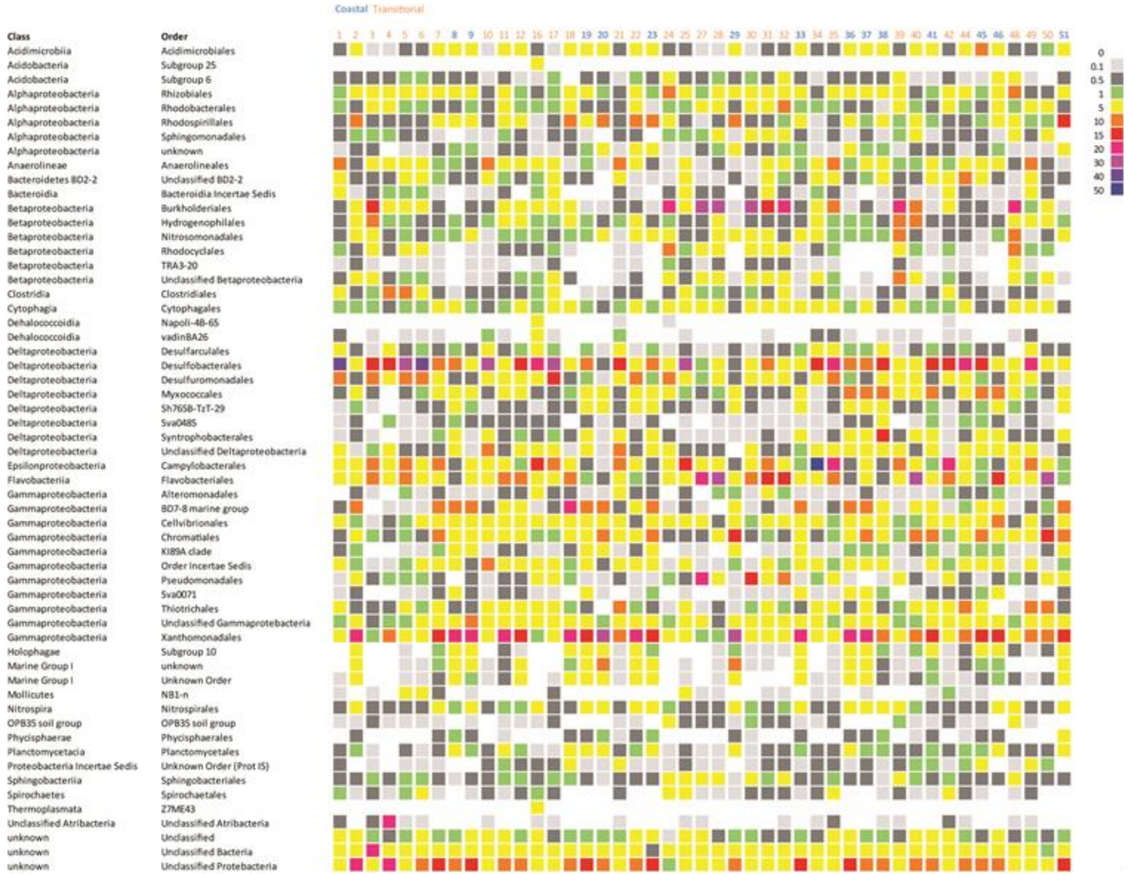

Figure S6.1. Clustering of bacterial assemblages at order level. The relative abundance of prokaryotic orders is expressed as a contribution of sequences affiliated with each order on the total number of sequences per sample, and such contribution is represented on a color scale (values reported in the upper right corner).

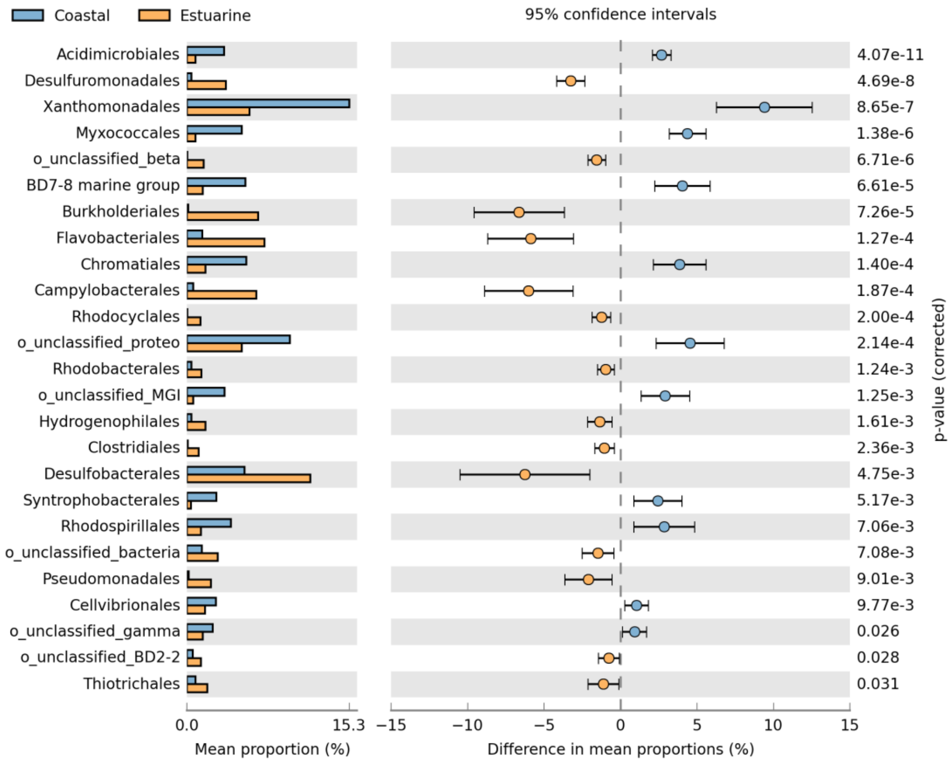

Figure S6.2. Bacterial taxa at order level contributing to significant differences (Welch's *t*-test; $p < 0.005$) between the estuarine and coastal stations.

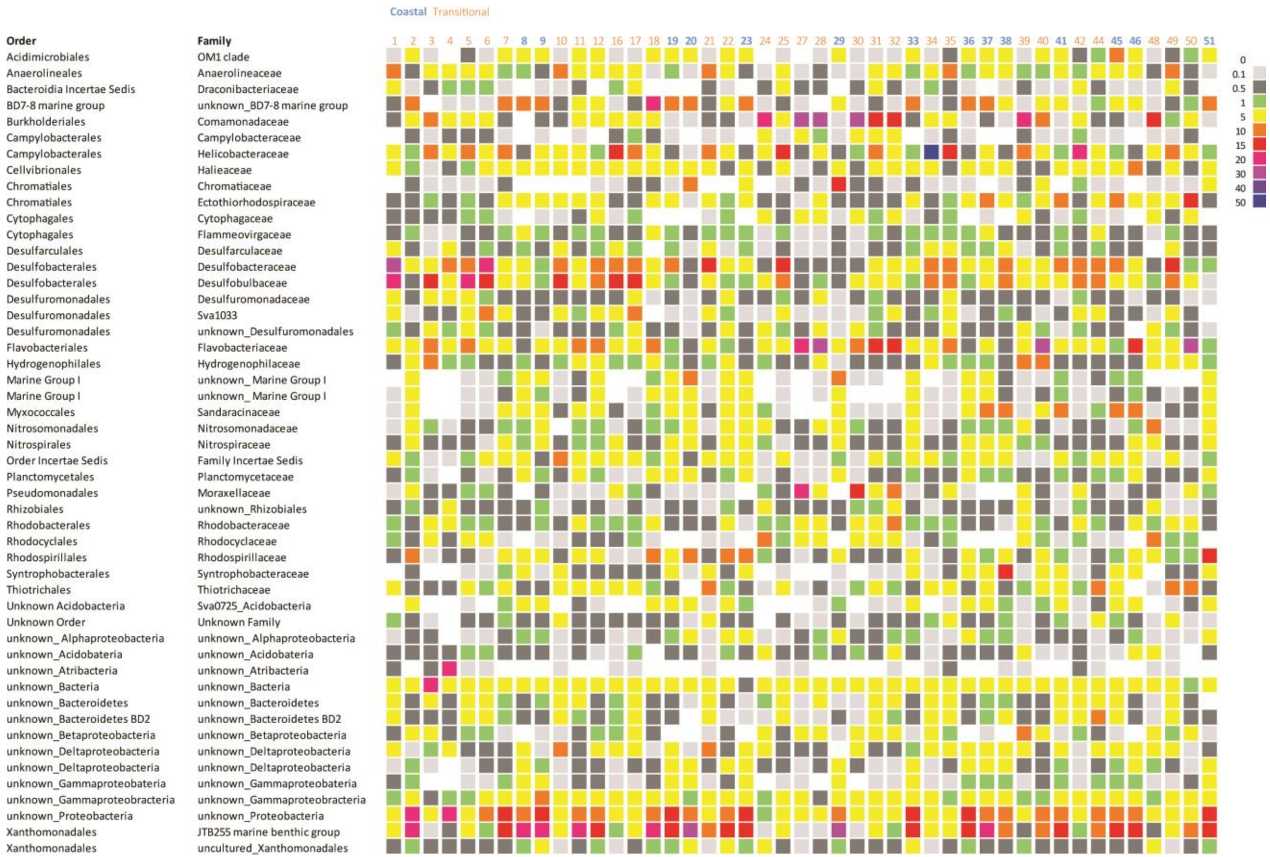

Figure S6.3. Clustering of bacterial communities at family level. The relative abundance of prokaryotic families is expressed as a contribution of sequences affiliated with each family on the total number of sequences per sample, and such contribution is represented on a color scale (values reported in the upper right corner).

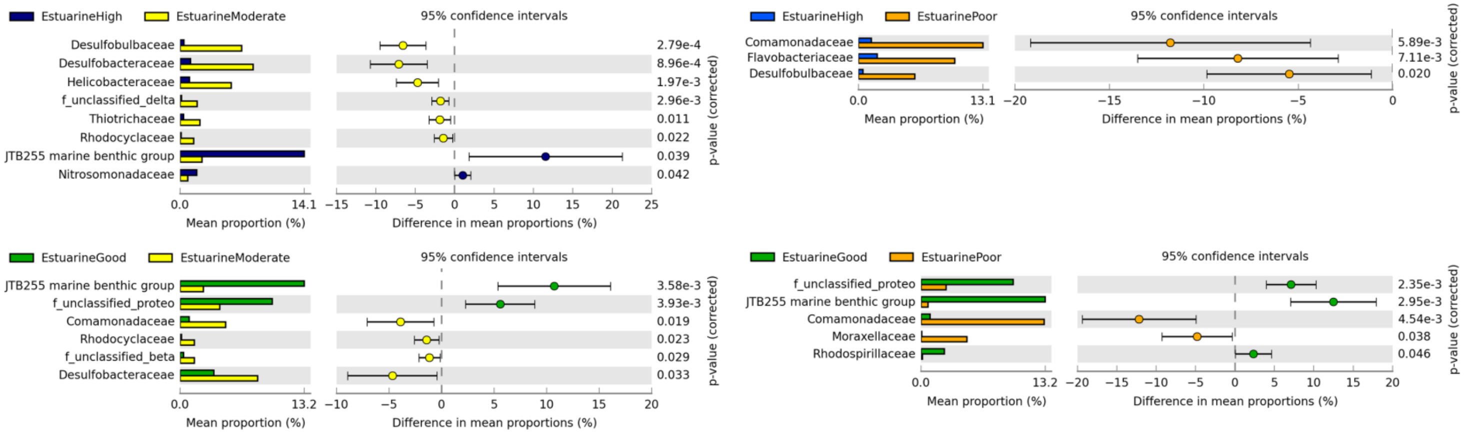

Figure S6.4: Detail of the relative contribution of the different bacterial taxa at the family level contributing to significant differences (Welch's *t*-test; $p < 0.005$) between the four classes (High, Good, Moderate and Poor) defined from the microgAMBI inferred for estuarine stations.

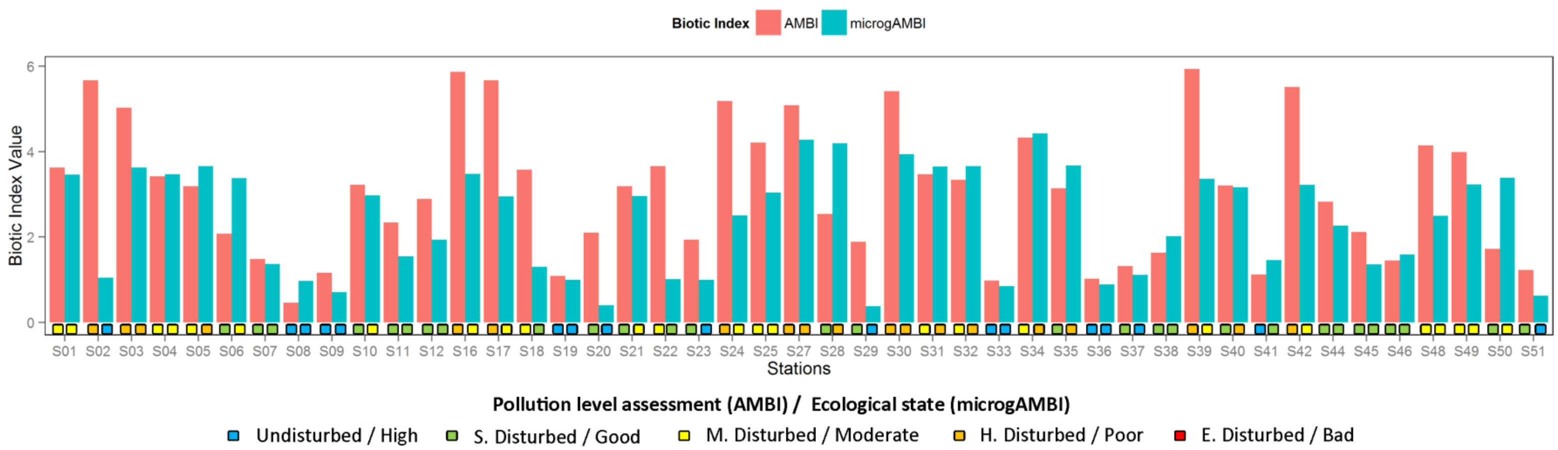

Figure S6.5: Comparison between AMBI and microgAMBI. Bars represent the obtained index values for each station (see Figure 6.1 for locations). Squared colors bellow bars show the pollution level assessment or ecological state assigned to each station considering AMBI and microgAMBI, respectively. S. Disturbed: Slightly disturbed; M. Disturbed: Moderately disturbed; H. Disturbed: Heavily disturbed, E. Disturbed: Extremely disturbed.

Table S1.1. Primer pairs tested for CO1 and 18S rRNA

Primer name	**5' – 3' Forward primer sequence**	**Target taxa**	Reference
CO1			
LCO1490	GGTCAACAAATCATAAAGATATTGG	Universal metazoa	(Folmer et al., 1994)
HC02198	TAAACTTCAGGGTGACCAAAAAATCA	Universal metazoa	(Folmer et al., 1994)
dgLCO	GGTCAACAAATCATAAAGAYATYGG	Universal Metazoa	(Meyer, 2003)
dgHCO	TAAACTTCAGGGTGACCAAARAAYCA	Universal Metazoa	(Meyer, 2003)
mlCOIintF	GGWACWGGWTGAACWGTWTAYCCYCC	Universal metazoa	(Leray et al., 2013)
mlCOIintR	GGRGGRTASACSGTTCASCCSGTSCC	Universal metazoa	(Leray et al., 2013)
Uni-MinibarF1	CAAAATCATAATGAAGGCATGAGC	Universal metazoa	(Meusnier et al., 2008)
Uni-MinibarR1	TCCACTAATCACAARGATATTGGTAC	Universal metazoa	(Meusnier et al., 2008)
Chelicerate-F1	TACTCTACTAATCATAAAGACATTGG	Arachnida, Arthropoda	(Barrett and Hebert, 2005)
Chelicerate-R1	CCTCCTCCTGAAGGGTCAAAAAATGA	Arachnida, Arthropoda	(Barrett and Hebert, 2005)
Chelicerate-R2	GGATGGCCAAAAAATCAAAATAAATG	Arachnida, Arthropoda	(Barrett and Hebert, 2005)
CrustDF1	GGTCWACAAAYCATAAAGAYATTGG	Crustacea, Arthropoda	(Radulovici et al., 2009)
CrustDR1	TAAACYTCAGGRTGACCRAARAAYCA	Crustacea, Arthropoda	(Radulovici et al., 2009)
CrustF1	TTTTCTACAAATCATAAAGACATTGG	Crustacea, Arthropoda	(Costa et al., 2007)
CrustF2	GGTTCTTCTCCACCAACCACAARGAYATHGG	Crustacea, Arthropoda	(Costa et al., 2007)
18S			
18 SA	AACCTGGTTGATCCTGCCAGT	Universal Metazoa	(Apakupakul et al., 1998)
18 SB	TGATCCTTCCGCAGGTTCACCT	Universal Metazoa	(Apakupakul et al., 1998)
SSUF04	GCTTGTAAAGATTAAGCC	Meiofauna (Nematodes)	(Blaxter et al., 1998)
SSUR22	GCCTGCTGCCTTCCTTGGA	Meiofauna (Nematodes)	(Blaxter et al., 1998)
NF1f	GGTGGTGCATGGCCGTTCTTAGTT	Meiofauna (Nematodes)	
18Sr2br	TACAAAGGGCAGGGACGTAAT	Meiofauna (Nematodes)	(Porazinska et al., 2009)
V4_F	CCAGCASCYGCGGTAATTCC	Protista	(Porazinska et al., 2009)
V4_R	ACTTTCGTTCTTGATYRA	Protista	(Guillou et al., 2012)
V9A_F	GTACACACCGCCCGTC	Protista	(Guillou et al., 2012)
V9A_R	TGATCCTTCTGCAGGTTCACCTAC	Protista	(Guillou et al., 2012)
V9B_F	TTGTACACACCGCCC	Protista	(Guillou et al., 2012)
V9B_R	CCTTCYGCAGGTTCACCTAC	Protista	(Amaral-Zettle et al., 2009)

18S1f	TACCTGGTTGATCCTGCCAGTAG	Decapoda, Arthropoda	(Amaral-Zettle et al., 2009)
18Sb2.9r	TATCTGATCGCCTTCGAACCTCT	Decapoda, Arthropoda	(Whiting, 2002)
18S5f	GCGAAAGCATTTGCCAAGAA	Decapoda, Arthropoda	(Whiting, 2002)
18S9r	GATCCTTCCGCAGGTTCACCTAC	Decapoda, Arthropoda	(Carranza et al., 1996)
MolF	GCCAGTAGCATATGCTTGTCTC	Mollusca, Bivalvia	(Carranza et al., 1996)
MolR	AGACTTGCCTCCAATGGATCC	Mollusca, Bivalvia	(Holland et al., 1991)
18eF	CTG GTT GAT CCT GCC AGT	Universal Metazoa	(Holland et al., 1991)
18rR	GTC CCC TTC CGT CAA TTY CTT TAA G	Mollusca, Bivalvia	(Hillis and Dixon, 1991)
18lR	GAA TTA CCG CGG CTG CTG GCA CC	Universal metazoa	(Passamaneck et al., 2004)
18R925D	GAT CYA AGA ATT TCA CCT CT	Annelida	(Halanych, 1998)
18F509	CCC CGT AAT TGG AAT GAG TAC A	Annelida	(Burnette et al., 2005)
18R1779	TGT TAC GAC TTT TAC TTC CTC TA	Annelida	(Struck et al., 2005)
#1F	CTGGTGCCAGCAGCCGCGGYAA	Universal Metazoa	(Struck et al., 2005)
#2F	AACTTAAAGRAATTGACGGA	Universal Metazoa	(Machida and Knowlton, 2012)
#3F	GYGGTGCATGGCCGTTSKTRGTT	Universal Metazoa	(Machida and Knowlton, 2012)
#4F	ATAACAGGTCWGTRATGCCCTYMG	Universal Metazoa	(Machida and Knowlton, 2012)
#2_RC	TCCGTCAATTYCTTTAAGTT	Universal Metazoa	(Machida and Knowlton, 2012)
#3_RC	AACYAMSAACGGCCATGCACCRC	Universal Metazoa	(Machida and Knowlton, 2012)
#4_RC	CKRAGGGCATYACWGACCTGTTAT	Universal Metazoa	(Machida and Knowlton, 2012)
#5_RC	GTGTGYACAAAGGBCAGGGAC	Universal Metazoa	(Machida and Knowlton, 2012)

Table S4.1. The 138 visually identified taxa from 11 sampling stations (St). Barcodes for taxa in bold were not available; in such cases, taxonomic level indicated below the “BOLDdb” column indicates the lowest taxonomic level found to represent that taxa in the BOLD database. Barcodes generated in this study are depicted with an asterisk and GenBank accession numbers (Acc. Num) are provided. Numbers below each station indicate number of individuals per taxa found.

Pyllum	Class	Family	Genus	Species	BOLDdb	Acc. Num	Station										
							1	2	3	4	5	6	7	8	9	10	11
ANNELIDA	Polychaeta	Acoetidae	*Panthalis*	*Panthalis oerstedi**		KT307675								1			
		Glyceridae	*Glycera*								1			1			
				Glycera alba									1				
				Glycera cf. alba							1						
				Glycera unicornis											2	1	
				Glycera cf. unicornis			1				1		1				
		Goniadidae	*Glycinde*	*Glycinde nordmanni*			1						1				
			Goniada	*Goniada cf. maculata*					1				1				
			Goniada	*Goniada maculata**		KT307646								1			
		Hesionidae	*Podarkeopsis*	*Podarkeopsis capensis**		KT307681						1					
		Nephtyidae	*Nephtys**			KT307664			1								
				*Nephtys hombergii**		KT307665							1				
				*Nephtys hystricis**		KT307666		1		2							
				*Nephtys incisa**		KT307667						1					
				Nephtys cf. incisa			1										
				*Nephtys kersivalensis**		KT307668	1						1				

		Phyllodocidae	*Phyllodoce**			KT307678	1										
				*Phyllodoce rosea**		KT307677							1				
			Pseudomystides	*Pseudomystides limbata**		KT307693										1	
		Pilargidae	*Ancistrosyllis*	*Ancistrosyllis groenlandica**		KT307624			1			1					
			Litocorsa	*Litocorsa stremma**		KT307655		1									1
			Pilargis	*Pilargis verrucosa**		KT307679							1				
		Polynoidae*				KT307684									1		
		Syllidae	*Sphaerosyllis*				1										
		Sigalionidae	*Labioleanira*	*Labioleanira yhleni**		KT307649						1		1			
			Gallardoneris	*Gallardoneris iberica**		KT307645		1			1	2					
			Ninoe	*Ninoe armoricana**		KT307669								2			1
		Poecilochaetidae	*Poecilochaetus*	*Poecilochaetus serpens**		KT307682			1								
		Sabellidae*							1								
		Sabellidae	*Euchone*	*Euchone rosea**		KT307641					1					1	
		Serpulidae	*Ditrupa*	***Ditrupa arietina***	Family								1				
		Oweniidae	*Galathowenia*	*Galathowenia oculata**		KT307644		3		2		4		1	9	5	1
		Scalibregmatidae										2					
			Scalibregma	*Scalibregma inflatum**		KT307695				1							
		Ampharetidae									1				1		
			Anobothrus	*Anobothrus gracilis*								1					
			Auchenoplax	*Auchenoplax crinita**		KT307632	1	3		1	1				1	3	
			Eclysippe	*Eclysippe vanelli*				1		2	1				1		

			Lysippe	*Lysippe labiata**		KT307657										1	
			Sosane	*Sosane sulcata**		KT307697					1				1		
		Cirratulidae	*Chaetozone*									1		1			
				Chaetozone carpenteri	Genus								2				
				Chaetozone setosa			1		3				7				
				*Chaetozone gibber**		KT307635	1		1				3				
			*Tharyx**			KT307703	1										
				*Tharyx tesselata**		KT307704						1				1	
		Flabelligeridae	*Diplocirrus*	***Diplocirrus glaucus***	Genus								1				
		Sternaspidae	*Sternaspis*	*Sternaspis scutata**		KT307702								1			
		Terebellidae	*Pista*	*Pista cristata**		KT307680				1	1						
			*Polycirrus**			KT307683		1		1							
		Trichobranchidae									1						
		Capitellidae								1			1				
			Mediomastus	***Mediomastus fragilis***	Genus							2				1	
			*Notomastus**			KT307670										2	
				Notomastus latericeus	Genus			1					1		1		
			Peresiella	***Peresiella clymenoides***	Family		1	1	1		1					1	
		Maldanidae									1						
			Chirimia	***Chirimia biceps***	Family						1					1	
			*Lumbriclymene**			KT307656				1							
				*Maldane glebifex**		KT307660					1			1			

		Maldanidae	*Petaloproctus*	***Petaloproctus terricola***	Family							1					
			*Praxillella**			KT307686		1								1	1
				*Praxillella gracilis**		KT307685						3		1	1		
		Magelonidae	*Magelona*	*Magelona lusitanica**		KT307658	1	1									
				*Magelona minuta**		KT307659	2		1								
		Spionidae*				KT307699									1		1
			*Aonides**			KT307626	1		2								
			*Laonice**			KT307650				1	1						
			Polydora							1				1			
			Prionospio	*Prionospio dubia**		KT307689										1	1
				*Prionospio ehlersi**		KT307690						4					
			Scolelepis												1		
			Spiophanes	*Spiophanes bombyx**		KT307700										1	
				*Spiophanes kroeyeri**		KT307701	1			1				1			
		Chaetopteridae						1									
			Phyllochaetopterus							1				1			
			Spiochaetopterus	*Spiochaetopterus costarum**		KT307698										1	
		Oenonidae	*Drilonereis*	*Drilonereis filum**		KT307638					2						
		Onuphidae	*Aponuphis*	***Aponuphis bilineata***	Family		1						1				
				Aponuphis fauveli	Family						1		2		1	1	
		Onuphidae						1									
			Paradiopatra	***Paradiopatra calliopae***	Genus						2	4		6	4	3	

			Paradiopatra	*Paradiopatra quadricuspis*													1
		Eunicidae	*Eunice*	*Eunice vittata**		KT307642					1					1	
			Marphysa	*Marphysa bellii**		KT307661											1
		Opheliidae	*Ophelina*	*Ophelina cylindricaudata**		KT307672					1	1					
		Paraonidae	*Aricidea**			KT307630									1	1	
				*Aricidea laubieri**		KT307628	1										
				*Aricidea mirunekoa**		KT307629											2
				*Aricidea wassi**		KT307631	1		2								
			Cirrophorus	*Cirrophorus branchiatus**		KT307637									1		
			Levinsenia	*Levinsenia gracilis**		KT307653	1	1									1
				*Levinsenia kantauriensis**		KT307654						3					
		Paraonidae	*Paradoneis*	*Paradoneis ilvana**		KT307676							2				
ARTHROPODA	Malacostraca	Alpheidae	*Alpheus*	***Alpheus glaber***	Genus											1	
		Callianassidae	*Callianassa*	***Callianassa subterranea***	Genus						2	1	1		1		
		Galatheidae	*Galathea**			KT307643	1										
		Goneplacidae	*Goneplax*	*Goneplax rhomboides*						1				1			
		Leucosiidae	*Ebalia**			KT307639					1						
		Processidae	*Processa*	*Processa nouveli**		KT307692	1										
		Paguridae	*Pagurus**			KT307674					1						
		Ampeliscidae	*Ampelisca*	*Ampelisca pectenata**		KT307620			1				1				
				*Ampelisca provincialis**		KT307621					1						
				*Ampelisca spinipes**		KT307622								1			

				*Ampelisca typica**		KT307623			1								
			Byblis	***Byblis guernei***	Genus			1									
		Leucothoidae	*Leucothoe*	*Leucothoe lilljeborgi**		KT307651									1		
		Pardaliscidae	*Halice*	*Halice walkeri**		KT307647	1										
			Nicippe	***Nicippe tumida***	Family			1		1							
		Unciolidae	*Unciolella*	*Unciolella lunata**		KT307706											2
		Cheirocratidae	*Cheirocratus*	*Cheirocratus intermedius**		KT307636				1							
		Arcturidae	*Arcturopsi*	*Arcturopsis giardi**		KT307627											1
		Cirolanidae	*Natatolana*	*Natatolana borealis**		KT307662	1		1								
		Bodotriidae	*Iphinoe*	*Iphinoe serrata**		KT307648	1		3				5		1		
		Diastylidae	*Vemakylindrus*	*Vemakylindrus cantabricus**		KT307707									1		
MOLLUSCA	Bivalvia	Cuspidariidae	*Cuspidaria*	*Cuspidaria cuspidata*											1		
		Nuculidae	*Nucula*	*Nucula sulcata*						1					1		
			Pronucula	***Pronucula tenuis***	Genus		1		2								
		Semelidae	*Abra*	*Abra alba**		KT307619		1									
				Abra nitida	Genus												1
		Thyasiridae	*Axinulus*	*Axinulus croulinensis**		KT307633							3				
			Mendicula	***Mendicula ferruginosa***	Family			1									1
			Thyasira	*Thyasira flexuosa**		KT307705		1									
		Veneridae	*Timoclea*	*Timoclea ovata*			1										
	Gastropoda	Cylichnidae	*Cylichna*	***Cylichna cylindracea***	Genus			1									
	Scaphopoda	Dentaliidae	***Entalis***		Family												2

ECHINODERMATA	Echinoidea	Brissidae	*Brissopsis*	*Brissopsis lyrifera**		KT307634											1
		Loveniidae	*Echinocardium**			KT307640				1							
	Holothuroidea								1								
	Ophiuroidea	Amphiuridae	*Amphiura*	***Amphiura cf. filiformis***	Genus						1						
		Ophiuridae	*Ophiura*	*Ophiura sp.*						1							
				*Ophiura texturata**		KT307673			1								
CNIDARIA	Anthozoa*					KT307625						1		1			
NEMATODA*						KT307663										3	
NEMERTEA							1			1		4	1				
SIPUNCULA	Sipunculidea	Phascolionidae	*Onchnesoma*	*Onchnesoma steenstrupii**		KT307671	2		1			1				2	1
		Sipunculidae	*Sipunculus*	*Sipunculus nudus**		KT307696					1		1				
CHAETOGNATA	Sagittoidea	Sagittidae	*Sagitta**			KT307694	1					1		1			

Table S4.2. Number of reads obtained per each DNA source and sampling station (St) after quality filtering for *mlCOI* and *folCOI* barcodes. Replicates are labeled with A, B and C. NA indicates samples that were not sequenced for failing at the amplification step.

mlCOI

DNA source	PCR Condition	St.1	St.2	St.3	St.4	St.5	St.6	Site7			St.8	St.9			St.10	St.11		
								A	B	C		A	B	C		A	B	C
eDNA	46 °C	38,808	68,726	71,436	79,977	NA	NA	106,096	149,491	85,812	105,986	86,044	94,284	126,359	108,831	95,252	88,600	156,341
Bulk DNA	46 °C	74,591	79,458	22,264	119,653	89,198	74,240	46,088	3,026	29,737	67,641	15,772	27,378	49,505	75,917	38,043	29,011	58,034
	50 °C	56,756	49,687	75,743	67,275	70,700	51,135	39,223	55,215	11,652	50,256	31,436	31,663	1,438	58,722	56,811	28,320	73,988
	TD	38	37,620	26,687	81,904	25,042	23,494	75,411	52,635	11,604	5,071	5,431	32,924	57,140	9,469	1,405	5,667	42,547
Pooled DNA	46 °C	64,900	65,598	41,021	72,418	NA	243,110	69,759	54,768	62,222	23,155	53,497	33,423	30,640	62,856	47,680	42,505	81,449
	50 °C	56,366	53,403	52,619	28,227	NA	72,845	45,190	30,603	85,674	57,565	57,649	35,624	94,808	49,888	62,455	12,760	43,203
	TD	85,616	62,229	45,766	13,802	NA	1,474	38,162	90,117	23,328	3,570	35,611	27,530	30,382	31,598	41,244	70,967	58,570

folCOI

DNA source	PCR Condition	St.1	St.2	St.3	St.4	St.5	St.6	St.7			St.8	St.9			St.10	St.11		
								A	B	C		A	B	C		A	B	C
eDNA	46 °C	68,022	89,970	37,244	59,588	43,107	51,420	53,203	43,181	75,954	70,245	94,632	68,243	81,134	48,016	14,844	41,485	37,612
Bulk DNA	46 °C	19,487	39,088	41,299	23,895	40,461	40,850	18,020	17,901	8,657	38,210	7,920	96,797	2,202	65,942	46,473	19,591	11,201
	50 °C	33,415	26,770	22,142	28,241	12,620	31,331	11,281	2,642	11,542	19,206	51,175	170	28,748	36,674	32,668	18,632	23,662
	TD	17,941	41,694	39,625	27,864	46,089	41,110	4,364	2,251	2,226	30,427	3,675	35,945	4,723	37,018	30,205	19,039	37,042
Pooled DNA	46 °C	31,336	21	NA	38,891	NA	35,828	12,509	4,544	3,798	48,567	27,873	51,667	35,490	NA	35,303	33,393	18,031
	50 °C	47,127	24,441	NA	48,777	NA	40,125	5,249	7,553	3,883	30,046	49,083	20,052	39,316	NA	48,332	18,206	12,298
	TD	19,467	38,300	NA	2,838	NA	9,702	7,889	3,081	1,976	31,642	19,656	20,737	40,877	NA	141	16,958	49,612

Table S4.3. Percentage of matches for each condition for the 9 visually identified phyla. Numbers in brackets represent total taxa found over all stations for each phylum.

Sample	Sediment		Complete specimens											
DNA method	eDNA		Bulk DNA						Pooled DNA					
Barcode	*mlCOI*	*folCOI*	*mlCOI*			*folCOI*			*mlCOI*			*folCOI*		
PCR Condition	46 °C	46 °C	46 °C	50 °C	TD	46 °C	50 °C	TD	46 °C	50°C	TD	46 °C	50 °C	TD
Mollusca (11)	11.5	11.5	78.8	78.8	60.2	21.1	23	19.2	58.3	48	25	4.1	2.7	0
Echinodermata (6)	0	0	66.6	66.6	50	0	0	0	90	90	90	0	0	12.5
Arthropoda (21)	0	0	30.6	29.2	28.6	32.1	28	28	25.1	27.9	25.1	29	30.4	24.2
Annelida (94)	4.1	3	67	63.4	53.8	54.8	50.9	52.5	64.7	62.4	53.1	48.2	53.3	48.9
Chaetognata (1)	0	0	66.6	66.6	33.3	0	0	0	100	33.3	0	0	0	0
Cnidaria (1)	0	0	100	100	50	50	50	50	100	100	0	100	100	100
Nematoda (1)	0	0	100	100	100	0	0	0	100	100	100	NA	NA	NA
Nemertea (1)	20	16.6	83.3	83.3	16.6	66.6	66.6	50	100	100	83.3	0	16.6	16.6
Sipuncula (2)	0	0	45.4	45.4	27.2	18.1	9	9	70	80	20	37.5	37.5	12.5

Table S6.1. Ecological role associated with prokaryotic taxa according to the available literature and potential relationship with pollution inputs.

Phyllum	Class	Order	Family	Genus	Role / indicator of	References
Proteobacteria	δ-Proteobacteria	Desulfobacterales	Desulfobacteraceae		SRB / Enriched OM sediment	(Miyatake et al., 2009; Elisabé et al., 2012)
Proteobacteria	δ-Proteobacteria	Desulfobacterales	Desulfobulbaceae		SRB / Enriched OM sediment	(Elisabé et al., 2012; Aranda et al., 2015)
Proteobacteria	δ-Proteobacteria	Desulfobacterales	Desulfarculaceae		SRB / Enriched OM sediment	(Elisabé et al., 2012)
Proteobacteria	δ-Proteobacteria	Desulfobacterales	Nitrospinaceae		AOB / Enriched OM sediment	(Ionescu et al., 2012)
Proteobacteria	δ-Proteobacteria	Desulfuromonadales	Desulfuromonadaceae		SRB / Enriched OM sediment	(Elisabé et al., 2012; Aranda et al., 2015)
Proteobacteria	δ-Proteobacteria	Desulfuromonadales	Geobacteraceae		SRB / Enriched OM sediment	(Holmes et al., 2002)
Proteobacteria	γ-Proteobacteria	Thiotrichales	Thiotrichaceae		SOB / Organic pollution	(Campbell et al., 2015)
Proteobacteria	γ-Proteobacteria	Thiotrichales	Piscirickettsiaceae		SOB / Organic pollution	(Zhang et al., 2016)
Atribacteria	Unkown	Unkown	Unkown		Anoxic methane-rich sediments	(Carr et al., 2015)
Proteobacteria	δ-Proteobacteria	Syntrophobacterales	Syntrophobacteraceae		SRB / Enriched OM sediment	(Muyzer and Stams, 2008; Zhang et al., 2008)
Proteobacteria	δ-Proteobacteria	Syntrophobacterales	Syntrophaceae		SRB / Enriched OM sediment	(Muyzer and Stams, 2008)
Proteobacteria	β-Proteobacteria	Nitrosomonadales	Nitrosomonadaceae		AOB / Organic enrichment	(Dang et al., 2010)
Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae		Nitrite oxidizer / Nitrogen input	(Dang et al., 2010)
Proteobacteria	β-Proteobacteria	Nitrosomonadales	Gallionellaceae		AOB / Organic enrichment	(Dang et al., 2010)
Proteobacteria	β-Proteobacteria	Rhodocyclales	Rhodocyclaceae		WTP	(Shchegolkova et al., 2016)

Proteobacteria	β-Proteobacteria	Hydrogenophilales	Hydrogenophilaceae		SOB / sulfide-rich wastewater	(Luo et al., 2011)
Proteobacteria	α-Proteobacteria	Rhodobacterales	Rhodobacteraceae		AS	(Ju and Zhang, 2015)
Chloroflexi	Anaerolineae	Anaerolineales	Anaerolineaceae		Methanogenic degradation of alkanes/ AS	(Ju and Zhang, 2015; Liang et al., 2015)
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae [1, 2]		AS	(Shchegolkova et al., 2016)
Proteobacteria	ε-Proteobacteria	Campylobacterales	Helicobacteraceae [1, 3, 4]		MWWTP	(Shchegolkova et al., 2016)
Proteobacteria	β-Proteobacteria	Burkholderiales	Comamonadaceae		Aromatic compounds biodegradation / WTP treating MW with petroleum products	(Perez-Pantoja et al., 2012; Shchegolkova et al., 2016)
Proteobacteria	β-Proteobacteria	Burkholderiales	Alcaligenaceae		Aromatic compounds biodegradation	(Perez-Pantoja et al., 2012)
Proteobacteria	β-Proteobacteria	Burkholderiales	Oxalobacteraceae		Aromatic compounds biodegradation	(Perez-Pantoja et al., 2012)
Proteobacteria	γ-Proteobacteria	Pseudomonadales	Pseudomonadaceae		WTP treating MW with petroleum products	(Shchegolkova et al., 2016)
Proteobacteria	ε-Proteobacteria	Campylobacterales	Helicobacteraceae	*Sulfuricurvum*	SOB / MW, AS or groundwater contaminated with petroleum	(Campbell et al., 2006; Haaijer et al., 2008)
Actinobacteria	Actinobacteria	Corynebacteriales	Mycobacteriaceae	*Mycobacterium*	Benzo[α]pyrene degraders	(Kappell et al., 2014)
Actinobacteria	Actinobacteria	Micrococcales	Micrococcaceae	*Arthrobacter*	Versatile aromatic hydrocarbon-degrader	(Jiang et al., 2015)
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	*Flavobacterium*	Benzo[α]pyrene degraders	(Kappell et al., 2014)

Proteobacteria	γ-Proteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*	Hydrocarbon-degrader / Incoming sewage	(Fondi et al., 2012; Shchegolkova et al., 2016)
Firmicutes	Bacilli	Bacillales	Bacillaceae	*Bacillus*	Hydrocarbon-degrader	(Fathepure, 2014)
Proteobacteria	α-Proteobacteria	Sphingomonadales	Sphingomonadaceae	*Novosphingobium*	Hydrocarbon-degrader	(Sohn et al., 2004)
Proteobacteria	δ-Proteobacteria	Desulfobacterales	Desulfobacteraceae	*Desulfococcus*	Hydrocarbon-degrader	(Kleindienst et al., 2014)
Proteobacteria	γ-Proteobacteria	Alteromonadales	Pseudoalteromonadaceae	*Pseudoalteromonas*	Hydrocarbon-degrader	(Hedlund and Staley, 2006)
Proteobacteria	γ-Proteobacteria	Alteromonadales	Shewanellaceae	*Shewanella*	Hydrocarbon-degrader	(Motoigi and Okuyama, 2011)
Proteobacteria	γ-Proteobacteria	Oceanospirillales	Oceanospirillaceae	*Neptunomonas*	Hydrocarbon-degrader	(Hedlun et al., 1999)
Proteobacteria	γ-Proteobacteria	Oceanospirillales	Oceanospirillaceae	*Oleispira*	Hydrocarbon-degrader	(Yakimov et al., 2003)
Proteobacteria	γ-Proteobacteria	Oceanospirillales	Oleiphilaceae	*Oleiphilus*	Hydrocarbon-degrader	(Yakimov et al., 2003)
Proteobacteria	γ-Proteobacteria	Pseudomonadales	Moraxellaceae	*Alkanindiges*	Hydrocarbon-degrader	(Bogan et al., 2003)
Proteobacteria	γ-Proteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	Benzo[α]pyrene degrader	(Kappell et al., 2014)
Proteobacteria	γ-Proteobacteria	Vibrionales	Vibrionaceae	*Vibrio*	Hydrocarbon-degrader	(Hedlund and Staley, 2001)
Proteobacteria	β-Proteobacteria	Burkholderiales	Comamonadaceae	*Burkholderia*	benzo[α]pyrene degrader	(Kappell et al., 2014)
Proteobacteria	ε-Proteobacteria	Campylobacterales	Campylobacteraceae	*Arcobacter* [4]	SOB / MW with petroleum products / Potential Pathogen	(Lehner et al., 2005; Aranda et al., 2015; Shchegolkova et al., 2016)
Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	*Paludibacter* [4]	Potential Pathogen	(Thomas et al., 2011)
Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Bifidobacterium* [4]	Potential	(Moubareck et al., 2005)

					Pathogen	
Firmicutes	Clostridia	Clostridiales [4]			Potential Pathogen	(Paredes et al., 2005)
Firmicutes	Clostridia	Clostridiales	Family XII	*Fusibacter* [4]	Potential Pathogen	(Paredes et al., 2005)
Firmicutes	Clostridia	Clostridiales	Defluviitaleaceae [4]		Potential Pathogen	(Paredes et al., 2005)

Table S6.2. List of prokaryotic taxa found in the 45 estuarine and coastal sampled stations with the ecological group (EG) assigned for this study. Taxonomy: Domain;Phylum;Class;Order;Family.

Taxonomy	EG	Taxonomy	EG
Archaea;Euryarchaeota;Thermoplasmata;Z7ME43;unknown	NA	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;MNG7	I
Archaea;Thaumarchaeota;Marine Group I;unknown;unknown	NA	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae	I
Archaea;Thaumarchaeota;Marine Group I;Unknown Order;Unknown Family	NA	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae	I
Archaea;Thaumarchaeota;Soil Crenarchaeotic Group(SCG);unknown;unknown	NA	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhodobiaceae	I
Bacteria;Acidobacteria;Acidobacteria;Subgroup 17;unknown	I	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;unknown	I
Bacteria;Acidobacteria;Acidobacteria;Subgroup 18;unknown	I	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Xanthobacteraceae	I
Bacteria;Acidobacteria;Acidobacteria;Subgroup 21;unknown	I	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae	III
Bacteria;Acidobacteria;Acidobacteria;Subgroup 25;unknown	I	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;MND8	I
Bacteria;Acidobacteria;Acidobacteria;Subgroup 4;Unknown	I	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;MSB-1E8	I
Bacteria;Acidobacteria;Acidobacteria;Subgroup 6;unknown	I	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae	I
Bacteria;Acidobacteria;Acidobacteria;Subgroup 9;unknown	I	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillales Incertae Sedis	I
Bacteria;Acidobacteria;Acidobacteria;unknown;unknown	I	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;unknown	I
Bacteria;Acidobacteria;Holophagae;Holophagales;Holophagaceae	I	Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;AKIW1012	I
Bacteria;Acidobacteria;Holophagae;Subgroup 10;ABS-19	I	Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;TK34	I
Bacteria;Acidobacteria;Holophagae;Subgroup 10;Sva0725	I	Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadac eae	III
Bacteria;Acidobacteria;Holophagae;Subgroup 23;unknown	I	Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;unknown	N A
Bacteria;Acidobacteria;Subgroup 22;unknown;unknown	I	Bacteria;Proteobacteria;Alphaproteobacteria;unknown;unknown	I
Bacteria;Acidobacteria;unknown;unknown;unknown	I	Bacteria;Proteobacteria;Betaproteobacteria;B1-7BS;unknown	I

Bacteria;Actinobacteria;Acidimicrobiia;Acidimicrobiales;Acidimicrobiaceae	I	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Alcaligenaceae	III
Bacteria;Actinobacteria;Acidimicrobiia;Acidimicrobiales;OM1 clade	I	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae	III
Bacteria;Actinobacteria;Acidimicrobiia;Acidimicrobiales;Sva0996 marine group	I	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae	III
Bacteria;Actinobacteria;Acidimicrobiia;Acidimicrobiales;uncultured	I	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;unknown	NA
Bacteria;Actinobacteria;Actinobacteria;Bifidobacteriales;Bifidobacteriaceae	III	Bacteria;Proteobacteria;Betaproteobacteria;Hydrogenophilales;Hydrogenophilaceae	III
Bacteria;Actinobacteria;Actinobacteria;Corynebacteriales;Mycobacteriaceae	III	Bacteria;Proteobacteria;Betaproteobacteria;Methylophilales;Methylophilaceae	I
Bacteria;Actinobacteria;Actinobacteria;Frankiales;Geodermatophilaceae	I	Bacteria;Proteobacteria;Betaproteobacteria;Nitrosomonadales;Gallionellaceae	III
Bacteria;Actinobacteria;Actinobacteria;Kineosporiales;Kineosporiaceae	I	Bacteria;Proteobacteria;Betaproteobacteria;Nitrosomonadales;Nitrosomonadaceae	III
Bacteria;Actinobacteria;Actinobacteria;Micrococcales;Intrasporangiaceae	I	Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae	III
Bacteria;Actinobacteria;Actinobacteria;Micrococcales;Microbacteriaceae	I	Bacteria;Proteobacteria;Betaproteobacteria;SC-I-84;unknown	I
Bacteria;Actinobacteria;Actinobacteria;Micrococcales;Micrococcaceae	III	Bacteria;Proteobacteria;Betaproteobacteria;TRA3-20;unknown	I
Bacteria;Actinobacteria;Actinobacteria;Micromonosporales;Micromonosporaceae	I	Bacteria;Proteobacteria;Betaproteobacteria;unknown;unknown	I
Bacteria;Actinobacteria;Thermoleophilia;Gaiellales;Gaiellaceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;43F-1404R;unknown	I
Bacteria;Actinobacteria;Thermoleophilia;Gaiellales;unknown	I	Bacteria;Proteobacteria;Deltaproteobacteria;Bdellovibrionales;Bacteriovoracaceae	I
Bacteria;Actinobacteria;Thermoleophilia;unknown;unknown	I	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfarculales;Desulfarculaceae	III
Bacteria;Atribacteria;unknown;unknown;Atribacteria	III	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobacteraceae	III
Bacteria;Bacteroidetes;Bacteroidetes BD2-2;unknown;unknown	I	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae	III
Bacteria;Bacteroidetes;Bacteroidetes vadinHA17;unknown;unknown	I	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfobacterales;Nitrospinaceae	III
Bacteria;Bacteroidetes;Bacteroidetes VC2.1 Bac22;unknown;unknown	I	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;BVA18	I
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Desulfuromonadaceae	III

Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Marinilabiaceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae	III
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae	III	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;M20-Pitesti	I
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Sva1033	I
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;unknown	NA
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;unknown	I	Bacteria;Proteobacteria;Deltaproteobacteria;GR-WP33-30;unknown	I
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidia Incertae Sedis;Draconibacteriaceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;BIrii41	I
Bacteria;Bacteroidetes;Cytophagia;Cytophagales;Cyclobacteriaceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;Cystobacteraceae	I
Bacteria;Bacteroidetes;Cytophagia;Cytophagales;Cytophagaceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;MidBa8	I
Bacteria;Bacteroidetes;Cytophagia;Cytophagales;Flammeovirgaceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;Phaselicystidaceae	I
Bacteria;Bacteroidetes;Cytophagia;Order II;Rhodothermaceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;Sandaracinaceae	I
Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Cryomorphaceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;unknown	I
Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Flavobacteriaceae	III	Bacteria;Proteobacteria;Deltaproteobacteria;SAR324 clade(Marine group B);unknown	I
Bacteria;Bacteroidetes;SB-5;unknown;unknown	I	Bacteria;Proteobacteria;Deltaproteobacteria;Sh765B-TzT-29;unknown	I
Bacteria;Bacteroidetes;Sphingobacteriia;Sphingobacteriales;Chitinophagaceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;Sva0485;unknown	I
Bacteria;Bacteroidetes;Sphingobacteriia;Sphingobacteriales;env.OPS 17	I	Bacteria;Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophaceae	III
Bacteria;Bacteroidetes;Sphingobacteriia;Sphingobacteriales;NS11-12 marine group	I	Bacteria;Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophobacteraceae	III
Bacteria;Bacteroidetes;Sphingobacteriia;Sphingobacteriales;Saprospiraceae	I	Bacteria;Proteobacteria;Deltaproteobacteria;unknown;unknown	NA
Bacteria;Bacteroidetes;Sphingobacteriia;Sphingobacteriales;unknown	I	Bacteria;Proteobacteria;Epsilonproteobacteria;Campylobacterales;Campylobacteraceae	III
Bacteria;Bacteroidetes;Sphingobacteriia;Sphingobacteriales;WCHB1-69	I	Bacteria;Proteobacteria;Epsilonproteobacteria;Campylobacterales;Helicobacteraceae	III
Bacteria;Bacteroidetes;unknown;unknown;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;1013-28-CG33;unknown	I

Bacteria;Bacteroidetes;WCHB1-32;unknown;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Aeromonadales;Aeromonadaceae	I
Bacteria;Chlorobi;Ignavibacteria;Ignavibacteriales;Ignavibacteriaceae	I	Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Alteromonadaceae	I
Bacteria;Chlorobi;Ignavibacteria;Ignavibacteriales;IheB3-7	I	Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Colwelliaceae	I
Bacteria;Chlorobi;Ignavibacteria;Ignavibacteriales;PHOS-HE36	I	Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Pseudoalteromonadaceae	III
Bacteria;Chlorobi;Ignavibacteria;Ignavibacteriales;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Psychromonadaceae	I
Bacteria;Chloroflexi;Anaerolineae;Anaerolineales;Anaerolineaceae	III	Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Shewanellaceae	III
Bacteria;Chloroflexi;Ardenticatenia;uncultured;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Arenicellales;Arenicellaceae	I
Bacteria;Chloroflexi;Dehalococcoidia;Napoli-4B-65;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;BD7-8 marine group;unknown	I
Bacteria;Chloroflexi;Dehalococcoidia;vadinBA26;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Cellvibrionales;Cellvibrionaceae	I
Bacteria;Chloroflexi;KD4-96;unknown;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Cellvibrionales;Halieaceae	I
Bacteria;Chloroflexi;MSB-5B2;unknown;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Cellvibrionales;Porticoccaceae	I
Bacteria;Chloroflexi;S085;unknown;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Cellvibrionales;Spongiibacteraceae	I
Bacteria;Cyanobacteria;Cyanobacteria;SubsectionII;FamilyII	I	Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Chromatiaceae	I
Bacteria;Cyanobacteria;Cyanobacteria;SubsectionIV;FamilyII	I	Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Ectothiorhodospiraceae	I
Bacteria;Cyanobacteria;Cyanobacteria;uncultured;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Granulosicoccaceae	I
Bacteria;Deferribacteres;Deferribacteres Incertae Sedis;Unknown Order;Unknown Family	I	Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;unknown	I
Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae	III	Bacteria;Proteobacteria;Gammaproteobacteria;CS-B046;unknown	I
Bacteria;Firmicutes;Bacilli;Lactobacillales;Carnobacteriaceae	I	Bacteria;Proteobacteria;Gammaproteobacteria;E01-9C-26 marine group;unknown	I
Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae	I	Bacteria;Proteobacteria;Gammaproteobacteria;Gammaproteobacteria Incertae Sedis;unknown	I
Bacteria;Firmicutes;Bacilli;unknown;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;KI89A clade;unknown	I

Bacteria;Firmicutes;Clostridia;Clostridiales;Defluviitaleaceae	III	Bacteria;Proteobacteria;Gammaproteobacteria;Methylococcales;Crenotrichaceae	I
Bacteria;Firmicutes;Clostridia;Clostridiales;Family XII	III	Bacteria;Proteobacteria;Gammaproteobacteria;Methylococcales;unknown	I
Bacteria;Firmicutes;Clostridia;Clostridiales;Family XIII	NA	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Oceanospirillacea e	III
Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae	NA	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Oceanospirillales Incertae Sedis	I
Bacteria;Firmicutes;Clostridia;Clostridiales;Peptostreptococcaceae	NA	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Oleiphilaceae	III
Bacteria;Firmicutes;Clostridia;Clostridiales;Ruminococcaceae	NA	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;OM182 clade	I
Bacteria;Firmicutes;Clostridia;Clostridiales;unknown	NA	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;ORI-860-26	I
Bacteria;Firmicutes;unknown;unknown;unknown	NA	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;SS1-B-06-26	I
Bacteria;Fusobacteria;Fusobacteriia;Fusobacteriales;Fusobacteriaceae	I	Bacteria;Proteobacteria;Gammaproteobacteria;Order Incertae Sedis;Family Incertae Sedis	I
Bacteria;Fusobacteria;Fusobacteriia;Fusobacteriales;Leptotrichiaceae	I	Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Moraxellaceae	III
Bacteria;Fusobacteria;Fusobacteriia;Fusobacteriales;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadac eae	III
Bacteria;Gemmatimonadetes;Gemmatimonadetes;BD2-11 terrestrial group;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;unknown	N A
Bacteria;Gemmatimonadetes;Gemmatimonadetes;Gemmatimonadales;Gemmati monadaceae	I	Bacteria;Proteobacteria;Gammaproteobacteria;Sva0071;unknown	I
Bacteria;Gemmatimonadetes;Gemmatimonadetes;PAUC43f marine benthic group;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Thiotrichales;Piscirickettsiaceae	III
Bacteria;Gemmatimonadetes;Gemmatimonadetes;unknown;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Thiotrichales;Thiotrichaceae	III
Bacteria;Gracilibacteria;unknown;unknown;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;unknown;unknown	N A
Bacteria;JL-ETNP-Z39;unknown;unknown;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Vibrionales;Vibrionaceae	III
Bacteria;Latescibacteria;unknown;unknown;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;JTB255 marine benthic group	I
Bacteria;Nitrospirae;Nitrospira;Nitrospirales;0319-6A21	I	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;uncultured	I

Bacteria;Nitrospirae;Nitrospira;Nitrospirales;Nitrospiraceae	III	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;unknown	I
Bacteria;Nitrospirae;Nitrospira;Nitrospirales;unknown	NA	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae	I
Bacteria;Planctomycetes;OM190;unknown;unknown	I	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadales Incertae Sedis	I
Bacteria;Planctomycetes;Phycisphaerae;Phycisphaerales;Phycisphaeraceae	I	Bacteria;Proteobacteria;Proteobacteria Incertae Sedis;Unknown Order;Unknown Family	I
Bacteria;Planctomycetes;Phycisphaerae;SHA-43;unknown	I	Bacteria;Proteobacteria;SPOTSOCT00m83;unknown;unknown	I
Bacteria;Planctomycetes;Pla3 lineage;unknown;unknown	I	Bacteria;Proteobacteria;unknown;unknown;unknown	NA
Bacteria;Planctomycetes;Planctomycetacia;Brocadiales;Brocadiaceae	I	Bacteria;Spirochaetae;Spirochaetes;Spirochaetales;Spirochaetaceae	I
Bacteria;Planctomycetes;Planctomycetacia;Planctomycetales;Planctomycetaceae	I	Bacteria;Tenericutes;Mollicutes;NB1-n;unknown	I
Bacteria;Planctomycetes;vadinHA49;unknown;unknown	I	Bacteria;Tenericutes;Mollicutes;unknown;unknown	I
Bacteria;Proteobacteria;AEGEAN-245;unknown;unknown	I	Bacteria;Verrucomicrobia;OPB35 soil group;unknown;unknown	I
Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacterales;Caulobacteraceae	I	Bacteria;Verrucomicrobia;Opitutae;Opitutales;Opitutaceae	I
Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacterales;Hyphomonadaceae	I	Bacteria;Verrucomicrobia;Opitutae;Puniceicoccales;Puniceicoccaceae	I
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;A0839	I	Bacteria;Verrucomicrobia;Spartobacteria;Chthoniobacterales;Chthoniobacteraceae	I
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae	I	Bacteria;Verrucomicrobia;Spartobacteria;Chthoniobacterales;DA101 soil group	I
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae	I	Bacteria;Verrucomicrobia;Spartobacteria;Chthoniobacterales;unknown	I
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;KF-JG30-B3	I	Bacteria;Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales;Verrucomicrobiaceae	I

Table S6.3. Environmental characterization and Pressure Index (PI) of each estuarine (E) and coastal (C) station. No data (ND). OM: Organic Matter; PAHs: Polycyclic aromatic hydrocarbons; PCBs: Polychlorinated biphenyls; Concentration of metals (mg kg $^{-1}$) and organic compounds (PAHs and PCBs) (µg kg $^{-1}$) is detailed. The Sediment Quality Guidelines value for each component is provided in parentheses.

Station	Salinity	Redox	% Sand	% OM	Zn (249)	Pb (78)	Hg (0.53)	Cd (1)	Cr (39)	Cu (55)	Ni (23)	∑PCBs (24.6)	∑PAHs (1607)	PI
1 (E)	0.1	265	86.1	5.7	134.8	46.2	0.1	0.11	27.5	51.8	30.4	10.3	357	1.31
2 (E)	0.3	574	99.9	2	116.9	77.9	0.09	0.16	19.1	74.4	50.5	7	180	1.17
3 (E)	29	299	99.3	1.3	147.6	223	1.5	0.62	96.3	163	75.9	16.1	341	2.15
4 (E)	33.1	-29	19.7	7.8	546.4	220.7	1.4	4.8	119	156.5	61.1	224	8100	3.76
5 (E)	33.7	137	86.1	5.2	321.6	233.4	2.1	1.1	103.8	119.8	70.8	197	63740	3.34
6 (E)	34.6	71	23.2	6.5	212.7	73.8	0.65	0.73	55.6	49.8	39.7	57.9	1139.9	2.25
7 (E)	35.1	450	98.8	1.7	197.9	70.4	0.95	0.21	21.2	41.6	33.4	7.8	589	1.38
8 (C)	35.2	439	98.2	1	115.9	55.8	0.24	0.12	18.3	24.5	30.7	7	180	0.91
9 (C)	35.1	472	99.4	1.8	328.4	108	0.29	0.16	29.4	82.8	45.1	7	764	1.54
10 (E)	2.6	175	18.8	3.6	120.6	30.5	0.17	0.28	39.4	28	29.4	12.5	985	1.36
11 (E)	13	608	93.6	0.7	65.3	12.6	0.12	0.05	22	13.3	22.7	7	576	0.64
12 (E)	8.5	593	99.7	0.4	51.9	7.5	0.07	0.04	25	10	20.2	7	182	0.51
16 (E)	0.2	97	49.3	10.5	69.2	12.7	0.16	0.12	47	18.8	53.1	7.8	413	1.44
17 (E)	2.4	91	32.5	4.6	107.9	14.5	0.17	0.04	54	28.8	43.3	17.1	811	1.45
18 (E)	21.4	422	99.3	0.8	63	7.2	0.08	0.09	26.7	11.8	28.1	7	180	0.67
19 (C)	35.7	333	98.2	1.2	167.2	55.4	0.67	0.22	22.5	28.4	24	7	237	1.15
20 (C)	35.7	464	96.3	1.5	ND	ND	ND	ND	ND	ND	ND	7	180	0.70
21 (E)	0.2	160	35.5	8.5	145.3	34.1	0.99	0.09	38.9	81.6	34.1	39.1	13694	2.24
22 (E)	0.3	296	96.7	1.2	90.6	56	0.25	0.08	32.9	33.6	36.7	7	241	1.08
23 (C)	35.7	466	98.4	1.1	90.8	37.2	0.13	0.13	23.3	11.1	19.6	7	185	0.74
24 (E)	0.2	232	85.6	5.9	122.2	47.2	0.09	0.11	34.7	47.6	41.6	7	1522	1.49
25 (E)	18.9	176	25.7	8	165.3	54.9	0.12	0.19	46.5	60.5	53.1	9.8	435	1.63
27 (E)	0.2	119	51.3	3	384.8	69.5	0.08	0.26	77.2	75.8	79.4	83.6	1264	2.17
28 (E)	0.2	512	97.5	1.9	430	93.2	0.14	0.45	111.7	121.2	113.2	45.9	1022.3	2.20

29 (C)	35.7	458	97.4	1.9	165	79.6	0.85	0.19	31.8	22.5	29.9	7	180	1.29
30 (E)	0.2	213	49.3	3.2	180.9	77	0.09	0.18	62.3	36.6	54.1	17.3	890	1.62
31 (E)	5.6	235	45.3	4.5	230.8	87.2	0.11	0.25	67.4	43.5	54.5	66.1	710	1.91
32 (E)	0.9	130	34.5	5.6	179.5	55	0.08	0.14	44.4	30.6	33.4	12	416	1.45
33 (C)	35.7	458	98.9	1.6	140.3	64.8	0.04	0.2	48.5	19.9	36.6	7	231	1.07
34 (E)	0.2	129	38.1	5	264.9	74.6	0.24	0.39	64.6	73.1	42.2	66.1	926	2.10
35 (E)	0.4	471	95.9	1	133.2	56.8	0.14	0.14	44.6	29.2	37.3	7	181	1.03
36 (C)	35.6	440	98.2	1.5	112.2	42.5	0.02	0.19	38.6	14	32.3	7	186	0.93
37 (C)	35.7	454	98.3	1.4	226.9	53.7	0.12	0.38	29.3	25.9	31.3	7	197	1.09
38 (C)	35.7	437	46.1	2.9	98.4	46.6	0.22	0.05	24.2	21.1	22.1	10.8	4573	1.29
39 (E)	0.1	376	90.9	1.8	361.8	242.9	0.37	0.3	45	42.6	32.6	11.3	201	1.63
40 (E)	5.6	325	51.8	5.9	206.4	115.2	3.4	0.39	71	48.7	36.3	7	185	2.03
41 (C)	35.2	294	86.2	2.1	130.8	44.1	0.11	0.27	38.6	19.3	28.6	11.3	985	1.21
42 (E)	33.7	99	28.5	10.8	1022.7	245.9	0.63	1.9	79.8	86.7	36.5	63.8	759	2.92
44 (E)	34.3	150	60.9	4.5	394.4	151.5	0.73	0.55	63.4	77	35.8	80.5	725	2.36
45 (C)	35.6	484	92.2	1.6	130	43.2	0.13	0.28	36.6	16.8	27.3	7.3	318.4	0.98
46 (C)	35.7	120	66	1.5	137.3	70.7	0.51	0.27	36.4	18	26.3	7.2	582	1.34
48 (E)	0.1	555	94.9	1	114	42.7	0.09	0.06	34.1	29.8	35.3	7	180	0.89
49 (E)	0.4	229	95.8	0.5	2244	1684	0.63	2.8	63.7	253.5	44.9	7	409	2.65
50 (E)	7.6	479	98.9	0.6	191.5	192	0.11	0.09	42.5	86.5	37.9	7.4	180	1.32
51 (C)	35.4	420	99.4	1.1	95.2	75.1	0.09	0.13	27.4	14.9	23.9	7	180	0.89

References for Supplementary Tables

Amaral-Zettle, L.A., McCliment, E.A., Ducklow, H.W., and Huse, S.M. (2009). A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. PLoS ONE.

Apakupakul, K., Siddall, M.E., and Burreson, E.M. (1998). Higher Level Relationships of Leeches (Annelida: Clitellata:

Euhirudinea) Based on Morphology and Gene Sequences. Molecular Phylogenetics and Evolution 12, 10.

Aranda, C.P., Valenzuela, C., Matamala, Y., Godoy, F.A., and Aranda, N. (2015). Sulphur-cycling bacteria and ciliated protozoans in a Beggiatoaceae mat covering organically enriched sediments beneath a salmon farm in a southern Chilean fjord. Mar Pollut Bull 100(1), 270-278.

Barrett, R.D.H., and Hebert, P.D.N. (2005). Identifying spiders through DNA barcodes. Canadian Journal of Zoology 83(3), 481-491.

Blaxter, M.L., De Ley, P., and Garey, J.R. (1998). A molecular evolutionary framework for the phylum Nematoda. Nature 392(71-75).

Bogan, B.W., Sullivan, W.R., Kayser, K.J., Derr, K.D., Aldrich, H.C., and Paterek, J.R. (2003). Alkanindiges illinoisensis gen. nov., sp. nov., an obligately hydrocarbonoclastic, aerobic squalane-degrading bacterium isolated from oilfield soils. Int J Syst Evol Microbiol 53(Pt 5), 1389-1395..

Burnette, A., Struck, T.H., and Halanych, K.M. (2005). Holopelagic Poeobius meseres (Poeobiidae, Annelida) is derived from within benthic Flabelligerid worms. Biological Bulletin 20, 208:213.

Campbell, B.J., Engel, A.S., Porter, M.L., and Takai, K. (2006). The versatile epsilon-proteobacteria: key players in sulphidic habitats. Nat Rev Microbiol 4(6), 458-468.

Campbell, A.M., Fleisher, J., Sinigalliano, C., White, J.R., and Lopez, J.V. (2015). Dynamics of marine bacterial community diversity of the coastal waters of the reefs, inlets, and wastewater outfalls of southeast Florida. Microbiologyopen 4(3), 390-408.

Carr, S.A., Orcutt, B.N., Mandernack, K.W., and Spear, J.R. (2015). Abundant Atribacteria in deep marine sediment from the Adelie Basin, Antarctica. Front Microbiol 6, 872.

Carranza, S., Giribet, G., Ribera, C., Baguña, J., and Riutort, M. (1996). Evidence that two types of 18S rDNA coexist in the genome of Dugesia (Schmidtea) mediterranea (Platyhelminthes, Turbellaria, Tricladida). Molecular Biology and Evolution 13(824-832).

Costa, F.O., deWaard, J.R., Boutillier, J., Ratnasingham, S., Dooh, R.T., Hajibabaei, M., et al. (2007). Biological identifications through DNA barcodes: the case of the Crustacea. Canadian Journal of Fisheries and Aquatic Sciences 64(2), 272-295.

Dang, H., Li, J., Chen, R., Wang, L., Guo, L., Zhang, Z., et al. (2010). Diversity, abundance, and spatial distribution of sediment ammonia-oxidizing Betaproteobacteria in response to environmental gradients and coastal eutrophication in Jiaozhou Bay, China. Appl Environ Microbiol 76(14), 4691-4702.

Elisabé, J.M., Brüchert, V., and Fuchs, B.M. (2012). Vertical shifts in the microbial community structure of organic-rich Namibian shelf sediments. African Journal of Microbiology Research 6(17), 3887-3897.

Fathepure, B.Z. (2014). Recent studies in microbial degradation of petroleum hydrocarbons in hypersaline environments. Front Microbiol 5, 173.

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3(5), 294-299.

Fondi, M., Orlandini, V., Emiliani, G., Papaleo, M.C., Maida, I., Perrin, E., et al. (2012). Draft Genome Sequence of the Hydrocarbon-Degrading and Emulsan-Producing Strain Acinetobacter venetianus RAG-1T. Genome Announc 194(17), 4771–4772.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., et al. (2012). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res 41(Database issue), D597-604.

Haaijer, S.C., Harhangi, H.R., Meijerink, B.B., Strous, M., Pol, A., Smolders, A.J., et al. (2008). Bacteria associated with iron seeps in a sulfur-rich, neutral pH, freshwater ecosystem. ISME J 2(12), 1231-1242.

Halanych, K.M. (1998). Considerations for reconstructing metazoan history: Signal, resolution, and hypothesis testing. American Zoologist 41, 38:929.

Hedlund, B.P., and Staley, J.T. (2001). Vibrio cyclotrophicus sp. nov., a polycyclic aromatic hydrocarbon (PAH)-degrading marine bacterium. Int J Syst Evol Microbiol. 51(Pt 1), 61-66.

Hedlund, B.P., and Staley, J.T. (2006). Isolation and characterization of Pseudoalteromonas strains with divergent polycyclic aromatic hydrocarbon catabolic properties. Environ Microbiol 8(1), 178-182.

Hedlun, B.P., Geiselbrecht, A.D., Bair, T.J., and Staley, J.T. (1999). Polycyclic Aromatic Hydrocarbon Degradation by a New Marine Bacterium, Neptunomonas naphthovorans gen. nov., sp. nov. Appl Environ Microbiol 65(1), 251–259.

Hillis, D.M., and Dixon, M.T. (1991). Ribosomal DNA: Molecular evolution and phylogenetic inference. Quarterly Review of Biology 53, 66:411.

Holmes, D.E., Finneran, K.T., O'Neil, R.A., and Lovley, D.R. (2002). Enrichment of Members of the Family Geobacteraceae Associated with Stimulation of Dissimilatory Metal Reduction in Uranium-Contaminated Aquifer Sediments. Appl Environ Microbiol 68(5), 2300-2306.

Holland, P.W.H., Hacker, A.M., and Williams, N.A. (1991). A molecular analysis of the phylogenetic affinities of Saccoglossus cambrensis brambell & cole (hemichordata). Phil. Trans. R. Soc. Land. B. 332, 185-189.

Ionescu, D., Sleber, C., Polerechy, L., Munwe, Y.Y., Lott, C., Häusler, S., et al. (2012). Microbial and Chemical Characterization of Underwater Fresh Water Springs in the Dead Sea. PLoS ONE 7(6:e38319).

Jiang, Y., Qu, Y., Xu, P., and Thang, H. (2015). Genome sequence of a versatile aromatic hydrocarbon-degrading bacterium, Arthrobacter sp. W1. Genome Announc 3((2):e00387-15).

Ju, F., and Zhang, T. (2015). Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. ISME J 9(3), 683-695.

Kappell, A.D., Wei, Y., Newton, R.J., Van Nostrand, J.D., Zhou, J., McLellan, S.L., et al. (2014). The polycyclic aromatic hydrocarbon degradation potential of Gulf of

Mexico native coastal microbial communities after the Deepwater Horizon oil spill. Front Microbiol 5, 205..

Kleindienst, S., Herbst, F.A., Stagars, M., von Netzer, F., von Bergen, M., Seifert, J., et al. (2014). Diverse sulfate-reducing bacteria of the Desulfosarcina/Desulfococcus clade are the key alkane degraders at marine seeps. ISME J 8(10), 2029-2044.

Lehner, A., Tasara, T., and Stephan, R. (2005). Relevant aspects of Arcobacter spp. as potential foodborne pathogen. Int J Food Microbiol 102(2), 127-135.

Leray, M., Yang, Y.J., Meyer, P.C., Mills, C.S., Agudelo, N., Ranwez, V., et al. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front Zool 10(34).

Liang, B., Wang, L.Y., Mbadinga, S.M., Liu, J.F., Yang, S.Z., Gu, J.D., et al. (2015). Anaerolineaceae and Methanosaeta turned to be the dominant microorganisms in alkanes-dependent methanogenic culture after long-term of incubation. AMB Express 5(1), 117.

Luo, J.F., Lin, W.T., and Guo, Y. (2011). Functional genes based analysis of sulfur-oxidizing bacteria community in sulfide removing bioreactor. Appl Microbiol Biotechnol 90(2), 769-778.

Machida, J.R., and Knowlton, N. (2012). PCR Primers for Metazoan Nuclear 18S and 28S Ribosomal DNA Sequences. PLoS ONE 7(9).

Meusnier, I., Singer, G., Landry, J., Hickey, D., Hebert, P., and Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. Bmc Genomics 9(10).

Meyer, C.P. (2003). Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. Biological Journal of the Linnean Society 79, 60.

Miyatake, T., MacGregor, B.J., and Boschker, H.T. (2009). Linking microbial community function to phylogeny of sulfate-reducing Deltaproteobacteria in marine sediments by combining stable isotope probing with magnetic-bead capture hybridization of 16S rRNA. Appl Environ Microbiol 75(15), 4927-4935.

Motoigi, T., and Okuyama, H. (2011). Fatty acid and hydrocarbon composition in tropical marine Shewanella amazonensis strain SB2B(T). J Basic Microbiol 51(5), 484-489.

Moubareck, C., Gavini, F., Vaugien, L., Butel, M.J., and Doucet-Populaire, F. (2005). Antimicrobial susceptibility of bifidobacteria. Journal of Antimicrobial Chemotherapy 55, 38–44.

Muyzer, G., and Stams, A.J. (2008). The ecology and biotechnology of sulphate-reducing bacteria. Nat Rev Microbiol 6(6), 441-454.

Paredes, C.J., Alsaker, K.V., and Papoutsakis, E.T. (2005). A comparative genomic view of clostridial sporulation and physiology. Nat Rev Micro 3(12), 969-978.

Passamaneck, Y.J., Schander, C., and Halanych, K.M. (2004). Investigation of molluscan phylogeny using large-subunit and small-subunit nuclear rRNA sequences. Molecular Phylogenetics and Evolution 38, 32:25.

Perez-Pantoja, D., Donoso, R., Agullo, L., Cordova, M., Seeger, M., Pieper, D.H., et al. (2012). Genomic analysis of the potential for aromatic compounds biodegradation in Burkholderiales. Environ Microbiol 14(5), 1091-1117.

Porazinska, D.L., Giblin-Davis, R.M., Faller, L., Farmerie, W., Kanzaki, N., Morris, K., et al. (2009). Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. Mol Ecol Resour 9(6), 1439-1450.

Radulovici, A.E., Sainte-Marie, B., and Dufresne, F. (2009). DNA barcoding of marine crustaceans from the estuary and gulf of st lawrence: A regional-scale approach. Molecular ecology resources 9(Suppl s1), 181-187.

Shchegolkova, N.M., Krasnov, G.S., Belova, A.A., Dmitriev, A.A., Kharitonov, S.L., Klimina, K.M., et al. (2016). Microbial Community Structure of Activated Sludge in Treatment Plants with Different Wastewater Compositions. Front Microbiol 7, 90.

Sohn, J.H., Kwon, K.k., Kang, J.H., Jung, H.B., and Kim, S.J. (2004). Novosphingobium pentaromativorans sp. nov., a high-molecular-mass polycyclic aromatic hydrocarbon-degrading bacterium isolated from estuarine sediment. Int J Syst Evol Microbiol. 54(Pt 5), 1483-1487.

Struck, T.H., Purschke, G., and Halanych, K.M. (2005). A scaleless scale worm: Molecular evidence for the phylogenetic placement of Pisione remota(Pisionidae, Annelida). Marine Biology Research 1(4), 243-253.

Thomas, F., Hehemann, J.H., Rebuffet, E., Czjzek, M., and Michel, G. (2011). Environmental and gut bacteroidetes: the food connection. Front Microbiol 2, 93.

Whiting, M.F. (2002). Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. Zoologica Scripta 31(93-104).

Yakimov, M.M., Giuliano, L., Gentile, G., Crisafi, E., Chernikova, T.N., Abraham, W.R., et al. (2003). Oleispira antarctica gen. nov., sp. nov., a novel hydrocarbonoclastic marine bacterium isolated from Antarctic coastal sea water. Int J Syst Evol Microbiol 53(Pt 3), 779-785.

Zhang, G., Haroon, M.F., Zhang, R., Hikmawan, T., and Stingla, U. (2016). Draft Genome Sequences of Two Thiomicrospira Strains Isolated from the Brine-Seawater Interface of Kebrit Deep in the Red Sea. Genome Announc 4((2):e00110-16).

Zhang, W., Song, L., Ki, J., Lau, C., Li, X., and Qian, P. (2008). Microbial diversity in polluted harbor sediments II: Sulfate-reducing bacterial community assessment using terminal restriction fragment length polymorphism and clone library of dsrAB gene. Estuar Coast Shelf S 76(3), 682-691.