Metabarcoding Analysis of the Eukaryotic Planktonic Communities of the Bilbao Estuary

David Abad García

Andone Estonba Rekalde
Aitor Albaina Vivanco

2018

Índice

Agradecimientos

En primer lugar me gustaría agradecer la dedicación de mis directores, Andone Estonba y Aitor Albaina, en el desarrollo de esta tesis. Su implicación a nivel de planificación, análisis e interpretación de resultados la han hecho posible. Gracias especialmente a Andone, por concederme esta segunda oportunidad y permitir que me incorporase a su grupo. Gracias a Aitor por sus incontables horas de trabajo, su mente enciclopédica y la disponibilidad 24/7.

En segundo lugar me gustaría darle las gracias a mi otra mitad de tesis, Mikel, porque sin él, esto no hubiese sido lo mismo. La verdad es que es una suerte poder llevarse tan bien con una persona con la que tienes que pasas la mayor parte de tu día. Por nuestros muestreos épicos, por las horas infinitas de laboratorio, por los errores del p*** Qiime, los congresos (y cachopos), las visitas a Asturias, los veranos de experimentos y la transferencia horizontal de conocimiento. FU-SIÓN.

En tercer lugar, y como es lógico, también debería mencionar a Imanol; compañero de compras por internet, estrenos de cine a horas intempestivas, sábados de películas/cena en casa y charlas sin fin en el labo. Gracias también a Montse por preocuparse siempre de nosotros. Al otro asturiano del pasillo, Álvaro. Al señorito andaluz, Llavero. Y a Esther, por un descenso del Sella “inolvidable”. Igualmente, gracias a Estitxu, a Ziortza y a la gente del labo de microbiología.

En cuarto lugar, no puedo olvidarme de dar las gracias a aquellas personas que me han ayudado/apoyado en algún momento de mi tesis: toda la gente que forma parte del grupo de Genomic Resources (Zarra, Montes, Otsanda, Jorge, etc), el resto de compañeros del laboratorio de genética y a los planctólogos (Aitor Laza e Ibón Uriarte).

Por último, también me gustaría agradecer a mi familia. A mis padres, por estar pendientes de mi futuro y apoyarme. A mi hermana, porque también sabe lo que es pasar por una tesis. Y a mi hermano, por preocuparse. Si no fuese por ellos no habría logrado llegar tan lejos. De la misma forma, también me gustaría darle las gracias a Elena por comprenderme, hacerme creer en mi mismo y animarme durante este último tramo.

Resumen

El estudio de la biodiversidad de las comunidades de plancton de los estuarios resulta de gran interés debido a que son parte fundamental de las cadenas tróficas acuáticas y a su capacidad de responder relativamente rápido ante cambios ambientales. Esta última característica sobretodo resulta de gran utilidad para la evaluación del estado de dichos ecosistemas ya que normalmente se encuentran sometidos a una gran presión antropogénica; es por ello que esta tesis se centra concretamente en la estuario de Bilbao, que llegó a ser uno de los más contaminados de España debido a la industrialización y creciente población de la ciudad. Debido a la mayor sensibilidad y resolución del metabarcoding, el objetivo principal de este trabajo es por tanto la de realizar una caracterización completa de la comunidad planctónica eucariota de este estuario, así como determinar tanto los patrones espaciales como temporales que la influencian. Con ello pretendemos conocer la composición taxonómica y abundancias relativas de dichas comunidades y, además, evaluar la adecuación del metabarcoding para la detección de alteraciones de esta comunidad, y como tal, su posible utilidad como herramienta rutinaria de monitoreo en la ría de Bilbao.

Introducción general

El plancton es esencial para el correcto funcionamiento de los ecosistemas acuáticos, ya que desempeña un papel crucial en las redes alimentarias y en los ciclos biogeoquímicos (Ward et al., 2012) mediante la transferencia de carbono de los productores primarios a los niveles tróficos más altos. Aparte de eso, sus tiempos de generación cortos les confiere la capacidad de respuesta rápida a cambios ambientales siendo esta la razón por la que los organismos planctónicos se vienen utilizando desde hace tiempo como indicadores en estudios ecológicos (Taylor et al., 2002). Es por ello que existen numerosos grupos de investigación dedicados a su estudio, como lo vienen a corroborar el incremento en publicaciones de estos últimos años (Lindeque et al. 2013, Logares et al. 2014, Hirai et al. 2015, Abad et al. 2016, Aguirre et al. 2017), debiéndose en gran medida a los avances tecnológicos que han permitido pasar del microscopio a la automatización mediante la aplicación de técnicas de secuenciación masiva.

1. Metabarcoding

1.1. Identificación de organismos basada en ADN

Una parte significativa de los estudios actuales de ecología planctónica requieren de la identificación de organismos durante la recolección de datos y normalmente se basan para ello en rasgos morfológicos observables mediante microscopía. Estas técnicas son a menudo difíciles debido a que las características de diagnóstico que se usan son limitadas, requieren un largo período de procesamiento y, además, una amplia experiencia debido a las similitudes entre especies (Lindeque et al., 2006) o a la presencia de especies crípticas (Chen y Hare, 2008). Además, muchas comunidades están a menudo compuestas de pocas especies muy abundantes y numerosas especies muy raras, lo que aumenta la dificultad para detectar e identificar todos los taxones (por ejemplo, Cheung et al., 2010). Es por ello que a lo largo de los años se han desarrollado varias estrategias basadas en ADN para aquellos casos en que la identificación basada en morfología resultaba problemática.

Los primeros métodos basados en ADN aparecieron a finales de la década de 1980 y se basaron en la hibridación de ADN utilizando sondas específicas (Gale y Crampton 1987) o mediante digestión previa con enzimas de restricción y posterior electroforesis (Curran y Webster 1987). Con la invención de la Reacción en Cadena de la Polimerasa (PCR; Mullis y Faloona 1987) y el diseño de primers universales (e.g. Taberlet et al., 1991), la identificación de especies pasó a ser por secuenciación Sanger (Cronin et al., 1991), primero con máquinas de un solo capilar (una muestra

por experimento) y posteriormente hasta varias decenas de capilares/muestras en paralelo.

Tabla 1. Lista de ventajas e inconvenientes de las distintas técnicas.

Técnica	Ventajas	Inconvenientes
Microscopía	Determinación de sexos Estados de desarrollo Parásitos asociados	Mucho tiempo / dinero Sinonimia Especies crípticas/foráneas
Barcoding/Sanger	Automatización Mayor resolución taxonómica	Un individuo por muestra Clonación para muestras de comunidad Bases de datos incompletas
Metabarcoding/NGS	Muestras de comunidad sin Necesidad de clonación Mayor sensibilidad Detección especies crípticas/foráneas Ahorro económico	Sesgos técnicos (contaminación, amplificación, Conocimientos de bioinformática, etc) Primers no universales Bases de datos incompletas Genes multicopia

Con esta última técnica apareció lo que se conoce como barcoding, que consiste en asignar a nivel de especie una espécimen o muestra (por ejemplo, un trozo de tejido o contenido estomacal) mediante la secuenciación de un fragmento de ADN corto normalizado ("DNA barcode" a modo de código de barras) y su comparación con una base de datos de secuencias de referencia (Hebert et al. 2003a). Entre las ventajas que tiene podemos incluir la de ser independiente de la experiencia taxonómica del usuario o permitir asignar nombres de especies a especímenes o muestras que son difíciles (o casi imposibles) de identificar de manera tradicional.

En este sentido, tradicionalmente se ha venido usando un fragmento de aproximadamente 650 pb (Folmer et al., 1994) del gen mitocondrial Citocromo c Oxidasa subunidad I (COI), que se proponía como suficiente para identificar al nivel de la especie la mayoría de los metazoos (Herbert et al., 2003b). Fue tal el impulso que generó este enfoque que en el año 2004 se creó el Consorcio "Barcode of Life" (CBOL) para crear una base de datos global de marcadores (barcodes) de biodiversidad para facilitar la identificación automatizada de especies. Uno de los argumentos esenciales para la difusión tan rápida que sufrió esta práctica se centra en el llamado "barcoding gap" que asume que la variación genética interespecífica excede la variación intraespecífica hasta tal punto que existe una distancia genética (o gap, en inglés) suficiente tal que permite la asignación de individuos no identificados a su especie con una tasa de error insignificante (Herbert Et al., 2003a). Como consecuencia, establecer el grado de divergencia de secuencia entre dos muestras por encima de un determinado umbral (generalmente al 3% de similitud de secuencia para las citadas

650 pb del barcode tradicional COI) indicaría distinción específica, mientras que la divergencia por debajo de dicho umbral indicaría la identidad taxonómica entre dichas muestras.

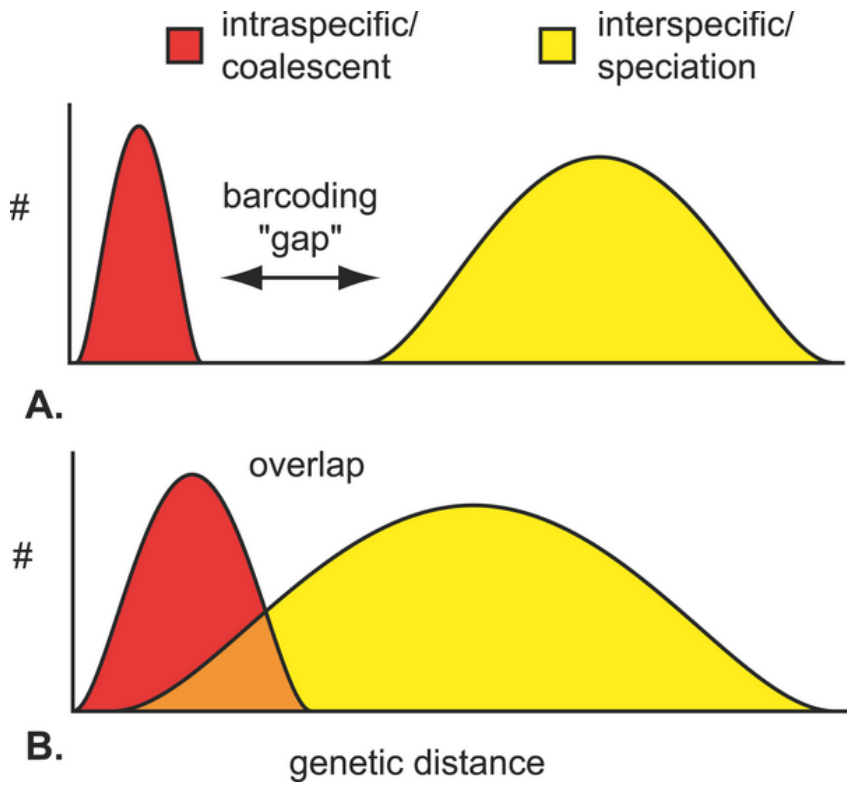


Figura 2. Esquema del "barcoding gap". La variación intraespecífica se muestra en rojo y la interespecífica en amarillo. En la gráfica A se muestra el ideal para el barcoding con una distribución discreta y un "gap", mientras que en la B se produce una superposición de las distribuciones sin "gap". Imagen extraída de Mayer & Paulay, 2005.

A raíz del barcoding también surgió una técnica que se aplicaba en muestras ambientales y que consistía en la clonación de los fragmentos amplificados previo a su secuenciación por Sanger (por ejemplo, Bowman & McCuaig 2003). Este proceso era muy laborioso y necesitaba de una gran cantidad de tiempo para llevarlo a cabo, por lo que se limitaba muchísimo el alcance de los estudios. Es por ello que el tamaño de muestras a caracterizar era muy limitado y la información sobre biodiversidad tenía una cobertura muy baja.

Finalmente, estas limitaciones quedaron superadas con la aparición de las tecnologías de secuenciación masiva (NGS por su siglas en inglés; Shendure & Ji 2008; Glenn 2011), la secuenciación de ADN ha experimentado impresionantes mejoras. Dichas plataformas pueden producir miles de millones de secuencias ("barcodes") en una sola ronda de secuenciación, lo que

conlleva una mejora sustancial en el número de "barcodes" secuenciados por unidad de tiempo y coste económico frente al enfoque tradicional de clonación seguida de método Sanger. Esta diferencia tan sustancial en la capacidad de secuenciación respecto a la generación anterior ha supuesto una revolución en muchas áreas de investigación científica, entre ellas la de la ecología (Medinger et al., 2010). En este sentido, las tecnologías de NGS poseen un enorme potencial para impulsar la adquisición de datos en la investigación sobre la biodiversidad, además de proporcionar una alternativa para superar los problemas asociados con el monitoreo basado en microscopía (Baird y Hajibabaei, 2012).

1.2. Origen del metabarcoding

Las tecnologías NGS junto con la necesidad de los ecologistas de identificar organismos a un alto nivel taxonómico han facilitado la aparición del metabarcoding. Este término se utiliza para designar a la identificación automatizada de múltiples especies a partir de muestras ambientales que contienen organismo enteros (comunidades) o ADN degradado (heces, contenidos estomacales, DNA extracelular, muestras forenses, etc). Más concretamente, una región corta de ADN de entre 100 y 500 pares de bases (barcode, a partir de ahora) se secuencia para una muestra de toda la comunidad y las secuencias obtenidas se utilizan para medir la biodiversidad a un coste asequible en términos no sólo económicos sino también de tiempo (por ejemplo Lindeque et al., 2013; Hirai et al., 2015). Aparte de esto, el metabarcoding es extremadamente sensible para la identificación de organismos, al nivel de la PCR en tiempo real (Zhan et al., 2015, Zimmermann et al., 2015), lo que resulta en una mejor capacidad para discriminar especies crípticas, estadios larvales (Lindeque et al., 2013) y especies con muy bajas abundancias que podrían ser pasadas por alto por los métodos tradicionales (Darling & Mahon, 2011).

Este enfoque usando tecnologías de NGS ya ha demostrado su potencial para el estudio de comunidades: trabajos recientes lo han aplicado para caracterizar diferentes grupos de organismos en ecosistemas acuáticos, incluyendo bacterias (por ejemplo Herlemann et al., 2011, Gilbert et al., 2012), protozoos (Bachy et al., 2013, Massana y otros, 2015), microalgas (Visco et al 2015, Eiler et al., 2013) y zooplancton (por ejemplo, Lindeque et al., 2013, Hirai et al., 2015). Estos estudios han revelado una riqueza taxonómica anteriormente oculta, incluyendo especies raras y parásitos (Lindeque et al., 2013, Logares et al., 2014). Además, proporcionaron estimaciones de biodiversidad mucho más altas y precisas que aplicando los métodos tradicionales (Bachy et al. 2013), destacando la alta sensibilidad y la mayor resolución taxonómica de la identificación por

metabarcoding. Finalmente, también se ha demostrado recientemente su aplicación para el monitoreo de especies alóctonas (Zaiko et al., 2015, Abad et al., 2016), destacando su potencial para detectar especies invasoras en estadios tempranos de colonización lo que permitiría aplicar políticas de erradicación/control más eficaces.

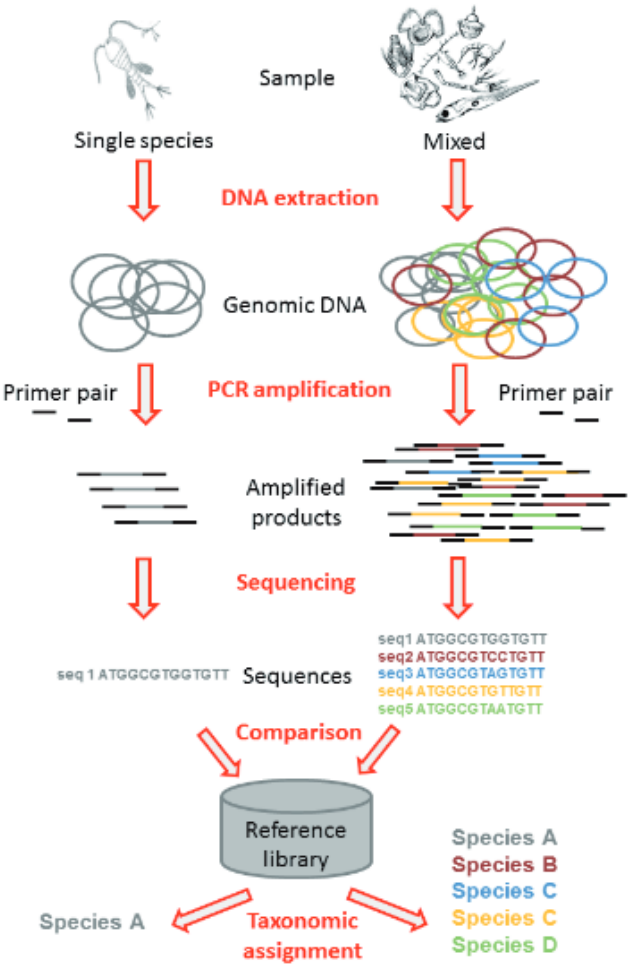


Figura 1. Se muestra el proceso de barcoding/sanger (izquierda) y el de metabarcoding/ngs (derecha). Imagen extraída de Corell & Rodriguez-Ezpeleta, 2014.

1.3. Limitaciones de la técnica

Como se ha mencionado anteriormente, el metabarcoding ha revolucionado la forma en que se realiza la caracterización de la biodiversidad a partir de muestras ambientales puesto que supone una alternativa de gran potencial para superar los problemas asociados a la identificación morfológica (Baird y Hajibabaei 2012) y permite la identificación de organismos con una mejor resolución taxonómica. Si bien esta técnica puede utilizarse como una herramienta de evaluación para la diversidad, tal como se implementa hoy en día todavía no está lista para reemplazar

completamente el análisis morfológico ya que tiene varias limitaciones.

En primer lugar, los sesgos técnicos introducidos durante el aislamiento del ADN influyen directamente en el resultado de la amplificación y secuenciación (Roh et al., 2006, Taberlet et al., 2012) debido a que el rendimiento del método de extracción varía con el tipo de organismo o incluso la etapa de desarrollo, además de que no todos se extraen con la misma eficiencia. El requisito de un paso de amplificación previo también resulta en la introducción de sesgos (Cline et al., 1996), teniendo típicamente dos orígenes: los errores durante la PCR debido a la tasa de error de la polimerasa, y los errores durante la secuenciación (González et al., 2012), que varían entre las diferentes tecnologías de secuenciación. Actualmente existen técnicas que se basan en la secuenciación de molécula individual (SMRT sequencing de PacBio), lo que permitiría saltarse el paso de PCR reduciendo así los sesgos a costa de una mayor tasa de error durante la secuenciación (Rhoads & Au, 2015).

Por otra parte, la precisión del metabarcoding para la asignación taxonómica depende en gran medida de la elección del marcador. Aunque se han publicado estudios que buscan primers universales (por ejemplo, Leray et al., 2013, Zhan et al., 2013), se ha detectado un balance entre la eficiencia de amplificación y el poder discriminatorio (Taylor y Harris 2012). Como muchas especies tienen que ser amplificadas durante la misma PCR es extremadamente importante que los primers usados para la amplificación sean altamente versátiles, es decir, que amplifiquen muchas moléculas diana diferentes con la misma eficiencia (universales) y se evite un sesgo producido por amplificación diferencial de unos grupos taxonómicos frente a otros. Pero encontrar un barcode adecuado, que posea una región de ADN variable lo suficientemente corta como para ser apta para metabarcoding (esta limitación de tamaño viene marcada por la propia capacidad de secuenciación de las tecnologías de NGS, constantemente sometida a mejoras por avance de la tecnología) y flanqueado por dos regiones altamente conservadas para anillar los primers, es complicado. En algunas ocasiones es difícil encontrarlos para ciertos grupos taxonómicos ya que la alta divergencia/variabilidad requerida impide identificar una región variable corta flanqueada por dos regiones conservadas. Es por ello que actualmente, con la bajada de precios de la secuenciación masiva, se baraja la posibilidad de hacer un “cocktail” de primers para suplir las carencias de la utilización de un único marcador.

Otra limitación importante del metabarcoding es la necesidad de bases de datos de referencia de

alta profundidad y calidad, es decir, de librerías que contienen secuencias específicas y "curadas" (editada y actualizada por expertos) de las especies obtenidas de la secuenciación a partir de especímenes taxonómicamente verificados (como pueden ser, en mayor o menor medida y no del todo inmunes a los errores, Silva para el 18S o BOLD para el COI). En este sentido, se ha demostrado que la adición de secuencias de especies locales relevantes que no tienen representación en las bases de datos disponibles aumenta significativamente la resolución taxonómica del método y es recomendable a la hora de diseñar este tipo de estudios (por ejemplo, Abad et al., 2016).

Tabla 2. Lista de ventajas e inconvenientes de los distintos primers.

Marcador	Primers más comunes	Tamaño (bp)	Ventajas	Inconvenientes
18S V1-2	SSU_F04 SSU_R22	~450	Situado al inicio del gen	Resolución intermedia
18S V4	TAReuk454FWD1 TAReukREV3	~400	Fragmento más utilizado Bastantes secuencias disponibles	Resolución intermedia
18S V9	1391f EukBr	~150	Tamaño reducido apto para ADN degradado	Baja resolución Situado al final del gen
COI (Leray et al 2015)	mlCOIintF jgHCO2198	~320	Resolución alta	No amplifica todos los grupos
COI (Meyer et al 2005)	dgLCO1490 dgHCO2198	~650	Resolución alta	No aptos a día de hoy para NGS por exceder el tamaño permitido
COI (Folmer et al 2004)	LCO-1490 HCO-2198	~650	Resolución alta	

Finalmente, aunque indudablemente constituye una ventaja a la hora de amplificar el ADN objetivos (frente al resto del genoma) la Variación de Número de Copias (CNV) asociada al ADN multicopia (rDNA, mtDNA) afecta las interpretaciones basadas en la abundancia relativa de las Unidades Taxonómicas Operacionales (OTUs, por sus siglas en inglés; Kembel et al., 2012). Sin embargo, se han identificado correlaciones entre la CNV y el tamaño del genoma en los eucariotas (Prokopowich et al., 2003) y entre la CNV y la longitud celular y el biovolumen en organismos unicelulares (Zhu et al., 2005, Godhe et al., 2008), lo que sugiere una forma potencial de abordar este problema en eucariotas. Mientras tanto, el metabarcoding con genes de copias múltiples permanecerá como un enfoque semicuantitativo (Amend et al., 2010; Albaina et al., 2016). Por otro lado, la naturaleza multicopia de estos barcodes también resulta ventajosa para el estudio del ADN degradado (heces, ADN extracelular, etc) ya que aumenta la probabilidad de amplificar al menos una de las copias (King et al., 2008).

2. Área de estudio

2.1. Estuarios

Los estuarios se encuentran entre los ecosistemas de mayor interés ecológico debido a su naturaleza transicional, dinámica particular y compleja, así como a su biodiversidad (McLusky y Elliott, 2004). A este respecto, los estuarios se caracterizan por gradientes de salinidad y dinámica de mareas que causan un movimiento neto de materia orgánica y otros nutrientes desde el interior del estuario hasta el mar (Herlemann et al., 2011). Por otro lado, parámetros fisicoquímicos como la salinidad, temperatura o la concentración de oxígeno disuelto son muy estables en las aguas oceánicas, pero varían ampliamente dentro de los estuarios debido al efecto de la entrada de agua dulce de los ríos (Kimmerer, 2002), mostrando generalmente un claro efecto estacional (Moran et al., 2013). Además, los estuarios se encuentran entre los hábitats más amenazados del mundo puesto que soportan una considerable presión antropogénica, como la contaminación o la introducción de especies alóctonas, debido a su cercanía a ciudades y puertos, lo que contribuye a la alteración del hábitat y cambios en la estructura y dinámica de la comunidades (Kennish, 2002).

2.2. Bilbao

El ría de Bilbao es un estuario principalmente euhalino (salinidad entre 30 y 35), que es un claro ejemplo de ecosistema con un pasado de alta presión antropogénica, ya que hasta hace relativamente poco tiempo era uno de los más contaminados en España. La morfología original del sistema fue fuertemente modificada desde mediados del siglo XIX por la utilización a gran escala de las zonas intermareales, lo que redujo el estuario original a un simple canal artificial (Cearreta et al. , 2000). Esta canalización cambió los patrones de circulación y rotación del agua, modificando tanto los procesos abióticos como los bióticos e incluyendo los patrones estacionales de la comunidad del plancton (Uriarte et al., 2014). Aparte de esto, los residuos procedentes de la ciudad y de las fábricas circundantes también modificaron el ecosistema del estuario de Bilbao durante los últimos siglos y, en particular, desde el comienzo de la era industrial (Cearreta et al., 2000). Como resultado, varias áreas desarrollaron condiciones hipóxicas o incluso anóxicas y gran parte de los organismos que las habitaban fueron eliminadas de la zona canalizada (González-Oreja y Saiz-Salinas, 1998). Aunque el estuario de Bilbao se encuentra actualmente siguiendo un programa de recuperación que fue iniciado en los años 90 (Borja et al., 2011), las concentraciones de contaminantes siguen siendo significativas (Cajaraville et al., 2016). Aún así, se han observado cambios significativos en la recuperación de la macrofauna bentónica (Borja et al., 2006) y en las comunidades de peces (Uriarte y Borja, 2009). Así mismo, también se ha reportado proceso de

recuperación para la comunidad del plancton (Albaina et al., 2009; Uriarte et al., 2015).

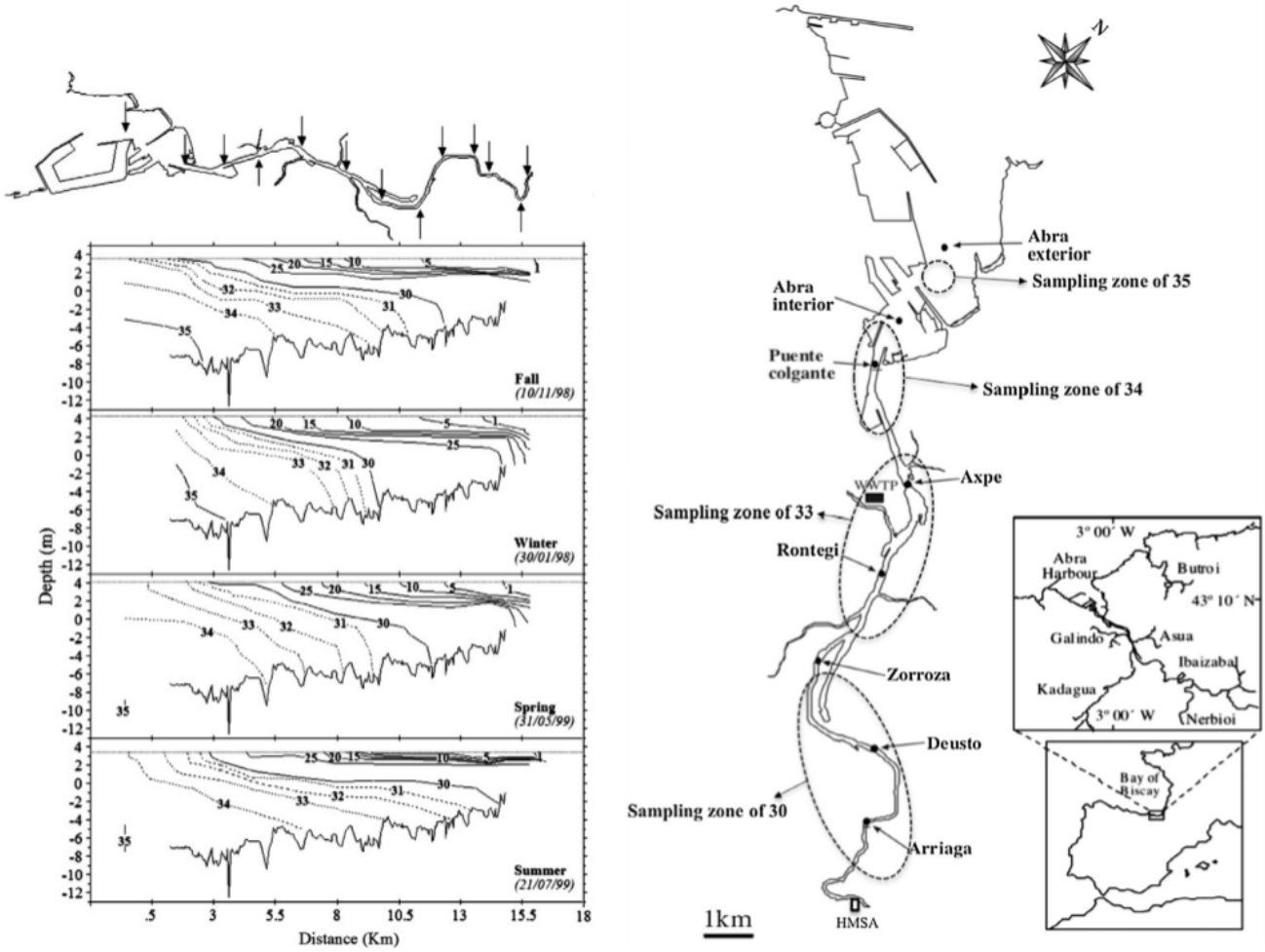


Figura 3. Distribución espacial de la salinidad (izquierda) y localización de los puntos de los puntos de muestreo en el estuario de Bilbao. Imágenes extraídas de Intxausti et al., 2012 y Villate et al., 2013 respectivamente.

2.3. Plancton

El aumento de la abundancia total de zooplancton en la zona interior (salinidad 30) del estuario de Bilbao observado a partir de 2001 coincidió con el inicio de una mejora sustancial en los niveles de oxígeno disuelto (Uriarte et al., 2015), lo que encaja con que estos organismos responden de forma rápida ante los cambios producidos en la calidad del agua (Uriarte y Villate, 2004, Mialet et al., 2011, Biancalana et al., 2012). Esta mejora en la oxigenación de la columna de agua se relacionó con la disminución de la contaminación derivada de las aguas residuales, particularmente cuando el tratamiento secundario comenzó a utilizarse en la planta de tratamiento de aguas residuales de Galindo en 2001 (García-Barcina et al., 2006; Villate et al. , 2013), lo que junto con el

declive industrial de la zona, ha provocado una mejoría general en los parámetros de calidad del agua y de los sedimentos (García-Barcina et al., 2006; Fernández-Ortiz de Vallejuelo et al., 2010). Aparte de que estos organismos pueden responder de manera diferente a la reducción y mitigación de la contaminación, las comunidades también pueden ser modificadas por la aparición de nuevas especies durante los procesos de cambio ambiental. Cabe destacar que en el estuario de Bilbao apenas hay estudios de zooplancton con tamaños corporales por debajo de 200µm (denominado nano- y micro-plancton) y que generalmente se centran en grupos taxonómicos más que en especies. En el caso del fitoplancton no hay estudios sobre los ciclos temporales y se centran en taxones concretos, por lo que no aportan una visión general de las comunidades de algas.

2.3.1 Zooplancton

Los copépodos son el grupo que más ha aumentado su abundancia durante las dos últimas décadas como resultado de la recuperación de la calidad de las aguas (Uriarte et al., 2015). Especies neríticas indígenas tales como *Acartia clausi* o *Paracalanus parvus* dominaban la comunidad y para el 2002 habían colonizado incluso hasta el estuario interno (salinidad 30). Al año siguiente se produjo un reemplazo parcial de esas especies por las poblaciones no indígenas de *Acartia tonsa* y *Oithona davisae* en dicha salinidad, lo que supuso un gran aumento en la proporción de copépodos dentro de la comunidad total debido a su tolerancia a ambientes con baja oxigenación. A partir de 2001 se detecta *Calanipeda aquadulcis*, que se convirtió en una de las más abundantes junto con *A. tonsa* (Uriarte et al., 2015).

Entre las especies de copépodos no indígenas que se establecieron con éxito en el estuario interno tras iniciarse el proceso de restauración encontramos *A. tonsa*, que se encuentra ampliamente distribuida por todo el mundo. Se observó por primera vez en el estuario interno de Bilbao en 2001 y se convirtió en dominante para 2003, desplazando a la especie nerítica congenérica *A. clausi* al estuario externo (Aravena et al., 2009). Esta separación espacial con predominancia *de A. clausi* y *A. tonsa* en aguas de mayor y menor salinidad, respectivamente, así como la segregación estacional en el estuario de Bilbao (Aravena et al., 2009) es un fenómeno también observado en otros estuarios del entorno (Azeiteiro et al., 2005; Gaudy et al., 2000). Otra de las especies que se establecieron con éxito en el estuario interno fue *Oithona davisae*, originaria de Asia (Mihneva y Stefanova, 2013) y que ahora se encuentra en algunos estuarios europeos, aunque inicialmente se identificaba erróneamente como *O. brevicornis* (Temnykh y Nishida, 2012). Tanto A. tonsa como *O. davisae* son capaces de habitar en aguas con concentraciones bajas de

oxígeno (Itoh et al., 2011; Roman et al., 1993) y presentan un patrones estacionales similares (Uriarte et al., 2015). Finalmente, la última especie que aumentó significativamente en el estuario de Bilbao durante los últimos años fue *Calanipeda aquaedulcis* (Uriarte et al., 2015). Esta especie fue citada por primera vez en el estuario de Bilbao en 2001 (Albaina et al., 2009) y desde entonces ha mostrado un gran incremento en abundancia (Uriarte et al., 2015).

2.3.2 Fitoplancton

Debido al carácter tóxico de algunas especies (capaces de provocar alertas sanitarias de primer orden y cuantiosas pérdidas económicas) y a la estrecha relación que existe entre ciertos taxones y determinadas condiciones ambientales, las comunidades de fitoplancton son una de las entidades biológicas consideradas como bioindicadores de la calidad del agua, y, de hecho, son de obligada monitorización de acuerdo a la Directiva Europea del agua (2008/56/CE y 2000/60/CE). Este crecimiento puntual de fitoplancton, también conocido como floraciones (o blooms, por su término en inglés), es un fenómeno que suele darse de manera habitual en estuarios repletos de nutrientes durante el verano, coincidiendo con un aumento en el tiempo de residencia del agua en conjunción con una mayor irradiación y temperatura (Paerl, 1996).

Aunque las comunidades de fitoplancton se desplazan a lo largo del estuario con la marea, la mayoría de los taxones son característicos de un área en particular. La parte externa del estuario de Bilbao contiene principalmente especies marinas, que en verano pueden ser desplazadas por la marea hacia el interior debido a la disminución de la descarga del río. Entre ellas, los más abundantes hasta la fecha pertenecen a los géneros *Pseudo-nitzschia*, *Chaetoceros*, *Heterosigma* y *Chrysochromulina* (aunque en ocasiones también se originan floraciones de géneros potencialmente dañinos). Además, varias especies de criptofitas, en las que principalmente domina el género *Teleaulax*, han sido aisladas del estuario externo (Laza-Martínez et al., 2012). Respecto al interior del estuario, la característica más destacada es la presencia de grandes cantidades de pequeñas diatomeas céntricas, la mayoría de ellas solitarias dominadas por los géneros tipo *Cyclotella* o *Thalassiosira* (Hevia-Orube et al., 2015), coincidiendo su presencia con las concentraciones mínimas de oxígeno en el estuario y los valores más altos de clorofila a.

3. Objetivos de la tesis

El estudio de la diversidad taxonómica de las comunidades planctónicas eucariotas de los estuarios resulta de gran interés puesto que son parte fundamental de las cadenas tróficas acuáticas y son capaces de responder relativamente rápido ante cambios ambientales. Esta última característica sobretodo es de gran utilidad para la evaluación del estado de "salud" de dichos ecosistemas ya que normalmente se encuentran sometidos a una gran presión antropogénica.

Esta tesis se centra concretamente en el estuario de Bilbao, uno de los más importantes del norte de España, que llegó a estar altamente contaminado debido a la industrialización y creciente población del área metropolitana (que acabaron prácticamente con la biodiversidad presente). La meta principal de este trabajo es por tanto la de obtener una caracterización completa de la comunidad planctónica eucariota de este estuario en dos salinidades diferentes (30 y 35), incluyendo tanto patrones espaciales como temporales. Con ello pretendemos conocer la composición taxonómica de la comunidad planctónica y sus abundancias relativas y, además, evaluar la adecuación del metabarcoding para la detección de alteraciones de esta comunidad, y como tal su utilidad como herramienta rutinaria de monitoreo en la ría de Bilbao.

La evolución de los organismos que componen la comunidad del estuario de Bilbao es conocida desde principios de los años 80 gracias a los diferentes esfuerzos de monitoreo realizados (Borja et al. 2006; Villate et al. 2013). Estos estudios hasta hoy han sido llevados a cabo mediante metodologías clásicas, es decir, la observación microscópica de la morfología de los organismos. Respecto a éstas, el metabarcoding proporcionaría (1) una mayor sensibilidad, lo que permitiría detectar especies que podrían haber pasado desapercibidas mediante métodos clásicos por sus bajas concentraciones, (2) una mayor resolución, de modo que alcanzaría a diferenciar organismos con características morfológicas similares o características diagnósticas limitadas y (3) además la capacidad de realizar los análisis taxonómicos con mayor rapidez y menor coste permitiendo un alto grado de automatización. Es por ello que en este estudio, implementaremos la técnica molecular del metabarcoding. Por lo tanto, los objetivos generales y específicos, que a su vez coinciden con los capítulos y artículos presentados en este trabajo, se describen a continuación:

1. Determinar la utilidad del metabarcoding como herramienta para el monitoreo de las comunidades eucariotas del plancton mediante su empleo para la caracterización de la composición taxonómica temporal y ambiental del plancton en la ría de Bilbao y su comparación con métodos tradicionales, incidiendo en dos aspectos: (1) Enriquecimiento de las bases de datos de referencia con secuencias de especies locales y (2) Evaluación de la sensibilidad para la detección de especies alóctonas.

2. Contribuir al desarrollo de un método de metabarcoding eficaz y fiable para el monitoreo de comunidades eucariotas de plancton en la ría de Bilbao midiendo la capacidad de dos marcadores nucleares (dos regiones diferentes y con distinta longitud de secuencia del gen 18S rDNA: 18S V1-2 y 18S V9) y uno mitocondrial (COI; Tabla 2 en la Introducción) para identificar la diversidad taxonómica de dichas comunidades, en base a indicadores como el número de taxones diferentes que pueden detectar, nivel de resolución al que pueden llegar, idoneidad para la cuantificación de abundancias, sesgos asociados, etc.

3. Definir los factores ambientales clave que impulsan los cambios en la estructura de la comunidad de plancton en sus fracciones de tamaño 0.22-20, 20-200 y >200 μm en la ría de Bilbao mediante (1) la realización de un seguimiento temporal de la composición de la comunidad en dos salinidades diferentes (2) la definición de correlaciones de esta composición con los distintos factores ambientales estudiados (3) puesta en contexto de los resultados obtenidos con los reportados anteriormente por estudios basados en microscopía (por ejemplo, Villate, 1994, Albaina et al., 2009).

Referencias

Abad D, Albaina A, Aguirre M, Laza-Martínez A and others (2016) Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. Mar Biol 163, 1–13. doi:10.1007/s00227-016-2920-0

Albaina A, Villate F, Uriarte I. (2009) Zooplankton communities in two contrasting Basque estuaries (1999–2001): reporting changes associated with ecosystem health. J. Plankton Res, 31: 739–752.

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: 1809–1824. doi:10.1002/ece3.1986

Aguirre M, Abad D, Albaina A, Cralle L, Goñi-Urriza MS, Estonba A, Zarraonaindia I (2017) Unraveling the environmental and anthropogenic drivers of bacterial community changes in the Estuary of Bilbao and its tributaries. PLOS ONE 12: e0178755

Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Mol Ecol, 19, 5555-5565.

Amorim Visco J, Apothoz-Perret-Gentil L, Cordonier A, Esling P , Pillet L, Pawlowski J. (2015) Environmental monitoring: inferring diatom index from next-generation sequencing data. Environ Sci Technol, 49: 7597-7605

Aravena G, Villate F, Uriarte I, Iriarte A, Ibáñez B (2009). Response of Acartia populations to environmental variability and effects of invasive congenerics in the estuary of Bilbao, Bay of Biscay. Estuar Coast Shelf Sci 83, 621–628. doi:10.1016/j.ecss.2009.05.013

Azeiteiro A, Marques SC, Miguel L, Vieira R and others (2005) Dynamics of the Acartia genus (Calanoida: Copepoda) in a temperate shallow estuary (the Mondego estuary). Acta Adriat 46 (1).

Bachy C, Dolan JR, López-García P, Deschamps P, Moreira D. (2013) Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes

and ITS regions using the conspicuous tintinnid ciliates as a case study. ISME J, 7(2): 244–255.

Baird DJ, Hajibabaei M (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Mol Ecol 21, 2039–2044.

Biancalana F., Menéndez M. C., Berasategui A. A., Fernández-Severini M. D., Hoffmeyer M. S. (2012) Sewage pollution effects on mesozooplankton structure in a shallow temperate estuary. Environ. Monit. Assess., 184, 3901–3913.

Borja A, Elliott M, Carstensen J, Heiskanen AS, van de Bund W. (2011) Marine management - Towards an integrated implementation of the European Marine Strategy Framework and the Water Framework Directives. Mar Pollut Bull, 60: 2175-2186.

Borja A, Muxika I, Franco J. (2006) Long-term recovery of soft-bottom benthos following urban and industrial sewage treatment in the Nervión estuary (southern Bay of Biscay). Mar Ecol Prog Ser, 313: 43-55.

Bowman, J. P., & McCuaig, R. D. (2003). Biodiversity, Community Structural Shifts, and Biogeography of Prokaryotes within Antarctic Continental Shelf Sediment. Applied and Environmental Microbiology, 69(5), 2463–2483. http://doi.org/10.1128/AEM.69.5.2463-2483.2003

Cajaraville MP, Orive E, Villate F, Laza-Martínez A and others (2016) Health status of the Bilbao estuary: A review of data from a multidisciplinary approach. Estuar. Coast. Shelf Sci., Special Issue: Functioning and dysfunctioning of Marine and Brackish Ecosystems 179, 124–134. doi:10.1016/j.ecss.2016.01.013

Cearreta A, Irabien MJ, Leorri E, Yusta I, Croudace IW, Cundy AB (2000) Recent Anthropogenic Impacts on the Bilbao Estuary, Northern Spain: Geochemical and Microfaunal Evidence. Estuar. Coast Shelf Sci 50, 571–592. doi:10.1006/ecss.1999.0582

Chen G, Hare MP (2008) Cryptic ecological diversification of a planktonic estuarine copepod, Acartia tonsa. Mol Ecol 17, 1451–1468.

Cheung MK, Au CH, Chu KH, Kwan HS, Wong CK (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. ISME J 4, 1053–1059. doi:10.1038/ismej.2010.26

Cline, J. et al. (1996) PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. Nucl. Acids Res. **24**, 3546–51.

Cronin MA, Palmisciano DA, Vyse ER, Cameron DG (1991) Mitochondrial DNA in wildlife forensic science: species identification of tissues. Wildlife Society Bulletin, 19, 94–105.

Curran J, Webster JM (1987) Identification of nematodes using restriction fragment length differences and species-specific DNA probes. Canadian Journal of Plant Pathology-Revue Canadienne de Phytopathologie, 9, 162–166.

Eiler A, Drakare S, Bertilsson S, Pernthaler J, Peura S, Rofner C, Simek K, Yang Y, Znachor P, Lindström ES. (2013) Unveiling Distribution Patterns of Freshwater Phytoplankton by a Next Generation Sequencing Based Approach. PLoS ONE, 8, e53516.

Fernández-Ortiz de Vallejuelo S, Arana G, de Diego A, Madariaga JM. (2010) Risk assessment of trace elements in sediments: the case of the estuary of the Nerbioi-Ibaizabal river (Basque Country). J. Hazard. Mater., 181, 565–573.

Folmer O, Black M, Hoeh W, Lutz R & Vrijenhoek R. 1994 DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3, 294–299.

Gale KR, Crampton JM (1987) DNA probes for species identification of mosquitos in the Anopheles gambiae complex. Medical and Veterinary Entomology, 1, 127–136.

García-Barcina J. M., González-Oreja J. A., De la Sota A. (2006) Assessing the improvement of the Bilbao estuary water quality in response to pollution abatement measures. Wat. Res., 40, 951–960.

Gaudy R, Cervetto G, Pagano M (2000) Comparison of the metabolism of Acartia clausi and A.

tonsa: influence of temperature and salinity. J. Exp. Mar Biol Ecol 247, 51–65. doi:10.1016/S0022-0981(00)00139-8

Gilbert JA, Steele JA, Caporaso JG, Steinbrük L, Reeder J, Temperton B, Huse S, McHardy AC et al. (2012) Defining seasonal marine microbial community dynamics. ISME J, 6: 298-308.

Godhe A, Asplund ME, Härnström K, Saravanan V, Tyagi A, Karunasagar I. (2008) Quantification of Diatom and Dinoflagellate Biomasses in Coastal Marine Seawater Samples by Real-Time PCR. Appl Environ Microbiol, 74, 7174–7182.

Gonzalez JM, Portillo MC, Belda-Ferre P, Mira A. (2012) Amplification by PCR Artificially Reduces the Proportion of the Rare Biosphere in Microbial Communities. PLoS ONE., 7(1):e29973. doi:10.1371/journal.pone.0029973.

Hebert PD, Cywinska A, Ball SL, deWaard JR: Biological identifications through DNA barcodes. Proc Biol Sci. 2003a, 270 (1512): 313-321. 10.1098/rspb.2002.2218.

Hebert PD, Ratnasingham S, deWaard JR: Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc Biol Sci. 2003b, 270 Suppl 1: S96-9. 10.1098/rsbl.2003.0025.

Herlemann DP, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J 5, 1571–1579. doi:10.1038/ismej.2011.41

Hevia-Orube J, Orive E, David H, Díez A and others (2016) Molecular and morphological analyses of solitary forms of brackish Thalassiosiroid diatoms (Coscinodiscophyceae), with emphasis on their phenotypic plasticity. Eur J Phycol 51, 11–30. doi:10.1080/09670262.2015.1077394

Hirai J, Kuriyama M, Ichikawa T, Hidaka K, Tsuda A (2015) A metagenetic approach for revealing community structure of marine planktonic copepods. Mol Ecol Res 15, 68-80.

Intxausti L, Villate F, Uriarte I, Iriarte A, Ameztoy I (2012) Size-related response of zooplankton to

hydroclimatic variability and water-quality in an organically polluted estuary of the Basque coast (Bay of Biscay). J Mar Syst 94, 87–96. doi:10.1016/j.jmarsys.2011.10.015

Itoh H, Tachibana A, Nomura H, Tanaka Y, Furota T, Ishimaru T (2011) Vertical distribution of planktonic copepods in Tokyo Bay in summer. Plankton Benthos Res 6, 129–134. doi:10.3800/pbr.6.129

Kembel SW, Wu M, Eisen JA, Green JL. (2012) Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS Comput Biol, 8(10):e1002743. doi: 10.1371/journal.pcbi.1002743.

Kennish MJ (2002) Environmental threats and environmental future of estuaries. Environ. Conserv. 29. doi:10.1017/S0376892902000061

Kimmerer WJ (2002) Effects of freshwater flow on abundance of estuarine organisms: physical effects or trophic linkages? Mar Ecol Prog Ser 243, 39–55. doi:10.3354/meps243039

King RA, Read DS, Traugott M, Symondson WOC (2008) Molecular analysis of predation: a review of best practice for DNA-based approaches. Molecular Ecology, **17**, 947−963.

Laza-Martínez, J. Arluzea, I. Miguel, E. Orive Morphological and molecular characterization of Teleaulax gracilis sp. nov. and T. minuta sp. nov. (Cryptophyceae) Phycologia, 51 (2012), pp. 649–661

Leray, Matthieu, Joy Y. Yang, Christopher P. Meyer, Suzanne C. Mills, Natalia Agudelo, Vincent Ranwez, Joel T. Boehm, and Ryuji J. Machida. 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." Frontiers in Zoology 10 (1): 1.

Lindeque PK, Hay SJ, Heath MR, Ingvarsdottir A and others (2006) Integrating conventional microscopy and molecular analysis to analyse the abundance and distribution of four Calanus congeners in the North Atlantic. J Plankton Res 28: 221-238.

Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, Atkinson A (2013) Next Generation

Sequencing Reveals the Hidden Diversity of Zooplankton Assemblages. PLoS ONE 8(11): e81327.

Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R, Claverie JM, Decelle J et al. (2014) Patterns of rare and abundant marine microbial eukaryotes. Curr Biol, 24: 813-821.

Massana R, Gober A, Audic S, Bass D, Bittner L, Boutte C, Chambouvet A, Christen R et al. (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. Environ Microbiol, doi: 10.1111/1462-2920.12955.

McLusky DS, Elliott M (2004) The Estuarine Ecosystem: Ecology, Threats and Management. OUP Oxford.

Meyer CP & Paulay G (2005). DNA Barcoding: Error Rates Based on Comprehensive Sampling. PLoS Biology, 3(12), e422. http://doi.org/10.1371/journal.pbio.0030422

Mialet B, Gouzou J, Azémar F, Maris T, Sossou C, Toumi N, Van Damme S, Meire P et al (2011) Response of zooplankton to improving water quality in the Scheldt estuary (Belgium). Estuar. Coast. Shelf Sci., 93, 47–57.

Mihneva V., Stefanova K. (2013) Non-native copepod Oithona davisae in Western Black Sea: seasonal and annual abundance patterns and effects of the invasion. Bioinvasions Rec., 2, 1–6.

Morán XAG, Ducklow HW, Erickson M (2013) Carbon fluxes through estuarine bacteria reflect coupling with phytoplankton. Mar Ecol Prog Ser 489, 75–85. doi:10.3354/meps10428

Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalysed chain reaction. Methods in Enzymology, 155, 335–350.

Paerl H (1996) A comparison of cyanobacterial bloom dynamics in freshwater, estuarine and marine environments Phycologia 35, pp. 25–35

Prokopowich CD, Gregory TR, Crease TJ. (2003) The correlation between rDNA copy number and

genome size in eukaryotes. Genome, 46: 48–50.

Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics. 2015;13:278–89. https://doi.org/10.1016/j.gpb.2015.08.002

Roh C, Villatte F, Kim BG, and Schmid RD. (2006) Comparative study of methods for extraction and purification of environmental DNA from soil and sludge samples. Appl. Biochem Biotechnol, 134**:**97-112.

Roman MR, Gauzens AL, Rhinehart WK, White JR (1993) Effects of low oxygen waters on Chesapeake Bay zooplankton. Limnol Oceanogr 38, 1603–1614. doi:10.4319/lo.1993.38.8.1603

Seoane S, Laza A, Urrutxurtu I, Orive E (2005) Phytoplankton Assemblages and Their Dominant Pigments in the Nervion River Estuary. Hydrobiologia 549, 1–13. doi:10.1007/s10750-005-1736-6

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology, 17, 1105–1109.

Taberlet, p., Coissac, e., Pompanon, f., Brochmann, c. and Willerslev, e. (2012), Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology, 21: 2045–2050. doi:10.1111/j.1365-294X.2012.05470.x

Taylor AH, Allen JI, Clark PA (2002) Extraction of a weak climatic signal by an ecosystem. Nature 416, 629–632. doi:10.1038/416629a

Taylor HR, and Harris WE. 2012. “An Emergent Science on the Brink of Irrelevance: A Review of the Past 8 Years of DNA Barcoding.” Molecular Ecology Resources 12 (3): 377–88. doi:10.1111/j.1755-0998.2012.03119.x.

Temnykh A, Nishida S. (2012) New record of the planktonic copepod Oithona davisae Ferrari and Orsi in the Black Sea with notes on the identity of “Oithona brevicornis“. Aquat. Invasions, 7, 425–431.

Uriarte I and Villate F. (2004) Differences in the abundance and distribution of copepods in two estuaries of the Basque coast (Bay of Biscay) in relation to pollution. J Plankton Res, 27(9): 863-

974.

Uriarte A and Borja A. (2009) Assessing fish quality status in transitional waters within the European water framework directive: setting boundary classes and responding to anthropogenic pressures. Estuar. Coast. Shelf Sci., 82, 214–224.

Uriarte I, Villate F, Iriarte A, Duque J, Ameztoy I (2014) Seasonal and axial variations of net water circulation and turnover in the estuary of Bilbao. Estuar Coast Shelf Sci 150: 312-324.

Uriarte I, Villate F, Iriarte A. (2015) Zooplankton recolonization of the inner estuary of Bilbao: influence of pollution abatement, climate and non-indigenous species. J. Plankton Res. doi: 10.1093/plankt/fbv060

Villate F, Iriarte A, Uriarte I, Intxausti L, de la Sota A. (2013) Dissolved oxygen in the rehabilitation phase of an estuary: influence of sewage pollution abatement and hydro-climatic factors. Mar Pollut Bull, 70(1-2):234-46.

Ward BA, Dutkiewicz SA, Jahn O, Follows MJ (2012) A size-structured food-web model of the global ocean. Limnol Oceanogr 57: 1877-1891.

Zaiko A, Samuiloviene A, Ardura A, Garcia-Vazquez E. (2015) Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. Mar Pollut Bull, 100(1):53-59. doi: 10.1016/j.marpolbul.2015.09.030.

Zhan A, Hulak M, Sylvester F, Huang X and others (2013) High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. Methods Ecol Evol 4: 558-565.

Zhu F, Massana R, Not F, Marie D, Vaulot D. (2005) Mapping of picoeukaryotes in marine ecosystems with a quantitative PCR of the 18S rRNA gene. FEMS Microbiol Ecol, 52(1): 79-92

Capítulo 1

PAPER 1: Is Metabarcoding Suitable for Estuarine Plankton Monitoring? A Comparative Study with Microscopy

David Abad, Aitor Albaina, Mikel Aguirre, Aitor Laza-Martínez, Ibon Uriarte, Arantza Iriarte, Fernando Villate, Andone Estonba

2016 Marine Biology 163: 149.

Abstract

Metabarcoding is becoming an increasingly valuable alternative approach to biodiversity assessment, due to the combination of extreme sensitivity and potential for the highest taxonomic resolution in a cost- and time-effective methodology. To evaluate the capacity of metabarcoding for estuarine plankton monitoring, a comparison between the results obtained with this approach were compared with those based on traditional taxonomic analysis (microscopy). Database incompleteness, one of the main limitations of metabarcoding, was somewhat overcome by the addition of DNA sequences for local species, which increased the taxonomic assignment success from 23.7% to 50.5%. When the communities were studied along with environmental variables, similar spatial and temporal trends of taxonomic diversity were observed for metabarcoding and microscopic studies of zooplankton, but not for phytoplankton. This is most likely attributable to the lack of representative sequences for phytoplankton species in current databases. In addition, there was high correspondence in community composition when comparing abundances estimated from metabarcoding and microscopy, suggesting semi-quantitative potential for metabarcoding. Furthermore, metabarcoding allowed the detection and identification of two non-indigenous species (NIS) found in the study area at abundances hardly detectable by microscopy. Overall, our results indicate that metabarcoding is a powerful approach with excellent possibilities for use in plankton monitoring, early detection of NIS and plankton biodiversity shifts.

Introduction

Plankton communities are essential for aquatic ecosystem functioning, playing a crucial role in food webs and biogeochemical cycles (Ward et al. 2012). Furthermore, due to their rapid response to environmental variation, planktonic organisms have been used as indicators of ecosystem change for monitoring purposes (e.g. Taylor et al. 2002). These features highlight the interest of studying plankton community structure, biodiversity, and responses to environmental factors.

On the one hand, phytoplankton biomass has been mainly estimated using Chlorophyll-*a* (Chl-*a*) as a proxy (e.g. Bricker et al. 2003). It has also been recently included in the Marine Strategy Framework Directive (MSFDe2008/56/EC) (e.g. Ferreira et al. 2011, Uriate et al. 2015). However, biodiversity of phytoplankton is very difficult to estimate and monitoring has usually been limited to certain groups (e.g. Amorim Visco et al. 2015; Eiler et al. 2013). On the other hand, and despite its ecological importance, zooplankton biodiversity is not yet included in European marine environmental policies (Borja et al. 2011). This is mostly related to constraints of microscopy-based identification, as the identification of morphological characteristics by light microscopy is complicated, time-consuming and requires wide expertise due to morphological similarities between species and restricted diagnostic features (e.g. Lindeque et al. 2011). Microscopy-based biodiversity assessment is also subjected to an unpredictable, but probably significant, bias due to the presence of cryptic species (e.g. Chen and Hare, 2008).

The advent of Next Generation Sequencing (NGS) technologies has provided an alternative to overcome issues associated with microscopy-based monitoring (Baird and Hajibabaei, 2012). These NGS technologies allow the use of metabarcoding, where a short DNA region is sequenced for a whole community sample and the obtained sequences are used to measure biodiversity at an affordable cost (e.g. Lindeque et al. 2013; Hirai et al. 2015). The reported high sensitivity of NGS (Zhan et al. 2013) makes this technique ideal for the detection of rare taxa. Therefore, NGS-based metabarcoding generates large amounts of biodiversity information and is capable of identifying species at any life stage in taxonomically complex assemblages (Comtet et al. 2015), including the precise identification of cryptic species and overlooked by traditional methods because they are either too fragile or too small. A particularly useful application of metabarcoding is the early detection of non-indigenous species (NIS), which are a cause of great concern in monitoring of vulnerable ecosystems, such as estuaries.

An essential requirement for metabarcoding studies is a deep and curated database of reference DNA sequences for identified specimens. Currently, this necessity represents one of the main drawbacks, since some groups of organisms have none or very few publicly-available sequences. In addition, it is also known that the Copy Number Variation (CNV) associated with rDNA genes could affect the abundance estimates when using metabarcoding (Kembel et al. 2012), explaining the lack of correlation between this approach and microscopy in some cases (e.g. Stoeck et al. 2014); other technical biases introduced during DNA extraction (Roh et al., 2006) or the PCR amplification step (Gonzalez et al. 2012) influence these estimates as well.

Recent studies have applied metabarcoding to characterize different groups of organisms in aquatic ecosystems, including bacteria (e.g. Herlemann et al. 2011; Gilbert et al. 2012), protozoa (Bachy et al. 2013; Massana et al. 2015), microalgae (Amorim Visco et al. 2015; Eiler et al. 2013), and zooplankton (e.g. Lindeque et al. 2013; Hirai et al. 2015). These studies have revealed previously hidden taxonomic richness, including rare species and parasites (Lindeque et al. 2013; Logares et al. 2014), and provided much higher biodiversity estimates than traditional methods (Bachy et al. 2013), highlighting the sensitivity and higher taxonomic resolution of NGS-based metabarcoding. Moreover, its application for NIS monitoring has been recently demonstrated (Zaiko et al. 2015c).

In contrast to many previous studies, which focused on a particular size fraction and/or limited number of taxonomic groups, we analyzed the entire eukaryotic plankton community (0.22-20, 20-200 and >200 μm size-fractions) of an estuary. We selected the V9 region of the 18S rDNA gene (18S V9) primarily because of its broad amplification range among eukaryotes (de Vargas et al. 2015; Albaina et al. 2016a), but also because the Earth Microbiome Project (EMP; http://www.earthmicrobiome.org) designed a protocol for Illumina platforms, that has markedly increased sequencing depth compared to the previously dominant NGS technology (i.e., Roche's 454) in the metabarcoding field (Mahé et al. 2015). Metabarcoding using the 18S V9 has recently allowed the characterization of marine plankton biodiversity within the Tara Oceans (http://www.embl.de/tara-oceans/start/; Massana, 2015) and Biomarks initiatives (http://www.biomarks.eu/; de Vargas et al. 2015).

The main objective of this study was to compare results of plankton community taxonomic composition based on metabarcoding versus microscopy in order to assess the usefulness of metabarcoding for estuarine plankton monitoring. We also used the 18S V9 to analyze the community spatio-temporal structure in relation to environmental parameters. Moreover, we studied the effects of completeness of the reference database on taxon assignment by adding local species, and the sensitivity of metabarcoding for NIS detection.

Methods

Study area

The estuary of Bilbao is located in the south-east part of the Bay of Biscay (within 43º23'N to 43º14'N and 3º07'W to 2º55'W). It is a small (~23 km long), narrow (25-270 m), shallow (0.5 - >10 m) and highly stratified channel that crosses urban and industrial settlements and drains into a wide coastal embayment (Uriarte et al. 2014). It was one of the most polluted estuaries in Europe, but since 1979 it has undergone a significant improvement of water/sediment quality and recovering of biodiversity (Borja et al. 2006; Villate et al. 2013). This transition has allowed the recolonization by a mixture of neritic and estuarine species (Albaina et al. 2009; Uriarte et al. 2015). Among them, there are NIS such as the copepods *Acartia tonsa* (Calanoida, Acartiidae), which was first described in this estuary in 2001, became dominant the following year (displacing congeneric species; Aravena et al. 2009); and *Pseudodiaptomus marinus* (Calanoida, Pseudodiaptomidae), which was recently collected for the first time in the estuary of Bilbao (Albaina et al. 2016b) and whose effect on the community cannot yet be predicted.

Sampling

Sampling was carried out in summer (June, July) and autumn (September, October) of 2013 from water at 30 and 35 ppt salinity during neap tides. Salinity (g/L), temperature (ºC), dissolved oxygen (DO; mg/L) and pH at each sampling point were measured with a YSI 556 MPS multi-parameter probe. Water transparency was measured with a Secchi Disk. Chl-*a* concentrations (mg/L) were calculated from spectrophotometric measurements on acetone extracts, following the monochromatic method with acidification (Jeffrey and Mantoura 1997). In addition, precipitation (ml/m^2) data was provided by the Hydrometeorology Service of the Regional Council of Bizkaia.

To obtain the 0.22-20 and 20-200 µm size fractions, a Niskin bottle was used to collect 10 L of water at each salinity (sampling depth depended on the water mass location). Samples were filtered through a 20 µm mesh (Millipore Nylon Nets) and, consecutively, approximately 1 L was filtered through a 0.22 µm Durapore Membrane (Millipore) using a Kitasato flask and a vacuum pump (Millipore). Meshes and membranes were kept in cryogenic tubes and frozen at -80ºC until further use for metabarcoding. For the microscopy analysis of phytoplankton samples, a 250 mL bottle with 1mL of Lugol was filled directly with water from the Niskin Bottle.

For the >200 μm size fraction, samples were collected at about 3 m depth by ~5 min horizontal tows of a 200 μm plankton net (mouth diameter: 0.25 m) equipped with a Hydrobios flow-meter. Once in the laboratory, each sample was divided using a plankton splitter. One half was kept in buffered formalin (4%) for microscopy analysis. One quarter was filtered through a 180 μm mesh (Millipore Nylon Nets) and preserved in ethanol for DNA barcoding analysis of some selected species. Finally, the remaining quarter was also filtered through a 180 μm mesh, kept in a falcon tube and frozen at -80ºC until further use for metabarcoding.

Microscopy

Both Lugol-fixed (non-filtered) and formalin-fixed (>200 μm) plankton samples were identified at the lowest taxonomic resolution possible. The phytoplankton community was characterized from the Lugol-fixed bottle samples through the Utermöhl or sedimentation method (Edler et al. 2010). Additionally, living sub-samples were observed under light microscopy on the day of sampling to determine the presence of species difficult to identify after fixation. Fixed phytoplankton cells from the settled samples were identified and counted under a Nikon Diaphot TMD (Nikon Corporation, Tokyo, Japan) inverted microscope. Heterotrophic dinoflagellates and some non-photosynthetic nanoflagellates such as kathablepharids and choanoflagellates were included in the analyses. Ten mL and 50 mL (to obtain sufficient abundances) were settled for 30 and 35 ppt salinities, respectively. The entire chamber area was examined at 100x magnification and transects were performed at 200x-400x. A minimum of 300 cells (average of 593 cells per sample) were counted. Species biomass was calculated using formulas from the Baltic Marine Environment Protection Commission (HELCOM; Olenina et al. 2006).

For zooplankton, identification of the >200 μm sample was carried out under an inverted stereo-microscope. A minimum of 100 individuals of the most abundant taxa was counted before finishing sub-sampling or, if not possible, the whole sample was examined. Absolute and relative abundances were computed for copepods. Biomass was also estimated for copepods based on the average size of individuals (http://copepodes.obs-banyuls.fr), assuming the Carbon content was 40% of the dry weight (Bamstedt, 1986) and following the formula from Gaudy and Boucher (1983).

DNA extraction

A modified salt protocol (Aljanabi and Martinez 1997) was used to extract the DNA from the 20 and 200 μm size-fractions. Meshes were defrosted and cleaned in a falcon tube with distilled water

injected through a wash bottle to remove any possible attached organisms or DNA-containing fragments. The filters were held with each tube lid and centrifuged at 4000 rpm for 30 min to create a pellet. The mesh and supernatant water were then carefully removed. Proteinase K (20 mg/mL) digestion was conducted on the pelleted organisms overnight and the extraction was continued according to the protocol. Samples from the 0.22 μm mesh were extracted following the instructions of the MOBIO PowerSoil® DNA Isolation Kit.

Extracted DNA was first quantified by spectrophotometry using Nanodrop (ND-1000; Thermo Scientific) and then by fluorometry using Qubit 1.0 (Thermo Scientific) to determine the amount of double-stranded DNA. DNA was stored at -20ºC until further processing.

DNA barcoding

After checking the microscopy results and previous studies on the area (Albaina, 2009; Uriarte, 2015), we noted the absence of some key/abundant species (e.g., *Pseudodiaptomus marinus*) of the estuary of Bilbao in publicly available databases. In order not to miss these taxa in our analysis of community composition by metabarcoding, we generated 18S V9 reference sequences (by Sanger sequencing) to include them in the database. Five copepods and three cladocerans species were isolated from the ethanol-preserved sample splits (Table 1; GenBank accession numbers KP768152-KP768156 and KR919779-KR919787). For *Acartia clausi*, *Euterpina acutifrons* and the cladocerans, 10 individuals were pooled in each extraction tube. We used the EMP primers 1391f (5'-GTACACACCGCCCGTC-3') and EukBr (5'-TGATCCTTCTGCAGGTTCACCTAC-3'), based on Stoeck (2010), for the amplification of the 18S V9 (~150 bp fragment). Polymerase Chain Reactions (PCRs) were performed in a 25 μL volume containing 7.5 μL of distilled water, 5.4 μL of Buffer (5x), 2.7 μL of $MgCl_2$ (25 mM), 2.7 μL of dNTPs (10 μM), 2.7 μL of each primer (10 μM), 0.3 μL of Taq Polymerase (Promega) and 1 μL of template DNA. PCR cycling included initial denaturation at 92 °C for 3 min, followed by 30 cycles of 45 sec denaturation at 92 °C, 1 min annealing at 57 °C and 1.5 min extension at 72 °C. A final extension step was performed at 72 °C for 10 min. The purified PCR products were sequenced in both directions on an ABI 31309 capillary electrophoresis Analyzer with ABI BigDye Terminator version 3.1 chemistry (Applied Biosystems) at the SGIKER (UPV/EHU).

Preparation of the custom reference databases

In order to illustrate the effects of reference database completeness, we studied the assignation rate in two versions of Silva clustered at 99% identity (http://www.arb-silva.de/; Quast et al. 2013) with and without the inclusion of the herein-generated 18S V9 sequences (Table 1). Silva 111 and 119 were the last two available releases at the time the study was performed; the difference in over 1 million reference sequences is explained by their publication dates (July 2012 and 2014, respectively). To further prove the effect of adding sequences corresponding to key/local species, we also included 9 copepod species inhabiting North East Atlantic (NEA) neritic waters (Laakmann et al. 2013; ESM_1) and representative sequences from *Oithona brevicornis*, *O. nana*, *O. similis* and *Oncaea media* (GenBank accession numbers: JF288757, HQ008734, KF153700 and AM114421; ESM_1), for which we could not find other reliable sources.

Species	GenBank Accession Numbers		
Acartia tonsa	KP768152	KP768153	KP768154
Acartia clausi	KR919781	KR919782	
Calanipeda aquaedulcis	KP768157	KP768158	
Euterpina acutifrons	KR919779	KR919780	
Pseudodiaptomus marinus	KP768155	KP768156	
Evadne nordmanni	KR919787		
Evadne spinifera	KR919783	KR919784	
Podon spp.	KR919785	KR919786	

Table 1 GenBank accession numbers of the 18S V9 sequences generated in this study

Metabarcoding , OTU definition and taxonomic assignation

Sequencing of the 18S V9 region was carried out at the Argonne National Laboratory (Lemont, IL, USA) following the EMP protocols and using Illumina MiSeq 2x150 bp. Raw reads were trimmed with Sickle Ver. 1.33 (Joshi and Fass 2011), using a quality threshold of 20. For paired-end merging, Pear Ver. 0.9.5 (Zhang et al. 2014) was used with a minimum overlapping of 15 bp and a cut-off P-value of 0.01. The barcodes from the sequences discarded in the previous steps were removed by fastq-barcode.pl (Smith, 2012). Chimeras were removed with UCHIME (Edgar et al. 2011), using a reference-based chimeric detection against Silva 119 custom.

Merged reads were processed using Qiime Ver. 1.9 (Caporaso et al. 2010): sequences were clustered into Operational Taxonomic Units (OTUs) with UCLUST (Edgar et al. 2010), using both

de novo and closed reference approaches with 97% and 99% identity thresholds. The *de novo* approach groups sequences based on sequence identity (Navas-Molina et al., 2013) and taxonomy is then assigned to the obtained representative sequences with BLAST (Altschul et al. 1990). The closed-reference approach matches sequences to an existing database of reference sequences (Silva, in our case); if a sequence fails to match the database, it is discarded (Navas-Molina et al., 2013). This approach assigns the taxonomy with UCLUST (Edgar et al. 2010).

Statistical analysis

Canonical Correspondence Analysis (CCA) of the OTUs showing >1% relative abundance were carried out using CANOCO Ver. 4.5 (ter Braak and Smilauer, 2002) to investigate the relationship between taxon abundances from metabarcoding versus microscopy (0.22-200 and >200 μm size-fractions) in relation to samples and measured environmental variables. The 18S V9 sequences from the 0.22 and 20 μm filters were merged (after rarification) so that the comparison with phytoplankton microscopy-based results (Utermöhl) can be possible. All the CCAs were constructed using relative abundance data, with square root transformations used to normalize the samples.

Spearman's rank correlation coefficient (ρ) was calculated for the comparison of relative abundances retrieved by microscopy (both counts and biomass measurements) and metabarcoding (18S V9 reads) using cor.test in R (R Core Team, 2015); the correlations were limited to taxa uncovered by both methods (ESM_2).

Results

Metabarcoding

Only 0.24% of the reads were discarded due to poor quality. Of the remaining, 89.89% were successfully merged. In all, 0.02% of the reads were eliminated due to their putative chimeric nature. Once the OTU table was constructed, 229 singletons were discarded from further analysis. Finally, the rarefaction curves (ESM_3) showed that the plateau was reached at 3000 reads in most of the samples.

The closed-reference approach produced 1174 and 831 OTUs for the 97 % and 99% identity thresholds, respectively. The comparison of the different databases (Table 2) showed that "Silva 119 custom" had the highest proportion of taxonomic assignments with 53.5%, whereas "Silva 111" only reached 23.7%. Although the assignment increased to 66.4 % with the 97% identity threshold, the 99% (as in Albaina et al. 2016a) was considered to be optimal, given its taxonomic resolution and stringency for the 18S V9 region (below the 1% machine error rate; Quail et al. 2012).

	Silva 111			**Silva 111 Custom**			**Silva 119**			**Silva 119 Custom**		
	0.20-20	**20-200**	**>200**	**0.20-20**	**20-200**	**>200**	**0.20-20**	**20-200**	**>200**	**0.20-20**	**20-200**	**>200**
June 30	28.21	5.25	14.46	40.96	67.99	87.34	55.60	5.63	14.67	55.69	68.12	87.34
June 35	50.71	17.38	24.26	55.62	80.59	86.81	55.26	22.96	48.81	60.09	80.52	86.49
July 30	42.38	1.16	13.69	42.42	10.79	59.68	23.95	0.98	14.85	23.99	10.36	59.47
July 35	46.03	35.28	88.17	46.05	43.39	89.68	53.61	51.20	91.24	53.62	57.81	92.64
Sept 30	22.53	0.75	24.97	22.57	21.67	33.7	22.78	6.55	29.91	22.80	21.68	33.71
Sept 35	38.21	21.30	10.58	38.23	72.84	86.58	54.06	24.55	12.81	54.08	73.71	87.13
Octo 30	30.36	2.31	13.35	30.63	10.16	79.31	35.11	2.44	76.93	35.14	8.85	79.31
Octo 35	25.05	6.63	6.54	25.48	39.69	35.48	42.18	16.38	19.58	42.59	49.41	39.62
Mean	35.44	11.26	24.5	37.75	43.39	69.82	42.82	16.34	38.60	43.50	46.31	70.71
Total	23.73			50.32			32.58			53.51		

Table 2 Percentage of sequences that were assigned to taxonomy using four different databases. Similarity threshold was set at at 99%. Total assignment percentage for each database is shown along with those for each specific size fraction (0.22-20, 20-200 and >200 μm), salinity (30 and 35 ppt) and sampling month (June-October)

When using the *de novo* approach instead, a taxonomic correspondence was obtained for approximately 100% of the reads, yielding a total of 2139 and 2318 OTUs for 97 % and 99 % identity thresholds respectively. However, due to the low identity values of the taxonomy obtained for the sequences that were unassigned in the closed-reference, we focused on the results produced by this latter method.

The resulting 831 OTUs (produced by the closed-reference at 99% similarity) were classified into 17 taxonomic ranks (those representing less than 1% abundance were not plotted), uncultured taxa and unassigned reads (Fig. 1). Interestingly, the percentage of reads for which the taxonomy was assigned was higher for salinity 35 (64.8%) than 30 ppt (42.2%). In general, the percentage of unassigned reads was lower as the size fraction increased (Table 2): 56.5, 53.7 and 29.3% for the 0.22-20, 20-200 and >200 µm, respectively .

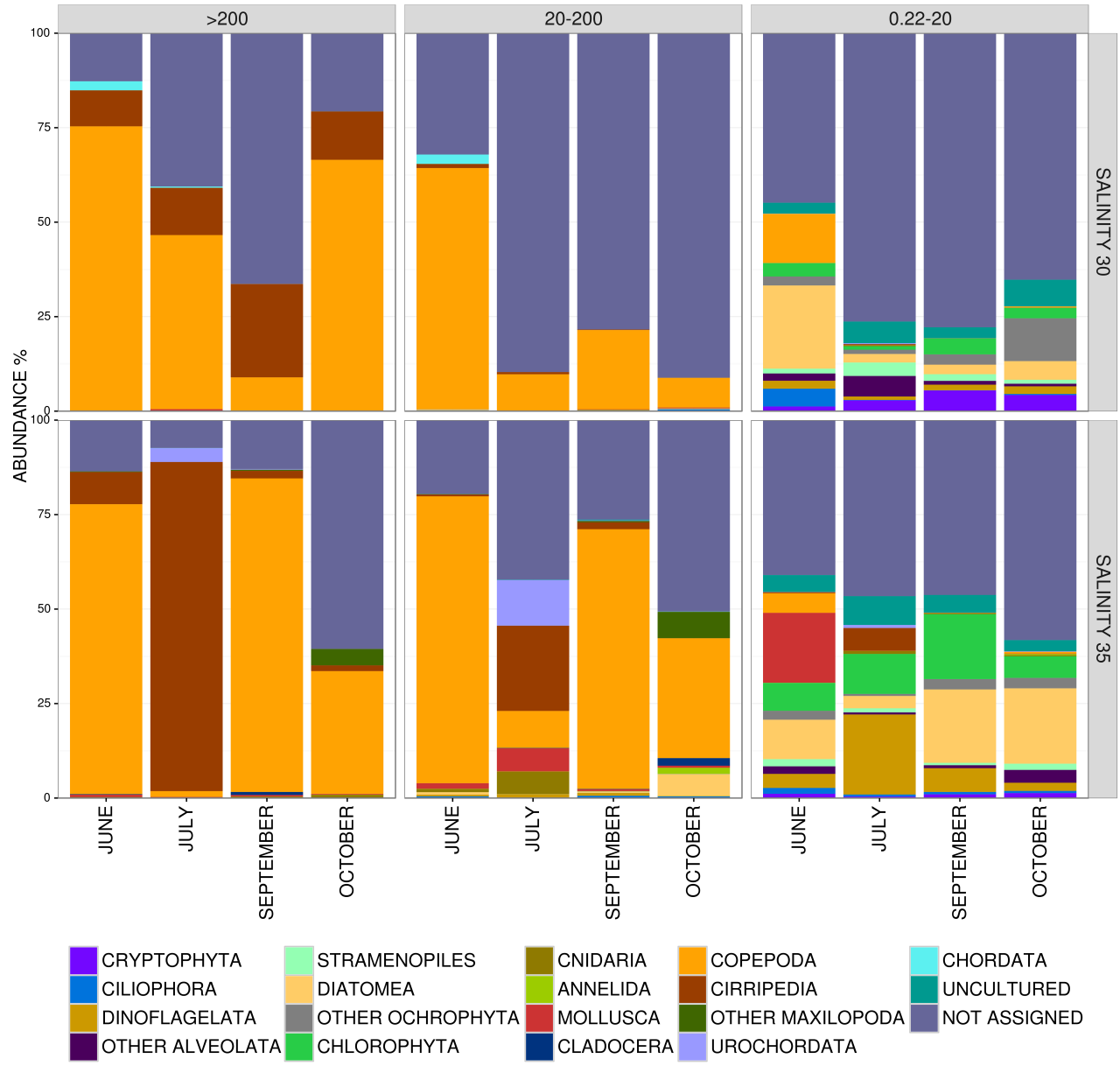


Fig. 1 Proportion of taxonomic ranks in each sample based on the metabarcoding approach. A total of 17 taxonomic ranks (>1% abundance) are shown. Samples are arranged by salinity (35 and 30 ppt) and plankton size-fraction (0.22-20, 20-200 and >200 µm)

Maxillopoda (mainly copepods and barnacles) predominated at the 20-200 and >200 μm size fractions, while a more diverse assemblage characterized the 0.22-20 μm one (Fig. 1). Copepods represented 2.3, 36 and 48.6% of the OTUs, while phytoplankton groups (e.g. Bacillariophyceae, Dinophyceae, Cryptophyceae) were 26.1, 1.5 and <0.1% of each size-fraction (0.22-20, 20-200 and >200 μm, respectively).

Metabarcoding vs. microscopy

The microscopic analysis identified 180 taxa for the Utermöhl method and 100 for the >200 μm zooplankton net. When the resulting taxa identified by metabarcoding and microscopy were compared, 44 of them were found in common. However, if the comparison was performed for the taxa with a >1% abundance in at least one of the samples (Table 3), only eleven taxa (three from phytoplankton and eight from zooplankton) were detected as abundant by both methodologies. Twelve taxa (six from phytoplankton and another six from zooplankton) were detected as abundant by microscopy but not as abundant in metabarcoding. Finally, two taxa (the diatom *Skeletonema menzellii* and the copepod *Centropages hamatus*) were detected as abundant by metabarcoding but not as abundant in microscopy.

While microscopy was unable to identify below genus level in *Thalassiosira*, metabarcoding was able to distinguish congeneric species (e.g. *T. allenii*, *T. guillardii*); the same occurred with the genus *Skeletonema* (*S. pseudocostatum*). Conversely, the microscopy-based analysis reported several species that were not identified using metabarcoding (e.g. *Apedinella radians, Teleaulax gracilis, Teleaulax minuta, Oithona davisae*).

The same spatial (salinity) and temporal (date; seasonal variation) trends were described for those species detected by both methodologies in the >200 μm size-fraction (Fig. 2a, b): while the higher dissolved oxygen (DO) and water transparency (SecchiDisk) values were associated with salinity, the highest precipitation could be associated with date (summer to autumn transition). The concentration of chlorophyll a (Chl-*a*) decreased with the advance of the season. However, in the 0.22-200 μm size-fraction neither approach identified a temporal pattern (Fig. 2c, d); a spatial pattern was discriminated only by microscopy (Fig. 2d).

METABARCODING	MICROSCOPY	ACRONYM	METABARCODING	MICROSCOPY	ACRONYM
	Centric diatoms	CENT		*Scenedesmus* spp.	SCEN*
	Chaetoceros curvisetus	CCUR		*Pyramimonas* spp.	PYRA
	Chaetoceros debilis	CDEB	*Ostreococcus tauri*		OTAU
	Chaetoceros socialis	CSOC	*Ulva intestinalis*		UINT
	Chaetoceros spp.	CHAE	Uncultured phytoplankton		UPHY
	Conticribra weissflogii	CWEI		Choanoflagellates	CHOA
Cyclotella choctawhatcheeana		CCHO		*Leucocryptos* spp.	LEUC
	Cyclotella meneghiniana	CMEN*	*Strombidium basimorphum*		SBAS
Leptocylindrus hargravesii		LHAR	*Chelophyes appendiculata*		CAPP
	L. danicus/hargravesii	LDAN		Unidentified Siphonophora	USIP
	Leptocylindrus aporus	LAPO*		*Evadne nordmanii*	ENOR*
	Leptocylindrus convexus	LCON*		*Evadne spinifera*	ESPI*
	Melosira varians	MVAR	*Podon* spp.		PODO
	Navicula radiosa	NRAD		*Podon intermedius*	PINT
	Pennate diatoms	PENN	*Balanus balanus*		BALA
	Proboscia alata	PALA	*Peltogaster paguri*		PPAG
	Pseudo-nitzschia multistriata	PMUL		Cirripedia nauplius larvae	CNAL
	Pseudo-nitzschia spp.	PSEU		Cirripedia cypris larvae	CCYL
	Skeletonema sp.	SKEL	*Acartia clausi*	*Acartia clausi*	ACLA*
Skeletonema menzellii		SMEN*	*Acartia tonsa*	*Acartia tonsa*	ATON*
Skeletonema pseudocostatum		SPSE		*Acartia* sp. (copepodite)	ASCO
Tenuicylindrus belgicus	*Tenuicylindrus belgicus*	TBEL*	*Calanipeda aquaedulcis*	*Calanipeda aquaedulcis*	CAQU*
	Thalassiosira sp.	THAL	*Calanus helgolandicus*		CHEL
Thalassiosira allenii		TALL	*Centropages hamatus*		CENT*
Thalassiosira delicatula		TDEL		*Cyclops* sp.	CYCL*
Thalassiosira guillardii		TGUI	*Euterpina acutifrons*	*Euterpina acutifrons*	EACU*
Thalassiosira lundiana		TLUN		*Oithona davisae*	OBRE
Heterosigma akashiwo		HAKA	*Oithona nana*	*Oithona nana*	ONAN*
	Apedinella radians	ASPI		*Oithona similis*	OSIM*
	Hemiselmis sp.	HEMI*		*Oncaea media*	OMED*
	Plagioselmis sp.	PLAG	*Paracalanus parvus*	*Paracalanus parvus*	PPAR*
Teleaulax acuta	*Teleaulax acuta*	TACU*		P-calanus (copepodite)	PCAL
Teleaulax amphioxeia	*Teleaulax amphioxeia*	TAMP*	*Pseudocalanus elongatus*		PELO
	Teleaulax gracilis	TGRA	*Pseudodiaptomus marinus*	*Pseudodiaptomus marinus*	PMAR*
	Teleaulax minuta	TMIN	*Temora longicornis*	*Temora longicornis*	TLON*
	Teleaulax spp.	TELE		Copepod nauplius	CNAU
	Chrysochromulina spp.	CHRY*		Unidentified brachiura larvae	UBRL
	Prymnesiales	PRYM	*Crassostrea gigas*		CGIG
	Gymnodiniales	GYMN	*Mytilus edulis*		MEDU
Gymnodinium aureolum		GAUR	*Littorina littorea*		LITT
	Gyrodinium flagellare	GFLA		Gastropod veliger larvae	GVEL
	Gyrodinium sp.	GYRO		Bivalve veliger larvae	BVEL
	Heterocapsa rotundata	HROT	Uncultured zooplankton		UZOO
	Heterocapsa sp.	HETE		*Oikopleura* sp.	OIKO*
	Katodinium spp.	KATO	*Sabellaria alveolata*		SALV
	Pfiesteria-like	PFIE		*Sagitta* sp.	SAGI
Micromonas pusilla		MPUS	*Scyliorhinus torazame*		STOR

Table 3 List of most abundant taxa from metabarcoding and microscopy. Only taxa with >1% abundance in at least one of the samples are shown. The acronyms listed here are used in the multivariate analysis. An asterisk marks those taxa identified by both methodologies, although not all of them are represented in the table (abundance lower than 1%)

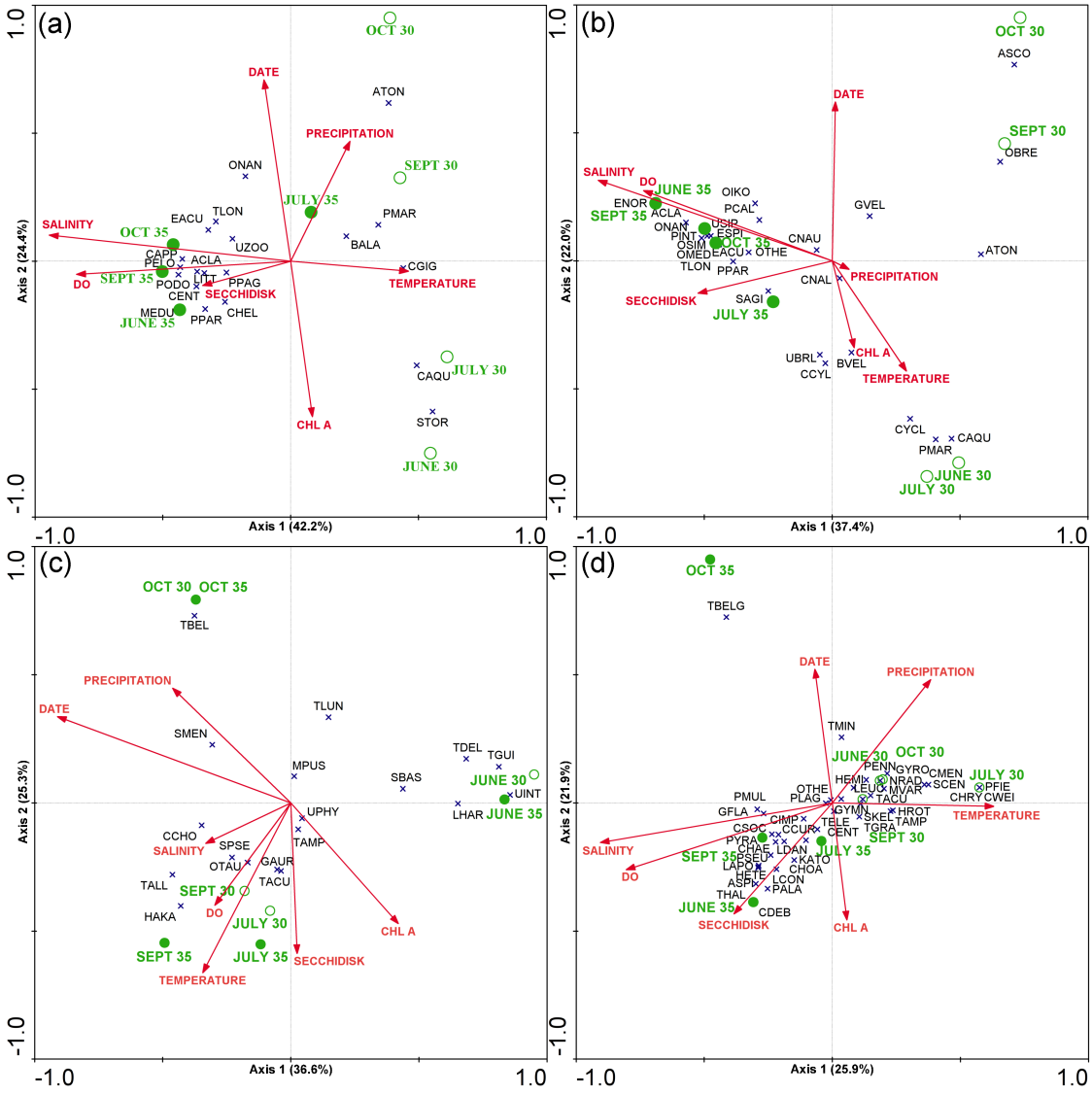


Fig. 2 Metabarcoding and microscopy CCA results. Only taxa with an abundance of 1% or higher in at least one sample were taken into account. Cross-marks identify taxa (see acronyms in Table 3). Sampling months are represented in green (35 ppt with filled circles). Environmental variables are showed as red arrows. (a) >200 µm metabarcoding, (b) >200 µm microscopy, (c) 0.22-200 µm metabarcoding and (d) 0.22-200 µm microscopy. Date is in Julian days

When comparing the relative abundance of all taxa within each particular sample obtained by both approaches (Table 4), significant correlations were reported in most *cases*. No difference was found between comparisons against microscopy-based counts or biomass (Table 4).

Fraction	Salinity (n)	Month	ρ (counts)	ρ (biomass)
>200	30 (6)	JUN	0.77*	0.89**
	30 (6)	JUL	0.95***	0.88*
	30 (6)	SEPT	0.65	0.65
	30 (6)	OCT	0.51	0.51
	35 (12)	JUN	0.63**	0.63**
	35 (12)	JUL	-0.27	-0.08
	35 (12)	SEPT	0.51*	0.58**
	35 (12)	OCT	0.52*	0.49*
0.22-200	30 (18)	JUN	0.48**	0.45*
	30 (18)	JUL	0.44*	0.48**
	30 (18)	SEPT	0.67***	0.69***
	30 (18)	OCT	0.75***	0.77***
	35 (25)	JUN	0.72***	0.73***
	35 (25)	JUL	0.55***	0.59***
	35 (25)	SEPT	0.58***	0.74***
	35 (25)	OCT	0.40**	0.44**

Table 4 Correlations between metabarcoding and microscopy-based analysis of community compositions. Spearman's rank correlation coefficient (ρ) and P-values are shown; $P < 0.01$ (***), $P < 0.05$ (**) and $P < 0.1$ (*). Relative abundances from metabarcoding were compared against both microscopy-based relative abundances and biomass. Each sample was defined by size-fraction (0.22-200 and >200 μm), salinity (30 and 35 ppt) and sampling month. Number of taxa (n) included in the correlations is specified after each salinity

Non-Indigenous Species (NIS)

We compared the performance of metabarcoding and microscopy to detect two NIS: *A. tonsa* and *P. marinus* in the >200 μm size fraction. While similar relative abundances were found for *A. tonsa* in the 30 ppt salinity by both approaches (Fig. 3a), it was only detected by metabarcoding in the 35 ppt salinity sample (Fig. 3b). Regarding *P. marinus*, metabarcoding was capable of detecting the species in all the samples, whilst its presence was detected by microscopy only in two of them (Fig. 3c, d). Finally, negative controls and extraction blanks showed no sequences corresponding to these two organisms giving further support to the herein reported data.

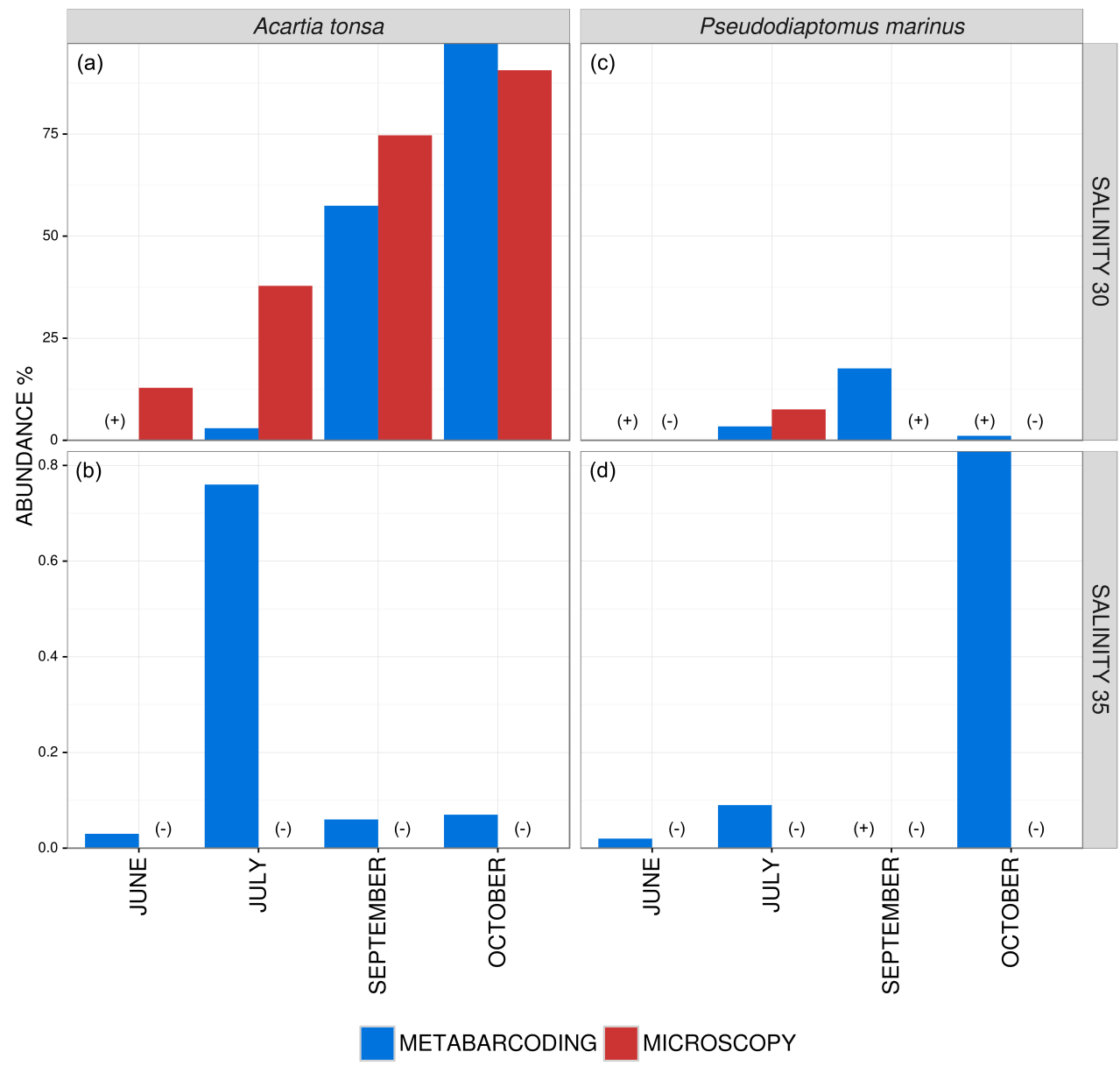


Fig. 3 Comparison of metabarcoding and microscopy when assessing two NIS. *Acartia tonsa* (a, b) and *Pseudodiaptomus marinus* (c, d) relative abundances in the >200 µm size fraction are divided by salinity (30 and 35 ppt). "+" stands for low detection percentages. "-" is showed when the species was not detected. Note that the y-axis scale is different for each salinity

Discussion

The Marine Strategy Framework Directive establishes a framework for marine environmental policy of the European exclusive economic zone (Ferreira et al. 2011), including a series of indicators that should be monitored to achieve the descriptors. Metabarcoding is especially valuable for some of these indicators, such as NIS management or biodiversity assessment (Bourlat et al. 2013). The performance of metabarcoding in monitoring plankton species, including two NIS, was critically assessed in the present study by comparing results with those of classical taxonomic analysis (microscopy). Our results indicate that metabarcoding is a promising alternative to traditional methods for early detection of NIS and plankton biodiversity shifts. Metabarcoding can be a useful tool for implementation in environmental policies, including the timely design of appropriate adaptation/mitigation measures. However, until more complete reference DNA sequence databases are publicly available, microscopic analysis should be performed in parallel at least for representative samples.

NIS detection

Although the high sensitivity of metabarcoding has been described elsewhere (e.g. Zhan et al. 2013; Pochon et al. 2013), its application for monitoring biological invasions has only been recently demonstrated (Zaiko et al. 2015c). In the present work, we confirm the suitability of metabarcoding for early detection of NIS at extremely low relative abundances (Fig. 3). The reasons behind this are: 1) the ability to analyze bigger sample volumes compared to microscopy-based methods, for which screening the whole sample requires great time and effort; and 2) the capacity to detect the presence of individuals at early life stages, such as eggs or nauplius larvae, whose identification is complicated with traditional methods (Comtet et al. 2015).

In this sense, plankton monitoring programs are not usually designed to provide an early-warning alert of NIS. However, the sensitivity of metabarcoding, combined with the relatively low time and cost associated to this technique (Kelly et al. 2014), results in a promising alternative approach for the rapid detection of plankton biodiversity shifts, opening the possibility for its implementation in environmental policies. As an example, Zaiko (Zaiko et al. 2015a-b) recently suggested the value of combining metabarcoding with current taxonomic analysis for the surveillance and management of ballast water, the main vector of most marine NIS introductions.

Quantitative nature of metabarcoding

Discrepancies between metabarcoding and microscopy-based relative abundances or biomass have been reported for plankton assemblages (Hirai et al. 2015; Massana et al. 2015; Stoeck et al. 2014; Sun et al. 2015), but studies evaluating the quantitative nature of this technique are still scarce. When comparing the relative abundances of all taxa within each sample, we showed that metabarcoding and microscopy data were correlated in most cases (Table 4). Lack of correlation could be explained by technical biases introduced during the DNA extraction (Roh et al., 2006), for which the method's performance can vary with organism type or even development stage, or PCR amplification step (Gonzalez et al. 2012), with a differential amplification that favors abundant taxa. But the Copy Number Variation (CNV) associated to rDNA has been suggested as one of the main factors affecting the quantitative value of metabarcoding (Kembel et al. 2012): incorporating CNV to the analysis can help to improve abundance estimates. There are also reported correlations between CNV and genome size in eukaryotes (Prokopowich et al. 2003) and, between CNV and cell length and biovolume in unicellular organisms (Zhu et al. 2005; Godhe et al. 2008), suggesting a potential way of addressing this drawback in eukaryotes. In the meantime, metabarcoding targeting multi-copy genes will remain as a semi-quantitative approach (Amend et al. 2010; Albaina et al. 2016a).

Metabarcoding for community ecology

Metabarcoding analysis of the plankton community replicated the temporal and spatial trends of the Bilbao estuary observed in the morphological (microscopic) analyses better for zooplankton than for phytoplankton. As expected, the main trends driving the community in the estuary of Bilbao are date (seasonal variation) and salinity (Uriarte and Villate, 2004). This somewhat reduced performance in the lowest size fractions (also shown in Figure 2) is probably related to the deficit of representative sequences for these organisms in current databases.

Metabarcoding was able to overcome the lack of resolution of microscopy for picoplankton (0.2-2 μm): among the most abundant OTUs, the smallest size fraction was represented by the mamiellophyceans *Micromonas pusilla* and *Ostreococcus tauri* (Table 3 and ESM_4), which are known as important components of the picoplankton in temperate waters (Romari and Vaulot 2004). Regarding the taxa that were only identified in the microscopy-based analysis (Table 3) there could be two possible explanations for their absence in the metabarcoding analysis: the taxa had no representative sequence in the database (e.g. *Teleaulax gracilis*, *Oithona davisae*) or the taxa were

present but not assigned (e.g. *Apedinella radians*, *Teleaulax minuta*). In relation to the latter, this could happen if the V9 region is missing or incomplete in the available representative sequence (*Teleaulax minuta*) or due to a possible local variability or misidentification (*Apedinella radians*). In this sense, the addition of local species with no previous representation in the database (as demonstrated in this study; Table 2) significantly increases the assignment success for locally collected field samples and is recommended when designing metabarcoding studies (e.g. Cowart et al. 2015).

Table 3 also showed that among the most abundant taxa identified by microscopy there were plankton developmental stages, such as copepodites or larvae. In the metabarcoding analysis those organisms would be assigned to a certain taxonomy, regardless the developmental stage. For example, the gastropod veliger larvae observed at the microscope could correspond to the benthic *Littorina littorea* or the Cirripedia nauplius/cypris larvae to *Balanus balanus* or *Peltogaster paguri*. This shows the taxonomic potential of metabarcoding versus microscopy and at the same time its limitation to be employed in studies where developmental stages need to be assessed.

Finally, a thorough revision of the OTUs uncovered surprising assignments such as the case of *Scyliorhinus torazame* (cloudy cat shark; Table 3 and ESM_4), relatively abundant in several samples (always higher in 30 than 35 ppt), which may be due to the incompleteness of the reference DNA sequence database and/or suboptimal taxonomic resolution of 18S V9 for this organism. In the same way, *Chelophyes appendiculata* (Table 3 and ESM_4) might most likely be *Muggiaea* spp., which are the only siphonophore species reported in the estuaries and coastal waters of the studied area (Villate et al. 2004). This calls for caution when reporting previously undetected species (such as NIS) using metabarcoding, because they may correspond to incorrectly assigned local species; further analysis should be performed to confirm the results.

Conclusion

Metabarcoding identified spatial and temporal trends similar to those resulting from morphological (microscopic) taxonomic analysis for zooplankton, but not for phytoplankton, probably due to the lack of representative sequences for the latter group of organisms in current databases. The addition of representative sequences from local species resulted in an improvement in taxonomic assignment success, highlighting the need for completing reference sequence databases in order to overcome these limitations. There was a high correspondence between this approach and microscopy-based abundances, suggesting the capacity of metabarcoding for semi-quantitative analysis of some taxonomic groups. Regarding the taxonomic resolution issue, while 18SV9 metabarcoding gives a broader range of taxa, its species-level resolution is not complete: a possible solution would include combining the results of 18S V9 with the 18S V1-V2 (avoiding therefore the introduction of a distinct Copy Number Variation bias) or with a high-resolution marker such as COI, for better discrimination between species. Furthermore, the superior sensitivity of this approach allowed the identification of Non-Indigenous Species at abundances barely detectable by microscopy. In conclusion, we think that metabarcoding is a rapid and cost-effective assessment tool that can be useful for the timely detection of NIS, which may allow the prevention or mitigation of their effects, and plankton biodiversity shifts.

Acknowledgements

The authors thank Ann Bucklin (University of Connecticut) for her comments on the manuscript. The authors also thank the SGIker (UPV/EHU) for the technical and human support provided, and the Hydro-meteorology Service of the Regional Council of Bizkaia for the precipitation data.

Compliance with Ethical Standards

Funding

AA and DA work was supported by a contract with the Euskampus Foundation (Euskampus Fundazioa) and a Basque Government doctoral fellowship (UPV/EHU "ZabaldUz" program), respectively. This study was funded by a Basque Government grant "Grupo de Investigación Consolidado del Sistema Universitario Vasco" to support activities of the Genomic Resources Research Group (IT-558-10), the Phytoplankton (IT-699-13) and Zooplankton (IT-778-13) Ecology Groups. Sampling was also partly financed by the Euskampus Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with animals performed by any of the authors.

Archiving of Data

Metabarcoding data (quality filtered, chimera-free merged reads) are available at Qiita repository (https://qiita.ucsd.edu/; ID 10518).

References

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

Albaina A, Uriarte I, Aguirre M, Abad D, Iriarte A, Villate F, Estonba A. (2016b) Insights on the origin of invasive copepods colonizing Basque estuaries; a DNA barcoding approach. *Mar Biodivers Rec*, in press.

Albaina A, Villate F, Uriarte I. (2009) Zooplankton communities in two contrasting Basque estuaries (1999–2001): reporting changes associated with ecosystem health. *J. Plankton Res*, 31: 739–752.

Aljanabi SM, Martinez I (1997) Universal and rapid salt extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res,* 25, 4692−4693.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic local alignment search tool. J Mol Biol, 215(3):403–410.

Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Mol Ecol*, 19, 5555-5565.

Amorim Visco J, Apothoz-Perret-Gentil L, Cordonier A, Esling P , Pillet L, Pawlowski J. (2015) Environmental monitoring: inferring diatom index from next-generation sequencing data. *Environ Sci Technol*, 49: 7597-7605

Aravena, G., Villate, F., Uriarte, I., Iriarte, A. and Ibáñez, B. (2009) Response of Acartia populations to environmental variability and effects of invasive congenerics in the estuary of Bilbao, Bay of Biscay. *Est. Coast. Shelf Sci.*, 83, 621-628.

Bachy C, Dolan JR, López-García P, Deschamps P, Moreira D. (2013) Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. *ISME J*, 7(2): 244–255.

Baird DJ, Hajibabaei M. (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by nextgeneration DNA sequencing. *Mol Ecol,* 21: 2039-2044.

Båmstedt,U. (1986) Chemical composition and energy content. In Corner,E.D.S. and O'Hara,S.C.M. (eds), *The Biological Chemistry of Marine Copepods.* Clarendon, Oxford, pp. 1–58.

Borja A, Elliott M, Carstensen J, Heiskanen AS, van de Bund W. (2011) Marine management - Towards an integrated implementation of the European Marine Strategy Framework and the Water Framework Directives. *Mar Pollut Bull*, 60: 2175-2186.

Borja A, Muxika I, Franco J. (2006) Long-term recovery of soft-bottom benthos following urban and industrial sewage treatment in the Nervión estuary (southern Bay of Biscay). *Mar Ecol Prog Ser*, 313: 43-55.

Bourlat SJ, Borja A, Gilbert J, Taylor MI, Davies N, Weisberg SB, Griffith JF, Lettierih T, *et al*. (2013) Genomics in marine monitoring: new opportunities for assessing marine health status. *Mar Pollut Bull*, 74(1):19-31.

Bricker SB, Ferreira JG, Simas T. (2003) An integrated methodology for assessment of estuarine trophic status. *Ecol Modell*, 169: 39-60.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.,*7(5):335-6. doi: 10.1038/nmeth.f.303

Chen G, Hare MP (2008) Cryptic ecological diversification of a planktonic estuarine copepod, *Acartia tonsa*. *Mol Ecol*, 17, 1451–1468.

Comtet T, Sandionigi A, Viard F, Casiraghi M. (2015) DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biol Invasions*, 17: 905-922.

Cowart DA, Pinheiro M, Mouchel O, Maguer M, Grall J, Miné J, Arnaud-Haond S. (2015)

Metabarcoding Is Powerful yet Still Blind: A Comparative Analysis of Morphological and Molecular Surveys of Seagrass Communities. *PLoS ONE*, 10, e0117562.

de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E , Berney C *et al*. (2015) Ocean Plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science* 348, 1261605–1261605. doi:10.1126/science.1261605

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460–2461.

Edgar RC, Haas BJ, Clememte JC, Quince C, Knight R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27, 2194–2200.

Edler L, Elbrächter M. (2010) The Utermöhl method for quantitative phytoplankton analysis. In: Microscopic and molecular methods for quantitative phytoplankton analysis. IOC Manuals and Guides.

Eiler A, Drakare S, Bertilsson S, Pernthaler J, Peura S, Rofner C, Simek K, Yang Y, Znachor P, Lindström ES. (2013) Unveiling Distribution Patterns of Freshwater Phytoplankton by a Next Generation Sequencing Based Approach. *PLoS ONE*, 8, e53516.

Ferreira JG, Andersen JH, Borja A, Bricker SB, Camp J, Cardoso da Silva M, Garcés E, Heiskaneng AS, *et al*. (2011) Overview of eutrophication indicators to assess environmental status within the European Marine Strategy Framework Directive. *Estuar Coast Shelf Sci*, 93: 117-131.

Gaudy R. and Boucher J. (1983) Relation between respiration, excretion (ammonia and inorganic phosphorus) and activity of amylase and trypsin in different species of pelagic copepods from an Indian Ocean equatorial area. *Mar Biol*, 75: 37-45.

Gilbert JA, Steele JA, Caporaso JG, Steinbrük L, Reeder J, Temperton B, Huse S, McHardy AC *et*

al. (2012) Defining seasonal marine microbial community dynamics. *ISME J*, 6: 298-308.

Godhe A, Asplund ME, Härnström K, Saravanan V, Tyagi A, Karunasagar I. (2008) Quantification of Diatom and Dinoflagellate Biomasses in Coastal Marine Seawater Samples by Real-Time PCR. *Appl Environ Microbiol*, 74, 7174–7182.

Gonzalez JM, Portillo MC, Belda-Ferre P, Mira A. (2012) Amplification by PCR Artificially Reduces the Proportion of the Rare Biosphere in Microbial Communities. *PLoS ONE*., 7(1):e29973. doi:10.1371/journal.pone.0029973.

Hall BG (2013) Building Phylogenetic Trees from Molecular Data with MEGA. *Mol Biol Evol* 30(5), 1229–1235.

Herlemann DP, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, Andersson AF. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J*, 5: 1571-1579.

Hirai J, Kuriyama M, Ichikawa T, Hidaka K, Tsuda A. (2015) A metagenetic approach for revealing community structure of marine planktonic copepods. *Mol Ecol Res*, 15, 68-80.

Jeffrey SW, Mantoura RFC. (1997) Development of pigment methods for oceanography: SCOR-supported working groups and objectives. In: Phytoplankton pigments in ceanography: guidelines to modern methods. Monographs on Oceanographic Methodology. UNESCO Publishing; 19-36.

Joshi NA, Fass JN (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.

Kelly RP, Port JA, Yamahara KM, Martone RG, Lowell N, Thomsen PF, Mach ME, Bennett M *et al*. (2014) Environmental monitoring. Harnessing DNA to improve environmental management. *Science*, 344:1455-1456.

Kembel SW, Wu M, Eisen JA, Green JL. (2012) Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput Biol*, 8(10):e1002743. doi:

10.1371/journal.pcbi.1002743.

Laakmann S, Gerdts G, Erler R, Knebelsberger T, Martínez Arbizu P, Raupach MJ. (2013) Comparison of molecular species identification for North Sea calanoid copepods (Crustacea) using proteome fingerprints and DNA sequences. *Mol Ecol Res*, 13: 862-876.

Lindeque PK, Hay SJ, Heath MR, Ingvarsdottir A, Rasmussen J, Smerdon GR, Waniek JJ. (2011) Integrating conventional microscopy and molecular analysis to analyse the abundance and distribution of four Calanus congeners in the North Atlantic. *J Plankton Res,* 28: 221-238.

Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, Atkinson A. (2013) Next Generation Sequencing Reveals the Hidden Diversity of Zooplankton Assemblages. *PLoS ONE.*, 8(11): e81327.

Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R, Claverie JM, Decelle J *et al*. (2014) Patterns of rare and abundant marine microbial eukaryotes. *Curr Biol*, 24: 813-821.

Mahé F, Mayor J, Bunge J, Chi J, Siemensmeyer T, Stoeck T, Wahl B, Paprotka T *et al*. (2015) Comparing High-throughput Platforms for Sequencing the V4 Region of SSU-rDNA in Environmental Microbial Eukaryotic Diversity Surveys. *J Eukaryot Microbiol*, 62: 338-45.

Massana R, Gober A, Audic S, Bass D, Bittner L, Boutte C, Chambouvet A, Christen R *et al*. (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environ Microbiol,* doi: 10.1111/1462-2920.12955.

Navas-Molina JA, Peralta-Sánchez JM, González A, McMurdie PJ, Vázquez-Baeza Y, Xu Z, Ursell LK, Lauber C, Zhou H, Song SJ, Huntley J, Ackermann GL, Berg-Lyons D, Holmes S, Caporaso JG, Knight R. (2013) Advancing our understanding of the human microbiome using QIIME. *Methods Enzymol*, 531: 371–444.

Olenina I, Hajdu S, Edler L, Andersson A, Wasmund N, Busch S, Göbel J, *et al*. (2006) Biovolumes and size-classes of phytoplankton in the Baltic Sea HELCOM. *Balt Sea Environ Proc*, 106.

Pochon X, Bott NJ, Smith KF, Wood SA. (2013) Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PLoS ONE*., 8: e73935.

Prokopowich CD, Gregory TR, Crease TJ. (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, 46: 48–50.

Quail M, Smith ME, Coupland P *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics*, 13, 341.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*, 41: 590-596.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Roh, C., F. Villatte, B.-G. Kim, and R. D. Schmid. (2006) Comparative study of methods for extraction and purification of environmental DNA from soil and sludge samples. *Appl. Biochem Biotechnol*, 134**:**97-112.

Romari K, Vaulot D. (2004) Composition and temporal variability of picoeukaryote communities at a coastal site of the English channel from 18SrDNA sequences. *Limnol Oceanogr*, 49: 784-798.

Smith D (2012) fastq-barcode.pl. [Software]. Available at https://gist.github.com/dansmith01/3920169.

Stoeck T, Bass D, Nebel M, Christen R, Jones MD, Breiner HW Richards TA. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol,* 19 Suppl 1:21-31.

Stoeck T, Breiner HW, Filker S, Ostermaier V, Kammerlander B, Sonntag B. (2014) A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. *Environ Microbiol*, 1: 430−444.

Sun C, Zhao Y, Li H, Dong Y, MacIsaac HJ, Zhan A. (2015) Unreliable quantitation of species abundance based on high-throughput sequencing data of zooplankton communities. *Aquat Biol*, 24:9-15.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M. Kumar S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*, 28, 2731-2739.

Taylor, AH; Allen, JI; Clark, PA. (2002) Extraction of a weak climatic signal by an ecosystem. *Nature*, 416: 629-632.

ter Braak CJF, Smilauer P (2002) *CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination (Version 4.5)*. Microcomputer Power, Ithaca.

Uriarte I, Villate F, Iriarte A, Duque J, Ameztoy I. (2014) Seasonal and axial variations of net water circulation and turnover in the estuary of Bilbao. *Estuar Coast Shelf Sci,*150: 312-324.

Uriarte I, Villate F, Iriarte A. (2015) Zooplankton recolonization of the inner estuary of Bilbao: influence of pollution abatement, climate and non-indigenous species. *J. Plankton Res.* doi: 10.1093/plankt/fbv060

Uriarte I and Villate F. (2004) Differences in the abundance and distribution of copepods in two estuaries of the Basque coast (Bay of Biscay) in relation to pollution. *J Plankton Res*, 27(9): 863-974.

Villate F, Iriarte A, Uriarte I, Intxausti L, de la Sota A. (2013) Dissolved oxygen in the rehabilitation phase of an estuary: influence of sewage pollution abatement and hydro-climatic factors. *Mar Pollut Bull*, 70(1-2):234-46.

Villate F, Uriarte I, Irigoien X, Beaugrand G, Cotano U. (2004) Zooplankton communities. In: Oceanography and marine environment of the Basque country. Elsevier, 395-423.

Ward BA, Dutkiewicz SA, Jahn O, Follows MJ. (2012) A size-structured food-web model of the global ocean. *Limnol Oceanogr*, 57: 1877-1891.

Zaiko A, Martinez JL, Ardura A, Clusa L, Borrell YJ, Samuiloviene A, Roca A, Garcia-Vazquez E. (2015a) Detecting nuisance species using NGST: Methodology shortcomings and possible application in ballast water monitoring. *Mar Environ Res,* pii: S0141-1136(15)30008-8.

Zaiko A, Martinez JL, Schmidt-Petersend J, Ribicic D, Samuiloviene A, Garcia-Vazquez E. (2015b) Metabarcoding approach for the ballast water surveillance – An advantageous solution or an awkward challenge? *Mar Pollut Bull*, 92(1-2): 25-34.

Zaiko A, Samuiloviene A, Ardura A, Garcia-Vazquez E. (2015c) Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. *Mar Pollut Bull*, 100(1):53-59. doi: 10.1016/j.marpolbul.2015.09.030.

Zhan A, Bailey SA, Heath DD, MacIsaac HJ. (2014) Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. *Mol Ecol Res*, 14: 1049-1059.

Zhan A, Hulak M, Sylvester F, Huang X. Adebayo AA, Abbott CL, Adamowicz SJ, Heath DD *et al*. (2013) High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods Ecol Evol*, 4: 558-565.

Zhang J, Kobert K, Flouri T, Stamatakis A. (2014) PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30: 614-620.

Zhu F, Massana R, Not F, Marie D, Vaulot D. (2005) Mapping of picoeukaryotes in marine ecosystems with a quantitative PCR of the 18S rRNA gene. *FEMS Microbiol Ecol,* 52(1): 79-92

ESM_1 GenBank accession numbers

Species	GenBank Accession Number	Source
Acartia tonsa	KP768152	Present study
Acartia clausi	KR919781	Present study
Calanipeda aquaedulcis	KP768157	Present study
Calanus helgolandicus	KP768135	Albaina *et al.* 2016
Euterpina acutifrons	KR919779	Present study
Pseudodiaptomus marinus	KP768155	Present study
Evadne nordmanni	KR919787	Present study
Evadne spinifera	KR919783	Present study
Podon spp.	KR919785	Present study
Acartia bifilosa	JX995289	Laakmann *et al.* 2013
Centropages typicus	JX995296	Laakmann *et al.* 2013
Eurytemora affinis	JX995299	Laakmann *et al.* 2013
Isias clavipes	JX995302	Laakmann *et al.* 2013
Anomalocera patersoni	JX995305	Laakmann *et al.* 2013
Temora longicornis	JX995308	Laakmann *et al.* 2013
Paracalanus parvus	JX995311	Laakmann *et al.* 2013
Pseudocalanus elongatus	JX995319	Laakmann *et al*. 2013
Pseudocalanus moultoni	JX995322	Laakmann *et al.* 2013
Oithona brevicornis	JF288757	GenBank
Oithona nana	HQ008734	GenBank
Oithona similis	KF153700	GenBank
Oncaea media	AM114421	GenBank

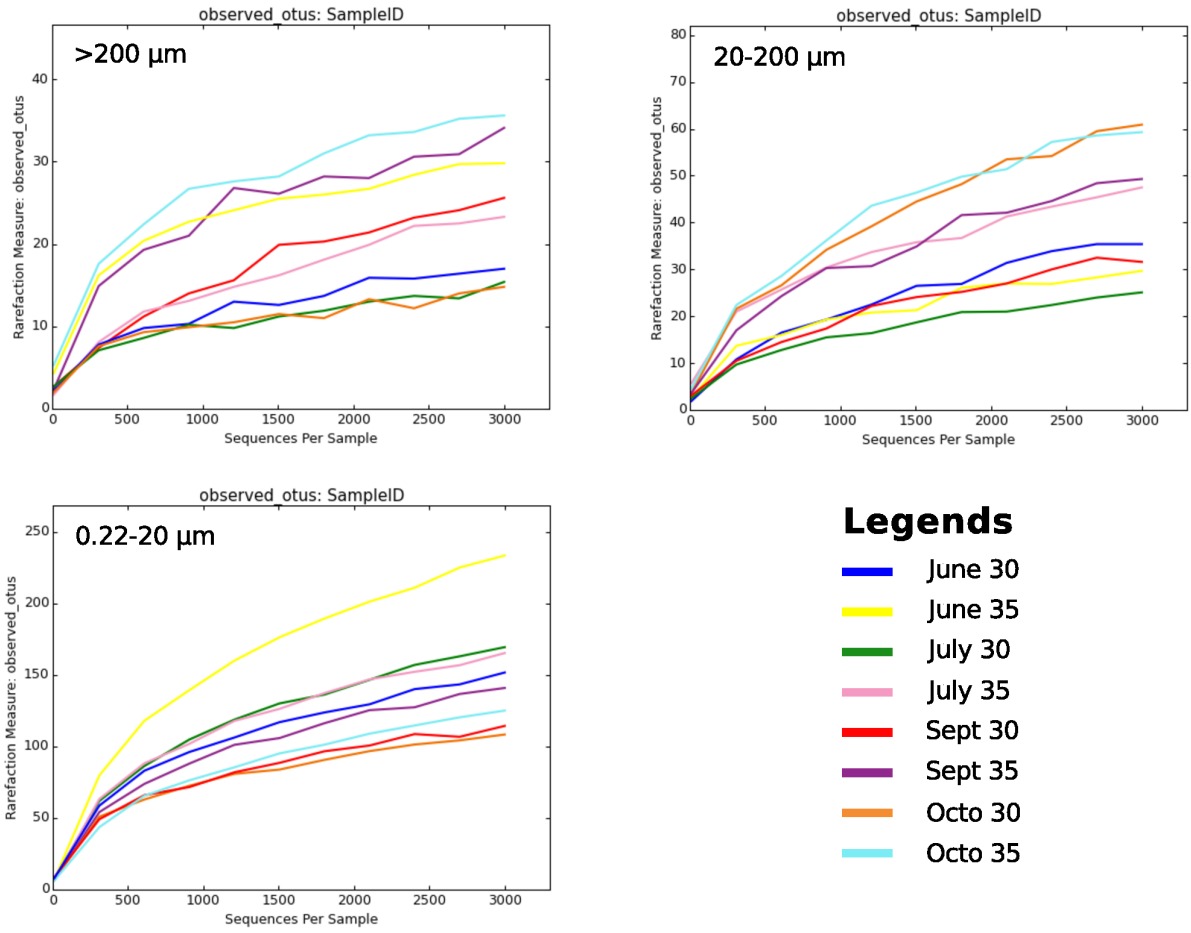
Source details for every taxon used to create the custom database is also included.

ESM_2 List of taxa used for the comparison between microscopy and metabarcoding

	SALINITY 30				SALINITY 35			
OTU ID	**JUNE**	**JULY**	**SEPT**	**OCTO**	**JUNE**	**JULY**	**SEPT**	**OCTO**
Acartia clausi	-	-	-	-	30.53	5.76	11.02	17.77
Acartia tonsa	12.86	37.83	74.69	90.63	0.00	0.00	0.00	0.02
Corycaeus sp.	-	-	-	-	0.00	1.44	0.22	0.26
Cyclops sp.	3.54	0.00	0.00	0.00	-	-	-	-
Calanipeda aquaedulcis	80.39	54.19	2.23	0.00	0.01	0.00	0.00	0.00
Centropages typicus	-	-	-	-	10.18	0.72	0.90	0.00
Euterpina acutifrons	-	-	-	-	0.00	2.16	3.81	7.80
Oithona nana	1.93	0.20	23.00	9.37	0.00	0.72	35.21	4.95
Oithona similis	-	-	-	-	14.09	3.60	0.00	1.30
Oncaea media	-	-	-	-	0.00	5.04	14.13	7.02
Other calanoids	-	-	-	-	44.81	80.58	34.71	59.35
Paracalanus parvus	0.64	0.20	0.00	0.00	-	-	-	-
Pseudodiaptomus marinus	0.00	7.57	0.08	0.00	0.00	0.00	0.01	0.00
Temora sp.	-	-	-	-	0.39	0.00	0.00	1.56
Pseudopediastrum boryanum	0.00	0.36	0.14	0.00	-	-	-	-
Central diatoms	1.84	0.01	4.01	6.14	9.74	0.00	7.70	0.01
Chaetoceros sp.	0.00	0.00	0.24	0.02	7.46	4.88	12.43	0.82
Conticribra weissflogii	0.00	13.93	0.00	0.00	-	-	-	-
Coscinodiscus sp.	-	-	-	-	0.02	0.02	0.00	0.00
Cyclotella meneghiniana	0.00	6.43	0.15	7.81	-	-	-	-
Dinophysis sp.	-	-	-	-	0.00	0.01	0.01	0.00
Gymnodinium sp.	9.22	5.36	5.11	0.56	3.57	12.71	2.09	1.46
Gyrodinium sp.	1.22	4.29	2.92	0.02	3.75	1.91	3.80	1.18
Hemiselmis sp.	-	-	-	-	0.00	0.00	2.61	0.20
Heterocapsa sp.	0.00	11.79	1.09	0.00	0.88	4.96	2.22	0.00
Heterosigma akashiwo	-	-	-	-	0.00	0.55	0.65	0.00
Karenia sp.	-	-	-	-	0.00	0.01	0.00	0.00
Leptocylindrus sp.	0.07	0.00	0.14	0.00	6.62	1.26	10.36	0.05
Minidiscus sp.	-	-	-	-	0.00	0.00	0.26	0.00
Navicula sp.	2.32	0.00	0.00	0.00	-	-	-	-
Nitzschia sp.	0.00	0.00	0.36	0.02	0.04	0.07	0.33	0.03
Pennate diatoms	80.55	6.54	1.47	14.18	1.32	1.65	0.39	1.76
Prymnesiophyceae	0.00	3.22	8.02	1.67	8.38	56.26	9.84	1.76
Pseudo-nitzschia sp.	1.11	0.01	0.00	0.03	40.36	0.08	16.58	0.20
Rhizosolenia sp.	-	-	-	-	0.00	0.46	0.52	0.01
Rhodomonas sp.	-	-	-	-	0.00	0.55	0.00	0.00
Scenedesmus sp.	0.00	3.30	0.05	2.92	-	-	-	-
Skeletonema sp.	3.67	4.03	0.07	10.60	0.18	0.17	8.49	0.00
Striatella unipunctata	-	-	-	-	0.00	0.00	0.00	0.01
Teleaulax/Plagioselmis sp.	0.00	40.73	76.22	55.78	10.14	13.79	15.41	1.18
Tenuicylindrus belgicus	0.00	0.00	0.00	0.27	0.00	0.00	4.96	91.35
Tetraselmis sp.	-	-	-	-	0.00	0.00	0.13	0.00
Thalassiosira sp.	-	-	-	-	7.53	0.00	1.18	0.00
Warnowia sp.	-	-	-	-	0.00	0.64	0.04	0.00

Microscopy-based relative abundances for each sample are also included

ESM_3 Rarefaction curves for the different plankton size-fractions



Alpha rarefaction plot generated with QIIME at the 99% identity threshold. Observed OTUs (left axis) plotted against sequencing depth (bottom axis; limited to the first 3,000 reads)

ESM_4 List of most abundant (> 1 % of the total 18S V9 reads) OTUs across the different plankton size-fractions in the metabarcoding approach. "+" stands for percentages higher than zero but below 0.01. "-" symbol corresponds to zero.

		>200 µm							20-200 µm							0.22-20 µm						
OTU ID	ACRONYM	JUNE 30	JUNE 35	JULY 30	JULY 35	SEPT 30	SEPT 35	OCTO 30	OCTO 35	JUNE 30	JUNE 35	JULY 30	JULY 35	SEPT 30	SEPT 35	OCTO 30	OCTO 35	JUNE 30	JUNE 35	JULY 30	JULY 35	SEPT 30	SEPT 35	OCTO 30	OCTO 35
Strombidium cf. basimorphum	SBAS	-	-	-	-	-	-	-	-	+	0,01	+	+	-	0,01	-	-	3,32	0,46	0,01	0,09	0,01	0,19	0,03	0,10
Gymnodinium aureolum	GAUR	-	-	-	+	+	+	-	-	-	-	-	+	+	0,03	+	+	0,40	0,54	0,07	10,54	1,06	2,04	0,31	0,25
Cyclotella choctawhatcheeana	CCHO	-	-	-	-	+	+	+	-	-	+	-	+	0,15	0,01	0,04	0,20	+	-	0,03	0,01	0,80	2,21	1,23	0,17
Skeletonema menzelliii	SMEN	-	-	-	+	+	-	-	+	-	-	-	+	+	0,03	0,01	0,02	0,01	+	0,12	0,18	0,89	1,02	2,15	2,82
Skeletonema pseudocostatum	SPSE	-	-	-	+	-	-	-	-	-	+	-	+	-	0,07	+	-	0,01	+	0,88	2,24	0,01	1,08	0,05	0,27
Thalassiosira allenii	TALL	-	-	-	+	-	+	-	-	-	-	-	-	-	0,03	+	-	-	-	-	+	0,01	12,77	0,06	0,20
Thalassiosira delicatula	TDEL	-	-	-	+	-	-	-	-	-	0,05	-	-	-	-	-	+	0,61	1,06	-	-	-	-	-	0,02
Thalassiosira guillardii	TGUI	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	9,20	0,08	0,01	+	-	-	0,01	0,02
Thalassiosira lundiana	TLUN	-	-	-	+	-	-	-	-	+	+	-	-	-	+	0,01	1,00	1,23	0,86	0,01	+	0,03	0,14	0,12	0,36
Leptocylindrus hargravesii	LHAR	-	-	-	+	-	-	-	-	-	0,15	-	+	-	-	-	-	3,36	2,61	-	0,20	-	0,06	+	0,01
Tenuicylindrus belgicus	TBEL	-	-	-	+	-	-	-	-	-	-	-	-	-	0,01	0,05	3,77	-	-	-	-	-	0,15	0,56	14,21
Heterosigma akashiwo	HAKA	-	-	-	+	-	-	-	-	-	-	-	-	-	0,01	-	-	-	-	-	0,06	-	1,94	+	0,01
Micromonas pusilla	MPUS	-	-	-	+	-	-	-	-	-	-	-	-	+	-	+	-	0,70	0,94	0,14	0,89	0,37	0,57	0,44	1,45
Ostreococcus tauri	OTAU	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	0,37	0,51	0,35	7,44	1,03	13,87	0,09	0,83
Ulva intestinalis	UINT	+	+	+	+	0,01	+	+	-	+	-	-	-	+	-	+	-	0,17	3,35	-	+	+	-	-	-
Teleaulax amphioxeia	TAMP	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	0,16	0,23	0,01	0,01	2,40	0,14	0,49	0,07
Teleaulax acuta	TACU	-	-	-	-	-	-	-	-	-	-	-	-	+	-	0,01	-	0,28	0,05	2,48	0,30	1,34	0,46	2,99	0,04
Uncultured phytoplankton	UPHY	-	+	-	-	-	+	-	-	+	0,11	-	0,49	0,01	0,31	0,01	0,02	11,35	12,50	13,76	17,93	9,08	9,06	19,52	8,42
Sabellaria alveolata	SALV	-	-	-	-	-	+	-	-	-	0,03	-	-	-	-	-	1,52	-	-	-	-	-	-	-	0,16
Podon sp.	PODO	-	0,11	+	0,01	+	0,67	-	0,01	-	+	-	0,11	-	0,10	-	2,01	-	-	-	-	-	-	-	-
Acartia clausi	ACLA	0,27	36,80	0,05	1,07	0,17	69,38	0,33	15,16	0,11	57,43	-	2,97	0,01	40,59	+	1,88	0,09	4,84	-	+	0,00	-	+	0,03
Acartia tonsa	ATON	0,11	0,02	1,36	0,01	4,90	0,05	63,46	0,02	0,11	0,00	0,16	-	5,80	0,01	1,42	+	-	-	0,01	-	0,02	+	0,23	-
Balanus balanus	BALA	9,43	4,68	12,45	86,03	24,59	0,64	12,74	1,37	1,06	0,56	0,58	22,07	0,14	1,09	-	0,02	-	-	0,27	5,58	-	+	-	0,01
Calanipeda aquaedulcis	CAQU	72,36	0,70	43,00	0,03	2,12	0,01	0,88	0,01	62,35	0,02	8,83	+	2,14	+	+	+	-	-	0,03	0,01	-	-	-	+
Centropages hamatus	CENT	0,04	15,73	0,01	0,07	0,01	1,58	0,01	1,94	-	0,01	-	-	+	2,35	-	-	+	+	-	-	+	-	-	0,01
Euterpina acutifrons	EACU	+	0,01	+	0,29	0,01	3,62	0,25	4,26	+	+	-	2,01	0,21	5,78	4,49	24,49	-	-	-	+	+	0,01	0,02	0,17
Oithona nana	ONAN	-	-	+	+	+	0,32	0,20	0,06	-	0,10	0,05	1,52	-	2,67	0,01	3,10	+	-	-	+	-	0,01	-	0,20
Paracalanus parvus	PPAR	+	5,38	+	+	0,01	+	+	0,11	0,13	5,40	0,09	1,56	-	+	-	+	12,64	0,06	-	+	+	-	-	+
Peltogaster paguri	PPAG	+	3,42	0,02	0,41	0,05	1,23	0,02	0,13	+	-	-	0,41	-	0,68	-	-	+	-	-	+	-	-	-	0,02
Pseudocalanus elongatus	PELO	+	3,51	+	+	+	0,03	0,01	6,20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
Pseudodiaptomus marinus	PMAR	+	0,02	1,56	+	1,50	+	0,72	0,26	0,03	-	0,49	-	12,76	-	1,90	1,56	+	-	-	-	0,02	-	0,01	-
Temora longicornis	TLON	+	1,07	-	+	+	0,12	0,53	2,79	+	0,99	0,01	+	+	0,01	+	0,01	0,01	+	-	0,01	+	0,01	+	0,01
Calanus helgolandicus	CHEL	1,70	13,02	0,02	0,06	0,04	7,11	0,03	1,05	0,02	11,91	0,03	1,58	0,04	17,18	0,02	0,08	0,07	0,15	0,01	0,01	0,01	0,01	+	0,02
Chelophyes appendiculata	CAPP	+	0,05	-	0,02	+	0,26	+	0,66	0,01	0,65	-	5,44	+	0,04	+	0,01	0,02	0,02	0,21	0,74	+	0,29	+	0,26
Scyliorhinus torazame	STOR	2,34	+	0,32	+	0,02	+	+	+	2,37	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+
Mytilus edulis	MEDU	-	0,13	+	+	+	+	-	-	-	1,38	-	-	-	0,04	+	0,16	+	17,81	-	-	-	-	-	0,01
Crassostrea gigas	CGIG	-	+	0,10	0,02	0,05	+	+	-	-	-	0,01	2,44	-	-	-	0,05	0,01	-	-	0,01	+	-	+	-
Littorina littorea	LITT	-	0,30	+	0,07	+	0,06	+	0,02	-	-	-	2,63	-	-	-	-	-	-	-	-	-	-	-	+
Uncultured zooplankton	UZOO	+	0,16	0,01	3,63	+	0,37	0,01	4,32	0,04	0,02	+	11,69	-	0,36	-	6,96	0,02	0,03	0,23	0,78	-	0,01	-	0,03
Not assigned	NASS	12,66	13,51	40,53	7,36	66,29	12,87	20,69	60,38	31,88	19,48	89,64	42,19	78,32	26,29	91,15	50,59	44,31	39,91	76,01	46,38	77,20	45,92	64,86	57,41
Others	OTHE	0,97	1,38	0,56	0,91	0,22	1,48	0,11	1,25	1,88	1,50	0,10	2,87	0,39	2,26	0,84	2,56	11,65	13,99	5,39	6,56	5,71	8,06	6,81	12,42
RICHNESS		**18**	**23**	**21**	**25**	**26**	**28**	**22**	**21**	**40**	**40**	**39**	**39**	**19**	**38**	**23**	**26**	**40**	**40**	**37**	**40**	**39**	**40**	**40**	**40**

Capítulo 2

PAPER 2: Comparison of Three Different Molecular Markers for Estuarine Plankton Monitoring by Means of DNA Metabarcoding

David Abad, Aitor Albaina, Mikel Aguirre and Andone Estonba

Manuscrito

Abstract

The advent of Next Generation Sequencing (NGS) technologies has provided an alternative to overcome the issues associated with microscopy-based monitoring. These NGS technologies allow the use of metabarcoding, where a short DNA region (marker) is sequenced for a whole communit sample and the obtained sequences are used to measure biodiversity at an affordable cost. In this study we performed a comparison between three different markers (two 18S regions, namely V1-V2 and V9, and COI) for biodiversity assessment of planktonic communities (>0.22 μm) in the estuary of Bilbao (Bay of Biscay). Our results showed that the power and efficiency varied among these markers, with the best overall performance obtained with the 18S V9. This study gives further support to the suitability of the 18S gene when profiling biodiversity in complex communities, especially in those such as plankton where a large number of taxa remain unknown and some dominant groups are difficult to amplify using the gold standard barcode gene (the Cytochrome c Oxidase subunit I, COI). In addition, correlations between sequences´ relative abundance and microscopy showed that the 18SV9 and COI are the best markers at least for the three targeted species in this study (*Acartia tonsa*, *Calanipeda aquaedulcis* and *Pseudodiaptomus marinus*). Finally, we consider that the shorter lentgh of the 18S V9 (~150bp versus ~450bp and ~350bp for the 18S V1-2 and COI, respectively) and the broader amplification range may also explain a higher success when DNA quality is severely compromised as environmental samples.

Introduction

Despite the ecological importance of planktonic communities and the significant effort devoted to characterize their biodiversity, species composition assessment usually relies on morphological characteristics observed by traditional methods, such as light microscopy, which are often difficult, time-consuming and require wide expertise due to similarities between species (including cryptic species) and restricted diagnostic features, especially in their development stages (Lindeque et al. 2006). Furthermore, to aggravate this problem, many plankton communities are comprised of a few very abundant species and numerous very rare species, increasing the difficulty to detect and identify all taxa (e.g. Cheung et al. 2010). In such a context, DNA metabarcoding has emerged as an alternative to traditional methods, allowing both automation of the process and potentially the highest taxonomic resolution for the characterization of species composition in bulk community samples because of its capability for generating large amounts of biodiversity information at an affordable cost (e.g. Fonseca et al. 2010; Lindeque et al. 2013; Hirai et al. 2015). Furthermore, the use of DNA metabarcoding is of particular interest with respect to eukaryotic plankton because of its reported high sensitivity (Zhan et al. 2013) that makes this technique ideal for the detection of species at any life stage, rare/cryptic taxa and low-abundance populations in complex eukaryote communities (e.g. Lindeque et al. 2013; Comtet et al. 2015). All of this advantages provided a way to overcome the mentioned issues associated with microscopy-based monitoring (Baird and Hajibabaei 2012).

Nonetheless, the accuracy of DNA metabarcoding for taxonomic assignment is highly dependent on marker (targeted DNA region) choice. Although several studies focused in marine communities have already proposed markers that present a combination between relatively conserved regions (as to allow designing “universal” primers) with variable regions capable, in principle, of discriminating species (e.g. Leray et al. 2013; Zhan et al. 2013), a trade-off between amplification efficiency and discriminatory power (Taylor and Harris 2012) is present suggesting the combination several markers as the sole effective approach. Thus, further progress, through a critical evaluation of different types of markers is crucial aiming for a reliable metabarcoding-based biodiversity assessment of aquatic communities. In this regard, we evaluated the suitability of 18S V9 for plankton biodiversity assessment in a previous study (Abad et al. 2016); although results were promising, the combination of a short amplicon (an advantage when targeting degraded DNA; e.g. Albaina et al. 2016a, 2016b) and a relatively low nucleotide variability among related taxa suggest

that this marker should be used in combination with other/s showing a higher discriminatory power. Because of this, in the present study we sequenced the same samples of the aforementioned study with another, longer, 18S region (V1-V2 one) along with the mitochondrial Cytochrome c Oxidase subunit I (COI; also longer and with a higher nucletotide variability among taxa). The combination of 18S and COI markers is promising due to their compatible characteristics (broad amplification and high taxonomic resolution in 18S and COI regions, respectively).

On the one hand, the mitochondrial COI gene (Folmer et al., 1994; Hebert et al., 2003) is one of the most sequenced regions for species diversity analysis purposes among marine animals (Bucklin et al., 2011). Due to its discriminatory power and the amount of publicly available sequence, it has become the gold standard barcode. It was assumed that with a 97% similarity clustering threshold, the obtained Operational Taxonomic Units (OTUs) corresponded to different species because of the so-called barcoding gap (Hebert et al. 2003a). Nonetheless, the length of this barcode (~700bp; Folmer et al., 1994) exceeds the one allowed by current Illumina technologies so it is necessary to use shorter COI regions (Leray et al., 2015) where not only the "universal" character of the primers is compromised but also the suitability of the 97% similarity threshold for species delimitation is called into question.

However, the higher variability associated to the COI region is counterbalanced by a relatively reduced barcode amplification range (Deagle et al., 2014); thus making this marker better suited for metabarcoding studies focused on concrete taxonomic groups (Zhan et al. 2014). This might explain the relatively short number of plankton biodiversity studies using COI that have been published so far (e.g. Machida et al., 2009;Zaiko et al., 2015b; Clarke et al. 2017).

The nuclear 18S ribosomal DNA (rDNA) gene, on the other hand, has commonly been employed as a marker in DNA metabarcoding because it shows consistent patterns of divergence across invertebrate and vertebrate taxa, and discriminates genera, families and higher taxonomic groups (Mallatt et al., 2004). To date different 18S hypervariable regions have been used, such as the V1-2 (hereafter 18S V1-2) and V9 (hereafter 18S V9) ones. Although having a reduced species-level taxonomic resolution, both markers have allowed to perform preliminary biodiversity profiling studies due to their broader taxonomic coverage (e.g. Zhan et al. 2014; Brown et al. 2015, de Vargas et al., 2015; Abad et al. 2016, Clarke et al. 2017).

In this study, we examined the efficiency and reliability of DNA metabarcoding for plankton biodiversity assessment using samples collected in the estuary of Bilbao (Bay of Biscay). We characterized the plankton communities associated with two different salinities (inner and outer estuary) using 18S V1-2, and COI markers. We compared the taxonomic resolution power of these two markers and the previously assessed 18S V9 one in order to determine which marker/combination of markers is more suitable for metabarcoding-based plankton monitoring. Furthermore, we also compared the relative abundances obtained with metabarcoding versus microscopy, with the focus on three targeted species (*Acartia tonsa*, *Calanipeda aquaedulcis* and *Pseudodiaptomus marinus*) because of their relevance for the ecosystem (the invasive nature of the first two and emerging of the latter).

Method

Sampling

The estuary of Bilbao is located in the south-east part of the Bay of Biscay (within 43º23'N to 43º14'N and 3º07'W to 2º55'W). It is a small (~23 km long), narrow (25-270 m), shallow (0.5 -10 m) and highly stratified channel that crosses urban and industrial settlements and drains into a wide coastal embayment (Uriarte et al. 2014).

Sampling was carried out in summer (June, July) and autumn (September, October) of 2013 from water at 30 and 35 ppt salinity during neap tides. Salinity (g/L) was measured with a YSI 556 MPS multi-parameter probe. Three plankton size-fractions were sampled (0.22-20, 20-200 and >200 μm). To obtain the 0.22-20 and 20-200 μm size fractions, a Niskin bottle was used to collect 10 L of water at each salinity (sampling depth depended on the water mass location). Samples were filtered through a 20 μm mesh (Millipore Nylon Nets) and, consecutively, approximately 1 L was filtered through a 0.22 μm Durapore Membrane (Millipore) using a Kitasato flask and a vacuum pump (Millipore). Meshes and membranes were kept in cryogenic tubes and frozen at -80ºC until further use for metabarcoding. For the microscopy analysis of phytoplankton samples, a 250 mL bottle with 1mL of Lugol was filled directly with water from the Niskin Bottle.

For the >200 μm size fraction, samples were collected at about 3 m depth by ~5 min horizontal tows of a 200 μm plankton net (mouth diameter: 0.25 m) equipped with a Hydrobios flow-meter. Once in the laboratory, each sample was divided using a plankton splitter. One half was kept in buffered formalin (4%) for microscopy analysis. One quarter was filtered through a 180 μm mesh (Millipore Nylon Nets) and preserved in ethanol for DNA barcoding analysis of some selected species. Finally, the remaining quarter was also filtered through a 180 μm mesh, kept in a falcon tube and frozen at -80ºC until further use for DNA metabarcoding.

DNA extraction

A modified salt protocol (Aljanabi and Martinez 1997) was used to extract the DNA from the 20 and 200 μm size-fractions. Meshes were defrosted and cleaned in a falcon tube with distilled water injected through a wash bottle to remove any possible attached organisms or DNA-containing fragments. The filters were held with each tube lid and centrifuged at 4000 rpm for 30 min to create a pellet. The mesh and supernatant water were then carefully removed. Proteinase K (20 mg/mL)

digestion was conducted on the pelleted organisms overnight and the extraction was continued according to the protocol. Samples from the 0.22 μm mesh were extracted following the instructions of the MOBIO PowerSoil® DNA Isolation Kit.

Extracted DNA was first quantified by spectrophotometry using Nanodrop (ND-1000; Thermo Scientific) and then by fluorometry using Qubit 1.0 (Thermo Scientific) to determine the amount of double-stranded DNA. DNA was stored at -20ºC until further processing.

DNA barcoding and database creation

After checking previous studies on the area (Albaina et al., 2009; Albaina et al., 2016a, Uriarte et al., 2015), we realized that some key/abundant species of the estuary of Bilbao were not present in publicly available databases (e.g *Pseudodiaptomus marinus*, *Calanipeda aquaedulcis*). In order not to miss these taxa among the community composition obtained by the metabarcoding approach, we generated 18S V1-2 and COI reference sequences (by Sanger sequencing) and included them in the database. The barcoding effort for the 18S V9 was performed in a previous study (Abad et al., 2016). 8 specimens were isolated from the ethanol replicas (S1_table). For the amplification of the COI fragment (~320 bp), we used the primers mlCOIintF (5'-GGWACWGGWTGAACWGTWTAYCCYCC-3') and jgHCO2198 (5'-TAIACYTCIGGRTGICCRAARAAYCA-3') described by Leray et al (2013). We used the primers from Fonseca et al (2010) SSU_F04 (5'-GCTTGTCTCAAAGATTAAGCC-3') and SSU_R22 (5'-GCCTGCTGCCTTCCTTGGA-3'), for the amplification of the 18S V1-2 (~450 bp fragment). Polymerase Chain Reactions (PCRs) were performed in a 25 μL volume containing 7.5 μL of distilled water, 5.4 μL of Buffer (5x), 2.7 μL of $MgCl_2$ (25 mM), 2.7 μL of dNTPs (10 μM), 2.7 μL of each primer (10 μM), 0.3 μL of Taq Polymerase (Promega) and 1 μL of template DNA. PCR cycling included initial denaturation at 92 °C for 3 min, followed by 30 cycles of 45 s denaturation at 92 °C, 1 min annealing at 57 °C and 1.5 min extension at 72 °C. A final extension step was performed at 72 °C for 10 min. The purified PCR products were sequenced in both directions on an ABI 31309 capillary electrophoresis Analyzer with ABI BigDye Terminator version 3.1 chemistry (Applied Biosystems) at the SGIKER (UPV/EHU). The obtained sequences were added to Silva 119 and a COI custom database. For the creation of the later, we downloaded all the COI sequences corresponding to vertebrates, invertebrates and plants from GenBank (search date: 18/12/15). Subsequently these files were merged and converted to Qiime format using in-house scripts.

DNA Metabarcoding, OTU definition and taxonomic assignation

Sequencing of the 18S V1-2and COI regions was carried out at the SGIKER (UPV/EHU) using Illumina MiSeq 2x300 bp. For the 18S V9, we used the sequences generated in Abad *et al* (2016). Raw reads were trimmed with Sickle v1.33 (Joshi and Fass 2011), using a quality threshold of 20. For paired-end merging, Pear v0.9.5 (Zhang et al. 2014) was used with a minimum overlapping of 15 bp and a cut-off P-value of 0.01. Chimeras were removed with UCHIME (Edgar et al. 2011), using a reference-based chimeric detection against our custom databases. Merged reads were processed using Qiime v1.9 (Caporaso et al. 2012). Sequences were clustered into Operational Taxonomic Units (OTUs) with UCLUST (Edgar 2010), using a *de novo* approach with a 97 and 99% identity threshold for 18S V1-2 and COI (Fonseca et al. 2010; Leray et al. 2013) and 99% for 18S V9 (Abad et al., 2016). OTU taxonomic assignment was performed with BLAST (Altschul et al. 1990), against our custom databases. Finally, singletons were removed from the OTU table.

Statistical analysis

Spearman's rank correlation coefficient (ρ) was calculated for the comparison of relative abundances retrieved by microscopy (extracted from Abad et al., 2016) and metabarcoding using cor.test in R (R Core Team, 2015); the correlations were limited to taxa uncovered by both methods. A Mantel test to analyze the differences in community structure obtained by the three markers included in this study was performed with Vegan (Oksanen et al. 2016). The bar charts were created with ggplot2 (Wickham 2009). Alpha diversities were calculated with Phyloseq (McMurdie and Holmes 2013). Detrended Correspondence Analysis (DCA) of the different samples were carried out using CANOCO Ver. 4.5 (ter Braak and Smilauer, 2002).

Results

Taxonomic composition

Once the raw sequences were quality filtered and merged, 2.33% of the reads were eliminated due to their putative chimeric nature. The 18S V1-2 marker at 97% (table 1) similarity was the approach with less taxa obtained (1801), whereas the COI at 99% produced the highest (3197).

Marker	Rep Seq	Singletons	Taxa
18S V1-2 97%	33709	396	1801
18S V1-2 99%	209306	481	2048
18S V9	180969	855	2318
COI 97%	87417	1074	2689
COI 99%	895410	1298	3197

Table 1. Summary for the different markers. Number of representative sequences (Rep Seq), unique sequences removed (singletons) and total taxa obtained are shown.

Regardless the salinity, in the smaller size-fraction (0.22-20 μm) there was a clear difference between the composition obtained by the two genes; while the assemblages of both 18S regions were very similar, the COI presented a different distribution. With respect to the larger size-fractions (20-200 and >200 μm), there is a more similar composition on the inner estuary communities (salinity 30) than in outer ones (salinity 35).

In the 18S V1-2, the presence of the category other Maxillopoda mostly in salinity 35 of the 20-200 and >200 μm size-fractions indicated that there were many OTUs (possibly copepods) that this region was not able to classify at a higher taxonomic level. As expected, abundances obtained with the two regions of the 18S gene (V1-2 and V9) were more similar than with COI. Maxillopoda (mainly copepods and barnacles) and Mollusca were the most frequently observed groups at the 20-200 and >200 μm size fractions, while a more diverse assemblage characterized the 0.22-20 μm one.

Particularly, the sample collected during July at salinity 35 were dominated by barnacles and this was correctly depicted by the 18S V9 and COI markers (and correlated with microscopy), but it was less appreciated with the 18S V1-2. The categories “other arthropoda” and “other metazoa” were more abundant in the COI. Malacostraca were recovered in higher abundance for the 18S V1-2.

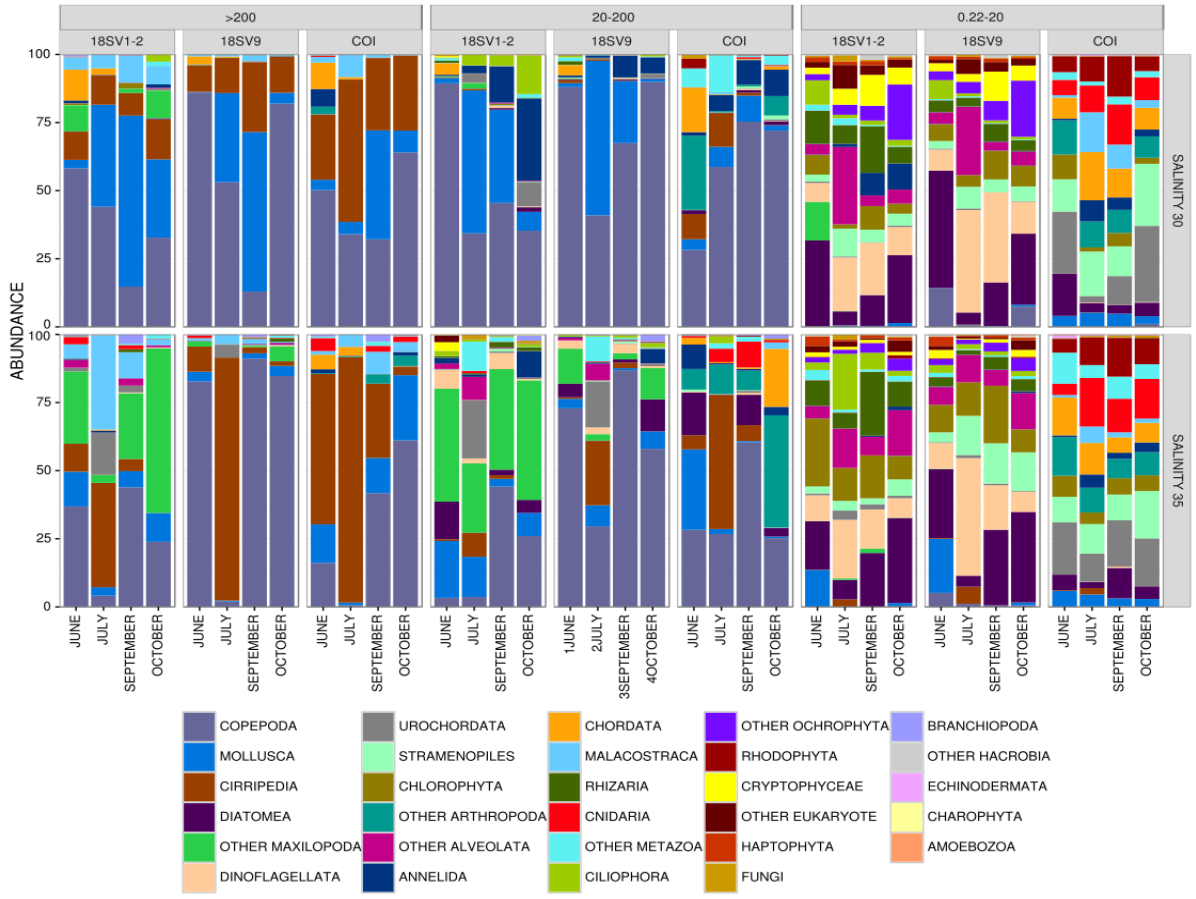


Fig 1. Proportion of taxonomic groups in each sample based on the DNA metabarcoding approach. A total of 29 taxonomic groups are shown in this graphic. Colored bar lengths correspond to their relative abundances. Samples are arranged by size-fraction (>200, 20-200 and 0.22-20 µm), marker (18S V1-2, 18S V9 and COI) and salinity (30 and 35). Only the 97% similarity threshold is represented for the 18S V1-2 and COI due to high similarity with the 99%. All values can be found in the S2_table.

The analysis from Fig 1 was also supported by the Detrended Correspondence Analysis (DCA) shown in Fig 2. For each of the markers and similarity thresholds, all the samples from the smaller size fraction (0.22-20 µm) were clustered together independently of the salinity. Instead, the samples from the larger ones (>200 and 20-200 µm) clustered together but separated by salinity.

Regarding the composition of the different taxonomic groups, the COI barely detected members from chloroplastida, cryptophyceae, ciliophora, dinoflagellata or rhizaria that were abundant with the 18S gene (Table 1). However, rhodophyceae, cnidaria, choradata, echinodermata and mollusca were detected in higher numbers for COI. In the case of the copepods, the COI at 99% similarity yielded an almost equal distribution to the 18S V9.

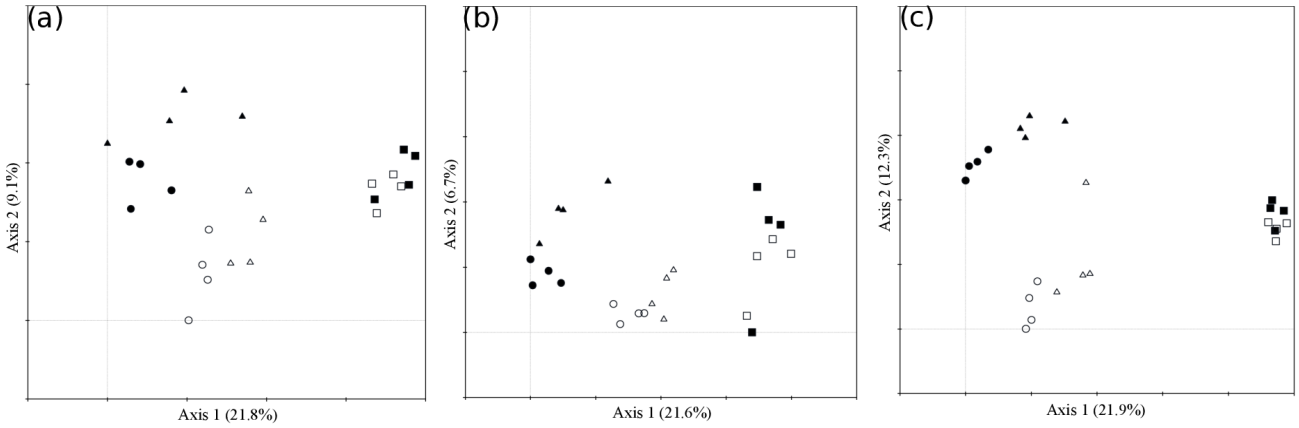


Fig 2. Detrended Correspondence Analysis (DCA) for the different markers and thresholds. Panel (a) shows the 18S V1-2 at 97%, (b) the 18S V9 at 99% similarity and (c) the COI gene at 97% similarity. Circles, triangles and squares represented the >200, 20-200 and 0.22-20 µm size fractions respectively. Empty and full symbols depicted 30 and 35 salinities respectively. Only the 97% similarity threshold is represented for the 18S V1-2 and COI due to high similarity with the 99%.

	18S V1-2 97%	18S V1-2 99%	18S V9	COI 97%	COI 99%
AMOEBOZOA	18	19	35	15	17
CHAROPHYTA	25	34	22	13	19
CHLOROPLASTIDA	99	133	105	11	13
RHODOPHYCEAE	28	32	50	308	357
CRYPTOPHYCEAE	32	34	39	1	1
HAPTOPHYTA	39	47	34	2	2
OTHER HACROBIA	4	10	3	0	0
OTHER EUKARYOTE	93	85	129	0	0
ANNELIDA	73	88	73	126	150
OTHER ARTHROPODA	21	22	78	261	315
BRANCHIOPODA	8	9	7	12	16
MALACOSTRACA	46	53	68	41	44
COPEPODA	58	67	93	73	92
CIRRIPEDA	9	10	22	19	20
OTHER MAXILOPODA	8	8	13	3	2
CNIDARIA	30	38	47	214	239
CHORDATA	9	10	31	194	267
ECHINODERMATA	5	6	7	62	68
MOLLUSCA	100	110	113	614	806
OTHER METAZOA	101	116	163	223	258
UROCHORDATA	16	16	14	5	5
FUNGI	139	170	135	76	76
CILIOPHORA	85	100	179	3	3
DINOFLAGELLATA	143	155	169	29	27
OTHER ALVEOLATA	94	108	110	1	0
RHIZARIA	136	141	158	3	3
STRAMENOPILES	183	206	168	193	209
OTHER OCHROPHYTA	40	43	95	115	112
DIATOMEA	159	178	158	72	76
Number of taxa	1801	2048	2318	2689	3197

Table 1. Number of taxa/OTUs included within the taxonomic groups for each marker. A total of 29 taxonomic groups are shown in this table. Below each marker with its corresponding similarity threshold is the number of taxa obtained for each category.

Marker comparison

As expected, Mantel test (Table 2) showed that the species assemblage depicted by both regions of the 18S gene had a significant correlation, whereas the 18S and COI ones were not correlated.

Fraction	V12/V9	V12/COI	V9/COI
R	0.6609	0.006877	-0.1879
Significance	0.001	0.484	0.973

Table 2. Mantel test between the communities/OTU table obtained by the different markers.

The 18S V9 showed the highest number of significant correlations with microscopy among markers (10 out 16 cases), followed by the 18S V1-2 (9 out 16 cases) and the COI gene (5 out 16 cases). For the 20-200 μm size-fraction, the COI performed better in salinity 30 whereas this corresponded to the 18S V9 in salinity 35. Likewise, this latter marker proved to work better for the 0.22-200 μm size-fraction.

Fraction	Salinity (n)	Month	V12/ABU	V9/ABU	COI/ABU
>200	30 (5)	JUN	0.22	0.67	0.97***
	30 (5)	JUL	0.89**	0.87**	0.97***
	30 (5)	SEPT	-0.05	0.40	1**
	30 (5)	OCT	-0.34	0.22	0.89**
	35 (11)	JUN	0.24	0.46	0.82
	35 (11)	JUL	-0.12	0.36	-0.02
	35 (11)	SEPT	0.13	0.76***	-0.04
	35 (11)	OCT	0.02	0.69**	0.31
0.22-200	30 (13)	JUN	0.44*	0.56**	0.13
	30 (13)	JUL	0.49*	0.33	0.26
	30 (13)	SEPT	0.77***	0.68***	-0.06
	30 (13)	OCT	0.50*	0.67***	-0.02
	35 (22)	JUN	0.70***	0.85***	0.39*
	35 (22)	JUL	0.58***	0.62***	0.11
	35 (22)	SEPT	0.64***	0.5**	0.1
	35 (22)	OCT	0.39**	0.52***	0.10

Table 4. Correlations between metabarcoding and microscopy-based analysis of community compositions. Spearman's rank correlation coefficient (ρ) and P-values are shown; $P < 0.01$ (***), $P < 0.05$ (**) and $P < 0.1$ (*). Relative abundances from DNA metabarcoding were compared against microscopy-based relative abundances. Each sample was defined by size-fraction (0.22-200 and >200 μm), salinity (30, 35) and sampled month (June, July, September and October). Number of taxa (n) included in the correlations is specified after each salinity data and was limited by the matches between the two approaches. Only the 97% similarity threshold is depicted for the 18S V1-2 and COI due to high similarity with the 99%

Figure 4 shows the relative abundances obtained from the DNA metabarcoding approach (the three markers) and microscopy for three targeted species (*A.tonsa*, *C. aquaedulcis* and *P. marinus*). In most cases, there were significant correlations among the values. For *C. aquaedulcis*, relative abundances based on the three markers showed significant correlations with microscopy-based ones except for COI in September and October in salinity 30. On the contrary, microscopy data for *P. marinus* and *A. tonsa* were only correlated for 18SV9 and COI markers.

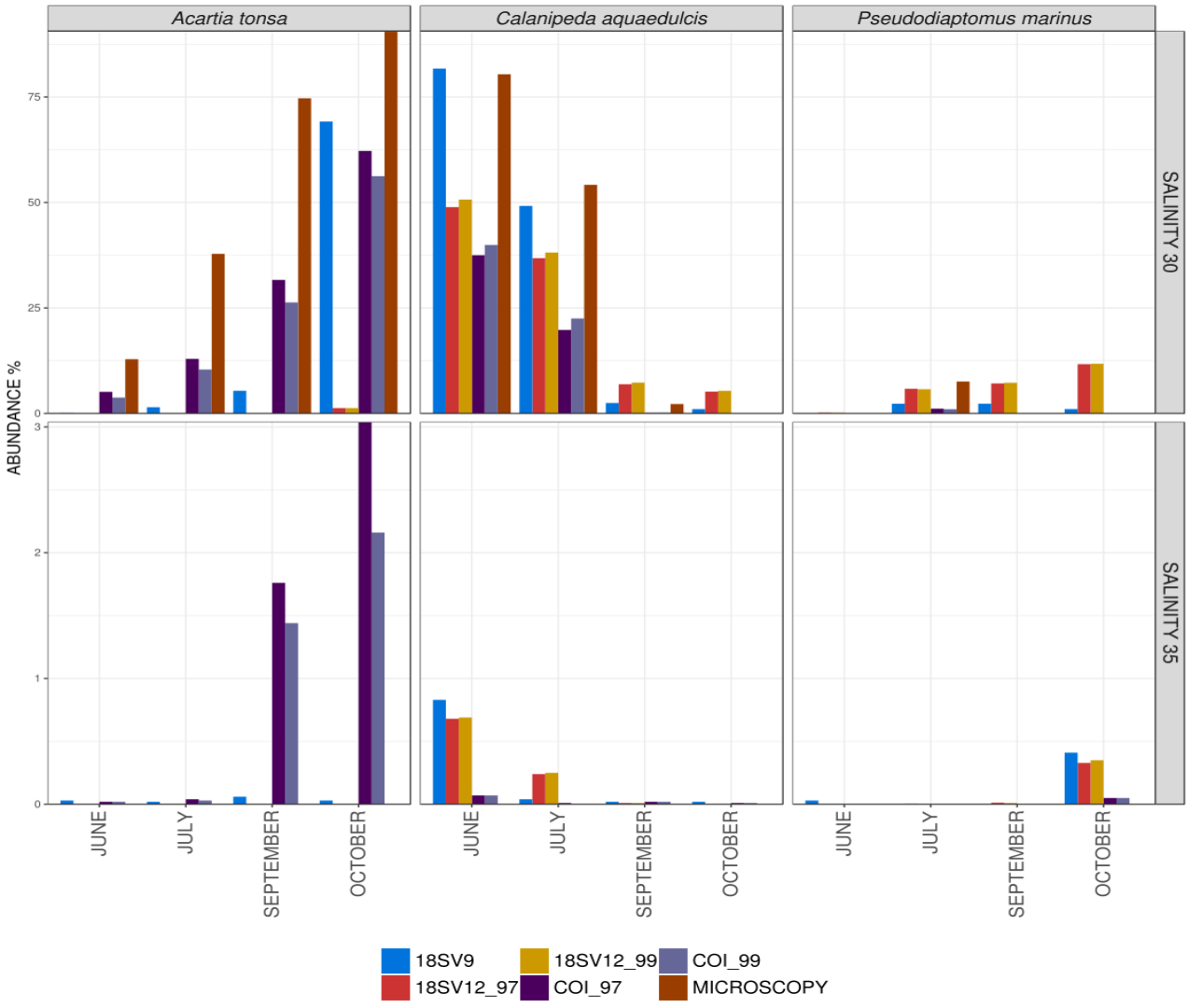


Fig 4. Relative abundances for three targeted species. The three markers and thresholds are depicted in different bars for each salinity.

Discussion

Metabarcoding has supposed a breakthrough for biodiversity assessment in environmental samples because it is a rapid and cost-effective method that produces lots of information at a relatively low cost. Furthermore, given its sensitivity, it is possible to detect low abundance/rare taxa (e.g. Fonseca et al., 2010; Lindeque et al., 2013) and organisms at any life stage (Comtet et al., 2015). Thus, it has been considered as an alternative to overcome the issues associated to morphological identification (Baird and Hajibabaei 2012).

However, despite its potential, there is still much work to do before metabarcoding can be considered for widespread adoption as it has several limitations in its current state. In this regard, marker choice is paramount (Coissac *et al*. 2012; Deagle *et al*. 2014); however, to date, there is no consensus about a certain marker and the combination of several markers seems to be compulsory because the taxonomic resolution and amplification efficiency varies largely depending on the target taxonomic group (Tang et al., 2012). In order to give insights on this matter, in the present study we compared the performance of three commonly used markers (two regions of the 18S rDNA and the COI gene) for biodiversity assessment in estuarine plankton community samples.

Cytochrome c Oxidase subunit I

The mitochondrial COI gene became the standard barcoding region (Folmer et al., 1994) for identifying and differentiating animal species (Hebert et al., 2003). Its high discriminatory power and better capacity to decipher true diversity have been confirmed by studies based on traditional Sanger sequencing (e.g. Tang et al., 2012). However, the higher variability associated to this gene reduces the amplification range and makes this marker better suited for studies focused on concrete taxonomic groups (Zhan et al. 2014). It has been shown that the lack of primer universality due to poorly conserved binding sites results in inadequate PCR amplification for some groups of organisms (Clarke et al., 2014; Deagle et al. 2014); this is the major reason why the COI gene has a narrower taxonomic coverage and is not recommended for its use in broad range biodiversity surveys that are based in metabarcoding (Deagle et al. 2014).

In this study, the comparison between markers revealed that the COI gene performed poorly for

picoplankton components mainly due to a combination of a lack of representative sequences in public databases and the mentioned inefficiency during the amplification step (Fig 2).

18S rDNA gene

Nowadays, the ramping number of DNA metabarcoding studies targeting regions within the nuclear 18S gene (from V1 to V9; e.g. Hadziavdic et al., 2014; Wu and Yu, 2015) signals 18S as the preferred option among molecular ecologists for biodiversity assessment in plankton samples (e.g. Hirai et al., 2015; Massana et al., 2015; Sun et al. 2015, Visco et al., 2015; Clarke et al. 2017). Although 18S markers have shown a reduced species taxonomic resolution compared to COI in some cases, they certainly provide a broader taxonomic coverage that allows to perform preliminary biodiversity profiling studies (Zhan et al. 2014; Amaral-Zettler et al. 2009). Among them, we included the V1-2 and V9 regions for this study. In this regard, our data clearly demonstrated that the 18S markers recovered more taxonomic groups (Fig. 2) than the COI gene (as previously shown in Zhan et al. 2014) in estuarine planktonic communities. However, several studies have suggested that the 18S likely underestimates true species diversity due to their relatively conserved nature and that it might be only adequate for higher taxonomic levels. Therefore, species level patterns should be interpreted with caution (e.g. Tang et al., 2012). Moreover, as expected, both 18S regions were highly correlated because they correspond to the same gene and are subjected to the same copy number variation (a significant advantage when aiming to produce, at least semi-, quantitative data).

The 18S V9 was the marker with most significant correlations with our microscopic dataset; this better performance of the smaller fragment could be explained because it was recently demonstrated that the 18S V1-2 fails to detect some species the *Acartia* genus (Mohrbeck et al., 2015), one of the most important (for its ecological value) and abundant organisms in many temperate estuaries such as the study area (Aravena et al., 2009). Furthermore, the 18S V9 region has proven sufficient resolution for this group of copepods as demonstrated in Abad et al. (2016), supporting the reliability of our results; the capacity to discriminate within the *Acartia* genus could be due to events of isolation that produced a higher nucleotide variability than in other copepods (Chen and Hare, 2008).

Nonetheless, there are some cases in which the reliability of the 18SV9 for copepods is compromised: a 100% identity was detected for two *Centropages* species and, more interestingly, to

eight copepod species corresponding to two different families, Aetideidae and Euchaetidae (Albaina et al., 2016b). Furthermore, Wu and colleagues (2015) reported that at least for copepods the V9 has a higher resolution at the genus level (with an identification success rate of about 80%) and the most divergent variable region in the copepod 18S rRNA gene. In addition, the 18S V9 small amplicon size (~150 bp including primers) results in an advantage when dealing with poor quality DNA such as eDNA (e.g. Kelly et al. 2014; Bohmann et al. 2014) or digested DNA (e.g. Pompanon et al., 2012; Albaina et al., 2016b) which are of huge interest for early detection of biodiversity shifts and food webs assessment, respectively.

Conclusions

Some of present results are in agreement with well-recognized characteristics of the 18S and COI markers identified in previous studies. Despite targeting two different regions of the 18S (V1-2 and V9) we found that they provide a reduced taxonomic resolution when comparing to the COI gene (Tang et al., 2012); even though, the best overall performance was obtained with the 18S V9 region. Instead, the taxonomic coverage was higher for the 18S corresponding to its more conserved (in an evolutionary context) nature; this would explain why some abundant taxa belonging to cryptophyceae, haptophyta, ciliophora or rhizaria were barely detected by COI (although a comparatively shortage of reference sequences for picoplanktonic organisms in the latter gene cannot be discarded). Note that it happens in less degree to the 18S gene for echinodermata, chordata, cnidaria, mollusca or rhodophyceae. These differences would account for the discrepancies in taxonomic compositions in the 0.22-20 size fraction for the COI and 18S (Table 2). Apart from this, in this study we also examined the relationship between microscopy and metabarcoding-based plankton relative abundances. Disparities between both techniques have been reported for plankton assemblages (Hirai et al. 2015; Massana et al. 2015; Stoeck et al. 2014; Sun et al. 2015). Interestingly, present results suggest the 18S V9 region provides a better proxy for Bilbao estuary planktonic communities than the other markers when considered separately. However, in future estudies, the combination of various markers should improve the results since it would help to overcome the issues associated with each of them.

In this regard, our results confirm previous findings by showing that different metabarcoding markers and thresholds provide slightly different views of genetic biodiversity, and that each of shortcomings and characteristics must be taken into account previous to the study. Finally, in order to properly evaluate markers´ performance (meaning which markers best match traditional datasets and highlighting potential shortcomings of markers), comparisons with morphology-based datasets are paramount and, currently, further work in this area (including the completion of reference sequence databases with the aid of taxonomists) is still needed as to validate metabarcoding as an alternative to traditional taxonomy methods.

Acknowledgements

The authors also thank the SGIker (UPV/EHU) for the technical and human support provided, and the Hydro-meteorology Service of the Regional Council of Bizkaia for the precipitation data. AA and DA work was supported by a contract with the Euskampus Foundation (Euskampus Fundazioa) and a Basque Government doctoral fellowship (UPV/EHU "ZabaldUz" program), respectively. This study was funded by a Basque Government grant "Grupo de Investigación Consolidado del Sistema Universitario Vasco" to support activities of the Genomic Resources Research Group (IT-558-10). Sampling was also partly financed by the Euskampus Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Albaina, A., F. Villate, and I. Uriarte. 2009. "Zooplankton Communities in Two Contrasting Basque Estuaries (1999-2001): Reporting Changes Associated with Ecosystem Health." Journal of Plankton Research 31 (7): 739–52. doi:10.1093/plankt/fbp025.

Albaina, A., Uriarte, I., Aguirre, M., Abad, D., Iriarte, A., Villate, F., Estonba, A., 2016a. Insights on the origin of invasive copepods colonizing Basque estuaries; a DNA barcoding approach. Marine Biodiversity Records 9, 51. doi:10.1186/s41200-016-0045-2

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

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic Local Alignment Search Tool." Journal of Molecular Biology 215 (3): 403–10. doi:10.1016/S0022-2836(05)80360-2.

Amaral-Zettler, Linda A., Elizabeth A. McCliment, Hugh W. Ducklow, and Susan M. Huse. 2009. "A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes." Edited by Gordon Langsley. PLoS ONE 4 (7): e6372. doi:10.1371/journal.pone.0006372.

Aravena, Guillermo, Fernando Villate, Ibon Uriarte, Arantza Iriarte, and Berta Ibáñez. 2009. "Response of Acartia Populations to Environmental Variability and Effects of Invasive Congenerics in the Estuary of Bilbao, Bay of Biscay." Estuarine, Coastal and Shelf Science 83 (4): 621–28. doi:10.1016/j.ecss.2009.05.013.

Baird, Donald J., and Mehrdad Hajibabaei. 2012. "Biomonitoring 2.0: A New Paradigm in Ecosystem Assessment Made Possible by next-Generation DNA Sequencing." Molecular Ecology 21 (8): 2039–2044.

Bohmann, Kristine, Alice Evans, M. Thomas P. Gilbert, Gary R. Carvalho, Simon Creer, Michael Knapp, Douglas W. Yu, and Mark de Bruyn. 2014. "Environmental DNA for Wildlife Biology and Biodiversity Monitoring." Trends in Ecology & Evolution 29 (6): 358–67.

doi:10.1016/j.tree.2014.04.003.

Bucklin A, Penelope K. Lindeque, Naiara Rodriguez-Ezpeleta, Aitor Albaina, and Maiju Lehtiniemi. 2016. “Metabarcoding of Marine Zooplankton: Prospects, Progress and Pitfalls.” Journal of Plankton Research 38 (3): 393–400. doi:10.1093/plankt/fbw023.

Caporaso, J. Gregory, Christian L. Lauber, William A. Walters, Donna Berg-Lyons, James Huntley, Noah Fierer, Sarah M. Owens, et al. 2012. “Ultra-High-Throughput Microbial Community Analysis on the Illumina HiSeq and MiSeq Platforms.” The ISME Journal 6 (8): 1621–24. doi:10.1038/ismej.2012.8.

Chen, Gang, and Matthew P. Hare. 2008. “Cryptic Ecological Diversification of a Planktonic Estuarine Copepod, Acartia Tonsa.” Molecular Ecology 17 (6): 1451–68. doi:10.1111/j.1365-294X.2007.03657.x.

Cheung, Man Kit, Chun Hang Au, Ka Hou Chu, Hoi Shan Kwan, and Chong Kim Wong. 2010. “Composition and Genetic Diversity of Picoeukaryotes in Subtropical Coastal Waters as Revealed by 454 Pyrosequencing.” The ISME Journal 4 (8): 1053–59. doi:10.1038/ismej.2010.26.

Clarke LJ, Soubrier J, Weyrich LS and Cooper A. 2014 Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. Mol Ecol Resour, 14: 1160-1170. doi:10.1111/1755-0998.12265

Clarke LJ, Beard JM, Swadling KM, Deagle BE. Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and Evolution. 2017;7(3):873-883. doi:10.1002/ece3.2667.

Comtet, Thierry, Anna Sandionigi, Frédérique Viard, and Maurizio Casiraghi. 2015. “DNA (Meta)barcoding of Biological Invasions: A Powerful Tool to Elucidate Invasion Processes and Help Managing Aliens.” Biological Invasions 17 (3): 905–22. doi:10.1007/s10530-015-0854-y.

Deagle, Bruce E., Simon N. Jarman, Eric Coissac, François Pompanon, and Pierre Taberlet. 2014. "DNA Metabarcoding and the Cytochrome c Oxidase Subunit I Marker: Not a Perfect Match." Biology Letters 10 (9): 20140562. doi:10.1098/rsbl.2014.0562.

Edgar, Robert C. 2010. "Search and Clustering Orders of Magnitude Faster than BLAST." Bioinformatics (Oxford, England) 26 (19): 2460–61. doi:10.1093/bioinformatics/btq461.

Edgar, Robert C., Brian J. Haas, Jose C. Clemente, Christopher Quince, and Rob Knight. 2011. "UCHIME Improves Sensitivity and Speed of Chimera Detection." Bioinformatics (Oxford, England) 27 (16): 2194–2200. doi:10.1093/bioinformatics/btr381.

Fonseca, Vera G., Gary R. Carvalho, Way Sung, Harriet F. Johnson, Deborah M. Power, Simon P. Neill, Margaret Packer, et al. 2010. "Second-Generation Environmental Sequencing Unmasks Marine Metazoan Biodiversity." Nature Communications 1 (7): 98. doi:10.1038/ncomms1095.

Godhe, Anna, Maria E. Asplund, Karolina Härnström, V. Saravanan, Anuj Tyagi, and Indrani Karunasagar. 2008. "Quantification of Diatom and Dinoflagellate Biomasses in Coastal Marine Seawater Samples by Real-Time PCR." Applied and Environmental Microbiology 74 (23): 7174–82. doi:10.1128/AEM.01298-08.

Gonzalez, Juan M., Maria C. Portillo, Pedro Belda-Ferre, and Alex Mira. 2012. "Amplification by PCR Artificially Reduces the Proportion of the Rare Biosphere in Microbial Communities." PLOS ONE 7 (1): e29973. doi:10.1371/journal.pone.0029973.

Hadziavdic K, Lekang K, Lanzen A, Jonassen I, Thompson EM, et al. (2014) Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. PLOS ONE 9(2): e87624. https://doi.org/10.1371/journal.pone.0087624

Hebert, Paul D. N., Alina Cywinska, Shelley L. Ball, and Jeremy R. deWaard. 2003. "Biological Identifications through DNA Barcodes." Proceedings. Biological Sciences / The Royal Society 270 (1512): 313–21. doi:10.1098/rspb.2002.2218.

Hirai, J., M. Kuriyama, T. Ichikawa, K. Hidaka, and A. Tsuda. 2015. "A Metagenetic Approach for Revealing Community Structure of Marine Planktonic Copepods." Molecular Ecology Resources

15 (1): 68–80. doi:10.1111/1755-0998.12294.

Huse, Susan M., Julie A. Huber, Hilary G. Morrison, Mitchell L. Sogin, and David Mark Welch. 2007. “Accuracy and Quality of Massively Parallel DNA Pyrosequencing.” Genome Biology 8 (7): R143. doi:10.1186/gb-2007-8-7-r143.

Joshi, and Fass. 2011. “Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files.” GitHub. https://github.com/najoshi/sickle.

Kelly, Ryan P., Jesse A. Port, Kevan M. Yamahara, Rebecca G. Martone, Natalie Lowell, Philip Francis Thomsen, Megan E. Mach, et al. 2014. “Environmental Monitoring. Harnessing DNA to Improve Environmental Management.” Science (New York, N.Y.) 344 (6191): 1455–56. doi:10.1126/science.1251156.

Kembel, Steven W., Martin Wu, Jonathan A. Eisen, and Jessica L. Green. 2012. “Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance.” PLOS Comput Biol 8 (10): e1002743. doi:10.1371/journal.pcbi.1002743.

Leray, Matthieu, Joy Y. Yang, Christopher P. Meyer, Suzanne C. Mills, Natalia Agudelo, Vincent Ranwez, Joel T. Boehm, and Ryuji J. Machida. 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.” Frontiers in Zoology 10 (1): 1.

Lindeque, P. K., S. J. Hay, M. R. Heath, A. Ingvarsdottir, J. Rasmussen, G. R. Smerdon, and J. J. Waniek. 2006. “Integrating Conventional Microscopy and Molecular Analysis to Analyse the Abundance and Distribution of Four Calanus Congeners in the North Atlantic.” Journal of Plankton Research 28 (2): 221–38. doi:10.1093/plankt/fbi115.

Lindeque, Penelope K., Helen E. Parry, Rachel A. Harmer, Paul J. Somerfield, and Angus Atkinson. 2013. “Next Generation Sequencing Reveals the Hidden Diversity of Zooplankton Assemblages.” Edited by Adrianna Ianora. PLoS ONE 8 (11): e81327. doi:10.1371/journal.pone.0081327.

Massana, Ramon, Angélique Gobet, Stéphane Audic, David Bass, Lucie Bittner, Christophe Boutte, Aurélie Chambouvet, et al. 2015. “Marine Protist Diversity in European Coastal Waters and Sediments as Revealed by High-Throughput Sequencing: Protist Diversity in European Coastal Areas.” Environmental Microbiology 17 (10): 4035–49. doi:10.1111/1462-2920.12955.

McMurdie, Paul J., and Susan Holmes. 2013. “Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data.” PLOS ONE 8 (4): e61217. doi:10.1371/journal.pone.0061217.

Mohrbeck, Inga, Michael J. Raupach, Pedro Martínez Arbizu, Thomas Knebelsberger, and Silke Laakmann. 2015. “High-Throughput Sequencing—The Key to Rapid Biodiversity Assessment of Marine Metazoa?” PloS One 10 (10): e0140342.

Oksanen, Jari, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O’Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, and Helene Wagner. 2016. Vegan: Community Ecology Package (version 2.3-4). https://cran.r-project.org/web/packages/vegan/index.html.

Pompanon, Francois, Bruce E. Deagle, William O. C. Symondson, David S. Brown, Simon N. Jarman, and Pierre Taberlet. 2012. “Who Is Eating What: Diet Assessment Using next Generation Sequencing: NGS DIET ANALYSIS.” Molecular Ecology 21 (8): 1931–50. doi:10.1111/j.1365-294X.2011.05403.x.

Prokopowich, Cheryl D, T Ryan Gregory, and Teresa J Crease. 2003. “The Correlation between rDNA Copy Number and Genome Size in Eukaryotes.” Genome 46 (1): 48–50. doi:10.1139/g02-103.

R Core Team. 2016. “R: A Language and Environment for Statistical Computing.” Text. GBIF.ORG. Accessed April 1. http://www.gbif.org/resource/81287.

Roh, Changhyun, Francois Villatte, Byung-Gee Kim, and Rolf D. Schmid. 2006. “Comparative Study of Methods for Extraction and Purification of Environmental DNA from Soil and Sludge Samples.” Applied Biochemistry and Biotechnology 134 (2): 97–112.

Stoeck, Thorsten, David Bass, Markus Nebel, Richard Christen, Meredith D. M. Jones, Hans-Werner Breiner, and Thomas A. Richards. 2010. "Multiple Marker Parallel Tag Environmental DNA Sequencing Reveals a Highly Complex Eukaryotic Community in Marine Anoxic Water." Molecular Ecology 19 (March): 21–31. doi:10.1111/j.1365-294X.2009.04480.x.

Tang, Cuong Q., Francesca Leasi, Ulrike Obertegger, Alexander Kieneke, Timothy G. Barraclough, and Diego Fontaneto. 2012. "The Widely Used Small Subunit 18S rDNA Molecule Greatly Underestimates True Diversity in Biodiversity Surveys of the Meiofauna." Proceedings of the National Academy of Sciences 109 (40): 16208–12. doi:10.1073/pnas.1209160109.

Taylor, H. R., and W. E. Harris. 2012. "An Emergent Science on the Brink of Irrelevance: A Review of the Past 8 Years of DNA Barcoding." Molecular Ecology Resources 12 (3): 377–88. doi:10.1111/j.1755-0998.2012.03119.x.

ter Braak CJF, Smilauer P (2002) CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination (Version 4.5). Microcomputer Power, Ithaca.

Uriarte, Ibon, Fernando Villate, and Arantza Iriarte. 2015. "Zooplankton Recolonization of the Inner Estuary of Bilbao: Influence of Pollution Abatement, Climate and Non-Indigenous Species." Journal of Plankton Research, July, fbv060. doi:10.1093/plankt/fbv060.

Uriarte, Ibon, Fernando Villate, Arantza Iriarte, Johanna Duque, and Iban Ameztoy. 2014. "Seasonal and Axial Variations of Net Water Circulation and Turnover in the Estuary of Bilbao." Estuarine, Coastal and Shelf Science, Special issue on Problems of Small Estuaries, 150, Part B (October): 312–24. doi:10.1016/j.ecss.2014.04.007.

Visco, Joana Amorim, Laure Apothéloz-Perret-Gentil, Arielle Cordonier, Philippe Esling, Loïc Pillet, and Jan Pawlowski. 2015. "Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data." Environmental Science & Technology 49 (13): 7597–7605. doi:10.1021/es506158m.

Wu S, Xiong J, Yu Y (2015) Taxonomic Resolutions Based on 18S rRNA Genes: A Case Study of

Subclass Copepoda. PLoS ONE 10(6): e0131498. doi:10.1371/journal.pone.0131498

Wickham, Hadley. 2009. ggplot2. New York, NY: Springer New York. http://link.springer.com/10.1007/978-0-387-98141-3.

Zhan, Aibin, Sarah A. Bailey, Daniel D. Heath, and Hugh J. Macisaac. 2014. “Performance Comparison of Genetic Markers for High-Throughput Sequencing-Based Biodiversity Assessment in Complex Communities.” Molecular Ecology Resources, April, n/a-n/a. doi:10.1111/1755-0998.12254.

Zhan, Aibin, Martin Hulák, Francisco Sylvester, Xiaoting Huang, Abisola A. Adebayo, Cathryn L. Abbott, Sarah J. Adamowicz, Daniel D. Heath, Melania E. Cristescu, and Hugh J. MacIsaac. 2013. “High Sensitivity of 454 Pyrosequencing for Detection of Rare Species in Aquatic Communities.” Edited by Sergei Kosakovsky Pond. Methods in Ecology and Evolution 4 (6): 558–65. doi:10.1111/2041-210X.12037.

Zhang, Jiajie, Kassian Kobert, Tomáš Flouri, and Alexandros Stamatakis. 2014. “PEAR: A Fast and Accurate Illumina Paired-End reAd mergeR.” Bioinformatics (Oxford, England) 30 (5): 614–20. doi:10.1093/bioinformatics/btt593.

Zhu, Fei, Ramon Massana, Fabrice Not, Dominique Marie, and Daniel Vaulot. 2005. “Mapping of Picoeucaryotes in Marine Ecosystems with Quantitative PCR of the 18S rRNA Gene.” FEMS Microbiology Ecology 52 (1): 79–92. doi:10.1016/j.femsec.2004.10.006.

S1_table. List of organisms sequenced by Sanger.

Species	Marker		Origin
Calanipeda aquaedulcis	18S V1-2		Present study (Bilbao estuary)
Euterpina acutifrons	18S V1-2		Present study (Bilbao estuary)
Pseudodiaptomus marinus	18S V1-2		Present study (Bilbao estuary)
Evadne nordmanni	18S V1-2	COI	Present study (Bilbao estuary)
Evadne spinifera	18S V1-2	COI	Present study (Bilbao estuary)
Podon spp.	18S V1-2		Present study (Bilbao estuary)

Source details for every taxon used to create the custom database is also included.

S2_table_a. Relative abundances of the 29 taxonomic groups for the 18S V1-2 are shown in this table. Samples are arranged by salinity (30 and 35) and month. A category with sequences that had no database match is also included "no blast hit".

	SALINITY 30											
	>200				20-200				0.22-20			
	June	July	September	October	June	July	September	October	June	July	September	October
AMOEBOZOA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.02
ANNELIDA	0.96	0.03	0.03	1.23	0.52	2.83	13.19	30.11	0.13	1.17	8.28	9.59
BRANCHIOPODA	0.62	0.00	0.01	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CHAROPHYTA	0.18	0.00	0.07	0.01	0.15	0.01	0.06	0.01	0.06	0.04	0.01	0.00
CHLOROPHYTA	0.58	0.02	1.94	0.10	0.10	0.02	0.55	0.08	7.17	1.58	8.65	3.67
CHORDATA	11.29	2.29	0.04	0.01	4.27	0.00	0.01	0.00	0.02	0.01	0.00	0.01
CILIOPHORA	0.01	0.00	0.04	2.54	0.26	3.97	3.82	14.10	8.97	1.28	1.61	2.21
CIRRIPEDA	10.37	10.73	8.25	14.69	0.17	0.57	0.10	0.01	0.00	0.07	0.00	0.00
CNIDARIA	0.12	0.12	0.09	0.00	0.00	0.01	0.07	0.00	0.06	0.02	0.00	0.00
COPEPODA	58.24	44.10	14.80	32.76	89.44	34.36	45.50	35.31	0.04	0.04	0.06	0.09
CRYPTOPHYCEAE	0.00	0.00	0.00	0.00	0.48	0.01	0.01	0.10	2.32	5.87	11.35	6.20
DIATOMEA	0.08	0.01	0.08	0.39	0.05	0.06	0.57	1.69	31.59	5.07	11.29	24.87
DINOFLAGELLATA	0.00	0.00	0.01	0.07	0.05	0.08	0.73	0.34	6.92	19.71	19.24	10.33
ECHINODERMATA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FUNGI	0.11	0.00	0.06	0.04	0.44	0.03	0.21	0.22	0.48	2.33	0.50	0.66
HAPTOPHYTA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.77	1.24	1.57	0.58
MALACOSTRACA	4.34	4.92	9.86	6.76	0.12	0.00	0.01	0.00	0.00	0.02	0.00	0.00
MOLLUSCA	3.07	37.44	62.77	28.74	1.94	52.47	34.15	6.92	0.09	0.50	0.33	1.35
OTHER ALVEOLATA	0.08	0.09	0.13	0.38	0.00	0.07	0.38	0.29	3.85	28.36	3.90	5.19
OTHER ARTHROPODA	0.06	0.03	0.01	0.04	0.00	0.02	0.00	0.19	0.17	0.02	0.00	0.01
OTHER EUKARYOTE	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.03	2.11	8.70	3.50	2.94
OTHER HACROBIA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41	0.24	1.90	0.62
OTHER MAXILOPODA	9.60	0.15	1.56	10.19	0.28	2.07	0.02	0.03	14.20	0.08	0.01	0.01
OTHER METAZOA	0.13	0.04	0.08	1.54	0.85	0.07	0.12	1.36	2.14	2.50	0.44	0.34
OTHER OCHROPHYTA	0.01	0.00	0.12	0.00	0.03	0.00	0.00	0.04	2.23	3.69	5.41	20.31
RHIZARIA	0.06	0.00	0.00	0.01	0.85	0.01	0.13	0.29	12.12	6.83	17.18	6.09
RHODOPHYTA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.01	0.00	0.01
STRAMENOPILES	0.07	0.00	0.02	0.02	0.02	0.01	0.13	0.21	2.51	10.22	4.68	4.37
UROCHORDATA	0.02	0.01	0.04	0.42	0.01	3.33	0.22	8.68	0.54	0.37	0.09	0.54
	SALINITY 35											
	>200				20-200				0.22-20			
	June	July	September	October	June	July	September	October	June	July	September	October
AMOEBOZOA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.03
ANNELIDA	0.45	0.72	0.03	0.05	2.04	0.94	0.10	9.73	0.11	0.01	0.42	1.34
BRANCHIOPODA	0.30	0.04	3.29	0.01	0.00	0.09	0.21	1.88	0.00	0.00	0.00	0.00
CHAROPHYTA	0.00	0.00	0.00	0.00	0.16	0.01	0.06	0.44	0.10	0.03	0.00	0.01
CHLOROPHYTA	0.02	0.02	0.00	0.00	0.01	0.02	0.12	0.00	24.96	12.11	15.77	8.61
CHORDATA	0.09	0.41	0.00	0.00	0.02	0.00	0.00	0.07	0.00	0.00	0.00	0.01
CILIOPHORA	0.00	0.00	0.00	0.00	1.85	0.76	0.42	1.08	2.78	20.49	6.15	1.90
CIRRIPEDA	10.26	38.29	4.48	0.05	0.63	8.73	1.34	0.00	0.10	2.69	0.00	0.00
CNIDARIA	2.63	0.00	1.51	0.14	0.00	0.87	0.04	0.00	0.07	0.08	0.03	0.04
COPEPODA	36.87	4.16	43.91	23.90	3.36	3.65	44.18	26.02	0.06	0.00	0.03	0.37
CRYPTOPHYCEAE	0.00	0.00	0.00	0.00	3.20	0.00	0.00	0.00	1.64	1.08	1.13	1.21
DIATOMEA	0.01	0.00	0.00	0.00	13.80	0.12	2.04	4.76	17.75	7.12	19.76	31.15
DINOFLAGELLATA	0.03	0.02	0.33	0.27	6.94	1.76	5.91	0.52	9.45	21.52	14.38	7.16
ECHINODERMATA	0.01	0.06	0.05	0.00	0.05	0.00	0.00	0.00	0.01	0.00	0.00	0.00
FUNGI	0.02	0.09	0.00	0.00	0.15	1.66	0.15	0.88	0.09	1.77	0.19	0.38
HAPTOPHYTA	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.05	3.51	0.62	1.14	1.01
MALACOSTRACA	5.09	34.48	9.55	2.16	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
MOLLUSCA	12.75	3.07	5.92	10.56	20.85	14.75	2.83	8.50	13.52	0.08	0.01	1.07
OTHER ALVEOLATA	2.84	0.07	2.66	0.73	2.05	8.49	0.27	0.05	4.52	14.42	6.80	16.68
OTHER ARTHROPODA	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.10
OTHER EUKARYOTE	0.19	0.00	0.01	0.00	2.49	0.11	0.04	0.07	2.35	2.01	0.61	4.19
OTHER HACROBIA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.05	0.02	0.75
OTHER MAXILOPODA	26.53	2.93	24.09	60.42	41.43	25.50	36.90	43.88	0.10	0.57	1.60	0.13
OTHER METAZOA	0.34	0.19	0.53	1.56	0.09	10.71	1.76	0.17	3.70	1.06	0.85	2.00
OTHER OCHROPHYTA	0.01	0.00	0.00	0.00	0.01	0.00	0.04	0.09	1.85	1.51	3.52	4.49
RHIZARIA	0.17	0.00	1.12	0.04	0.58	0.40	1.99	1.33	9.33	5.81	23.41	9.18
RHODOPHYTA	0.05	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.23	0.00	0.09	1.26
STRAMENOPILES	0.00	0.01	0.00	0.00	0.08	0.09	1.03	0.00	2.58	3.57	2.18	6.00
UROCHORDATA	1.30	15.43	2.51	0.08	0.20	21.34	0.54	0.47	0.66	3.37	1.91	0.93

S2_table_b. Relative abundances of the 29 taxonomic groups for the 18S V9 are shown in this table. Samples are arranged by salinity (30 and 35) and month. A category with sequences that had no database match is also included "no blast hit".

	SALINITY 30											
	>200				20-200				0,22-20			
	June	July	September	October	June	July	September	October	June	July	September	October
AMOEBOZOA	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.05	0.04	0.02	0.04
ANNELIDA	0.09	0.01	0.04	0.17	0.76	0.87	7.89	5.92	0.15	0.12	0.13	0.17
BRANCHIOPODA	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
CHAROPHYTA	0.01	0.00	0.01	0.00	0.06	0.00	0.01	0.01	0.04	0.03	0.00	0.01
CHLOROPHYTA	0.02	0.00	0.12	0.01	0.03	0.00	0.06	0.05	6.14	4.29	10.47	7.61
CHORDATA	2.97	0.42	0.02	0.00	3.80	0.00	0.00	0.00	0.05	0.01	0.00	0.02
CILIOPHORA	0.02	0.00	0.00	0.00	0.75	0.03	0.05	0.63	6.99	1.07	0.98	1.00
CIRRIPEDA	9.74	12.89	25.66	13.19	1.13	0.60	0.15	0.00	0.00	0.28	0.00	0.00
CNIDARIA	0.02	0.04	0.06	0.00	0.00	0.01	0.12	0.00	0.05	0.03	0.01	0.02
COPEPODA	86.04	53.13	12.95	81.97	88.02	40.95	67.50	89.97	14.29	0.38	0.11	7.48
CRYPTOPHYCEAE	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.02	2.90	2.86	10.71	5.52
DIATOMEA	0.02	0.01	0.06	0.02	0.38	0.00	0.38	0.27	43.00	4.42	15.98	26.06
DINOFLAGELLATA	0.00	0.00	0.00	0.00	0.05	0.01	0.40	0.07	7.70	37.59	33.06	11.74
ECHINODERMATA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FUNGI	0.02	0.01	0.03	0.02	0.05	0.01	0.03	0.01	0.22	0.70	0.25	0.44
HAPTOPHYTA	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	1.37	0.88	1.32	0.68
MALACOSTRACA	0.52	0.65	2.25	0.22	0.09	0.01	0.02	0.01	0.01	0.05	0.01	0.00
MOLLUSCA	0.23	32.69	58.56	3.98	1.46	56.75	22.80	1.06	0.04	0.21	0.23	0.67
OTHER ALVEOLATA	0.00	0.00	0.05	0.04	0.02	0.02	0.06	0.06	4.23	24.99	3.19	5.18
OTHER ARTHROPODA	0.14	0.07	0.02	0.08	0.03	0.01	0.01	0.02	0.12	0.01	0.09	0.03
OTHER EUKARYOTE	0.03	0.00	0.01	0.02	0.09	0.01	0.02	0.02	1.21	5.44	3.52	2.45
OTHER HACROBIA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.31	0.07	1.18	0.45
OTHER MAXILOPODA	0.04	0.01	0.02	0.06	0.28	0.02	0.04	0.04	0.07	0.00	0.00	0.00
OTHER METAZOA	0.02	0.02	0.03	0.09	1.38	0.22	0.14	0.26	0.43	0.47	0.29	0.26
OTHER OCHROPHYTA	0.01	0.00	0.04	0.00	0.05	0.01	0.03	0.02	3.16	4.25	7.16	20.59
RHIZARIA	0.02	0.01	0.03	0.09	1.31	0.02	0.06	0.04	4.19	3.23	6.49	3.95
RHODOPHYTA	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.15	0.01	0.01	0.01
STRAMENOPILES	0.01	0.00	0.01	0.00	0.17	0.00	0.07	0.05	2.80	7.96	4.71	5.44
UROCHORDATA	0.00	0.01	0.01	0.02	0.02	0.47	0.12	1.44	0.33	0.59	0.04	0.17
	SALINITY 35											
	>200				20-200				0,22-20			
AMOEBOZOA	June	July	September	October	June	July	September	October	June	July	September	October
ANNELIDA	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.03	0.04	0.03	0.34
BRANCHIOPODA	0.09	0.11	0.08	0.10	0.30	0.49	0.06	5.27	0.25	0.09	0.16	0.66
CHAROPHYTA	0.17	0.01	0.85	0.01	0.00	0.12	0.12	2.50	0.00	0.00	0.00	0.00
CHLOROPHYTA	0.03	0.00	0.01	0.00	0.01	0.00	0.01	0.02	0.05	0.02	0.00	0.06
CHORDATA	0.00	0.00	0.00	0.00	0.01	0.02	0.05	0.04	10.06	12.29	21.08	8.44
CILIOPHORA	0.04	0.08	0.00	0.01	0.00	0.01	0.00	0.07	0.02	0.01	0.00	0.01
CIRRIPEDA	0.00	0.00	0.00	0.01	0.64	0.43	0.92	1.67	2.87	1.78	1.45	1.90
CNIDARIA	9.27	89.20	2.20	1.64	0.58	23.60	2.08	0.02	0.28	6.28	0.00	0.03
COPEPODA	0.99	0.00	0.18	0.22	0.00	0.34	0.04	0.01	0.06	0.04	0.01	0.04
CRYPTOPHYCEAE	82.77	1.99	91.13	84.74	72.92	29.56	86.86	58.04	5.16	1.11	0.47	0.58
DIATOMEA	0.00	0.00	0.06	0.02	0.01	0.00	0.00	0.01	3.00	0.37	1.26	2.50
DINOFLAGELLATA	0.01	0.01	0.02	0.07	4.96	0.13	1.18	11.77	25.26	4.03	27.74	33.20
ECHINODERMATA	0.01	0.01	0.05	0.41	2.91	2.48	3.39	0.72	9.49	43.16	16.34	7.39
FUNGI	0.00	0.01	0.01	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HAPTOPHYTA	0.01	0.01	0.01	0.01	0.02	0.05	0.03	0.09	0.14	0.30	0.08	0.16
MALACOSTRACA	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	3.35	0.35	1.01	0.96
MOLLUSCA	0.87	3.41	1.77	0.34	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.01
OTHER ALVEOLATA	3.55	0.35	1.95	3.73	3.45	7.75	0.74	6.37	19.84	0.03	0.10	1.05
OTHER ARTHROPODA	0.06	0.02	0.28	0.70	1.02	6.09	0.32	0.16	6.51	10.04	5.82	13.31
OTHER EUKARYOTE	0.01	0.07	0.02	0.06	0.00	0.02	0.00	0.03	0.02	0.01	0.02	0.02
OTHER HACROBIA	0.01	0.00	0.08	0.21	0.04	0.01	0.03	0.10	1.49	0.74	0.51	3.55
OTHER MAXILOPODA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.62	0.06	0.08	0.59
OTHER METAZOA	1.89	0.03	0.56	5.43	12.88	2.37	2.28	11.47	0.17	0.00	0.00	0.03
OTHER OCHROPHYTA	0.10	0.03	0.11	0.92	0.02	9.17	0.56	0.31	1.53	0.38	0.50	1.08
RHIZARIA	0.00	0.00	0.00	0.00	0.01	0.09	0.02	0.18	2.23	1.66	3.19	4.71
RHODOPHYTA	0.02	0.01	0.30	1.20	0.03	0.04	0.30	0.18	3.37	1.65	4.68	4.41
STRAMENOPILES	0.01	0.01	0.01	0.04	0.00	0.00	0.02	0.12	0.31	0.02	0.04	0.52
UROCHORDATA	0.01	0.01	0.06	0.05	0.04	0.29	0.74	0.13	3.60	14.40	14.80	14.19
	0.09	4.60	0.25	0.07	0.08	16.91	0.22	0.69	0.29	1.13	0.61	0.26

S2_table_c. Relative abundances of the 29 taxonomic groups for the COI are shown in this table. Samples are arranged by salinity (30 and 35) and month. A category with sequences that had no database match is also included "no blast hit".

	SALINITY 30											
	>200				20-200				0,22-20			
	June	July	September	October	June	July	September	October	June	July	September	October
AMOEBOZOA	0.02	0.00	0.00	0.00	0.00	0.00	0.06	0.11	0.13	0.06	0.04	0.08
ANNELIDA	6.52	0.02	0.00	0.03	1.30	5.82	9.03	9.69	0.60	7.76	4.56	2.52
BRANCHIOPODA	0.04	0.00	0.00	0.00	0.07	0.00	0.00	0.01	0.04	0.02	0.02	0.04
CHAROPHYTA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CHLOROPHYTA	0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.05	9.09	1.52	4.89	2.21
CHORDATA	9.68	0.83	0.04	0.04	16.40	0.30	0.45	1.69	30.57	20.03	21.24	35.91
CILIOPHORA	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CIRRIPEDA	23.85	52.25	26.56	27.55	9.26	12.39	1.09	0.09	0.03	0.11	0.01	0.01
CNIDARIA	0.51	0.07	0.02	0.01	0.07	0.00	0.62	0.40	5.61	9.83	14.76	8.41
COPEPODA	50.24	34.03	32.17	64.02	28.30	58.73	75.32	72.04	0.14	0.13	0.16	1.14
CRYPTOPHYCEAE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DIATOMEA	0.17	0.00	0.04	0.01	1.41	0.09	0.99	1.42	15.40	3.38	3.08	4.96
DINOFLAGELLATA	0.00	0.00	0.00	0.00	0.00	0.06	0.47	0.19	0.06	0.21	0.07	0.22
ECHINODERMATA	0.02	0.00	0.00	0.01	0.08	0.01	0.01	0.03	0.19	0.46	0.32	0.24
FUNGI	0.09	0.00	0.03	0.00	1.45	0.12	0.02	0.01	0.23	0.17	0.19	0.30
HAPTOPHYTA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MALACOSTRACA	2.09	8.17	1.04	0.17	0.06	0.78	0.12	0.15	0.90	14.63	8.82	2.79
MOLLUSCA	3.77	4.51	39.97	8.04	3.88	7.38	9.56	1.96	3.89	5.15	4.72	2.81
OTHER ALVEOLATA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OTHER ARTHROPODA	2.66	0.06	0.09	0.03	27.23	0.50	0.65	7.12	12.53	9.54	8.47	7.86
OTHER EUKARYOTE	0.00	0.03	0.00	0.00	0.00	0.05	0.12	0.06	0.18	0.09	0.13	0.16
OTHER HACROBIA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OTHER MAXILOPODA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OTHER METAZOA	0.19	0.02	0.02	0.08	6.92	13.64	0.65	3.43	2.83	1.46	2.84	2.25
OTHER OCHROPHYTA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RHIZARIA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RHODOPHYTA	0.04	0.00	0.00	0.01	3.49	0.06	0.12	0.20	5.71	9.12	14.76	5.29
STRAMENOPILES	0.11	0.00	0.02	0.00	0.10	0.07	0.72	1.35	11.85	16.32	10.92	22.80
UROCHORDATA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	SALINITY 35											
	>200				20-200				0,22-20			
	June	July	September	October	June	July	September	October	June	July	September	October
AMOEBOZOA	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.21	0.09	0.03	0.00
ANNELIDA	1.41	0.21	0.05	1.32	8.95	0.49	0.62	3.09	0.68	4.86	2.26	3.57
BRANCHIOPODA	0.66	0.03	2.55	0.15	0.01	0.17	0.39	1.52	0.04	0.02	0.02	0.01
CHAROPHYTA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CHLOROPHYTA	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	7.73	4.22	5.98	5.76
CHORDATA	5.36	3.28	0.07	0.07	2.49	0.43	1.26	21.30	32.96	21.89	22.65	24.74
CILIOPHORA	0.00	0.01	0.04	0.00	0.00	2.57	0.05	0.00	0.00	0.00	0.00	0.00
CIRRIPEDA	55.16	90.10	27.22	3.21	5.22	49.24	5.81	0.01	0.26	2.33	0.02	0.01
CNIDARIA	4.75	0.01	2.17	2.09	0.68	4.94	9.42	0.69	4.11	17.82	12.30	14.57
COPEPODA	16.17	0.66	41.76	61.15	28.31	26.74	60.41	25.23	0.07	0.17	0.22	0.07
CRYPTOPHYCEAE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DIATOMEA	0.00	0.00	0.01	0.00	15.72	0.07	11.14	3.16	5.74	2.21	11.16	4.62
DINOFLAGELLATA	0.01	0.00	0.02	0.00	0.32	0.17	0.76	0.07	0.03	0.12	0.49	0.06
ECHINODERMATA	0.05	0.01	0.05	0.07	0.01	0.00	0.01	0.03	1.08	0.40	0.22	0.29
FUNGI	0.00	0.00	0.00	0.00	0.00	0.37	0.03	0.44	0.15	0.47	0.82	0.97
HAPTOPHYTA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MALACOSTRACA	1.35	4.45	8.16	3.58	0.01	0.00	0.03	2.38	0.98	6.03	1.82	1.56
MOLLUSCA	14.14	0.94	12.94	23.91	29.51	1.84	0.49	0.59	5.82	4.41	2.92	2.85
OTHER ALVEOLATA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OTHER ARTHROPODA	0.32	0.26	3.42	3.84	7.56	10.85	7.14	41.20	14.11	9.10	7.12	8.48
OTHER EUKARYOTE	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.16	0.04	0.27	0.60
OTHER HACROBIA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OTHER MAXILOPODA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OTHER METAZOA	0.56	0.00	1.51	0.30	0.33	1.99	1.79	0.01	11.43	4.75	8.14	5.40
OTHER OCHROPHYTA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RHIZARIA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RHODOPHYTA	0.04	0.02	0.03	0.18	0.09	0.11	0.36	0.19	5.01	10.19	14.11	9.04
STRAMENOPILES	0.00	0.00	0.01	0.12	0.68	0.02	0.20	0.08	9.40	10.88	9.45	17.39
UROCHORDATA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Capítulo 3

PAPER 3: 18S V9 Metabarcoding Correctly Depicts Plankton Estuarine Community Drivers

David Abad, Aitor Albaina, Mikel Aguirre and Andone Estonba

2017 Marine Ecology Progress Series, 584:31-43

Abstract

Metabarcoding is a time-saving and cost-effective approach that promises to overcome issues associated with traditional plankton taxonomy (i.e., lack of specialized personnel, time-consuming methodologies, difficulties in assignment of larval stages and detection of cryptic species). In this regard, we applied metabarcoding using the18S rDNA V9 region to samples collected throughout one year from the Estuary of Bilbao (Basque Country, Spain) in order to characterize the annual cycle of the eukaryotic plankton community. We found clear patterns of spatial and seasonal environmental variability that drive the distribution and abundance of the plankton assemblage throughout the year, thus confirming results of previous studies using microscopic identification of the planktonic species. Our results also suggest that the low oxygen period during summer in the inner part of the estuary (salinity 30) and the thermal variation from winter to summer are among the main environmental drivers of the plankton community of the Estuary of Bilbao. Finally, we report misidentification of some species (e.g. *Cyclopina gracilis*, *Maristentor dinoferus*), which highlights the need for more comprehensive reference sequence databases in order to overcome this limitation of metabarcoding.

Introduction

Estuaries are ecosystems of high interest for ecological and conservation studies because of their transitional nature, complex dynamics and species richness (McLusky & Elliott 2004). These environments are characterized by salinity gradients and physicochemical parameters, such as temperature or dissolved oxygen concentration, that show a clear seasonality (e.g. Moran et al. 2013). Moreover, estuaries are among the most threatened habitats worldwide; their closeness to cities and harbors contributes to habitat alteration and changes in the structure and dynamics of biotic communities (Kennish 2002), mainly due to pollution and/or introduction of non-indigenous species.

In terms of abundance and biomass, estuarine water masses are dominated by planktonic communities. They are essential for the functioning of the ecosystem, playing a crucial role in food webs and biogeochemical cycles (Ward et al. 2012). Although previous studies have provided evidence that salinity is one of the main variables driving variation of these communities (e.g. Kimmerer 2002, Muylaert et al. 2009), precipitation or temperature variation have also been related to shifts in the structure and composition of estuarine plankton assemblages (e.g. Shen et al. 2011). The ability of plankton to rapidly respond to these environmental shifts is precisely the reason why they have been used as indicators of ecosystem change for monitoring purposes (Taylor et al. 2002).

Ecosystem monitoring programs rely on robust information regarding species composition. Until recently, the identification of planktonic organisms has relied on the observation of morphological characteristics by means of microscopy. Apart from the complexity and expertise required for this task, many plankton communities are often comprised of a few dominant species and numerous very rare species which increases the difficulty to detect and identify all taxa (e.g. Cheung et al. 2010). In this context, metabarcoding has emerged during the last few years as a promising approach for the characterization of species composition in a diverse range of aquatic community samples (e.g. Lindeque et al. 2013, Logares et al. 2014, Hirai et al. 2015, Abad et al. 2016, Aguirre et al. 2017): the capability to generate millions of sequences from a sole sample at affordable costs along with its high sensitivity (capable of detecting DNA traces) and, at least, comparable taxonomic resolution provided an alternative to surmount the issues associated with traditional monitoring (Baird & Hajibabaei 2012, Zhan et al. 2013).

The Estuary of Bilbao, situated in the Basque Country (south Bay of Biscay), is a ~20 kilometers long channel that crosses a metropolitan area of about 1 million inhabitants and several industrial zones before flowing into Cantabrian Sea (Uriarte et al. 2014). The land reclamation (especially since the mid-19th century) together with the pollution coming from the city of Bilbao and surrounding factories reduced the original estuary and modified the ecosystem (Cearreta et al. 2000), altering abiotic processes and seasonal patterns in the planktonic community (Uriarte et al. 2014). All this factors transformed the Estuary of Bilbao in one of the most polluted estuaries in Europe. However, since 1979 the estuary has been subjected to a Comprehensive Plan for the Sanitation of the Metropolitan Area of Bilbao. Although the pollutant concentrations are still significant, it has resulted in a notable improvement of water and sediment quality, and recovery of biodiversity (see Cajaraville et al. 2016, for a review).

Except for short periods of high river discharge, euhaline waters (salinity >30) dominate within the estuary (Villate et al. 2013). Finally, the seasonal patterns of this estuary are determined mostly by temperature and precipitation; between November-May the temperatures are lower and precipitation is higher than during the rest of the year.

In the present study, we used the hyper-variable V9 region of the nuclear 18S rDNA gene (hereafter 18S V9) to characterize the planktonic eukaryotic community associated with the inner (salinity 30) and outer (salinity 35) areas of the Estuary of Bilbao. We conducted temporal monitoring by collecting samples throughout an annual cycle in order to define the key determinants that drive seasonal changes in plankton community structure. Finally, we describe the effect of the different seasonal periods on these communities and compare the findings with those previously reported by microscopy-based surveys (e.g. Villate 1994, Albaina et al. 2009).

Materials and methods

Sampling

Sampling was carried out from September 2013 to September 2014 in areas with 30 and 35 salinities during neap tides; January and February collections in 35 salinity were not possible due to bad weather conditions. As the distribution and depth of each salinity mass varied from season to season (Fig.2 in Intxausti et al. 2012 for further information), measurements were made every 0.5 m depth in order to define the water column profile. Samples were then collected with Niskin bottles and a 200 μm mesh net when the desired salinity mass was reached (sampling depths ranging from 2 to 10 m). Once in the laboratory, three plankton size-fractions were obtained (0.22-20 μm, 20-200 μm, and >200 μm). While the latter came directly from the plankton net, the Niskin bottle samples were pre-screened with a 200 μm mesh prior to the processing of the two lower size fractions (see Abad et al. 2016 for further details). Apart from this, water samples for chlorophyll *a* determination (Jeffrey & Mantoura 1997) were collected with Niskin bottles also at each salinity. Furthermore, the values for the different environmental variables and physico-chemical parameters (temperature, precipitation, pH, …) were measured.

Metabarcoding

DNA was extracted using a modified salt protocol for the 20-200 μm and >200 μm size fractions, and a commercial kit (MOBIO PowerSoil®) for the 0.22 μm filters. The 18S V9 region (~150-bp) was amplified using the primers 1391f and EukBr from Stoeck et al. 2010. Sequencing data of the samples corresponding to September and October 2013 have already been published (Abad et al. 2016). The rest of the samples were sequenced at the SGIKER facilities of the University of the Basque Country (UPV/EHU) using Illumina MiSeq 2x150 bp (sequencing information is available at the Sequence Read Archive (SRA); https:// www. ncbi. nlm. nih. gov/ sra; PRJNA385805).

Raw reads were pre-processed (trimming, paired-end merging and chimera removal) with Sickle v1.33 (quality threshold = 20; Joshi & Fass 2011), Pear v0.9.5 (minimum overlap of 15 bp and a cut-off P-value of 0.01; Zhang et al. 2014) and UCHIME (using our custom database; Edgar et al. 2011), respectively. The resulting reads were clustered into Operational Taxonomic Units (OTUs) with UCLUST (Edgar 2010) in Qiime v1.9 (Caporaso et al. 2010), using a *de novo* approach with a 99% identity threshold (Abad et al. 2016). OTU taxonomic assignment was performed with BLAST (Altschul et al. 1990) with a minimum of 90% identity, against a Silva 119 custom database (with

the addition of representative sequences from key local species from the Estuary of Bilbao; Abad et al. 2016). Finally, a core community analysis was performed to detect the OTUs present in at least 90% of the samples collected from each water mass over the annual cycle.

Statistical analysis

A supervised learning analysis (confusion matrix; Table 1) was performed using a Random Forest classifier (Knights et al. 2011) with OTUs as predictors and size fractions as class labels: this method uses a subset of samples to train a model that identifies unique features within communities to predict putative similarities among size fraction. In addition, to determine the community dissimilarity for each sample, a Bray-Curtis distance network was carried out using the Phyloseq v1.14 (McMurdie and Holmes 2013) R package. Alpha diversities were also calculated with Phyloseq. The bar charts representing relative abundances (Fig. 2) and alpha diversities (Fig. 5) were created with ggplot2 (Wickham 2009) in R (R Core Team 2017).

Finally, the Canonical Correspondence Analysis (CCA) of the OTUs showing >5% relative abundance of reads in a particular sample were carried out using CANOCO v4.5 (ter Braak & Smilauer 2002); square root transformations were used to normalize data among samples.

Results

OTU assignment

Only 0.74% of the reads were lost after quality filtering and 0.14% were eliminated due to their putative chimeric nature, resulting in 3848144 total reads (64136 ± 20729 reads/sample). After read clustering, 4984 OTUs with assigned taxonomy were obtained after clustering. In all, 1859 singletons were discarded, yielding a total of 3125 OTUs for further analysis.

The core community analysis revealed that only 11 and 8 OTUs were present in at least 90% of the samples throughout the annual cycle collected from 30 and 35 salinities, respectively. Six of these OTUs (*Acartia clausi*, *Acartia tonsa*, *Calanipeda aquaedulcis*, *Cyclopina gracilis*, *Stomatolepas praegustator* and Appendicularia) were shared between the two salinities, whereas only five (*Hyperamoeba flagellata,* Chrysophyceae, Paraphysomonas, *Adula californiensis*, *Polydora ciliata*) and two OTUs (*Paracalanus parvus* and Maxillopoda) were unique for the 30 and 35 samples, respectively.

Size-fraction similarity

The Bray-Curtis distance network (Fig 1) showed that the communities from the >200 and 20-200 μm size-fractions were more similar to each other than to the 0.22-20 μm one. To further support this result, the machine-learning based classification was carried out to determine the variability of each size fraction: the model showed that the 0.22-20 μm size-fraction had a high similarity and that all the samples grouped together. Although the samples from the other two size fractions were usually classified together (as shown in Table 1), there were some errors in classification (four for the >200 μm and three for the 20-200 μm; 0.167 and 0.125 class error respectively,

Size	>200	20-200	0.22-20	Class. error
>200	0.6524 ± 0.1515	0.3348 ± 0.1423	0.0128 ± 0.0135	0.166666667
20-200	0.3387 ± 0.1201	0.5828 ± 0.0964	0.0785 ± 0.0602	0.125
0.22-20	0.0567 ± 0.0398	0.1487 ± 0.0772	0.7946 ± 0.1129	0

Table 1: Confusion matrix for the size-fraction of the Estuary of Bilbao based on all samples collected during the annual cycle. The classification error from machine learning was defined by the proportion of samples that were not clustered into their own size-fraction; the higher the value of the classification error, the lower the similarity of that size. Values are ±SD

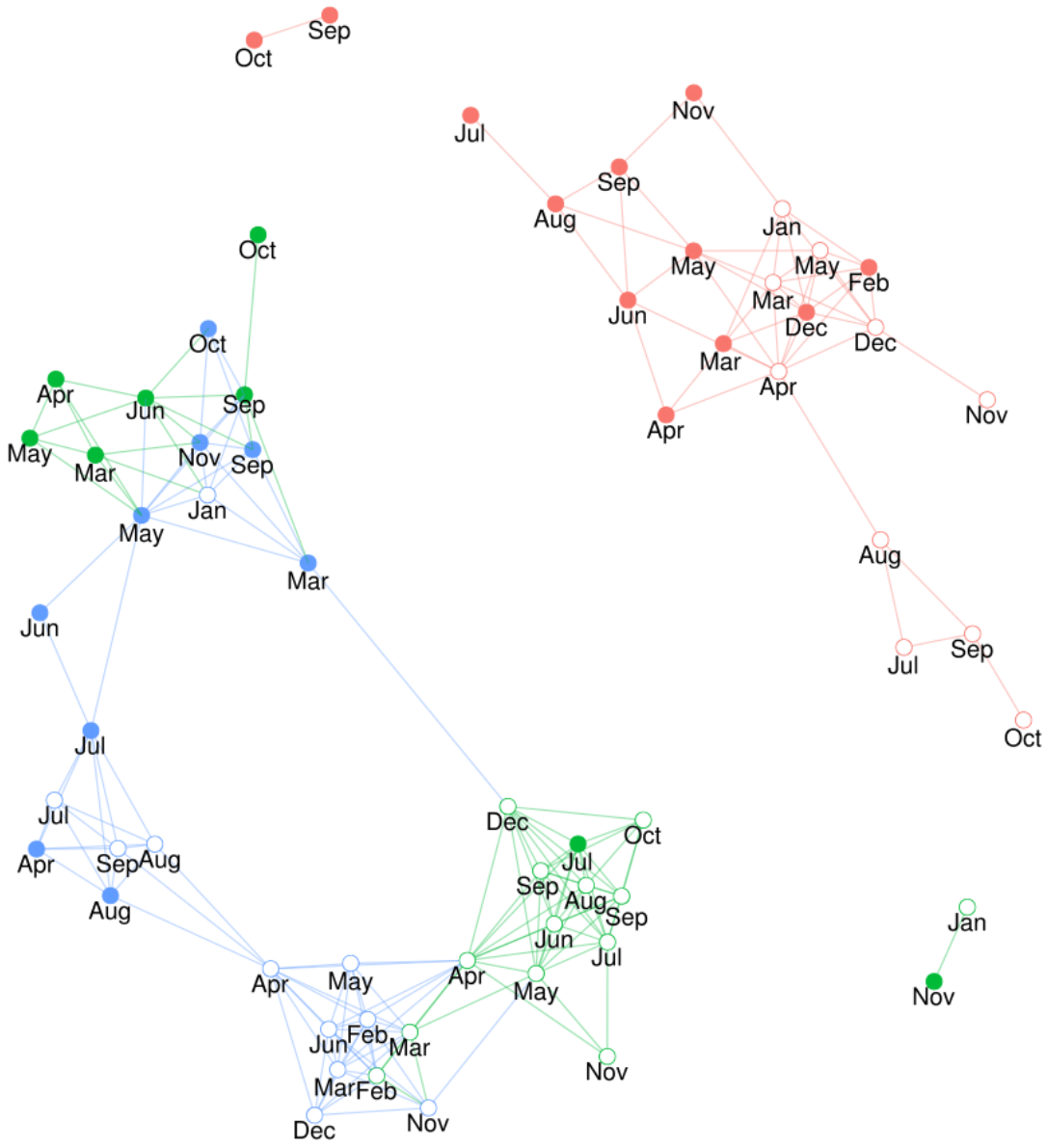


Fig 1. Bray distance network

Each node represents a specific sample (empty and full for 30 and 35 salinity, respectively). Different colors are used for the >200, 20-200 and 0.22-20 µm size fractions (blue, green and red, respectively).

Taxonomic composition

The OTUs identified by metabarcoding were assigned to 29 categories (Fig. 2; see Table S1 for detailed relative abundances). Maxillopoda was the most frequently observed group in the 20-200 and >200 µm size fractions for samples from both salinities: more concretely, copepods represented a 51.7% and 57.1% of the OTUs, respectively. Chrysophyceae (11.7%) and Cirripedia (15.6%) were

the second most abundant groups in the 20-200 µm and >200 size fractions, respectively.

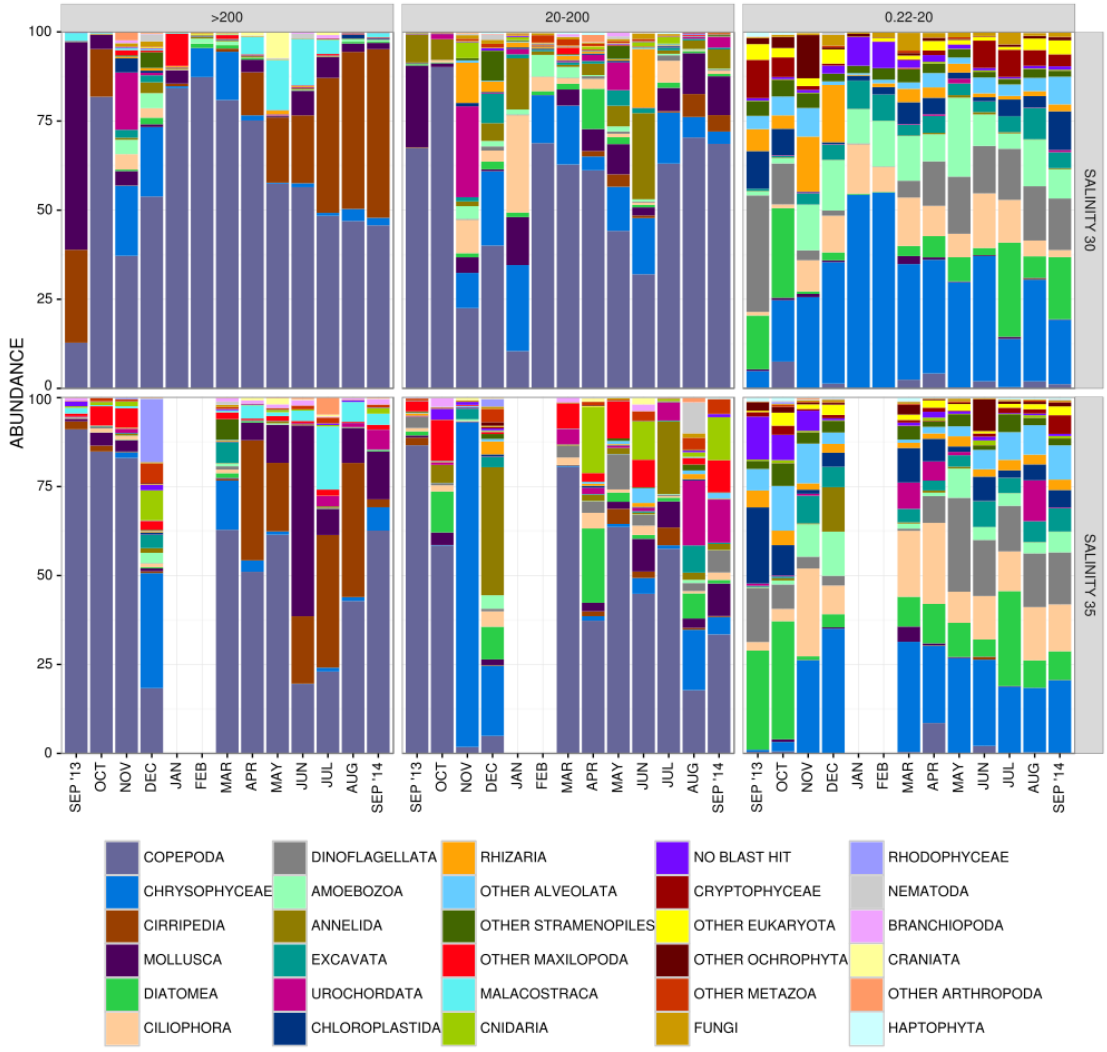


Fig 2. Proportions of taxonomic groups by sample A total of 29 taxonomic groups are shown in this graphic. Samples are arranged by size-fraction (>200, 20-200, and 0.22-20 µm) and salinity (30 and 35). A category with sequences that had no database match is labeled "no blast hit". Samples from January and February in 35 salinity are missing due to bad weather.

The >200 and 20-200 µm communities

Cyclopina gracilis was the most abundant copepod species in salinity 30 (16.1% of the total relative abundance combined for both size-fractions), followed closely by *Calanipeda aquaedulcis* (15.4%); the third most abundant species was *Acartia tonsa* (11.6%). For this salinity, there was a clear succession between *A. tonsa* and *C. gracilis* from the end of summer to the beginning of autumn, followed by the dominance of *C. aquaedulcis* during late winter and spring (Fig. 3); there

was also a peak of *A. clausi* during winter (3.5%).

Acartia clausi dominated the samples from salinity 35, accounting for 18.3% of the total combined abundance, while *Paracalanus parvus* was the secondary dominant species at 8.8%. There was also replacement of both species by *Acartia tonsa* (3.6%) and *Cyclopina gracilis* (1.7%) during spring and summer-autumn, respectively, but the pattern was not as clear as in the sample from salinity 30 (Fig. 3).

Although copepods dominated these assemblages, a barnacle bloom was observed in the >200 µm size-fraction in both salinities during April/May (Fig. 2) and was dominated mainly by the species Stomatolepas praegustator (13.1%).

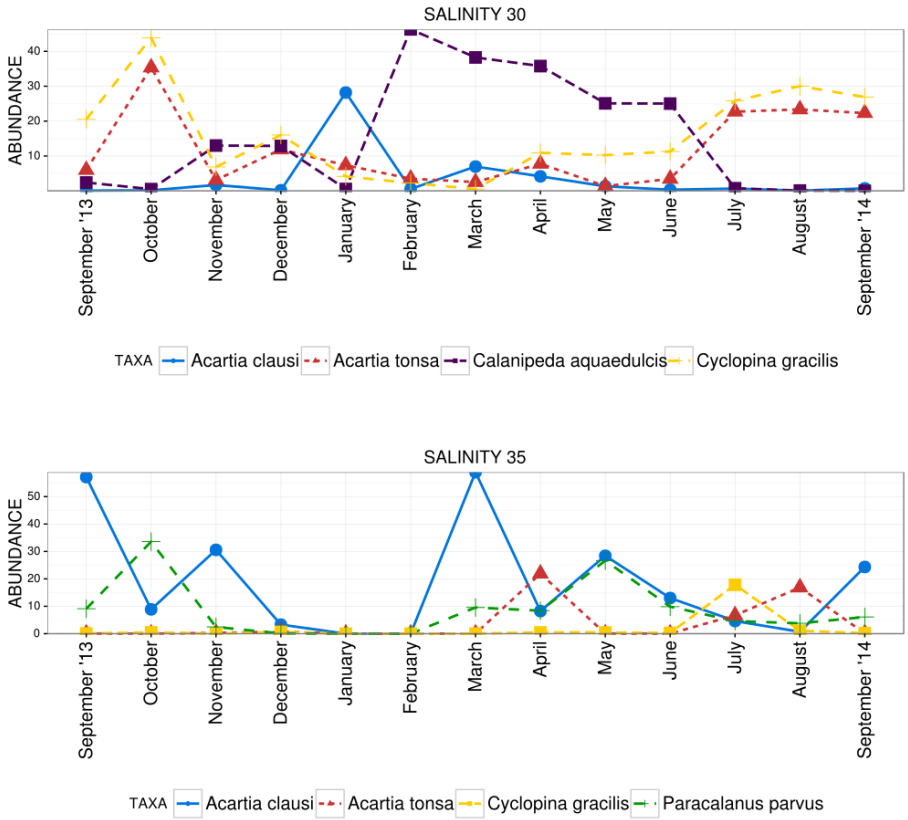


Fig 3. Relative abundances (percentage of reads) of copepods during the annual cycle The six most abundant copepods of the community are shown by sample, arranged by salinity. Each species has a distinct color and symbol shape. The data included here corresponds to the 20-200 and >200 µm size fractions.

The 0.22-20 µm community

A more diverse assemblage characterized the 0.22-20 µm size-fraction, as shown in the taxonomic composition (Fig. 2); furthermore, the network analysis (Fig. 1) also showed that the communities from both salinities were quite similar. Chrysophytes (23.3% of the total relative abundance for the combined salinities) were the most abundant group of the whole community. Phytoplankton components such as diatoms and cryptophytes accounted for 9.7% and 2.3% of the total community, respectively. Dinoflagellates (10%) and ciliates (9.4%) were also abundant.

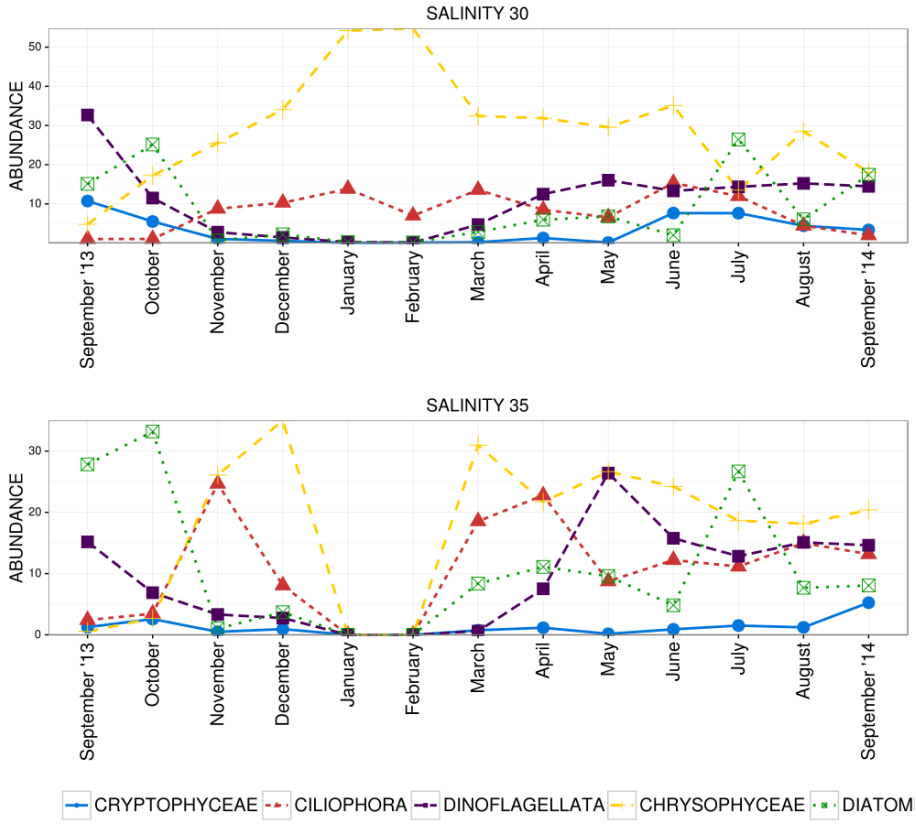


Fig 4. Annual cycle of the 0.22-20 µm size-fraction community Relative abundances (percentage of reads) of the five most abundant groups from this size fraction are shown in this plot. Samples are arranged by salinity. Groups are indicated by symbols of different colors and shapes.

The dominant diatoms were the species *Papiliocellulus elegans* (2%) and the genus *Skeletonema* (1.8%). Among the dinoflagellates, the genera *Gyrodinium* (5.4%) and *Protoperidinium* (1.7%) were the most abundant. The species *Maristentor dinoferus* was the main ciliate, accounting for

6.4% of the community. Finally, the heterotrophic genus *Paraphysomonas* (23.4%) was the most abundant in these samples and the dominant among the chrysophyceans. In this size fraction the chrysophytes became the dominant group in both salinities during the whole year, with punctual exceptions of diatoms (Fig 4). The similarity between communities was also reflected in the annual cycle (Fig 4); the only exception was the ciliate group in salinity 35 during winter and the some occasional peaks from other groups (e.g. dinoflagellates or diatoms).

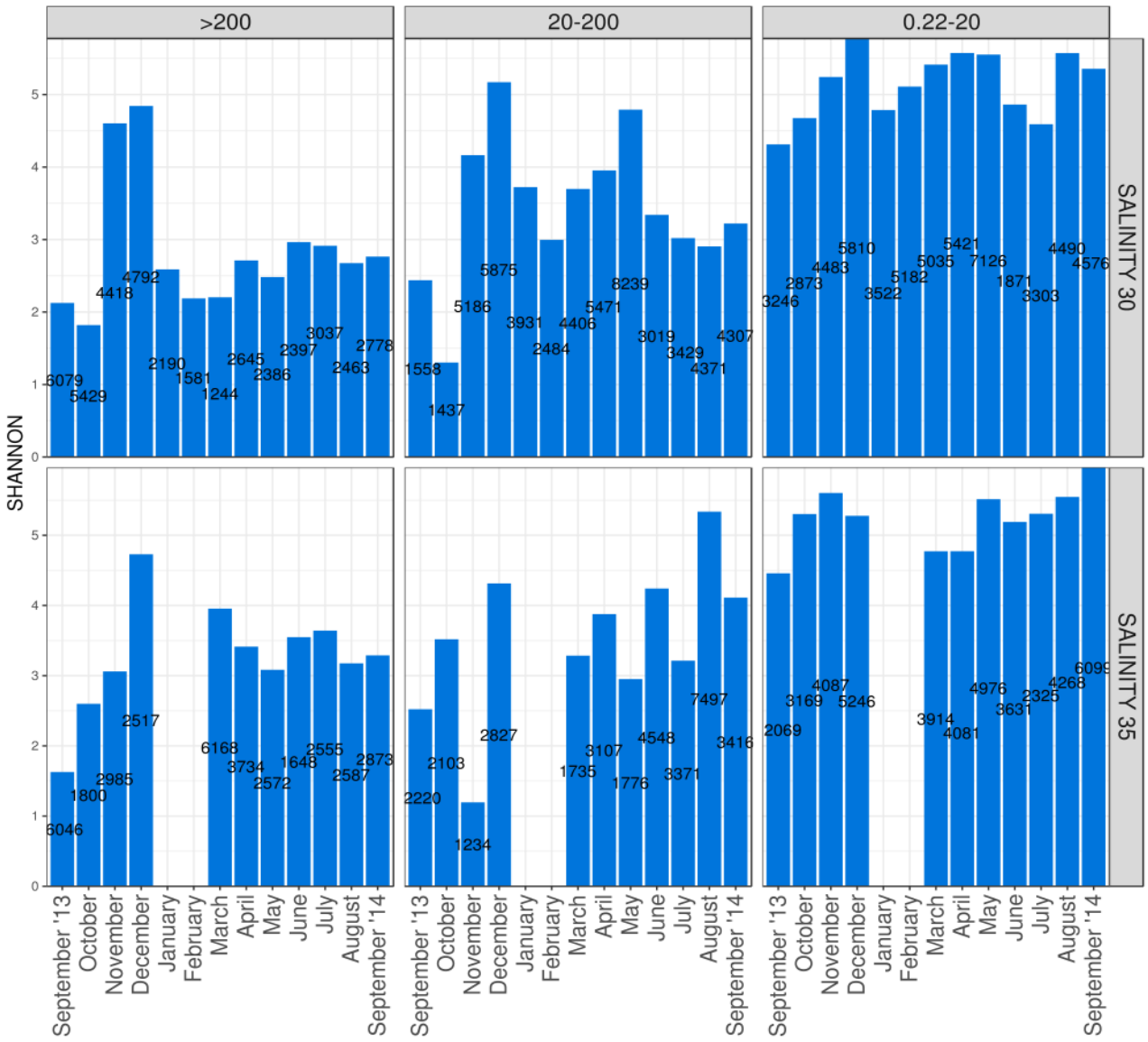


Fig 5. Alpha diversities over the annual cycle from September 2013 to September 2014. Values shown are the Shannon diversity index for each sample, based on OTU relative abundances. Samples are organized by size fraction (0.22-20, 20-200 and >200 µm), salinity (30, 35) and month sampled. Number of OTUs per sample is included within each column.

The Shannon index for each sample is represented in Figure 5. Overall, there are higher values and more homogeneous diversities throughout the year in the 0.22-20 μm size fraction. Instead, the 20-200 and the >200 μm samples showed greater fluctuation in the samples from salinity 35 than in salinity 30 ones.

Environmental drivers

A total of 64 taxa, which contributed to a minimum of 5% relative abundance in at least one sample, made up 80.3% of the total community throughout the annual cycle. The selected taxa for the CCA (Fig. 6) consisted primarily of copepods, diatoms and mollusks (see Table S2 for listing of groups).

According to the results of the forward selection procedure in CCA, all selected environmental variables (pH, DO, temperature, salinity, Chl-*a*, and precipitation) were significantly correlated with the most abundant OTUs of the plankton community (Fig. 6). Specifically, axis 1 explained 33.9% of the species-environment relation; this axis was strongly determined by DO, pH and salinity. Axis 2 explained 20.3% of the variation in the species data, which was determined by temperature, precipitation and Chl-*a*. The CCA analysis also showed that there is a clear spatial (salinity) and temporal (seasonal) separation for the most abundant taxa throughout the year: the samples from summer and autumn were grouped together for each salinity, as well as those from winter and spring. Higher dissolved oxygen (DO) and pH values were associated with salinity; the highest Chl-*a* peak was linked to temperature. As expected, precipitation varies in opposition with temperature.

According to the CCA of community composition and environmental variables (Fig. 6), the copepods *A. tonsa* and *C. gracilis*, barnacle *S. praegustator,* dinoflagellate genus *Gyrodinum* and diatom genus *Skeletonema* were inversely correlated with salinity and precipitation, but positively correlated with temperature; the copepod *C. aquaedulcis* and the chrysophyte genus *Paraphysomonas* had negative associations with temperature and salinity. The ciliate *M. dinoferus* and copepod *A. clausi* were positively correlated with salinity and precipitation, but negatively correlated with temperature. Finally, the copepod *P. parvus* was related positively with salinity.

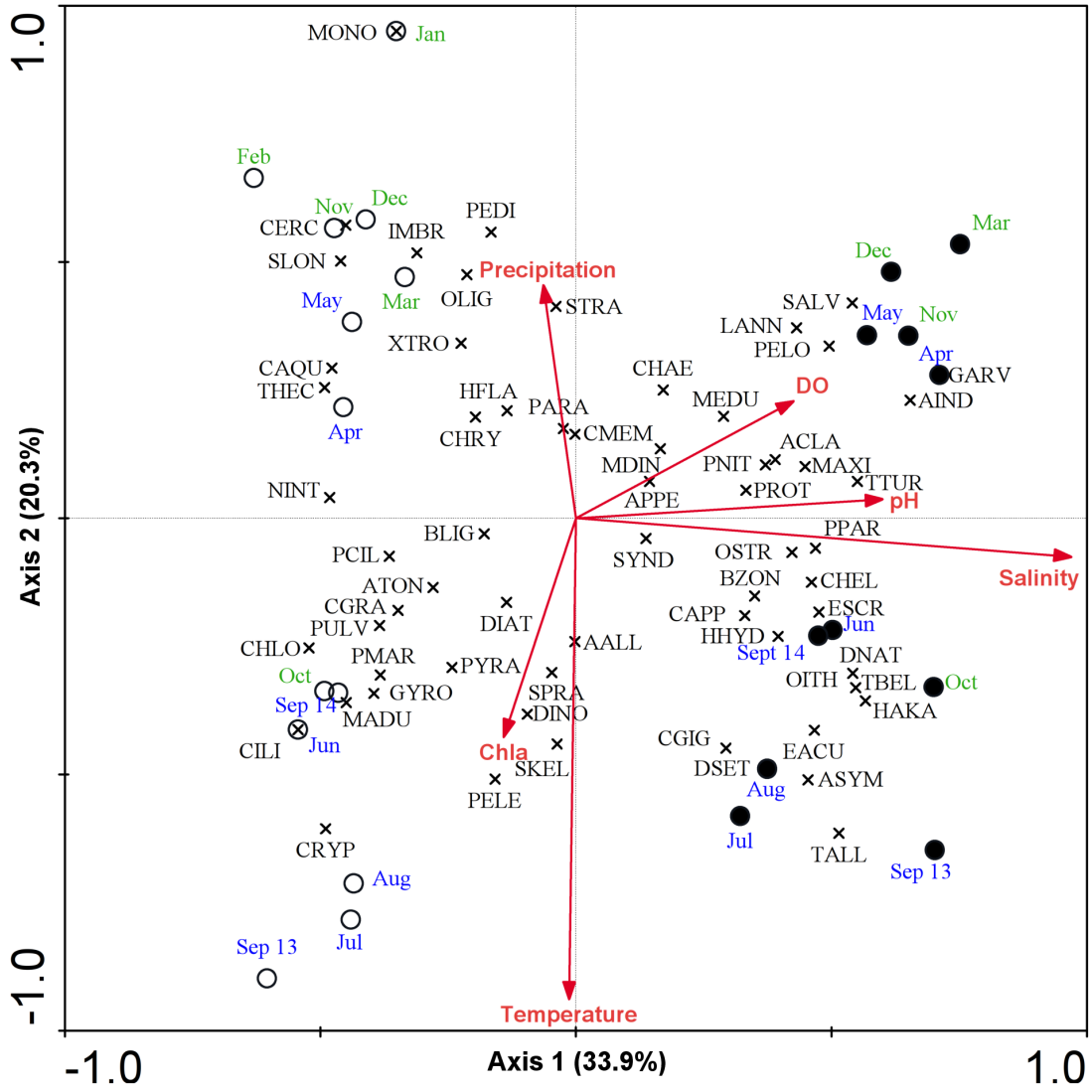


Fig 6. Multivariate analysis of the most abundant OTUs and environmental variables. Only OTUs representing > 5 % of the abundance in any size fraction in at least one sample were included in the analysis. OTUs are identified as cross-marks (see acronyms in SI_table_2); sampling stations are depicted by circles (empty and full for 30 and 35 salinity, respectively) with the particular sampling month depicted in blue (summer and spring stations) or green (winter and autumn stations). Environmental variable gradients are represented by red arrows.

Discussion

>200 and 20-200 µm community

Although the Estuary of Bilbao used to be one of the most polluted in Europe, its water/sediment quality has improved significantly and biodiversity has recovered well since 1979 (e.g. Villate et al. 2013). This transition from a polluted to a rehabilitated area has allowed the recolonization of the water column by a mixture of neritic and brackish-water species, including non-indigenous species (Aravena et al. 2009, Albaina et al. 2009, 2016a, Uriarte et al. 2015).

Our study shows that in this estuary there is clear dominance of the *Acartia* complex copepod species among the mesozooplankton, as demonstrated in previous morphological studies (e.g. Villate et al. 1994, Uriarte & Villate 2004, Albaina et al. 2009, Aravena et al. 2009, Uriarte et al. 2015). In our case, the 18S V9 region allowed us to decipher the current status of this complex, in which *A. tonsa* dominates during most of the year in samples collected from salinity 30, while *A. clausi* is the most abundant in salinity 35. This spatial separation of *A. clausi* and *A. tonsa* in higher and lower salinity waters, as well as the seasonal segregation, agrees with previous studies of the area (Fig 3; Aravena et al. 2009) and has also been observed in other estuaries (e.g. Azeiteiro et al. 2005, Gaudy et al. 2000). 18S V9 genetic marker has shown to provide sufficient taxonomic resolution for *Acartia* species (Abad et al. 2016); the observed discriminatory power probably related to resulting from isolation in this brackish-water genus (Chen & Hare 2008).

Furthermore, the copepod genera *Paracalanus, Clausocalanus, Pseudocalanus* and *Ctenocalanus* usually represente a large percentage of the abundance in planktonic communities in temperate waters and are commonly grouped due to identification difficulties (e.g. Albaina & Irigoien 2004, Gonçalves et al. 2012). Interestingly, the metabarcoding approach does not have the bias regarding the development stage or the cryptic species issue that affects this category, so potentially it would be able to estimate more accurately their abundance. Nonetheless, there are some cases in which the reliability of the 18S V9 for copepods is compromised: a 100% identity was detected between two *Centropages* species, as well as among eight copepod species corresponding to two sister families, Aetideidae and Euchaetidae (Albaina et al. 2016b).

Such inaccuracies of OTU assignment by metabarcoding could explain the finding of *Cyclopina*

gracilis for which there was no previous record in the area (Villate et al. 1997, Albaina et al. 2009, Uriarte et al. 2015). Given its abundance, this is most likely an error in taxonomic assignment, due to the absence of a comprehensive SILVA reference database. A search of the GenBank repository revealed that the sequence belonging to this OTU is most likely *Oithona davisae* (accession number: KJ814022) for which there is a recent citation in the Estuary of Bilbao (Uriarte et al. 2015). The *Oithona* genus is one of the main constituents of the >100 μm copepod assemblage in this system (Intxausti et al. 2012) and hence mostly falls into the microzooplankton, a fraction that has been less studied to date. The difficulty of identifying early stages of this genus implies that these organisms would be classified as *Oithona* spp. rather than to a particular species. As expected, this OTU is more abundant in the 20-200 μm size fraction (Table S2) and presents similar patterns in seasonal abundance, with peaks during summer/autumn, as described in previous studies of the area for *Oithona* spp (Intxausti et al. 2012, Uriarte et al. 2015).

Difficulties of identifying developmental stages and cryptic species are more evident within the microplankton fraction (20-200 μm): for example, in a zooplankton study of the Estuary of Bilbao carried out by Intxausti and colleagues, the identified organisms were grouped in broad taxonomic categories since some of the larval and immature forms (nauplii and copepodites) that dominated this lower size-fraction could not be assigned to species without time-consuming examination (Intxausti et al. 2012). Metabarcoding does not have this limitation, and is thus capable of assigning early stages to a certain taxonomic classification, as long as there is a reference sequence for the organism in the database. This is particularly useful for detecting Non-Indigenous Species at very low abundances (e.g. Abad et al. 2016).

Our finding of another abundant copepod species *Calanipeda aquaedulcis* also agreed with previous studies (Aravena et al. 2009, Albaina et al. 2009, Uriarte et al. 2015). This species has contributed significantly to the increase in the total number of copepods in the Estuary of Bilbao during the last few years (Uriarte et al. 2015). *C. aquaedulcis* is known to attain peak abundances from March (Uriarte et al. 2015) and, as shown in this study, until June. Apart from this, the seasonal succession of the inner estuary zooplankton assemblage in the present study corresponded to a low oxygen period that is commonly reported during part of the summer following stratification (Intxausti et al. 2012). This condition, along with an increase in temperature may have favored the

settlement of two species with a higher tolerance to some degree of hypoxia: *A. tonsa* and *O. davisae* (Itoh et al. 2011, Roman et al. 1993). During winter and spring, the dominance shifts to *C. aquaedulcis*, which is considered to be eurythermal but usually prefers cooler temperatures (Frisch et al. 2006).

0.22-20 µm community

Previous studies of the picoplankton in the Estuary of Bilbao have focused mainly on taxonomic or phylogenetic analysis of specific groups (Seoane et al. 2005, Laza-Martinez et al. 2007, Orive et al. 2010, Hevia-Orube et al. 2016), since the time and cost constraints of morphological identification prevented studies entailing analysis of samples with sufficient volume required to detect whole community´s spatial and temporal cycles. Metabarcoding using the 18S V9 region, although subject to the aforementioned taxonomic resolution limitations (but see also the genus *Acartia* case), allowed us to analyze the entire community assemblage through a year, and thereby to reveal previously unreported patterns of variation.

Our results showed that the chrysophytes are the most abundant group throughout the year: the heterotrophic *Paraphysomonas* was the dominant genus, not unusual for partially eutrophic estuaries (Bazin et al. 2014). Colorless chrysomonads, such as *Paraphysomonas*, are the major phagotrophs in freshwater and soil food webs, but they are also widespread in marine environments (Scoble & Cavalier-Smith 2014), and have been found in the Bay of Biscay (Artolozaga et al. 2000) and the Estuary of Bilbao (Cajaraville et al. 2016).

Furthermore, the naked dinoflagellate genus *Gyrodinium* is among the least-known groups of marine protists (Kubiszyn & Wiktor 2015). In contrast, *Protoperidinium* is a large and ubiquitous genus of marine heterotrophic dinoflagellates, whose species typically follow diatom blooms and generally exhibit coastal distributions (Taylor 1990). Both genera were previously described in other studies of the area, but were not followed during a complete year, as in this study (e.g. Seoane et al. 2005).

Among the diatoms, the tiny *Papiliocellulus elegans* is a marine organism commonly found in coastal environments. Its small size requires electron microscopy for its identification, so this

species has typically been overlooked and the extent of its habitat is not yet well known, although it has been regarded as possibly planktonic (Round et al. 1990). It could be present in the Estuary of Bilbao but it has not been previously reported. On the other hand, the genus *Skeletonema* occurs in coastal waters throughout the world, where it can be extremely common (Round et al. 1990) and is usually found in this estuary (Seoane et al. 2005, Laza-Martinez et al. 2007, Hevia-Orube et al. 2016).

Finally, the benthic species *Maristentor dinoferus* was the dominant ciliate but, as in the case of *C. gracilis*, this is most likely an error of taxonomic assignment, because this organism was recently discovered on coral reefs (Lobban et al. 2002). A GenBank search resulted in matches of the sequences belonging to this OTU to uncultured phytoplankton, so it is entirely possible that this could be another species. Completion of a reference database is needed to solve problems associated with taxonomic identification by metabarcoding (e.g. Abad et al. 2016, Albaina et al. 2016b).

In this size fraction, there is little variation between the community compositions of both salinities, suggesting that the low oxygen period does not have the same influence as in the zooplankton. *Paraphysomonas*, the most abundant group throughout the year, are important feeders on bacteria (but not exclusively restricted to them) and peaked during the coldest months in the Estuary of Bilbao (November-December to March), when the lack of nutrients and sunlight prevents the proliferation of autotrophic phytoplankton and turbulence can increase grazing rates of protozoa on bacteria (Rose & Caron 2007). Diatoms, on the other hand, showed peaks in abundance during the summer (July to October), when the temperature was higher and precipitation resulted in nutrient input from the tributaries. Non-photosynthetic species of dinoflagellates feed on diatoms or other protists (Jeong et al. 2010), which would explain why they begin to be more abundant during spring (Cajaraville et al. 2016). Finally, ciliates seem to have a peak during winter in samples collected from salinity 35 (Fig 3), but we cannot be certain due to the lack of data from January and February.

Conclusions

The metabarcoding analysis of the plankton communities present in the Estuary of Bilbao revealed that their distribution and abundance throughout the year were due to spatial and seasonal environmental variability, confirming results of previous studies using traditional techniques. The low oxygen period during summer in salinity 30 and the thermal variation from winter to summer are among the main environmental drivers of zooplankton, while temperature and precipitation are for phytoplankton. Furthermore, we also reported misidentification of some species (e.g. *Cyclopina gracilis*, *Maristentor dinoferus*), which highlights the need of completing reference sequence databases as to overcome this limitation. In the light of these results we think that metabarcoding can be useful for plankton monitoring, but that the findings obtained should be interpreted carefully until further improvement of the approach.

Acknowledgements

The authors thank Ann Bucklin (University of Connecticut) for her comments on the manuscript. The authors also thank the SGIker (UPV/EHU) for the technical and human support provided, and the Hydro-meteorology Service of the Regional Council of Bizkaia for the precipitation data.

Funding

AA and DA work was supported by a contract with the Euskampus Foundation (Euskampus Fundazioa) and a Basque Government doctoral fellowship (UPV/EHU "ZabaldUz" program), respectively. This study was funded by a Basque Government grant "Grupo de Investigación Consolidado del Sistema Universitario Vasco" to support activities of the Genomic Resources Research Group (IT-558-10). Sampling was also partly financed by the Euskampus Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Abad D, Albaina A, Aguirre M, Laza-Martínez A and others (2016) Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. Mar Biol 163:149, 1–13. doi:10.1007/s00227-016-2920-0

Aguirre M, Abad D, Albaina A, Cralle L, Goñi-Urriza MS, Estonba A, Zarraonaindia I (2017) Unraveling the environmental and anthropogenic drivers of bacterial community changes in the Estuary of Bilbao and its tributaries. PLOS ONE 12: e0178755

Albaina A, Irigoien X (2004) Relationships between frontal structures and zooplankton communities along a cross-shelf transect in the Bay of Biscay (1995 to 2003). Mar Ecol Prog Ser 284.

Albaina A, Villate F, Uriarte I (2009) Zooplankton communities in two contrasting Basque estuaries (1999-2001): reporting changes associated with ecosystem health. J Plankton Res 31, 739–752. doi:10.1093/plankt/fbp025

Albaina A, Uriarte I, Aguirre M, Abad D, Iriarte A, Villate F, Estonba A (2016a) Insights on the origin of invasive copepods colonizing Basque estuaries; a DNA barcoding approach. Marine Biodiversity Records 9, 51. doi:10.1186/s41200-016-0045-2

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

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Res 25, 4692–4693.

Almeida MA, Cunha MA, Dias JM (2007) Bacterial productivity distribution during a rainy year in an estuarine system. Microb Ecol 53, 208–220. doi:10.1007/s00248-005-0082-6

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J

Mol Biol 215, 403–410. doi:10.1016/S0022-2836(05)80360-2

Aravena G, Villate F, Uriarte I, Iriarte A, Ibáñez B (2009). Response of Acartia populations to environmental variability and effects of invasive congenerics in the estuary of Bilbao, Bay of Biscay. Estuar Coast Shelf Sci 83, 621–628. doi:10.1016/j.ecss.2009.05.013

Azeiteiro UM, Marques SC, Vieira LMR, Pastorinho MRD, Ré PAB, Pereira MJ, Morgado FMR (2005) Dynamics of the Acartia genus (Calanoida: Copepoda) in a temperate shallow estuary (the Mondego estuary). Acta Adriat 46 (1).

Baird DJ, Hajibabaei M (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Mol Ecol 21, 2039–2044.

Bazin P, Jouenne F, Friedl T, Deton-Cabanillas AF, Le Roy B, Véron B (2014) Phytoplankton Diversity and Community Composition along the Estuarine Gradient of a Temperate Macrotidal Ecosystem: Combined Morphological and Molecular Approaches. PLoS ONE 9, e94110. doi:10.1371/journal.pone.0094110

Cajaraville MP, Orive E, Villate F, Laza-Martínez A and others (2016) Health status of the Bilbao estuary: A review of data from a multidisciplinary approach. Estuar. Coast. Shelf Sci., Special Issue: Functioning and dysfunctioning of Marine and Brackish Ecosystems 179, 124–134. doi:10.1016/j.ecss.2016.01.013

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K and others (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7(5):335-6. doi: 10.1038/nmeth.f.303

Cearreta A, Irabien MJ, Leorri E, Yusta I, Croudace IW, Cundy AB (2000) Recent Anthropogenic Impacts on the Bilbao Estuary, Northern Spain: Geochemical and Microfaunal Evidence. Estuar. Coast Shelf Sci 50, 571–592. doi:10.1006/ecss.1999.0582

Chen G, Hare MP (2008) Cryptic ecological diversification of a planktonic estuarine copepod, Acartia tonsa. Mol Ecol 17, 1451–1468.

Cheung MK, Au CH, Chu KH, Kwan HS, Wong CK (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. ISME J 4, 1053–1059. doi:10.1038/ismej.2010.26

Comtet T, Sandionigi A, Viard F, Casiraghi M (2015) DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. Biol Invasions 17: 905-922.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.

Edgar RC, Haas BJ, Clememte JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200.

Fonseca VG, Carvalho GR, Sung W, Johnson HF and others (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. Nat Commun 1, 98. doi:10.1038/ncomms1095

Frisch D, Rodríguez-Pérez H, Green AJ (2006) Invasion of artificial ponds in Doñana Natural Park, southwest Spain, by an exotic estuarine copepod. Aquat. Conserv. Mar Freshw Ecosyst 16, 483–492. doi:10.1002/aqc.718

Gaudy R, Cervetto G, Pagano M (2000) Comparison of the metabolism of Acartia clausi and A. tonsa: influence of temperature and salinity. J. Exp. Mar Biol Ecol 247, 51–65. doi:10.1016/S0022-0981(00)00139-8

Gonçalves AM, Pardal MA, Marques SC, Mendes S, Fernández-Gómez MJ, Galindo-Villardón MP, Azeiteiro UM (2012) Responses of Copepoda life-history stages to climatic variability in a Southern-European temperate estuary. Zool Stud 51, 321–335.

Herlemann DP, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J 5, 1571–1579. doi:10.1038/ismej.2011.41

Hevia-Orube J, Orive E, David H, Díez A, Laza-Martínez A, Miguel I, Seoane S (2016) Molecular and morphological analyses of solitary forms of brackish Thalassiosiroid diatoms (Coscinodiscophyceae), with emphasis on their phenotypic plasticity. Eur J Phycol 51, 11–30. doi:10.1080/09670262.2015.1077394

Hirai J, Kuriyama M, Ichikawa T, Hidaka K, Tsuda A (2015) A metagenetic approach for revealing community structure of marine planktonic copepods. Mol Ecol Res 15, 68-80.

Intxausti L, Villate F, Uriarte I, Iriarte A, Ameztoy I (2012) Size-related response of zooplankton to hydroclimatic variability and water-quality in an organically polluted estuary of the Basque coast (Bay of Biscay). J Mar Syst 94, 87–96. doi:10.1016/j.jmarsys.2011.10.015

Itoh H, Tachibana A, Nomura H, Tanaka Y, Furota T, Ishimaru T (2011) Vertical distribution of planktonic copepods in Tokyo Bay in summer. Plankton Benthos Res 6, 129–134. doi:10.3800/pbr.6.129

Jeffrey SW, Mantoura RFC. (1997) Development of pigment methods for oceanography: SCOR-supported working groups and objectives. In: Phytoplankton pigments in ceanography: guidelines to modern methods. Monographs on Oceanographic Methodology. UNESCO Publishing; 19-36.

Jeong HJ, Yoo YD, Kim JS, Seong KA, Kang NS, Kim TH (2010) Growth, feeding and ecological roles of the mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. Ocean Sci J 45, 65–91. doi:10.1007/s12601-010-0007-2

Joshi NA, Fass JN (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (v1.33). https://github.com/najoshi/sickle

Kennish MJ (2002) Environmental threats and environmental future of estuaries. Environ Conserv 29:78-107 doi:10.1017/S0376892902000061

Kimmerer WJ (2002) Effects of freshwater flow on abundance of estuarine organisms: physical effects or trophic linkages? Mar Ecol Prog Ser 243, 39–55. doi:10.3354/meps243039

Knights D, Costello EK, Knight R (2011) Supervised classification of human microbiota. FEMS

Microbiol Rev 35, 343–359

Kubiszyn AM, Wiktor JM (2015) The Gymnodinium and Gyrodinium (Dinoflagellata: Gymnodiniaceae) of the West Spitsbergen waters (1999–2010): biodiversity and morphological description of unidentified species. Polar Biol 1–9. doi:10.1007/s00300-015-1764-2

Laza-Martinez A, Seoane S, Zapata M, Orive E (2007) Phytoplankton pigment patterns in a temperate estuary: from unialgal cultures to natural assemblages. J. Plankton Res 29, 913–929. doi:10.1093/plankt/fbm069

Lindeque PK, Hay SJ, Heath MR, Ingvarsdottir A and others (2006) Integrating conventional microscopy and molecular analysis to analyse the abundance and distribution of four Calanus congeners in the North Atlantic. J Plankton Res 28: 221-238.

Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, Atkinson A (2013) Next Generation Sequencing Reveals the Hidden Diversity of Zooplankton Assemblages. PLoS ONE 8(11): e81327.

Lobban CM, Simpson A, Pochon X, Pawlowski J, Foissner W (2002) Maristentor dinoferus n. gen., n. sp., a giant heterotrich ciliate (Spirotrichea: Heterotrichida) with zooxanthellae, from coral reefs on Guam, Mariana Islands. Mar Biol 140, 411–423. doi:10.1007/s002270100690

McLusky DS, Elliott M (2004) The Estuarine Ecosystem: Ecology, Threats and Management. OUP Oxford.

McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE 8, e61217. doi:10.1371/journal.pone.0061217

Medinger R, Nolte V, Pandey RV, Jost S and others (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol Ecol 19, 32–40. doi:10.1111/j.1365-294X.2009.04478.x

Morán XAG, Ducklow HW, Erickson M (2013) Carbon fluxes through estuarine bacteria reflect coupling with phytoplankton. Mar Ecol Prog Ser 489, 75–85. doi:10.3354/meps10428

Orive E, Laza-Martinez A, Seoane S, Alonso A, Andrade R, Miguel I (2010) Diversity of Pseudo-nitzschia in the southeastern Bay of Biscay. Diatom Research 25, 125–145.

R Core Team (2017). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Roman MR, Gauzens AL, Rhinehart WK, White JR (1993) Effects of low oxygen waters on Chesapeake Bay zooplankton. Limnol Oceanogr 38, 1603–1614. doi:10.4319/lo.1993.38.8.1603

Rose JM, Caron DA (2007) Does low temperature constrain the growth rates of heterotrophic protists? Evidence and implications for algal blooms in cold waters. Limnol Oceanogr 52, 886–895. doi:10.4319/lo.2007.52.2.0886

Round FE, Crawford RM, Mann DG (1990) The Diatoms - Biology & Morphology of the genera. Cambridge University Press.

Scoble JM, Cavalier-Smith T (2014) Scale evolution in Paraphysomonadida (Chrysophyceae): Sequence phylogeny and revised taxonomy of Paraphysomonas, new genus Clathromonas, and 25 new species. Eur J Protistol 50, 551–592. doi:10.1016/j.ejop.2014.08.001

Seoane S, Laza A, Urrutxurtu I, Orive E (2005) Phytoplankton Assemblages and Their Dominant Pigments in the Nervion River Estuary. Hydrobiologia 549, 1–13. doi:10.1007/s10750-005-1736-6

Stoeck T, Bass D, Nebel M, Christen R, Jones MD, Breiner HW Richards TA. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol, 19 Suppl 1:21-31.

Taylor AH, Allen JI, Clark PA (2002) Extraction of a weak climatic signal by an ecosystem. Nature 416, 629–632. doi:10.1038/416629a

Taylor FJR (1990) Phylum Dinoflagellata. In: Margulis L, Corliss JO, Melkonian M, Chapman DJ (eds), Handbook of Protoctista. Jones and Bartlett Publishers, Boston.

ter Braak CJF, Smilauer P (2002) CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination (v4.5). Microcomputer Power, Ithaca.

Uriarte I, Villate F (2004) Effects of pollution on zooplankton abundance and distribution in two estuaries of the Basque coast (Bay of Biscay). Mar Poll Bull 49: 220-228

Uriarte I, Villate F, Iriarte A, Duque J, Ameztoy I (2014) Seasonal and axial variations of net water circulation and turnover in the estuary of Bilbao. Estuar Coast Shelf Sci 150: 312-324.

Uriarte I, Villate F, Iriarte A (2016) Zooplankton recolonization of the inner estuary of Bilbao: influence of pollution abatement, climate and non-indigenous species. J Plankton Res doi: 10.1093/plankt/fbv060

Villate F (1994) Temporal variability of the spatial distribution of the zooplankton community in a coastal embayment of the Basque country in relation to physical phenomena. Hydrobiologia 288, 79–95.

Villate F, Moral M, Valencia V (1997) Mesozooplankton community indicates climate changes in a shelf area of the inner Bay of Biscay throughout 1988 to 1990. J Plankton Res 19, 1617–1636.

Villate F, Iriarte A, Uriarte I, Intxausti L, de la Sota A (2013) Dissolved oxygen in the rehabilitation phase of an estuary: Influence of sewage pollution abatement and hydro-climatic factors. Mar Pollut Bull 70, 234–246. doi:10.1016/j.marpolbul.2013.03.010

Ward BA, Dutkiewicz SA, Jahn O, Follows MJ (2012) A size-structured food-web model of the global ocean. Limnol Oceanogr 57: 1877-1891.

Wickham H (2009) ggplot2. Springer New York, New York, NY.

Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: A fast and accurate Illumina *Paired-End reAd mergeR. Bioinformatics 30: 614-620.*

Table S1_a. Relative abundances of the 29 taxonomic groups for the 200µm are shown in this table. Samples are arranged by salinity (30 and 35) and month. A category with sequences that had no database match is also included "no blast hit". Samples from January and February in 35 salinity were not collected due to bad weather.

							SALINITY 30						
#GROUP ID	September '13	October	November	December	January	February	March	April	May	June	July	August	September '14
AMOEBOZOA	0.00	0.00	4.04	4.21	0.09	1.38	1.03	0.09	0.03	0.06	0.01	0.66	0.45
CHLOROPLASTIDA	0.13	0.00	3.97	0.61	0.04	0.09	0.12	0.09	0.06	0.05	0.03	1.06	0.32
RHODOPHYCEAE	0.00	0.00	0.02	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
CRYPTOPHYCEAE	0.00	0.00	0.03	0.02	0.06	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00
EXCAVATA	0.00	0.02	2.23	1.98	0.05	0.10	0.09	0.03	0.02	0.06	0.03	0.10	0.08
HAPTOPHYTA	0.00	0.00	0.01	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ANNELIDA	0.02	0.16	0.57	3.03	0.36	0.46	0.22	0.36	0.06	0.31	0.16	0.17	0.09
BRANCHIOPODA	0.00	0.00	0.72	0.11	0.00	0.00	0.00	0.00	0.01	0.46	0.86	0.00	0.00
MALACOSTRACA	2.20	0.22	0.57	0.37	0.22	0.01	0.17	4.88	14.05	13.02	3.96	0.08	1.15
OTHER ARTHROPODA	0.01	0.01	2.34	0.33	0.07	0.05	0.04	0.03	0.04	0.03	0.00	0.00	0.01
CNIDARIA	0.01	0.01	0.07	0.04	0.02	0.14	0.07	0.04	0.04	0.09	0.12	0.02	0.02
MOLLUSCA	58.28	3.95	4.02	0.77	3.78	0.03	0.03	3.54	0.33	6.88	5.84	2.36	1.75
NEMATODA	0.00	0.06	0.07	1.98	0.02	0.11	0.18	0.17	0.40	1.32	0.01	0.07	0.04
CRANIATA	0.02	0.00	0.01	0.07	0.00	0.00	0.06	0.96	7.28	0.02	0.88	0.01	0.01
OTHER METAZOA	0.02	0.07	0.11	0.15	0.01	0.07	0.07	0.03	0.02	0.03	0.16	0.03	0.14
UROCHORDATA	0.01	0.02	16.06	0.69	0.14	0.06	0.07	0.15	0.29	0.25	0.07	0.03	0.11
FUNGI	0.02	0.00	0.84	1.40	0.02	0.21	0.13	0.03	0.02	0.00	0.04	0.07	0.03
OTHER ALVEOLATA	0.05	0.03	0.08	0.05	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.02	0.01
CILIOPHORA	0.02	0.03	4.32	2.67	0.16	0.08	0.04	0.01	0.03	0.09	0.04	0.02	0.02
DINOFLAGELLATA	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.02	0.00	0.01	0.01
RHIZARIA	0.06	0.05	0.42	0.71	0.29	0.11	0.07	0.03	1.03	0.27	0.01	0.01	0.03
OTHER STRAMENOPILES	0.01	0.00	0.36	4.45	0.03	0.12	0.11	0.02	0.01	0.02	0.02	0.03	0.03
CHRYSOPHYCEAE	0.00	0.00	19.68	19.45	0.53	8.11	13.53	1.47	0.27	1.11	0.82	3.39	2.11
DIATOMEA	0.05	0.01	0.42	1.84	0.04	1.15	0.94	0.47	0.06	0.17	0.09	0.48	0.32
OTHER OCHROPHYTA	0.04	0.00	0.02	0.01	0.00	0.01	0.00	0.03	0.00	0.01	0.00	0.10	0.00
OTHER EUKARYOTA	0.00	0.00	0.06	0.30	0.05	0.23	0.13	0.08	0.07	0.00	0.00	0.00	0.00
CIRRIPEDIA	26.07	13.36	0.07	0.05	0.74	0.03	0.88	12.08	18.30	19.05	37.87	44.05	47.30
OTHER MAXILOPODA	0.02	0.05	1.44	0.45	9.04	0.02	1.06	0.22	0.11	0.20	0.40	0.23	0.17
COPEPODA	12.83	81.84	37.21	53.84	84.24	87.36	80.91	75.15	57.44	56.45	48.43	46.97	45.76
NO BLAST HIT	0.14	0.08	0.22	0.27	0.00	0.03	0.02	0.02	0.00	0.01	0.15	0.04	0.04
							SALINITY 35						
#GROUP ID	September '13	October	November	December	January	February	March	April	May	June	July	August	September '14
AMOEBOZOA	0.00	0.00	0.12	2.90	-	-	0.79	0.01	0.33	0.00	0.01	0.03	0.03
CHLOROPLASTIDA	0.01	0.00	0.02	0.61	-	-	0.13	0.02	0.04	0.00	0.01	0.04	0.00
RHODOPHYCEAE	0.01	0.03	0.41	17.81	-	-	0.01	0.00	0.06	0.01	0.01	0.00	0.00
CRYPTOPHYCEAE	0.06	0.02	0.05	0.11	-	-	0.02	0.00	0.00	0.00	0.00	0.00	0.00
EXCAVATA	0.09	0.22	0.25	3.74	-	-	5.88	0.14	0.07	0.02	0.04	0.03	0.10
HAPTOPHYTA	0.00	0.00	0.00	0.00	-	-	0.01	0.00	0.00	0.00	0.00	0.00	0.00
ANNELIDA	0.02	0.09	0.46	1.32	-	-	0.12	0.12	0.50	0.28	0.39	0.05	0.02
BRANCHIOPODA	0.83	0.01	0.00	0.00	-	-	0.94	0.89	0.97	1.54	0.24	0.49	1.57
MALACOSTRACA	1.61	0.33	0.58	0.15	-	-	0.54	3.71	1.08	2.91	17.75	5.56	3.13
OTHER ARTHROPODA	0.08	0.06	0.02	0.14	-	-	0.10	0.00	0.15	0.39	4.75	0.03	0.20
CNIDARIA	0.43	0.91	1.25	8.44	-	-	0.01	0.11	0.71	0.36	0.41	0.02	1.67
MOLLUSCA	0.75	3.63	3.19	0.83	-	-	0.31	4.92	10.71	53.63	7.35	9.84	13.51
NEMATODA	0.01	0.00	0.01	0.08	-	-	0.95	0.01	0.01	0.00	0.04	0.03	0.02
CRANIATA	0.00	0.01	0.00	0.10	-	-	0.02	0.76	1.72	0.25	0.54	0.06	0.12
OTHER METAZOA	0.10	0.89	0.52	5.54	-	-	0.53	0.15	0.36	1.06	1.63	0.10	0.75
UROCHORDATA	0.24	0.07	1.21	0.07	-	-	0.06	0.07	0.05	0.35	2.88	1.14	5.55
FUNGI	0.00	0.00	0.01	0.34	-	-	0.57	0.03	0.01	0.00	0.01	0.04	0.03
OTHER ALVEOLATA	0.00	0.01	0.02	0.11	-	-	0.18	0.01	0.00	0.00	0.01	0.01	0.01
CILIOPHORA	0.26	1.13	1.01	1.05	-	-	1.04	0.06	0.03	0.01	0.02	0.03	0.03
DINOFLAGELLATA	0.05	0.39	0.04	0.10	-	-	0.91	0.00	0.01	0.01	0.20	0.00	0.10
RHIZARIA	0.08	0.08	0.24	0.37	-	-	0.25	0.01	0.01	0.01	0.05	0.03	0.03
OTHER STRAMENOPILES	0.00	0.00	0.06	0.25	-	-	5.84	0.01	0.02	0.01	0.02	0.02	0.01
CHRYSOPHYCEAE	0.00	0.00	1.58	32.33	-	-	13.83	3.20	0.92	0.09	1.10	1.11	6.57
DIATOMEA	0.02	0.06	0.12	0.32	-	-	1.31	0.26	0.05	0.02	0.07	0.10	0.26
OTHER OCHROPHYTA	0.00	0.00	0.04	0.25	-	-	0.01	0.00	0.02	0.00	0.00	0.01	0.00
OTHER EUKARYOTA	0.01	0.04	0.01	0.11	-	-	0.03	0.00	0.00	0.00	0.01	0.00	0.00
CIRRIPEDIA	2.23	1.66	0.19	0.50	-	-	0.44	33.78	19.27	18.97	37.28	37.70	2.22
OTHER MAXILOPODA	0.71	5.38	5.42	2.52	-	-	1.90	0.45	1.34	0.54	1.72	0.25	1.31
COPEPODA	91.12	84.88	83.10	18.36	-	-	62.88	51.06	61.49	19.51	23.02	42.86	62.63
NO BLAST HIT	1.30	0.11	0.04	1.57	-	-	0.36	0.18	0.07	0.00	0.45	0.42	0.11

Table S1_b. Relative abundances of the 29 taxonomic groups for the 20-200µm are shown in this table. Samples are arranged by salinity (30 and 35) and month. A category with sequences that had no database match is also included "no blast hit". Samples from January and February in 35 salinity were not collected due to bad weather.

							SALINITY 30						
#GROUP ID	September '13	October	November	December	January	February	March	April	May	June	July	August	September '14
AMOEBOZOA	0.00	0.00	3.57	1.53	1.39	6.08	3.04	0.32	1.43	0.65	0.41	0.23	0.26
CHLOROPLASTIDA	0.08	0.06	0.93	0.30	0.44	0.25	0.23	0.38	0.47	0.10	0.29	0.11	0.15
RHODOPHYCEAE	0.01	0.00	0.01	0.03	0.01	0.01	0.01	0.02	0.01	0.01	0.00	0.01	0.00
CRYPTOPHYCEAE	0.02	0.02	0.05	0.05	0.04	0.02	0.00	0.02	0.01	0.02	0.05	0.01	0.01
EXCAVATA	0.01	0.01	1.19	8.22	1.14	0.49	0.75	0.77	4.36	1.32	0.80	0.48	0.51
HAPTOPHYTA	0.00	0.04	0.14	0.52	0.05	0.06	0.59	0.85	0.09	0.05	0.48	0.03	0.01
ANNELIDA	7.77	5.81	1.32	4.99	14.36	0.98	0.56	3.19	5.79	24.17	0.96	2.41	5.40
BRANCHIOPODA	0.00	0.00	0.02	0.10	0.03	0.03	0.70	0.01	1.04	0.00	0.00	0.01	0.01
MALACOSTRACA	0.01	0.00	0.00	0.00	0.04	0.00	0.01	0.00	0.01	0.01	0.17	0.17	0.01
OTHER ARTHROPODA	0.01	0.02	0.25	0.15	0.03	0.03	0.01	1.94	0.07	0.19	0.01	0.00	0.01
CNIDARIA	0.13	0.01	4.46	0.05	0.86	0.30	0.17	0.61	0.07	1.26	1.89	0.03	0.05
MOLLUSCA	22.93	1.05	4.36	0.50	13.45	0.04	4.50	6.16	8.47	2.26	6.53	11.46	10.89
NEMATODA	0.08	0.01	0.37	1.38	0.17	0.15	0.13	0.13	0.74	0.15	0.02	0.03	0.08
CRANIATA	0.00	0.00	0.11	0.01	0.04	0.04	0.01	0.01	0.02	0.03	0.07	0.04	0.00
OTHER METAZOA	0.05	0.22	0.64	1.05	0.26	2.07	1.73	0.70	0.51	0.61	0.09	0.19	0.33
UROCHORDATA	0.12	1.42	25.51	0.22	1.47	0.04	0.36	1.71	7.80	0.07	0.17	0.48	2.95
FUNGI	0.02	0.02	0.99	1.13	0.36	0.59	0.65	0.29	0.58	1.52	0.51	0.06	0.19
OTHER ALVEOLATA	0.06	0.05	0.69	0.24	0.25	0.12	0.18	0.21	0.15	0.38	0.17	0.08	0.09
CILIOPHORA	0.05	0.62	9.37	3.08	27.30	4.01	1.54	2.78	0.97	0.65	6.11	0.26	0.90
DINOFLAGELLATA	0.38	0.06	0.26	1.18	0.22	0.09	0.15	0.74	0.64	0.26	1.54	0.07	0.31
RHIZARIA	0.05	0.06	11.27	2.92	1.44	0.36	1.22	0.50	0.35	16.52	0.18	0.05	0.10
OTHER STRAMENOPILES	0.06	0.03	0.60	8.39	0.31	0.14	0.28	0.22	3.79	0.17	0.12	0.06	0.33
CHRYSOPHYCEAE	0.03	0.02	9.87	20.85	24.13	13.43	16.52	3.89	12.37	15.76	14.34	5.75	3.45
DIATOMEA	0.37	0.26	1.15	2.18	1.28	1.09	1.62	11.29	1.94	0.72	1.49	0.60	0.66
OTHER OCHROPHYTA	0.00	0.00	0.03	0.03	0.04	0.00	0.01	0.04	0.01	0.00	0.01	0.00	0.01
OTHER EUKARYOTA	0.00	0.01	0.14	0.13	0.15	0.21	0.22	0.20	0.14	0.04	0.03	0.01	0.01
CIRRIPEDIA	0.15	0.00	0.02	0.03	0.00	0.02	0.02	1.54	3.49	0.74	0.39	6.44	4.62
OTHER MAXILOPODA	0.05	0.01	0.01	0.00	0.00	0.28	1.83	0.22	0.19	0.13	0.00	0.26	0.02
COPEPODA	67.48	90.14	22.55	40.09	10.48	68.82	62.82	61.19	44.20	32.03	63.07	70.40	68.60
NO BLAST HIT	0.06	0.04	0.13	0.65	0.27	0.25	0.15	0.07	0.29	0.18	0.10	0.27	0.05
							SALINITY 35						
#GROUP ID	September '13	October	November	December	January	February	March	April	May	June	July	August	September '14
AMOEBOZOA	0.01	0.06	0.40	3.78	-	-	0.03	0.11	0.06	0.13	0.15	0.92	0.10
CHLOROPLASTIDA	0.06	0.07	0.13	0.59	-	-	0.05	0.55	0.14	0.23	0.14	0.37	0.12
RHODOPHYCEAE	0.02	0.10	0.00	2.03	-	-	0.00	0.01	0.01	0.01	0.00	0.02	0.03
CRYPTOPHYCEAE	0.02	0.07	0.00	0.48	-	-	0.01	0.01	0.00	0.03	0.00	0.03	0.02
EXCAVATA	0.01	0.01	2.97	2.93	-	-	0.04	0.17	0.09	0.43	0.20	7.67	0.34
HAPTOPHYTA	0.02	0.01	0.00	0.02	-	-	0.00	0.14	0.08	0.04	0.01	0.03	0.18
ANNELIDA	0.03	5.03	0.06	35.99	-	-	0.31	1.76	1.60	0.50	20.45	2.01	1.74
BRANCHIOPODA	0.11	2.41	0.02	0.03	-	-	1.28	0.00	0.00	1.87	0.04	0.07	0.00
MALACOSTRACA	0.01	0.01	0.05	0.22	-	-	0.01	0.01	0.00	0.01	0.00	0.28	0.00
OTHER ARTHROPODA	0.00	0.01	0.00	0.08	-	-	0.00	0.10	0.00	0.00	0.00	1.25	0.00
CNIDARIA	0.10	0.04	0.01	0.22	-	-	0.08	18.60	0.07	10.81	0.61	1.29	12.08
MOLLUSCA	0.62	3.52	0.13	1.57	-	-	0.05	2.39	1.98	9.21	7.33	2.62	9.13
NEMATODA	0.07	0.12	0.07	0.17	-	-	0.04	0.01	0.01	0.01	0.02	8.62	0.20
CRANIATA	0.00	0.07	0.65	0.16	-	-	0.00	0.88	0.05	1.84	0.01	0.04	0.01
OTHER METAZOA	0.53	0.16	0.05	3.74	-	-	0.11	0.88	0.29	2.59	0.15	3.28	4.02
UROCHORDATA	0.20	0.70	0.25	0.09	-	-	4.28	1.78	0.62	0.96	5.09	18.24	12.12
FUNGI	0.02	0.07	0.37	0.65	-	-	0.01	0.04	0.39	0.05	0.03	1.24	0.07
OTHER ALVEOLATA	0.21	0.16	0.29	0.68	-	-	0.06	0.51	0.93	4.19	0.05	1.02	1.59
CILIOPHORA	0.94	1.73	0.17	4.33	-	-	2.08	4.39	0.86	2.62	0.45	0.65	2.09
DINOFLAGELLATA	3.17	0.65	0.02	0.74	-	-	3.31	3.24	9.93	3.01	0.42	2.24	6.20
RHIZARIA	0.65	0.14	0.05	3.66	-	-	0.01	0.59	0.28	1.10	0.10	1.42	0.16
OTHER STRAMENOPILES	0.34	0.08	0.03	1.39	-	-	0.02	0.06	0.70	0.22	0.09	1.70	0.17
CHRYSOPHYCEAE	0.01	0.12	91.29	19.74	-	-	0.33	1.28	0.85	4.45	1.00	16.98	4.89
DIATOMEA	1.15	11.52	0.02	9.19	-	-	0.05	20.89	2.61	1.10	1.06	7.08	1.00
OTHER OCHROPHYTA	0.02	0.06	0.00	1.02	-	-	0.03	0.22	0.05	0.11	0.00	0.16	0.03
OTHER EUKARYOTA	0.01	0.04	0.00	0.25	-	-	0.01	0.28	0.02	0.03	0.02	0.15	0.04
CIRRIPEDIA	2.13	0.03	0.05	0.17	-	-	0.14	1.37	4.25	1.80	4.99	0.53	0.29
OTHER MAXILOPODA	2.85	11.51	0.24	0.58	-	-	7.12	2.41	10.42	7.73	0.03	1.85	9.04
COPEPODA	86.64	58.46	1.83	4.94	-	-	80.55	37.31	63.66	44.87	57.52	17.80	33.46
NO BLAST HIT	0.07	3.05	0.85	0.55	-	-	0.00	0.02	0.05	0.04	0.01	0.45	0.87

Table S1_c. Relative abundances of the 29 taxonomic groups for the 0.22-20µm are shown in this table. Samples are arranged by salinity (30 and 35) and month. A category with sequences that had no database match is also included "no blast hit". Samples from January and February in 35 salinity were not collected due to bad weather.

							SALINITY 30						
#GROUP ID	September '13	October	November	December	January	February	March	April	May	June	July	August	September '14
AMOEBOZOA	1.16	1.52	12.81	14.16	9.61	12.80	12.53	7.60	22.15	8.83	4.28	13.07	8.21
CHLOROPLASTIDA	10.49	7.43	0.17	0.51	1.35	2.08	6.14	4.62	1.71	1.13	4.64	3.16	10.86
RHODOPHYCEAE	0.01	0.01	0.00	0.03	0.00	0.00	0.03	0.01	0.01	0.00	0.01	0.00	0.01
CRYPTOPHYCEAE	10.73	5.50	1.06	0.56	0.06	0.13	0.21	1.29	0.13	7.66	7.63	4.42	3.35
EXCAVATA	0.63	0.44	3.17	4.31	7.91	7.35	3.01	4.75	3.27	2.96	4.46	8.82	4.48
HAPTOPHYTA	1.32	0.71	0.18	0.10	0.03	0.03	0.12	0.17	0.06	0.01	0.02	0.10	0.09
ANNELIDA	0.12	0.16	0.05	0.08	0.10	0.04	0.17	0.39	0.71	0.65	0.38	0.12	0.30
BRANCHIOPODA	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
MALACOSTRACA	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.03	0.01	0.00	0.00
OTHER ARTHROPODA	0.12	0.06	0.06	0.01	0.00	0.01	0.01	0.03	0.01	0.01	0.08	0.08	0.09
CNIDARIA	0.10	0.13	0.13	0.05	0.00	0.00	0.06	0.47	0.09	0.34	0.21	0.19	0.30
MOLLUSCA	0.46	0.66	0.91	0.49	0.05	0.01	2.25	0.61	0.20	0.26	0.57	0.23	0.12
NEMATODA	0.02	0.03	0.04	0.65	0.04	0.05	0.06	0.03	0.06	0.07	0.02	0.07	0.02
CRANIATA	0.00	0.02	0.04	0.01	0.00	0.02	0.02	0.02	0.03	0.00	0.01	0.05	0.00
OTHER METAZOA	0.04	0.06	0.05	0.49	0.01	0.04	0.68	0.11	0.45	0.12	0.06	0.08	0.14
UROCHORDATA	0.04	0.17	0.28	0.11	0.05	0.00	0.19	0.47	1.34	0.07	0.04	0.12	0.61
FUNGI	0.25	0.57	0.51	3.03	0.44	1.42	4.99	1.29	1.29	1.04	3.33	0.79	1.14
OTHER ALVEOLATA	3.73	5.29	8.04	1.01	0.44	1.02	1.78	4.02	1.73	4.45	3.21	3.73	7.76
CILIOPHORA	1.01	1.05	8.77	10.30	13.83	7.01	13.57	8.49	6.58	15.33	11.99	4.43	1.98
DINOFLAGELLATA	32.69	11.50	2.78	1.44	0.24	0.12	4.71	12.48	16.04	13.32	14.36	15.24	14.47
RHIZARIA	6.15	3.76	15.43	16.06	0.26	1.06	3.77	2.91	2.50	1.10	1.02	1.39	1.95
OTHER STRAMENOPILES	4.15	4.68	4.16	2.13	2.09	3.24	3.80	2.67	2.35	1.58	1.38	2.42	1.64
CHRYSOPHYCEAE	4.79	17.24	25.55	34.08	54.24	54.77	32.43	31.90	29.56	35.10	13.50	28.52	18.07
DIATOMEA	15.10	25.14	0.68	2.23	0.29	0.16	2.83	5.96	6.78	1.99	26.41	6.05	17.43
OTHER OCHROPHYTA	1.65	3.06	12.11	0.57	0.77	0.14	0.71	0.85	0.32	0.14	0.63	0.68	0.86
OTHER EUKARYOTA	4.41	2.63	2.17	3.97	0.08	0.92	1.04	2.68	1.24	1.03	0.82	3.23	3.92
CIRRIPEDIA	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.04	0.02	0.01	0.04	0.27	0.02
OTHER MAXILOPODA	0.00	0.04	0.01	0.01	0.00	0.00	0.25	0.01	0.02	0.02	0.00	0.00	0.01
COPEPODA	0.10	7.53	0.04	1.36	0.08	0.16	2.46	4.21	0.25	2.02	0.38	2.00	1.20
NO BLAST HIT	0.73	0.61	0.79	2.25	8.02	7.35	2.16	1.88	1.10	0.71	0.50	0.77	0.96
							SALINITY 35						
#GROUP ID	September '13	October	November	December	January	February	March	April	May	June	July	August	September '14
AMOEBOZOA	0.15	1.12	9.12	12.40	-	-	1.42	0.97	8.23	3.59	3.52	3.10	5.77
CHLOROPLASTIDA	21.41	8.52	1.14	3.94	-	-	9.72	6.37	1.53	6.72	2.03	4.49	4.92
RHODOPHYCEAE	0.04	0.53	0.01	0.17	-	-	0.07	0.00	0.02	0.01	0.00	0.02	0.01
CRYPTOPHYCEAE	1.26	2.58	0.50	0.94	-	-	0.75	1.16	0.17	0.90	1.52	1.23	5.24
EXCAVATA	0.34	0.47	7.77	5.73	-	-	3.78	2.79	3.21	7.33	4.29	5.86	6.53
HAPTOPHYTA	0.97	1.09	0.09	0.29	-	-	0.98	0.15	0.14	0.09	0.48	0.41	0.43
ANNELIDA	0.16	0.66	0.26	12.50	-	-	0.37	0.60	0.52	0.06	0.05	0.09	0.09
BRANCHIOPODA	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MALACOSTRACA	0.00	0.01	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OTHER ARTHROPODA	0.02	0.01	0.36	0.05	-	-	0.00	0.03	0.03	0.01	0.04	0.02	0.07
CNIDARIA	0.37	0.84	0.39	0.16	-	-	0.03	0.84	0.08	1.34	0.07	0.71	0.68
MOLLUSCA	0.02	0.68	0.08	0.37	-	-	4.27	0.53	0.29	0.06	0.15	0.03	0.08
NEMATODA	0.07	0.29	0.30	0.06	-	-	0.10	0.04	0.06	0.02	0.04	0.05	0.33
CRANIATA	0.00	0.02	0.01	0.00	-	-	0.02	0.00	0.01	0.00	0.01	0.00	0.00
OTHER METAZOA	0.05	0.66	0.23	0.21	-	-	0.30	0.06	0.35	0.13	0.10	0.42	0.22
UROCHORDATA	0.62	0.28	0.53	0.01	-	-	7.38	5.39	0.95	0.09	0.05	11.48	0.22
FUNGI	0.09	0.20	0.38	0.69	-	-	0.52	0.12	0.78	0.24	0.37	0.22	0.38
OTHER ALVEOLATA	6.05	12.56	11.16	3.20	-	-	0.89	1.85	2.81	5.45	7.80	8.27	9.65
CILIOPHORA	2.42	3.46	24.68	8.07	-	-	18.55	22.78	8.74	12.22	11.15	14.99	13.18
DINOFLAGELLATA	15.18	6.88	3.33	2.74	-	-	0.61	7.53	26.42	15.77	12.84	15.10	14.63
RHIZARIA	4.75	4.07	1.79	2.48	-	-	1.47	1.50	2.85	2.08	2.94	2.74	2.91
OTHER STRAMENOPILES	2.27	6.39	3.19	3.28	-	-	3.82	1.64	3.41	1.28	5.04	2.07	1.92
CHRYSOPHYCEAE	0.57	2.68	26.10	35.01	-	-	30.98	21.73	26.64	24.23	18.65	18.15	20.40
DIATOMEA	27.84	33.20	1.02	3.70	-	-	8.38	11.09	9.59	4.83	26.69	7.68	8.07
OTHER OCHROPHYTA	2.41	1.62	0.49	0.49	-	-	2.76	0.34	0.39	8.89	0.71	0.66	1.01
OTHER EUKARYOTA	0.39	3.65	1.28	2.88	-	-	1.50	2.06	1.86	0.65	1.05	1.41	2.48
CIRRIPEDIA	0.00	0.04	0.01	0.01	-	-	0.01	0.13	0.02	0.78	0.00	0.01	0.01
OTHER MAXILOPODA	0.00	0.04	0.03	0.01	-	-	0.08	0.11	0.02	0.03	0.00	0.01	0.01
COPEPODA	0.47	0.56	0.08	0.06	-	-	0.41	8.56	0.17	2.14	0.15	0.29	0.16
NO BLAST HIT	12.09	7.10	5.65	0.56	-	-	0.84	1.63	0.73	1.08	0.24	0.48	0.60

Table S2_a. List of most abundant taxa (> 5 % of abundance for any size-fraction in at least one sample) for the 200µm. Acronyms used in the CCA are shown.

								SALINITY 30					
#OTU ID	ACRONYM	September '13	October	November	December	January	February	March	April	May	June	July	August	September '14
Hyperamoeba flagellata	HFLA	0.00	0.00	3.99	3.92	0.08	1.36	1.02	0.09	0.03	0.06	0.01	0.66	0.45
Ostreococcus	OSTR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pyramimonas	PYRA	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
Cryptomonadales	CRYP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Carpediemonas membranifera	CMEM	0.00	0.00	1.19	0.97	0.01	0.06	0.04	0.02	0.01	0.04	0.01	0.06	0.04
Boccardiella ligerica	BLIG	0.01	0.06	0.04	1.49	0.00	0.07	0.01	0.00	0.02	0.24	0.11	0.08	0.00
Polydora ciliata	PCIL	0.00	0.01	0.11	1.31	0.23	0.02	0.03	0.02	0.00	0.03	0.03	0.03	0.00
Sabellaria alveolata	SALV	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Ambidexter symmetricus	ASYM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Neomysis integer	NINT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.66	13.26	12.90	0.01	0.00	0.01
Armatobalanus allium	AALL	0.34	0.21	0.00	0.00	0.50	0.00	0.09	1.41	2.26	2.77	4.96	5.60	6.87
Cyclopina gracilis	CGRA	2.00	7.64	0.37	4.27	0.78	0.71	0.09	0.70	1.70	0.34	1.08	3.15	2.66
Euchaeta indica	AIND	0.00	0.00	0.01	0.01	0.32	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Oithona sp. 2 New Caledonia-RJ	OITH	0.00	0.21	0.03	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.02
Paracalanus parvus	PPAR	0.29	1.06	3.14	0.08	3.78	0.03	0.34	0.49	0.46	0.41	3.78	0.02	0.51
Pseudocalanus elongatus	PELO	0.01	0.01	0.00	0.00	4.59	0.02	0.10	0.01	0.02	0.00	0.00	0.00	0.00
Scolecithricella longispinosa	SLON	0.03	0.02	0.05	1.95	3.86	43.90	33.77	11.70	14.37	0.03	0.01	0.00	0.10
Stomatolepas praegustator	SPRA	25.71	13.12	0.02	0.01	0.22	0.02	0.80	10.67	16.03	16.11	32.87	38.45	40.36
Temora turbinata	TTUR	0.00	0.57	0.00	0.00	0.00	0.00	0.20	0.11	0.00	0.12	0.00	0.00	0.50
uncultured eukaryote	MAXI	0.01	0.04	1.43	0.24	9.04	0.02	1.05	0.15	0.02	0.15	0.33	0.01	0.07
Chelophyes appendiculata	CAPP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.10	0.00	0.01
Garveia sp. CC-2005	GARV	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00
Xenopus (Silurana) tropicalis (we	XTRO	0.02	0.00	0.00	0.00	0.00	0.00	0.06	0.96	7.26	0.01	0.06	0.01	0.01
Adula aff. californiensis MO-201	MADU	33.33	0.36	1.37	0.08	0.00	0.00	0.00	0.07	0.20	0.50	0.10	1.21	0.56
Mytilus edulis	MEDU	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.42	0.02	0.03	0.24
Crassostrea gigas (Pacific oyste	CGIG	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.84	0.11	0.00
Batillaria zonalis	BZON	0.19	0.32	0.18	0.00	0.54	0.00	0.00	0.00	0.00	1.58	1.75	0.23	0.19
Euplica scripta	ESCR	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.38	0.09	0.00	0.00
Peringia ulvae	PULV	23.88	3.19	0.56	0.64	3.00	0.02	0.03	3.33	0.09	0.98	0.22	0.32	0.13
Haminoea hydatis	HHYD	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.18	0.07	0.00	0.00
Daptonema setosum	DSET	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Appendicularia	APPE	0.01	0.02	16.03	0.69	0.14	0.06	0.07	0.15	0.29	0.03	0.07	0.03	0.03
Doliolum nationalis	DNAT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cryptocaryon	CILI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Monodinium sp. HCB-2005	MONO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Oligotrichia	OLIG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maristentor dinoferus	MDIN	0.00	0.00	0.03	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.01
Dinophyceae	DINO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Gyrodinium	GYRO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Protoperidinium	PROT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Syndiniales	SYND	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cercozoa	CERC	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Chlorarachniophyta	CHLO	0.05	0.00	0.01	0.05	0.09	0.00	0.00	0.00	1.00	0.25	0.00	0.00	0.00
Imbricatea	IMBR	0.00	0.00	0.10	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Thecofilosea	THEC	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Stramenopiles	STRA	0.00	0.00	0.04	0.03	0.01	0.02	0.02	0.01	0.01	0.00	0.01	0.02	0.02
Chrysophyceae	CHRY	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Paraphysomonas	PARA	0.00	0.00	19.67	19.43	0.53	8.11	13.53	1.47	0.27	1.11	0.82	3.39	2.11
Pseudo-nitzschia australis	PNIT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Chaetoceros	CHAE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Lauderia annulata	LANN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Papiliocellulus elegans	PELE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01
Skeletonema	SKEL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
Thalassiosira allenii	TALL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tenuicylindrus belgicus	TBEL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Diatomea	DIAT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pedinellales	PEDI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Heterosigma akashiwo	HAKA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Acartia_clausi	ACLA	0.18	0.35	3.38	0.32	56.22	0.74	1.90	7.30	0.05	0.59	1.01	0.03	0.46
Acartia_tonsa	ATON	5.35	69.17	6.16	20.52	12.55	5.48	2.75	12.18	2.14	6.29	40.43	39.48	39.16
Calanipeda_aquaedulcis	CAQU	2.46	1.03	21.78	18.76	1.10	36.14	40.41	42.38	38.28	43.46	0.28	0.18	0.03
Calanus_helgolandicus	CHEL	0.04	0.01	0.00	0.01	0.00	0.00	1.13	0.00	0.00	0.00	1.09	0.02	0.00
Euterpina_acutifrons	EACU	0.01	0.27	0.00	0.18	0.11	0.00	0.00	0.00	0.00	0.00	0.07	0.01	0.04
Pseudodiaptomus_marinus	PMAR	2.27	1.04	0.16	5.64	0.00	0.02	0.00	0.10	0.23	2.83	0.37	3.93	2.27
								SALINITY 35						
#OTU ID	ACRONYM	September '13	October	November	December	January	February	March	April	May	June	July	August	September '14
Hyperamoeba flagellata	HFLA	0.00	0.00	0.12	2.83	-	-	0.77	0.01	0.33	0.00	0.01	0.03	0.03
Ostreococcus	OSTR	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pyramimonas	PYRA	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cryptomonadales	CRYP	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Carpediemonas membranifera	CMEM	0.00	0.00	0.13	1.97	-	-	4.86	0.03	0.01	0.01	0.02	0.02	0.04
Boccardiella ligerica	BLIG	0.00	0.00	0.01	0.40	-	-	0.00	0.07	0.00	0.00	0.04	0.03	0.00
Polydora ciliata	PCIL	0.00	0.01	0.00	0.02	-	-	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Sabellaria alveolata	SALV	0.00	0.00	0.00	0.14	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ambidexter symmetricus	ASYM	1.09	0.00	0.00	0.00	-	-	0.06	0.01	0.23	0.67	5.29	0.00	0.45
Neomysis integer	NINT	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Armatobalanus allium	AALL	0.20	0.06	0.06	0.30	-	-	0.36	6.45	5.61	2.39	5.34	5.28	0.54
Cyclopina gracilis	CGRA	0.01	0.31	0.01	0.46	-	-	0.01	0.96	0.06	0.02	0.13	2.01	0.05
Euchaeta indica	AIND	0.00	0.53	5.98	0.76	-	-	0.00	0.00	0.02	0.00	0.01	0.00	0.00
Oithona sp. 2 New Caledonia-RJ	OITH	0.34	0.06	1.49	0.02	-	-	0.00	0.03	0.00	0.00	0.03	0.12	0.51
Paracalanus parvus	PPAR	4.86	47.14	4.89	0.15	-	-	1.97	3.20	10.12	1.11	6.25	1.90	2.60
Pseudocalanus elongatus	PELO	0.03	7.21	1.86	5.06	-	-	0.11	0.00	0.11	0.00	0.00	0.00	0.00
Scolecithricella longispinosa	SLON	0.00	0.00	0.01	0.01	-	-	0.01	0.00	0.20	0.02	0.01	0.00	0.07
Stomatolepas praegustator	SPRA	1.92	1.58	0.00	0.12	-	-	0.08	27.27	13.01	11.98	31.78	31.83	1.24
Temora turbinata	TTUR	0.13	3.20	0.65	0.00	-	-	0.80	0.00	4.05	6.35	0.01	0.47	1.60
uncultured eukaryote	MAXI	0.71	5.31	5.28	2.52	-	-	1.89	0.37	1.30	0.51	1.65	0.18	1.30
Chelophyes appendiculata	CAPP	0.26	0.69	1.17	0.14	-	-	0.00	0.10	0.05	0.13	0.37	0.02	1.55
Garveia sp. CC-2005	GARV	0.01	0.01	0.02	0.11	-	-	0.00	0.00	0.12	0.00	0.00	0.00	0.00
Xenopus (Silurana) tropicalis (we	XTRO	0.00	0.00	0.00	0.00	-	-	0.01	0.05	1.19	0.13	0.06	0.02	0.09
Adula aff. californiensis MO-201	MADU	0.02	0.00	0.00	0.00	-	-	0.00	0.08	0.00	0.01	0.00	0.05	0.00
Mytilus edulis	MEDU	0.00	1.08	0.26	0.20	-	-	0.13	0.02	1.92	0.98	0.02	0.16	1.47
Crassostrea gigas (Pacific oyste	CGIG	0.00	0.00	0.00	0.00	-	-	0.00	0.66	0.00	0.00	0.73	5.44	0.14
Batillaria zonalis	BZON	0.31	1.51	0.85	0.11	-	-	0.02	1.58	4.55	7.19	0.81	1.14	1.27
Euplica scripta	ESCR	0.06	0.03	0.00	0.00	-	-	0.02	0.05	1.30	5.26	0.60	0.03	1.65
Peringia ulvae	PULV	0.03	0.60	0.90	0.00	-	-	0.06	0.19	0.22	0.11	0.01	0.05	0.32
Haminoea hydatis	HHYD	0.05	0.00	0.01	0.00	-	-	0.00	0.05	0.93	28.82	1.81	0.31	0.81
Daptonema setosum	DSET	0.00	0.00	0.00	0.00	-	-	0.30	0.00	0.00	0.00	0.00	0.00	0.00
Appendicularia	APPE	0.24	0.06	1.11	0.04	-	-	0.06	0.07	0.04	0.34	2.87	1.14	4.91
Doliolum nationalis	DNAT	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.64
Cryptocaryon	CILI	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Monodinium sp. HCB-2005	MONO	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Oligotrichia	OLIG	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maristentor dinoferus	MDIN	0.00	0.00	0.01	0.04	-	-	0.89	0.00	0.00	0.00	0.00	0.00	0.01
Dinophyceae	DINO	0.00	0.00	0.00	0.00	-	-	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Gyrodinium	GYRO	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Protoperidinium	PROT	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Syndiniales	SYND	0.00	0.00	0.00	0.01	-	-	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Cercozoa	CERC	0.00	0.00	0.00	0.00	-	-	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Chlorarachniophyta	CHLO	0.00	0.00	0.07	0.03	-	-	0.01	0.00	0.00	0.00	0.00	0.01	0.02
Imbricatea	IMBR	0.00	0.00	0.00	0.00	-	-	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Thecofilosea	THEC	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Stramenopiles	STRA	0.00	0.00	0.02	0.12	-	-	5.80	0.01	0.02	0.01	0.01	0.02	0.01
Chrysophyceae	CHRY	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Paraphysomonas	PARA	0.00	0.00	1.58	32.32	-	-	13.82	3.19	0.92	0.09	1.10	1.10	6.57
Pseudo-nitzschia australis	PNIT	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Chaetoceros	CHAE	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lauderia annulata	LANN	0.00	0.00	0.00	0.02	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Papiliocellulus elegans	PELE	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Skeletonema	SKEL	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Thalassiosira allenii	TALL	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tenuicylindrus belgicus	TBEL	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Diatomea	DIAT	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pedinellales	PEDI	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Heterosigma akashiwo	HAKA	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Acartia_clausi	ACLA	71.27	15.67	59.83	6.37	-	-	57.23	1.28	41.97	7.20	9.14	0.77	48.29
Acartia_tonsa	ATON	0.06	0.03	0.42	0.22	-	-	0.00	43.81	0.07	0.04	0.44	33.70	0.17
Calanipeda_aquaedulcis	CAQU	0.01	0.01	0.16	0.01	-	-	0.01	0.21	0.20	0.06	0.04	0.00	0.02
Calanus_helgolandicus	CHEL	7.35	0.98	1.88	4.57	-	-	2.26	0.91	2.15	1.36	4.95	0.48	0.78
Euterpina_acutifrons	EACU	4.04	4.57	0.38	0.02	-	-	0.00	0.07	0.02	0.02	1.72	2.63	5.74
Pseudodiaptomus_marinus	PMAR	0.00	0.41	0.00	0.02	-	-	0.00	0.39	0.00	0.03	0.00	0.46	0.00

Table S2_b. List of most abundant taxa (> 5 % of abundance for any size-fraction in at least one sample) for the 20-200µm. Acronyms used in the CCA are shown.

								SALINITY 30						
#OTU ID	ACRONYM	September '13	October	November	December	January	February	March	April	May	June	July	August	September '14
Hyperamoeba flagellata	HFLA	0.00	0.00	3.47	1.44	1.26	5.99	2.96	0.27	1.40	0.64	0.39	0.22	0.26
Ostreococcus	OSTR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
Pyramimonas	PYRA	0.01	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.03	0.08
Cryptomonadales	CRYP	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.02	0.00	0.00
Carpediemonas membranifera	CMEM	0.00	0.00	0.79	6.81	0.57	0.11	0.24	0.42	3.59	0.89	0.41	0.22	0.36
Boccardiella ligerica	BLIG	0.37	0.68	0.00	0.54	0.54	0.14	0.00	0.00	4.00	13.56	0.11	1.46	0.77
Polydora ciliata	PCIL	7.40	4.83	0.14	4.37	7.66	0.67	0.22	2.56	1.70	10.15	0.35	0.69	4.02
Sabellaria alveolata	SALV	0.00	0.00	0.00	0.00	0.81	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ambidexter symmetricus	ASYM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.00
Neomysis integer	NINT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Armatobalanus allium	AALL	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.20	0.52	0.12	0.08	1.94	1.29
Cyclopina gracilis	CGRA	39.06	80.14	13.38	27.79	7.52	3.82	0.94	21.15	18.83	22.28	50.59	56.90	51.20
Euchaeta indica	AIND	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	0.00	0.00
Oithona sp. 2 New Caledonia-RJH-2004	OITH	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Paracalanus parvus	PPAR	0.00	0.00	0.00	0.01	0.02	0.03	0.14	0.99	0.05	0.02	0.09	0.14	0.44
Pseudocalanus elongatus	PELO	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Scolecithricella longispinosa	SLON	0.00	0.00	0.01	0.15	0.00	6.45	7.87	3.43	5.43	0.02	0.00	0.01	0.00
Stomatolepas praegustator	SPRA	0.15	0.00	0.01	0.02	0.00	0.01	0.01	0.99	2.46	0.62	0.32	4.48	3.32
Temora turbinata	TTUR	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
uncultured eukaryote	MAXI	0.00	0.01	0.01	0.00	0.00	0.01	1.83	0.20	0.00	0.02	0.00	0.24	0.01
Chelophyes appendiculata	CAPP	0.00	0.00	2.36	0.00	0.07	0.00	0.00	0.00	0.00	0.03	1.65	0.02	0.03
Garveia sp. CC-2005	GARV	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.22	0.01	0.00	0.00	0.00	0.00
Xenopus (Silurana) tropicalis (western clawed frog)	XTRO	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.00	0.00
Adula aff. californiensis MO-2010	MADU	18.01	0.80	3.13	0.35	0.18	0.02	3.63	4.05	6.70	1.83	5.22	8.25	7.74
Mytilus edulis	MEDU	0.00	0.00	0.00	0.00	12.62	0.00	0.00	0.01	0.00	0.00	0.00	0.71	0.00
Crassostrea gigas (Pacific oyster)	CGIG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.04	0.01
Batillaria zonalis	BZON	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.27	0.00	0.00	0.03	0.39
Euplica scripta	ESCR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.01	0.00
Peringia ulvae	PULV	4.19	0.00	0.00	0.07	0.00	0.00	0.00	1.19	0.00	0.00	0.00	0.01	0.01
Haminoea hydatis	HHYD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.07	0.01	0.00
Daptonema setosum	DSET	0.00	0.00	0.00	0.34	0.00	0.00	0.00	0.02	0.20	0.00	0.00	0.00	0.02
Appendicularia	APPE	0.12	1.42	25.48	0.13	1.47	0.01	0.34	1.71	7.78	0.05	0.17	0.45	2.92
Doliolum nationalis	DNAT	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.04	0.02
Cryptocaryon	CILI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Monodinium sp. HCB-2005	MONO	0.00	0.00	0.00	0.00	14.05	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00
Oligotrichia	OLIG	0.00	0.00	0.22	0.00	6.31	0.75	0.13	0.07	0.01	0.02	0.01	0.00	0.00
Maristentor dinoferus	MDIN	0.00	0.00	0.36	1.28	0.19	0.02	0.03	0.08	0.28	0.00	0.13	0.02	0.03
Dinophyceae	DINO	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
Gyrodinium	GYRO	0.01	0.00	0.03	0.00	0.04	0.00	0.01	0.27	0.01	0.09	0.09	0.00	0.03
Protoperidinium	PROT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Syndiniales	SYND	0.00	0.00	0.40	0.04	0.02	0.05	0.14	0.05	0.01	0.00	0.01	0.00	0.01
Cercozoa	CERC	0.00	0.00	1.54	0.09	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Chlorarachniophyta	CHLO	0.02	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.01	16.36	0.02	0.01	0.02
Imbricatea	IMBR	0.00	0.00	0.66	0.08	0.17	0.09	0.01	0.01	0.03	0.00	0.01	0.00	0.00
Thecofilosea	THEC	0.00	0.00	8.08	0.01	0.03	0.00	0.06	0.00	0.00	0.01	0.00	0.00	0.00
Stramenopiles	STRA	0.00	0.00	0.11	8.04	0.12	0.02	0.02	0.03	3.69	0.03	0.04	0.04	0.30
Chrysophyceae	CHRY	0.00	0.00	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
Paraphysomonas	PARA	0.02	0.01	9.75	20.83	24.10	13.42	16.50	3.81	12.35	15.60	14.25	5.75	3.45
Pseudo-nitzschia australis	PNIT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.23	0.00	0.00	0.00	0.00	0.00
Chaetoceros	CHAE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.66	0.00	0.00	0.01	0.00	0.00
Lauderia annulata	LANN	0.00	0.00	0.00	0.01	0.00	0.00	0.01	1.05	0.00	0.00	0.00	0.00	0.00
Papiliocellulus elegans	PELE	0.01	0.00	0.01	0.00	0.00	0.01	0.00	0.21	0.33	0.03	0.53	0.00	0.03
Skeletonema	SKEL	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.40	0.01	0.00	0.00	0.00	0.12
Thalassiosira allenii	TALL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tenuicylindrus belgicus	TBEL	0.00	0.06	0.00	0.00	0.10	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Diatomea	DIAT	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pedinellales	PEDI	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Heterosigma akashiwo	HAKA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Acartia_clausi	ACLA	0.01	0.00	0.01	0.04	0.18	0.38	12.05	1.04	2.59	0.08	0.33	0.11	0.97
Acartia_tonsa	ATON	6.63	1.58	0.12	3.25	2.15	1.52	2.32	3.36	0.61	0.61	4.99	7.25	5.46
Calanipeda_aquaedulcis	CAQU	2.36	0.00	4.23	7.02	0.01	56.47	36.11	29.19	11.91	6.61	1.23	0.03	0.00
Calanus_helgolandicus	CHEL	0.05	0.03	0.00	0.00	0.00	0.00	3.01	0.00	0.00	0.00	0.01	0.00	0.00
Euterpina_acutifrons	EACU	0.25	4.68	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.35
Pseudodiaptomus_marinus	PMAR	18.21	2.90	3.05	0.70	0.55	0.00	0.00	0.63	2.20	0.01	5.58	5.70	9.40
								SALINITY 35						
#OTU ID	ACRONYM	September '13	October	November	December	January	February	March	April	May	June	July	August	September '14
Hyperamoeba flagellata	HFLA	0.00	0.00	0.40	3.58	-	-	0.03	0.10	0.05	0.11	0.14	0.85	0.08
Ostreococcus	OSTR	0.01	0.00	0.00	0.01	-	-	0.00	0.03	0.00	0.03	0.00	0.00	0.02
Pyramimonas	PYRA	0.00	0.03	0.00	0.04	-	-	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Cryptomonadales	CRYP	0.00	0.00	0.00	0.04	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Carpediemonas membranifera	CMEM	0.00	0.00	0.40	1.90	-	-	0.03	0.04	0.07	0.23	0.10	5.48	0.15
Boccardiella ligerica	BLIG	0.00	0.00	0.00	5.66	-	-	0.00	0.00	0.12	0.11	7.65	0.00	0.00
Polydora ciliata	PCIL	0.00	0.01	0.00	0.40	-	-	0.00	0.12	0.32	0.02	12.29	0.00	0.01
Sabellaria alveolata	SALV	0.00	1.68	0.00	28.38	-	-	0.00	0.00	0.01	0.00	0.00	0.00	0.06
Ambidexter symmetricus	ASYM	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Neomysis integer	NINT	0.00	0.00	0.00	0.01	-	-	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Armatobalanus allium	AALL	0.22	0.00	0.01	0.11	-	-	0.10	0.47	0.81	0.29	0.68	0.11	0.02
Cyclopina gracilis	CGRA	0.16	0.64	0.02	0.83	-	-	0.05	0.01	0.83	0.51	35.57	0.18	0.29
Euchaeta indica	AIND	0.00	0.01	0.03	0.06	-	-	0.03	0.05	0.00	0.00	0.02	0.00	0.01
Oithona sp. 2 New Caledonia-RJH-2004	OITH	2.87	3.49	0.14	0.39	-	-	0.01	0.00	0.00	0.02	0.02	3.92	15.02
Paracalanus parvus	PPAR	13.35	20.04	0.02	0.03	-	-	17.15	13.59	43.17	18.61	2.97	5.67	9.59
Pseudocalanus elongatus	PELO	0.00	0.00	0.00	0.00	-	-	0.65	1.84	0.00	0.00	0.00	0.00	0.00
Scolecithricella longispinosa	SLON	0.00	0.00	0.00	0.04	-	-	0.03	0.01	0.00	0.01	0.00	0.00	0.01
Stomatolepas praegustator	SPRA	1.91	0.02	0.04	0.06	-	-	0.03	0.68	3.26	1.21	4.31	0.40	0.08
Temora turbinata	TTUR	0.01	0.01	0.00	0.01	-	-	0.67	3.55	1.77	4.68	0.00	0.05	0.00
uncultured eukaryote	MAXI	2.85	11.51	0.24	0.58	-	-	7.11	2.40	9.91	7.73	0.00	0.62	9.04
Chelophyes appendiculata	CAPP	0.05	0.01	0.01	0.04	-	-	0.00	0.00	0.00	9.41	0.00	0.57	10.38
Garveia sp. CC-2005	GARV	0.00	0.00	0.01	0.01	-	-	0.00	18.08	0.00	0.00	0.00	0.01	0.00
Xenopus (Silurana) tropicalis (western clawed frog)	XTRO	0.00	0.00	0.00	0.03	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Adula aff. californiensis MO-2010	MADU	0.08	0.00	0.00	0.02	-	-	0.01	0.03	0.12	0.17	3.72	0.04	0.00
Mytilus edulis	MEDU	0.08	2.19	0.00	0.92	-	-	0.00	1.30	0.79	0.68	0.01	0.26	0.06
Crassostrea gigas (Pacific oyster)	CGIG	0.00	0.06	0.00	0.00	-	-	0.00	0.00	0.00	0.14	0.75	0.66	0.00
Batillaria zonalis	BZON	0.00	0.83	0.00	0.00	-	-	0.00	0.06	0.14	1.50	0.51	0.02	3.10
Euplica scripta	ESCR	0.00	0.00	0.00	0.00	-	-	0.04	0.12	0.48	0.00	0.03	0.00	0.00
Peringia ulvae	PULV	0.00	0.00	0.00	0.45	-	-	0.00	0.00	0.00	0.00	0.09	0.10	0.00
Haminoea hydatis	HHYD	0.00	0.00	0.00	0.00	-	-	0.00	0.26	0.05	3.98	0.01	0.04	0.45
Daptonema setosum	DSET	0.00	0.00	0.00	0.00	-	-	0.01	0.01	0.00	0.00	0.00	6.19	0.00
Appendicularia	APPE	0.20	0.70	0.25	0.04	-	-	4.28	1.78	0.62	0.96	5.08	18.24	1.26
Doliolum nationalis	DNAT	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.01	0.00	0.00	0.00	10.85
Cryptocaryon	CILI	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Monodinium sp. HCB-2005	MONO	0.00	0.00	0.00	0.00	-	-	0.00	0.22	0.01	0.00	0.00	0.00	0.00
Oligotrichia	OLIG	0.01	0.00	0.00	0.02	-	-	0.00	0.06	0.01	0.06	0.00	0.00	0.00
Maristentor dinoferus	MDIN	0.00	0.00	0.06	1.37	-	-	0.03	0.35	0.02	0.13	0.01	0.08	0.07
Dinophyceae	DINO	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.01	0.00	0.02	0.00
Gyrodinium	GYRO	0.00	0.00	0.00	0.00	-	-	0.01	0.01	0.00	0.01	0.01	0.01	0.00
Protoperidinium	PROT	0.00	0.00	0.00	0.00	-	-	0.00	0.00	3.61	0.04	0.00	0.07	0.59
Syndiniales	SYND	0.04	0.00	0.00	0.07	-	-	0.05	0.05	0.08	0.01	0.01	0.07	0.17
Cercozoa	CERC	0.00	0.00	0.00	0.10	-	-	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Chlorarachniophyta	CHLO	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.03	0.29	0.02	0.47	0.04
Imbricatea	IMBR	0.00	0.00	0.00	0.09	-	-	0.00	0.00	0.00	0.00	0.00	0.09	0.00
Thecofilosea	THEC	0.00	0.00	0.00	0.05	-	-	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Stramenopiles	STRA	0.00	0.00	0.03	0.93	-	-	0.01	0.02	0.06	0.03	0.02	1.44	0.03
Chrysophyceae	CHRY	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Paraphysomonas	PARA	0.00	0.02	91.29	19.71	-	-	0.33	1.14	0.63	4.38	0.98	16.85	4.89
Pseudo-nitzschia australis	PNIT	0.04	0.02	0.00	0.00	-	-	0.00	9.13	0.01	0.02	0.11	0.00	0.01
Chaetoceros	CHAE	0.01	0.13	0.00	0.00	-	-	0.00	2.44	0.02	0.02	0.00	0.00	0.01
Lauderia annulata	LANN	0.00	0.12	0.00	6.80	-	-	0.00	1.55	0.01	0.04	0.00	0.00	0.01
Papiliocellulus elegans	PELE	0.01	0.00	0.00	0.01	-	-	0.00	0.00	0.01	0.00	0.27	0.00	0.02
Skeletonema	SKEL	0.12	0.01	0.00	0.00	-	-	0.00	0.09	0.01	0.00	0.00	0.03	0.40
Thalassiosira allenii	TALL	0.03	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.10	0.00	0.00
Tenuicylindrus belgicus	TBEL	0.02	6.04	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Diatomea	DIAT	0.00	0.00	0.00	0.02	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pedinellales	PEDI	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Heterosigma akashiwo	HAKA	0.02	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.09	0.00	0.01	0.00
Acartia_clausi	ACLA	43.03	2.08	1.35	0.22	-	-	60.60	15.22	14.89	18.82	0.10	0.62	0.39
Acartia_tonsa	ATON	0.02	0.00	0.11	1.23	-	-	0.03	0.01	0.01	0.04	12.97	0.04	0.03
Calanipeda_aquaedulcis	CAQU	0.00	0.00	0.02	0.44	-	-	0.06	0.03	0.21	0.02	0.01	0.01	0.00
Calanus_helgolandicus	CHEL	18.26	0.09	0.00	0.11	-	-	0.00	0.00	0.00	0.23	0.00	0.02	0.00
Euterpina_acutifrons	EACU	6.25	28.08	0.00	0.01	-	-	0.01	0.00	0.05	0.00	1.31	4.69	6.63
Pseudodiaptomus_marinus	PMAR	0.00	2.02	0.00	0.04	-	-	0.02	0.00	0.00	0.01	4.42	0.00	0.00

Table S2_c. List of most abundant taxa (> 5 % of abundance for any size-fraction in at least one sample) for the 0.22-20µm. Acronyms used in the CCA are shown.

								SALINITY 30						
#OTU ID	ACRONYM	September '13	October	November	December	January	February	March	April	May	June	July	August	September '14
Hyperamoeba flagellata	HFLA	1.14	1.45	12.68	13.76	9.59	12.73	11.99	7.49	22.05	8.75	4.23	12.91	8.14
Ostreococcus	OSTR	1.22	0.10	0.00	0.08	0.13	0.09	0.51	2.78	0.07	0.03	0.00	0.35	0.28
Pyramimonas	PYRA	1.54	4.15	0.00	0.04	0.00	0.00	5.37	0.55	0.11	0.56	4.23	1.02	9.12
Cryptomonadales	CRYP	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.92	6.46	0.08	0.01
Carpediemonas membranifera	CMEM	0.11	0.06	0.93	1.96	5.95	4.57	1.24	2.47	1.64	1.74	3.65	6.59	3.18
Boccardiella ligerica	BLIG	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.03	0.02	0.02	0.00	0.00
Polydora ciliata	PCIL	0.08	0.02	0.00	0.01	0.01	0.00	0.00	0.05	0.01	0.23	0.00	0.00	0.02
Sabellaria alveolata	SALV	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Ambidexter symmetricus	ASYM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Neomysis integer	NINT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Armatobalanus allium	AALL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.05	0.00
Cyclopina gracilis	CGRA	0.03	7.11	0.01	0.90	0.01	0.00	0.01	2.62	0.07	1.82	0.24	1.90	0.27
Euchaeta indica	AIND	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Oithona sp. 2 New Caledonia-RJH-2004	OITH	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Paracalanus parvus	PPAR	0.00	0.00	0.01	0.00	0.00	0.01	0.10	0.18	0.04	0.05	0.04	0.03	0.10
Pseudocalanus elongatus	PELO	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00
Scolecithricella longispinosa	SLON	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.03	0.04	0.02	0.00	0.00	0.00
Stomatolepas praegustator	SPRA	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.03	0.01	0.01	0.04	0.22	0.02
Temora turbinata	TTUR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
uncultured eukaryote	MAXI	0.00	0.04	0.01	0.00	0.00	0.00	0.23	0.01	0.02	0.02	0.00	0.00	0.01
Chelophyes appendiculata	CAPP	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.23	0.16	0.12	0.19
Garveia sp. CC-2005	GARV	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.43	0.04	0.00	0.00	0.00	0.01
Xenopus (Silurana) tropicalis (western clawed frog)	XTRO	0.00	0.02	0.03	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.02	0.00
Adula aff. californiensis MO-2010	MADU	0.19	0.58	0.09	0.01	0.01	0.00	1.47	0.18	0.07	0.17	0.44	0.14	0.08
Mytilus edulis	MEDU	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.02	0.08	0.00	0.00	0.01	0.00
Crassostrea gigas (Pacific oyster)	CGIG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
Batillaria zonalis	BZON	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00
Euplica scripta	ESCR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Peringia ulvae	PULV	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.00
Haminoea hydatis	HHYD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Daptonema setosum	DSET	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Appendicularia	APPE	0.04	0.17	0.28	0.09	0.05	0.00	0.19	0.47	1.34	0.07	0.04	0.12	0.60
Doliolum nationalis	DNAT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Cryptocaryon	CILI	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.07	0.01	9.90	0.11	0.02	0.01
Monodinium sp. HCB-2005	MONO	0.00	0.00	0.01	0.00	0.77	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.00
Oligotrichia	OLIG	0.02	0.03	1.58	0.15	1.26	0.29	2.17	0.68	0.21	0.21	0.06	0.00	0.02
Maristentor dinoferus	MDIN	0.01	0.04	0.91	3.33	10.89	5.55	4.49	4.89	3.52	2.74	6.20	3.11	1.69
Dinophyceae	DINO	0.21	7.68	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.15	0.03	0.49	4.02
Gyrodinium	GYRO	28.70	1.36	1.20	0.39	0.12	0.01	1.66	7.68	1.06	11.61	12.27	12.42	8.96
Protoperidinium	PROT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.34	0.00	0.00	0.00	0.00
Syndiniales	SYND	0.70	0.39	6.17	0.35	0.04	0.03	0.89	1.60	0.60	0.26	0.21	0.83	0.57
Cercozoa	CERC	0.00	0.00	6.68	4.76	0.04	0.00	0.23	0.08	0.05	0.00	0.00	0.00	0.00
Chlorarachniophyta	CHLO	0.00	0.10	0.02	0.01	0.00	0.00	0.02	0.02	0.06	0.04	0.00	0.02	0.01
Imbricatea	IMBR	0.02	0.03	6.11	4.42	0.03	0.13	0.40	0.10	0.94	0.00	0.05	0.00	0.01
Thecofilosea	THEC	1.13	0.39	0.88	0.02	0.01	0.01	0.19	0.70	0.02	0.07	0.01	0.26	0.23
Stramenopiles	STRA	0.01	0.03	0.35	0.41	1.65	2.57	0.39	0.86	0.57	0.70	0.19	0.33	0.27
Chrysophyceae	CHRY	0.54	8.39	3.11	0.88	0.70	0.88	0.90	0.38	0.05	0.00	0.04	0.02	0.17
Paraphysomonas	PARA	1.95	2.38	18.89	31.84	53.08	53.49	30.53	31.11	28.58	33.99	12.84	28.22	17.65
Pseudo-nitzschia australis	PNIT	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.05	0.02	0.00	0.00	0.00	0.00
Chaetoceros	CHAE	0.01	0.03	0.00	0.01	0.00	0.00	0.00	2.91	0.23	0.00	0.00	0.00	0.00
Lauderia annulata	LANN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.01	0.01	0.00	0.00	0.00
Papiliocellulus elegans	PELE	0.87	1.60	0.03	0.05	0.00	0.00	0.01	0.39	1.31	0.05	23.89	4.02	6.29
Skeletonema	SKEL	3.87	17.14	0.01	0.01	0.00	0.01	0.00	0.07	0.01	0.00	0.01	0.40	7.71
Thalassiosira allenii	TALL	0.02	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Tenuicylindrus belgicus	TBEL	0.00	0.74	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Diatomea	DIAT	8.76	0.36	0.01	0.65	0.01	0.00	0.01	0.01	0.02	0.10	0.43	0.05	0.51
Pedinellales	PEDI	0.05	0.37	9.03	0.09	0.01	0.01	0.18	0.11	0.01	0.00	0.00	0.09	0.31
Heterosigma akashiwo	HAKA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01
Acartia_clausi	ACLA	0.00	0.00	0.00	0.02	0.01	0.00	1.86	0.26	0.04	0.01	0.05	0.01	0.01
Acartia_tonsa	ATON	0.02	0.26	0.00	0.37	0.00	0.01	0.01	0.26	0.03	0.01	0.03	0.05	0.79
Calanipeda_aquaedulcis	CAQU	0.00	0.00	0.00	0.03	0.03	0.12	0.14	0.79	0.02	0.01	0.01	0.01	0.00
Calanus_helgolandicus	CHEL	0.01	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00
Euterpina_acutifrons	EACU	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pseudodiaptomus_marinus	PMAR	0.03	0.02	0.00	0.03	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01
								SALINITY 35						
#OTU ID	ACRONYM	September '13	October	November	December	January	February	March	April	May	June	July	August	September '14
Hyperamoeba flagellata	HFLA	0.10	0.79	8.75	12.34	-	-	1.37	0.96	8.16	3.48	3.44	2.89	5.68
Ostreococcus	OSTR	14.90	0.90	0.08	1.07	-	-	8.78	4.67	0.27	2.24	0.36	0.84	1.08
Pyramimonas	PYRA	0.48	0.05	0.00	0.45	-	-	0.03	0.88	0.18	0.87	0.76	0.36	1.20
Cryptomonadales	CRYP	0.02	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.02	0.00	0.00
Carpediemonas membranifera	CMEM	0.03	0.09	6.24	4.08	-	-	0.86	1.13	1.63	4.77	2.17	3.31	4.81
Boccardiella ligerica	BLIG	0.14	0.00	0.00	0.00	-	-	0.00	0.00	0.03	0.00	0.00	0.00	0.00
Polydora ciliata	PCIL	0.00	0.01	0.00	0.01	-	-	0.00	0.00	0.01	0.01	0.00	0.00	0.00
Sabellaria alveolata	SALV	0.00	0.17	0.07	12.29	-	-	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Ambidexter symmetricus	ASYM	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Neomysis integer	NINT	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Armatobalanus allium	AALL	0.00	0.01	0.00	0.01	-	-	0.00	0.00	0.00	0.14	0.00	0.00	0.00
Cyclopina gracilis	CGRA	0.04	0.04	0.01	0.02	-	-	0.02	0.04	0.04	2.05	0.07	0.07	0.02
Euchaeta indica	AIND	0.00	0.00	0.00	0.00	-	-	0.00	0.02	0.00	0.00	0.00	0.00	0.00
Oithona sp. 2 New Caledonia-RJH-2004	OITH	0.01	0.22	0.02	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.08	0.07
Paracalanus parvus	PPAR	0.39	0.03	0.00	0.01	-	-	0.29	2.85	0.04	0.03	0.01	0.03	0.02
Pseudocalanus elongatus	PELO	0.00	0.00	0.00	0.00	-	-	0.00	1.80	0.00	0.00	0.00	0.00	0.00
Scolecithricella longispinosa	SLON	0.00	0.00	0.01	0.00	-	-	0.00	0.01	0.01	0.00	0.01	0.01	0.00
Stomatolepas praegustator	SPRA	0.00	0.03	0.01	0.00	-	-	0.01	0.01	0.01	0.63	0.00	0.01	0.00
Temora turbinata	TTUR	0.01	0.01	0.00	0.00	-	-	0.00	1.78	0.00	0.00	0.00	0.00	0.00
uncultured eukaryote	MAXI	0.00	0.03	0.03	0.01	-	-	0.08	0.11	0.01	0.02	0.00	0.01	0.01
Chelophyes appendiculata	CAPP	0.32	0.28	0.26	0.03	-	-	0.00	0.00	0.00	0.75	0.03	0.50	0.56
Garveia sp. CC-2005	GARV	0.00	0.01	0.00	0.00	-	-	0.00	0.82	0.03	0.00	0.00	0.00	0.00
Xenopus (Silurana) tropicalis (western clawed frog)	XTRO	0.00	0.01	0.01	0.00	-	-	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Adula aff. californiensis MO-2010	MADU	0.01	0.00	0.01	0.00	-	-	0.42	0.01	0.07	0.00	0.00	0.00	0.00
Mytilus edulis	MEDU	0.00	0.30	0.01	0.04	-	-	0.13	0.45	0.13	0.00	0.00	0.00	0.00
Crassostrea gigas (Pacific oyster)	CGIG	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.07	0.01	0.00
Batillaria zonalis	BZON	0.00	0.01	0.00	0.00	-	-	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Euplica scripta	ESCR	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.01	0.00	0.01	0.00	0.00
Peringia ulvae	PULV	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Haminoea hydatis	HHYD	0.00	0.00	0.00	0.00	-	-	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Daptonema setosum	DSET	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Appendicularia	APPE	0.62	0.17	0.52	0.01	-	-	7.38	5.39	0.93	0.09	0.05	11.48	0.21
Doliolum nationalis	DNAT	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Cryptocaryon	CILI	0.00	0.08	0.02	0.00	-	-	0.01	0.30	0.01	0.18	0.26	0.00	0.00
Monodinium sp. HCB-2005	MONO	0.00	0.00	0.00	0.00	-	-	0.01	0.02	0.01	0.00	0.10	0.00	0.00
Oligotrichia	OLIG	0.23	0.22	0.09	0.02	-	-	0.03	1.10	0.36	0.37	2.20	0.49	0.38
Maristentor dinoferus	MDIN	0.97	1.39	22.33	5.91	-	-	16.43	19.17	5.36	8.46	3.60	12.32	9.80
Dinophyceae	DINO	1.18	0.07	0.00	0.02	-	-	0.01	0.15	0.02	0.98	0.39	0.84	0.19
Gyrodinium	GYRO	0.27	0.27	1.69	0.17	-	-	0.01	0.18	0.73	0.33	0.25	0.27	0.26
Protoperidinium	PROT	0.01	0.00	0.00	0.00	-	-	0.01	0.01	15.56	4.17	0.00	6.17	5.25
Syndiniales	SYND	1.20	1.33	4.60	1.19	-	-	0.08	0.61	0.82	1.30	3.86	1.64	1.97
Cercozoa	CERC	0.00	0.08	0.26	0.15	-	-	0.01	0.02	0.04	0.07	0.01	0.01	0.00
Chlorarachniophyta	CHLO	0.32	0.09	0.03	0.01	-	-	0.00	0.00	0.06	0.02	0.01	0.04	0.06
Imbricatea	IMBR	0.00	0.05	0.18	0.08	-	-	0.02	0.02	0.59	0.00	0.00	0.00	0.00
Thecofilosea	THEC	0.04	0.34	0.03	0.01	-	-	0.08	0.26	0.04	0.03	0.16	0.07	0.03
Stramenopiles	STRA	0.01	0.06	1.50	0.75	-	-	0.17	0.11	0.18	0.05	0.06	0.11	0.23
Chrysophyceae	CHRY	0.01	0.85	0.06	0.62	-	-	0.62	0.04	0.04	0.03	0.01	0.03	0.03
Paraphysomonas	PARA	0.44	0.84	25.87	34.06	-	-	28.76	21.56	25.53	24.05	17.58	17.84	20.17
Pseudo-nitzschia australis	PNIT	0.38	0.13	0.00	0.00	-	-	0.00	1.53	0.03	0.03	1.56	0.01	0.03
Chaetoceros	CHAE	0.26	0.08	0.00	0.00	-	-	0.00	5.47	0.37	0.13	0.00	0.21	0.01
Lauderia annulata	LANN	0.01	0.01	0.00	0.02	-	-	0.00	0.12	0.03	0.00	0.02	0.00	0.02
Papiliocellulus elegans	PELE	2.45	0.20	0.01	0.00	-	-	0.03	0.02	0.75	0.23	3.15	1.96	0.89
Skeletonema	SKEL	2.91	5.41	0.08	0.10	-	-	0.02	0.04	0.02	0.03	0.07	0.70	3.33
Thalassiosira allenii	TALL	13.71	0.22	0.01	0.00	-	-	0.00	0.00	0.00	0.00	13.44	0.09	0.03
Tenuicylindrus belgicus	TBEL	0.23	18.79	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Diatomea	DIAT	0.03	0.09	0.01	1.11	-	-	1.98	0.28	0.02	0.01	0.06	0.02	0.05
Pedinellales	PEDI	0.00	0.05	0.01	0.05	-	-	0.72	0.01	0.02	0.00	0.00	0.00	0.00
Heterosigma akashiwo	HAKA	2.10	0.01	0.00	0.01	-	-	0.00	0.00	0.00	8.59	0.27	0.15	0.09
Acartia_clausi	ACLA	0.00	0.04	0.01	0.01	-	-	0.06	1.99	0.04	0.03	0.02	0.01	0.02
Acartia_tonsa	ATON	0.00	0.00	0.01	0.00	-	-	0.00	0.05	0.00	0.01	0.01	0.01	0.01
Calanipeda_aquaedulcis	CAQU	0.00	0.00	0.02	0.00	-	-	0.01	0.01	0.02	0.00	0.01	0.02	0.00
Calanus_helgolandicus	CHEL	0.01	0.02	0.00	0.01	-	-	0.00	0.00	0.00	0.00	0.01	0.01	0.00
Euterpina_acutifrons	EACU	0.01	0.19	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.05	0.00
Pseudodiaptomus_marinus	PMAR	0.00	0.00	0.00	0.00	-	-	0.00	0.00	0.00	0.00	0.00	0.01	0.00

Discusión general

El metabarcoding ha supuesto un avance para la evaluación de la biodiversidad en muestras ambientales porque es un método rápido que produce una gran cantidad de información a un costo relativamente bajo. Además, representa una alternativa potencial para superar los problemas asociados a la identificación morfológica (Baird & Hajibabaei 2012): dada su sensibilidad, es posible detectar taxones que son raros o se encuentran en baja abundancia (Fonseca et al., 2010; Lindeque et al., 2013) , organismos en cualquier etapa de desarrollo (huevos, larvas, …), especies crípticas (Comtet et al., 2015) e incluso ADN extracelular (Guardiola et al., 2015).

Es por ello que en la primera parte de esta tesis nos hemos dedicado a comparar esta técnica molecular con la microscopia. Dicha comparación ha evidenciado las ventajas del metabarcoding, ya que nos ha permitido establecer patrones espaciales y temporales semejantes a los estudios clásicos, pero además ha permitido detectar especies no indígenas cuya abundancia era tan baja que hubiera sido difícilmente detectable por microscopia. Además, el metabarcoding ha resultado ser válido para medir las abundancias relativas de los diferentes taxones, abriendo la posibilidad de aplicarse como método semicuantitativo.

Sin embargo, a pesar de su potencial, todavía hay mucho trabajo por hacer antes de que esta técnica pueda ser considerada para su adopción generalizada en estudios de ecología planctónica, ya que tiene varias limitaciones en su estado actual. Por ejemplo, un requisito esencial para estos estudios es la accesibilidad a una base de datos completa y "curada" de secuencias de ADN de referencia para especímenes identificados. Esta necesidad no está cubierta y hoy por hoy representa uno de los principales inconvenientes para la aplicación rutinaria del metabarcoding, ya que algunos grupos de organismos como el fitoplancton tienen relativamente pocas secuencias disponibles en las bases de datos públicas. En este sentido, esta tesis ha servido para completar las bases de datos públicas aportando secuencias de especies comunes en la ría de Bilbao, y por extensión comunes en estuarios templados. Así, hemos subido a Genbank las secuencias 18SV9 correspondientes a cinco especies de copépodos y tres de cladóceros; la secuencia 18SV1-2 de dos copépodos y cuatro cladóceros , y la secuencia COI de dos cladóceros.

Otra de las limitaciones del metabarcoding es su dependencia de la PCR, y por ende la elección de primers puede constituir una fuente de sesgo importante. A este respecto, el presente trabajo ha puesto de relieve la falta de adecuación de la región del 18S V1-2 para el estudio de las

comunidades de la ría de Bilbao, ya que falla a la hora de amplificar una especie del género *Acartia* que es muy abundante en esta ría y su cuantificación relativa resulta altamente sesgada. En su defecto proponemos el 18S V9 para posteriores estudios de la ria de Bilbao, y por extensión de estuarios templados.

A lo largo de esta discusión general se evalúan los distintos aspectos de esta técnica en lo que concierne al monitoreo de pláncton, haciendo hincapié en ciertas ventajas e inconvenientes para su posible aplicación práctica.

1. Detección precoz de especies foráneas

Aunque la alta sensibilidad del metabarcoding se ha descrito en otros estudios (por ejemplo, Zhan et al., 2013, Pochon et al., 2013), su aplicación para el monitoreo de invasiones biológicas sólo se ha demostrado recientemente (Zaiko et al., 2015c). En el primer capítulo de esta tesis, se confirma la idoneidad de esta técnica para la detección temprana de especies foráneas en abundancias relativas extremadamente bajas. Más concretamente se detectaron con mayor sensibilidad que la microscopia dos especies de copépodos foráneas (*Acartia tonsa* y *Pseudodiaptomus marinus*). La primera fue descrita en el 2001 y actualmente se encuentra entre las más abundantes, mientras que la segunda fue observada por primera vez en 2010 (Uriarte et al., 2015). Lo que implica que de realizarse estos análisis de forma rutinaria, se podría afrontar con mayores probabilidades de éxito/margen de tiempo las posibles consecuencias y opciones de manejo de estas especies en los ecosistemas.

Aparte de su sensibilidad, existen dos razones principales que hacen del metabarcoding muy útil para la detección temprana de especies invasoras: 1) la capacidad de analizar grandes volúmenes de muestra en comparación con los métodos basados en microscopía (varios litros, en vez de mililitros) lo que permite recolectar taxones invasores para su posterior análisis aunque su abundancia sea extremadamente baja y 2) la posibilidad de detectar la presencia de individuos en estadios tempranos de la vida, como huevos o larvas, o incluso con restos de ADN extracelular, cuya identificación es complicada o imposible, respectivamente, con los métodos tradicionales (Comtet et al., 2015).

Finalmente, se ha de mencionar que los programas de monitoreo no suelen estar diseñados para proporcionar una alerta temprana de especies invasoras lo que complica presentar una respuesta eficaz. Sin embargo, en esta tesis hemos visto que todos los factores mencionados en combinación con su velocidad y costes relativamente bajos (Kelly et al., 2014), hacen del metabarcoding un enfoque alternativo prometedor para la detección temprana de dichas especies. De esta forma, si se incluyese en programas de monitoreo, podrían tomarse medidas preventivas antes de que supusiese un cambio en el ecosistema. Como nota, dos estudios recientes (Zaiko et al., 2015a-b) también han sugerido el valor de combinar esta técnica con el análisis taxonómico actual para la vigilancia del agua de lastre: el principal vector de la mayoría de las introducciones marinas (Golasch et al., 2000).

2. Metabarcoding como método cuantitativo

Hasta la fecha, para las comunidades de plancton se han descrito discrepancias entre el los valores de abundancias relativas o de biomasa estimados mediante microscopía y metabarcoding (Hirai et al., 2015, Massana et al., 2015, Stoeck et al., 2014, Sun et al., 2015); aún así, los estudios que evalúan la naturaleza cuantitativa de esta técnica siguen siendo escasos. En un esfuerzo por llenar este vacío, en la primera parte de esta tesis se comparan las abundancias relativas de taxones comunes entre la microscopía y el metabarcoding, viéndose que se correlacionaron en la mayoría de los casos en nuestro estudio (13 de 20 casos).

La falta de correlación en los 7 casos restantes podría explicarse por los sesgos técnicos introducidos durante la extracción de ADN (Roh et al., 2006), que puede variar con el tipo de organismo o incluso con la etapa de desarrollo, y/o durante la etapa de amplificación por PCR (Gonzalez et al., 2012), que puede favorecer la sobrerepresentación de ciertos taxones. Sin embargo, se ha sugerido que la Variación de Número de Copias (CNV) asociada al rDNA es uno de los principales factores que afectan al valor cuantitativo del metabarcoding (Kembel et al., 2012) y que la incorporación de medidas correctoras adecuadas al análisis puede ayudar a mejorar las estimaciones de abundancia.

A raíz de estos resultados podemos concluir que el metabarcoding, aunque indudablemente aporte información de gran utilidad, seguirá siendo un método semicuantitativo mientras que no sea posible solventar los problemas citados (incluyendo los inherentes a la naturaleza multicopia de los

genes comúnmente utilizados) (Amend et al., 2010; Albaina et al., 2016); es por ello que recomendamos que por ahora estos estudios sean combinados con métodos tradicionales en mayor o menor medida.

3. Análisis de marcador adecuado para metabarcoding

3.1. Subunidad I del Citocromo c Oxidasa

Hace algunos años, el gen mitocondrial del COI se convirtió en la región estándar utilizada en barcoding (Folmer et al., 1994) para identificar y diferenciar especies animales (Hebert et al., 2003), y se esperaba que fuera un marcador candidato ideal para la evaluación de la biodiversidad mediante metabarcoding. Sin embargo, la mayor variabilidad de secuencia de ADN (que le confiere ventaja a lla hora de discriminar especies) asociada a este gen conlleva que los primers tengan sitios de unión mal conservados (Clarke et al., 2014; Deagle et al., 2014), lo que reduce el rango de amplificación de este marcador y pone en cuestión su valor universal. Es por ello que creemos, al igual que otro estudio previo, que este marcador es más adecuado para estudios enfocados en grupos taxonómicos concretos (Zhan et al., 2014), y no aquellos centrados en evaluar/medir la biodiversidad total del ecosistema.

En la segunda parte de esta tesis, la comparación entre los diferentes marcadores testados reveló que el COI tiene un rendimiento más pobre para ciertos grupos taxonómicos, entre los que se incluyen especies claves del plancton estuárico y componentes de pico-plankton. Esto ha de asociarse principalmente a una combinación de la mencionada ineficiencia durante la etapa de amplificación y a una falta de secuencias representativas en bases de datos públicas.

3.2. 18S rDNA

Hoy en día, el gen del 18S nuclear es el marcador más comúnmente utilizado en metabarcoding para la evaluación de la biodiversidad en muestras ambientales eucariotas (por ejemplo, Hirai et al., 2015, Massana et al., 2015, Visco et al., 2015). Existen diferentes regiones hipervariables, en la segunda parte de esta tesis se incluyeron la V1-2 (Fonseca et al., 2010) y la V9 (Stoeck et al., 2010) para comparar su eficacia. Ciertamente proporcionan una cobertura taxonómica más amplia que el COI por lo que son teóricamente idóneos para realizar estudios preliminares de perfiles de biodiversidad total (Zhan et al., 2014). Ahora bien, el hecho de que el 18S cubra un rango

taxonómico tan amplio, se ve contrarrestado por una reducción en la resolución taxonómica de ciertos grupos en comparación con el COI. Más aún, varios estudios han sugerido que el 18S probablemente subestima la verdadera diversidad de especies debido a su naturaleza relativamente conservada y que podría ser sólo adecuado para niveles taxonómicos más altos, y que por lo tanto, los patrones de nivel de especies deben ser interpretados con precaución (por ejemplo, Tang et al., 2012). En este sentido, la segunda parte de esta tesis demostró claramente que el 18S era capaz de identificar más grupos taxonómicos que el COI (como se mostró anteriormente en Zhan et al., 2014) en comunidades planctónicas estuarinas. Además, para nuestro estudio en concreto, la resolución de este marcador resulta ser suficiente ya que las comunidades de la ría de Bilbao están dominadas por pocas especies genéticamente bien diferenciadas entre ellas.

De entre las dos regiones 18S estudiadas, la 18SV9 fué la que presentó un mayor número de correlaciones significativas con los datos microscópicos en nuestro estudio. Además, y al contrario que el 18SV1-2, fué capaz de detectar y diferenciar todas las especies del género *Acartia* de la ría de Bilbao, uno de los organismos más abundantes en muchos estuarios templados tales como el área de estudio (Aravena et al., 2009). Estas razones junto con el pequeño tamaño del amplicón del V9 (~ 150 pb incluyendo primers) da como resultado una ventaja cuando se trata con ADN de relativa baja calidad tal como el obtenido en muestras ambientales, por lo que consideramos que esta región presenta gran potencial para la detección temprana de cambios en la biodiversidad planctónica de estuarios templados.

4. El metabarcoding para análisis de comunidades planctónicas

En el primer capítulo de esta tesis se llevó a cabo un análisis en paralelo usando la técnica tradicional y el metabarcoding en un número reducido de muestras en donde reportamos que el metabarcoding era capaz de replicar las tendencias temporales y espaciales del estuario de Bilbao observadas en los análisis morfológicos (microscópicos), resultando más comparable en el caso del zooplancton que en el fitoplancton. Aún más, aún con el hándicap del déficit existente en las bases de datos públicas de secuencias representativas para la fracción 0,22-20 μm del plancton, el metabarcoding fue capaz de superar a la microscopía en la caracterización del picoplancton.

También cabe destacar que en los análisis de microscopía se encontraron entre muestras analizadas especímenes en diferentes estadíos de desarrollo, tales como copepoditos o larvas, que

en algunos casos no fue posible asignar a un taxón concreto. En cambio, el metabarcoding es capaz de asignar estos organismos a una determinado OTU independientemente de la etapa de desarrollo: esto constituye una de las principales ventajas de esta técnica, ya que permite identificar taxones que son difíciles de asignar en estadíos tempranos.

Por otro lado, en cuanto a los taxones que sólo fueron identificados en el análisis por microscopía, se plantean dos posibles explicaciones al porqué de su ausencia en el análisis mediante metabarcoding: o bien dichos taxones no fueron amplificados o no fueron asignados. En relación con esta última posibilidad, esto podría suceder si no existe al menos una secuencia representativa para ese taxón en concreto en la base de datos de referencia o existe pero se encuentra incompleta. También podría deberse a una posible variabilidad local (de la secuencia o de la región de los primers), una identificación incorrecta por errores durante el análisis bioinformático o una base de datos no suficientemente curada. Para solucionar parte de este problema, en nuestros estudios secuenciamos la región 18SV9 correspondientes a cinco especies de copépodos y tres de cladóceros, la 18SV1-2 de dos copépodos y cuatro cladóceros, y el COI de dos cladóceros presentes en la ría de Bilbao. De esta forma, hemos podido mejorar la asignación taxonómica obtenida mediante metabarcoding (de un 32.6 a un 53.5%; veáse tabla 2 del capítulo 1).

4.1. Ciclo anual del zooplancton

En el tercer y último capítulo de esta tesis realizamos un análisis de las fracciones de tamaño de 20-200 μm y >200, lo que resulta novedoso frente al resto de estudios realizados hasta la fecha puesto que habitualmente suelen centrarse en grandes grupos de organismos (con mucha abundancia de huevos, larvas o copepoditos que no son asignados generalmente). Nuestro estudio confirmó que en el mesozooplancton de este estuario existe una clara dominancia de las especies del complejo *Acartia*, como se ha ido viendo en estudios morfológicos previos (Villate et al., 1994, Uriarte & Villate 2004, Albaina et al., 2009, Aravena et al. 2009, Uriarte et al., 2015). Además, en nuestro caso, la región 18S V9 nos permitió descifrar el estado actual de este complejo, en el que *A. tonsa* domina durante la mayor parte del año en muestras de salinidad 30, mientras que *A. clausi* es la de mayor abundancia en las de salinidad 35 (Fig. 3 del capítulo 3). La separación espacial de *A. clausi* y *A. tonsa* en aguas de salinidad superior e inferior, así como la segregación estacional, también concuerda con estudios previos del área (Fig. 3, Aravena et al., 2009).

Otra de las especies que se encontró en abundancias altas es *Oithona davisae*, de la que también existen citas recientes en el Estuario de Bilbao (Intxausti et al., 2012; Uriarte et al., 2015) y presenta patrones estacionales con picos durante el verano/otoño. Curiosamente, en un principio, la secuencia correspondiente a esta especie fue erróneamente asignada con la base de datos Silva como *Cyclopina gracilis*, pero dado que no existía registro alguno de esta especie en esta área (Villate et al., 1997, Albaina et al., 2009, Uriarte et al., 2015), se realizó una búsqueda específica en el repositorio de GenBank que reveló la mala identificación.

Nuestro hallazgo de otra abundante especie de copépodos *Calanipeda aquaedulcis* también estuvo de acuerdo con estudios previos (Aravena et al., 2009, Albaina et al., 2009, Uriarte et al., 2015). Esta especie ha contribuido significativamente al aumento del número total de copépodos en el estuario de Bilbao durante los últimos años (Uriarte et al., 2015). Se sabe que esta especie alcanza su máxima abundancia a partir de marzo (Uriarte et al., 2015) y, como se muestra en este estudio, hasta junio.

Finalmente observamos la sucesión estacional del zooplancton en el interior del estuario, en la que se produce la dominancia de *C. aquaedulcis* de febrero a junio siendo sustituida por *A. tonsa* y *O. davisae* durante el resto del año. Esta sucesión coincide con al período de baja oxigenación del agua que todavía se registra en el estuario de Bilbao durante parte del verano siguiendo a la estratificación (Intxausti et al., 2012). Estas condiciones, junto con un aumento de la temperatura parecen pudieron haber favorecido el asentamiento de especies con una mayor tolerancia a cierto grado de hipoxia: tanto *A. tonsa* como *O. davisae* han encontrado que habitan aguas bajas en oxígeno (Itoh et al., 2011, Roman et al., 1993). Durante el invierno y la primavera, la dominancia se desplaza a *C. aquaedulcis*, que se considera euriterma, pero usualmente con preferencia por temperaturas más frías (Frisch et al., 2006).

4.2. Ciclo anual del fitoplancton

En este mismo capítulo, el uso de la región del 18S V9, aunque sujeto a las mencionadas limitaciones de resolución, también permitió el análisis de todo el conjunto de la comunidad de fitoplancton a largo de un ciclo anual (fracción de tamaño de 0.22-20 μm). Nuestros resultados mostraron que los crisófitas son el grupo más abundante durante todo el año: entre estas las *Paraphysomonas* fue el género dominante, algo no inusual en los estuarios parcialmente eutróficos

(Bazin et al., 2014), también se encontraron previamente en el Golfo de Bizkaia (Artolozaga et al., 2000) y el Estuario de Bilbao (Cajaraville et al., 2016).

Entre lo reportado aquí destacamos al el género de dinoflagelados *Gyrodinium,* uno de los grupos menos conocidos de protistas marinos (Kubiszyn & Wiktor 2015) y, en cambio, el *Protoperidium*, un género grande y omnipresente de dinoflagelados heterotróficos marinos, que suele seguir floraciones de diatomeas y generalmente exhiben distribuciones costeras (Taylor 1990). Aunque ambos géneros habían sido descritos previamente en la ría de Bilbao, su dinámica temporal no se había seguido durante un año completo como en este estudio (por ejemplo, Seoane et al., 2005). Así, nosotros observamos cómo su abundancia aumenta a partir de abril y hasta septiembre, coincidiendo con el inicio de la primaver y el verano.

Entre las diatomeas, destaca el haber citado por primera vez en la ría de Bilbao al pequeño *Papiliocellulus elegans*: un organismo marino que se encuentra en ambientes costeros y es posiblemente planctónico pero que por su pequeño tamaño requiere microscopía electrónica para su identificación. Es por ello que esta especie ha sido típicamente pasada por alto y la extensión de su hábitat aún no se conoce suficientemente (Round et al., 1990), lo que podría explicar el hecho de que hasta el presente estudio con metabarcoding no se haya reportado previamente la presencia de este organismo en el Estuario de Bilbao. En cambio, el género *Skeletonema* reportado en este estudio, de más sencilla identificación y que se localiza típicamente en aguas costeras de todo el mundo (Round et al., 1990), sí que se ha citado de forma regular en este estuario (Seoane et al., 2005, Laza -Martinez et al., 2007, Hevia-Orube et al., 2016).

En lo referente a los ciliados encontramos, la especie bentónica *Maristentor dinoferus* resultó la dominante pero, al igual que en caso de *C. gracilis*, esto puede corresponder probablemente a un error de asignación taxonómica, ya que este organismo fuerecientemente descubierto en arrecifes de coral (Lobban et al., 2002). Una búsqueda específica en GenBank lo sitúa como fitoplancton sin cultivar, por lo que es totalmente posible que sea una especie para la que todavía no hay secuencias en la base de datos.

Finalmente, en la fracción de tamaño 0.22-20 μm existe poca variación entre las composiciones comunitarias del interior y el exterior del estuario, lo que sugiere que el período de bajo oxígeno no

tiene la misma influencia que en el zooplancton >20 μ . Así, *Paraphysomonas*, el grupo más abundante a lo largo del año, se alimentan principalmente de bacterias (pero no exclusivamente) y alcanzan su máximo durante los meses más fríos en el estuario de Bilbao (noviembre-diciembre a marzo), cuando la falta de nutrientes y la luz solar impide la proliferación de fitoplancton autotrófico y la turbulencia pueden aumentar las tasas de alimentación de protozoos en las bacterias (Rose & Caron 2007). Las diatomeas, una especie más asociada a alta disponibilidades de nutrientes por el contrario, mostraron picos en abundancia durante el verano (julio a octubre), cuando la temperatura fue mayor y la precipitación supuso el aporte de nutrientes de los tributarios. Las especies no fotosintéticas de dinoflagelados se alimentan de diatomeas u otros protistas (Jeong et al., 2010), lo que explicaría por qué comenzaron a ser abundantes durante la primavera (Cajaraville et al., 2016). Por último, aunque los ciliados parecen tener un pico durante el invierno en las muestras recogidas de la salinidad 35, no podemos estar seguros debido a la falta de datos para los meses de enero y febrero.

5. El futuro del metabarcoding

Sobre la base de la discusión anterior queda claro que aunque el metabarcoding todavía tiene ciertas limitaciones como su exposición a los errores asociados a la PCR, la limitaciones de los primers disponibles o la necesidad de completar y "curar" las bases de datos existentes, puede ser una herramienta útil para su implementación en las políticas ambientales, favoreciendo el diseño oportuno de medidas adecuadas de adaptación/mitigación en el caso de la coloniozación por especies foráneas tal y como sesugiere del presente estudio.

Una posibilidad sencilla de evitar los problemas asociados a la PCR sería secuenciar directamente el extracto de ADN mediante metagenómica y/o secuenciación de molécula individual (SMRT sequencing de PacBio; Rhoads & Au, 2015). En cuanto a la cuestión de la resolución taxonómica, mientras que la región V9 amplifica una gama más amplia de taxones, su resolución a nivel de especie no es completa: una posible solución incluiría combinar los resultados de 18S V9 con 18S V1-V2 (evitando así la introducción de un CNV diferente) o con un marcador de alta resolución como COI, para una mejor discriminación entre especies. Una combinación o "cocktail" de primers con una resolución taxonómica más alta (capítulo 2).

En este sentido, la adición de especies locales sin representación previa en la base de datos (como se

demuestra en este estudio, Tabla 2) aumenta significativamente el éxito de la asignación para muestras de campo recolectadas localmente y se recomienda cuando se diseñan estudios de metabarcoding (por ejemplo, Cowart et al., 2015). Pero de momento, hasta que no se disponga de bases de datos de secuencias de ADN de referencia más completas, se recomienda combinar/complementar el metabarcoding con el análisis microscópico en paralelo, al menos para muestras representativas.

Referencias

Albaina A, Villate F, Uriarte I. (2009) Zooplankton communities in two contrasting Basque estuaries (1999–2001): reporting changes associated with ecosystem health. J. Plankton Res, 31: 739–752.

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: 1809–1824. doi:10.1002/ece3.1986

Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Mol Ecol, 19, 5555-5565.

Aravena G, Villate F, Uriarte I, Iriarte A, Ibáñez B (2009). Response of Acartia populations to environmental variability and effects of invasive congenerics in the estuary of Bilbao, Bay of Biscay. Estuar Coast Shelf Sci 83, 621–628. doi:10.1016/j.ecss.2009.05.013

Baird DJ, Hajibabaei M (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Mol Ecol 21, 2039–2044.

Bazin P, Jouenne F, Friedl T, Deton-Cabanillas AF, Le Roy B, Véron B (2014) Phytoplankton Diversity and Community Composition along the Estuarine Gradient of a Temperate Macrotidal Ecosystem: Combined Morphological and Molecular Approaches. PLoS ONE 9, e94110. doi:10.1371/journal.pone.0094110

Cajaraville MP, Orive E, Villate F, Laza-Martínez A and others (2016) Health status of the Bilbao estuary: A review of data from a multidisciplinary approach. Estuar. Coast. Shelf Sci., Special Issue: Functioning and dysfunctioning of Marine and Brackish Ecosystems 179, 124–134. doi:10.1016/j.ecss.2016.01.013

Clarke LJ, Soubrier J, Weyrich LS and Cooper A. 2014 Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. Mol Ecol Resour, 14: 1160-1170. doi:10.1111/1755-0998.12265

Comtet, Thierry, Anna Sandionigi, Frédérique Viard, and Maurizio Casiraghi. 2015. "DNA (Meta)barcoding of Biological Invasions: A Powerful Tool to Elucidate Invasion Processes and Help Managing Aliens." Biological Invasions 17 (3): 905–22. doi:10.1007/s10530-015-0854-y.

Cowart DA, Pinheiro M, Mouchel O, Maguer M, Grall J, Miné J, Arnaud-Haond S. (2015) Metabarcoding Is Powerful yet Still Blind: A Comparative Analysis of Morphological and Molecular Surveys of Seagrass Communities. *PLoS ONE, 10, e0117562.*

Deagle, Bruce E., Simon N. Jarman, Eric Coissac, François Pompanon, and Pierre Taberlet. 2014. "DNA Metabarcoding and the Cytochrome c Oxidase Subunit I Marker: Not a Perfect Match." Biology Letters 10 (9): 20140562. doi:10.1098/rsbl.2014.0562.

Folmer O, Black M, Hoeh W, Lutz R & Vrijenhoek R. 1994 DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3, 294–299.

Fonseca, Vera G., Gary R. Carvalho, Way Sung, Harriet F. Johnson, Deborah M. Power, Simon P. Neill, Margaret Packer, et al. 2010. "Second-Generation Environmental Sequencing Unmasks Marine Metazoan Biodiversity." Nature Communications 1 (7): 98. doi:10.1038/ncomms1095.

Frisch D, Rodríguez-Pérez H, Green AJ (2006) Invasion of artificial ponds in Doñana Natural Park, southwest Spain, by an exotic estuarine copepod. Aquat. Conserv. Mar Freshw Ecosyst 16, 483–492. doi:10.1002/aqc.718

Gollasch S, Lenz J, Dammer M, Andres HG (2000) Survival of tropical ballast water organisms during a cruise from the Indian Ocean to the North Sea. J Plankton Res 22:923–937

Gonzalez JM, Portillo MC, Belda-Ferre P, Mira A. (2012) Amplification by PCR Artificially Reduces the Proportion of the Rare Biosphere in Microbial Communities. PLoS ONE., 7(1):e29973. doi:10.1371/journal.pone.0029973.

Guardiola M, Uriz MJ, Taberlet P, Coissac E, Wangensteen OS, et al. (2015) Deep-Sea, Deep-

Sequencing: Metabarcoding Extracellular DNA from Sediments of Marine Canyons. PLOS ONE 10(10): e0139633. https://doi.org/10.1371/journal.pone.0139633

Hebert PD, Cywinska A, Ball SL, deWaard JR: Biological identifications through DNA barcodes. Proc Biol Sci. 2003a, 270 (1512): 313-321. 10.1098/rspb.2002.2218.

Hevia-Orube J, Orive E, David H, Díez A and others (2016) Molecular and morphological analyses of solitary forms of brackish Thalassiosiroid diatoms (Coscinodiscophyceae), with emphasis on their phenotypic plasticity. Eur J Phycol 51, 11–30. doi:10.1080/09670262.2015.1077394

Hirai J, Kuriyama M, Ichikawa T, Hidaka K, Tsuda A (2015) A metagenetic approach for revealing community structure of marine planktonic copepods. Mol Ecol Res 15, 68-80.

Intxausti L, Villate F, Uriarte I, Iriarte A, Ameztoy I (2012) Size-related response of zooplankton to hydroclimatic variability and water-quality in an organically polluted estuary of the Basque coast (Bay of Biscay). J Mar Syst 94, 87–96. doi:10.1016/j.jmarsys.2011.10.015

Itoh H, Tachibana A, Nomura H, Tanaka Y, Furota T, Ishimaru T (2011) Vertical distribution of planktonic copepods in Tokyo Bay in summer. Plankton Benthos Res 6, 129–134. doi:10.3800/pbr.6.129

Jeong HJ, Yoo YD, Kim JS, Seong KA, Kang NS, Kim TH (2010) Growth, feeding and ecological roles of the mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. Ocean Sci J 45, 65–91. doi:10.1007/s12601-010-0007-2

Kembel SW, Wu M, Eisen JA, Green JL. (2012) Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS Comput Biol, 8(10):e1002743. doi: 10.1371/journal.pcbi.1002743.

Kelly RP, Port JA, Yamahara KM, Martone RG, Lowell N, Thomsen PF, Mach ME, Bennett M et al. (2014) Environmental monitoring. Harnessing DNA to improve environmental management. Science, 344:1455-1456.

Kubiszyn AM, Wiktor JM (2015) The Gymnodinium and Gyrodinium (Dinoflagellata: Gymnodiniaceae) of the West Spitsbergen waters (1999–2010): biodiversity and morphological description of unidentified species. Polar Biol 1–9. doi:10.1007/s00300-015-1764-2

Laza-Martinez A, Seoane S, Zapata M, Orive E (2007) Phytoplankton pigment patterns in a temperate estuary: from unialgal cultures to natural assemblages. J. Plankton Res 29, 913–929. doi:10.1093/plankt/fbm069

Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, Atkinson A (2013) Next Generation Sequencing Reveals the Hidden Diversity of Zooplankton Assemblages. PLoS ONE 8(11): e81327.

Lobban CM, Simpson A, Pochon X, Pawlowski J, Foissner W (2002) Maristentor dinoferus n. gen., n. sp., a giant heterotrich ciliate (Spirotrichea: Heterotrichida) with zooxanthellae, from coral reefs on Guam, Mariana Islands. Mar Biol 140, 411–423. doi:10.1007/s002270100690

Massana R, Gober A, Audic S, Bass D, Bittner L, Boutte C, Chambouvet A, Christen R et al. (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. Environ Microbiol, doi: 10.1111/1462-2920.12955.

Pochon X, Bott NJ, Smith KF, Wood SA. (2013) Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. PLoS ONE., 8: e73935.

Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics. 2015;13:278–89. https://doi.org/10.1016/j.gpb.2015.08.002

Roh C, Villatte F, Kim BG, and Schmid RD. (2006) Comparative study of methods for extraction and purification of environmental DNA from soil and sludge samples. Appl. Biochem Biotechnol, 134:97-112.

Roman MR, Gauzens AL, Rhinehart WK, White JR (1993) Effects of low oxygen waters on Chesapeake Bay zooplankton. Limnol Oceanogr 38, 1603–1614. doi:10.4319/lo.1993.38.8.1603

Rose JM, Caron DA (2007) *Does low temperature constrain the growth rates of heterotrophic protists? Evidence and implications for algal blooms in cold waters. Limnol Oceanogr 52, 886–895. doi:10.4319/lo.2007.52.2.0886*

Round FE, Crawford RM, Mann DG (1990) The Diatoms - Biology & Morphology of the genera. Cambridge University Press.

Seoane S, Laza A, Urrutxurtu I, Orive E (2005) Phytoplankton Assemblages and Their Dominant Pigments in the Nervion River Estuary. Hydrobiologia 549, 1–13. doi:10.1007/s10750-005-1736-6

Stoeck T, Bass D, Nebel M, Christen R, Jones MD, Breiner HW Richards TA. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol, 19 Suppl 1:21-31.

Stoeck T, Breiner HW, Filker S, Ostermaier V, Kammerlander B, Sonntag B. (2014) A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. Environ Microbiol, 1: 430−444.

Sun C, Zhao Y, Li H, Dong Y, MacIsaac HJ, Zhan A. (2015) Unreliable quantitation of species abundance based on high-throughput sequencing data of zooplankton communities. Aquat Biol, 24:9-15.

Tang, Cuong Q., Francesca Leasi, Ulrike Obertegger, Alexander Kieneke, Timothy G. Barraclough, and Diego Fontaneto. 2012. “The Widely Used Small Subunit 18S rDNA Molecule Greatly Underestimates True Diversity in Biodiversity Surveys of the Meiofauna.” Proceedings of the National Academy of Sciences 109 (40): 16208–12. doi:10.1073/pnas.1209160109.

Taylor FJR (1990) Phylum Dinoflagellata. In: Margulis L, Corliss JO, Melkonian M, Chapman DJ (eds), Handbook of Protoctista. Jones and Bartlett Publishers, Boston.

Uriarte I and Villate F. (2004) Differences in the abundance and distribution of copepods in two estuaries of the Basque coast (Bay of Biscay) in relation to pollution. J Plankton Res, 27(9): 863-974.

Uriarte I, Villate F, Iriarte A. (2015) Zooplankton recolonization of the inner estuary of Bilbao: influence of pollution abatement, climate and non-indigenous species. J. Plankton Res. doi: 10.1093/plankt/fbv060

Villate F (1994) Temporal variability of the spatial distribution of the zooplankton community in a coastal embayment of the Basque country in relation to physical phenomena. *Hydrobiologia 288, 79–95.*

Villate F, Moral M, Valencia V (1997) Mesozooplankton community indicates climate changes in a shelf area of the inner Bay of Biscay throughout 1988 to 1990. J Plankton Res 19, 1617–1636.

Visco J, Apothoz-Perret-Gentil L, Cordonier A, Esling P , Pillet L, Pawlowski J. (2015) Environmental monitoring: inferring diatom index from next-generation sequencing data. Environ Sci Technol, 49: 7597-7605

Zaiko A, Martinez JL, Ardura A, Clusa L, Borrell YJ, Samuiloviene A, Roca A, Garcia-Vazquez E. (2015a) Detecting nuisance species using NGST: Methodology shortcomings and possible application in ballast water monitoring. Mar Environ Res, pii: S0141-1136(15)30008-8.

Zaiko A, Martinez JL, Schmidt-Petersend J, Ribicic D, Samuiloviene A, Garcia-Vazquez E. (2015b) Metabarcoding approach for the ballast water surveillance – An advantageous solution or an awkward challenge? Mar Pollut Bull, 92(1-2): 25-34.

Zaiko A, Samuiloviene A, Ardura A, Garcia-Vazquez E. (2015c) Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. Mar Pollut Bull, 100(1):53-59. doi: 10.1016/j.marpolbul.2015.09.030.

Zhan A, Hulak M, Sylvester F, Huang X and others (2013) High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. Methods Ecol Evol 4: 558-565.

Conclusiones

1. El metabarcoding identificó tendencias espaciales y temporales similares a las obtenidas en el análisis morfológico (microscopía) del zooplancton, pero no del fitoplancton. Para este último grupo, el metabarcoding fue capaz de distinguir especies congenéricas (*Thalassiosira)* o organismos muy pequeños como para ser detectados por microscopía (*Micromonas pusilla* o *Ostreococcus tauri*). Sin embargo, el metabarcording también detectó un menor número de taxones en géneros específicos (por ejemplo, a nivel de los géneros *Chaetoceros* o *Teleaulax*), debido a la falta de secuencias representativas para este grupo de organismos en las bases de datos actuales. Todo ello evidencia tanto el potencial del metabarcoding para el monitoreo del zooplancton como la necesidad, a día de hoy, de colaboración entre taxónomos clásicos y moleculares en aras de mejorar las bases de datos (especialmente para el fitoplancton).

2. El metabarcoding mostró una alta correspondencia al comparar las abundancias relativas obtenidas con este enfoque con las de la microscopía, por lo que concluimos que el metabarcoding representa una alternativa metodológica eficiente para el análisis semicuantitativo de, al menos, ciertos grupos taxonómicos muy relevantes en la ría de Bilbao (por ejemplo el genero de copépodos *Acartia*, la especie *Oithona davisae* o las diatomeas centrales).

3. La sensibilidad superior del metabarcoding permitió la identificación de especies foráneas en abundancias apenas detectables por microscopía. Este hecho combinado con el relativamente bajo tiempo y costo asociados de esta técnica, la identifica como una herramienta alternativa prometedora de alerta temprana de especies no indígenas de alto valor/potencial en política/gestión del medio ambiente.

4. Los distintos marcadores y umbrales de similaridad utilizados en esta tesis, cada uno con sus ventajas e inconvenientes, ofrecen vistas ligeramente diferentes de la diversidad taxonómica, por lo que sugerimos que la elección de uno u otro debe realizarse de acuerdo al objetivo de cada estudio concreto y de la comunidad propia del área. En cuanto a la ría de Bilbao, el gen del 18S es el que mejores resultados generales produce debido a su su superior capacidad de amplificación de taxones (mayor carácter "universal" de sus primers). Concretamente, el 18S V9 obtuvo las correlaciones más significativas con los datos microscópicos

evidenciando así su gran potencial para el monitoreo de las comunidades de plancton en la ria de Bilbao. Por otro lado, el 18S V1-2, a pesar de su mayor capacidad de discriminacion taxonómica, no fué capaz de detectar algunas de las especies más características y abundantes en el estuario de Bilbao y en en estuarios templados del Atlántico Norte (por ejemplo, especies del género Acartia).

5. Sin embargo, la baja capacidad de discriminación taxonómica del 18S V9 hace que una combinación de éste con un barcode de mayor capacidad discriminatoria (como el COI o el 18S V1-V2) represente una solución de mayor potencial para su aplicación generalizada en otros ecosistemas.

6. El metabarcoding confirmó que la estructura de las comunidades de plancton eucariota del estuario de Bilbao fluctúan a lo largo del año y está determinada por la variabilidad espacial (salinidad) y estacional experimentada, confirmando los resultados de estudios previos llevados a cabo con técnicas tradicionales. Más concretamente, el período de baja oxigenación que se produce durante el verano en el interior de la ría y la variación térmica se encuentran entre los principales factores ambientales que afectan al zooplancton, mientras que en el caso del fitoplancton, además de la temperatura, la precipitación también es un factor determinante. Sin embargo, este estudio mostró que la fracción tradicionalmente ignorada del plancton de tamaños 0.22-20 μm no respondía a ciclos comparables/tan marcados.

7. Mientras que las bases de datos no estén completas y curadas, la taxonomía obtenida por metabarcoding debería ser validada mediante taxonomía clásica. Un efecto directo de esta falta de completitud son las identificaciones erróneas (debido a que en ausencia de la secuencia correspondiente, son asignadas a otra especie dentro del rango de similaridad). Es por ello que sugerimos analizar de forma exhaustiva las asignaciones taxonómicas obtenidas mediante metabarcoding antes de sacar conclusiones definitivas. Además, el metabarcoding representará un método semicuantitativo hasta que las limitaciones aquí citadas no sean superadas, especialmente las asociadas con la CNV.

8. Finalmente, en esta tesis demostramos que para las comunidades de plancton eucariota de la ría de Bilbao, el metabarcoding, a pesar de sus limitaciones, es una técnica con un gran potencial. Tanto su alta sensibilidad que hace posible aplicación la detección de especies no indígenas a bajos niveles de abundancia, como su capacidad de detectar cambios en las estructuras de las comunidades y patrones tanto estacionales como espaciales, la convierten en una técnica muy útil para ser implementada en el monitoreo del plancton de este sistema.