Nereis. Revista Iberoamericana Interdisciplinar de Métodos, Modelización y Simulación	9	81-89	Universidad Católica de Valencia San Vicente Mártir	Valencia (España)	ISSN 1888-8550

Retos del empleo de la secuenciación masiva de nueva generación (NGS) de la macrofauna bentónica para la evaluación de la calidad ambiental marina

Challenges of the use of new generation massive sequencing (NGS) of the benthic macrofauna for the evaluation of the marine environment quality

Fecha de recepción y aceptación: 14 de diciembre de 2016 y 23 de enero de 2017

Gerardo J. Martí-Chillón[1*], Ana de Luis[1*], Mónica Díez-Díaz[1*], Javier Torres Gavilá[2], José Rafael García-March[2], José Tena Medialdea[2] y Francisco M. Codoñer[1]

[1] Departamento de Ciencias Aplicadas y Tecnológicas. Facultad de Veterinaria y Ciencias Experimentales. Universidad Católica de Valencia San Vicente Mártir.
[2] Instituto de Investigación en Medioambiente y Ciencia Marina (IMEDMAR). Universidad Católica de Valencia San Vicente Mártir.
[*] Estos autores han contribuido de forma equivalente a este trabajo.
Correspondencia: Calle Guillem de Castro, 94. 46001 Valencia, España. *E-mails*: ana.deluis@ucv.es, monica.diez@ucv.es

RESUMEN

La Directiva Marco del Agua 2000/60/CE obliga al diagnóstico ambiental del ecosistema marino, incluyendo la evaluación de las especies de macroinvertebrados considerados bioindicadores presentes en el medio. Hasta la fecha, este tipo de determinaciones se realizan mediante la identificación taxonómica *de visu* de la macrofauna bentónica presente en las muestras y el cálculo de bioíndices asociados, un proceso costoso en términos de tiempo y financiación y, en algunos casos, subjetivo por precisar de un equipo humano altamente especializado y por la dificultad de identificar correctamente determinadas especies. En este sentido, las técnicas de DNA *barcoding* permiten identificar de forma fiable organismos empleando técnicas de secuenciación de DNA y evitando las desventajas de la identificación morfotaxonómica. Por otro lado, el reciente desarrollo de técnicas de secuenciación masiva de DNA de nueva generación (NGS) ha permitido el desarrollo del DNA *metabarcoding*, o caracterización de poblaciones de organismos presentes en una muestra empleando datos genómicos. Este trabajo plantea los retos fundamentales que presenta, a día de hoy, el análisis de organismos bioindicadores de calidad ambiental marina a través de las técnicas de secuenciación NGS.

PALABRAS CLAVE: NGS, DNA *metabarcoding, Directiva Marco del Agua, índices biológicos, bioindicadores*.

ABSTRACT

The Water Framework Directive 2000/60/EC regulates the environmental diagnosis of the marine ecosystem, including the evaluation of species of bioindicator macroinvertebrates present in the environment. To date, these types of determinations are carried out through the morphotaxonomic identification of the benthic macrofauna present in the samples and the calculation of associated biotic indexes, a process that is time-consuming and resource-intensive, being in some cases inaccurate due to the requirement of highly specialized human resources and the difficulty of correctly identifying certain species. In this respect, DNA barcoding techniques allow the reliable identification of organisms using DNA sequencing techniques and avoiding the disadvantages of morphotaxonomic identification. On the other hand, the recent development of New Generation DNA Sequencing techniques (NGS) has allowed the development of DNA metabarcoding, i.e. the characterization of populations of organisms present in a sample using genomic data. This paper shows the fundamental challenges to be overcome in order to establish a NGS sequencing-based assessment of the marine environmental quality.

KEYWORDS: NGS, DNA *metabarcoding, Water Framework Directive, biotic index, bioindicator*.

INTRODUCCIÓN

El análisis biótico del medio marino se centra actualmente en el estudio de comunidades de especies y permite obtener unos índices biológicos predefinidos que categorizan la calidad ambiental en función de la fauna existente, y definen la disposición y el estado del medio. Actualmente la identificación de especies bioindicadoras se realiza mayoritariamente a través de técnicas taxonómicas. Sin embargo, el desarrollo de las técnicas de secuenciación masiva de DNA de nueva generación (*Next-Generation Sequencing* -NGS- en inglés) plantea una alternativa para la determinación de la calidad medioambiental a través del análisis del material genético de muestras ambientales como método diferenciador de las distintas especies presentes. Esto supondría un importante avance respecto a la identificación de organismos *de visu*, más costosa en términos temporales.

El objetivo de este trabajo es analizar los retos fundamentales que presenta, a día de hoy, el análisis de organismos bioindicadores de calidad ambiental marina a través de las técnicas de secuenciación NGS.

REGLAMENTO ACTUAL, BIOINDICADORES Y MUESTREO

La caracterización ecológica y de calidad de las aguas viene dada actualmente por el establecimiento de parámetros biológicos que se fundamentan en la variedad y cantidad de fauna presente. Esta información queda reflejada en el *Water Framework Directive (WFD)*, resuelta en la Directiva de 2000/60/CE de la Comisión Europea [1, 2], al igual que queda identificado en el *Ecological Quality Status* (EcoQ) [3, 4]. En el primer documento citado, se decretan una serie de obligaciones y recomendaciones a seguir por los Estados Europeos, promoviendo la protección y defensa de las aguas, entre otros, la necesidad de redacción de una legislación comunitaria que garantice una calidad ecológica y la disminución de factores contaminantes que aumentan el deterioro de éstas [1]. Por lo tanto, tienen como objetivo principal el establecimiento de un marco para la protección y conservación de los ecosistemas acuáticos.

Se incluye el requisito de determinar diferentes indicadores de calidad del estado ecológico sobre ríos, lagos y aguas costeras, informando así de la composición de la comunidad de especies bioindicadoras presentes así como de indicadores químicos y físicos que afectan, a su vez, a los indicadores biológicos. Los principales indicadores de calidad biológicos son el fitoplancton, macroalgas y angiospermas y la comunidad bentónica de invertebrados. Se ha de obtener datos de investigación que permitan esclarecer bases informativas de los tipos de taxones presentes y adecuados a cada agua afectada, con la intención de obtener una mayor precisión y alta fiabilidad en la categorización de indicadores de calidad como sistemas de control. Éstos deben garantizar una aportación de información que pueda ser usada, siendo extrapolada y comparada con investigaciones similares [1].

Los Estados miembros deben aportar programas de reconocimiento y seguimiento del estado ecológico de las aguas superficiales (ríos, lagos, aguas de transición, aguas costeras), además de la localización de dichos puntos de control como de la periodicidad de los estudios. En algunos casos, los proyectos de determinación de calidad de las aguas se ven involucrados en programas de investigación, al no haber información disponible referente al tema [1].

Por lo tanto, es importante seleccionar puntos de control en zonas específicas que permitan obtener datos generales periódicos o puntuales, por condicionantes de contaminación. En función del grado de polución que afecte a las masas de agua, se esclarece el número de puntos de control por zona afectada, con el objeto de determinar el estado de aquellas que se sospeche del incumplimiento de los requisitos medioambientales, para un control rutinario y para evaluar los cambios generados por medidas de remediación. La periodicidad de estos controles va a venir estipulada por el tipo de análisis (indicador), la fiabilidad, la precisión, y de la masa de agua de estudio, siendo estos análisis, los mínimos controles necesarios. No se puede exceder de los tiempos propuestos, a menos de haber una causa justificada en base a conocimientos técnicos o la apreciación de especialistas. En el caso de los controles de calidad relacionados con indicadores biológicos de macroinvertebrados para aguas costeras, la periodicidad de análisis será de tres años, con posibilidad de verse reducida por distintos condicionantes [1].

En base a los resultados obtenidos en los controles, se dispone una clasificación del estado ecológico señalado por los valores biológicos y fisicoquímicos de los indicadores. En el caso de la fauna bentónica de invertebrados, la presencia de

taxones sensibles a las alteraciones de las condiciones y un grado de diversidad normal en la región, son indicativos de un buen estado de calidad. Por otro lado, en el caso de presencia de taxones indicadores de contaminación (mayor diversidad de la habitual) o ausencia de taxones sensibles, nos informan de que la calidad de esas aguas son relativamente malas [1].

Esta jerarquía del estado ecológico tiene que ser representativa y estandarizada dentro de unos valores comunes. Dicha clasificación se divide en: muy buen estado (azul) para sitios no contaminados, bueno (verde) para regiones parcialmente contaminadas, aceptable (amarillo); para zonas moderadamente contaminadas, deficiente (naranja) lugares con altos niveles de contaminación y malo (rojo) a zonas muy contaminadas [1, 3]. De esta manera, la atribución de dichos colores permite conocer de forma visual qué regiones de un mapa están más o menos contaminadas, con el uso de esta escala de colores asignada por los valores dados por estudio de los bioindicadores [1].

Así, ha sido necesaria la creación de índices biológicos que sean capaces de transformar los valores obtenidos experimentalmente en datos objetivos para la clasificación de las aguas en función de su calidad y contaminación.

ÍNDICES BIOLÓGICOS Y ESPECIES BIOINDICADORAS

Los índices biológicos permiten la transformación de los datos que nos aporta el estudio de los grupos de la macrofauna bentónica a un lenguaje matemático [3]. Gracias a la utilización de esta información a partir de índices biológicos, es posible evaluar el efecto de un tipo de estrés (contaminación) sobre los ecosistemas bentónicos. Los índices requieren que cada una de las especies identificadas sea asignada a un grupo ecológico, proceso muy delicado ya que es en este momento en el que se definen todos los organismos de cada una de las regiones y en las que se basará el estudio [2].

Existe gran variedad de índices utilizados en estudios ecológicos, que varían en función del tamaño de muestra, la metodología de muestreo, grupos zoológicos implicados o los procedimientos de identificación. Parte de éstos, se desarrollan para seguir la Directiva Europea Marco del Agua (WFD), ya comentada, de modo que a partir de la fauna de macroinvertebrados bentónicos es posible asignar un valor de calidad ambiental. De entre ellos, destaca el índice de diversidad de Shannon-Wiener [4].

Los índices se basan en los datos de tolerancia al estrés que aportan las especies recolectadas en los muestreos. Si bien no hay ningún índice que se pueda usar de forma general, algunos como el AMBI (*AZTI Marine Biotic Index*, 2000), se ha extendido a muchos países de Europa usado de forma oficial, y testado su actividad en América, África y Oceanía [5]. Este índice se calcula en base al número de individuos de cada especie presentes, centrándose en el uso de macroinvertebrados bentónicos marinos como indicadores de la salud de un ecosistema [6]. Se fundamenta en la tolerancia de especies presentes a diferentes niveles de estrés ambiental. Actualmente incluye una base de datos con 6000 especies categorizadas con asignación a cinco grupos ecológicos: sensible a la presión y presentes en condiciones no contaminadas (EG1), indiferente al enriquecimiento orgánico y constante (EG_2), tolerante (EG_3), oportunista de segundo orden (EG_4) y de primer orden (EG_5) [6, 7]. El valor del índice de calidad ambiental se obtiene a partir de la siguiente expresión:

$$\text{AMBI} = \frac{(0)(\%EG_1) + (1.5)(\%EG_2) + (3)(\%EG_3) + (4.5)(\%EG_4) + (6)(\%EG_5)}{100} \qquad [2]$$

Su valor oscila entre 0 (estado ecológico correcto) y 6 (estado ecológico malo), siendo 7 cuando el sedimento es azoico. A partir de este índice, se puede clasificar a cada zona en función de su calidad ambiental, considerándose no contaminado (de 0 a 1.2), ligeramente contaminado (1.3 hasta 3.3), moderadamente contaminado (3.4 a 5), altamente contaminado (5.1 a 6) y extremadamente contaminado (6.1 hasta 7) [5].

Recientemente se ha propuesto el uso de este índice, en base a la presencia/ausencia de especies (p/a AMBI) y de la cantidad de biomasa (BAMBI), para la evaluación cuantitativa del número de organismos de cada una de las especies presentes en la muestra [5]. De tal forma, que se pueda por un lado detectar la presencia de organismos y por otro, conocer en qué proporción se encuentra cada especie, y ver así, si los valores están dentro de los límites considerados como normales. Por otro lado, el uso de las nuevas tecnologías de secuenciación permite la combinación de herramientas, caso de la técnica conocida como gAMBI, que combina información de secuencias genéticas con la presencia y ausencia de las especies más frecuentes [5].

La combinación de AMBI, Shannon-Wiener (H') y de riqueza de especies (S) generan un índice multimétrico conocido como M-AMBI, que incluye parámetros de abundancia y composición medibles sobre la fauna bentónica de invertebrados [8]. Según una reciente investigación Subida *et al.* (2012), aunque esta combinación de parámetros representó, una menor capacidad de respuesta frente a un gradiente de contaminación que el AMBI. Este índice multimétrico suele estar más influido por valores de diversidad, donde está más correlacionado con sus componentes, que AMBI con los suyos propios.

Otros de los índices bióticos utilizados con frecuencia son entre otros: BENTIX, MEDOCC (*MEDiterranean OCCidental*), ITI (*Infaunal Trophic Index*) [8], BOPA (*Benthic Opportunistic Polychaetes Amphipods index*) [2] e IOBS (*Oligochaete Index of Sediment Bioindication*) [18], dejando a un lado a los índices de menor calibre como: IP (*Index Pllution*), BQI (*Biological Quality Index*) [4], DKI, UK [2], NQI1 y NSI (*Norwegian Sensivity and Quality Indices*) o el ES (*Stage index*) [8].

ITI es un índice biótico trófico, que al igual que AMBI, se desarrolló para un estudio morfotaxonómico. Se diseñó para el estudio de aguas costeras contaminadas por factores orgánicos, clasificando las especies en cuatro grupos. Su mayor limitación, se basa en la dificultad para determinar en detalle el hábito trófico de muchas especies, marcado por la falta de suficiente información sobre el tema [4].

El uso de BENTIX se articula en tres grupos ecológicos, que posteriormente se reducen a dos en comparación a los 5 de AMBI. En este caso, especies sensibles (GS) y (GT) incluye los tolerantes y los oportunistas, cuyos valores van de 6 a 2 respectivamente [2]. Esta reducción de grupos no afecta al cálculo del índice, en cambio tiene la ventaja de evitar un agrupamiento erróneo al acotar las posibilidades de cálculo y elección de grupo. En este sentido BENTIX se aproxima al planteamiento de ITI, puesto que ambos tienen menos grupos que AMBI, si bien BENTIX posee un mayor valor descriptivo al utilizar factores más determinantes en el cálculo con su fórmula [4]. La ecuación para calcular BENTIX es la siguiente:

$$\text{BENTIX} = \frac{(6)(\%GS) + (2)(\%GT)}{100} \qquad [2]$$

En comparación con AMBI, el término GS incluye a EG_1 (no contaminado) y EG_2 (indiferente), mientras que GT se forma de EG_3 (tolerante), EG_4 (oportunistas de segundo orden) y EG_5 (oportunistas de primer orden). Por ello, el grupo de especies muy afectadas a los cambios se les considera como no tolerantes o sensibles (GS), mientras que el resto se les atribuye el concepto de tolerantes a modificaciones en el ambiente [2].

Por su parte, MEDOCC es un índice basado en indicativos de polución y especies sensibles, caracterizado por clasificar a las especies en cuatro grupos, cuyos valores de cálculo varían de 0 (especies sensibles están presentes) hasta 6 (especies oportunistas predominantes). Los cuatro grupos para categorizar se dividen en: sensibles (EG_I), indiferentes (EG_{II}), tolerantes (EG_{III}) y oportunistas (EG_{IV}) [2]. La ecuación para su cálculo es:

$$\text{MEDOCC} = \frac{(0)(\%EG_I) + (2)(\%EG_{II}) + (4)\%EG_{III}) + (6)(\%EG_{IV})}{100} \qquad [2]$$

Este índice se desarrolló para ajustar los parámetros de AMBI a las condiciones del mar Mediterráneo. Tiene como objetivo establecer una separación entre las diferentes clases, con la posibilidad de generar curvas de abundancia relativa de cada grupo ecológico sin delimitar con valores fijos [9].

Otro de los importantes índices bióticos a destacar es BOPA, especialmente validado para el estudio de aguas altamente cargadas en materia orgánica, si bien la ausencia o escasez de poliquetos y especies no oportunistas, hacen que su valor sea bajo. No siendo muy recomendable su uso en ambientes con buenos recursos ecológicos y de calidad, sin embargo se revela eficaz en comunidades que contengan un número mayor de 20 especies [3].

El valor de este índice se calcula a través de la siguiente fórmula; donde f_P es el ratio de poliquetos oportunistas frente a todos los individuos presentes en la muestra, y f_A la frecuencia de anfípodos, siendo la diferencia entre el número de anfípodos individuales excluyendo los oportunistas *Jassa* frente al total de individuos restante [3].

$$\text{BOPA} = \log\,[f_P \,/\, (f_A + 1) + 1] \qquad [2]$$

Los valores dados tras el cálculo se establecen entre el rango desde 0 cuando f_P es 0, a 0.30103 cuando f_A equivale a 0. Por lo tanto, en el cálculo de dicho índice intervienen únicamente dos grupos taxonómicos, los poliquetos oportunistas y

los anfípodos, de tal forma que se reduce considerablemente el esfuerzo de identificación de macrofauna, si bien lo hace poco eficaz en comunidades pobres de estos grupos [3].

En este índice se introduce un tercer parámetro, la frecuencia f_X, que permite afinar el cálculo. Este valor representa el ratio entre: el número de organismos no poliquetos ni anfípodos, incluyendo los del género *Jassa*, frente al número total de individuos de la muestra. Este valor puede conocerse sabiendo las anteriores funciones, con:

$$f_X + f_P + f_A = 1 \qquad [3]$$

Para poder cumplir con los requisitos de la WFD ha sido necesario adaptar el modelo del ratio de poliquetos oportunistas /anfípodos [3].

AMBI se introdujo para la categorización de la Costa Atlántica basado en la sensibilidad y tolerancia de las especies frente a diferentes condiciones [4]. A pesar de tener gran repercusión en la asignación de la calidad de las aguas utilizando bioindicadores ecológicos, presenta diferencias con sus predecesores que determinan la utilidad de cada uno de ellos para cada estudio definido.

Es destacable la variación de grupos utilizados en cada índice, así AMBI agrupa en 5 frente a los 4 de MEDOCC o los 2 de BENTIX [2, 3]. La organización en un menor número de grupos puede ofrecer ventajas a la hora de eliminar errores de clasificación, simplificar los estudios y disminuir los tiempos de análisis de muestras, sin embargo hay que considerar que puede suponer pérdida de poder discriminatorio y de sensibilidad. Hay que considerar también la dependencia de factores externos y la capacidad de designar las cualidades de un ecosistema, basándose en la habilidad de diferenciar entre las especies de cada clase ecológica [4]. Para establecer diferencias entre los índices, habría que abordar más estudios de comparación de resultados utilizando más de un tipo de índice sobre las mismas muestras.

En el estudio de Subida *et al*. (2012), los resultados para costas del mar Mediterráneo, esclarecen que la mayor correlación de datos final fue dada por AMBI, BENTIX y MEDOCC, considerando que el segundo se desarrolló como adaptación de AMBI para estas regiones.

En ambientes de alta perturbación, el número de macroinvertebrados bentónicos identificados son equivalentes para todos los índices, con un alto grado de especies oportunistas. En áreas de menor perturbación, los datos demuestran que la correlación entre los índices varía, debido al mayor número de especies o por la dependencia de otros factores abióticos que condicionan los resultados, como salinidad o propiedades del sedimento [2].

Los índices AMBI y BENTIX necesitan de un mayor esfuerzo taxonómico, y en ocasiones dificultad en la asignación correcta de los organismos de una comunidad [3]. Además los resultados que aporta AMBI son difíciles de comparar a lo largo del tiempo al igual que BENTIX, en comparación con BOPA, puesto que éste utiliza los valores de frecuencia, permitiendo así la no utilización de unidades para la expresión de los datos [3].

Otra diferencia entre los índices anteriores y BOPA, es la dificultad de cálculo, siendo este más sencillo de obtener, ya que no requiere tanta precisión en relación al número y variedad de especies [3]. De tal forma, este índice puede ser de gran utilidad de uso para programas de investigación enfocados en el conocimiento de comunidades bentónicas de las cuales se posee poca información taxonómica. BOPA, a pesar de ser un índice que reduce su análisis al estudio de poliquetos y anfípodos [2], reduce los costos puesto que minimiza la identificación de especies disminuyendo el tiempo y errores durante la identificación. Así, permite una rápida monitorización de la calidad de las aguas y posibles actuaciones de remediación [3]. El desarrollo y elección de la herramienta apropiada para el análisis y evaluación de la información que aporta la fauna bentónica es crucial para la clasificación y determinación correcta de las comunidades [4].

LIMITACIONES DEL ANÁLISIS MORFOTAXONÓMICO DE LA FAUNA BIOINDICADORA

Una vez se conocen cuáles son los bioindicadores principales de la calidad de las aguas, se debe fijar un protocolo de identificación de cada una de las especies presentes en las muestras. Tradicionalmente, la elaboración de estos estudios se ha realizado a través de un estudio *de visu*, es decir, los organismos aislados se han reconocido a través de un análisis morfotaxonómico, detallando propiedades estructurales y tamaños con los cuales poder clasificar cada organismo encontrado

[10]. Este proceso precisa de un amplio conocimiento en materia taxonómica junto con un alto grado de experiencia, por lo que supone la necesidad de contar con la ayuda de personal técnico muy especializado [11, 12].

El diagnóstico morfológico también se puede ver afectado por una posible identificación incorrecta debida a la plasticidad fenotípica y la variabilidad genética dentro de una misma especie, el parecido físico de las estructuras usadas para la categorización y el pequeño tamaño de los organismos a identificar [13]. Así, el análisis morfológico requiere del empleo de un tiempo considerable, siendo habitual el establecimiento de programas de varios meses, lo que impide la categorización y elaboración de unos resultados a corto plazo. Por todo ello, desde hace unos años se está trabajando en desarrollar nuevas herramientas que reduzcan tiempo y costes en la monitorización ambiental marina, como el empleo de las técnicas de secuenciación de DNA [8].

IDENTIFICACIÓN DE ORGANISMOS BIOINDICADORES POR DNA *BARCODING.*

En la última década se han desarrollado técnicas moleculares basadas en el análisis de DNA que permiten una rápida identificación de especies [14]. Estas herramientas reciben el nombre de DNA *barcoding* (código de barras del DNA, en inglés) y su combinación con los estudios taxonómicos clásicos ha dado lugar a la llamada "taxonomía integrada" [15]. De hecho, un gran número de los trabajos taxonómicos publicados en la actualidad incluyen estudios moleculares [16 y 17, entre otros]. El DNA *barcoding* consiste en la amplificación de regiones cortas del genoma del individuo que se pretende identificar y su posterior comparación con bases de datos genómicas existentes para determinar a qué especie taxonómica pertenece. En conjunto, las técnicas del DNA *barcoding* permiten identificar la fauna presente de forma más rápida que la identificación morfológica tradicional, aunque precisan la puesta a punto de una metodología concreta para cada especie o población a estudiar [15, 18].

Los marcadores genéticos más empleados hasta la fecha para la distinción de especies incluyen el DNA mitocondrial correspondiente a la subunidad I del enzima citocromo oxidasa (COI) [17] para caracterización de animales [5, 13], o la identificación de plantas mediante el empleo de genes de los cloroplastos *matK* y *rbcl* [11] y el gen de la subunidad ribosomal 18S [19]. En algunos estudios se mide la capacidad de resolución taxonómica que tienen estos genes, y en algunos casos éstos no proporcionan suficiente información para una correcta discriminación, como es el caso del 18S en eucariotas marinos [11]. Por ello se ha promovido el análisis de otras regiones génicas como el RNA ribosómico mitocondrial 12S (mt12S) en investigación en nemátodos [15, 20] o el RNA ribosómico mitocondrial 16S (mt16S) [20]. A la hora de determinar qué genes se van a estudiar es importante conocer el grado de conservación de las secuencias entre las distintas especies que se pretenden estudiar.

Distintos estudios recientes, han evaluado la capacidad de los distintos genes marcadores para la diferenciación de especies. Algunos estudios muestran que COI es un buen marcador de biodiversidad cuando se emplean técnicas tradicionales de secuenciación de DNA (Sanger), pero su uso para tecnologías más avanzadas de secuenciación es limitado. Del mismo modo, 18S parece funcionar bien con tecnologías básicas de secuenciación de DNA, siendo además capaz de recoger mayor información taxonómica que otros marcadores [20, 21, 22]. Por otro lado, estudios comparativos entre mt12S y COI muestran que el primero es más corto por lo que es más fácil de amplificar y más apropiado para el análisis de muestras peor conservadas o que presentan daño químico [15]. Con todo ello, en estudios recientes, se propone que la combinación de varios genes fomenta una mayor rigurosidad de los resultados y de su interpretación [5, 15].

Para el empleo de la técnica del DNA *barcoding* es fundamental un diseño adecuado de cebadores o *primers* para la amplificación por PCR de las regiones de interés [10]. Previo al diseño de *primers* se ha de establecer qué organismos se pretende estudiar y se ha de realizar una búsqueda *in silico* de secuencias de DNA que se han de alinear y comparar mediante herramientas informáticas como BLAST [23]. De esta manera se consigue identificar las regiones conservadas de los genes marcadores elegidos, para proceder, así, al diseño específico de los *primers* [24]. Durante este ejercicio, se mantienen controlados el ratio de evolución de los genes [21], la repetición a lo largo del genoma, y la posibilidad de que los *primers* pueden amplificar múltiples regiones, generando una mayor abundancia de lecturas de secuencias [14].

En términos generales los *primers* deben comprender el mayor número de organismos del grupo taxonómico que se pretende diferenciar, flanquear los genes de estudio y ser específicos para regiones puntuales que mantengan sus secuencias conservadas. De este modo se consigue minimizar el número de *primers* y simplificar, por tanto, el proceso [21].

Por último, para el empleo de la técnica de DNA *barcoding* – tanto para el diseño de primers como para la comparación de secuencias obtenidas [10] – es fundamental que las bases de datos genómicos de los genes elegidos sean completas o, en su defecto, se generen estos datos *ad hoc* y previamente a realizar el estudio.

IDENTIFICACIÓN DE ESPECIES BIOINDICADORAS MEDIANTE SECUENCIACIÓN MASIVA DE NUEVA GENERACIÓN (NGS): DNA *METABARCODING*

Desde 2005, el desarrollo de técnicas de secuenciación masiva de DNA (*High-Troughput Sequencing –HTS- o Next-Generation Sequencing* –NGS-, en inglés) ha revolucionado el DNA *barcoding,* introduciendo el concepto de DNA *metabarcoding* [12], también conocido como *community barcoding* [25]. El DNA *metabarcoding* consiste en analizar directamente una comunidad de individuos de distintas especies a partir de muestras complejas, realizando secuenciación masiva del DNA de la biota presente y contrastando los resultados obtenidos con las bases de datos génicas disponibles para identificar los organismos presentes [11]. El uso de plataformas de secuenciación NGS a través de plataformas como *454 pyrosequencing* [22] e *Illumina MiSeq* [26], permiten la obtención de datos exhaustivos (*deep-sequencing*, en inglés), lo que aporta una mayor fiabilidad en los resultados de secuenciación obtenidos y permite identificar especies que de otro modo pasarían desapercibidas [25]. Existe también la posibilidad de aislar el DNA extracelular o libre en el medio, variante denominada *environmental barcoding* o eDNA [12], y que permite obtener el material genético directamente de muestras ambientales (p.ej. aguas) sin necesidad de que los organismos estén vivos o incluso presentes, lo que es útil para muestras con pocos indicios directos de vida biológica [18]. En cualquier caso, es necesario optimizar previamente protocolos específicos para aislar el material genético total presente en las muestras ambientales [10, 18, 25].

Se han empleado las técnicas de NGS para el estudio de la biodiversidad en distintos ambientes terrestres [11], de agua dulce [14, 18] o en sedimentos marinos [18], entre otros trabajos. Recientemente las técnicas NGS también se han extendido al análisis de la fauna macroinvertebrada empleada en la biomonitorización ambiental acuática, tanto en agua dulce [14, 27, 28] como en el medio marino [5, 6, 8], con el objetivo de reducir los tiempos y costes asociados al cálculo de índices bióticos por taxonomía clásica. En este sentido, Lejzerowicz *et al.* (2015) mostraron recientemente que las técnicas de secuenciación masiva eran tan efectivas como el análisis taxonómico para la evaluación de la fauna bioindicadora marina, encontrando un alto grado de correlación estadística entre ambos métodos [8].

Sin embargo, en estos trabajos se muestra que existen varias limitaciones fundamentales para desarrollar, a día de hoy, un nuevo índice biótico válido basado en datos genéticos. Por un lado, las secuencias genómicas disponibles en las bases de datos públicas son insuficientes para la identificación de un buen número de especies [10] e impiden correlacionar los resultados obtenidos con los índices bióticos AMBI o BENTIX [5, 8], que tienen en cuenta un elevado número de especies bioindicadoras. Así, actualmente existen iniciativas para enriquecer las bases de datos con el objetivo de desarrollar un índice AMBI basado en datos genéticos (gAMBI) [5]. Otra posibilidad para reducir el número de especies biológicas a monitorizar sería la adaptación de índices bióticos simplificados que consideren un menor número de taxones, como BOPA, bioíndice basado exclusivamente en el estudio de poliquetos oportunistas y anfípodos [2] que, por tanto, permita hacer uso de bases de datos genómicos más sencillas. No obstante, su uso estaría restringido a determinadas regiones y/o condiciones ambientales [3].

Por otro lado, sería necesario optimizar las regiones genómicas marcadoras y *primers* asociados para la diferenciación eficaz de especies e, idealmente, emplear varios genes marcadores a la vez, puesto que hasta ahora ningún marcador empleado, tal como COI o 18SrRNA, puede considerarse universal [8].

Otra de las posibles limitaciones del empleo de datos genómicos para el cálculo de bioíndices sería la dificultad que presenta el análisis de DNA para obtener datos cuantitativos de la fauna presente, por un lado porque el DNA puede ser estable en el medio aunque el organismo ya no se encuentre con vida y, por otro, porque los organismos con mayor tamaño podrían presentar más células y por tanto más contenido en DNA que otros organismos más pequeños. Esta di-

ficultad podría ser superada empleando índices basados en presencia/ausencia de especies y no tanto en la cuantificación de individuos [5].

Por último, cabe destacar que la implementación de bioíndices de origen genómico implicaría, además, el establecimiento de protocolos estandarizados de toma de muestras y extracción de DNA de distintos sustratos marinos, y la realización de estudios estadísticos de correlación entre los valores genómicos obtenidos y el estado ecológico del medio. Sin embargo, dado el rápido avance tecnológico tanto en la secuenciación de DNA como en las herramientas de análisis de los datos obtenidos, que a su vez vienen acompañados de una reducción en los costes, es de esperar que a medio plazo podamos contar con herramientas fiables de monitorización ambiental a partir de datos genómicos.

CONCLUSIÓN

Distintos trabajos recientes han mostrado el potencial de las nuevas tecnologías de secuenciación de DNA para la evaluación de la fauna bioindicadora en muestras ambientales, en sustitución al análisis morfotaxonómico exhaustivo que se requiere para el cálculo actual de los índices biológicos vigentes y que resulta costoso en términos temporales y económicos. Sin embargo, el empleo a nivel práctico de las técnicas NGS para la monitorización ambiental marina exige superar tres retos fundamentales: (i) la creación de bases de datos genómicos completas incluyendo la macrofauna bentónica bioindicadora de cada hábitat, (ii) el desarrollo de marcadores genéticos fiables y universales, y (iii) el establecimiento de protocolos experimentales y metodología para correlacionar los datos genómicos obtenidos con el estado ambiental del medio.

AGRADECIMIENTOS

Los autores agradecen la financiación recibida por parte de la Universidad Católica de Valencia San Vicente Mártir en la convocatoria 2016-17 de ayudas para proyectos de investigación.

BIBLIOGRAFÍA

[1] European Commission. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 Establishing a Framework for Community Action in the Field of Water Policy. Official Journal 22 December 2000 L 327/1. European Commission, Brussels (2000).

[2] M. D. Subida, P. Drake, E. Jordana, B. Mavric, S. Pinedo, N. Simboura, J. Torres, F. Salas, Response of different biotic indices to gradients of organic enrichment in Mediterranean coastal waters: Implications of non-monotonic responses of diversity measures, *Ecol. Indic*. **19** 2012 106–117.

[3] J. C. Dauvin, T. Ruellet, Polychaete/amphipod ratio revisited, *Mar. Pollut. Bull*. **55** (2007) 215–224.

[4] N. Simboura, A. Zenetos, Benthic indicators to use in Ecological Quality classification of Mediterranean soft bottom marine ecosystems, including a new Biotic Index, *Mediterr. Mar. Sci*. **3** (2002) 77–111.

[5] E. Aylagas, A. Borja, N. Rodríguez-Ezpeleta, Environmental status assessment using DNA metabarcoding: towards a genetics based Marine Biotic Index (gAMBI), *PLoS One* **9** (2014) e90529.

[6] A. Borja, J. Franco, V. Pérez, Marine Biotic Index to Establish the Ecological Quality of Soft-Bottom Benthos Within European Estuarine and Coastal Environments, *Mar. Pollut. Bull*. **40** (2000) 1100–1114.

[7] E. Aylagas, N. Rodríguez-Ezpeleta, Analysis of Illumina MiSeq Metabarcoding Data: Application to Benthic Indices for Environmental Monitoring, *Methods Mol. Biol*. **1452** (2016) 237–249.

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

[9] S. Pinedo, E. Jordana, Spain (Catalonia and Balearic Islands), in: A.C., Heiskanen, A.-S. (Eds.), Water Framework Directive Intercalibration Technical Report. Part 3: Coastal and Transitional waters. *JRC Scientific and Technical reports* (2007) 62–70.

[10] A. Dell'Anno, L. Carugati, C. Corinaldesi, G. Riccioni, R. Danovaro, Unveiling the Biodiversity of Deep-Sea Nematodes through Metabarcoding: Are We Ready to Bypass the Classical Taxonomy?, *PLoS One* **10** (2015) e0144928.

[11] M. Leray, N. Knowlton, Censusing marine eukaryotic diversity in the twenty-first century, *Philos. Trans. R. Soc. B Biol. Sci*. **371** (2016) 20150331.

[12] D. W. Yu, Y. Ji, B. C. Emerson, X. Wang, C. Ye, C. Yang, Z. Ding, Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring, *Methods Ecol. Evol*. **3** (2012) 613–623.

[13] P. D. N. Hebert, A. Cywinska, S. L. Ball, J. R. DeWaard, Biological identifications through DNA barcodes, *Proc. Biol. Sci.* **270** (2003) 313–321.

[14] M. Hajibabaei, S. Shokralla, X. Zhou, G. A. C. Singer, D. J. Baird, Environmental Barcoding: A NextGeneration Sequencing Approach for Biomonitoring Applications Using River Benthos, *PLoS One* **6** (2011) e17497.

[15] E. Ferri, M. Barbuto, O. Bain, A. Galimberti, S. Uni, R. Guerrero, H. Ferté, C. Bandi, C. Martin, M. Casiraghi, Integrated taxonomy: traditional approach and DNA barcoding for the identification of filarioid worms and related parasites (Nematoda), *Front. Zool* **6** (2009) 1.

[16] S. E. Miller, DNA barcoding and the renaissance of taxonomy, *Proc. Natl. Acad. Sci*. **104** (2007) 4775–4776.

[17] C. J. Geraci, M. A. Al-Saffar, X. Zhou, DNA barcoding facilitates description of unknown faunas: a case study on Trichoptera in the headwaters of the Tigris River, Iraq, *North Am. Benthol. Soc*. **30** (2011) 163–173.

[18] R. Vivien, F. Lejzerowicz, J. Pawlowski, Next-Generation Sequencing of Aquatic Oligochaetes: Comparison of Experimental Communities, *PLoS One* **11** (2016) e0148644.

[19] M. Guardiola, M. J. Uriz, P. Taberlet, E. Coissac, O. S. Wangensteen, X. Turon, Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Marine Canyons, *PLoS One* **10** (2015) e0139633.

[20] L. Yang, Z. Tan, D. Wang, L. Xue, M. Guan, T. Huang, R. Li, Species identification through mitochondrial rRNA genetic analysis, Sci. Rep. 4 (2014) 4089.

[21] R. J. Machida, M. Kweskin, N. Knowlton, PCR Primers for Metazoan Mitochondrial 12S Ribosomal DNA Sequences, *PLoS One* **7** (2012).

[22] A. Zhan, S. A. Bailey, D. D. Heath, H. J. Macisaac, Performance comparison of genetic markers for highthroughput sequencing-based biodiversity assessment in complex communities, *Mol. Ecol. Resour* **14** (2014) 1049–1059.

[23] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool, *J. Mol. Biol*. **215** (1990) 403–410.

[24] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen T. L. Madden, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, *BMC Bioinformatics* **13** (2012) 134.

[25] V. Elbrecht, F. Leese, Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass—Sequence Relationships with an Innovative Metabarcoding Protocol, *PLoS One* **10** (2015) e0130324.

[26] S. Shokralla, T. M. Porter, J. F. Gibson, R. Dobosz, D. H. Janzen, W. Hallwachs, G. B. Golding, M. Hajibabaei, Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform, *Sci. Rep*. **5** (2015) 9687.

[27] M. E. Carew, V. J. Pettigrove, L. Metzeling, A. A.Hoffmann. Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Front Zool.* **10** (2013) e45.

[28] J. F. Gibson, S. Shokralla, C. Curry, D. J. Baird, W. A. Monk, I. King I, M. Hajibabaei. Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. Fontaneto D, editor. *PLoS One* **10** (2016) e0138432.