

Nota de Investigación/Research Note

Ejercicio interlaboratorio de bioensayos marinos para la evaluación de la calidad ambiental de sedimentos costeros. IV. Ensayo de toxicidad sobre sedimento con crustáceos anfípodos

Interlaboratory assessment of marine bioassays to evaluate the environmental quality of coastal sediments in Spain. IV. Whole sediment toxicity test using crustacean amphipods

MC Casado-Martínez^{1*}, R Beiras², MJ Belzunce³, MA González-Castromil⁴,
L Marín-Guirao⁵, JF Postma⁶, I Riba⁷, TA DelValls¹

¹ Departamento de Química Física, Facultad de Ciencias del Mar y Ambientales, Polígono Río San Pedro s/n, 11510 Puerto Real, Cádiz, España. * E-mail: mcarmen.casado@uca.es

² Area de Ecoloxia, Facultade de Ciencias do Mar, Universidade de Vigo, Vigo, Galicia, España

³ AZTI-Fundación, Unidad de Investigación Marina, Herrera kaia, Portualdea z/g, 20110 Pasajes, Guipúzcoa, España

⁴ CIS Centro de Investigaciones Submarinas, Laboratorio de Calidad y Medioambiental, Vía Nobel 9, 15890 Santiago de Compostela, A Coruña, España

⁵ Departamento de Ecología e Hidrología, Facultad de Biología, Universidad de Murcia, 30100 Murcia, España

⁶ AquaSense Laboratories, PO Box 95125, 1090 HC Amsterdam, The Netherlands

⁷ Consejo Superior de Investigaciones Científicas, Instituto de Ciencias Marinas de Andalucía, Polígono Río San Pedro s/n, 11510 Puerto Real, Cádiz, España

Resumen

Se estudió la precisión interlaboratorio del bioensayo con crustáceos anfípodos. Se recogieron nueve series de datos procedentes de distintos laboratorios europeos y de cuatro especies distintas (*Ampelisca brevicornis*, *Corophium volutator*, *Corophium multisetosum* y *Microdeutopus gryllotalpa*), todas ellas usadas previamente para la caracterización de la toxicidad de sedimentos en España. Esos resultados se estudiaron para evaluar la variabilidad interlaboratorio de acuerdo a las diferentes especies y los distintos tiempos de almacenamiento de las muestras antes del ensayo. Los resultados mostraron coeficientes de variación similares a los encontrados en estudios previos y permitieron una clasificación homogénea de las muestras entre los laboratorios que completaron el ejercicio con éxito. No se identificaron tendencias significativas debidas al uso de distintas especies y, al parecer, el factor que puede afectar más críticamente la clasificación de las muestras y, por lo tanto, el uso de los resultados para la toma de decisiones, es el tiempo de almacenamiento de los sedimentos previamente al desarrollo del ensayo.

Palabras clave: material de dragado, bioensayo, calidad de sedimentos, España.

Abstract

Interlaboratory variability was studied for the bioassay using crustacean amphipods. Nine series of data were obtained from different laboratories using four different species (*Ampelisca brevicornis*, *Corophium volutator*, *Corophium multisetosum* and *Microdeutopus gryllotalpa*), all previously used for sediment toxicity assessment in Spain. Results were studied for interlaboratory variability and according to different confounding factors. The coefficients of variation were similar to those previously reported for this bioassay and sample classification was homogeneous among the laboratories that successfully completed the exercise. No significant tendencies related to test species were identified and it seems that the factor most critically affecting test results and the classification of samples is storage time before testing.

Key words: dredged material, bioassay, sediment quality, Spain.

Introducción

El bioensayo con crustáceos anfípodos se ha convertido en un ensayo de referencia para la caracterización de sedimentos contaminados y material de dragado, y es usado rutinariamente para evaluar los efectos biológicos potenciales de este tipo de muestras ambientales. Existen en la actualidad protocolos estandarizados (ASTM 1991, Environment Canada 1992, USEPA 1994, RIKZ 1999) para distintas especies (p.e.

Introduction

The bioassay using crustacean amphipods has become a benchmark for contaminated marine sediments and dredged material characterization. It is routinely used to measure the biological effects of sediment samples and standard operating procedures already exist (ASTM 1991, Environment Canada 1992, USEPA 1994, RIKZ 1999) for some species (e.g., *Corophium volutator*, *Ampelisca abdita*, *Rhepoxynius*

Corophium volutator, *Ampelisca abdita*, *Rhepoxynius abronius*) pero se han utilizado otras especies autóctonas con algunas modificaciones a las condiciones de ensayo (Costa *et al.* 1998, Onorati *et al.* 1999). En España se han usado con éxito distintas especies para la caracterización de sedimentos y materiales de dragado: *Corophium volutator* (Casado-Martínez *et al.* 2004), *C. multisetosum* (Belzunce *et al.* 2004), *Ampelisca brevicornis* (Riba *et al.* 2003, 2004) y *Microdeutopus gryllotalpa* (DelValls *et al.* 1998, César *et al.* 2002). La principal razón de la variabilidad en la selección de la especie con la que se realiza el ensayo radica en la distribución a lo largo de las costas españolas de cada una de ellas, ya que su abundancia depende de la zona geográfica y de su disponibilidad a lo largo del año. Aunque todas ellas parecen recomendables para la caracterización de materiales de dragado y el efecto de la sensibilidad en la toma de decisiones puede ser evitado mediante el uso de análisis estadísticos para la clasificación de las muestras como tóxicas o no tóxicas, otros factores pueden ser causantes de toxicidad y deben tenerse en cuenta para el análisis de los resultados. Entre éstos, el tamaño de grano parece tener efectos en algunas especies de anfípodos. Por ejemplo *Rhepoxynius abronius* parece verse afectado por sedimentos muy finos y aunque estas diferencias pueden ser mínimas cuando se estudian sedimentos contaminados, la supervivencia puede llegar a reducirse en un 15% cuando los porcentajes de limos y arcillas son del 80% aun en ausencia de contaminación (Tay *et al.* 1998). Cabe mencionar que estos porcentajes de sedimentos de tamaños finos son comunes en muestras de materiales de dragado (DelValls *et al.* 2001, 2003).

Para el ejercicio se distribuyó un protocolo general a cada uno de los laboratorios participantes ya que la mayoría de ellos disponía de personal experimentado en el uso de este ensayo de toxicidad. Los principales factores que pueden causar variabilidad en los resultados interlaboratorio son el uso de distintas especies de anfípodos de distintas poblaciones que, por lo tanto, estuvieron sujetas a distinta aclimatación, manipulación y mantenimiento de los organismos. Los objetivos principales fueron evaluar la habilidad de los distintos laboratorios para caracterizar sedimentos contaminados como tóxicos o no tóxicos, determinar diferencias en la supervivencia media registrada en los distintos laboratorios y detectar las causas de estas diferencias. En caso de introducir cualquier modificación al protocolo común, ésta se incluyó en la memoria de los resultados. Otros factores, como el tiempo de almacenamiento, han sido estudiados para intentar esclarecer su posible influencia en la toxicidad registrada.

Material y métodos

Los anfípodos fueron recogidos por los laboratorios participantes en las zonas donde cada uno los recogía normalmente para sus ensayos de toxicidad. En todos los casos se trataba de áreas libres de contaminación. Los organismos se aclimataron a las condiciones del laboratorio en acuarios de 20 L con unos 3 cm de sedimento aproximadamente y agua de mar, y con

abronius), though other amphipod species can be used with minor modifications (Costa *et al.* 1998, Onorati *et al.* 1999). In Spain, the following species have been successfully used for sediment and dredged material toxicity assessment: *Corophium volutator* (Casado-Martínez *et al.* 2004), *C. multisetosum* (Belzunce *et al.* 2004), *Ampelisca brevicornis* (Riba *et al.* 2003, 2004) and *Microdeutopus gryllotalpa* (DelValls *et al.* 1998, César *et al.* 2002). The main reason for this variability in test species selection is their distribution along the coasts of Spain, the presence or absence of certain species depending on the geographical zone and their year-round availability. Even though all of them are recommended for dredged material characterization and the effect of different sensitivity for making pass/fail decisions in regulatory programs can be avoided by using statistical analyses for sample toxicity classification, other factors, such as particle size distribution, may cause toxic effects. For example, *Rhepoxynius abronius* can be affected by very fine sediments and even if the effect on toxicity is lower when testing contaminated sediments, survival can decrease by 15% when the silt-clay content is greater than 80% (Tay *et al.* 1998), which is common for dredged sediment samples (DelValls *et al.* 2001, 2003).

For this study, a protocol based on standard procedures was sent together with the sediment samples to all the participating laboratories, most having staff trained in sediment toxicity assessment using amphipods. Variability factors included different species of amphipods and different populations, acclimation and handling. The objectives were to evaluate the ability of the different laboratories to characterize contaminated sediments according to toxicity, to determine differences in mean survival between laboratories and to detect the causes of such differences. Any change made to the general protocol was reported with the test results. Other factors such as time storage have been studied to try to identify their contribution to sediment toxicity.

Material and methods

Toxicity tests were applied to six sediment samples provided by the University of Cádiz in 2003 (Casado-Martínez *et al.* 2006). The amphipods used in the bioassays were obtained by each laboratory from where they normally collect them for routine toxicity analysis, in all cases from clean areas free of contamination. Organisms were acclimated to laboratory conditions in 20-L aquaria with a 3-cm sediment bottom layer and clean seawater, and aeration was adjusted to ensure enough dissolved oxygen without disturbing the sediment. They were fed a mixture of microalgae or commercial food (depending on the laboratory) once a week and water was changed on the following day. During acclimation, the dissolved oxygen concentration, pH, salinity and temperature were controlled.

The bioassay using amphipods was a 10-day static test on whole sediment. Test parameters and conditions are given in table 1, though some slight modifications were introduced depending on the test species. Sediments were added to the test

Tabla 1. Parámetros y condiciones a seguir para el desarrollo del test con crustáceos anfípodos en el laboratorio.
Table 1. Parameters and conditions to develop the test using crustacean amphipods in the laboratory.

Parameters	Conditions
1. Test type	Static; on whole sediment
2. Temperature	15–20°C (depending on the species)
3. Salinity	32–40
6. Photoperiod	Natural of the season; also continuous light
7. Test chambers	Glass, 2 L (recommended cylindrical and covered)
8. Volume of sediment	250 mL (or 1:4 sediment/water)
9. Volume of overlying water	1 L (or 1:4 sediment/water)
10. Water renewal	No
11. Size and state of organisms	<i>Ampelisca brevicornis</i> , 3–5 mm; <i>Corophium volutator</i> , 5 mm or larger; <i>Mycrodeutopus gryllotalpa</i> , 5 mm or larger; <i>Corophium multisetosum</i> , 10 mm
12. Number of organisms per chamber	20
13. Number of replicates	3–5
14. Feeding regime	No
15. Aeration	12 h before introducing the organisms, to ensure dissolved oxygen concentrations equal or higher than 90% saturation
16. Overlying water	Clean seawater
17. Water quality	pH, ammonia, salinity and dissolved oxygen at the beginning and at the end of the test
18. Test duration	10 days
19. Endpoints	Survival
20. Test acceptability	90% survival in the negative toxicity control

aireación ajustada para no remover el sedimento del fondo. Los organismos se alimentaron con una mezcla de microalgas o alimento comercial (dependiendo del laboratorio) una vez por semana y el agua sobrenadante se cambió al día siguiente. Durante el periodo de aclimatación se controlaron el pH, la concentración de oxígeno disuelto, la salinidad y la temperatura.

El ensayo con anfípodos se desarrolló a 10 días de exposición en condiciones estáticas y sobre el sedimento en bruto. Los parámetros para la realización del ensayo en el laboratorio se incluyen en la tabla 1, aunque se tuvieron en cuenta ciertas modificaciones en las condiciones dependiendo de la especie utilizada. Los sedimentos se añadieron a las cámaras de exposición hasta una profundidad de 3 cm y posteriormente se les añadió agua de mar en relación 1:4 v/v sedimento:agua. Una vez que los sedimentos hubieron decantado se inició la aireación al menos durante 12 h previamente a la introducción de los organismos. Se usaron 20 anfípodos por replica y un mínimo de tres replicas por muestra. Pasados los 10 días de exposición se registró el porcentaje de supervivencia mediante el tamizado (0.5 mm) de las muestras y se calculó el porcentaje medio de mortalidad en cada una. Los organismos desaparecidos se consideraron como muertos.

Para evaluar la variación interlaboratorio se ha calculado el coeficiente de variación (CV) para cada muestra como el

vessels until attaining a 3-cm bottom layer and clean seawater was added at a ratio of 1:4 v/v (sediment:water). They were left to settle and then aerated for at least 12 h before amphipods were added. Twenty randomly-selected amphipods were used per replicate and each sample consisted of a minimum of three replicates. After the 10-day exposure period, the percentage of surviving organisms was estimated by sieving the samples through a 0.5-mm mesh. Mean mortality was calculated for each sample. Missing organisms were considered dead.

To determine the interlaboratory precision, the coefficient of variation (CV) was calculated for each sample by dividing the standard deviation (SD) by the mean for all laboratories (X):

$$CV (\%) = \frac{SD}{X} \cdot 100$$

The effect of interlaboratory variability on the sample classification was also evaluated according to the individual test results for each laboratory. Percentage of mortality was normalized by angular transformation and treated as a normal distribution. Statistically significant differences ($P \leq 0.05$) between a negative toxicity control (sample A) and the sediment samples were established by means of an ANOVA and Tukey's test using the SPSS 11.0 program.

cociente entre la desviación estándar (SD) y la media de los laboratorios (X):

$$CV (\%) = \frac{SD}{X} \cdot 100$$

Además se estudió el efecto de esta variabilidad en la clasificación de las muestras según los resultados del ensayo en cada laboratorio individualmente. El porcentaje de mortalidad se normalizó mediante transformación angular y fue tratado como una distribución normal para posteriormente buscar diferencias entre el control negativo de toxicidad (sedimento A) y cada una de las muestras. Las muestras significativamente homogéneas ($P \leq 0.05$) fueron identificadas mediante un test ANOVA y test de Tukey mediante el programa estadístico SPSS 11.0.

Resultados

En total se recogieron datos de nueve laboratorios participantes. Los laboratorios (identificados del 1 al 9), las fechas de ensayo y las especies utilizadas por cada uno de ellos se presentan en la tabla 2. Se encontraron diferencias importantes en el tiempo de almacenaje de las muestras de sedimento previamente al inicio del ensayo, muy por encima de los tiempos recomendados que son de tres a cuatro semanas (USEPA 1994), sin embargo, por diversas razones este tiempo varió desde las tres semanas hasta tres meses para algunos laboratorios. Los resultados de mortalidad media y la desviación estándar se presentan en la figura 1, en la que se aprecia la disminución de la toxicidad registrada por los laboratorios 7, 8 y 9 para todas las muestras excepto la F. Todos los laboratorios excepto el 6 y el 9 alcanzaron los valores mínimos de supervivencia en el sedimento control para la aceptabilidad del ensayo (90%). Por ello los resultados de los laboratorios 6 y 9 no fueron considerados para el cálculo de los parámetros estadísticos. Los parámetros de calidad del agua (concentración de oxígeno disuelto, pH, salinidad, temperatura y amonio total) estuvieron en el rango permitido para las especies.

Los resultados del análisis para identificar las muestras estadísticamente homogéneas ($P \leq 0.05$) se muestran en la figura 2. Como puede observarse, todos los laboratorios clasifican a la muestra A como no tóxica y a las muestras F y C como las más tóxicas. El resto de las muestras (B, D y E) podrían ser clasificadas como moderadamente tóxicas aunque se observan diferencias en sus valores de mortalidad y en su clasificación respecto al resto de muestras.

Los resultados del estudio interlaboratorio se resumen en la tabla 3. Si incluimos para el cálculo todos los resultados excepto aquellos que no cumplen los criterios de aceptabilidad para el control negativo, el coeficiente de variación interlaboratorio fue alto para la muestra A (114.5%) debido principalmente a los bajos valores de mortalidad en todos los laboratorios, que distorsionan el valor de la estimación de la

Results

Data sets were collected from nine participating laboratories (numbered from 1 to 9); the test dates and test species are shown in table 2. Important differences were reported for the sample storage period before toxicity testing, some laboratories failing available guidelines (USEPA 1994). Recommended storage time is three to four weeks, but for diverse reasons it ranged from three weeks to three months. Mean mortality results and standard deviation are represented in figure 1. A decrease in sediment toxicity can be observed for laboratories 7, 8 and 9 for all the samples except F. All the laboratories except 6 and 9 registered mortality values for the negative toxicity control under the maximum guideline for test acceptability (10%); hence, the results of these two laboratories were not considered for further analyses. Water quality parameters (dissolved oxygen concentration, pH, salinity, temperature and total ammonia) were in the range where no adverse effects are expected on test organisms.

The results of the statistical analysis to identify homogeneous responses ($P \leq 0.05$) are shown in figure 2. All laboratories classified sample A as not toxic and samples F and C as the most toxic. Samples B, D and E could be considered moderately toxic, though some differences were observed regarding the percentage of mortality and the classification relative to the rest of the samples.

The results of the interlaboratory precision study are summarized in table 3. When the statistics were calculated including all the results except those not meeting the acceptance criteria, a high CV was obtained for sample A (114.5%), mainly because of the low mean mortality that increased this value. The rest of the samples had CVs ranging from 19.8% (sample F) to 69.8% (sample D). When the test results from laboratories that reported storage periods longer than four weeks were also excluded from the calculations, the CVs ranged from 10.5% (sample F) to 81.8% (sample A), with a

Tabla 2. Días transcurridos desde la recogida de los sedimentos y el inicio del ensayo, y especie utilizada en cada laboratorio.

Table 2. Days after sampling when the test was started and species used at each participating laboratory.

Laboratory	Days after sampling	Species
1	1 day	<i>Ampelisca brevicornis</i>
2	1 day	<i>Corophium</i> sp.
3	2 days	<i>Corophium multisetosum</i>
4	8 days	<i>Ampelisca brevicornis</i>
5	15 days	<i>Mycrodeutopus gryllotalpa</i>
6	60 days	<i>Corophium</i> sp.
7	65 days	<i>Corophium multisetosum</i>
8	90 days	<i>Corophium volutator</i>
9	110 days	<i>Corophium</i> sp.

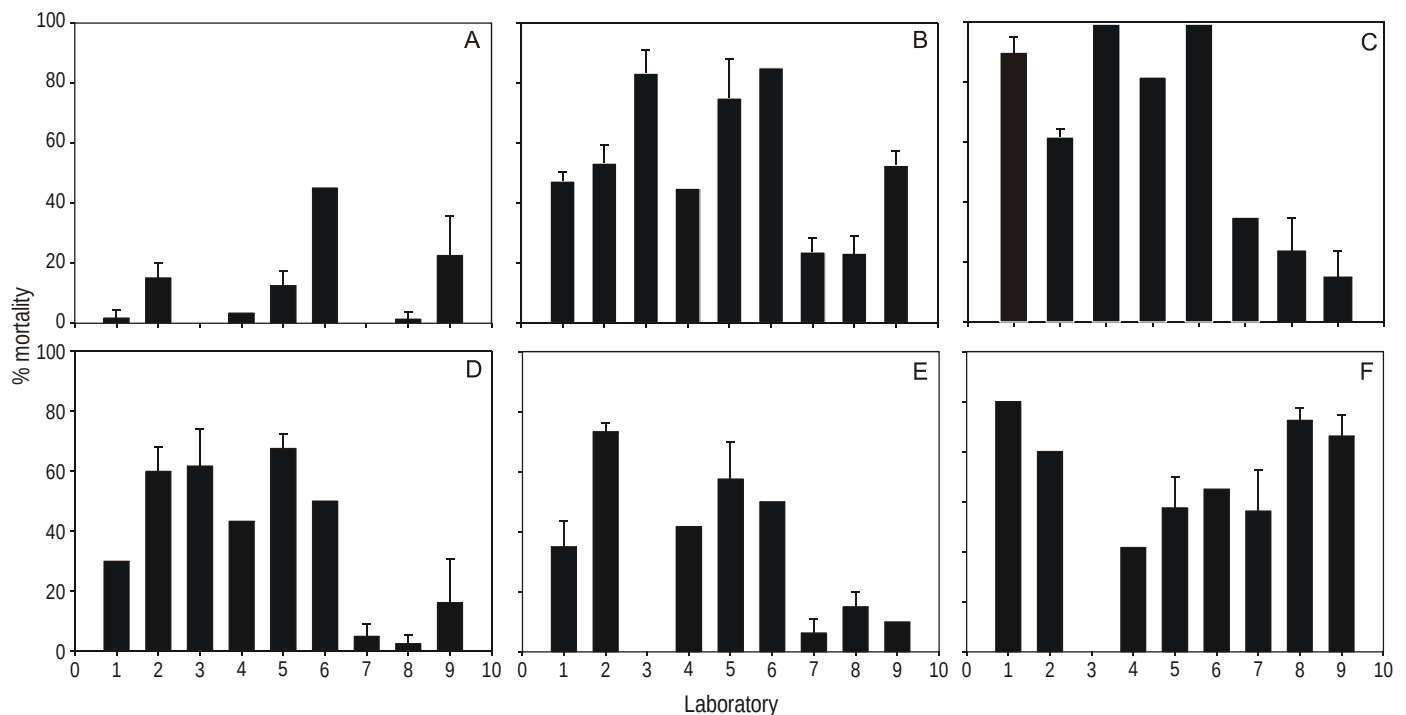


Figura 1. Resultados en porcentaje medio de mortalidad y desviación estándar para cada laboratorio y cada una de las muestras. Los laboratorios están numerados de 1 a 9 y las muestras de A a F.

Figure 1. Mean percentage of mortality and standard deviation for each laboratory and each sample. Laboratories are numbered from 1 to 9 and samples from A to F.

variabilidad mediante este método. Con el resto de las muestras el CV varía entre 69% y 19%, para las muestras D y F respectivamente. Si se calculan estos parámetros excluyendo también los laboratorios que desarrollaron el ensayo fuera del tiempo máximo de almacenamiento recomendado (cuatro semanas) los CV se encuentran entre 81% para la muestra A y 10.5% para la muestra F, con un CV medio del 33% (23% si excluimos la muestra A). Con todas las muestras se obtuvieron valores medios de mortalidad mayores a 50% excepto con la muestra A.

Discusión

El bioensayo con crustáceos anfípodos se realizó con éxito en casi todos los laboratorios. Excepto dos laboratorios (los laboratorios 6 y 9) que superaron los valores de mortalidad permitidos para el control de toxicidad negativo, el resto de los resultados se consideraron aceptables y no se encontraron factores importantes de confusión para el análisis de dichos resultados. Si el laboratorio tiene cierto nivel de experiencia previa, como en la mayoría de los casos, el ensayo puede ser desarrollado con facilidad. Cabe mencionar que en uno de los laboratorios cuyos resultados quedaron invalidados por la elevada mortalidad en el control negativo de toxicidad era la primera vez que se realizaba el ensayo.

El ensayo parece adecuado para la caracterización en laboratorio de la toxicidad de muestras de dragado ya que los

1	A 1.6	D 30.0	E 35.0	B 48.3	C 90.0	F 100
2	A 15.0	B 53.3	D 60.0	C 61.7	E 73.3	F 80.0
3	D 61.67	B 83.33	C 100			
4	A 3.33	E 41.67	D 43.33	B 45	C 81.67	F 100
5	A 12.5	E 57.5	D 67.5	B 75	C 100	F 100
7	A 0.0	D 5.0	E 6.2	B 23.7	C 23.7	F 56.2
8	A 1.2	D 2.5	E 15.0	C 15.0	B 23.3	F 92.5

Figura 2. Resultado del análisis estadístico de los porcentajes de mortalidad de cada laboratorio (excepto aquellos cuya supervivencia en el control no cumplía los criterios de aceptabilidad) del ejercicio de intercalibración del bioensayo con anfípodos ($P \leq 0.05$).

Figure 2. Results of the statistical analysis of the mortality percentages for each laboratory (except those whose control survival did not meet the acceptance criteria) during the intercalibration exercise of the bioassay using amphipods ($P \leq 0.05$).

Tabla 3. Variabilidad interlaboratorio de los resultados de mortalidad media del ensayo de toxicidad con crustáceos anfípodos. n.a. = dato no disponible.
Table 3. Interlaboratory variability of mean mortality from sediment toxicity tests with crustacean amphipods; n.a. = not available.

Laboratory	Sample					
	A	B	C	D	E	F
1	1.6	47.3	90.0	30.0	35.0	100.0
2	15.0	53.3	61.7	60.0	73.3	80.0
3	n.a.	83.3	100.0	61.7	n.a.	n.a.
4	3.33	45.0	81.7	43.3	41.7	100.0
5	12.5	75.0	100.0	67.5	57.5	100.0
6	45.0	85.0	35.0	50.0	50.0	65.0
7	0.0	23.8	23.8	5.0	6.25	56.3
8	1.3	23.3	15.0	2.5	15.0	92.5
9	22.5	52.5	n.a.	16.3	10.0	86.3
SD ^a	6.43	23.0	35.4	26.9	25.3	17.5
Xm ^a	5.62	50.2	67.4	38.6	38.1	88.1
CV ^a	114.5	45.9	52.5	69.8	66.2	19.8
SD ^b	6.63	17.3	15.9	15.5	17.1	10.0
Xm ^b	8.11	60.8	86.7	52.5	51.9	95.0
CV ^b	81.8	28.4	18.4	29.4	33.0	10.5

^a SD, Xm and CV include all data points except those for laboratories 6 and 9.

^b SD, Xm and CV include only sediments meeting the control survival criterion and not tested at a later date.

laboratorios clasificaron eficazmente las muestras de acuerdo con la contaminación presente (Casado-Martínez *et al.* 2006). Dentro de la reglamentación para el proceso de este tipo de sedimentos se han usado distintas aproximaciones para la clasificación de las muestras según la respuesta del ensayo de toxicidad. Algunos países han optado por criterios absolutos en forma de valores de mortalidad que clasifican las muestras como tóxicas si éstos son superados (por ejemplo el Reino Unido considera tóxicas las muestras con mortalidades superiores al 40% en el ensayo con *Corophium volutator*). Otras agencias utilizan el doble criterio de un valor mínimo de mortalidad superior a un sedimento de referencia o control y además la diferencia estadística (den Besten 2003). En este caso se ha utilizado el criterio del 20% de mortalidad superior al sedimento control y la diferencia estadística para considerar una muestra como tóxica (USEPA 1998, Environment Canada 2000). Como se ha mencionado anteriormente, la clasificación de las muestras según su toxicidad fue similar en todos los laboratorios aunque ciertamente existe variabilidad en los porcentajes de mortalidad registrados en cada uno de ellos. Una de las causas de esta variabilidad es el tiempo de almacenamiento de las muestras que parece provocar la disminución de la toxicidad medida. El laboratorio 7, que realizó el ensayo cuatro semanas después del muestreo de los sedimentos, obtuvo resultados de toxicidad para las muestras D y E estadísticamente similares ($P \leq 0.05$) al control de toxicidad negativo y las muestras B y C podrían ser consideradas moderadamente tóxicas. Cuando el tiempo de almacenamiento aumenta a más de cuatro semanas, como es el caso del laboratorio 8, sólo los

mean CV of 33% (23% excluding A, the negative toxicity control). Mean mortalities greater than 50% were obtained for all the samples except for A.

Discussion

The bioassay using crustacean amphipods was carried out successfully by most of the laboratories. Except for two laboratories (6 and 9) that failed the acceptance criteria due to higher mortalities in the negative toxicity control, the test results were acceptable since no interfering factors were found. One of the laboratories that did not meet the acceptance criteria was performing the test for the first time.

These results indicate that the test is suitable for dredged material characterization since laboratories classify the samples based on sediment contamination. Different approaches have been used for regulatory processes to classify sediment samples according to the toxic response. Some countries are using an absolute mortality value to classify the samples as toxic. In the UK, for example, samples are considered toxic when the mortality values are higher than 40% in the test using *C. volutator*. Other agencies use two criteria: a minimum percentage of mortality higher than the control or reference sediment and the statistically significant difference (den Besten 2003). In this exercise samples were classified as toxic based on 20% mortality higher than the control sediment and the statistical difference (USEPA 1998, Environment Canada 2000). As mentioned before, the classification of samples according to the toxicity was similar for all laboratories despite some

efectos producidos por la muestra F serían considerados como estadísticamente diferentes al sedimento control ($P \leq 0.05$), mientras que las otras cuatro muestras C, E, B y D se considerarían, al igual que la muestra A ($P \leq 0.05$), no tóxicas. Dado que el aumento del tiempo de almacenamiento parece afectar la toxicidad registrada y, por lo tanto, la clasificación de los materiales, éste debería ser un factor crítico a considerar por las agencias reguladoras para aceptar los resultados. Aunque los resultados de este ejercicio deben tomarse con precaución, cabe destacar que los resultados de los ensayos realizados tras ocho semanas desde la recogida de las muestras ya no obtuvieron la misma clasificación de los sedimentos, clasificando como no tóxicos materiales de categoría III con altas concentraciones químicas (Casado-Martínez *et al.* 2006).

En cuanto a la variabilidad de los resultados, el valor de CV encontrado con la muestra A, considerada control de toxicidad negativo, es notablemente superior a los valores encontrados con el resto de las muestras. Esto no significa que esta variabilidad afecte al análisis de los resultados obtenidos ya que los resultados clasifican claramente en todos los casos la muestra como no tóxica, sino más bien refleja los inconvenientes del uso de este valor para estimar la variabilidad, ya que si expresásemos los resultados como porcentaje de supervivencia, el CV tomaría valores muy elevados con aquellas muestras con valores medios más bajos, que en este caso corresponderían a las muestras más tóxicas. Del mismo modo parece haber mayor variabilidad en el ensayo de muestras medianamente tóxicas, la cual disminuye al aumentar la toxicidad: las muestras D y E, con valores de mortalidad de ~50%, obtuvieron CV ~30%; con la muestra B la mortalidad fue de 60% y el CV de 28%; mientras que las muestras C y F, las más tóxicas de las analizadas con valores medios de mortalidad de 85% y 95%, registraron menor variabilidad con CVs de 18% y 10%, respectivamente.

La posibilidad de falsos negativos, entendidos como sedimentos tóxicos que son clasificados por un ensayo como no tóxicos, debe ser estudiada por sus implicaciones ambientales. En este sentido, parece que el mayor factor que puede hacer aumentar el número de falsos negativos es el aumento del tiempo de almacenamiento. Por el contrario los falsos positivos, entendidos como los sedimentos no tóxicos clasificados como tóxicos, son también de importancia cuando se deben usar los resultados de estos ensayos para la toma de decisiones, especialmente si han de ser utilizados como principal herramienta legislativa en procesos de gestión. Para minimizar el número de decisiones erróneas y, por lo tanto, los costes ambientales y económicos asociados a este tipo de actividades, deben conocerse los factores que pueden aumentar la toxicidad de las muestras; es decir, los factores que pueden aumentar el número de falsos positivos asociados al método. No sólo la variabilidad y la naturaleza de los sedimentos van a causar efectos tóxicos. Existen numerosos factores que pueden enmascarar la respuesta tóxica, como por ejemplo cambios en el pH o la presencia de determinados compuestos como sulfuros o amonio, aunque una correcta caracterización de la

variabilidad en the mortality values. One of the reasons for this variability is the storage period, which apparently causes a decrease in sediment toxicity. Laboratory 7 tested the sediments after four weeks and classified samples D and E as statistically similar ($P \leq 0.05$) to the negative toxicity control, while samples B and C could be considered moderately toxic. If the storage time is longer than four weeks, as occurred at laboratory 8, only sample F is considered statistically different from the control ($P \leq 0.05$) and samples C, E, B and D would be classified as not toxic ($P \leq 0.05$). Since an increase in sample storage time seems to change the toxic response and its classification, it should be considered a determining factor to accept test results for regulatory purposes. Although these results must be taken with caution, we could say that the results obtained after a certain storage time may not classify the sediments similarly, and even sediment classified as category III due to high contaminant concentrations (Casado-Martínez *et al.* 2006) would be considered not toxic.

The results of the variability study produced the highest CV for sample A, the negative toxicity control. This variability did not affect the analysis of the results because all laboratories classified this sample as not toxic, but the inconvenience of using this statistic to estimate variability is evident. If the results were to be expressed as survival values, the CV would be extremely high for those samples reporting the lowest mean values that would be attributed to the most toxic samples. At the same time, it seems that moderately toxic samples are related to a higher variability, which decreases with an increase in toxicity: samples D and E, with mean mortality values of 50%, obtained CVs of ~30%; sample B had 60% mortality and a CV of 28%; and samples C and F, the most toxic samples with mortalities of 85% and 95%, respectively, had CVs of 18% and 10%, respectively.

The possible false nontoxic sediments (toxic sediments classified as not toxic according to test results) should be studied to avoid adverse environmental effects related to contaminated sediments and dredged material disposal. Storage time seems to be an important factor that can increase the number of these false negatives. On the other hand, false toxic sediments (not toxic sediments classified as toxic according to test results) are also important when using a bioassay for decision-making and especially if they are to be used as principal regulatory tool for managing dredged materials. If we want to minimize incorrect decisions and the related environmental and economic costs, we have to minimize the number of false positives related to the method and to study the factors that are actually increasing their number. It is not only the variability and nature of sediments that cause toxicity. Many other factors can also influence the toxic response, such as pH changes or certain compounds such as hydrogen sulphides or ammonia. A correct water quality characterization can be helpful for early identification and to ensure that concentrations are among the acceptance criteria. Another possible confounding factor is the grain size distribution, since the species used in this study live in sediments with different characteristics: *A. brevicornis* and

calidad del agua puede ayudar a identificar estos factores y asegurar que se encuentran dentro de los límites de tolerancia para la especie. Otro posible factor de confusión es la distribución de tamaños de grano del sedimento. Las distintas especies utilizadas se encuentran normalmente en sedimentos de distinta naturaleza: las especies *Ampelisca brevicornis* y *Corophium multisetosum* son características de sedimentos arenosos (<5% de limos <63 µm) pero *Corophium volutator* predomina en sedimentos limosos (contenido en limos <63 µm mayor al 20%). Se han realizado distintos estudios para identificar posibles efectos adversos de este factor y para establecer los rangos óptimos para determinadas especies. Un estudio realizado por la agencia ambiental canadiense incluso estableció una clasificación de las especies recomendadas en ese país según el grado de sensibilidad al tamaño de grano (Environment Canada 1992). Dado que los resultados de este ejercicio no pueden llevar a conclusiones de este tipo, se recomienda el estudio de los resultados en cada caso y de acuerdo a los factores que pueden afectar a la especie utilizada hasta que no se disponga de rangos de aplicabilidad y/o recomendaciones explícitas por parte de las agencias competentes.

En general, se puede considerar que los resultados de este primer ejercicio son similares a los obtenidos previamente en otros ejercicios interlaboratorio aún cuando se incluyen los resultados de ensayos realizados tras 8 y 12 semanas del muestreo de los sedimentos (CV = 61%). Cuando se cumplen todos los criterios de aceptabilidad, el valor medio de CV de 33% está incluso por debajo del valor de 46% desarrollado por Parkhurst *et al.* (1992) tras la revisión de la variabilidad en ensayos agudos realizados con compuestos químicos individuales. Este último valor, que se desarrolló con resultados para distintas especies, puede ser considerado como el CV esperado cuando se realizan ensayos de toxicidad debido al factor biológico (Still *et al.* 2000). Además, la variabilidad en nuestros resultados es similar a la obtenida con *C. volutator* por Bowmer (1993), quien registró CVs entre 39% y 52% aunque con sedimentos contaminados artificialmente y gran experiencia de los laboratorios en el desarrollo del ensayo. En el mismo rango, Schlekert *et al.* (1995) estudiaron la variabilidad para el ensayo con *Eohaustorius estuarius* (CV de 1.5–45%), *Leptocheirus plumosus* (CV de 5–65%) y *Ampelisca abdita* (CV de 5–146%) y Mearns *et al.* (1986) con *Rhepoxynius abronius* obtuvieron CVs de 6–81%. Aunque la posible influencia de la experiencia en el desarrollo del ensayo en los resultados y en su interpretación puede evitarse mediante cursos de perfeccionamiento, es más difícil encontrar especies comunes para las distintas zonas geográficas; no obstante, parece que el uso del doble criterio para la clasificación de las muestras como tóxicas puede evitar la variabilidad debida al uso de distintas especies con distintas sensibilidades.

Agradecimientos

Los autores expresan su agradecimiento a Pablo Vidal por su ayuda durante el muestreo y la preparación de los

C. multisetosum are usually found in sandy sediments (<5% silt <63 µm), but *C. volutator* predominantly occurs in silty sediments (content of silt <63 µm higher than 20%). Different studies have focused on identifying the possible adverse effects due to this factor and to establish the guidelines for each species. The Canadian Environmental Agency published a classification of test species according to their preferences for grain size distribution (Environment Canada 1992). Since our results are not so conclusive, it would be recommendable to study the results case by case and together with the factors affecting the test species until applicability ranges can be established by the interested agencies.

The results of this first interlaboratory exercise in Spain are similar to previous test results, even including those reported after 8 and 12 weeks (CV = 61%). Mean CV was 33% for the laboratories that met the acceptance criteria; this is lower than the value of 46% reported by Parkhurst *et al.* (1992) after reviewing the variability of acute responses to particular chemicals. This value, developed using different species, can be considered the expected CV when using biological toxicity tests (Still *et al.* 2000). Similar results were reported by Bowmer (1993) using *C. volutator* (CVs of 39–52%), spiked sediments and highly trained laboratories. In the same range, Schlekert *et al.* (1995) used *Eohaustorius estuarius* (CVs of 1.5–45%), *Leptocheirus plumosus* (CVs of 5–65%) and *Ampelisca abdita* (CVs of 5–146%), and Mearns *et al.* (1986) used *Rhepoxynius abronius* (CVs of 6–81%). The variability associated with the different experience of the laboratories can be avoided through training courses and further experience, but it is more difficult to find common available species; thus, it seems that the double criterion method used for sediment toxicity classification is suitable to avoid differences in test species selection and different sensitivities.

Acknowledgements

The authors thank Pablo Vidal for his help with the sampling and sample handling, and Maria José Salamanca, Nuria Fernández, Laura Martín and Natalia Jiménez for their help during the toxicity tests. This study was carried out under a joint research project between CEDEX and the University of Cádiz. Part of the work was funded by the Spanish Ministry of Science and Technology (REN2002_01699/TECNO). The first author was supported by a grant (FPI) from the Spanish Ministry of Education and Science. We thank A. Luque for the useful comments on the final manuscript.

sedimentos, así como a María José Salamanca, Nuria Fernández, Laura Martín y Natalia Jiménez por la ayuda durante los ensayos de toxicidad. Este estudio se realizó como parte de un proyecto conjunto entre el CEDEX y la Universidad de Cádiz. Parte del trabajo fue financiado por el Ministerio Español de Ciencia y Tecnología (REN2002_01699/TECNO). MC Casado-Martínez agradece la financiación mediante una beca del programa Nacional de Formación de Personal

Investigador al Ministerio de Educación y Ciencia. Nuestro agradecimiento a A Luque por sus comentarios durante la preparación del manuscrito final.

Referencias

- ASTM. 1991. Standard guide for conducting 10 day static sediment toxicity test with marine and estuarine amphipods. In: Annual Book of ASTM Standards. Vol. 11.04, E 1367-90. American Society of Testing and Materials, Philadelphia, PA, pp. 1052–1076.
- Belzunce MJ, Franco J, Castro R, Pérez V, Aierbe E, Borja A, Muxika I, Jiménez-Tenorio N, Casado-Martínez MC, DelValls A. 2004. The use of integrative methods for the evaluation of dredged material from Spanish ports: A case study. 13th European SETAC Congress (Society of Environmental Toxicology and Chemistry), Prague.
- Bowmer CT. 1993. Method for the assessment of acute toxicity of contaminated sediment using the burrowing heart urchin *Echinocardium cordatum*. Test guideline for OSPAR sediment re-worker ring test. Organization of Applied Scientific Research TNO, Delft, the Netherlands. Report IMW R 93/317.
- Casado-Martínez MC, Buceta JL, DelValls TA. 2004. The use of the bioassay using amphipods to characterize dredged material from Spanish ports: Comparative study of *Ampelisca brevicornis* and *Corophium* sp. 6th International Symposium on Sediment Quality Assessment. Watershed and Sediment Management, from Source to Sink. Antwerp.
- Casado-Martínez MC, Buceta JL, Forja JM, DelValls TA. 2006. Interlaboratory assessment of marine bioassays to evaluate the environmental quality of coastal sediments in Spain. I. Exercise description and sediment quality. *Cienc. Mar.* (this issue).
- César A, Marín-Guirao L, Vita R, Marín A. 2002. Sensitivity of Mediterranean amphipods and sea urchins to reference toxicants. *Cienc. Mar.* 28: 407–417.
- Costa FO, Correia AD, Costa MH. 1998. Acute marine sediment toxicity: A potential new test with the amphipod *Gammarus locusta*. *Ecotoxicol. Environ. Saf.* 40: 81–87.
- DelValls TA, Forja JM, Gómez-Parra A. 1998. The use of multivariate analysis in linking sediment contamination and toxicity data to establish sediment quality guidelines: An example in the Gulf of Cádiz (Spain). *Cienc. Mar.* 24: 127–154.
- DelValls TA, Casado-Martínez MC, Riba I, Martín-Díaz ML, Forja JM, García-Luque E, Gómez-Parra A. 2001. Investigación conjunta sobre la viabilidad de utilizar ensayos ecotoxicológicos para la evaluación de la calidad ambiental del material de dragado. Technical Report for CEDEX, December, 2001, Puerto Real. Cádiz.
- DelValls TA, Casado-Martínez MC, Riba I, Martín-Díaz ML, Forja JM, García-Luque E, Gómez-Parra A. 2003. Investigación conjunta sobre la viabilidad de utilizar ensayos ecotoxicológicos para la evaluación de la calidad ambiental del material de dragado. Technical Report for CEDEX, Noviembre, 2003, Puerto Real. Cádiz.
- Den Besten PJ, Deckere E, Babut MP, Power B, DelValls TA, Zago C, Oen AMP, Heise S. 2003. Biological effects-based sediment quality in ecological risk assessment for European waters. *J. Soils Sed.* 3: 144–162.
- Environment Canada. 1992. Biological test method: Acute test for sediment toxicity using marine and estuarine amphipods. Rep. EPS 1/RM/26. Environmental Protection, Conservation and Protection, Ottawa, Ontario.
- Environment Canada. 2000. Report on biological toxicity tests using pollution gradient studies. Sydney Harbour. Ministry of Public Works and Government Service, Canada. Rep. EPS 3/AT/2, 104 pp.
- Mearns JA, Swartz RC, Cummins JM, Plesha P, Chapman PM. 1986. Interlaboratory comparison of a sediment toxicity test using the marine amphipod *Rhepoxynius abronius*. *Mar. Environ. Res.* 51: 113–129.
- Onorati F, Bigongiari N, Pellegrini D, Giuliani S. 1999. The suitability of *Corophium orientale* (Crustacea, Amphipoda) in harbour sediment toxicity bioassessment. *Aquat. Ecosyst. Health Manage.* 2: 465–473.
- Parkhurst BR, Warren-Hicks WJ, Noel L. 1992. Performance characteristics of effluent toxicity tests: Summarization and evaluation of data. *Environ. Toxicol. Chem.* 11: 771–791.
- Riba I, Zitko V, Forja JM, DelValls TA. 2003. Deriving sediment quality guidelines in the Guadalquivir Estuary associated with the Aznalcóllar mining spill: A comparison of different approaches. *Cienc. Mar.* 29: 261–264.
- Riba I, Casado-Martínez MC, Forja JM, DelValls TA. 2004. Sediment quality on the Atlantic coast of Spain. *Environ. Toxicol. Chem.* 23: 271–282.
- RIKZ. 1999. The 10d marine amphipod *Corophium volutator* mortality sediment toxicity test. Standard Operating Procedure Nr: SPECIE-01. National Institute for Coastal and Marine Management (RIKZ), 17 pp.
- Schlekat CE, Scott KJ, Swartz RC, Albrecht B, Antrim L, Doe K, Douglas S, Ferretti JA, Hansen DJ. 1995. Interlaboratory comparison of a 10-day sediment toxicity test method using *Ampelisca abdita*, *Eohaustorius estuarius* and *Leptocheirus plumosus*. *Environ. Toxicol. Chem.* 14: 2163–2174.
- Still I, Rabke S, Candler J. 2000. Development of a standardized reference sediment to improve the usefulness of marine benthic toxicity testing as a regulatory tool. *Environ. Toxicol.* 15: 406–416.
- Tay KL, Doe KG, MacDonald AJ, Lee K. 1998. The influence of particle size, ammonia, and sulfide on toxicology of dredged materials for ocean disposal. Chapter 39. In: Wells PG, Lee K, Blaise C (eds.), *Microscale Testing in Aquatic Toxicology: Advances, Techniques and Practice*. CRC Press, New York, pp. 559–574.
- USEPA. 1994. Methods for assessing the toxicity of sediment-associated contaminants with estuarine and marine amphipods. United States Environmental Protection Agency, EPA/600/R-94/025.
- USEPA. 1998. Evaluation of dredge material proposed for discharge in waters of the US. Testing manual (The Inland Testing Manual). United States Environmental Protection Agency, EPA/823/F/98/005.

Recibido en noviembre de 2004;
aceptado en septiembre de 2005