

---

# Relative performance of the Southern California Benthic Response Index using species abundance and presence-only data: Relevance to DNA Barcoding

---

*J. Ananda Ranasinghe, Eric D. Stein, Peter E. Miller<sup>1</sup> and Stephen B. Weisberg*

## ABSTRACT

Benthic indices used to assess sediment quality are based on having abundance data for all species in the community. Molecular genetic approaches, including DNA barcoding provide promise for improving the speed and accuracy of taxonomic identifications. Future application of DNA barcoding by analyzing bulk environmental samples through next-generation sequencing will enable rapid species identification, but this gain may be offset by the inability of these methods to quantify species abundances. To address this issue we evaluated the performance of the Southern California Benthic Response Index (BRI) when species abundance data were removed from its calculation. The presence BRI was created by eliminating abundance weighting while preserving species identity. The relationship between the presence and abundance BRI was highly significant, with an  $r^2 = 0.98$ . The presence BRI also validated almost equally to the abundance BRI when applied to the spatial and the temporal monitoring data used to validate the original BRI. Simulations were conducted to assess how large the barcode library must be for an effective index, with changes in the  $r^2$  between the presence and abundance BRI determined as taxa were systematically removed from calculation of the presence BRI. The  $r^2$

remained above 0.8 with fewer than 400 species in the library.

## INTRODUCTION

Marine benthic macrofauna are frequently used as indicators of environmental condition because they reside in sediments where contaminants accumulate and their immobility allows them to integrate exposure at a site (Diaz *et al.* 2004, Borja *et al.* 2009, Pinto *et al.* 2009). Benthic community composition is typically summarized using benthic indices that allow easy communication of complex biological information as a single number that ranks sites on a scale from good to bad. These index values allow managers to prioritize impacted sites, track trends over time, or correlate biological responses with stressor data.

Challenges in using benthos as indicators are the cost, time and error associated with identifying the biota. Organisms must be manually separated from the sediment, which can take more than a day and is subject to underestimation as some remain hidden among the debris (Ranasinghe *et al.* 2003). Every captured organism must then be identified, typically to species, which requires a number of highly skilled taxonomists that provide expertise over the range of different taxonomic groups. This adds substantial cost and is subject to error, particularly when the

---

<sup>1</sup>Canadian Centre for DNA Barcoding, Biodiversity Institute of Ontario, University of Guelph, ON, Canada

specimens are damaged or immature life stages are present.

DNA barcoding has the potential to increase the speed, accuracy and resolution of species identification (Borisenko *et al.* 2009, Stoeckle and Hebert 2008, Janzen *et al.* 2009). Barcoding involves identifying species based on a short gene sequence from a standardized portion in the genome, and for animals this is the mitochondrial cytochrome c oxidase 1 gene (CO1). Using standard molecular biology tools, DNA is extracted from the specimen tissue, and a 658 base pair region of the CO1 gene is amplified by polymerase chain reaction and sequenced (Hebert *et al.* 2003). DNA from unknown specimens collected in benthic samples can be identified by comparing their barcode sequences to the Barcode of Life Database (BOLD, <http://www.boldsystems.org>) reference library.

DNA barcoding using Sanger sequencing, as it is currently practiced, has the potential for more accurately and rapidly enumerating benthic organisms, at a cost that is comparable (or slightly higher) than traditional morphology-based methods. Next-generation sequencing has the potential to reduce costs and increase the speed of obtaining results by analyzing bulk environmental samples. Natural samples would be homogenized, tissue lysed, DNA extracted and the entire sample analyzed for species presence based on DNA barcodes. Although this method promises much greater speed and lower costs (Hajibabaei *et al.* 2011), a drawback is that it produces presence-only information whereas prevailing benthic indices require abundance for each species. Another consideration is that barcoding provides identification of unknown specimens by matching them to species in the reference library. Therefore, the reference library must contain a sufficient number of species to allow environmental samples to be analyzed to a level where there is enough taxonomic resolution to apply commonly used benthic indices. Here we redevelop the benthic index presently used for regulatory assessments in southern California as a presence-only index and compare its performance to that of its abundance-based counterpart. Second, we evaluate the minimum number of taxa needed in a reference library to calculate a credible presence-based index.

## METHODS

The index used for comparison between abundance and presence-absence derivation was the Smith *et al.* (2001) Benthic Response Index (BRI). The BRI is based on the abundance-weighted average pollution tolerance of species in a sample. The pollution tolerance scores (“p-values”) are developed using ordination analysis to place sites along a pollution disturbance gradient, with pollution tolerance scores assigned to each species based on the position of its peak abundance along the gradient. Lower scores indicate sensitive species and higher scores indicate pollution tolerant species.

The presence BRI (BRI-P) was calculated using the p-values of Smith *et al.* (2001), but the average tolerance score was calculated as the sum of species tolerance scores divided by the number of taxa without abundance weighting. Assessment thresholds for the resulting index were then developed by regression against the BRI and selecting thresholds corresponding to the original BRI assessment categories: i) natural benthic assemblages; ii) the loss of biodiversity, above which 25% of the species pool occurring in reference samples no longer occurred; iii) loss of community function, where echinoderms and arthropods were lost from the assemblage; and iv) defaunation, where 90% of the species pool in reference samples no longer occurred.

### Assessing Performance of the Presence BRI

Performance of the BRI-P was assessed using the two data sets used to validate the original BRI, both of which were independent of the calibration data used to develop the indices. The first tested whether the BRI-P reproduced known temporal gradients of benthic conditions over several years near a southern California waste-water outfall using data from two Los Angeles County Sanitation Districts monitoring sites which were sampled monthly since 1972. The first site, Station 6C (located 2220 m from the outfall) was severely impacted in the early 1970s and has improved since that time (Stull *et al.* 1986, Stull 1995). The second site, Station 0C (located 14,720 m from the outfall) was less affected than Station 6C, but has also improved. The hypothesis was that BRI-P index values should decrease over time at Stations 6C and 0C and that index values will be higher and decrease more at Station 6C than at Station 0C.

The second tested whether the index reproduced a known spatial gradient between two stations on the 60 m isobath from the Orange County Sanitation Districts outfall. Previous studies have shown that Station 0 located near the outfall has altered species composition in comparison to reference Station Con, which is located 7840 m from the outfall (County Sanitation Districts of Orange County 1991).

### Effect of Reducing the Number of Taxa

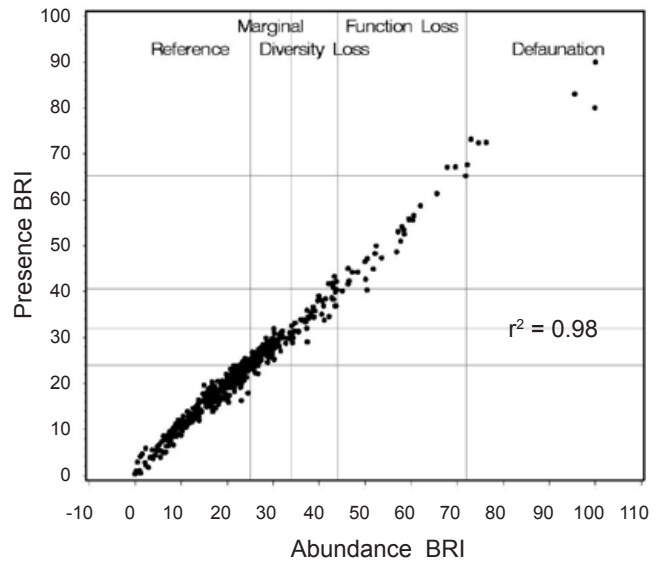
To assess how large the barcode library must be for an effective index, changes in the  $r^2$  between the BRI-P and the BRI were determined as taxa were systematically removed from calculation of the BRI-P. To accomplish this, taxa were ranked according to their p-values and taxa were selected for removal at even intervals. This process was repeated with increasing percentages of taxa removed from the calculation.

## RESULTS

There was a strong linear relationship between the BRI-P and BRI (Figure 1) with an  $r^2$  of 0.98 ( $p < 0.0001$ ). Application of the BRI-P assessment thresholds calculated by applying the equation from the calibration linear regression analysis (Table 1) resulted in 92% of the 493 calibration samples assigned to the same assessment category by both indices. Where assessment categories disagreed, the samples were always in adjacent categories, with no samples disagreeing by more than one category. The high agreement reflected the strong linear relationship in Figure 1.

Application of the BRI-P to the validation data resulted in patterns almost identical to the original abundance BRI results for both validation data sets. For the Los Angeles County outfall temporal data (Figure 2), the BRI-P accurately reflected the severe impacts at Los Angeles County Station 6C in the early 1970's, as well as the substantial improvement that occurred over time. The BRI-P also correctly identified Station 0C, located 14.7 km from the outfall, as less affected than Station 6C with only slight biodiversity loss in the early 1970's and improvement to reference conditions in the 1990's.

The BRI-P also correctly identified the spatial patterns near the Orange County outfall (Figure 3), with Station 0 located adjacent to the outfall having poorer benthic condition relative to Station



**Figure 1. Relationship between Presence BRI (BRI-P) and the original abundance BRI. Vertical and horizontal lines indicate assessment thresholds for the BRI (Smith *et al.* 2001) and BRI-P (Table 1), respectively.**

Con, which is located 7.8 km away. The BRI-P also retained similar among-station differences across all seasons, as was observed in the original BRI validation.

Reducing the number of species included in BRI-P calculations resulted in an  $r^2$  of greater than 0.8 relative to the abundance BRI when 20% of the species with tolerance scores were removed (Table 2). The  $r^2$  remained above 0.7 even when as many as 50% of the species were removed, though at that level of removal a number of samples included no taxa with tolerance scores, preventing BRI-P calculation.

## DISCUSSION

The benthic community gradients used to validate the original BRI were reproduced well by the BRI-P, with an  $r^2$  between the two indices of 0.98. There were minor reductions in classification efficiency using presence-only information, but no changes in the patterns or magnitudes of differences among sites. The reductions in index performance were also small relative to previous attempts to reduce processing cost by identifying taxa only to genus or family level (Ferraro and Cole 1995, Dethier and Schoch 2006, Bevilacqua *et al.* 2009). In contrast, barcoding provides the potential for improving index resolution by identifying cryptic species and clarifying species complexes that are

**Table 1. Presence BRI assessment categories, assessment thresholds, and percentage of samples in agreement with abundance BRI assessment category in the 493 sample calibration data.**

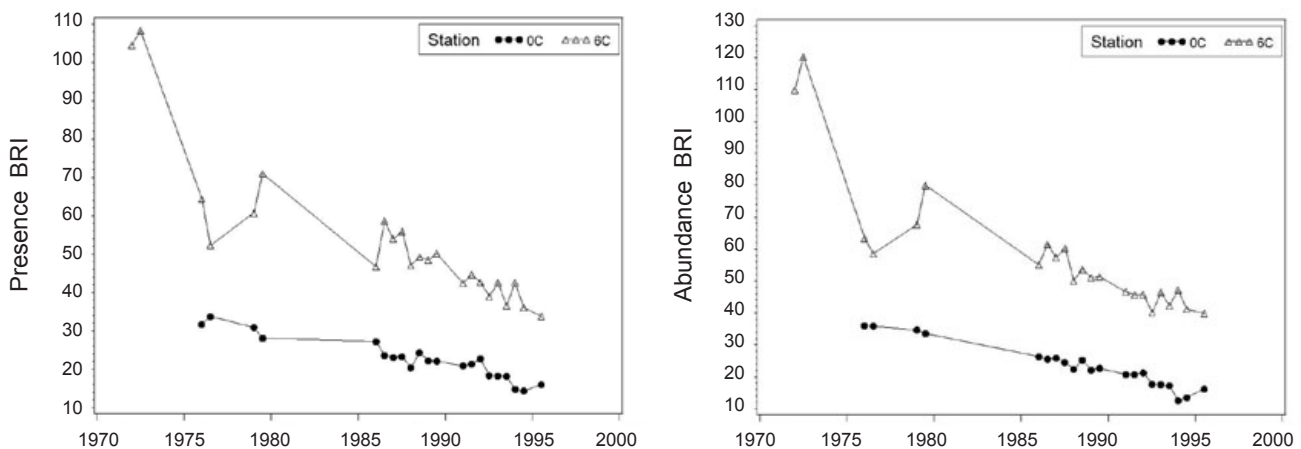
| Assessment Category     | BRI-P                  | Samples    | Samples in same BRI category |             |
|-------------------------|------------------------|------------|------------------------------|-------------|
|                         |                        |            | n                            | %           |
| Reference               | $\leq 24$              | 304        | 298                          | 98          |
| Marginal                | $>24$ to $\leq 32$     | 114        | 92                           | 80.7        |
| Diversity Loss          | $>32$ to $\leq 40.7$   | 34         | 31                           | 91.2        |
| Community Function Loss | $>40.7$ to $\leq 65.3$ | 32         | 25                           | 78.1        |
| Defaunation             | $\geq 65.3$            | 9          | 7                            | 77.8        |
| <b>Total</b>            |                        | <b>493</b> | <b>453</b>                   | <b>91.9</b> |

inseparable by traditional morphological taxonomy methods (Hajibabaei *et al.* 2007, Ratnasingham and Hebert 2007).

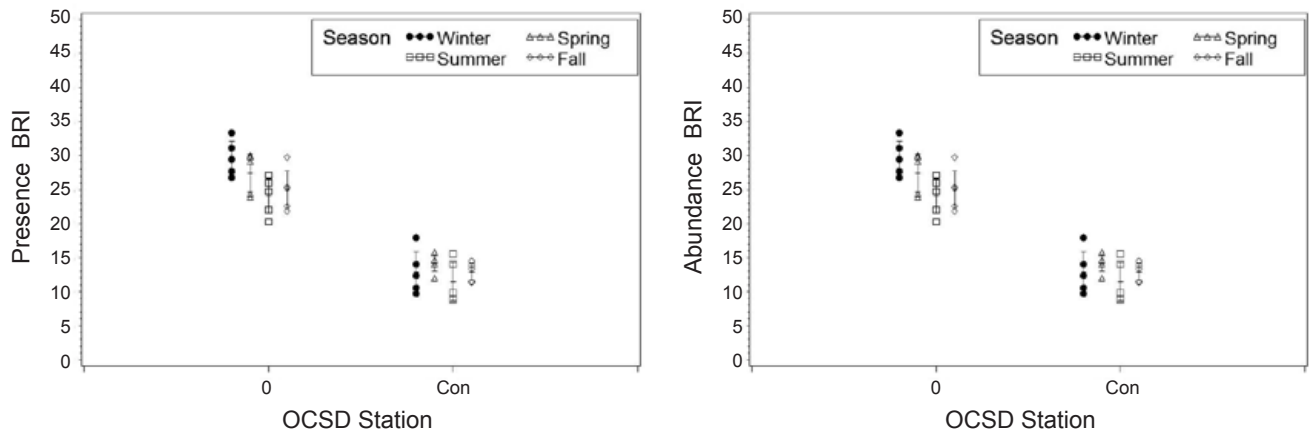
Similarity in performance between the BRI-P and the traditional BRI results because the index relies on composition of the whole community, not just the dominant taxa. Smith *et al.* (2001) found that even removing the top ten most abundant taxa had minimal effect on the performance of the abundance BRI. Reliance on the entire community is enhanced in the BRI by use of a cube-root transformation which lessens the influence of abundant species, selected by Smith *et al.* (2001) using an optimization algorithm to maximize index performance. This is

similar to the findings of Warwick *et al.* (2010) and Teixeira *et al.* (2012) who found that performance of the AZTI Marine Biological Index (AMBI; Borja *et al.* 2000), a European analog to the BRI that is also based on average pollution tolerance of organisms in the sample, is enhanced when including transformations that reduce the influence of dominant taxa.

While the presence based approach worked well for the BRI, it is unclear whether it will be equally successful in all situations. For instance, this study was conducted in euhaline water where there was an average of 67 taxa per sample; the presence-only index may be less sensitive in oligohaline waters



**Figure 2. Application of the Presence BRI and BRI to outfall monitoring data from 1972 to 1995 at Los Angeles County Station 6C (2.2 km from the outfall) and Station 0C (14.7 km from the outfall).**



**Figure 3. Application of the Presence BRI and BRI to 1990 seasonal outfall monitoring data at Orange County Station 0 (adjacent to the outfall) and Station Con (8 km from the outfall).**

where index development is more challenging because there are typically fewer than ten taxa per sample, even at reference sites (Dauvin and Ruellet 2009, Thompson *et al.* In Press). It might also not work as well in locations with more severe benthic community effects, such as in Chesapeake Bay where hypoxia leads to substantial reductions in benthic abundance. The disturbance gradient effects in this study were mostly limited to replacement of pollution sensitive with pollution tolerant taxa and indices that rely more heavily on dominance and diversity measures, which require abundance information, may be necessary to capture the larger range of effects. Still, species sensitivity and tolerance to disturbance are more robust measures of benthic

community condition than abundance, diversity, or other community measures in multimetric indices (Weisberg *et al.* 1997, Alden *et al.* 2002, Borja *et al.* 2008).

There are many aspects of DNA barcoding that need to be investigated before it can be adopted for biological assessments, such as improved understanding of intraspecific genetic variation within groups currently considered as a single species, and sample preservation and handling methods. However, the lack of quantification in species identification does not seem to be an impediment nor does creation of a locally relevant DNA barcode reference library. Our results indicate

**Table 2. Effect of reductions in numbers of taxa. BRI-P values could not be calculated for stations where no species with tolerance scores were present in the data after the reduction.**

| Percentage of species with tolerance scores included in BRI-P | Number of species with tolerance scores included in BRI-P | BRI-P vs. BRI Coefficient of determination ( $r^2$ ) | Number of stations with BRI-P values |
|---|---|--|--------------------------------------|
| 10  | 46  | 0.44   | 380                                  |
| 20  | 92  | 0.57   | 457                                  |
| 30  | 139   | 0.63   | 488                                  |
| 40  | 185   | 0.68   | 490                                  |
| 50  | 231   | 0.80   | 493                                  |
| 60  | 277   | 0.75   | 493                                  |
| 70  | 323   | 0.77   | 492                                  |
| 80  | 370   | 0.81   | 493                                  |
| 90  | 416   | 0.84   | 493                                  |
| 100   | 462   | 0.98   | 493                                  |

that acceptable results are likely with reference barcodes for less than 400 taxa. The highest priority is to collect specimens for the library that have known tolerance values and which are influential on the BRI score. Currently approximately 16% of the 520 taxa with tolerance scores are in the reference library and efforts are currently underway along the Pacific Coast to expand this number. Developing the reference library is a relatively achievable goal and, combined with the encouraging results for the BRI-P, represent progress towards integrating molecular methods into routine bioassessment.

## LITERATURE CITED

- Alden, R.W., III, D.M. Dauer, J.A. Ranasinghe, L.C. Scott and R.J. Llansó. 2002. Statistical verification of the Chesapeake Bay benthic index of biotic integrity. *Environmetrics* 13:473-498.
- Bevilacqua, S, S. Frascchetti, L Musco and A Terlizzi. 2009. Taxonomic sufficiency in the detection of natural and human-induced changes in marine assemblages: A comparison of habitats and taxonomic groups. *Marine Pollution Bulletin* 58:1850-1859.
- Borisenko, A.V., J.E. Sones and P.D.N. Hebert. 2009. The front-end logistics of DNA barcoding: challenges and prospects. *Molecular Ecology Resources* 9:27-34.
- Borja, Á., D.M. Dauer, R. Diaz, R.J. Llansó, I. Muxika, J.G. Rodriguez and L. Schaffner. 2008. Assessing estuarine benthic quality conditions in Chesapeake Bay: A comparison of three indices. *Ecological Indicators* 8:395-403.
- Borja, Á., J. Franco and V. Perez. 2000. A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. *Marine Pollution Bulletin* 40:1100-1114.
- Borja, Á., J.A. Ranasinghe and S.B. Weisberg. 2009. Assessing ecological integrity in marine waters using multiple indices and ecosystem components: Challenges for the future. *Marine Pollution Bulletin* 59:1-4.
- County Sanitation Districts of Orange County (CSDOC). 1991. 1991 Annual Report. County Sanitation Districts of Orange County. Fountain Valley, CA.
- Dauvin, J.C. and T. Ruellet. 2009. The estuarine quality paradox: Is it possible to define an ecological quality status for specific modified and naturally stressed estuarine ecosystems? *Marine Pollution Bulletin* 59:38-47.
- Dethier, M.N. and G.C. Schoch. 2006. Taxonomic sufficiency in distinguishing natural spatial patterns on an estuarine shoreline. *Marine Ecology-Progress Series* 306:41-49.
- Diaz, R.J., M. Solan and R.M. Valente. 2004. A review of approaches for classifying benthic habitats and evaluating habitat quality. *Journal of Environmental Management* 73:165-181.
- Ferraro, S.P. and F.A. Cole. 1995. Taxonomic level sufficient for assessing pollution impacts on the southern California bight macrobenthos - revisited. *Environmental Toxicology and Chemistry* 14:1031-1040.
- Hajibabaei M., S. Shokralla, X. Zhou, G.A.C. Singer and D.J. Baird. 2011. Environmental Barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE* 6(4): e17497. doi:10.1371/journal.pone.0017497.
- Hajibabaei, M., G.A.C. Singer, P.D.N. Hebert and D.A. Hickey. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics* 23:167-172.
- Hebert, P.D.N., A. Cywinska, S.L. Ball and J.R. DeWaard. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences* 270:313-321.
- Janzen, D.H., W. Hallwachs, P. Blandin, J.M. Burns, J.M. Cadiou, I. Chacon, T. Dapkey, A.R. Deans, M.E. Epstein, B. Espinoza, J.G. Franclemont, W.A. Haber, M. Hajibabaei, J.P.W. Hall, P.D.N. Hebert, I.D. Gauld, D.J. Harvey, A. Hausmann, I.J. Kitching, D. Lafontaine, J.F. Landry, C. Lemaire, J.Y. Miller, J.S. Miller, L. Miller, S.E. Miller, J. Montero, E. Munroe, S.R. Green, S. Ratnasingham, J.E. Rawlins, R.K. Robbins, J.J. Rodriguez, R. Rougerie, M.J. Sharkey, M.A. Smith, M.A. Solis, J.B. Sullivan, P. Thiaucourt, D.B. Wahl, S.J. Weller, J.B. Whitfield, K.R. Willmott, D.M. Wood, N.E. Woodley and J.J. Wilson. 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources* 9:1-26.

- Martínez-Crego, B., T. Alcoverroa, and J. Romero. 2010. Biotic indices for assessing the status of coastal waters: a review of strengths and weaknesses. *Journal of Environmental Monitoring* 12:1013–1028.
- Pinto, R.I.C., J. Patricio, A. Baeta, B.D. Fath, J.M. Neto and J.C. Marques. 2009. Review and evaluation of estuarine biotic indices to assess benthic condition. *Ecological Indicators* 9:1-25.
- Ranasinghe, J.A., D.E. Montagne, S.B. Weisberg, M. Bergen, and R.G. Velarde. 2003. Variability in the identification and enumeration of marine benthic invertebrate samples and its effect on benthic assessment measures. *Environmental Monitoring and Assessment* 81:199-206.
- Ratnasingham, S. and P.D.N. Hebert. 2007. BOLD: The Barcode of Life Data System ([www.barcodinglife.org](http://www.barcodinglife.org)). *Molecular Ecology Notes* 7:355-364.
- Smith, R.W., M. Bergen, S.B. Weisberg, D.B. Cadien, A. Dalkey, D.E. Montagne, J.K. Stull and R.G. Velarde. 2001. Benthic response index for assessing infaunal communities on the southern California mainland shelf. *Ecological Applications* 11:1073-1087.
- Stoeckle, M.Y. and P.D.N. Hebert. 2008. Barcode of life. *Scientific American* 299:82-88.
- Stull, J.K., C.I. Haydock, R.W. Smith and D.E. Montagne. 1986. Long-term changes in the benthic community on the coastal shelf of Palos Verdes, Southern California. *Marine Biology* 91:539-551.
- Stull, J.K. 1995. Two decades of marine biological monitoring, Palos Verdes, California, 1972 to 1992. *Bulletin of the Southern California Academy of Sciences* 94:21-45.
- Teixeira, H., S.B. Weisberg, A. Borja, J.A. Ranasinghe, D.B. Cadien, R.G. Velarde, L.L. Lovell, D. Pasko, C.A. Phillips, D.E. Montagne, K.J. Ritter, F. Salas and J.C. Marques. 2012. Calibration and validation of the AZTI's Marine Biotic Index (AMBI) for Southern California marine bays. *Ecological Indicators* 12:84-95.
- Thompson, B., S.B. Weisberg, A.R. Melwani, S. Lowe, J.A. Ranasinghe, D.B. Cadien, D.M. Dauer, R.J. Diaz, W. Fields, M. Kellogg, D.E. Montagne, P.R. Ode, D.J. Reish and P.N. Slattery. In Press. Low levels of agreement among experts using best professional judgment to assess benthic condition in the San Francisco estuary and delta. *Ecological Indicators*.
- Warwick, R.M., K.R. Clarke and P.J. Somerfield. 2010. Exploring the marine biotic index (AMBI): variations on a theme by Ángel Borja. *Marine Pollution Bulletin* 60:554-559.
- Weisberg, S.B., J.A. Ranasinghe, L.C. Schaffner, R.J. Diaz, D.M. Dauer and J.B. Frithsen. 1997. An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. *Estuaries* 20:149-158.