Taxonomic structure and the effectiveness of surrogates in environmental monitoring: a lesson from polychaetes

Luigi Musco*, Antonio Terlizzi, Margherita Licciano, Adriana Giangrande

Laboratorio di Zoologia e Biologia Marina, CoNISMa, DiSTeBA, Università del Salento, Via Lecce-Monteroni, 73100 Lecce, Italy

ABSTRACT: Multivariate approaches for the quantification of changes in biological assemblages can provide a powerful tool for assessing human impacts on ecological systems. Taxonomic sufficiency (TS) has been proposed as a short-cut method in environmental monitoring. However, issues about the efficacy of taxonomic surrogates in depicting spatial patterns and taxonomic structure of marine assemblages are still a source of debate. Here, we followed the logic of beyond-BACI (before, after, control, impact) designs and their modifications to deal with cases in which no data have been collected before the purported impact. By using, as a case study, the analysis of the effects of sewage pollution on spatial patterns of polychaete assemblages along a stretch of rocky coast in the Ionian Sea (south Italy), this study aimed to (1) investigate whether the assemblage's taxonomic structure influences the results of the analyses based on coarser taxonomy, and (2) test the efficiency of an alternative cost-effective method to TS. The analyses highlighted the impact as a modification of natural spatial patterns of assemblages at the scale of location, site and replicate units. Results were consistent when analysing data at species and genus level. The use of other surrogates failed in fully depicting modifications of natural spatial patterns as a consequence of the impact. The surrogate performance depended on the assemblage's taxonomic structure, and was related to the information provided by the mono-specific higher taxa (i.e. species distribution). Differences among sites at the impacted location were underestimated after lumping species distributions together. Our data suggest that the analysis at species level of a correctly selected taxon (i.e. a single family) can represent a valid cost-effective alternative to using all species-based data sets. The results stress that only an *a priori* taxonomic screening may address the correct approach for depicting changes in assemblages in relation to human impacts.

KEY WORDS: Taxonomic sufficiency · Species redundancy · Spatial scales · Sewage pollution · Polychaeta · Syllidae · Asymmetrical designs

INTRODUCTION

It is generally acknowledged that whole biological assemblages are more sensitive to environmental changes than individual indicator species or other univariate indices, and that measures of assemblage patterns can provide the most useful information for interpreting the consequences of human disturbances onto ecological systems (Underwood & Peterson 1988). New statistical tools have recently been developed for the analysis of multispecies responses in structured ecological experiments (Anderson 2001). Describing assemblages at high levels of taxonomic resolution (species) is logistically difficult, requiring the counting of all organisms and their taxonomic identification. Accurate taxonomic analysis is often unachievable due to the costs in processing samples and, above all, the lack of taxonomic expertise. As a general rule, the coarser the taxonomic resolution, the cheaper the research cost (Ferraro & Cole 1995).

Taxonomic sufficiency (TS) (Ellis 1985) implies 'identifying organisms only to a level of taxonomic resolution sufficient to satisfy the objective of a study' (Bertrand et al. 2006) and has been proposed as the ideal short-cut procedure in community perturbation studies. TS is based on evidence that abundance and diversity of species are related to abundance and diversity of the same organisms identified to supra-specific taxonomic levels (genus, family, order, etc.) (e.g. Balmford et al. 1996).

Following Somerfield & Clarke (1995), if the pattern of community change is marked, interpretable results are possible independent of both the examined component of the benthos and the taxonomic level at which the analyses are carried out. However, the extent to which aggregation to higher taxonomic levels affects the results depends on the distribution of species amongst higher taxa in the original samples.

Current evidence suggests that family is a sufficient taxonomic level, but this statement stems from too limited a number of case studies, and an array of possible biases in TS results have not been fully evaluated. In particular, the relationships among TS, spatial scale, habitat features and assemblage structure are still too ambiguous to make generalisations (Terlizzi et al. 2003).

Higher taxa may be efficient surrogates for species in describing spatial patterns, and their use may prevent problems due to high variability of species distribution, especially at small spatial scales (Chapman 1998, Anderson et al. 2005). However, the effectiveness of taxonomic surrogates for analyses of variation at different spatial scales warrants further study (Anderson et al. 2005, Terlizzi et al. 2009).

Another not yet properly addressed limitation of the TS approach is the consideration of processes of sympatric speciation, which occur frequently in the marine environment (Hellberg 1998). Such processes lead to the presence of great numbers of species within some genera (or families) as a result of a high intra-area radiation. Sympatric speciation is often reduced or absent in habitats of short evolutionary time (e.g. brackish waters or other stressed environments) (Giangrande et al. 2005). Patterns of biodiversity and taxonomic relatedness of species within given taxa are therefore context-dependent, an issue which could limit generalisations about the efficiency of a TS approach.

Bertrand et al. (2006) criticised the use of taxonomic surrogates of species in palaeontology, environmental monitoring and conservation studies. They suggested that results of TS-based analyses depend on the level of taxonomic knowledge for any investigated group that, in its turn, represents a 250 yr heterogeneous mixture of ideas. Taxonomy in fact evolves, and the systematic position and status of species (and even higher taxa) continuously change after revisions (e.g. Giangrande & Licciano 2008). As a consequence, using different classifications of the same organisms, both valid according to different plausible phylogenetic hypotheses, may result in very different conclusions on the diversity of the studied assemblages. As pointed out by Bertrand et al. (2006), several authors suggested that the treatment of the same taxonomic ranks as equivalent units might allow for spurious comparisons.

As an alternative to TS, the counting and identification at species level of a subset of the whole assemblage (e.g. a class or a family) has been proposed as a cost-effective method (e.g. Olsgard et al. 2003, Giangrande et al. 2004, 2005). However, some authors suggest caution in the use of subgroups of indicator species (e.g. Anderson et al. 2005, Bertrand et al. 2006). Moreover, it has been suggested that this method should be planned after a preliminary faunal survey of the study area (Włodarska-Kowalczuk & Kędra 2007).

Here, we explored the TS approach using polychaetes as an indicator of the effects of human impact. For the analyses, we followed the logic of beyond-BACI (before, after, control, impact) designs (Underwood 1991, 1994) and their modifications to deal with cases in which no data have been collected before the purported impact (Glasby 1997, Terlizzi et al. 2005a). We focused on a sewage outfall as an example of human disturbance, extending the information already available in the literature about the impact of the same source of disturbance on other components of the biota such as sessile macrobenthos, some meiofaunal components, fish and mollusc assemblages (Terlizzi et al. 2002, 2005b, Guidetti et al. 2003, Fraschetti et al. 2006). Specifically, the present study aimed to (1) establish whether and how the taxonomic structure of the assemblage may influence the results of TS-based analyses, and (2) verify the descriptive efficiency of the species belonging to the most representative taxonomic subsets of Polychaeta in the analysed assemblage.

Finally, in order to provide further generalisations about the concepts of redundancy and sufficiency of taxa, we re-analysed the descriptive efficiency of the best-performing taxonomic surrogate (i.e. genus) by excluding variables consisting of monospecific genera and considering only those variables effectively derived by lumping species counts together.

MATERIALS AND METHODS

Study area and experimental design. The study area is located along the southwestern coast of Apulia (SE Italy, Ionian Sea) and is characterised by wave-exposed calcarenitic rocky plateaus extending from the water surface to about 10 m depth on fine sand

with a gentle-medium slope (Terlizzi et al. 2002). Tidal range in the area is in the order of a few cm. Water temperature ranges from 13.4°C in February (mean annual minimum) to 24.8°C in August (mean annual maximum). Sampling was undertaken at the beginning of November 2002 at the outfall location (hereinafter indicated as I) and 2 control or reference locations (C1 and C2, hereinafter indicated as Cs). Following the logic of beyond-BACI designs, multiple controls were adopted in order to quantify patterns of natural variability and tease them apart from the variability induced by the source of disturbance (i.e. the outfall) as evidence of impact. Controls were chosen at random from a set of possible locations separated by at least 3 km and providing comparable environmental conditions to those occurring at the outfall in terms of slope, water temperature (about 19.5°C during the sampling period), wave exposure and type of substrate. They were also chosen to be located on either side of the outfall (Fig. 1). In general, Location I is characterised by complex algal assemblages dominated by *Colpomenia sinuosa*, *Gelidium* spp., *Pterocladiella* spp. and filamentous brown and green algae. At both control locations, the sessile assemblages are mainly dominated by the algae *Dictyota dichothoma*, *Padina pavonica*, *Acetabularia acetabulum* and *Cystoseira* spp. Specific information about the type and volumes of discharged waste waters as well as a more detailed account of spatio-temporal patterns of abundance and species composition of sessile assemblages in relation to the presence of the outfall is reported elsewhere (Terlizzi et al. 2005a). At each of the 3 locations, 3 sites separated by 80 to 100 m were chosen. Sampling in sites within locations allowed an evaluation, at a lower scale, of spatial patterns of variation in natural (i.e. sites within Location Cs) versus disturbed conditions (i.e. sites within Location I). At Location I, 1 site (Ib) was located immediately adjacent to the point of discharge and the other 2 were on its right and left, respectively. At all sites, assemblages were sampled at a depth of 3 to 4 m on sloping rocky surfaces in 9 replicates, yielding a total of 81 units of observation.

Sampling procedure and taxonomic discrimination. In order to limit the loss of small, cryptic and motile specimens, sampling was obtained by SCUBA diving, using an airlift sampler while scraping off the organisms within a 20 × 20 cm metal frame placed on the rocky surface. The material was collected into a 0.4 mm mesh bag. After collection, samples were fixed in 4% buffered formalin solution. Polychaetes were then sorted under magnification, preserved in 70% alcohol and identified to species level. Taxonomy followed Rouse & Pleijel (2001).

Statistical analyses. ANOVA was used to compare the total number of individuals, species, genera and
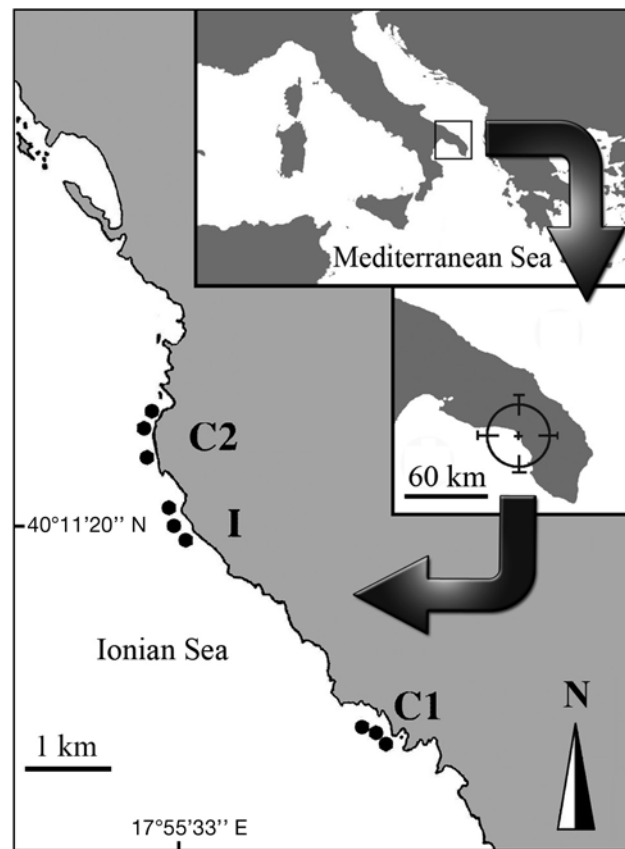

Fig. 1. Study area and sampling locations. I = location exposed to sewage discharges; C1, C2 = control locations; ● = sampling site

families and the abundance of the dominant species at I-vs.-Cs. For all analyses, the model consisted of 2 factors: Location (L) (1 disturbed and 2 control locations) and Site (S) (3 levels, random, nested in Location), with n = 9 replicates. Given the presence of a single purportedly impacted location, this is an asymmetrical design (sensu Underwood 1991, 1994). Thus, for the analyses, in order to test for the main effect of I-vs.-Cs, the Location term was partitioned into 2 portions: the 1 degree-of-freedom contrast I-vs.-Cs and the variability between Cs. The term Site(Location) (S(L)) was similarly divided into S(I) and S(Cs). Finally, the residual variation was divided into 2 components: the residual variability for observations within Location I (Res I), and the residual variability for observations within Location Cs (Res Cs). Appropriate denominators for F-ratios were identified following the logic of beyond-BACI designs and their modifications to deal with cases in which no data before the occurrence of the disturbance are provided (see particularly Glasby 1997, Terlizzi et al. 2005a). ANOVAs were done using the GMAV 5 computer program (University of Sydney, Australia).

Permutational multivariate ANOVA (PERMANOVA) based on Bray-Curtis dissimilarities on untransformed data was used to estimate differences in the structure of multivariate assemblages between the Locations I and Cs, following, in a multivariate context, the same design described in the previous paragraph. PERMANOVAs were also used to test differences in the structure of multivariate assemblages within Location I between Site Ib (the site closest to the outfall) and the other 2 adjacent Sites Ia and Ic (Bishop et al. 2002). Each term in the analysis was tested using 4999 random permutations of the appropriate units (Anderson & ter Braak 2002). When the number of possible permutable units was not enough to get a reasonable test by permutation, a p-value was obtained using a Monte Carlo random sample from the asymptotic permutation distribution (Anderson & Robinson 2003). The analyses were done using the computer programs DISTLM.exe and PERMANOVA. exe (available at www.stat.auckland. ac.nz/~mja/Programs.htm).

Multivariate patterns of assemblages were visualised by non-metric multidimensional scaling (NMDS) of the sites' centroids in Bray-Curtis space. The calculation of centroids and the NMDS plotting were done using the routines contained in the PRIMER v6 computer program (Clarke & Gorley 2006, Anderson et al. 2008).

A regression analysis was also performed in order to test the correlation between the number of species and the number of genera (or families) per replicate.

RESULTS

A total of 5908 specimens belonging to 19 families, 59 genera and 106 species were collected. Syllidae (33% of the individuals), Nereididae (24%), Eunicidae (15%), Opheliidae (10%) and Sabellidae (6%) were the most important families in terms of abundance of individuals (Fig. 2a). The majority (85%) of the total individuals belonged to 10 of the 59 genera (Fig. 2b). *Syllis* (29%), *Platynereis* (15%), *Lysidice* (10%) and *Polyophthalmus* (10%) were the most abundant genera. *Platynereis dumerilii*
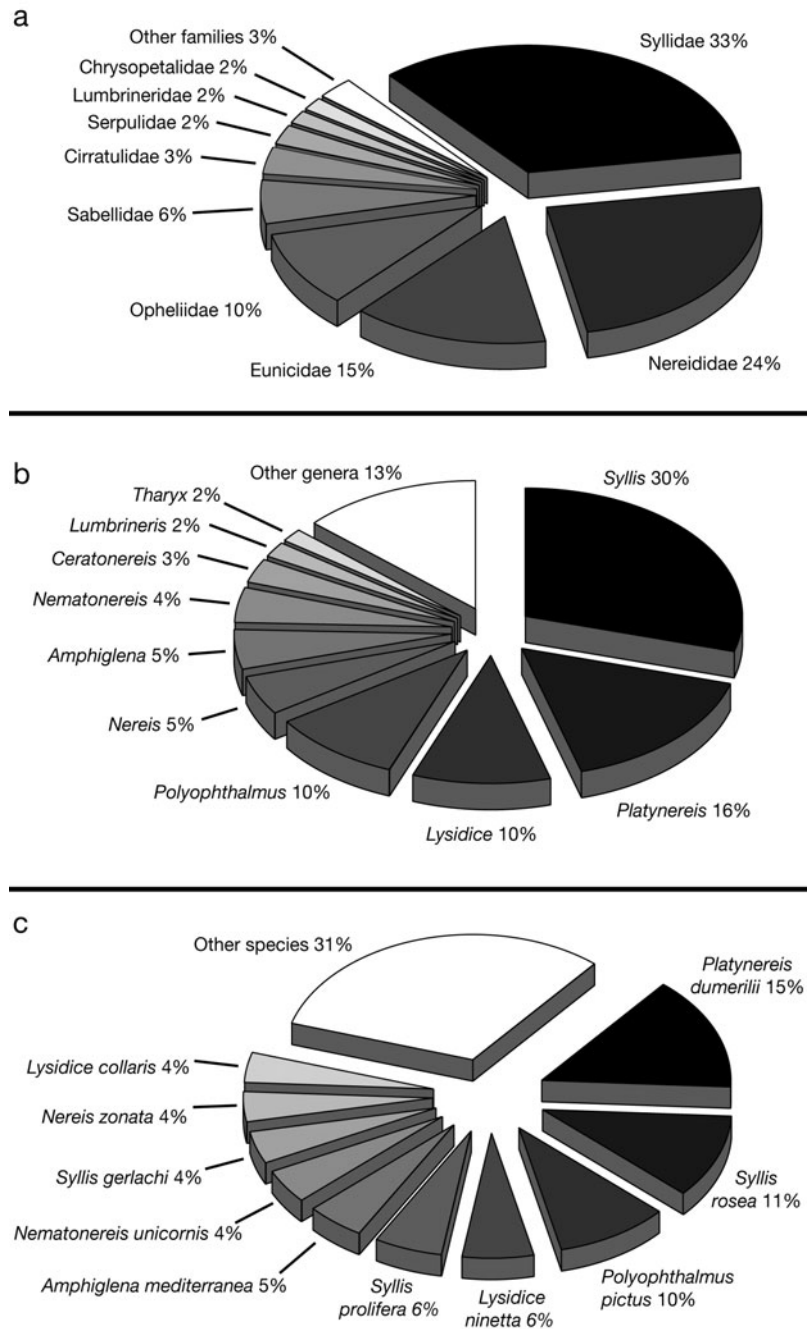

Fig. 2. Relative abundance (%) of the polychaete (a) families, (b) genera and (c) species sampled in the study area

(15%) ranked first among species followed by *S. rosea* and *Polyophthalmus pictus*; among the 10 most abundant species, 3 belonged to genus *Syllis* and 2 to *Lysidice* (Fig. 2c).

The asymmetrical ANOVAs indicated significant differences between I-vs.-Cs, in the mean numbers of species (Fig. 3b) and genera (Fig. 3c). There were no significant differences between the Locations I and Cs
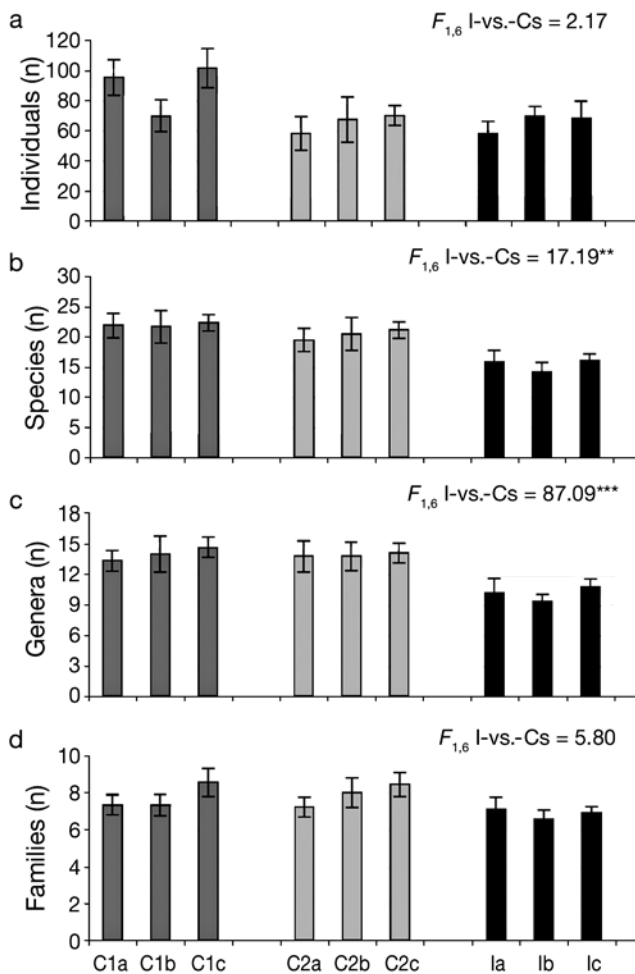
Fig. 3. Mean number (± SE, n = 9) of polychaete (a) individuals, (b) species, (c) genera and (d) families at the impacted (I) vs. control sites (C1, C2). For each graph, the output of the ANOVA *F*-test for the significances of the I-vs.-Cs term is provided (**p < 0.01, ***p < 0.001)

in the mean numbers of individuals (Fig. 3a) and families (Fig. 3d).

Patterns of mean abundances of the 10 dominant polychaete species at the 9 sampled sites and test results for their differences in abundance between the locations I and Cs are shown in Fig. 4. Opposite distribution patterns were found in the 2 most abundant species, namely *Platynereis dumerilii* and *Syllis rosea*, the first being significantly more abundant at Location I (particularly in Site Ib, the site closest to the outfall discharge), and the second being significantly more abundant, on average, in control locations, with the lowest values recorded at Site Ib. *Lysidice ninetta* was among the species showing the most striking difference in abundance between the locations I and Cs, being common at Cs and rather rare at I. *Polyophthalmus pictus* was among the most abundant species at Location I while its presence in Cs was occa-

sional. The 3 *Syllis* species (*S. rosea*, *S. gerlachi* and *S. prolifera*) showed idiosyncratic response to the presence of the outfall. Unlike *S. rosea*, the mean abundance of both *S. gerlachi* and *S. prolifera* did not differ, on average, between the locations I and Cs. The abundance of *S. gerlachi* was however particularly high at Site Ib, where the mean abundance of *S. prolifera* was particularly low (Fig. 4). The mean abundance of *Nereis zonata* did not differ significantly between I-vs.-Cs and its distribution pattern was similar to that observed for *S. prolifera*. Similar outcomes emerged from the analysis of the distribution of *L. collaris*, which differed from the congeneric *L. ninetta*.

Species vs. higher taxa correlation

There was a significant correlation between number of species per replicate and genera per replicate (Fig. 5a) ($R^2 = 0.8808$; $p < 0.001$) (mean ± SD species: genus ratio per replicate = 1.52 ± 0.19). Considering only syllids (Fig. 5b), the same correlation (species–genera) was weaker ($R^2 = 0.6014$; $p < 0.001$) (mean ± SD species:genus ratio per replicate = 3.36 ± 1.61) and, as a consequence, the correlation species–genera of the remaining assemblage (Fig. 5c) (polychaetes excluding Syllidae) was stronger ($R^2 = 0.9341$; $p < 0.001$) (mean ± SD species:genus ratio per replicate = 1.12 ± 0.11).

The correlation species-families (Fig. 6a) was weak ($R^2 = 0.5670$; $p < 0.001$) (mean ± SD species:family ratio per replicate = 2.57 ± 0.59). The same correlation (species-families) was stronger when the Syllidae were excluded from the analysis (Fig. 6b) ($R^2 = 0.6367$; $p < 0.001$) (mean ± SD species:family ratio per replicate = 1.72 ± 0.42).

Multivariate analyses

PERMANOVAs investigating the full spatial design for the whole assemblage at species, genus and family level of taxonomic resolution revealed the significance of the I-vs.-Cs contrast term, indicating that, independent of the taxonomic level adopted, the assemblage at the impact location differed significantly from those observed, on average, at the control locations (Table 1). With the exception of the I location in the analysis of polychaetes at the family level, significant differences were also observed among sites within both controls and the I location (i.e. the significance of the terms S(Cs) and S(I) in Table 1). The outcome of the PERMANOVA was also well-depicted by the NMDS plots (Fig. 7), which showed that, although the magnitude of the difference between I-vs.-Cs might differ at
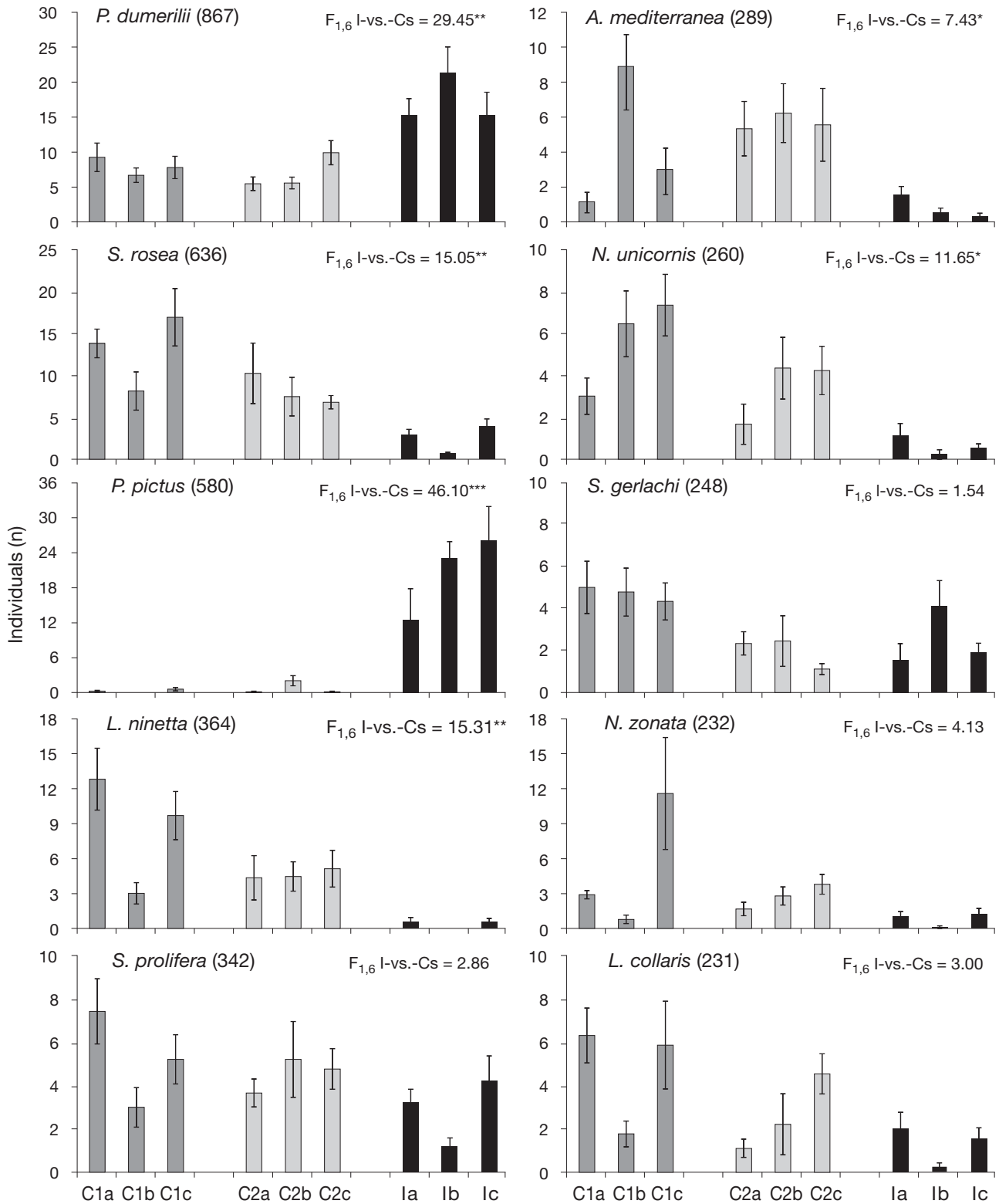
Fig. 4. Mean number of individuals (± SE, n = 9) of the 10 most abundant polychaete species — (a) *Platynereis dumerilii*, (b) *Syllis rosea*, (c) *Polyophthalmus pictus*, (d) *Lysidice ninetta*, (e) *S. prolifera*, (f) *Amphiglena mediterranea*, (g) *Nematonereis unicornis*, (h) *S. gerlachi*, (i) *Nereis zonata* and (j) *L. collaris* — at the impacted (I) vs. control sites (C1, C2). For each species, the total number of individuals is given in brackets. For each graph, the output of the ANOVA *F*-test for the significances of the I-vs.-Cs term is provided (*p < 0.05; **p < 0.01; ***p < 0.001)
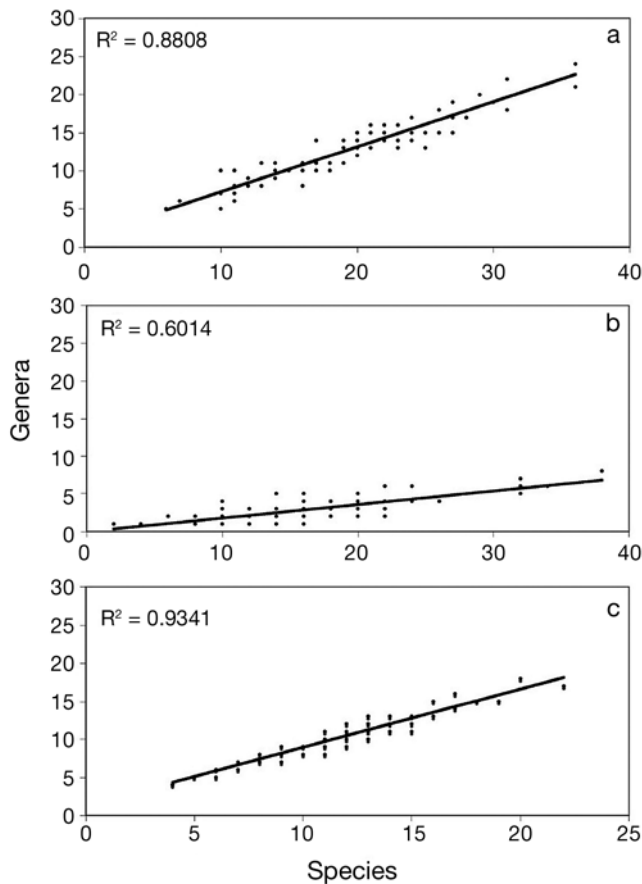
Fig. 5. Regression analysis between species and genera of (a) the whole polychaete assemblage, (b) only Syllidae, and (c) the whole polychaete assemblage excluding Syllidae

Fig. 6. Regression analysis between species and families for (a) the whole polychaete assemblage, (b) polychaetes excluding Syllidae

different levels of taxonomic resolution, there was always consistency in the direction of this difference, i.e. the assemblage composition at Location I were distinct from the Cs ones at species (Fig. 7a), genus (Fig. 7b) and family (Fig. 7c) levels.

The significance of the I-vs.-Cs contrast term also resulted from the PERMANOVAs investigating the full spatial design for the assemblage species subsets of Eunicida (Arabellidae + Dorvilleidae + Eunicidae + Lumbrineridae, 14 species), Sabellida (Sabellidae + Serpulidae, 18 species) and both Syllidae species (49) and genera (18) (Table 1). There were no significant differences in the multivariate structure of the assemblages between Cs locations. No clear evidence of differences between I-vs.-Cs in patterns of site-to-site variation emerged from the analyses of data subsets. With an exception for the analyses of syllid genera where both the S(I) and the S(Cs) terms were not significant, differences among sites were always significant in both Cs and I conditions.

PERMANOVAs investigating the reduced spatial design (Table 2) for Location I (for simplicity, Site Ib is
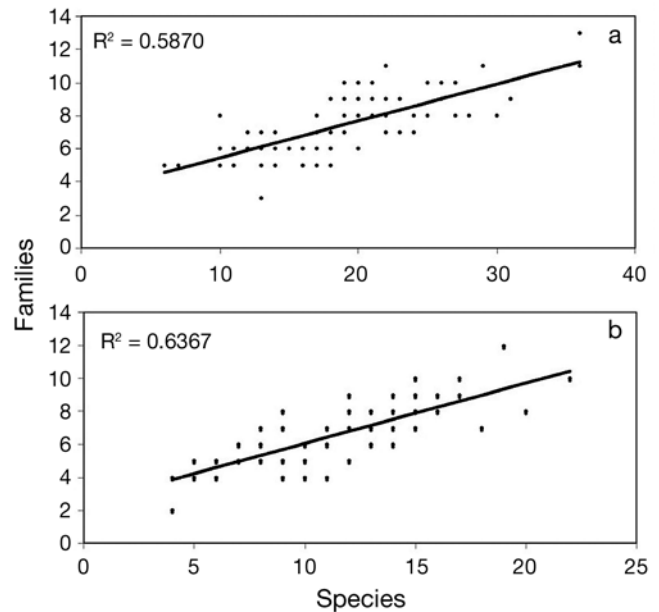
renamed I' and Sites Ia and Ic are renamed Cs') showed that, whatever the taxonomic level or the assemblage subset adopted, the Cs' sites were not significantly different. Instead, the I'-vs.-Cs' contrast term was significant for the whole assemblage identified at species and genus level and for the species of Eunicida, Sabellida and Syllidae. This indicates that the site closest to the outfall (I') differed significantly from the other 2 located 100 m distant from the point of sewage discharge. The I'-vs.-Cs' term was not significant for the whole assemblage identified at family level and for the syllids identified at genus level.

A posteriori analysis of the taxonomic surrogate (genus) efficiency

The analysis of the assemblage's taxonomic structure revealed that 41 of the 59 collected genera were locally represented by only 1 species (hereinafter named mono-specific genera). Table 3 shows the results of the PERMANOVAs investigating the pattern of spatial distribution of the remaining 18 genera recorded. These 18 variables (taxa) are the ones effectively derived from the lumping together of 2 or more variables (species distributions) of the original dataset. For the full spatial design (Table 3), the structure of these *a posteriori* made polychaete assemblages differed between I-vs.-Cs while no differences between Cs was observed. Considering the reduced spatial design (Table 3), the terms Cs' and I'-vs.-Cs' were both not significant.

Table 1. PERMANOVAs testing differences in the spatial distribution of the analysed assemblages (all polychaetes classified at different taxonomic levels, only Eunicida and Sabellida species, and only Syllidae species and genera) at the impacted (I) vs. control locations (Cs). For each data set, the number of analysed taxa is indicated within parentheses. Analyses based on Bray-Curtis dissimilarity. Each test was done using 4999 permutations of appropriate units, p-values given in italics were obtained using 4999 Monte Carlo samples from the asymptotic permutation distribution. The term used for the denominator mean square in each case is given in column MS_{denom}. L = location, S = site

Source of variation	df	Polychaete species (106) MS	F	p_{perm}	Polychaete genera (59) MS	F	p_{perm}	Polychaete families (19) MS	F	p_{perm}	Syllidae genera (18) MS	F	p_{perm}	MS_{denom}	Permutable units
L	2	24558.53	6.858	*0.0002*	20219.18	7.396	*0.0002*	14581.81	7.219	*0.0002*	6416.21	4.576	*0.0052*	S(L)	9
Cs	1	7366.67	1.820	*0.0702*	5631.82	1.808	*0.1090*	4978.50	2.164	*0.1052*	3595.38	2.204	*0.1326*	S(Cs)	6
I-vs.-Cs	1	41750.39	11.659	*0.0002*	34806.53	12.732	*0.0002*	24185.11	11.973	*0.0002*	9237.03	8.888	*0.0006*	S(L)	9
S(L)	6	3580.96	2.441	*0.0002*	2733.68	2.459	*0.0002*	2019.95	2.333	*0.0006*	1402.25	1.349	*0.1548*	res	81
S(Cs)	4	4047.64	2.662	*0.0002*	3114.75	2.762	*0.0002*	2300.77	2.744	*0.0018*	1630.99	1.540	*0.1050*	resCs	54
S(I)	2	2647.60	1.948	*0.0148*	1971.56	1.827	*0.0396*	1458.31	1.584	*0.1014*	944.79	0.946	*0.4412*	resI	27
Residuals	72	1466.82			1111.60			865.84			1039.28				
resCs	48	1520.70			1127.72			838.45			1059.43				
resI	24	1359.07			1079.36			920.61			998.99				

Source of variation	df	Eunicida species (14) MS	F	p_{perm}	Sabellida species (18) MS	F	p_{perm}	Syllidae species (49) MS	F	p_{perm}	MS_{denom}	Permutable units
L	2	19516.84	3.717	*0.0026*	11396.53	1.999	*0.0404*	13391.26	3.702	*0.0002*	S(L)	9
Cs	1	7701.43	2.004	*0.1108*	3192.95	0.613	*0.7100*	7647.49	2.249	*0.0610*	S(Cs)	6
I-vs.-Cs	1	31332.24	5.966	*0.0012*	19600.12	3.438	*0.0086*	19135.03	5.290	*0.0006*	S(L)	9
S(L)	6	5251.45	2.306	*0.0002*	5700.29	1.952	*0.0008*	3617.53	2.261	*0.0002*	res	81
S(Cs)	4	3843.15	2.174	*0.0016*	5208.04	1.926	*0.0070*	3400.80	2.364	*0.0006*	resCs	54
S(I)	2	8068.06	2.446	*0.0144*	6684.80	1.996	*0.0146*	4050.99	2.106	*0.0058*	resI	27
Residuals	72	2277.80			2919.64			1600.33				
resCs	48	1767.72			2704.71			1438.87				
resI	24	3297.96			3349.50			1923.23				

DISCUSSION

Polychaete assemblages had significantly different structures at a location exposed to sewage discharge in comparison to those at control locations. Observations were not replicated in time, since temporal variations in sewage impact were not the main object of the present study. Lack of temporal replication, thus, might limit general inference concerning sewage impact and the efficiency of TS in its detection. Nevertheless, our outcomes concur with previous studies exploring the effects of the same source of disturbance on other assemblages. In particular, studies carried out in the same area, in different periods of the year, consistently suggested the potential of the analysed source of disturbance to affect, at a hierarchy of spatial scales, the patterns of spatial distribution of sessile benthos (Terlizzi et al. 2002, 2005a), meiofauna (Fraschetti et al. 2006), fish (Guidetti et al. 2003) and molluscs (Terlizzi et al. 2005b). The present analysis also indicated that the differences in assemblages between the locations I and Cs were significantly larger than the variation between controls, considering both species (whole assemblage and its subsets) and other coarser levels of taxonomic discrimination.

Such differences might be directly linked to the source of disturbance, which could determine changes in abiotic factors (e.g. sedimentation rate, hydrodynamics, salinity, organic enrichment, associated oxygen demands, or a lingering effect due to the disturbance caused by the construction of the outfall), and/or indirectly originated from changes in the natural biotic features (e.g. background faunal assemblages including habitat-forming species). For example, the massive presence of algae typical of organically polluted environments at Location I (Terlizzi et al. 2002) might have influenced the structure of associated polychaete assemblages. Given the descriptive nature of the present study data, however, our intent was not to highlight cause-effect mechanisms underlying the observed patterns of spatial differences, but rather to analyse the efficiency of taxonomic surrogates in depicting such patterns.
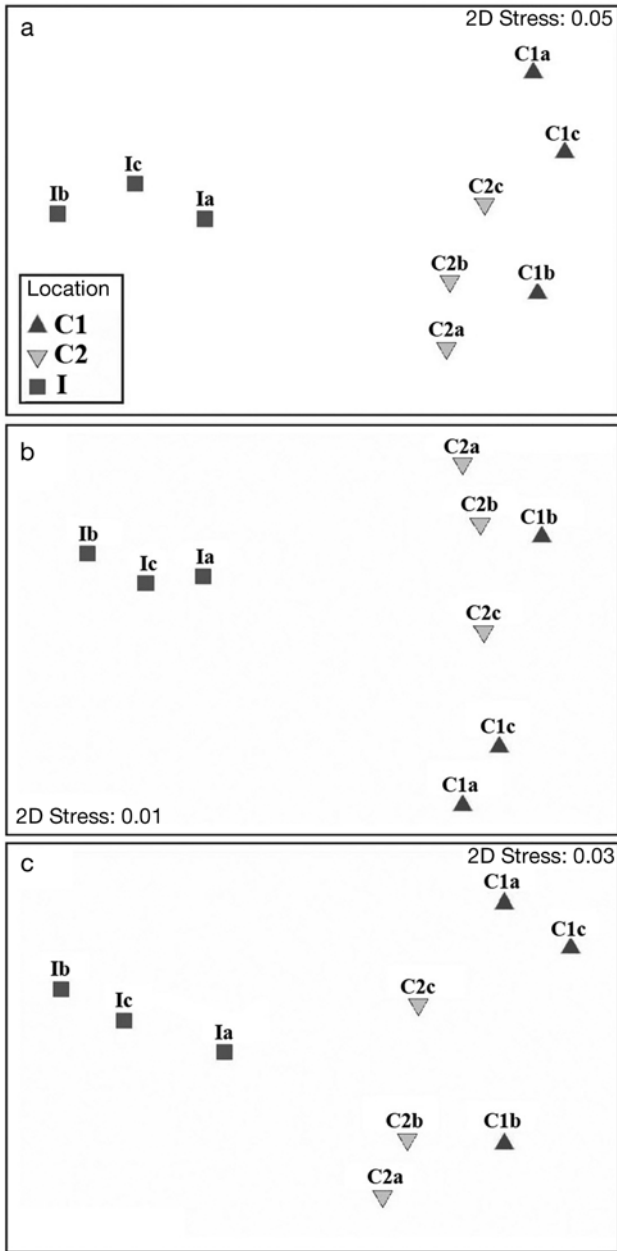
Fig. 7. NMDS plots on the basis of the Bray-Curtis dissimilarity measure of centroids of the 9 observations for each site at the impacted (I) vs. control locations (C1, C2). The centroids represent the polychaete assemblages classified at (a) species, (b) genus and (c) family level

Table 2. PERMANOVAs testing differences in the spatial distribution of the analysed assemblages (all polychaetes classified at different taxonomic levels, only Eunicida and Sabellida species, and only Syllidae species and genera (Cs') and the 2 control sites (Cs') at Location I. For each data set, the number of analysed taxa is indicated within parentheses. Detailed information on the analyses is given in the Table 1 legend. S = site

Source of variation	df	MS	F	p_perm	MS	F	p_perm	MS	F	p_perm	MS	F	p_perm	MS_denom	Permutable units
		Polychaete species (66)			Polychaete genera (45)			Polychaete families (15)							
S	2	2647.60	1.948	0.0148	1971.56	1.827	0.0396	1458.31	1.584	0.1014				res	27
Cs'	1	1869.33	1.195	0.2712	1447.11	1.169	0.2964	1320.04	1.241	0.2602				resCs'	18
I'-vs.-Cs'	1	3425.86	2.521	0.0130	2496.01	2.312	0.0276	1596.58	1.734	0.1216				res	27
Residuals	24	1359.07			1079.36			920.61							
resCs'	16	1564.34			1238.01			1146.43							
resI'	8	948.54			762.07			468.97							
		Eunicida species (9)			Sabellida species (11)			Syllidae species (30)			Syllidae genera (13)			
S	2	8068.06	2.446	0.0144	6684.80	1.996	0.0146	4050.99	2.106	0.0058	944.79	0.946	0.4412	res	27
Cs'	1	4082.52	1.272	0.2880	5697.37	1.554	0.1438	1100.03	0.641	0.7922	402.82	0.449	0.7590	resCs'	18
I'-vs.-Cs'	1	12053.59	3.655	0.0070	7672.23	2.291	0.0188	7001.96	3.641	0.0002	1486.76	1.488	0.2076	res	27
Residuals	24	3297.96			3349.50			1923.23			998.99				
resCs'	16	3210.83			3667.38			1715.90			897.77				
resI'	8	3472.22			2713.75			2337.90			1201.43				

In some cases, species lumping at higher taxonomic levels failed to detect patterns of differences at the scale of 100s of metres. At the scale of sites, in fact, differences were observed only when analysing data at either species or genus level. The performance of taxonomic surrogates depended on the taxonomic structure of the assemblage. In particular, at least for the scale of sites and the considered time of sampling, the

Table 3. PERMANOVAs testing differences in the spatial distribution of the polychaete assemblage classified at genus level excluding the mono-specific genera. Analysis testing differences at the location exposed to sewage pollution (I) and the control locations (Cs), and differences at the closest site to the sewage discharge (I′) and the 2 control sites (Cs′) at Location I. Analyses based on Bray-Curtis dissimilarity for respectively 18 and 16 taxa. Detailed information on the analyses is given in Table 1. L = location, S = site

Source of variation	df	MS	F	p_{perm}	MS_{denom}	Permutable units
At sewage discharge						
L	2	13136.30	5.129	*0.0008*	S(L)	9
Cs	1	5076.16	1.838	*0.1598*	S(Cs)	6
I-vs.-Cs	1	21196.43	8.276	*0.0006*	S(L)	9
S(L)	6	2561.14	2.457	0.0006	res	81
S(Cs)	4	2761.26	2.673	0.0010	resCs	54
S(I)	2	2160.90	2.038	0.0684	resI	27
Residuals	72	1042.20				
resCs	48	1033.06				
resI	24	1060.48				
Near sewage discharge						
S	2	2160.90	2.038	0.0684	res	27
Cs′	1	352.07	0.338	0.9438	resCs′	18
I′-vs.-Cs′	1	1559.89	1.471	0.2024	res	27
Residuals	24	1060.48				
resCs′	16	1043.21				
resI′	8	1095.03				

generic level was effective due to the information provided by mono-specific genera only. Indeed, differences among sites at the impacted location were not evident when data at genus level actually derived by lumping species distributions together (i.e. excluding mono-specific genera) (see Table 3).

Some crucial differences, however, characterised the data sets under consideration. Using species (the whole polychaete assemblage or its parts) and polychaete genera allowed the appreciation of differences at the scale of both site and location. When analysing the whole polychaete assemblage at family level, and the Syllidae at genus level, differences in assemblages were still apparent in the main contrast I-vs.-Cs, but no effect of the outfall in modifying patterns of site-to-site variation between I′-vs.-Cs′ was recorded. Using these 2 last surrogates limited the descriptive efficiency of the data set in detecting the impact. This supports the view that the efficiency of surrogates may be scale-dependent.

Our approach also provided information on how the efficiency of taxonomic surrogates may be influenced by the taxonomic structure of the assemblage. The least effective taxonomic surrogates (family, syllid genera) were characterised by the lowest R^2 values in the correlation between species and higher taxa. As expected, the higher the surrogate taxonomic level, the weaker the correlation. If the Syllidae were excluded from the analyses, the R^2 values rose on account of the taxonomic structure of the family in the area, since their

mean number of species per genus was higher than in the rest of the polychaetes. Excluding the syllids from the data set reduced the species:genus ratio to close to 1 (i.e. from 1.61 to 1.12). That is, some taxa (e.g. Syllidae or *Syllis*) more strongly influenced results obtained by TS than others of the same taxonomic rank. The information derived from the 41 collected mono-specific genera (e.g. *Polyophthalmus*, *Platynereis*) allowed the taxonomic surrogate 'genus' to work as effective bioindicator of differences in assemblages. For example, *Polyophthalmus pictus* was the only representative of the genus and even of the family Opheliidae; obviously, the use of TS did not produce any loss of knowledge on its distribution pattern. On the contrary, the information derived from syllids was clearly affected, since they were effective indicators at species level, but the worst ones (among those used herein) at genus level.

The case of syllids (1/3 of the collected polychaetes) is an example of how the TS approach may alter the original information. Among the 22 collected *Syllis* spp., both distribution patterns of 'sensitive' and 'tolerant' species were represented as well as 'opportunistic' ones (e.g. *S. krohni*, data not shown). The 3 most abundant *Syllis* spp. in the assemblages were characterised by different distributions at impacted versus control locations. It is clear that species of the same genus distributed independently of each other and their aggregation into a single taxonomic entity resulted in a heavy loss of ecological information. Differences in the distribution patterns between the 2 *Lysidice* spp. (*L. ninetta* and *L. collaris*) indicate that the case of *Syllis* spp. is not an exception within the analysed assemblage. According to Dauvin et al. (2003, p. 554), '… only in the theoretical case in which species of the same genus show different responses to a perturbation, and those species are dominant in the community, will the use of TS be inappropriate for pollution monitoring'. The present data show that this theoretical possibility can be valid in practice.

The information provided by species is not (or at least not always) redundant as is commonly thought (Gray & Pearson 1982, Gray et al. 1988, Chapman 1998). The present results imply that the efficiency of TS might be supported by the locally mono-specific higher taxa, which preserve the information provided by the original distribution of species. In our case, since lumping taxa together always limited the surrogate

efficiency, it might be concluded that, at least for polychaetes, only high levels of mono-specificity of higher taxa may allow TS to effectively work.

Włodarska-Kowalczuk & Kędra (2007) argued that the high species:higher taxa ratio does not directly induce ineffective TS performance. However, an assemblage with a high ratio is more likely to be characterised by a low number of mono-specific taxa, depending on its taxonomic structure. The present results suggest that the higher the number of mono-specific taxa retained after the TS lumping process, the higher the unaltered proportion of the original information and, consequently, the surrogate efficiency.

A further criticism of the abuse of surrogates in ecological studies arises from the concept of the redundancy of species and how they can coexist (Hutchinson 1959, Boero et al. 2004). Taxonomic sufficiency is essentially based on the implicit assumption that phylogenetically close species are probably ecologically similar. Species having similar ecological requirements, in fact, may compete for the same resources. If a stressor produces slightly different effects on slightly different species, the natural balance in competitive interactions might favour one of the competitors. It is often the case, however, that species of the same genus do have different body sizes. From this point of view, as remarked by Basset & De Angelis (2007), species coexistence is possible when different body sizes allow for different resource use from the same environmental patch. It is predicted, thus, that species of the same genus can coexist, with few specimens of a species with large body size sharing resources with many specimens of a co-generic species of smaller body size. This fine-scale resource partitioning as well as most of the connections between biodiversity and ecosystem functioning are lost when TS is used.

CONCLUSIONS

Approximations of the faunal features of an area should not be made *a priori*: the species-level information on the local fauna should be achieved before applying TS (Chapman 1998, Olsgard & Somerfield 2000, Gage 2001, Terlizzi et al. 2003). This represents an additional cost that should be taken into account in the TS cost-benefit analyses.

A note of caution should also be devoted to the use of subgroups of species as surrogates of the whole faunal assemblage. Here, the use of the species of Eunicida, Sabellida and Syllidae led to similar outcomes as for the whole polychaete assemblages. The analysis of target species groups could represent a valid cost-effective alternative to TS (Olsgard et al. 2003, Giangrande et al. 2005). However, they should be chosen depending on the habitat and biogeographic location. In particular, Syllidae can be an effective choice for the Mediterranean hard bottom environments where they are widely distributed, abundant and diverse in terms of species richness and ecological requirements (Giangrande et al. 2005, Musco & Giangrande 2005).

LITERATURE CITED

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. Aust J Ecol 26:32–46

Anderson MJ, Robinson J (2003) Generalised discriminant analysis based on distances. Aust NZ J Stat 45:301–318

Anderson MJ, ter Braak CJF (2002) Permutation tests for multifactorial analysis of variance. J Stat Comput Simul 73:85–113

Anderson MJ, Connell SD, Gillanders BM, Diebel CE, Blom WM, Saunders JE, Landers TJ (2005) Relationships between taxonomic resolution and spatial scales of multivariate variation. J Anim Ecol 74:636–646

Anderson MJ, Gorley RN, Clarke KR (2008) PERMANOVA+ for PRIMER: guide to software and statistical methods. PRIMER-E, Plymouth

Balmford A, Green MJB, Murray MG (1996) Using higher-taxon richness as a surrogate for species richness: I. Regional tests. Proc R Soc Lond B Biol Sci 263:1267–1274

Basset A, De Angelis L (2007) Body size mediated coexistence of consumers competing for resources in space. Oikos 116:1363–1377

Bertrand Y, Pleijel F, Rouse GW (2006) Taxonomic surrogacy in biodiversity assessments, and the meaning of Linnaean ranks. Syst Biodivers 4:149–159

Bishop MJ, Underwood AJ, Archambault P (2002) Sewage and environmental impacts on rocky shores: necessity of identifying relevant spatial scales. Mar Ecol Prog Ser 236: 121–128

Boero F, Belmonte G, Bussotti S, Fanelli G and others (2004) From biodiversity and ecosystem functioning to the roots of ecological complexity. Ecol Complex 1:101–109

Chapman MG (1998) Relationships between spatial patterns of benthic assemblages in a mangrove forest using different levels of taxonomic resolution. Mar Ecol Prog Ser 162: 71–78

Clarke KR, Gorley RN (2006) PRIMER V6: user manual/tutorial. PRIMER-E, Plymouth

Dauvin JC, Gomez Gesteira L, Salvande Fraga M (2003) Taxonomic sufficiency: an overview of its use in the monitoring of sublittoral benthic communities after oil spills. Mar Pollut Bull 46:552–555

Ellis D (1985) Taxonomic sufficiency in pollution assessment. Mar Pollut Bull 16:459

Ferraro SP, Cole FA (1995) Taxonomic level sufficient for assessing pollution impacts on the southern California Bight macrobenthos — revisited. Environ Toxicol Chem 14: 1031–1040

Fraschetti S, Gambi C, Giangrande A, Musco L, Terlizzi A, Danovaro R (2006) Structural and functional response of meiofauna rocky assemblages exposed to sewage pollution. Mar Pollut Bull 52:540–548

Gage JD (2001) Deep-sea benthic community and environmental impact assessment at the Atlantic Frontier. Cont Shelf Res 21:957–986

Giangrande A, Licciano M (2008) Revision of the species of *Megalomma* (Polychaeta, Sabellidae) from the Mediterranean Sea, with the description of *M. messapicum* n. sp. Ital J Zool 75:207–217

Giangrande A, Delos AL, Musco L, Licciano M, Pierri C (2004) Polychaete assemblages of rocky shore along the South Adriatic coast (Mediterranean Sea). Cah Biol Mar 45:85–95

Giangrande A, Licciano M, Musco L (2005) Polychaetes as environmental indicators revisited. Mar Pollut Bull 50:1153–1162

Glasby TM (1997) Analysing data from post-impact studies using asymmetrical analyses of variance: a case study of epibiota on marinas. Aust J Ecol 22:448–459

Gray JS, Pearson TH (1982) Objective selection of sensitive species indicative of pollution-induced change in benthic communities. I. Comparative methodology. Mar Ecol Prog Ser 9:111–119

Gray JS, Aschan M, Carr MR, Clarke KR and others (1988) Analysis of community attributes of the benthic macrofauna of the Frierfjord/Langesundfjord and in a mesocosm experiment. Mar Ecol Prog Ser 46:151–165

Guidetti P, Terlizzi A, Fraschetti S, Boero F (2003) Changes in Mediterranean rocky-reef fish assemblages exposed to sewage pollution. Mar Ecol Prog Ser 253:269–278

Hellberg ME (1998) Sympatric sea shells along the sea's shore: the geography of speciation in the marine gastropods *Tegula.* Evolution 52:1311–1324

Hutchinson GE (1959) Homage to Santa Rosalia; or, why are there so many kinds of animals? Am Nat 93:145–159

Musco L, Giangrande A (2005) Mediterranean Syllidae (Annelida: Polychaeta) revisited: species diversity, biogeography and faunal fidelity to environmental features. Mar Ecol Prog Ser 304:143–153

Olsgard F, Somerfield PJ (2000) Surrogates in marine benthic investigations — which taxonomic unit to target? J Aquat Ecosyst Stress Recovery 7:25–42

Olsgard F, Brattegard T, Holthe T (2003) Polychaetes as surrogates for marine biodiversity: lower taxonomic resolution and indicator groups. Biodivers Conserv 12:1033–1049

Rouse GW, Pleijel F (2001) Polychaetes. Oxford University Press, Oxford

Somerfield PJ, Clarke KR (1995) Taxonomic levels, in marine community studies, revisited. Mar Ecol Prog Ser 127:113–119

Terlizzi A, Fraschetti S, Guidetti P, Boero F (2002) The effects of sewage discharge on shallow hard bottom sessile assemblages. Mar Pollut Bull 44:544–550

Terlizzi A, Bevilacqua S, Fraschetti S, Boero F (2003) Taxonomic sufficiency and the increasing insufficiency of taxonomic expertise. Mar Pollut Bull 46:556–561

Terlizzi A, Benedetti-Cecchi L, Bevilacqua S, Fraschetti S, Guidetti P, Anderson MJ (2005a) Multivariate and univariate analyses in environmental impact assessment: a case study of Mediterranean subtidal sessile assemblages. Mar Ecol Prog Ser 289:27–42

Terlizzi A, Scuderi D, Fraschetti S, Anderson MJ (2005b) Quantifying effects of pollution on biodiversity: a case study of highly diverse molluscan assemblages in the Mediterranean. Mar Biol 148:293–305

Terlizzi A, Anderson MJ, Bevilacqua S, Fraschetti S, Włodarska-Kowalczuk M, Ellingsen KE (2009) Beta diversity and taxonomic sufficiency: Do higher-level taxa reflect heterogeneity in species composition? Divers Distrib 15:450–458

Underwood AJ (1991) Beyond BACI: experimental designs for detecting human environmental impacts on temporal variations in natural populations. Aust J Mar Freshw Res 42:569–587

Underwood AJ (1994) On beyond BACI: sampling designs that might reliably detect environmental disturbances. Ecol Appl 4:3–15

Underwood AJ, Peterson CH (1988) Towards an ecological framework for investigating pollution. Mar Ecol Prog Ser 46:227–234

Włodarska-Kowalczuk M, Kędra M (2007) Surrogacy in natural patterns of benthic distribution and diversity: selected taxa versus lower taxonomic resolution. Mar Ecol Prog Ser 351:53–63