Zaiko A., Samulioviene A., Ardura A., Garcia-Vazquez E. 2015. Metabarcoding approach for non-indigenous species surveillance in marine coastal waters. Marine Pollution Bulletin, 100:53-59. doi:10.1016/j.marpolbul.2015.09.030

Access to underlying data: http://doi.pangaea.de/10.1594/PANGAEA.853664

TITLE: "Metabarcoding approach for non-indigenous species surveillance in marine coastal waters"

Anastasija Zaiko[1,2], Aurelija Samuiloviene[1], Alba Ardura [1,3], Eva Garcia-Vazquez[3]

[1] Marine Science and Technology Centre, Klaipeda University, H. Manto 84, LT 92294, Klaipeda, Lithuania

[2] Coastal and Freshwater Group, Cawthron Institute, 98 Halifax Street East, 7010 Nelson, New Zealand

[3] Department of Functional Biology, University of Oviedo. C/ Julian Claveria s/n. 33006-Oviedo, Spain

Corresponding author: Anastasija Zaiko, Marine Science and Technology Centre, Klaipeda University, H. Manto 84, LT 92294, Klaipeda, Lithuania, Tel.: +370 46 398855, Fax.: +370 46 398845, e-mail: anastasija@corpi.ku.lt

Zaiko A., Samulioviene A., Ardura A., Garcia-Vazquez E. 2015. Metabarcoding approach for non-indigenous species surveillance in marine coastal waters. Marine Pollution Bulletin, 100:53-59. doi:10.1016/j.marpolbul.2015.09.030
Access to underlying data: http://doi.pangaea.de/10.1594/PANGAEA.853664

Abstract

Here we applied High-Throughput Sequencing (HTS) metabarcoding for the surveillance of plankton communities within the SE Baltic Sea coastal zone. Results were compared to those from routine monitoring survey and morphological analyses. Four of five non-indigenous species found in the samples were identified exclusively by metabarcoding. All of them are considered as invasive in the Baltic Sea with reported impacts on ecosystem and biodiversity. This study indicates that despite some current limitations, HTS metabarcoding can provide information on presence of exotics and advantageously complement conventional approach, without exceeding the regular monitoring effort. Such a combination of HTS metabarcoding and observational records could be recommended, even in the currently immature status of HTS, for assisting early detection of marine pests and delivery of the non-indigenous species-related environmental status metrics.

Keywords: COI, high-throughput sequencing, non-indigenous species, Baltic Sea

Timely detection and accurate identification of marine species is a prerequisite for the efficient pest spread prevention or mitigation (Lehtiniemi et al. 2015). In many cases it is not an easy task though. The reported distribution of a species often says more about the distribution of scientists than it does about the species itself (Rilov and Crooks 2009). Most marine surveillance programs require considerable taxonomic expertise, especially for identifying cryptic species and those at the larval stage. On the other hand, the taxonomic expertise is becoming less and less available and its importance is continuously underestimated (Hopkins and Freckleton 2002; Kim and Byrne 2006). As a result, misidentifications of organisms may occur and non-indigenous species (NIS) may therefore

remain undetected for extended periods until becoming abundant and widespread marine pests (Freire et al. 2014; Zaiko et al. 2014). In order to enhance the opportunities and efficiency of management responses (e.g. mitigation of further spread by strengthening the control on domestic pathways, raising public awareness, etc.), it is extremely important to detect NIS at the initial stage of incursion, when population is confined to a small area and a low density (Olenin et al. 2011; Pochon et al. 2013).

Emerging molecular techniques are advancing rapidly and provide promising tools for species identification from environmental samples (Pochon et al. 2013; Wood et al. 2013; Zaiko et al. 2015a,b). DNA barcoding and metabarcoding have a particular potential to provide more accurate and standardized, high resolution taxonomic data (Hajibabaei et al. 2007; Ji et al. 2013; Wood et al. 2013). DNA barcoding is a method of taxonomical assignment of a specimen based on sequencing a short standardized DNA fragment (molecular marker or barcode), specific for a species (Hajibabaei et al. 2007). Metabarcoding applies the principle of DNA barcoding with extension to communities and biotic assemblages rather than single individuals (Hajibabaei et al. 2011; Wood et al. 2013). The recent development of High-Throughput Sequencing (HTS) technologies has introduced new opportunities in marine environmental surveillance due to the massive sequencing capacity, allowing multiple samples to be processed faster and cheaper compared to the traditional morphological and Sanger sequencing based approaches (Ji et al. 2013; Pochon et al. 2013). However, these techniques are far from perfect. Although they can be highly informative, there are possible biases due to preferential annealing of expectedly universal primers in some species than in others. For example, using the mitochondrial *cytochrome c oxidase subunit I* (COI) gene as a DNA barcode some aquatic groups such as copepoda and cladocera may be overlooked (Creer et al. 2010; Tang et al. 2012; Zhan et al. 2014). On the other hand,

current reference databases do not contain a balanced suite of representatives of all taxonomic groups (Ardura et al. 2013; Ratnasingham and Hebert 2013). Other problems are intrinsic biases for some type of sequences in HTS platforms (Frey et al. 2014, Zaiko et al. 2015a). Notwithstanding it, HTS technologies are nowadays increasingly employed for biodiversity surveys together with other independent methods (Taberlet et al. 2012, Aylagas et al. 2014, Zaiko et al. 2015b), in order to better identifying and eventually correcting their current biases.

Here we present a pilot application of metabarcoding approach for the NIS surveillance in zooplankton communities. We compare the results to those from the conventional approach and discuss potential implications for NIS monitoring in coastal waters. In our study we challenged the hypothesis if metabarcoding in combination with HTS could provide consistent results with those obtained from the routine monitoring programs (in terms of NIS detection), without exceeding the regular sampling effort. For the HTS analysis we employed the universal COI gene as a molecular marker because sequence data of this gene are amongst the most voluminous components of the public databases (Medlin and Kooistra 2010; Ratnasingham and Hebert 2007), although as commented above it can be expected that some taxonomic groups may be overlooked with this marker (Creer et al. 2010; Tang et al. 2012; Zhan et al. 2014). The study was conducted in the south-eastern part of the Baltic Proper, within the Lithuanian coastal zone of the Baltic Sea and the Klaipeda strait area (Fig. 1). Although the Baltic Sea is often referred to as a "low biodiversity" system (Dippner et al. 2000; Leppäkoski et al. 2002) and considered as one of the most intensively studied regional seas in the world (Ojaveer et al. 2010), there are still substantial gaps in biodiversity knowledge and relevant surveillance programs. For instance, in routine zooplankton monitoring only dominant species of certain groups and size fractions are identified and

counted, disregarding many important taxa, particularly small-sized and cryptic forms (Ojaveer et al. 2010; Hällfors et al. 2013). This and insufficient spatial and temporal resolution of the monitoring programs (Patricio et al. 2014) may handicap the environmental status assessments relevant to biodiversity and NIS components. Therefore coastal waters of the Baltic Sea were considered as a good model location for our study. The ultimate aim was to determine the improvement of NIS detection achieved by combining visual and HTS methods, and the taxonomic biases of the each one, if any. In other words, we wanted to know the potential utility of combining two (currently imperfect) methods for NIS detection, since the perfect definitive methodology does not seem to exist yet.

Materials and Methods

Sample collection and handling

For assessing the performance of the metabarcoding in NIS surveillance, we initiated an alternative sampling campaign in the three areas traditionally sampled within the Lithuanian national monitoring (LNM) program during the boreal vegetation season (May-August): Klaipeda Strait - KS; the "sea gates" - SG - entrance to the strait from the sea side and in the open sea at the northern edge of the Curonian Lagoon plume area - PA (Fig. 1).

Six samples for HTS were taken at those locations in 2013 (KS on June and August; SG on May, June and August; PA on May). For sample collection, a plankton net (55 cm diameter, 80 μm mesh size) was towed vertically 8-10 meters, depending on the depth. The concentrated samples (approximately 10 ml) were kept on ice until delivered to the laboratory (1-4 hours), then filtered through the sterile 0.2μm Nuclepore™ membrane, which was thereafter preserved with 96% ethanol for the future bulk DNA extraction.

Zaiko A., Samulioviene A., Ardura A., Garcia-Vazquez E. 2015. Metabarcoding approach for non-indigenous species surveillance in marine coastal waters. Marine Pollution Bulletin, 100:53-59. doi:10.1016/j.marpolbul.2015.09.030
Access to underlying data: http://doi.pangaea.de/10.1594/PANGAEA.853664

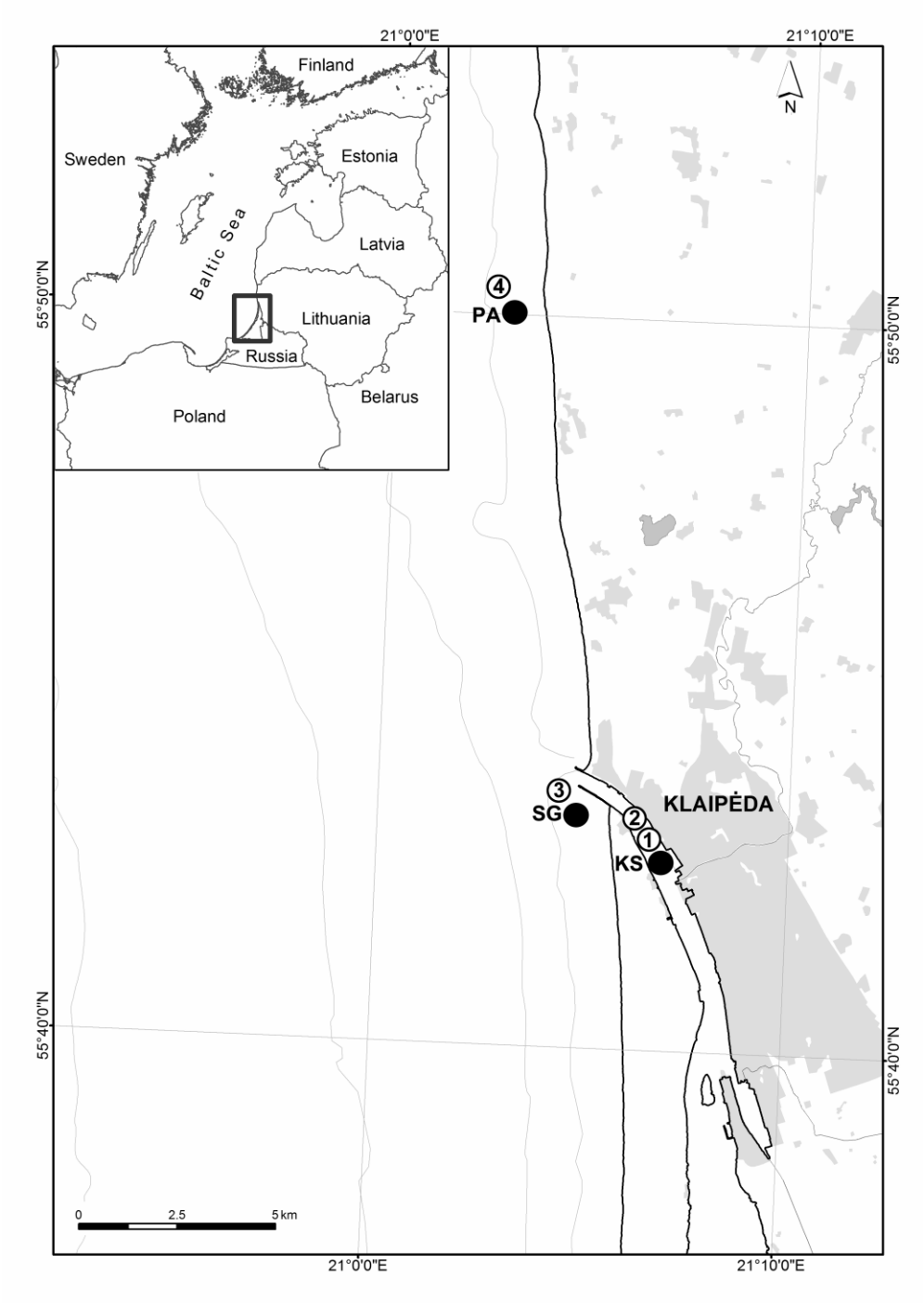

Figure 1. Sampling sites: filled dots – metabarcoding samples, open dots – Lithuanian national monitoring (LNM) locations (indicated by numbers). Labels indicate sampling site name used in this account: KS - Klaipeda Strait, SG - sea gates, PA - plume area.

For comparative analyses, we used eight conventional zooplankton samples collected independently by EPA Marine Research Department as a part of the LNM program in 2013:

at KS on May, June, July and August; SG on May and August; PA on May and August. Samples were collected by vertical hauls using WP-2 net (57 cm diameter, 100 μm mesh size) from about 5 m above bottom to the water surface (HELCOM 2005). Samples were concentrated to approximately 330 ml volume and fixed with 4% formaldehyde solution onboard. In the laboratory, all individuals found in the samples were identified to species or genus level (if not possible - to higher taxonomic units).

The sampling time did not coincide exactly in both campaigns (LNM and HTS), but corresponding samples used for comparisons were collected on approximate dates. Sampling effort (as the number of samples analyzed by each method) and environmental conditions at the sampling locations are shown in Table 1.

Table 1. Sampling effort as number of samples analyzed by each method; environmental conditions over the observation period; number of species/taxa detected (proportion of NIS in parentheses), in three sampling sites – Klaipeda strait (KS), sea gates (SG) and plume area (PA).

		KS	SG	PA
Sampling effort	LNM	4	2	2
	HTS	2	3	1
Environmental conditions	Salinity, PSU	0.2-1.9	6.2-6.6	4.8-6.2
	Temperature, C˚	17-23	11-21	11-22
	pH	7.2-7.5	7.1-8.1	6.0-8.8
Species (NIS proportion)	LNM	23 (0)	13 (0.08)	12 (0.08)
	HTS	24 (0.13)	24 (0.17)	16 (0.13)

Molecular and bioinformatics analyses

The precipitates from membrane filters were removed with sterile blades and DNA was extracted from the filters using QIAamp DNA Mini Kit (Qiagen) following the manufacturer's instructions. In order not to overload the spin columns, each sample was homogenized and split into 3 or 4 subsamples depending on the amount of material. No further homogenization was applied as the tissues were lysed enzymatically as described in

the protocol. Extracted DNA subsamples were then pooled back for the further analysis. The DNA was quantified by a fluorescence-based quantification method (Picogreen, Invitrogen). The purity of DNA was assessed by the NanoDrop instrument.

The modified universal COI primers (Geller et al. 2013) were used for PCR amplification of a fragment of approx. 658 base pairs (bp) within the mitochondrial gene coding for the cytochrome oxidase subunit I (COI). PCR reactions were undertaken by Macrogen based on the original protocol described by Geller et al. (2013). They were sequenced using a Genome Sequencer FLX (Roche) by Macrogen (Korea). The initial GS FLX data processing was performed using the Roche GS FLX software (v2.9). The software uses tag (barcode) sequences to segregate the reads from each sample, by matching the initial and final bases of the reads to the known tag sequences used in the preparation of the libraries. Zero base errors were allowed in this sorting by tag step.

Raw data were then processed using MOTHUR v.1.34.4 software (Schloss et al. 2009). Reads having low quality (mean Phread scores ≥20) or any ambiguous bases, shorter that 400 bp, with >10-base long homopolymers, were removed from the downstream analysis. Retained sequences were grouped into a single file and de-replicated into unique sequences which were then clustered into operational taxonomic units (OTUs) and taxonomically assigned based on the curated reference database described below. Prior to sequence clustering and assignment, chimera checks were undertaken using Uchime 4.2 in *de novo* mode (Edgar et al. 2011) and sequences with Uchime scores above 0.3 were discarded. The singletons (OTUs represented by less than 2 sequences) were also discarded from the downstream analysis.

A database was built as a basis for global alignments of the environmental sequences by isolating all entries of aquatic invertebrates (sequences and taxonomy) from the BOLD

database (Ratnasingham and Hebert 2010). For taxonomic classification Wang method was applied with minimum bootstrap value set at 97%. The curated dataset of assigned taxa was further used for biodiversity analysis.

Statistical analyses

Given the nature of HTS data for eukaryotes (one single cell carries many mitochondrial DNA sequences, and in environmental DNA there is a mixture of multi- and unicellular organisms), only presence-absence data were considered. Untransformed data (number of reads per OTU per sample) were used to generate OTU rarefaction curves using Vegan package in R software (Oksanen et al. 2014). For comparing the datasets obtained by conventional monitoring and metabarcoding we have considered the presence of a species within a site as a unit (i.e. the number of + in Table 1), taking into account only those taxa identified to a taxonomic level (i.e. species, family). Non-parametric Chi-Square of contingency test was employed for comparison of proportion of detected NIS between two methods (HTS and morphological analysis).

With presence-absence data, the taxonomic composition (at species and family levels) in analyzed samples was related to the multivariate effects of species identification method (HTS *versus* morphological analysis), sampling site (KS, SG and PA), and the sampling time (May, June, July and August) using PRIMER 6 & PERMANOVA software package (PRIMER-E, Ltd., UK). The statistical differences between the factor levels were assessed using repeated measure permutational analysis of variance (PERMANOVA) with method and site entered as fixed factors, and time was nested within the site level. The permutation of residuals under a reduced model was applied. The results of the analysis were verified by a visual assessment of patterns in the nonmetric multidimensional scaling (MDS) plots based on Jaccard similarities. The contribution of species presence to similarities within the factor

levels and dissimilarities between the groups was estimated via two-way crossed similarity percentages (SIMPER) analysis.

Results

DNA extractions from all samples yielded good quality DNA as indicated by the nanophotometer absorbance ratios 260/280 which ranged between 1.7 and 2.1. The concentration of DNA extracted ranged between 6.3 and 65.7 ng/µl. The raw sequencing data of the positive PCR amplicons produced 118,869 reads with average read length of 376 bp. The initial filtering of sequences with ambiguous bases, barcodes errors or mismatches has resulted in removal of 18% of reads. After the further denoising, stringent filtering and normalization, 19,524 sequences de-replicated into 10,920 unique sequences with the average read length of 498 bp, remained suitable for the downstream analysis and taxonomy assignment. These sequences were clustered into 387 OTUs using a 97% sequence similarity threshold (Suppl. 1). The mean number of OTUs retrieved from the sample was 130 ± 42. Only 18 OTUs (ca. 5%) could not be identified using the curated subset of the BOLD database. The remaining were assigned to various taxonomic levels. The rarefaction curves indicated that in terms of numbers of sequences generated, all samples had an adequate size for the further species richness analysis (Fig. 2).

Sequences from the curated HTS dataset were assigned to 32 taxa belonging to 4 phyla, 6 classes, 8 orders, 15 families, 19 genera and 19 species (Table 2). In LNM samples there were 30 taxa morphologically identified by an expert, belonging to 2 phyla, 3 classes, 6 orders, 20 families, 24 genera and 21 species (Table 2). Seventeen taxa identified at a species or genus level were shared between both datasets. Considering the results of each method by taxonomic group, in the LNM dataset, the highest occurrence (presence per site) was reported for rotifers, whereas in HTS data there was no marked difference among phyla. Generally,

the number of detected taxa in metabarcoding samples tended to be higher than in LNM samples (Table 1); for example Mollusca and Annelida were detected only with this method (Table 2). At a lower taxonomic level, more genera of Maxillopoda and Rotifers were detected in LNM.

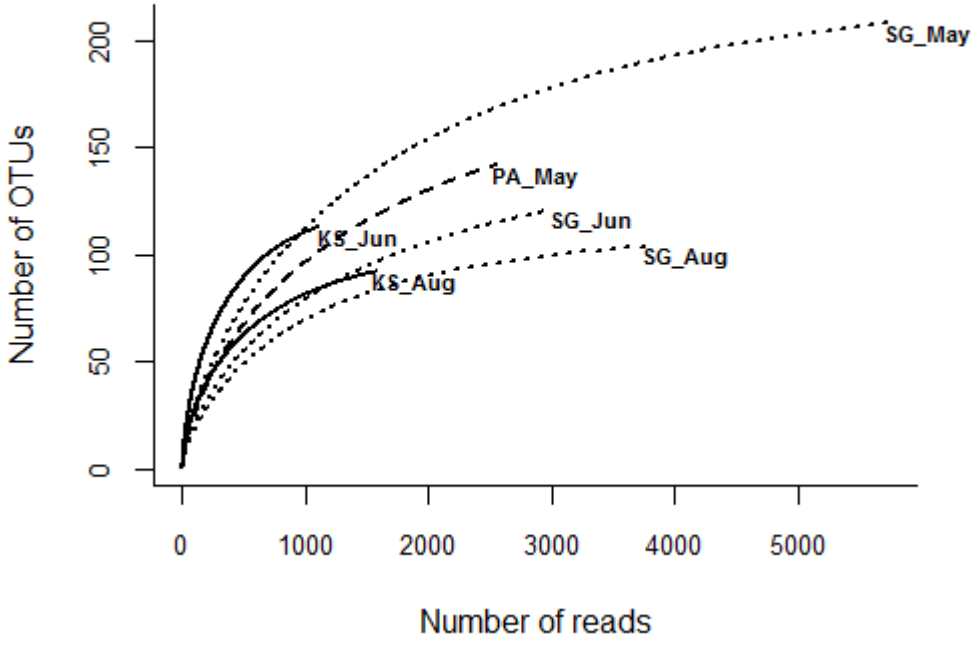

Figure 2. Rarefaction curves plotting the number of reads by the number of OTUs for each of the 6 samples analyzed using high-throughput sequencing. Labels next to the curves indicate the sampling site and the sampling month (see Table 1, Figure 1 and the text for details).

There were 21 OTUs (1,061 sequences aligned with up to 100% similarity (Suppl. 1) to 5 non-indigenous species considered as invasive in the Baltic Sea with reported impacts on biodiversity and ecosystem functioning (Zaiko et al. 2011). Despite higher LNM sampling effort in KS and PA (Table 1), most of NIS in the samples were identified exclusively by the HTS, except for Ponto-Caspian water flea *Cercopagis pengoi* which was also morphologically detected from the LNM samples (Table 2). Differences in the proportion of

Zaiko A., Samulioviene A., Ardura A., Garcia-Vazquez E. 2015. Metabarcoding approach for non-indigenous species surveillance in marine coastal waters. Marine Pollution Bulletin, 100:53-59. doi:10.1016/j.marpolbul.2015.09.030
Access to underlying data: http://doi.pangaea.de/10.1594/PANGAEA.853664

NIS detected by two methods were significant (estimated Chi-Square of contingency was 11.32, P<0.001, df=1).

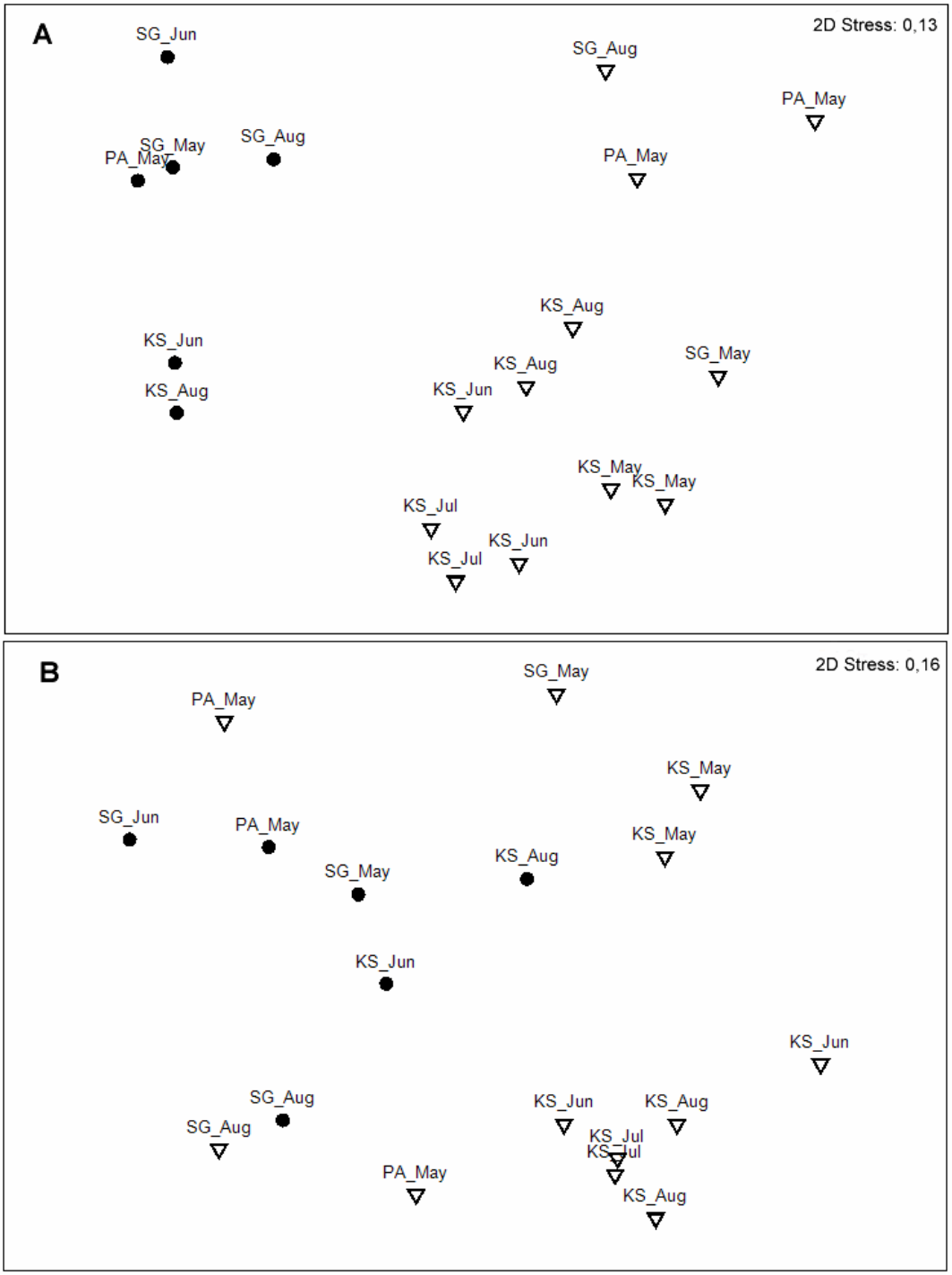

Figure 3. MDS plots showing similarities in the composition of taxa (A- at a species (lowest assigned taxonomy) level, B- at a family level) identified by to approaches at three sampling locations (KS – Klaipeda Strait, SG – sea gates, PA – plume area). Black dots correspond to HTS samples, open triangles – LNM samples, labels indicate sampling site and month.

Multidimentional scaling (MDS) analysis performed on presence-absence data at a species (lowest assigned taxonomy) level showed an apparent clustering between methods and some separation between sampling sites (Fig. 3a), particularly samples from the Klaipeda Strait area tended to aggregate more closely. However, at a family level the distinction between methods was more blurred (Fig. 3b).

PERMANOVA confirmed statistically significant difference in taxonomic composition at the species level both between methods and between sampling events (Table 3A). The average dissimilarity between LNM and metabarcoding samples was 73%. From the location perspective, samples from Klaipeda strait area grouped together the best with 75% average similarity and highest distinctness with samples from the plume area site (67%). At a family level, the effect of the method was demoted (average dissimilarity between methods dropped to 48%) with sampling event remaining the only significant factor affecting the taxonomic composition (Table 3B).

Discussion

The direct comparison of the biodiversity registered from LNM and HTS campaigns in this study should be performed with caution, since taxonomic data were not retrieved from exactly the same samples. Therefore certain mismatches in species composition between the analyzed datasets could be caused by local dynamics of the zooplankton communities, since zooplankton community is characterized by a pronounced degree of unpredictability (Caroppo et al. 2013). However in the current study we did not intend to oppose the two methods, but rather to compare their performance in routine surveillance and assess their complementarity and potential for NIS detection. The results should be read in this context only.

Comparison of the zooplankton surveillance results obtained by two approaches indicated some discrepancies in retrieved species data, although the taxonomical composition was rather consistent at higher levels. All the genera and species identified in our survey had been reported also from other studies in the area (Aleksandrov et al. 2009; Telesh et al. 2009). Noticeably, among the species identified exclusively by metabarcoding, five were the benthic organisms with planktonic larval stage (*Dreissena polymorpha, Hydrobia ulvae, Marenzelleria neglecta, M. viridis* and *Mytilus sp*.). Usually, larvae of benthos are not identified in monitoring samples due to their cryptic morphology and lack of specific taxonomical expertise. This implies potential biosecurity risks, since many invasive sessile organisms have dispersive planktonic stage (like three of the aforementioned species – *D. polymorpha, M. neglecta* and *M. viridis*), and might be overlooked in morphologically analyzed monitoring samples. It means that these and other non-indigenous species might spread unnoticeably in the region, until their presence and impacts become apparent and irreversible (Lehtiniemi et al. 2015).

In all three sampling locations there were sequences attributed with high confidence to the invasive polychaete *M. viridis*. Based on the results of the earlier molecular identification and areal distribution assessment of three sibling *Marenzelleria* species within the Baltic Sea (Blank et al. 2008; Michalek and Werner 2012), only *M. neglecta* was unambiguously reported from the eastern and south-eastern regions (where Lithuanian cost belongs to). Our findings might contribute to the update of the current distributional maps of the species as well as national inventories of the non-indigenous organisms. However further ground-truthing studies are required to verify the particular distribution of these two species in the benthic habitats.

Expanding the reference databases seems to be urgent for a better performance of the metabarcoding approach (Ratnasingham and Hebert 2007). Five of the species detected in LNM and not represented in HTS data (*Eubosmina longispina*, *Eudiaptomus gracilis*, *Kellicottia longispina*, *Phyllognathopus poludosus* and *Synchaeta monopus*) did not have reference barcode sequences in the BOLD database at the time of this study. Presumably, those species could remain unassigned in our data. The limitation of the reference database and limited resolution of the morphological assignments could explain the better correspondence of LNM and HTS data at a higher taxonomic (family) level. This indicates the necessity of the further improvements of the both approaches in order to advance their performance and complementarity in the monitoring surveys.

The use of other DNA regions as barcodes could overcome the problem of deficient recognition of some taxonomic groups by COI primers (Creer et al. 2010; Tang et al. 2012; Zhan et al. 2014) and improve the number of retrieved species and overall taxonomic resolution of the assignment, but to date all databases are still incomplete and some taxonomic groups are underrepresented (e.g. Ratnasingham and Hebert 2013). Therefore when considering the metabarcoding approach as a tool for routine surveillance programs, it is strongly recommended to compile an operational reference database curated cooperatively by the regional taxonomic and phylogenetic experts.

As evidenced by our study, despite some current limitations, metabarcoding in combination with HTS has provided more information on presence-absence of NIS comparing to conventional approach with sampling effort comparable to that usual for marine monitoring in the Baltic Sea coastal areas (on average two to three sampling events per season at three sampling localities (HELCOM 2005)).

As a complementary monitoring measure, HTS is advantageous for determining the identities of marine NIS, uncovering new or earlier overlooked invasions, monitoring invasion dynamics, assessing and predicting the secondary spread and thus NIS effect on recipient communities (Wood et al. 2013; Lehtiniemi et al. 2015; Zaiko et al. 2015b). HTS data obtained from the non-targeted metabarcoding survey can provide information on the number of NIS in a given area and their temporal and spatial occurrence necessary for the environmental status assessment within the MSFD (Lehtiniemi et al. 2015). As expected, with constantly improving technology and understanding of DNA fate in the environment (Kelly et al. 2014), molecular monitoring techniques will become more quantitative thus allowing more accurate assessment of marine biodiversity. This approach does not require particular taxonomic expertise, allows precise identification of cryptic life stages (eggs or larvae) as well as detection of rare and sparsely distributed organisms. Even at the current stage of development, metabarcoding and HTS techniques can supplement observational records, experimental studies and field surveys in order to provide essential information on the marine ecosystem and support competent and efficient management.

Acknowledgements

This study was supported by the DEVOTES (DEVelopment Of innovative Tools for understanding marine biodiversity and assessing Good Environmental Status) project funded by the European Union under the 7th Framework Programme, 'The Ocean of Tomorrow' Theme (Grant Agreement No. 308392), www.devotes-project.eu; the Spanish Grant MINECO CGL2013-42415-R (with a contribution from the Marine Observatory of Asturias and GRUPIN14-093); and BIO-C3 (Biodiversity changed investigating causes, consequences and management implications) within the BONUS, the joint Baltic Sea Research and

Development programme, funded by the EU 7th Framework Programme and Research Council of Lithuania (Grant Agreement No. BONUS-1/2014). We thank Lithuanian EPA Marine Research Department for the monitoring datasets provided. We also thank Y.J. Borrell (University of Oviedo) for his help with HTS samples.

References

1. Aleksandrov, S. V., N. N. Zhigalova, and A. S. Zezera. 2009. Long-term dynamics of zooplankton in the southeastern Baltic Sea. Russian Journal of Marine Biology 35:296-304.
2. Aylagas E., A. Borja, N. Rodriguez-Ezpeleta. 2014. Metabarcoding: towards a genetics based marine biotic index (gAMBI). PLoS ONE. 9:e90529.
3. Blank, M., A. O. Laine, K. Jurss, and R. Bastrop. 2008. Molecular identification key based on PCR/RFLP for three polychaete sibling species of the genus Marenzelleria, and the species' current distribution in the Baltic Sea. Helgoland Marine Research 62:129-141.
4. Caroppo, C., I. Buttino, E. Camatti, G. Caruso, R. De Angelis, C. Facca, F. Giovanardi, L. Lazzara, O. Mangoni, and E. Magaletti. 2013. State of the art and perspectives on the use of planktonic communities as indicators of environmental status in relation to EU Marine Strategy Framework Directive. Biologia Marina Mediterranea 20(1): 65-73.
5. Dippner, J.W., G. Kornilovs, and L. Sidrevics. 2000. Long-term variability of mesozooplankton in the Central Baltic Sea. Journal of Marine Systems 25: 23-31.
6. Edgar R.C., B.J. Haas, J.C. Clemente, C. Quince, R. Knight. 2011. UCHIME improves sensitivity and speed of chimera detection, Bioinformatics 27:2194-2200.
7. Freire, M., G. N. Genzano, S. Neumann-Leitao, and C. D. Perez. 2014. The non-indigenous medusa Blackfordia virginica (Hydrozoa, Leptothecata) in tropical Brazil: 50 years of unnoticed presence. Biological Invasions 16:1-5.
8. Frey, K.G., J.E. Herrera-Galeano, C.L. Redden, T.V. Luu, S.L. Servetas, A.J. Mateczun, V.P. Mokashi, K.A. Bishop-Lilly. 2014. Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. BMC Genomics 15: 96.
9. Geller, J., C. Meyer, M. Parker, and H. Hawk. 2013. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. Molecular Ecology 13:851-861.
10. Hajibabaei, M., S. Shokralla, X. Zhou, G. A. Singer, and D. J. Baird. 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS One 6:e17497.
11. Hajibabaei, M., G. A. Singer, P. D. N. Hebert, and D. A. Hickey. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends in Genetics 23:167-171.
12. Hällfors, H., L. Uusitalo, and U. Lips. 2013. Report on gaps in monitoring programs. GES-REG WP4, 28 pp.

13. HELCOM. 2005. Manual for marine monitoring in the COMBINE programme of HELCOM, Part C. http://sea.helcom.fi/Monas/CombineManual2/PartC/CFrame.htm.
14. Hopkins, G. W., and R. P. Freckleton. 2002. Declines in the numbers of amateur and professional taxonomists: implications for conservation. Animal Conservation 5:245-249.
15. Ji, Y., L. Ashton, S. M. Pedley, D. P. Edwards, Y. Tang, A. Nakamura, R. Kitching, P. M. Dolman, P. Woodcock, F. A. Edwards, T. H. Larsen, W. W. Hsu, S. Benedick, K. C. Hamer, D. S. Wilcove, C. Bruce, X. Wang, T. Levi, M. Lott, B. C. Emerson, and D. W. Yu. 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecology Letters 16:1245-1257.
16. Kelly, R. P., J. A. Port, K. M. Yamahara, R. G. Martone, N. Lowell, P. F. Thomsen, M. E. Mach, M. Bennet, E. Prahler, M. R. Caldwell, and L. B. Crowder. 2014. Harnessing DNA to improve environmental management. Science 344:1455-1456.
17. Kim, K. C., and L. B. Byrne. 2006. Biodiversity loss and the taxonomic bottleneck: emerging biodiversity science. Ecological Research 21:794-810.
18. Lehtiniemi, M., H. Ojaveer, M. David, B. Galil, S. Gollasch, C. McKenzie, D. Minchin, A. Occhipinti-Ambrogi, S. Olenin and J. Pederson. 2015. Dose of truth – Monitoring marine non-indigenous species to serve legislative requirements. Marine Policy 54: 26-35.
19. Leppäkoski, E., S. Gollasch, P. Gruszka, H. Ojaveer, S. Olenin and V. Panov. 2002. The Baltic a sea of invaders. Canadian Journal of Fisheries and Aquatic Sciences 59: 1175-1188.
20. Michalek, M., and M. Werner. 2012. Abundance and distribution of Marenzelleria species in the Baltic Sea. in. HELCOM Baltic Sea Environment Fact Sheets. http://www.helcom.fi/baltic-sea-trends/environment-fact-sheets/.
21. Neigel, J., A. Domingo, and J. Stake. 2007. DNA barcoding as a tool for coral reef conservation. Coral Reefs 26:487-499.
22. Ojaveer, H., A. Jaanus, B.R. MacKenzie, G. Martin, S. Olenin, T. Radziejewska, I. Telesh, M.L. Zettler, and A. Zaiko. 2010. Status of Biodiversity in the Baltic Sea. PLoS ONE Collection: Marine Biodiversity and Biogeography Regional Comparisons of Global Issues 5: e12467.
23. Oksanen, J., F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R.B. O'Hara, G.L. Simpson, P. Solymos, M.H.H. Stevens, H. Wagner. 2014. Vegan: community ecology package version 22-0. Available from: http:// cran.r-project.org/web/packages/vegan/vegan.pdf
24. Olenin, S., M. Elliott, I. Bysveen, P. F. Culverhouse, D. Daunys, G. B. J. Dubelaar, S. Gollasch, P. Goulletquer, A. Jelmert, Y. Kantor, K. Bringsvor Mezeth, D. Minchin, A. Occhipinti-Ambrogi, I. Olenina, and J. Vandekerkhove. 2011. Recommendations on methods for the detection and control of biological pollution in marine coastal waters. Marine Pollution Bulletin 62:2598-2604.
25. Patrício J., Little S., Mazik K., Thomson S., Zampoukas N., Teixeira H., Solaun O., Uyarria M., Papadopoulou N., Kaboglu G. et al. 2014. SWOT Analysis of Monitoring Networks in Europe. DEVOTES Deliverable 1.4 (Public Report), 100 pp.
26. Pochon, X., N. J. Bott, K. F. Smith, and S. A. Wood. 2013. Evaluating detection limits of Next-Generation Sequencing for the surveillance and monitoring of international marine pests. PLoS One 8:e73935. doi:73910.71371/journal.pone.0073935.

27. Ratnasingham, S., and P. D. N. Hebert. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). Molecular Ecology Notes 7:355-364.
28. Ratnasingham, S., and P. D. N. Hebert. 2013. A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. PLoS ONE 8(8): e66213.
29. Rilov G, Crooks JA (2009) Biological invasions in marine ecosystems: ecological, management, and geographic perspectives, Vol 204. Springer, Berlin, Germany.
30. Schloss, P.D., S. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, B.B. Oakley, D.H. Parks, C.J. Robinson, J.W. Sahl, B. Stres, G.G. Thallinger, D.J. Van Horn, and C.F. Weber. 2009. Introducing mother: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75(23): 7537-7541.
31. Schmieder, R., R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863-864.
32. Taberlet, .P, E. Coissac, M. Hajibabaei, L.H. Rieseberg. 2012. Environmental DNA. Molecular Ecology 21:1789–1793.
33. Telesh, I., L. Postel, R. Heerkloss, E. Mironova, and S. Skarlato. 2009. Zooplankton of the Open Baltic Sea: Extended Atlas. Marine Science Reports 76, Leibniz institute for Baltic Sea Research, Warnemunde.
34. Wood, S. A., K. F. Smith, J. C. Banks, L. A. Tremblay, L. Rhodes, D. Mountfort, S. C. Cary, and X. Pochon. 2013. Molecular genetic tools for environmental monitoring of New Zealand's aquatic habitats, past, present and the future. New Zealand Journal of Marine and Freshwater Research 47:90-119.
35. Zaiko, A., M. Lehtiniemi, A. Narscius, and S. Olenin. 2011. Assessment of bioinvasion impacts on a regional scale: a comparative approach. Biological Invasions 13:1739-1765.
36. Zaiko, A., J.L. Martinez, A.Ardura, L. Clusa, Y.J. Borrell, A. Samuiloviene, A. Roca, E. Garcia-Vazquez. 2015a. Detecting nuisance species using NGST: methodology shortcomings and possible application in ballast water monitoring. Marine Environmental Research 1-9, in press.
37. Zaiko, A., J.L. Martinez, J. Schmidt-Petersen, D. Ribicic, A. Samuiloviene and E. Garcia-Vazquez. 2015b. Metabarcoding approach for the ballast water surveillance – An advantageous solution or an awkward challenge? Marine Pollution Bulletin, doi:10.1016/j.marpolbul.2015.01.008.
38. Zaiko, A., D. Minchin, and S. Olenin. 2014. "The day after tomorrow": anatomy of an 'r' strategist aquatic invasion. Aquatic Invasions 9:145-155.

Zaiko A., Samulioviene A., Ardura A., Garcia-Vazquez E. 2015. Metabarcoding approach for non-indigenous species surveillance in marine coastal waters. Marine Pollution Bulletin, 100:53-59.
doi:10.1016/j.marpolbul.2015.09.030
Access to underlying data: http://doi.pangaea.de/10.1594/PANGAEA.853664

Table 2. Taxa identified from metabarcoding samples using high-throughput sequencing (HTS) and Lithuanian national monitoring (LNM) samples at three sampling locations: KS – Klaipeda Strait, SG – Sea Gates, PA – northern edge of the Plume Area. Genera or species found by both methods shown in bold. An asterisk denotes species reported as non-indigenous in the Baltic Sea.

Phylum	Class	Order	Family	Genus	Species	HTS			LNM		
						KS	SG	PA	KS	SG	PA
Annelida	Polychaeta	Spionida	Spionidae	*Marenzelleria*	*neglecta**	-	+	+	-	-	-
					*viridis**	+	+	+	-	-	-
Arthropoda	Branchiopoda	Diplostraca	Bosminidae	***Bosmina***	***unclassified***	+	+	+	+	-	-
				Eubosmina	*longispina*	-	-	-	-	+	+
			Cercopagididae	***Cercopagis***	***pengoi****	-	+	-	-	+	+
			Chydoridae	***Chydorus***	***sphaericus***	+	-	-	+	-	-
					unclassified	+	-	-	-	-	-
			Daphniidae	***Daphnia***	***cucullata***	+	-	-	+	-	-
					galeata	+	-	-	+	-	-
					unclassified	+	-	-	+	-	-
			Leptodoridae	***Leptodora***	***kindtii***	+	-	-	+	-	-
			Podonidae	***Evadne***	*nordmanni*	-	-	-	-	-	+
					spinifera	+	+	+	-	-	-
					unclassified	+	+	+	-	-	-
				Pleopis	*polyphemoides*	+	+	+	-	-	-
					unclassified	+	+	+	-	-	-
				Podon	***leukartii***	-	+	+	-	+	-
					unclassified	-	+	+	-	-	-
			Sididae	*Diaphanosoma*	*brachyurum*	-	-	-	+	-	-
			unclassified	*unclassified*	*unclassified*	-	+	+	-	-	-
	Maxillopoda	Calanoida	Acartiidae	***Acartia***	***longiremis***	-	+	-	+	+	+
					*tonsa**	+	+	-	-	-	-
					unclassified	-	+	-	+	+	+
			Centropagidae	***Centropages***	*hamatus*	+	+	+	-	-	-
					unclassified	-	-	-	-	-	+
			Diaptomidae	*Diaptomus*	*unclassified*	-	-	-	+	-	-

				Eudiaptomus	*gracilis*	-	-	-	+	-	-
			Paracalanidae	*Paracalanus*	*unclassified*	-	-	-	-	+	-
			Temoridae	***Temora***	*longicornis*	+	+	+	-	-	-
					unclassified	-	-	-	-	-	+
		Cyclopoida	Cyclopidae	*Mesocyclops*	*leuckarti*	+	+	-	-	-	-
				unclassified	*unclassified*	-	-	-	+	+	+
			Oithonidae	*Oithona*	*similis*	-	-	-	+	-	-
		Harpacticoida	Phyllognathopodidae	*Phyllognathopus*	*paludosus*	-	-	-	+	-	-
	unclassified	unclassified	unclassified	*unclassified*	*unclassified*	+	+	+	-	-	-
Mollusca	Bivalvia	Mytiloida	Mytilidae	*Mytilus*	*unclassified*	-	+	-	-	-	-
		Veneroida	Dreissenidae	*Dreissena*	*polymorpha**	+	-	-	-	-	-
	Gastropoda	Littorinimorpha	Hydrobiidae	*Hydrobia*	*ulvae*	+	-	-	-	-	-
Rotifera	Monogononta	Flosculariaceae	Conochilidae	*Conochilus*	*unicornis*	-	-	-	+	-	-
			Trochosphaeridae	*Filinia*	*longiseta*	-	-	-	+	-	-
		Ploima	Asplanchnidae	*Asplanchna*	*priodonta*	-	-	-	+	+	+
			Brachionidae	***Brachionus***	***calcyciflorus***	+	-	-	+	+	-
				Kellicottia	*longispina*	-	-	-	+	+	-
				Keratella	***cochlearis***	+	+	+	+	+	+
					quadrata	+	+	+	+	+	+
					unclassified	-	+	-	-	-	-
				unclassified	*unclassified*	+	+	-	-	-	-
			Synchaetidae	*Polyarthra*	*sp.*	-	-	-	+	+	-
				Synchaeta	*monopus*	-	-	-	+	-	+
					unclassified	+	+	+	-	-	-

Zaiko A., Samulioviene A., Ardura A., Garcia-Vazquez E. 2015. Metabarcoding approach for non-indigenous species surveillance in marine coastal waters. Marine Pollution Bulletin, 100:53-59. doi:10.1016/j.marpolbul.2015.09.030
Access to underlying data: http://doi.pangaea.de/10.1594/PANGAEA.853664

Table 3. PERMANOVA results showing of statistical difference in taxonomic composition (A- at a species level; B- at a family level) between two methods, three sampling sites and sampling events (sampling time nested within the site level). Iitalicized values indicate statistically significant effects. Based on the presence-absence transformations, Jaccard similarity matrix.

A

Source	df	SS	MS	Pseudo-F	P	Unique perms
Method	1	10250	10250	5.46	*0.03*	997
Site	2	6824	3412	1.86	0.08	971
Time(Site)	6	13870	2312	2.73	*0.01*	998
Method x Site	2	4587	2294	1.21	0.41	999
Method x Time(Site)	2	3948	1974	2.33	0.08	999
Residual	4	3389	847			
Total	17	50026				

B

Source	df	SS	MS	Pseudo-F	P	Unique perms
Method	1	3751	3751	3.26	0.09	996
Site	2	6178	3089	1.59	0.15	979
Time(Site)	6	15009	2501	3.24	*0.004*	999
Method x Site	2	2758	1379	1.19	0.42	999
Method x Time(Site)	2	2375	1188	1.54	0.25	999
Residual	4	3089	772			
Total	17	38043				

Zaiko A., Samulioviene A., Ardura A., Garcia-Vazquez E. 2015. Metabarcoding approach for non-indigenous species surveillance in marine coastal waters. Marine Pollution Bulletin, 100:53-59. doi:10.1016/j.marpolbul.2015.09.030
Access to underlying data: http://doi.pangaea.de/10.1594/PANGAEA.853664

Supplementary material:

Suppl. 1: FASTA-formatted list of OTUs classified based on the HTS data analysis with the corresponding consensus sequence.