High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems

Franck Lejzerowicz, Philippe Esling, Loïc Pillet, Thomas A. Wilding, Kenneth D. Black, Jan Pawlowski

SCIENTIFIC REP⚙RTS

High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems

Franck Lejzerowicz[1,†], Philippe Esling[1,2], Loïc Pillet[1,3], Thomas A. Wilding[4], Kenneth D. Black[4] & Jan Pawlowski[1]

Environmental diversity surveys are crucial for the bioassessment of anthropogenic impacts on marine ecosystems. Traditional benthic monitoring relying on morphotaxonomic inventories of macrofaunal communities is expensive, time-consuming and expertise-demanding. High-throughput sequencing of environmental DNA barcodes (metabarcoding) offers an alternative to describe biological communities. However, whether the metabarcoding approach meets the quality standards of benthic monitoring remains to be tested. Here, we compared morphological and eDNA/RNA-based inventories of metazoans from samples collected at 10 stations around a fish farm in Scotland, including near-cage and distant zones. For each of 5 replicate samples per station, we sequenced the V4 region of the 18S rRNA gene using the Illumina technology. After filtering, we obtained 841,766 metazoan sequences clustered in 163 Operational Taxonomic Units (OTUs). We assigned the OTUs by combining local BLAST searches with phylogenetic analyses. We calculated two commonly used indices: the Infaunal Trophic Index and the AZTI Marine Biotic Index. We found that the molecular data faithfully reflect the morphology-based indices and provides an equivalent assessment of the impact associated with fish farms activities. We advocate that future benthic monitoring should integrate metabarcoding as a rapid and accurate tool for the evaluation of the quality of marine benthic ecosystems.

Aquaculture is a rapidly growing industry[1], which impact on marine benthic ecosystems needs to be evaluated quickly and efficiently[2]. This is traditionally done using physico-chemical measurements and the response of benthic biological communities[3,4]. The latter approach is referred to as benthic monitoring and consists of making morphotaxonomic inventories of macro-invertebrates from which various indices are calculated[5]. Beyond generic alpha-diversity measures such as the Shannon diversity H' or species richness S, specific biotic indices have been formalized in order to ascribe samples into environmental quality classes. These indices include the Infaunal Trophic Index (ITI[6]), the AZTI Marine Biotic Index (AMBI[7]), the Norwegian Sensitivity and Quality Indices (NQI1[8], NSI[9]) or the Enrichment Stage index (ES[10]). Their formulas include taxon- or cohort-specific weights empirically defined from the autecology of macrofaunal species. The rapid development of salmon farming activities led the main producing countries (Norway, Scotland, Canada, New Zealand) to adopt specific regulations using different reference biotic indices: ITI in Scotland, NSI, NQI1 and AMBI in Norway, ES in New Zealand.

[1]Department of Genetics and Evolution, University of Geneva, Switzerland. [2]IRCAM, UMR 9912, Université Pierre et Marie Curie, Paris, France. [3]ADMM UMR 7144, CNRS, Station Biologique de Roscoff, 29682 Roscoff, France. [4]SAMS, Scottish Marine Institute, Oban, Argyll, UK. †Present address: Department of Genetics and Evolution, University of Geneva, Sciences 3, 30, Quai Ernest Ansermet, CH-1211 Geneva 4, Switzerland. Correspondence and requests for materials should be addressed to F.L. (email: franck.lejzerowicz@unige.ch)

The realization of morphotaxonomic inventories involves the morphological identification of numerous sorted specimens, which is extremely time consuming and taxonomic-expertise demanding. As results take typically several months, it is not possible to respond in a timely manner for effective adaptive management. High-throughput sequencing (HTS) of taxonomic markers enriched from environmental DNA (eDNA) (i.e. metabarcoding) offers an alternative to the morphotaxonomy-based biomonitoring[11–13]. This approach has already been used extensively for exploring the microbial and meiofaunal diversity in various environments[14]. It has also been successful for assessing the quality of freshwater environments, based on the HTS of diatoms[15,16] and aquatic insects[17,18]. However, only few studies examined the application of metabarcoding to marine ecosystems[19–21].

The west coast of Scotland is characterized by sheltered sea lochs, which have been exploited by salmon farmers since the 1980s. Salmon farms consist of a series of nets or pens hanging 10–20 m below the sea surface. Fish farms typically consisting of between 4 and 20 pens are located in areas sheltered from severe storms, but exposed to moderate current flows. The pens are usually aligned with the predominant current flow. Fish-faeces, and uneaten fish-feed, fall down through the water column and accumulate on the seabed around the farm, usually in an ellipsoid shape with the major axis occurring in the direction of the main current. The culture of salmon impacts the benthic environment primarily as a consequence of the accumulation of farm-related detritus (uneaten feed, faeces) around the farm. The detritus increases the biological oxygen demand of the sediment and, if this demand is not met, the sediment becomes hypoxic. Sedimentary hypoxia following organic enrichment is typically associated with the replacement of relatively few large, long-lived, burrowing species by numerous small, short-lived opportunistic species[22,23].

Here, we compare the metabarcoding and morphological approaches in their ability to indicate environmental quality gradients occurring around fish farms. The morphotaxonomic approach is usually restricted to benthic macrofaunal taxa whereas the sequencing of eDNA/RNA molecules can extend the taxonomic analysis by including the meiofauna. From both morphotaxonomic inventories and normalized eDNA/RNA sequence data, we reconstruct both the ITI and AMBI indices for 10 stations located at different distances from salmon cages. We compared the indices inferred from molecular and morphotaxonomic diversity datasets and evaluate how these two different views on the benthic communities impact the assessment of the quality of environmental samples.

Material and Methods

Sampling. A total of 10 macrobenthic stations were sampled (Supplementary Table 1). Station 1–9 were distributed along a transect (bearing 240°), extending 400 m from the most southerly (cage centre: −5.500, 56.502, decimal °, WGS84) of 9 circular salmon cages located on the east side of the Isle of Lismore, on the west coast of Scotland. The samples of stations 1–9 were taken in-line with the cages and with the dominant current-flow. Station 10 was situated perpendicularly to the other samples (bearing 135°) but since the dispersion of detritus around cages is elliptical along the water currents axis, it was treated as distant station.

At each station, one macrobenthic sample was collected using a Van-veen grab, from which the redox was measured, five sediment replicates were subsampled for metabarcoding, and the remaining sediment (i.e. the 2 first centimetres over a 0.1 m² area) was treated for morphotaxonomic inventory. The location of each grab was recorded by noting the position of the boat's A-frame (via a dedicated A-frame mounted dGPS aerial) from which the grab was lowered vertically to the seabed. The position was noted as soon as the grab reached the bottom (as indicated by a slackening of the winch wire), the survey vessel regaining its position (if necessary) prior to recovery. The distance to the fish-cage was determined using the boat's radar.

Redox was measured immediately following collection using a redox probe (Model CMPtr 106/300 mm; Russel pH Ltd, Auchtermuchty, UK). Prior to use, the probe was checked against a standard solution[24]. The probe was inserted 10 mm into the sediment and the redox value recorded once the reading had stabilized (generally after two to three minutes). Whilst the redox was being measured, the five sediment replicates were sub-sampled from the top 2 cm of the grab using disposable spatulas and immersed into 6 ml of LifeGuard Preservation Solution (MoBio), in order to preserve labile RNA molecules. Once sub-sampling had been completed the sediment was washed through a 1 mm sieve and the residue fixed in 4% borax-buffered formaldehyde prior to macrobenthic sorting and counting. The sieve-retained fauna were identified to species level under the National Marine Biological Quality Control Scheme (NMBAQCS)[25] by Myriad Taxonomy (Campbeltown, Argyll).

Molecular analyses. We extracted the total environmental RNA and DNA content of each of the fifty sub-samples using the PowerSoil RNA kit in combination with the DNA Elution Accessory kit (MoBio), according to the manufacturer instructions. The RNA molecules were treated to remove carried-over DNA contaminants and reverse-transcribed to obtain complementary DNA (cDNA) as previously[21]. Then, we enriched the 50 DNA and 50 cDNA extracts for the V4 region of the SSU rRNA gene by PCR amplification. The PCR were realized with the eukaryotic primers pair TAReuk454FWD1 (5′ – CCAGCASCYGCGGTAATTCC – 3′) and TAReukREV3 (5′ – ACTTTCGTTCTTGATYRA – 3′) according to previously published thermo-cycling conditions[26] and PCR reactors[21]. We used tagged PCR primers to label and multiplex PCR products in one HTS library (Supplementary Table 2) following

previous primer design and workflow[21] and according to a desaturated Latin Square Design in order to reduce the impact of sequence-to-sample misidentifications[27]. We then quantified the amount of amplicons generated by each reaction using relative gel electrophoresis band intensities in order to pool the PCR products in equimolar quantities. We prepared one HTS library from the pool of PCR products according to the instructions of the TruSeq Nano DNA LT Sample Prep kit (Illumina). We then sequenced the resulting library on a MiSeq instrument for 502 cycles (251 cycles paired-end) using the MiSeq Reagent Nano Kit v2.

Bioinformatics.   We quality-filtered and assembled the paired-end reads into full-length sequences following the stringent approach described previously[21]. Then, we performed the de-multiplexing of these sequences into their samples of origin. During the de-multiplexing, we filtered sequence-to-sample misidentifications (i.e. cross-contaminants) due to the mistagging phenomenon as described in the recently published method accounting for unexpected tagged primers[27]. We then dereplicated the filtered set of sequences into Individual Sequence Units (ISUs) and we removed singletons. We considered as singletons every ISU represented by only one read throughout the entire dataset (i.e. we would keep an ISU represented by one read in more than one sample). Then, we extracted all ISUs matching any entry of a subset of the PR2 reference database[28] containing all Metazoa V4 sequences (23,999 records). We performed BLASTn v. 2.2.25+ searches[29] as follows: blastn –word_size 20 –max_ target_seqs 50 –perc_identity 70 –strand plus. We then used MOTHUR v.1.33.3[30] to compute pairwise global alignments (Needleman-Wunsch algorithm) and we built Operational Taxonomic Units (OTUs) using a 3% sequence dissimilarity threshold (average linkage clustering). We chose the threshold of 3% in order to avoid inflated diversity estimates, as it has been shown that the 18S rDNA marker reduces the magnitude of diversity estimates, particularly for the meiofauna[31]. We removed the chimeric OTUs originating from the artificial recombination of different sequences by manual inspection of all the candidates identified by Uchime v.4.2[32] in both "self" and "reference" modes using the following parameters: –abskew 1 –minh 0.3 –xn 5 –minchunk 32.

We then assigned the OTU reference sequences using another round of BLAST searches and phylogenetics. Briefly, we kept the taxonomic consensus of all metazoan reference sequences that best match an OTU sequence along decreasingly stringent combinations of identity (from 100 to 90%) and coverage (100 to 80%) thresholds for all BLAST high scoring pairs (HSPs). If no genus or species could be assigned using HSPs, we used PhyML v.3.0[33] to build trees from all metazoan reference sequences matching the OTU sequence – but this time along increasingly stringent thresholds – until a supported clade (bootstrap superior or equal to 80/100) containing the OTU sequence was found. In each tree we incorporated three extra sequences belonging to the closest family (or order depending on the taxonomies of the BLAST results) and sharing less than 30% identity with the OTU sequence. For the assignment of the OTU, we kept the taxonomy shared among all the reference sequences constituting the supported clade, but only if we obtained a more precise taxonomy than the taxonomy obtained after the BLAST search.

Biotic indices.   For both morphological and molecular data (but separately for DNA and RNA), we calculated the Infaunal Trophic Index (ITI) as well as three alternative versions of the AMBI[7]. This allows comparing the performance of using either the sequence abundance (H-AMBI) or the OTU richness (S-AMBI) information[34], or both as it is commonly calculated from morphotaxonomic data (M-AMBI). The taxon-specific bioindicator values for ITI and AMBI were extracted from previous works[35–37] and from the AZTI software v.5.0, respectively. Because of the uneven distribution of sequences among the samples, we performed a normalization of the OTU-to-sample dataset prior to the indices computations[38]. Briefly, we randomly subsampled the OTUs of each sample replicate 100 times (with replacement), picking a number of reads corresponding to the median of the number of reads per sample (n = 4102). We kept the average number of reads per OTU and no OTUs represented by less than 1.01 reads.

Results
Morphotaxonomic analyses.   In total, 18,351 specimens representing 116 taxa (including 98 genera) were sorted from 10 grab samples (Supplementary Table 3). On average (±standard deviation), 17.2 (±6.67) species occur at stations close to the cages (within 60 meters), and 47.25 (±8.6) species occur at remote stations. The number of specimens ranges from 366 to 7083 in AZE stations (st. 1 to 6), and from 217 to 320 in distant stations. Both the Shannon H' and Pielou J indices indicate a lower diversity close to cages (H' = 0.77 ± 0.24 and J = 0.27 ± 0.07) as compared to remote stations (H' = 3.2 ± 0.08 and J = 0.83 ± 0.035). The benthic communities in AZE stations are dominated by the annelids: *Capitella spp*, (76.7 ± 11%), *Tubificoides benedii* (1.66 ± 2.13%) and *Malacoceros fuliginosus* (3.03 ± 2.55%) and unidentified nematods (15.8 ± 9.4%). In distant stations (st. 7 to 10), only five specimens belonging to these taxa could be found. The seven species which dominate in the distant stations belong to diverse phyla, including one gastropod (10.79 ± 4.40%), one bivalve (8.47 ± 2.99%), one Echinoderm (5.03 ± 2.31%) and four annelids orders: Capitellida (4.34 ± 2.41%), Terebellida (8.57 ± 2.40%), Spionida (1.93 ± 0.71%) and Phyllodocida (9.15 ± 1.10%).
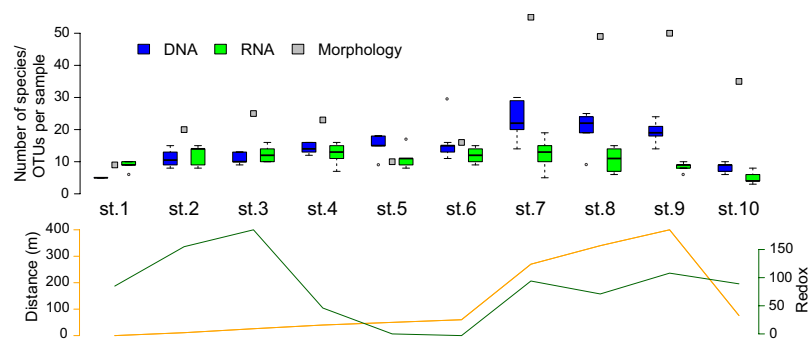
Figure 1. Morphological species and molecular OTU richness. For each station the number of taxa identified morphologically (grey squares) as well as the number of OTUs obtained using DNA (blue boxes) and RNA (green boxes) are shown. The boxplots entail up to 5 samples (PCR replicates) per station for the molecular data. In the lower panel the distance to the cage in meters (yellow) and the redox value (green) for each station are indicated.

HTS data statistics. We obtained about 4.5 million eukaryotic reads distributed across 100 samples (5 DNA and 5 cDNA replicates at 10 sampling stations), from which a subset of 583,574 and 295,727 sequence reads correspond to Metazoa, for DNA and RNA respectively (Supplementary Table 4). We discarded five DNA samples including four from station 1 and one from station 2 because after filtering they contained no sequence or less than three metazoan sequences, respectively. We found significantly more metazoan sequences in the DNA than in the RNA samples (Friedman rank sum test excluding samples 1 and 2 because of missing sample pairs, p-value = 0.008), as well as in two out of three stations situated far from the cages (stations 7 and 9) (Pairwise Wilcoxon Mann-Whitney tests with FDR correction: p-value = 0.036, Supplementary Fig. 1 and Supplementary Table 5). Similarly, the OTU richness is systematically higher in DNA samples than in RNA samples, irrespective of the threshold used for OTUs clustering (Friedman rank sum test excluding samples 1 and 2 because of missing sample pairs, p-value: $1.54 \ 10^{-5}$).

We used the sequence dataset of 163 OTUs clustered at 3% dissimilarity for further analyses of metazoan communities. The OTU richness is significantly higher at remote stations than close to the cages, but only for DNA samples (Fig. 1, Supplementary Table 6). For the RNA data, the OTU richness does not show any particular pattern, except for the most distant station 9 appearing less OTU-rich than the stations close to the cages. Interestingly, even in the DNA data the increase of the OTU richness observed in distant stations is much lower than the number of morphologically identified species.

Taxonomic composition. The molecular assemblage of 163 metazoan OTUs is dominated by annelids (28.2%), Platyhelminthes (20.8%), nematodes (17.8%) and arthropods (14.1%). Other major metazoan phyla, such as molluscs, cnidarians or echinoderms are represented by relatively few OTUs (from 1.22 to 4.29%). Four phyla (Bryozoa, Entoprocta, Priapulida and Sipuncula) occurring in the morphotaxonomic inventory could not be detected in the HTS data. Although represented with reduced OTU diversity, we could detect the presence of Hemichordates and several small-sized phyla (Gastrotricha, Kinorhyncha and Rotifera) that are not reported in the morphotaxonomic study. The most striking discrepancy between morphological and molecular data is that numerous OTUs could be assigned to Platyhelminthes and Acoelomorpha whereas these meiofaunal taxa are not included in the morphotaxonomic inventories (Supplementary Tables 3 and 7).

There are also important differences between morphotaxonomic and metabarcoding analyses at lower taxonomic levels. Although the annelids dominate both assemblages, their richness at the family level is much lower in the HTS data (12 families) compared to the morphological inventory (26 families). This difference is even higher for molluscs, with only 7 genera detected with HTS versus 24 genera with morphological examination. Similarly, none of the 11 Malacostraca species identified morphologically are present in the HTS crustacean assemblage dominated by the copepods and ostracods. Yet, the proportions of the taxa that can be found using both techniques are fairly similar. For instance, out of 11 shared genera, only the proportions of Arthropoda and Mollusca genera are highly skewed (Supplementary Fig. 2). This is also the case for the 15 shared families, that are represented by a majority of the OTUs (60.2%) and species (53.5%) assigned to a family.

In spite of these differences, the congruence between morphological and HTS data is high for the most abundant morphotaxa. The genus *Capitella* by far dominates both morphological and molecular datasets. The next three most abundant morphotaxa (Nemertea, *Malacoceros fuliginous*, *Tubificoides*) represented by more than 200 specimens are also present in the HTS data. Nevertheless, the proportion of taxa present in both morphological and HTS datasets decreases rapidly for the rare ones. In total, less than 20% of morphotaxa are found in the HTS data.
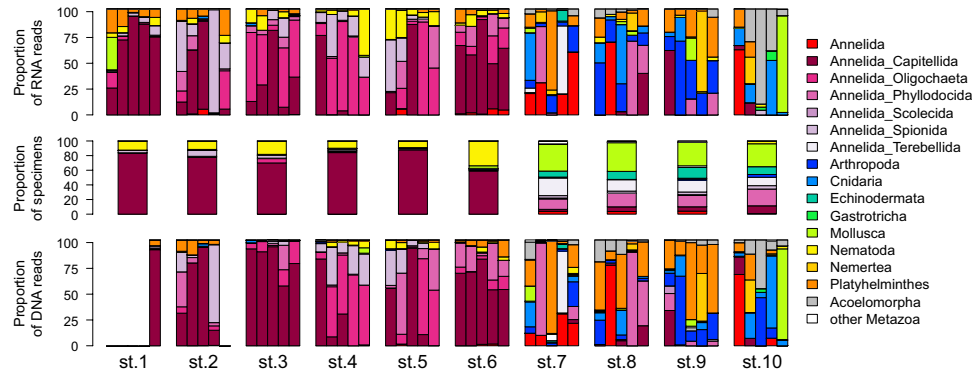
Figure 2. Taxonomic composition. For each of the samples grouped per station the proportions of taxa identified morphologically (middle panel) as well as using RNA (upper panel) and DNA (lower panel) are shown. The proportions of taxa at the phylum level and at the order level for Annelida taxa are shown.
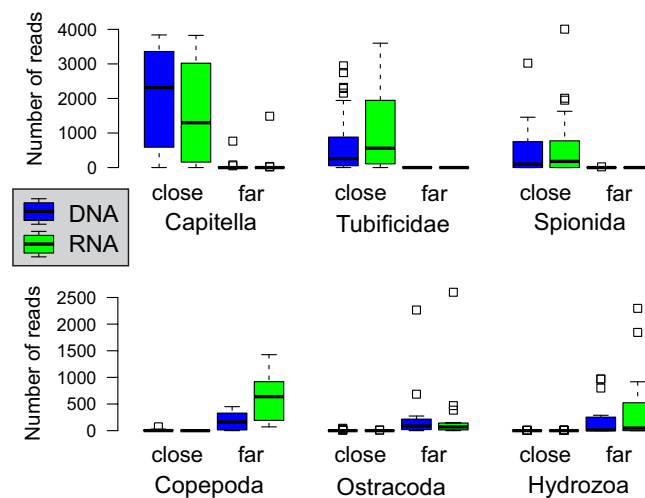
Figure 3. Sequence reads abundances of selected bioindicator taxa binned per class of distance to the cage. The two bins correspond to the stations found at or closer than 60 meters from a cage (close) and to the stations located over 60 meters from a cage (far). The numbers of DNA- (blue boxes) and RNA- (green boxes) derived reads are shown separately.

Metazoan OTUs distribution. We analyzed the distribution of OTUs in different stations, separately for DNA and RNA, and compared it to the morphotaxonomic inventories (Fig. 2). The metazoan assemblage in AZE samples collected close to the cages is very different from the distant samples. In the AZE samples the dominant taxa are *Capitella*, Tubificidae, *Malacoceros* (Spionida), an unassigned species of Cirratulidae (OTU6) and the nematodes (Fig. 3). The genus *Capitella* is present in all AZE samples where it accounts for up to 93.7% of sample sequences. However, it is rare in distant samples, with relative sequence abundances usually lower than 2% (excepted in one sample of station 9). The main Tubificidae, Cirratulidae and *Malacoceros* OTUs are also abundantly sequenced in AZE samples, representing up to 87.8%, 48.8% and 97.6% of sample sequences, respectively. Interestingly, the presence of Cirratulidae is restricted from 25 to 60 m off the cage while *Malacoceros* only occur in the samples within 11 m. Both OTUs are absent from distant samples. The nematode OTUs exceptionally reach the abundance of 43.7% in a sample, but in general never exceed 10% in AZE samples (on average 5.04 ± 7.78% SD) and 5% in distant samples.

The distant samples are characterized by highly diversified assemblages compared to the AZE samples. The replicates taken at the same distant station rarely present the same taxonomic composition, as evidenced by sample dissimilarities computed based on the presence/absence of OTUs (Supplementary Fig. 3). When compared along with the distant samples, the AZE samples all seem similar, but the replicates remain more similar when the sequence abundance information is used (especially for DNA). Moreover, there is a good congruence between the sequence abundance observed in DNA and RNA datasets (Fig. 2).
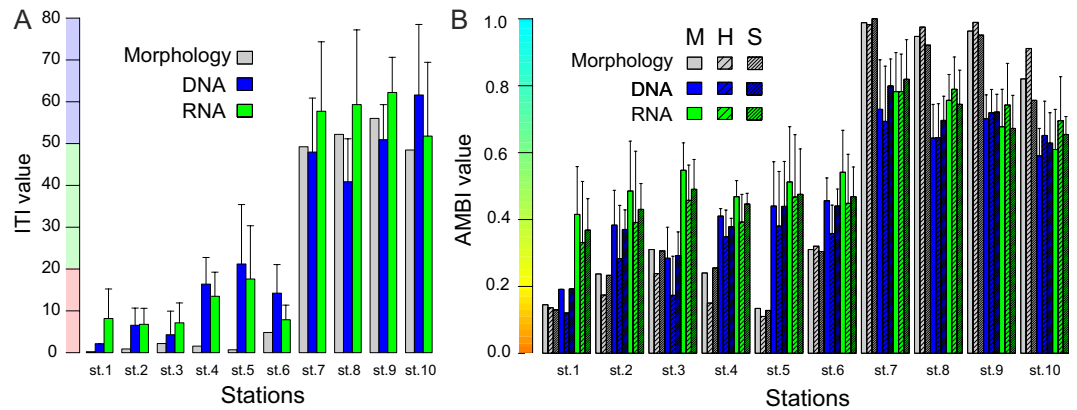
Figure 4. Infaunal Trophic Index (ITI) and AZTI's Marine Biotic Index (AMBI). (**A**) Only one ITI value could be calculated for the morphologically derived values (grey bars) while the mean value of up to 5 samples PCR replicates is shown for both DNA (blue bars) and RNA (green bars). For the molecular data, the error bars indicate the standard deviation obtained after 1000 ITI calculations based on the OTUs diversity of one randomly picked sample per station out of up to 5 PCR replicate samples sequenced per station. For the AMBI calculations, the error bars indicate the standard deviation across the sample replicates. The coloured background indicates the three discrete environmental quality values corresponding to ITI values. (**B**) For each station is indicated the AMBI values calculated based either on the Shannon diversity (H-AMBI; sparsely hatched bars), the species richness (S-AMBI; densely hatched bars) or both (M-AMBI; plain bars). The ramping coloured background indicates the continuous environmental quality depicted by the AMBI.

The distant samples are rarely dominated by a single OTU, except the OTU assigned to the bivalve genus *Corbula*, which comprises 91% of one of the replicates of the station 10. The taxonomic groups most commonly sequenced in the distant samples are Copepods, Ostracods, Hydrozoans, Nemerteans, Acoelomorpha and Platyhelminthes (Fig. 3). Some of these groups are totally absent from AZE samples (Acoelomorpha, Nemertea), while others can be found but with relative sequence abundances generally below 1% in a sample. Interestingly, some OTUs can be very abundant in a single distant sample and absent (Ampharetidae OTU21) or moderately common (*Phyllodoce maculata*) in other samples.

The distribution of OTUs matches relatively well the morphotaxonomic inventories. In morphological counts, the AZE samples are dominated by Capitellids (58.5–87.8%) like in the HTS data. The Tubificidae and Spionidae (*Malacoceros*) are also present in both datasets, but their abundance is much higher in the HTS data. In contrast, the nematodes show reverse pattern, being much more abundant in morphological counts (9.1–33.3%) than in the DNA/RNA samples from the AZE stations. This high congruence between morphological and HTS data in AZE stations is clearly less pronounced in the distant samples. The morphological inventories of these samples are largely dominated by molluscs (31.3–39.3%) and ophiuroids (10.7–14.7). Both groups are represented by few OTUs, which relative abundance is generally low (except *Corbula gibba)*. Several taxonomic groups represented by the OTUs common in distant stations, such as Hydrozoa or Acoelomorpha, are absent from morphotaxonomic inventories. On the other hand, some abundant OTUs, particularly within annelids, possibly correspond to the sorted morphospecies but could only be identified to the family level.

Biotic indices. The ITI and AMBI (M-AMBI, H-AMBI and S-AMBI) values inferred from the molecular data reflect similar ecological conditions to the corresponding values inferred from the reference morphological data (Fig. 4). The morphology-based ITI values are extremely low for the AZE stations and very high for the distant stations, indicating a clear separation between the strongly impacted conditions (ITI < 20) of the former and the low impact of the fish farms on the benthic communities living father than 300 m from the cages (ITI > 50). The same clear-cut difference between AZE and distant stations are observed in values of AMBI, regardless of whether the index calculation is based on sequence abundance (H-AMBI), OTU richness (S-AMBI) or both (M-AMBI) and of the fact that only half of the sequenced taxa are associated with an AMBI ecological group (Supplementary Fig. 4). Interestingly, the station 10, situated on the other side of the fish farm, at distance of 76 m, shows ITI and AMBI values similar to the distant stations, presumably because this station was oriented perpendicularly to the direction of the residual current, corresponding to the main depositionary axis of the fish farm.

The correlation between values inferred from morphological and molecular data is very high for both ITI (DNA: $R^2 = 0.866$, RNA: $R^2 = 0.974$) and AMBI indices (H-AMBI DNA: $R^2 = 0.821$, RNA: $R^2 = 0.898$; S-AMBI DNA: $R^2 = 0.899$, RNA: $R^2 = 0.855$; M-AMBI DNA: $R^2 = 0.811$, RNA: $R^2 = 0.868$) (Fig. 5). In the case of AZE stations, the values of ITI and AMBI indices inferred from DNA/RNA data are higher than
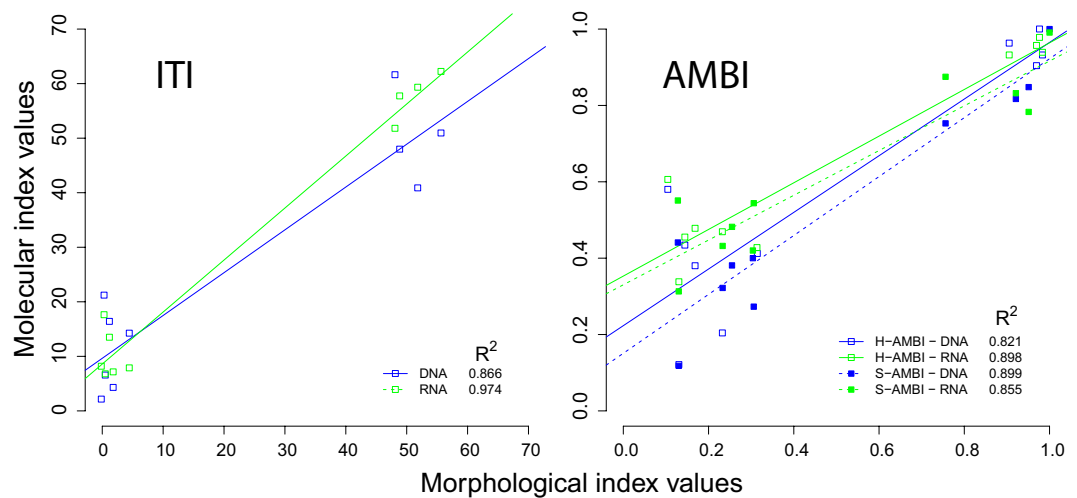
Figure 5. Relationship between biotic index values derived morphological and molecular data. The index values are shown separately for the ITI (left panel) and AMBI (right panel) indices. The relationships are represented for both the H-AMBI and S-AMBI calculations. Linear regressions and associated coefficients of determination are indicated. All correlations are highly significant (p < 0.001).

those based on morphological analyses. This difference is particularly pronounced in the case of station 5 (50 m off the cage), with ITI, H-AMBI, S-AMBI and M-AMBI values inferred from DNA being 30.8, 3.45, and 3.4 and 3.3 times higher, respectively. In general, the correlation seems better in the case of RNA than DNA. Interestingly, the index values inferred from RNA are higher than those inferred from DNA data in the three closest stations from the cage (1 to 3) as well as in the three distant stations (7 to 9) but only for ITI.

Discussion

This study confirms the usefulness of metabarcoding to estimate the biotic indices routinely used in benthic monitoring of marine ecosystems. The outcome of the traditional morphotaxonomic approach is similar to the HTS eDNA approach, even though both involve very different sampling volumes and rely on the contrasting diversity of different set of taxa. Our results are promising but need to be interpreted with caution in order to understand the challenges of the new approach and fully appreciate its potential.

Applying metabarcoding for benthic monitoring offers numerous practical advantages. Current developments towards automation and reduction of analytic steps will greatly simplify the use of DNA sequences as species identifiers and accelerate benthic biomonitoring surveys. It will make analysis independent of taxonomic expertise, overcoming the issue of taxonomic impediment and misidentification biases[39]. Moreover, the metabarcoding will allow extending the range of potential bio-indicators to meiofauna[40,41] and protists[21,42].

Compared to the morphological approach, metabarcoding provides a more holistic view of the metazoan taxonomic diversity, regardless of the size and developmental stage. Our HTS data not only include the macrofaunal species that dominate in the morphological samples, but also small-sized (<1 mm) species, extending the scope of analysis to the much broader meiofaunal diversity. In fact, taxa such as harpacticoid copepods, ostracods and many minor groups (gastrotriches, kinorhynchans, rotifers) are currently not included in the morphology-based bioassessments. Yet, some of them (kinorhynchans, turbellarians, ostracods) have been shown to be good indicators for assessing the impact of finfish[43] and shellfish farming[44]. Hence, metabarcoding might be the only way to account for these meiofaunal-size organisms.

It is not surprising that the metabarcoding data is enriched with meiofauna sequences. Indeed, it is more likely that small rather than large organisms would be captured in 2-grams sediment samples used for molecular analyses. The macrofaunal species sequences may well originate from tissue fragments, mucus, eggs or larvae[45]. However, the majority of macrofaunal DNA likely originates from extracellular DNA. Large quantities of extracellular DNA are preserved in the sediment[46]. This extracellular DNA accumulates over seasons and thus integrates the diversity of several population turnovers, including pioneer species that may constitute the short-time response to environmental perturbation. This is supported by the fact that the DNA-based biotic indices better reflect the sites environmental quality than those based on short-lived RNA molecules. RNA reflects the active fraction of the diversity[47], and thus might be less efficient than DNA to capture the macrofaunal diversity. It seems that DNA buffers the high natural variability observed between biological replicates[41]. Nevertheless, the presence of extracellular DNA might represent a major strength of the metabarcoding approach as for the detection

of macrofaunal taxa, especially annelids that are well represented in our samples. We obtained fewer sequences for other macrofaunal groups, such as molluscs and echinoderms. This may be due to the presence of shells or hard walls, which impede the diffusion of their DNA into the surrounding environment or the absence of mucus and small-sized benthic developmental stages.

Beyond its many advantages and ease of use, the routine application of metabarcoding for benthic monitoring requires overcoming some limitations. The main shortcomings involve the incompleteness of reference sequence databases as well as the fragmented knowledge on meiofaunal autecology. Despite considerable barcoding efforts, reference sequences are still very rare for benthic meiofaunal species. Given the prevalence of meiofauna in molecular assemblages, it is crucial to further describe potential meiofaunal bioindicator taxa not only through their genetic identification but also to specify their ecological values. Indeed, only half of the taxa detected with HTS could be ascribed to an ecological group and a unique ecological value is often assigned to an entire meiofaunal phylum. For instance, the nematodes form a hyper diverse group[48], but all are ascribed to the same AMBI ecological group. Given the immense phylogenetic diversity of meiofaunal groups, the relevance of these values and thus of the inferred indices is doubtful.

It is also important that molecular databases include more than one gene. At present, most of benthic metazoans are represented either by 18S rRNA gene or COI gene. Although these markers have different advantages and offer different taxonomic resolutions, both suffer similar limitations related to database incompleteness and primer specificities, and ideally should be coupled as in a recent comparative study[49]. We chose the V4 fragment of the 18S because it is shorter and easier to PCR amplify and sequence using Illumina technology. However, the resolution of the V4 region is limited for species-level assignments and has been shown to provide less accurate diversity estimates than COI[31]. Nevertheless, some key species could only be detected using the 18S marker, as *Malacoceros fulginosus*, for which there is no reference COI sequence. With upcoming extensions of Illumina sequencing read lengths, it will be possible to sequence the full COI barcode, which might be more informative for future metazoan-based metabarcoding. Alternatively, improvements are being proposed towards the design of new primers targeting shorter COI fragments[50] or new amplification strategies[51]. However, the COI marker is a protein-coding gene and thus remains less suitable than the 18S rDNA marker for the design of universal primers[40]. Moreover, the use of COI is hampered by difficulties in assigning higher taxonomic level to those sequences that lack close correspondence in the reference database[52]. Finally, it is necessary to expand the dimension of DNA sequence databases by gathering knowledge on gene copy numbers and polymorphisms. With this new information on intra-genomic variation in hand, it will be possible to refine sequence taxonomic assignments and quantitative ecological inferences.

Another important challenge is to develop biotic indices specifically for HTS data and assign appropriate scores to species given their autecology[53]. The currently used ITI and AMBI formulas have been developed for morphotaxonomic inventories of marine species. In their formulas, the ecological weight of each taxon morphologically isolated and identified in an environmental sample is used as a factor of its abundance in the sample. However, HTS sequence abundance data depend on many technical and biological biases and its exploitation for quantitative analyses remains a major issue. In fact, the relative abundance of DNA template molecules can be obtained if rigorous HTS data filtering is undertaken[27], and useful relative abundance information could be drawn from analyses performed at coarse taxonomic levels[54], which is inherent to the use of the 18S marker[31]. In our comparative study, the sequence abundance is not completely disconnected from the abundance of specimens, as shown by the high similarities of sequence proportions for the taxa found with both approaches, as well as among station replicates and in terms of relative abundances of the dominant taxa (e.g. *Capitella*). Such up-weighting of the dominant bioindicator taxa certainly improved the reconstruction of biotic indices. Biotic indices may integrate multiple diversity metrics (M-AMBI[55]). However, it has been shown that relying on the richness only (S-AMBI) or on the Shannon diversity only (H-AMBI) performs equally well and avoids unnecessary statistical noise[34]. Therefore, we recommend the use of the sequence abundance information for biomonitoring purposes, and thus the use of H-AMBI over S-AMBI.

Beyond the use of alpha-diversity metrics, beta-diversity patterns could be incorporated in a HTS index, as recently proposed in the novel index designed for coralligenous macroalgal assemblages[56]. For example, accounting for the dispersion propensity and distance of bioindicator species when studying communities sampled along distance-to-cage gradients remain a major challenge[3]. Further surveys along smoother environmental gradients are needed as our sampling strategy involves a categorical jump from impacted to non-impacted (and therefore high correlations values). In fact, this is crucial for DNA-based studies given the ability of extracellular DNA to be carried over great distances by water currents and turbulences. It is also important to reconsider the HTS data normalization step, as statistical models accounting for sequence abundances heteroscedasticities could replace rarefaction approaches[57].

To conclude, our study shows that the metabarcoding has potential to revolutionize benthic monitoring surveys. Its implementation will require some efforts, especially concerning the adaptation of biotic indices to molecular data. However, the advantages provided by the standardization and automation of the eDNA-based benthic monitoring fully justifies the further developments of this approach.

References

1. Merino, G. *et al.* Can marine fisheries and aquaculture meet fish demand from a growing human population in a changing climate? *Glob. Environ. Change* **22,** 795–806 (2012).
2. Kalantzi, I. & Karakassis, I. Benthic impacts of fish farming: meta-analysis of community and geochemical data. *Mar. Pollut. Bull.* **52,** 484–493 (2006).
3. Lee, S., Hartstein, N. D., Wong, K. Y. & Jeffs, A. Assessment of the production and dispersal of faecal waste from the sea-cage aquaculture of spiny lobsters. *Aquac. Res.* doi: 10.1111/are.12618 (2014).
4. Huang, Y. C. A., Huang, S. C., Hsieh, H. J., Meng, P. J. & Chen, C. A. Changes in sedimentation, sediment characteristics, and benthic macrofaunal assemblages around marine cage culture under seasonal monsoon scales in a shallow-water bay in Taiwan. *J. Exp. Mar. Bio. Ecol.* **422,** 55–63 (2012).
5. Borja, A. *et al.* Assessing the suitability of a range of benthic indices in the evaluation of environmental impact of fin and shellfish aquaculture located in sites across Europe. *Aquaculture* **293,** 231–240 (2009).
6. Maurer, D., Nguyen, H., Robertson, G. & Gerlinger, T. The Infaunal Trophic Index (ITI): its suitability for marine environmental monitoring. *Ecol. Appl.* **9,** 699–713 (1999).
7. Borja, A., Franco, J. & Pérez, V. A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. *Mar. Pollut. Bull.* **40,** 1100–1114 (2000).
8. Rygg, B. Developing indices for quality status classification of marine soft-bottom fauna in Norway. In: NIVA report;5208. Norsk institutt for vannforskning (2006).
9. Rygg, B. & Norling, K. Norwegian Sensitivity Index (NSI) for marine macroinvertebrates, and an update of Indicator Species Index (ISI). In: NIVA-rapport;6475. Norsk institutt for vannforskning (2013).
10. Keeley, N. B., Forrest, B. M., Crawford, C. & Macleod, C. K. Exploiting salmon farm benthic enrichment gradients to evaluate the regional performance of biotic indices and environmental indicators. *Ecol. Indic.* **23,** 453–466 (2012).
11. Aylagas, E., Borja, A. & Rodríguez-Ezpeleta, N. Environmental status assessment using DNA metabarcoding: towards a genetics based marine biotic index (gAMBI). *PLoS One* **9,** e90529 (2014).
12. Bohmann, K. *et al.* Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* **29,** 358–367 (2014).
13. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21,** 2045–2050 (2012).
14. Creer, S. & Sinniger, F. Cosmopolitanism of microbial eukaryotes in the global deep seas. *Mol. Ecol.* **21,** 1033–1035 (2012).
15. Zimmermann, J., Glöckner, G., Jahn, R., Enke, N. & Gemeinholzer, B. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Mol. Ecol. Res.* doi: 10.1111/1755-0998.12336 (in the press).
16. Kermarrec, L. *et al.* A next-generation sequencing approach to river biomonitoring using benthic diatoms. *Freshw. Sci.* **33,** 349–363 (2014).
17. Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. & Baird, D. J. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One* **6,** e17497 (2011).
18. Yu, D.W. *et al.* Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol. Evol.* **3,** 613–623 (2012).
19. Chariton, A. A., Court, L. N., Hartley, D. M., Colloff, M. J. & Hardy, C. M. Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Front. Ecol. Environ.* **8,** 233–238 (2010).
20. Bik, H. M., Halanych, K. M., Sharma, J. & Thomas, W. K. Dramatic shifts in benthic microbial eukaryote communities following the Deepwater Horizon oil spill. *PLoS One* **7,** e38550 (2012).
21. Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T. & Wilding, T. A. Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Mol. Ecol. Resour.* **14,** 1129–1140 (2014).
22. Black, K. D. The environmental interactions associated with fish culture. In: *Biology of Farmed Fish.* ed. Black, K. D. & Pickering, A. D. Sheffield, Sheffield Academic Press: pp. 284–326 (1998).
23. Wilding, T. A., Cromey, C. J., Nickell, T. D. & Hughes, D. J. Salmon farm impacts on muddy-sediment megabenthic assemblages on the west coast of Scotland. *Aquac. Environ. Interact.* **2,** 145–156 (2012).
24. Zobell, C. E. Studies on redox potential of marine sediments. *Bulletin of the American Association of Petrology and Geology* **30,** 477–513 (1946).
25. Worsfold, T. & Hall, D. National marine biological analytical quality control scheme: guidelines for processing marine macrobenthic invertebrate samples: a processing requirements protocol version 1.0. *NMBAQC* http://www.nmbaqcs.org/media/9732/nmbaqc%20-%20inv%20-%20prp%20-%20v1.0%20june2010.pdf (2010) (Date of access: 21/03/2015).
26. Stoeck, T. *et al.* Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19,** 21–31 (2010).
27. Esling, P., Lejzerowicz, F. & Pawlowski, J. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Res.* doi: 10.1093/nar/gkv107 (in the press).
28. Guillou, L. *et al.* The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* gks1160. doi: 10.1093/nar/gks1160 (2012).
29. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
30. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–7541 (2009).
31. Tang, C. Q. *et al.* The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proc. Natl. Acad. Sci. USA* **109,** 16208–16212 (2012).
32. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27,** 2194–2200 (2011).
33. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321 (2010).
34. Sigovini, M., Keppel, E. & Tagliapietra, D. M-AMBI revisited: looking inside a widely-used benthic index. *Hydrobiologia* **717,** 41–50 (2013).
35. Fauchard, K. & Jumars, P. A. The diet of worms: A study of polychaete feeding guilds. *Oceanogr. Mar. Biol. Ann. Rev.* **17,** 193–284 (1979).
36. Word, J. Q. The Infaunal Trophic Index. In: Southern California Coastal Water Research Project Annual Report. El Segundo. California. pp. 19–40 (1978).
37. Word, J. Q. Classification of benthic invertebrates into infaunal trophic index feeding groups. In: *Coastal Water Research Project Biennial Report.* pp. 103–121 (1980).
38. de Cárcer, D. A., Denman, S. E., McSweeney, C. & Morrison, M. Evaluation of subsampling-based normalization strategies for tagged high-throughput sequencing data sets from gut microbiomes. *Appl. Environ. Microbiol.* **77,** 8795–8798 (2011).

39. de Carvalho, M. R. *et al.* Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. *Evol. Biol.* **34,** 140–143 (2007).
40. Carugati, L., Corinaldesi, C., Dell'Anno, A. & Danovaro, R. Metagenetic tools for the census of marine meiofaunal biodiversity: An overview. *Mar. Genomics* (2015) doi: 10.1016/j.margen.2015.04.010.
41. Chariton, A. A *et al.* Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. *Environ. Pollut.* **203,** 165–174 (2015).
42. Amorim Visco, J. *et al.* Environmental monitoring: inferring diatom index from next-generation sequencing data. *Environ. Sci. Technol.* (2015) doi: 10.1021/es506158m.
43. Grego, M., De Troch, M., Forte, J. & Malej, A. Main meiofauna taxa as an indicator for assessing the spatial and seasonal impact of fish farming. *Mar. Poll. Bull.* **58,** 1178–1186 (2009).
44. Mirto, S., La Rosa, T., Danovaro, R. & Mazzola, A. Microbial and meiofaunal response to intensive mussel-farm biodeposition in coastal sediments of the Western Mediterranean. *Mar. Poll. Bull.* **40,** 244–252 (2000).
45. Taberlet, P. *et al.* Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Mol. Ecol.* **21,** 1816–1820 (2012).
46. Corinaldesi, C., Beolchini, F. & Dell'Anno, A. Damage and degradation rates of extracellular DNA in marine sediments: implications for the preservation of gene sequences. *Mol. Ecol.* **17,** 3939–3951 (2008).
47. Coolen, M. J. & Orsi, W. D. The transcriptional response of microbial communities in thawing Alaskan permafrost soils. *Front. Microbiol.* **6** (2015).
48. Blaxter, M., Floyd, R. & Abebe, E. Molecular barcoding for nematode identification and diversity studies. *J. Nematol.* **35,** 326 (2003).
49. Cowart, D. A. *et al.* Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS ONE* **10,** e0117562 (2015) doi: 10.1371/journal.pone.0117562.
50. Leray, M. *et al.* A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front. Zool.* **10,** 34 (2013).
51. Shokralla, S. *et al.* Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Sci. Rep.* **5** (2015).
52. Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F. & Taberlet, P. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biol. Lett.* **10,** 20140562 (2014).
53. Haddad, N. M. *et al.* Species' traits predict the effects of disturbance and productivity on diversity. *Ecol. Lett.* **11,** 348–356 (2008).
54. Leray, M. & Knowlton, N. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proc. Natl. Acad. Sci. USA* **112,** 2076–2081 (2015).
55. Borja, A. *et al.* Using M-AMBI in assessing benthic quality within the water framework directive: some remarks and recommendations. *Mar. Pollut. Bull.* **56,** 1377–1379 (2008).
56. Cecchi, E., Gennaro, P., Piazzi, L., Ricevuto, E. & Serena, F. Development of a new biotic index for ecological status assessment of Italian coastal waters based on coralligenous macroalgal assemblages. *Eur. J. Phycol.* **49,** 298–312 (2014).
57. McMurdie, P. J. & Holmes, S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10,** e1003531 (2014).

Acknowledgments

Author Contributions

Conception and design of the work: F.L., J.P., L.P. and T.W. Acquisition of the data: F.L. and L.P. Analysis and interpretation of the data: F.L., P.E., J.P., T.W. and K.B. Contribution of reagents/materials/analysis tools: P.E., J.P. and T.W. Writing of the paper: F.L., J.P., P.E., T.W. and K.B. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Lejzerowicz, F. *et al.* High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. *Sci. Rep.* **5**, 13932; doi: 10.1038/srep13932 (2015).