

Are genomic indices effective alternatives to morphology based benthic indices in biomonitoring studies? Perspectives from a major harbour and marine protected area.

Heidy Q. Dias and Soniya Sukumaran*

CSIR-National Institute of Oceanography,
Regional Centre Andheri (W), Mumbai 400 053, India

*Corresponding author. Tel.: +91 22 2635 9605; fax: +91 22 2636 4627.

E-mail address: soniya@nio.org (S. Sukumaran).

Abstract

Ecological assessments are currently being conducted by traditional morpho-taxonomical identification techniques that are time-consuming and often inaccurate. Biomonitoring programs are increasingly being complemented by the more rapid and efficient DNA barcoding approach. We compared the congruency of morpho-taxonomic (AMBI - AZTI's Marine Biotic Index) and genomic (gAMBI) benthic indices in ecological quality status (EcoQS) assignment in Mumbai harbour and Malvan Marine Protected area (MPA). The study, first of its kind to adopt the gAMBI tool in the selected milieu, contributed substantial number of macrobenthic cytochrome c oxidase subunit I gene (COI) sequences that were previously unavailable in the reference library, adding sufficient genetic resources for establishing ecostatus. AMBI and gAMBI values based on presence/absence data related significantly with those derived from abundance data matrices. Taxonomic and genomic indices derived ecostatus corresponded sufficiently well despite minor discrepancies, underscoring the viability of gAMBI as a superior alternative to AMBI in monitoring studies.

Keywords: Macroinvertebrates, DNA barcoding, AMBI, gAMBI, Ecological quality status, Biomonitoring

1. Introduction

Amelioration of coastal ecosystems has been the prime focus during the past decade since these systems are under a constant threat of multi-stressors. Protection of these ecosystems is the principal goal of the European bodies such as the Water Framework Directive (WFD; 2000/60/EC) and Marine Strategy Framework Directive (MSFD; 2008/56/EC). In this context, the universal index such as AZTI Marine Biotic Index (AMBI) (Borja et al., 2000) has been widely adopted worldwide for ecological quality status assessments and proven to give promising results in various milieu (Borja et al., 2019). The AMBI was solely formulated using the macroinvertebrates, as they responded to disturbance by means of different adaptive strategies (Borja et al., 2019). The macrobenthic organisms have long been used as bioindicators of ecosystem health owing to their characteristics, functions and ease in handling (Aylagas et al., 2014; Dias et al., 2018).

For bioassessment studies, a frequently used method is the classical visual morpho-taxonomy based classification that requires in-depth taxonomic expertise in species identification. This approach has numerous drawbacks as they are extensively laborious, time-consuming, error-prone leading to misidentification and also expensive (Ranasinghe et al., 2012; Aylagas et al., 2020). As the benthic indices are mainly calculated on the basis of species sensitivity to assign an area to a disturbance class (high to bad), the species identity has to be precise. Any erroneous classification of species may in turn lead to potentially inaccurate ecological quality assessments resulting in mismanagement of ecosystems (Stein et al., 2014a). Besides, the identification is often compromised when there is cryptic diversity, juvenile forms and damaged specimens (Cordier et al., 2021) which necessitates the removal of these species from the datasets. To overcome the hurdle of misidentification and improving the speed and efficacy of biomonitoring programs, genomic methods are now being employed as an alternative to morpho-taxonomy methodology (Ranasinghe et al., 2012). DNA barcoding is a method proposed for explicit identification of species based on a short fragment of the mitochondrial DNA from a standardised genetic marker (Hebert et al., 2003). From the available set of genes, the ~658bp region of the mitochondrial cytochrome c oxidase subunit I gene (COI) has been employed widely for its higher efficiency and faster substitution rate (Aylagas et al., 2014; Lobo et al., 2016).

The adoption of genomic approach in ecological quality status assessment studies is still in nascent stages and very few works have been conducted, that too only in temperate waters (Aylagas et al., 2014; 2018). Assessment of EcoQS was performed using multitrophic Marine biotic index via the metabarcoding approach using foraminifera, bacteria and general eukaryotes by Keeley et al. (2018). There are two methods available for bioassessments: 1) DNA barcoding of individual species

using Sanger sequencing (Hebert et al., 2003); 2) DNA metabarcoding of more than one species (bulk samples) using next-generation sequencing techniques (Fonseca et al., 2010). In both the cases, sequences need to be compared to a reference library that contains correspondence between the sequences/barcodes and taxonomical identification (Aylagas et al., 2014). Although the latter method appears easier and faster, the inherent requirement of comprehensive reference databases containing all the detected species sequences is a major deterrent (Ranasinghe et al., 2012; Aylagas et al., 2016; Dormontt et al., 2018). Next-generation sequencing will also require well-built sequence repositories produced by DNA barcoding of individual species to establish taxonomic identities (Stein et al., 2014a). Besides, the metabarcoding approach provides only the presence-absence information of species that can create a lacuna regarding the species abundance data which is the basic requirement for calculation of the benthic indices. Another drawback is that large and abundant species are easily detected due to higher contribution of tissue as compared to smaller and rare species (Deagle et al., 2018). This bias is assumed to occur more in areas with higher diversity that have small-sized species in contrast to areas with low diversity (Van den Bulcke et al., 2021). When the DNA concentration is low and the sample mixture is of non-target DNA and environmental particles, it influences the oligo efficiency (Van den Bulcke et al., 2021). Besides, the presence of extracellular DNA in an environmental sample also could give biased outcome (Torti et al., 2015). Considering the various complexities in application of bioinformatic tools and the cost for trials, it might not be feasible to conduct metabarcoding studies until the reference libraries are well defined for all possible species.

Prior to the implementation of barcoding approaches in ecological monitoring programs, it is essential to compare the results with that of visual identification so that valid biotic indices and the inferences are derived from genetic methods (Aylagas et al., 2016). Hence, the traditional DNA barcoding of single species with prior visual identification seems to be the most practical method at present to verify the right hit and to deposit a new species sequence in the database (Borja et al., 2016). AMBI was formulated based on abundance-weighted tolerances of the species by assigning them to ecological groups (sensitive to opportunist). To overcome the shortcomings associated in the calculation of AMBI, genomic version of AMBI (gAMBI) was initiated by Aylagas et al. (2014) by attempting metabarcoding approach to assess the ecological quality status. The gAMBI holds the same concept as that of the standard AMBI calculation except the species are identified based on molecular approaches. In this study, the genomic-based AMBI was estimated by undertaking the DNA barcoding of each individual species. Also, AMBI mandates identification and enumeration of all species present in a sample which is not yet possible by metabarcoding (Yu et al., 2012). Next

generation sequencing has immense potential for future bio-monitoring exercises and hence effectiveness of biotic indices based on presence-only data juxtaposed against those estimated on abundance based indices is required to be examined in different geographic areas.

The Indian coast, a tropical region, is subjected to multifarious anthropogenic stressors and of late, continuous efforts are being expended to evaluate the Ecological Quality Status (EcoQS) using the standard benthic indices. However, these approaches are time-consuming, expensive, manpower intensive and face the problem of diminishing trained taxonomists (Prathapan et al., 2006, Xie et al., 2021). Hence, immediate attention is required to improvise the biomonitoring tools with an aim to accelerate the ecological assessment process. Despite the fact that the molecular methods have existed for more than a decade, very few attempts have been made to barcode macrobenthic species with the focus being largely on the polychaeta taxon (Vijapure et al., 2019a and references therein). Although the polychaetes dominate largely in the benthic realm (Joydas and Damodaran, 2009) and are used as ecosystem health indicators, the other important macrobenthic groups (crustaceans, molluscs, sipunculids) also contribute to the benthic processes. No attempt has been made to date for assessing the ecological integrity using the genomic methods in tropical waters. The heterogeneous Indian coast harbours high marine biodiversity (Vijapure and Sukumaran, 2019). Therefore, it is vital to achieve maximum taxonomic coverage by barcoding the majority of macrobenthic species which can effortlessly aid speed and accuracy of monitoring assessments in Indian waters (Vijapure et al., 2019a). The present study was undertaken in two areas of contrasting characteristics, an active major harbour (Mumbai) and a Marine Protected Area (Malvan) along the Northwest coast of India. The main aims were 1) to fortify the sequences of macrobenthic species in the reference library as a stepping block towards the implementation of gAMBI, 2) to assess the ecological quality status using benthic indices based on morpho-taxonomic (AMBI) and genomic (gAMBI) methods, 3) to test the level of congruence between both methods in establishing the EcoQS and determining the efficacy of gAMBI in the resolution of EcoQS. Furthermore, based on the results obtained we have elaborated on the merits and demerits of using gAMBI which can be useful for taking conscious decisions in terms of index efficacy, cost-effectiveness and acceleration in monitoring programs for future management purposes.

2. Material and Methods

2.1 Study site

The two areas, namely Mumbai and Malvan, were chosen along the 720 km Maharashtra coastal stretch on the Northwest coast of India (Fig. 1). Mumbai (18°52'26.32" N 72°51'31.03" E)

has two major ports (Mumbai and Jawaharlal Nehru) located on the shores of the Thane Creek which forms a semi-enclosed basin that open into the Arabian Sea. The Mumbai harbour is one of the busiest ports and is of paramount importance for the country's trade with intensive shipping activities and several industrial setups (Mandal and Harkantara, 2013; Pradhan et al., 2021). This area is recipient of sewage releases, industrial wastes, port-related activities and ongoing vessel movements (Zingde et al., 1981; Sukumaran and Saraladevi, 2009). Malvan (16°02'48.48" N 73°27'24.16" E) is a protected area that was declared as a Marine Sanctuary in 1987 by the Indian Government to protect the high biodiversity including corals of the region. Recently, the Malvan coast has been under threat due to fishing activities, rapid urbanisation and a boom in tourism (Vijapure et al., 2019b). The geographical area where both study sites are located experiences typical weather with three prevailing seasons: premonsoon (March-May), monsoon (June-September) and postmonsoon (October-February).

2.2 Sampling strategy and processing

Three sampling surveys were conducted during premonsoon, monsoon and postmonsoon during 2017-2018 at both locations. A total of 18 stations were sampled at Mumbai (12 stations) and Malvan (6 stations). Bottom water was collected using a Niskin sampler for the analyses of abiotic variables. Dissolved oxygen (DO) was fixed immediately after collection with Winklers A and Winklers B following the method of Grasshoff et al. (1999). Salinity was measured using a salinometer and the nutrients such as NO_2^- -N, NO_3^- -N and NH_4^+ -N were measured following the standard protocols (Grasshoff et al., 1999). Sediment samples were collected using a van Veen grab of 0.04 m² bite area at each station for the analyses of texture and total organic carbon (TOC) and were kept in ice before transporting to the laboratory. Sediment texture was analysed using the wet-sieving pipette method (Buchanan, 1984). The TOC content was analysed after homogenization by a TOC analyzer (SHIMADZU SSM-5000A series). For the macrobenthos, sediment samples were collected in quadruplicates using a van Veen grab (0.04 m² bite area). Over the 3 sampling seasons, a total of 216 grabs were collected from 18 locations, of which 72 grabs were deployed exclusively for molecular work. The sediment samples were sieved immediately using a 0.5 mm mesh size and the retained material of the morpho-taxonomic samples were stored in 5% formalin containing rose bengal stain and transported to the laboratory for taxonomic analysis. The macrobenthic samples collected for molecular studies were sorted immediately by manually picking live organisms, fixed in 96% ethanol and stored at -20°C until DNA extraction (<6 months). The morpho-taxonomic samples were further sieved, sorted, counted and identified up to the lowest taxonomic resolution level using compound microscopes with the help of available identification keys. The ethanol stored samples

were taxonomically identified up to species level prior to DNA extraction and abundance was calculated similar to the morpho-taxonomic dataset. Following this, one specimen with sufficient tissue material each of single species per station was taken for further analysis by DNA barcoding.

2.3 DNA extraction, amplification and sequencing

The total genomic DNA was extracted from the prior-identified macroinvertebrates (Polychaeta, Amphipoda, Tanaidacea, Cumacea, Pelecypoda, Sipuncula) from the tissue sample of each soft-bodied individual and single leg of crustacean using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's procedure and the DNA quality was tested. The barcode region of the 658bp of COI gene was amplified using the universal primer set given by Folmer et al. (1994): HCO2198 (TAAACTTCAGGGTGACCAAAAAATCA) and LCO1490 (GGTCAACAAATCATAAAGATATTGG). For the preparation of PCR reaction mixture, 2.5 µl of Qiagen PCR buffer, 0.5 µl dNTPs, 19.5 µl ddH₂O, 0.5 µl of each forward and reverse primer (10 mM), 0.3 µl of (1.5 mM) MgCl₂, 0.2 µl Taq polymerase and 1 µl DNA template were taken making it to a final volume of 25 µl. Two different PCR thermal profiles were set for amplification of COI gene from the macrobenthic species (Table 1). Program 1 was applied for most of the polychaete species and program 2 was applied to the rest of the macroinvertebrates. The final PCR products were purified via gel extraction or column method using Invitrogen Purelink Quick Gel extraction and PCR purification kit as per the manufacturer's instruction manual. Further, the purified PCR products were sequenced bidirectionally by the ABi 3730XL Genetic Analyzer: 96 capillary sequencer. The final sequences were deposited in GenBank with accession numbers MZ504161 – MZ504235 (Table 2).

2.4 Calculation of indices: AMBI and gAMBI

The schematic workflow for the assessment of EcoQS using benthic indices is represented in Fig. 2. The focus was on barcoding single species with facilities that are easily available. Post identification through morphological features as well as DNA barcoding, the species abundances were estimated by enumerating counts of each species for both sets. The species were then assigned to the five ecological groups according to their sensitivity such as EGI (sensitive species), EGII (indifferent species), EGIII (tolerant species), EGIV (second-order opportunist species) and EGV (first-order opportunist species) following the AZTI's AMBI species list of December 2020. In case any particular species did not have an assigned EG, that of the genus was adopted (Borja and Tunberg, 2011). We calculated both morphology-based AMBI and genomic-based gAMBI for each station of the two areas using the AMBI software 6.0 (<http://ambi.azti.es>). We also calculated both

AMBI and gAMBI based on species abundance and presence/absence (p/a) data to compare the concurrences between disturbance classes. The index values designated the disturbance class based on the threshold boundaries proposed by Borja et al. (2000; 2007): ‘High’, 0.0-1.2; ‘Good’, 1.2–3.3; ‘Moderate’, 3.3–5.0; ‘Poor’, 5.0-6.0; and ‘Bad’, 6.0-7.0.

2.5 Data analyses

To compare the similarities between both morpho-taxonomic and molecular approaches, the percentages of species commonly identified by both methods were estimated. To compare the outcomes of both benthic indices, the number of sites assigning similar EcoQS was determined and the level of agreement between both methods was assessed following Mulik et al. (2020). The univariate analyses were carried out using the software package PRIMER v.7 (Clarke and Gorley, 2015).

3 Results

3.1 Hydro-sedimentological nature

The physico-chemical parameters are presented in Table A1. The average DO concentrations were high at Malvan (4.1 – 6.7 mg l⁻¹) stations as compared to Mumbai (3.1 – 5.9 mg l⁻¹). Nutrient concentrations were much higher in Mumbai, while it was remarkably lower in Malvan stations. Similarly, the SS displayed high values in Mumbai as compared to Malvan. The sediment nature in Mumbai was predominantly composed of silt, while Malvan was mostly dominated by sand, except for the outer station that was majorly silty. The average C_{org} content varied from 1.7 – 2.6% in Mumbai and from 1.0 – 3.4 % in Malvan.

3.2 Macrobenthic species community

From the morpho-taxonomic dataset, a total of 20371 and 42552 individuals belonging to 84 and 115 species were collected at Mumbai and Malvan respectively (Table A2). The macrobenthic species community in Mumbai was dominated by *Paraprionospio patiens* (18.2%), *Cossura coasta* (10.9%), *Magelona cincta* (6.2%), *Mediomastus capensis* (6.0%) and *Cheiriphotis megacheles* (5.1%). In Malvan, the macrobenthic community comprised of *Cossura coasta* (9.1%), *Mediomastus sp.* (8.7%), *Iphinoe sp.* (8.6%), *Maera othonides* (7.6%) and *Aricidea capensis* (6.1%). The Shannon-Wiener diversity displayed a range of 1.2 to 3.5 at Mumbai and 2.4 to 3.8 at Malvan indicating higher diversity in the latter case.

DNA barcoding was attempted for 115 species of which only 75 species (Table 2) were successfully sequenced for the barcode region of COI showing >65% success rate at Mumbai (28

species) and Malvan (35 species). Around 19% did not give any PCR product, 11% were sequenced but did not give good quality reads and 5% could not be extracted for DNA and were mostly singletons. The polychaete species formed a major part of the community in terms of dominance at both Mumbai and Malvan. Overall, 38 species of Polychaeta, 7 species of Amphipoda, 2 species of Pelecypoda, 1 species of Cumacea, 1 species of Tanaidacea, 1 species of Sipunculida and 2 species of Brachyura were sequenced. Brachyurans were not included in the calculation of benthic indices as per the index guidelines. In terms of abundance based on the genomic dataset, the species that dominated in Mumbai were *Paraprionospio patiens* (28.2%), *Magelona cincta* (17.1%), *Eunice* sp. (10.7%), *Polycirrus swakopianus* (10.3%) and *Mediomastus capensis* (4.7%), while in Malvan, *Iphinoe* sp. (20.4%), Maldanidae (10.6%), *Lumbrineris* sp. (7.9%), *Cossura coasta* (5.7%) and *Magelona crenulifrons* (4.3%) were the dominant species. The morpho-taxonomic and DNA barcoding approaches showed an overlap of 29 (34%) and 34 (30%) common species in Mumbai and Malvan respectively.

3.3 Taxonomic (AMBI) and genomic (gAMBI) benthic indices

The macrobenthic species derived from both morphological and molecular datasets were ascribed to an ecological group (EG) following the AZTI's species library. Based on morphology, the Mumbai region indicated the EG percentages in the following order EGIV (37.7%) > EGI (20.3%) > EGIII (18.1%) > EGII (16.6%) > EGV (7.4%), while Malvan showed higher percentages as EGI (47.1%) > EGII (19.2%) > EGIV (18.0%) > EGIII (11.5%) > EGV (4.3%). The EG assigned to species obtained via molecular approach depicted different trends in Mumbai as compared to the morpho-taxonomic EG such as EGII (38.5%) > EGI (32.2%) > EGIV (23.4%) > EGIII (6.0%). At Malvan, barcoding based method indicated similar trend as morphology based EG in terms of dominance i.e. EGI (44.6%) > EGII (30.3%) > EGIV (14.6%) > EGIII (10.4%). EGV species could not be identified from the genomic datasets in both areas due to sequencing constraints. However, the EG varied along with the stations for both approaches depending on the species occurrence (Fig. A2).

For reliable EcoQS assessments, AMBI and gAMBI was calculated based on both species abundance and presence/absence (p/a) data matrices derived from morpho-taxonomic and genetic methods. Considering the morphological species data (Fig. 3), AMBI values between 2.0 (Good EcoQS; st-2) and 4.3 (Moderate EcoQS; st-5) were obtained for Mumbai and from 0.8 (High EcoQS; st-2) and 3.3 (Good EcoQS; st-4) in Malvan. Similarly, gAMBI ranged from 0.5 (High EcoQS; st-6) and 4.2 (Moderate EcoQS; st-1) in Mumbai and 0.4 (High EcoQS; st-6) and 3.3 (Good EcoQS; st-5) in Malvan. More discrepancies between the two methods were visualised among the stations of

Mumbai, though between adjacent categories of EcoQS. Comparatively, the AMBI and gAMBI concurred in most of the stations of Malvan than Mumbai (Fig. 3a and b). p/a AMBI showed a range of 1.9 (Good EcoQS; st-10) and 4.1 (Moderate EcoQS; st-5) in Mumbai, whereas in Malvan, it ranged from 1.0 (High EcoQS' st-5) to 2.2 (Good EcoQS; st-3) (Fig. 4). The p/a gAMBI gave a range from 0.8 (High EcoQS; st-6 and st-9) to 2.8 (Good EcoQS; st-5 and st-8) and from 0.7 (High EcoQS' st-6) to 3.5 (Moderate EcoQS; st-5) in Mumbai and Malvan respectively (Fig. 4a and b). Overall, the ecostatus of most stations at Mumbai and Malvan remained unchanged irrespective of the abundance based or p/a based indices applied.

3.4 Relationship between morpho-taxonomic and genomic indices

In Fig. 5, both AMBI and gAMBI indices based on abundance and p/a data suggested different relationships amongst them. In Mumbai, there was a strong relationship between AMBI and p/a AMBI ($r^2=0.838$) and gAMBI and p/a gAMBI ($r^2=0.774$). In Malvan, gAMBI and p/a gAMBI ($r^2=0.796$) were strongly related, while AMBI and p/a AMBI exhibited weaker relationship ($r^2=0.51$). AMBI and gAMBI in both areas showed weak relationship as the points were extremely scattered and the slope attained was less. The level of agreement between the indices based on abundance, genomic abundance and presence/absence datasets for both Mumbai and Malvan is shown in Table 3. Stations 7 and 10 of Mumbai area were not included in Table 3 as gAMBI values could not be calculated and therefore comparison was not possible for these stations. Table 3 indicated that the level of agreement was high in the acceptable categories (high and good) in Mumbai (80%) and Malvan (96%). However, a small percentage of unacceptable category i.e moderate EcoQS was obtained by AMBI, gAMBI and p/a AMBI in Mumbai and by p/a gAMBI in Malvan. No station was assigned a poor or bad EcoQS by any index in the study area.

4. Discussion

Genomic based assessments have garnered attention in recent years as a more reliable method for bioassessments as compared to traditional approaches (Hajibabei et al., 2011). Although previous studies (Vijapure et al. 2019a and references therein) have contributed substantially to the barcoding of macrobenthic species, assessments based entirely on DNA barcoding of species have not been implemented as yet in the Indian subcontinent. Hence, this is the first study in this biogeographical region to apply a genomic approach for species identification and assess the EcoQS via the genomic index gAMBI. The primary objective was to derive reliable conclusions using the gAMBI and evaluate its efficacy vis-à-vis the morphology based AMBI based on abundance and p/a matrices.

4.1 Environmental background

The abiotic variables did not vary spatially in both areas. DO was high throughout and was above the required threshold for both the harbour and protected area suggesting good oxidizing conditions in Mumbai (Hardikar et al., 2021) and high environmental quality in Malvan (Vijapure et al., 2019b). Although the Mumbai harbour area had slightly higher levels of nutrients attributed to the inputs from various sources as compared to Malvan, highly degraded water quality was not observed. This is in line with the study conducted by Pradhan et al. (2021), wherein it was suggested that the harbour area was good to moderately polluted. On the other hand, Malvan is a marine protected area known to portray nearly pristine conditions due to less human interference and no point source discharges (Vijapure et al. 2019b). The high total organic carbon content in Mumbai can be attributed to the silty nature of sediments having a larger surface area leading to the accumulation of C_{org} (Zhou et al., 2007). In Malvan, the sandy sediments retained less C_{org} content except at station 4 where high silt composition naturally contributed to the high organic content which was also observed by Vijapure et al. (2019b).

4.2 Macrobenthic community

The macrobenthic community in the Malvan region was richer in diversity as compared to Mumbai harbour. Mumbai revealed the high composition of opportunistic groups (EGIV) due to the dominance of polychaetes such as *Paraprionospio patiens* and *Cossura coasta*. In a previous study conducted in the same harbour area (Sukumaran and Saraladevi, 2009), *Paraprionospio* sp. was reported as the dominant species which indicated unhealthy conditions. Malvan had more sensitive species (EG I) owing to the high diversity of polychaetes along with other macrobenthic groups. Past studies conducted in Malvan also reported similar species assemblages and rich diversity (Sukumaran et al., 2016; Vijapure et al., 2019b).

More than half of the species encountered in this study were successfully barcoded with more species barcodes obtained in the Malvan area owing to the high extant diversity. As observed in previous studies carried out by several authors (Satheeshkumar and Jagadeesan, 2010; Magesh et al., 2012; Sekar et al., 2016; Sigamani et al., 2016; Vijapure et al., 2019a), only the polychaete species were barcoded in India. Altogether, around 54 polychaete species from various regions of India have been barcoded till date. In the present study, a total of 38 polychaete species were barcoded successfully out of which 20 polychaete species were new additions to the GenBank. Considering all the barcoded species presently available, 11 polychaete species corresponded with the species barcodes provided by Vijapure et al. (2019a) and 2 species by Sigamani et al. (2016). Furthermore,

this is the first study to barcode other macrobenthic species such as amphipods (7 species), tanaids (1 species), cumaceans (1 species), pelecypods (2 species), sipunculids (1 species) and brachyurans (2 species) from the Indian subcontinent. It is pertinent to note here that out of a total of 75 barcoded species, 41 were submitted for the first time forming important additions to the GenBank reference library. Thus, with the inclusion of the species barcoded in this study, the expanded reference library for macrobenthic species stands at a record of 129 species from the Indian subcontinent.

The average sequencing success for the COI barcode region of >65% in this study is somewhat comparable to that obtained in similar studies conducted elsewhere (Schenk et al., 2019). In terms of species dominance based on the molecular data as well as morpho-taxonomic dataset, a similar trend of *Paraprionospio patiens* dominating the macrobenthic community was observed in Mumbai. Conversely, an opposite trend to that obtained by morphology was observed in Malvan. The cumacean *Iphinoe* sp. (EGI) was indicated as the dominant one along with other polychaete species by molecular data apropos of *Cossura coasta* (EGV) discerned by the taxonomy data. In Mumbai, there was a partial mismatch between the EGs owing to other factors such as non-detection of some species via the genomic method. This occurs as for many species DNA do not get amplified and as a result, are not sequenced due to various molecular constraints (Cavaliere et al., 2021).

4.3 Ecological quality status assessment

In Mumbai, the EcoQS assigned by AMBI was between Good and Moderate depending on the station and the differential impacts in the harbour area. The Mumbai harbour has been reported to have an abundance of opportunistic species signifying that the area was under considerable stress (Sukumaran and Saraladevi, 2009). However, in our study, both opportunistic as well as sensitive species were present indicating that the stress might not be severe presently as also opined by Pradhan et al. (2021). Malvan was assigned mostly High and Good EcoQS by AMBI, as it was less impacted which also concurred with the abiotic variables. Such trends were expected at Malvan where the impact was minimal and high diversity was evidenced. The real question that arises is whether the gAMBI also produced similar results and if there was a congruence in the assignment of the EcoQS. At Mumbai, identical EcoQS were conferred to 5 stations, whereas at 4 other stations, variations were minor marked by a shift between adjacent categories (Moderate vs Good; Good vs High). The variation in EcoQS was significant only at station 6 in Mumbai (Moderate vs High). Variations in EcoQS at Mumbai resulted due to the presence of EGIV species (*Cossura coasta* and *Kirkegaardia* sp.) identified by morpho-taxonomic method, while the genomic set had only the presence of sensitive species (*Magelona cincta* and *Eunice* sp.). Areas categorized below Good (Moderate, Poor, Bad) requires corrective action and management interventions for propelling these

sites towards achieving a good ecological status (Water Framework Directive; Cavaliere et al., 2021). However, if only two categories were considered, acceptable (High+Good) and unacceptable (Moderate+Poor+Bad), abundance based gAMBI showed high concurrence (70%) with abundance based AMBI at Mumbai. In Malvan, similarity between abundance based AMBI and genomic abundance based gAMBI results was considerable with all stations presenting High to Good status, thus falling in acceptable category. The only deviation at Malvan was observed at station 5, where the EcoQS assigned by the AMBI (High) and gAMBI (Good) differed, though the categories were adjacent. Such discrepancies occurred due to the constraints in DNA methods in sequencing certain species resulting in non-inclusion of these species in the analyses while computing gAMBI. Nevertheless, comparable EcoQS evaluations were derived by AMBI and gAMBI in both the areas as the level of agreement shown between both was reasonable (Table 3).

The EcoQS computed on the basis of p/a data indicated that most of the stations at Mumbai had similar EcoQS as that of abundance based EcoQS, except for stations 1 and 3 where Good EcoQS was indicated by p/a data as opposed to the Moderate EcoQS assigned by abundance AMBI. This occurs due to non-reliance on the dominant species and consideration only of the presence of species which can lead to erroneous classifications (Muxika et al., 2012; Aylagas et al., 2018). The EcoQS based on p/a indices in Malvan indicated that most stations were in High to Good categories, similar to that of abundance-based indices, proving its competence. This might be due to its less impacted nature supporting even distribution of species. This observation stands valid as many authors suggested that the p/a is more robust than abundance alone (Alden et al., 2002; Ranasinghe et al., 2012). Robustness of p/a index depends on the variation between areas, species composition, pollution impact and ecological sensitivity of species. At Malvan, p/a based gAMBI assigned 4 out of the 6 stations similar to p/a based AMBI, another station (station 6) varied between adjacent categories, whereas at station 5 difference was prominent (High vs Moderate). The p/a based AMBI and gAMBI categorised all stations in 'acceptable' group, except station 5 which was designated moderate by p/a gAMBI.

The relationship was positive and highly significant (close to 1) between AMBI and p/a AMBI, gAMBI and p/a gAMBI in Mumbai and between gAMBI and p/a gAMBI in Malvan (Fig. 6). The level of agreement between these indices was high largely in the acceptable categories with the exception of a few stations falling in the unacceptable category (Moderate EcoQS). The successful application of gAMBI is entirely dependent on the species coverage detected by DNA barcoding, which if not sufficient, can make it cumbersome to compute gAMBI particularly in stressed areas. However, this study revealed that 30% to 34% species were common between both taxonomic and

genomic approaches which gave a satisfactory evaluation. For the efficient working of gAMBI, the main aspect is the assignation of species to an ecological group wherein, the species detected by DNA barcoding should be present in the AMBI species library. Aylagas et al. (2014) suggested that minimum 15% of species present in the AMBI list should suffice for a reliable p/a index calculation. In this study, all the species encountered by molecular approach were present in the AMBI species list, thus strengthening the use of index. Secondly, most of the dominant species were detected by DNA barcoding in this study which is important to give an accurate EcoQS by rating their ecological sensitivity.

4.4 gAMBI as a future tool for EcoQS assessments

The assessments based on molecular methods have shown a success rate in the temperate waters using the metabarcoding approach which are deemed to be cost-effective, reliable and faster (Aylagas et al., 2020). However, this approach on macrobenthos has not been attempted in India due to various reasons, for instance the absence of an updated reference repository, HTS costs and diminishing taxonomic experts which was also highlighted by Vijapure et al. (2019a). Sigamani et al. (2016) have suggested that for the successful application of gAMBI, the majority of the species spectrum present in the AMBI list need to be barcoded on priority basis. In this study, more than 65% of species were barcoded successfully and therefore the gAMBI application gave satisfactory results. This genomic approach applied in both Mumbai harbour and Malvan MPA provided valid inferences, reliably differentiating a moderately polluted zone (Mumbai harbour) from lowly impacted zone (Malvan MPA), thus proving its efficiency in establishing an appropriate EcoQS. The merits of gAMBI applicability include accuracy in species identification leading to proper ecological group assignation, higher efficiency than traditional morphology-based index, sufficiency of p/a data for index computation, acceleration in biomonitoring work etc. However, in our study, gAMBI tended to overestimate the ecological quality status of 5 stations at Mumbai and at 1 station at Malvan into the adjacent higher categories as some key EGIV species like *Namalycastis senegalensis*, *Cossura coasta*, *Parheteromastus tenuis* could not be successfully barcoded. Thus, there is still a need to improve and conduct more barcoding studies in stressed ecosystems to gain a complete understanding of the area. It is also important to target the species frequently occurring in an area and if these species are present in the library, it can provide valuable inferences in biomonitoring studies (Pérez-Burillo et al., 2020). Aylagas et al. (2014) suggested to barcode the most frequent species which even in low numbers can suffice for the calculation of p/a indices. Major pitfalls in the implementation of gAMBI in routine biomonitoring can arise when the target species are not present in the reference repository and species are not detected due to sequencing

constraints, resulting in overestimation or underestimation of EcoQS in naturally stressed ecosystems. Although metabarcoding has received greater recognition recently for identification of multiple species, it is often hampered by the absence of quality reference libraries (Dormontt et al., 2018). Tropical ecosystems are highly biodiverse and hence efforts may be initially focussed on barcoding the priority species identified through an initial visual taxonomy based abundance data (Pérez-Burillo et al., 2020). DNA barcoding of individual species properly identified by experts and then generating reference libraries by including error free barcodes is the first logical step towards the successful application of metabarcoding in monitoring programs of any geographical area (Behrens-Chapuis et al., 2021). In the future, this can be extended to the metabarcoding approach with an up-to-date reference repository and baseline data from the area of interest. Such an approach will facilitate faster evaluation of larger geographical areas with higher degrees of accuracy resulting in major savings of man-hours.

Conclusion

In the present study, the application of DNA barcoding of macrobenthic species in Mumbai harbour and Malvan Marine Protected Area provided valuable insights into the ecological quality status assessments. A well curated comprehensive DNA barcode reference repository will enable application of HTS techniques in future studies, that is currently constrained by the availability of sufficient reliable sequence databases. In monitoring programs, species identities are pre-requisite to estimate biotic indices based on tolerance values of individual species for which robust DNA reference libraries are inevitable (Stein et al., 2014). Our study facilitated the deposition of 75 sequences in GenBank belonging to Polychaeta, Amphipoda, Cumacea, Tanaidacea, Pelecypoda, Sipunculida and Brachyura species. This provides an impetus towards strengthening existing reference barcode libraries for the region that could ease adoption of metabarcoding techniques in future monitoring programs. The EcoQS of both areas was assessed using the genomic AMBI and compared to the traditional morphology-based AMBI. This is the first study in India to sequence macrobenthic groups other than polychaetes and to apply gAMBI for EcoQS evaluation. The EcoQS fell in the range of High to Moderate disturbance class in both areas as per AMBI and gAMBI depending on the abundance and p/a datasets. Although the species compositions obtained by the visual and genomic approaches presented 30-34% resemblance, both methods gave comparable results in EcoQS at most stations and the minor misclassifications observed at others were oscillations between adjacent classes. Thus, the application of gAMBI corresponded with the traditional AMBI in most of the stations, thus proving its efficiency and reliability in bioassessments. The congruency of both morpho-taxonomic and genomic approaches further supports the application

of such metrics in routine biomonitoring. It can be concluded that the 30% to 34% species coverage was adequate for the successful ecological assessment in this study. Thus, visual identification based biotic indices can be reliably supplanted by gAMBI in monitoring regions that lack sufficient taxonomic expertise and monetary resources. Further, we recommend more studies in extremely stressed ecosystems of different geographical areas to test the utility of DNA-based methods before fully transitioning to these methods.

Presently, AMBI estimation mandates information on the identity of all individuals present in the sample with their respective abundances. Adoption of metabarcoding, a superior method to DNA barcoding, may not be suitable for estimating AMBI in the present form due to its limitations in providing reliable abundance data (Aygalas et al., 2014; Dowle et al., 2016). However, metabarcoding data can provide relevant information i.e species occurrence data for estimation of p/a AMBI which does not require species abundance. The present study indicated that p/a AMBI and p/a gAMBI mirrored the ecological quality status inferred by AMBI and gAMBI respectively. This gives confidence that metabarcoding can be adopted in future for improving the speed, accuracy and cost-effectiveness of ecological monitoring programs once sufficient barcode reference databases are built for most frequent species in an area.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to the Director, CSIR-NIO for providing us the facilities to carry out this work. HQD is thankful to CSIR for awarding a Senior Research Fellowship that gave her the opportunity to carry out the present work. This is CSIR-NIO contribution no.----.

References

- Alden III, R.W., Dauer, D.M, Ranasinghe, J.A., Scott, L.C., Llanso', R.J., 2002. Statistical verification of the Chesapeake Bay benthic index of biotic integrity. *Environmetrics* 13, 473–498.
- Aygalas, E., Borja, A., Irigoien, X., Rodríguez-Ezpeleta, N., 2016. Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. *Front. Mar. Sci.* 3.
- Aygalas, E., Borja, Á., Muxika, I. and Rodríguez-Ezpeleta, N., 2018. Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. *Ecol.Indic.* 95, 194-202.

- Aylagas, E., Borja, A., Pochon, X., Zaiko, A., Keeley, N., Bruce, K., Hong, P., Ruiz, G.M., Stein, E.D., Theroux, S. and Geraldini, N., 2020. Translational molecular ecology in practice: linking DNA-based methods to actionable marine environmental management. *Sci. Total Environ.* 744, 140780.
- Aylagas, E., Borja, A., Rodríguez-Ezpeleta, N., 2014. Environmental status assessment using DNA metabarcoding: towards a genetics based Marine Biotic Index (gAMBI). *PLoS ONE* 9, e90529.
- Behrens-Chapuis, S., Herder, F., Geiger, M.F., 2021. Adding DNA barcoding to stream monitoring protocols – What’s the additional value and congruence between morphological and molecular identification approaches? *PLoS ONE* 16(1), e0244598. <https://doi.org/10.1371/journal.pone.0244598>.
- Borja, A., Josefson, A.B., Miles, A., Muxika, I., Olsgard, F., Phillips, G., Rodríguez, J.G., Rygg, B., 2007. An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European Water Framework Directive. *Mar. Pollut. Bull.* 55, 42–52. <https://doi.org/10.1016/j.marpolbul.2006.08.018>.
- Borja, A., Tunberg, B.G., 2011. Assessing benthic health in stressed subtropical estuaries, eastern Florida, USA using AMBI and M-AMBI. *Ecol. Indic.* 11, 295–303. <https://doi.org/10.1016/j.ecolind.2010.05.007>.
- Borja, A., Chust, G., Muxika, I., 2019. Forever young: the successful story of a marine biotic index. *Adv. Mar. Biol.* 82, 93. <https://doi.org/10.1016/bs.amb.2019.05.001>.
- Borja, A., Elliott, M., Andersen, J.H., Berg, T., Carstensen, J., Halpern, B.S., Heiskanen, A.-S., Korpinen, S., Lowndes, J.S.S., Martin, G., Rodriguez-Ezpeleta, N., 2016. Overview of integrative assessment of marine systems: the ecosystem approach in practice. *Front. Mar. Sci.* 3. <http://dx.doi.org/10.3389/fmars.2016.00020>.
- Borja, A., Franco, J., Perez, V., 2000. A marine biotic index to the establish ecology quality of soft-bottom benthos within European estuarine coastal environments. *Mar. Pollut. Bull.* 40, 1100–1114. [https://doi.org/10.1016/S0025-326X\(00\)00061-8](https://doi.org/10.1016/S0025-326X(00)00061-8).
- Buchanan, J.B., 1984. Sediment analysis. In: Holme NA, McIntyre AD (eds) *Methods for the study of marine benthos*. Blackwell, Oxford, pp 41–63.
- Cavaliere, M., Angeles, I.B., Montresor, M., Bucci, C., Brociani, L., Balassi, E., Margiotta, F., Francescangeli, F., Bouchet, V.M.P., Pawlowski, J. and Frontalini, F., 2021. Assessing the ecological quality status of the highly polluted Bagnoli area (Tyrrhenian Sea, Italy) using foraminiferal eDNA metabarcoding. *Sci. Total Environ.* 790, 147871.
- Clarke, K.R., Gorley, R.N., 2015. *PRIMER v7: User Manual/Tutorial*. PRIMER-E, Plymouth.
- Cordier, T., Alonso-Sáez, L., Apothéloz-Perret-Gentil, L., Aylagas, E., Bohan, D.A., Bouchez, A., Chariton, A., Creer, S., Frühe, L., Keck, F. and Keeley, N., 2021. Ecosystems monitoring powered by environmental genomics: a review of current strategies with an implementation roadmap. *Mol. Ecol.* 30(13), 2937-2958.

- Deagle, B.E., Clarke, L.J., Kitchener, J.A., Polanowski, A.M., Davidson, A.T., 2018. Genetic monitoring of open ocean biodiversity: An evaluation of DNA metabarcoding for processing continuous plankton recorder samples. *Mol. Ecol. Resour.* 18, 391–406.
- Dias, H.Q., Sukumaran, S., Srinivas, T., Mulik, J., 2018. Ecological quality status evaluation of a monsoonal tropical estuary using benthic indices: comparison via a seasonal approach. *Environ. Sci. Pollut. Res.* 25, 22672–22688. <https://doi.org/10.1007/s11356-018-2344-0>.
- Dormontt, E.E., Van Dijk, K.J., Bell, K.L., Biffin, E., Breed, M.F., Byrne, M., Caddy-Retalic, S., Encinas-Viso, F., Nevill, P.G., Shapcott, A. and Young, J.M., 2018. Advancing DNA barcoding and metabarcoding applications for plants requires systematic analysis of herbarium collections—an Australian perspective. *Front. Ecol. Evol.* 6, 134.
- Dowle, E.J., Pochon, X.C., Banks, J., Shearer, K., Wood, S.A, 2016. Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: a case study using freshwater macroinvertebrates. *Mol. Ecol. Resour.* 16(5),1240–1254.
- Fonseca, V.G., Carvalho, G.R., Sung, W., Johnson, H.F., Power, D.M., Neill, S.P., Packer, M., Blaxter, M.L., Lambhead, P.J.D., Thomas, W.K., Creer, S., 2010. Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nat. Commun.*, 1-8.
- Grasshoff, K., Ehrhardt, M., Kremling, K., 1999. Methods of seawater analysis (Verlag Chemie). In: 1–419.
- Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A., Baird, D.J., 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE* 6(4), e17497.
- Hardikar, R., Haridevi, C.K., Ram, A., Parthipan, V., 2021. Distribution of size-fractionated phytoplankton biomass from the anthropogenically stressed tropical creek (Thane creek, India). *Reg. Stud. Mar. Sci.* 41, 101577.
- Hebert, P.D., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270, 313–321.
- Joydas, T.V., Damodaran, R., 2009. Infaunal macrobenthos along the shelf waters of the west coast of India Arabian Sea. *Indian J. Mar. Sci.* 38, 191–204.
- Keeley, N., Wood, S.A., Pochon, X., 2018. Development and preliminary validation of a multi-trophic metabarcoding biotic index for monitoring benthic organic enrichment. *Ecol. Indic.* 85, 1044-1057.
- Leese, F., Altermatt, F., Bouchez, A., Ekrem, T., Hering, D., Meissner, K., Mergen, P., Pawlowski, J., Piggott, J.J., Rimet, F., Steinke, D., 2016. DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe.
- Lobo, J., Teixeira, M.A., Borges, L.M., Ferreira, M.S., Hollatz, C., Gomes, P.T., Sousa, R., Ravara, A., Costa, M.H., Costa, F.O., 2016. Starting a DNA barcode reference library for shallow water polychaetes from the southern European Atlantic coast. *Mol. Ecol. Resour.* 16(1), 298–313.

- Magesh, M., Kvist, S., Glasby, C.J., 2012. Description and phylogeny of *Namalycastis jaya* sp. n. (Polychaeta, Nereididae, Namanereidinae) from the southwest coast of India. *ZooKeys* 238, 31–43.
- Mandal, S., Harkantra, S.N., 2013. Changes in the soft-bottom macrobenthic diversity and community structure from the ports of Mumbai, India. *Environ. Monit. Assess.* 185(1), 653–672.
- Mulik, J., Sukumaran, S., Dias, H.Q., 2020. Is the benthic index AMBI impervious to seasonality and data transformations while evaluating the ecological status of an anthropized monsoonal estuary?. *Ocean Coast. Manag.* 186, 05080.
- Muxika, I., Somerfield, P.J., Borja, Á., Warwick, R.M., 2012. Assessing proposed modifications to the AZTI marine biotic index (AMBI), using biomass and production. *Ecol. Indic.* 12(1), 96–104. <https://doi.org/10.1016/j.ecolind.2011.04.030>.
- Pérez-Burillo, J., Trobajo, R., Vasselon, V., Rimet, F., Bouchez, A., Mann, D.G., 2020. Evaluation and sensitivity analysis of diatom DNA metabarcoding for WFD bioassessment of Mediterranean rivers. *Sci. Total Environ.* 727, 138445.
- Pradhan, U.K., Ram, A., Sahu, B., Udayakrishnan, P.B., Lonkar, S., Kumbhar, L., 2021. Multivariate assessment of outflow water quality in highly urbanized creek system: implication of natural and anthropogenic processes. *Environ. Sci. Pollut. Res.* 28(36), 49725–49742.
- Prathapan, K.D., Dharma Rajan, P., Narendran, T.C., Viraktamath, C.A., Subramanian, K.A., Aravind, N.A., Poorani, J., 2006. Biological Diversity Act 2002: shadow of permit-raj over research. *Curr. Sci.* 91(8), 1006–1007.
- Ranasinghe, J.A., Stein, E.D., Miller, P.E., Weisberg, S.B., 2012. Performance of Two Southern California Benthic Community Condition Indices Using Species Abundance and Presence-Only Data: Relevance to DNA Barcoding. *PLoS one*, e40875.
- Satheeshkumar, P., Jagadeesan, L., 2010. Phylogenetic position and genetic diversity of Neridae—polychaeta based on molecular data from 16S r RNA sequences. *Middle-East J. Sci. Res.* 6(6), 550–555.
- Sekar, V., Rajasekaran, R., Prasannakumar, C., Sankar, R., Sridhar, R., Sachithanandam, V., 2016. Morphological and COI sequence based characterisation of marine polychaete species from Great Nicobar Island India. In: *DNA barcoding in marine perspectives*. Springer, Cham, pp. 89–111.
- Sigamani, S., Perumal, M., Thivakaran, G.A., Thangavel, B., Kandasamy, K., 2016. DNA barcoding of macrofauna act as a tool for assessing marine ecosystem. *Mar. Pollut. Bull.* 152, 107891. <https://doi.org/10.1016/j.marpolbul.2016.07.017>.
- Stein, E.D., White, B.P., Mazor, R.D., Jackson, J.K., Battle, J.M., Miller, P.E., Pilgrim, E.M. and Sweeney, B.W., 2014a. Does DNA barcoding improve performance of traditional stream bioassessment metrics?. *Freshw. Sci.* 33(1), 302–311.

- Stein, E.D., Martinez, M.C., Stiles, S., Miller, P.E., Zakharov, E.V., 2014b. Is DNA barcoding actually cheaper and faster than traditional morphological methods: results from a survey of freshwater bioassessment efforts in the United States. *PLoS One* 9, e95525. <https://doi.org/10.1371/journal.pone.0095525.t003>.
- Sukumaran, S., Vijapure, T., Kubal, P., Mulik, J., Rokade, M.A., Salvi, S., Thomas, J., Naidu, V.S., 2016. Polychaete community of a marine protected area along the west coast of India—prior and post the tropical cyclone Phyan. *PLoS ONE* 11(8), e0159368.
- Sukumaran, S., Saraladevi, K., 2009. Polychaete diversity and its relevance in the rapid environmental assessment of Mumbai Port. *Curr. Sci.* 97(10), 1439-1444.
- Torti, A., Lever, M.A., Jørgensen, B.B., 2015. Origin, dynamics, and implications of extracellular DNA pools in marine sediments. *Mar. Genom.* 24, 185–196.
- Van den Bulcke, L., De Backer, A., Ampe, B., Maes, S., Wittoeck, J., Waegeman, W., Hostens, K. and Derycke, S., 2021. Towards harmonization of DNA metabarcoding for monitoring marine macrobenthos: the effect of technical replicates and pooled DNA extractions on species detection. *Metabarcoding Metagenom.* 5, 71107.
- Vijapure, T., Sukumaran, S., 2019. Optimization of the taxonomic resolution of an indicator taxon for cost-effective ecological monitoring: Perspectives from a heterogeneous tropical coastline. *J. Environ. Manag.* 247, 474-483.
- Vijapure, T., Sukumaran, S., Manohar, C.S., 2019a. Molecular characterization and phylogenetics of Indian polychaete fauna: scope for implementation in ecological monitoring. *Aquat. Ecol.* 53(4), 665-677.
- Vijapure, T., Sukumaran, S., Neetu, S., Chandel, K., 2019b. Macrobenthos at marine hotspots along the northwest Indian inner shelf: Patterns and drivers. *Mar. Environ. Res.* 144, 111-124. <https://doi.org/10.1016/j.marenvres.2018.12.007>.
- Xie, R., Zhao, G., Yang, J., Wang, Z., Xu, Y., Zhang, X., Wang, Z., 2021. eDNA metabarcoding revealed differential structures of aquatic communities in a dynamic freshwater ecosystem shaped by habitat heterogeneity. *Environ. Res.* 201, 111602. <https://doi.org/10.1016/j.envres.2021.111602>.
- Zhou, J., Wu, Y., Kang, Q., Zhang, J., 2007. Spatial variations of carbon, nitrogen and phosphorus in the salt marsh sediments of the Yangtze estuary in China. *Estuar. Coast. Shelf Sci.* 71, 47-59. <https://doi.org/10.1016/j.ecss.2006.08.012>.
- Zingde, M. D. and Desai, B. N., Mercury in Thana Creek, Bombay harbour. *Mar. Pollut. Bull.*, 1981, 12, 237–241.

Tables

Table 1 PCR amplification thermal conditions set for DNA barcoding.

PCR program 1	
Temperature ↓ Time	94°C → [94°C → 48°C → 72°C] → 72°C ----cool---- 4°C 15mins 15 sec 15 sec 1min 5mins end └──────────┘ -----30 cycles-----
PCR program 2	
Temperature ↓ Time	95°C → [94°C → 45°C - 48°C → 72°C] → 72°C ----cool---- 4°C 15mins 30 sec 30 sec 1min 10mins end └──────────┘ -----35 cycles-----

Table 2 List of barcoded species along with NCBI accession numbers from Mumbai and Malvan.

* denotes the species uploaded for the first time in the GenBank.

Species name	NCBI accession number	Species name	NCBI accession number
<i>Magelona cincta</i> *	MZ504161	<i>Isolda pulchella</i>	MZ504199
<i>Nereis</i> sp.	MZ504162	<i>Lumbrinereis</i> sp. *	MZ504200
<i>Goniada maculate</i> *	MZ504163	<i>Euclymene</i> sp. *	MZ504201
<i>Dendronereis arborifera</i> *	MZ504164	<i>Glycera</i> sp.2	MZ504202
<i>Ninoe</i> sp. *	MZ504165	<i>Phyllodoce</i> sp. *	MZ504203
<i>Paraprionospio patiens</i>	MZ504166	<i>Paraprionospio</i> sp. *	MZ504204
<i>Cossura coasta</i> *	MZ504167	<i>Aglaophamus dibranchis</i>	MZ504205
<i>Glycera longipinnis</i>	MZ504168	<i>Lumbrineris</i> sp. *	MZ504206
<i>Eunice</i> sp. *	MZ504169	<i>Armandia</i> sp. *	MZ504207
<i>Eunice</i> sp. *	MZ504170	<i>Glycera</i> sp.	MZ504208
<i>Leiochone</i> sp.	MZ504171	<i>Nereis</i> sp.	MZ504209
<i>Eunice</i> sp. *	MZ504172	<i>Syllis</i> sp. *	MZ504210
<i>Glycera natalensis</i> *	MZ504173	<i>Prionospio</i> sp. *	MZ504211
<i>Goniada emerita</i> *	MZ504174	<i>Scoloplos</i> sp. *	MZ504212
<i>Glycera longipinnis</i>	MZ504175	<i>Micronephtys</i> sp. *	MZ504213
<i>Mediomastus capensis</i> *	MZ504176	<i>Glycera</i> sp.2	MZ504214
<i>Paraprionospio</i> sp. *	MZ504177	<i>Cossura coasta</i> *	MZ504215
<i>Paraprionospio</i> sp. *	MZ504178	<i>Syllis</i> sp. *	MZ504216
<i>Mediomastus capensis</i> *	MZ504179	<i>Lumbrineris</i> sp. *	MZ504217
<i>Aglaophamus dibranchis</i>	MZ504180	<i>Euclymene</i> sp.2 *	MZ504218
<i>Glycera</i> sp.1	MZ504181	<i>Ampelisca</i> sp. *	MZ504219
<i>Polycirrus swakopianus</i> *	MZ504182	<i>Stenothoe</i> sp. *	MZ504220
<i>Glycera</i> sp.	MZ504183	<i>Ampelisca</i> sp. *	MZ504221
<i>Phyllodoce malmgreni</i> *	MZ504184	Gammaridae *	MZ504222
<i>Paleanotus chrysolepis</i> *	MZ504185	<i>Kalliapseudes</i> sp. *	MZ504223
<i>Paraprionospio</i> sp. *	MZ504186	<i>Iphinoe</i> sp. *	MZ504224
<i>Goniada</i> sp. *	MZ504187	<i>Iphinoe</i> sp. *	MZ504225
Polynoidae *	MZ504188	<i>Iphinoe</i> sp. *	MZ504226
<i>Magelona crenulifrons</i> *	MZ504189	Brachyura *	MZ504227
<i>Lepidonotus</i> sp. *	MZ504190	Mytilidae *	MZ504228
<i>Sthenelais boa</i>	MZ504191	<i>Grandidierella</i> sp. *	MZ504229
<i>Nephtys</i> sp.	MZ504192	Aoridae *	MZ504230
<i>Terebellides stroemi</i> *	MZ504193	Ischyroceridae *	MZ504231
<i>Magelona crenulifrons</i> *	MZ504194	<i>Grandidierella</i> sp. *	MZ504232
<i>Goniadiella</i> sp. *	MZ504195	Brachyura *	MZ504233
<i>Aphelochaeta filiformis</i>	MZ504196	Arcidae *	MZ504234
Maldanidae *	MZ504197	Golfingiidae *	MZ504235
<i>Isolda pulchella</i>	MZ504198		

Table 3 Comparison between EcoQS of AMBI and gAMBI based on abundance data and presence/absence (p/a) data in Mumbai and Malvan. ✓ denotes agreement between EcoQS of AMBI and gAMBI; X denotes disagreement between EcoQS of AMBI and gAMBI and **Adj** denotes EcoQS between adjacent categories of AMBI and gAMBI. Blue=High, Green=Good, Yellow=Moderate

Area	Station	Based on abundance			Based on p/a			
		AMBI	gAMBI		AMBI	gAMBI		
Mumbai	1	Moderate	Moderate	✓	Good	Good	✓	
	2	Good	Good	✓	Good	Good	✓	
	3	Moderate	Moderate	✓	Good	Good	✓	
	4	Good	Good	✓	Good	Good	✓	
	5	Moderate	Good	Adj	Moderate	Good	Adj	
	6	Moderate	High	X	Moderate	High	X	
	8	Moderate	Good	Adj	Moderate	Good	Adj	
	9	Good	High	Adj	Good	High	Adj	
	11	Good	Good	✓	Good	Good	✓	
	12	Good	High	Adj	Good	Good	✓	
	Malvan	1	Good	Good	✓	Good	Good	✓
		2	High	High	✓	Good	Good	✓
3		Good	Good	✓	Good	Good	✓	
4		Good	Good	✓	Good	Good	✓	
5		High	Good	Adj	High	Moderate	X	
6		High	High	✓	High	Good	Adj	