Molecular characterization and phylogenetics of Indian polychaete fauna: scope for implementation in ecological monitoring

Tejal Vijapure · Soniya Sukumaran · Cathrine S. Manohar

Abstract DNA barcodes are increasingly applied to ascertain the taxonomic identification to improve the speed and accuracy of ecological monitoring programmes. The success of integrating molecular approach in routine surveys ultimately depends on the coverage of reference libraries that require constant upgradation. The present molecular study was aimed at strengthening the genetic database of Polychaeta, which at present is poorly constructed. The current effort is first of its kind that covered a large geographical area along the northwest India. The study has contributed in building a comprehensive COI database of polychaete taxocene that included new records of one family, four genera and six species. The phylogenetic analysis revealed presence of 19 distinct clades, each comprising of individual family with studied polychaete species and conspecific/congeneric reference sequences. This is the first analysis that revealed a close relationship between Longosomatidae and Cirratulidae, rather than Spioniform polychaetes. Thus, the phylogenetic information was useful in distinguishing the polychaete species in the study region. Molecular analysis also facilitated the identification of potentially new *Streblospio* sp. that displayed close morphological as well as genetic affinity with *S. gynobranchiata,* with an inter-specific distance of 0.11. The present study proves the effectiveness of molecular characterization and phylogenetics in delineating the Indian polychaete species complex for ecological monitoring. The reference database can aid the high-throughput biomonitoring programmes in future.

Keywords Ecological monitoring · DNA barcoding · Polychaeta · Northwest Indian coast

Handling Editor: Télesphore Sime-Ngando.

T. Vijapure · S. Sukumaran (✉)
CSIR-National Institute of Oceanography, Regional Centre, Andheri (W), Mumbai 400 053, India
e-mail: soniya@nio.org

C. S. Manohar
CSIR-National Institute of Oceanography,
Dona Paula, Goa 403 004, India

Introduction

A gradual and continual diversity loss in the marine realm, more so along the coasts that are exposed to a multitude of anthropogenic pressures (Halpern et al. 2008), has propelled the need to regularly monitor the ecological quality of marine and estuarine waters (Erpenbeck et al. 2016). This is generally achieved by establishing species inventories that serve as a baseline against which future perturbations in terms of loss

of diversity can be measured. Consequently, suitable management strategies can be applied in areas of diminishing species richness or in zones where imminent diversity loss is predicted (Thomsen and Willerslev 2015). The biomonitoring mainly relies upon morphological taxonomy (Aylagas et al. 2018) which is often criticized as a labour-intensive, time-consuming and expensive exercise (Lejzerowicz et al. 2015; Bevilacqua and Terlizzi 2016). To circumvent these issues, integrating genetics in bioassessment procedures to complement the traditional taxonomy is a recent and innovative approach (Bik et al. 2012; Sigamani et al. 2016) well explained by Borja et al. (2016).

Macrozoobenthos are potential biological indicators that are extensively employed in monitoring programmes (De Jonge et al. 2006; Aguado-Giménez et al. 2015). A reliable surrogate taxon that can represent the entire benthic assemblage and help in the rapid assessment of an ecosystem health is Polychaeta (Soares-Gomes et al. 2012). Polychaetes comprise of both sensitive and tolerant species occurring in a continuum of habitats from pristine to polluted (Del-Pilar-Ruso et al. 2009). They form one of the dominant groups of macrozoobenthos that are highly diverse with regard to morphologies and ecologies (Struck et al. 2007). Difficulty in morphotaxonomy with lack of taxonomical expertise and sufficient identification keys for this group (Zhou et al. 2010) has lead into overlooking or mis-identification of species. Polychaeta constitutes a diverse component of marine invertebrate community with many species still awaiting description (Lobo et al. 2016). In addition, chances of fragmentation of body parts are high during collection for these soft-bodied worms that may lead to missing body parts crucial for identification (Lobo et al. 2017). To avoid these impediments, the use of complementary tool "DNA barcoding" is a recommended alternative. It can stimulate taxonomic (Goldberg et al. 2016) as well as systematic studies by using a small segment of the genome from damaged individuals, irrespective of life history stages and establish distinct evolutionary lineages (Vogler and Monaghan 2007).

DNA sequences are 'born digital' that can be reserved freely in public reference library and can be referred in future whenever necessary (Bourlat et al. 2013). DNA barcoding based on the sequencing of a small segment of the mitochondrial DNA is one possible way of simplifying (Dasmahapatra and Mallet 2006) and speeding up the evaluation and identification of taxa (Aylagas et al. 2014). Integrating the traditional taxonomy with modern molecular techniques will achieve efficiency in the species identification (Bhadury et al. 2006; Bik 2017). Among the bundle of genes, cytochrome c oxidase subunit I (COI) is the potential marker used in molecular taxonomy (Sun et al. 2012) and has a voluminous database in the reference libraries (Zaiko et al. 2015). Several studies have also suggested the efficacy of COI gene to distinguish polychaete species (Pleijel et al. 2009; Barroso et al. 2010; Nygren and Pleijel 2010; Maturana et al. 2011; Lobo et al. 2016) and their phylogenetics (Norlinder et al. 2012; Kobayashi et al. 2017). The efforts required to barcode the recognized annelid fauna is around seven times the current pace of molecular analyses (Kvist 2013). Hence, to meet this requirement, it is important that molecular barcoding be undertaken in various geographical areas to revitalize taxonomy with genetics around the world. DNA barcoding of polychaetes has also become an integral part of several large-scale projects (Hardy et al. 2011; Brasier et al. 2016; Neal et al. 2018) to improve the efficiency and accuracy of taxon identification. It is also relevant to verify the results of genomic taxonomy vis-à-vis the morphological taxonomy to utilize the technique as a regular tool in environmental monitoring programmes (Borja et al. 2016).

India ranks among one of the 12 mega-diversity countries (Myers et al. 2000) and is one of the few Asian countries that maintains a huge inventory of coastal and marine biodiversity (Venkataraman and Wafar 2005). Estimating the species diversity in a region that exhibits taxonomically complex communities will require a considerable level of taxonomic expertise (Amini-Yekta et al. 2017). Unfortunately, the dearth of trained taxonomists in India, especially for invertebrate groups, is a major concern (Prathapan et al. 2006). Among the macroinvertebrate community, polychaetes dominate the sediments in a range of habitats along Indian waters (Joydas and Damodaran 2009) with approximate species count reaching to 883 (Venkataraman and Wafar 2005). Despite the taxonomic diversity, the few molecular studies that have been carried on the Indian polychaete fauna have either focused on a limited number of species or covered restricted geographic areas (Satheeshkumar and Jagadeesan 2010; Magesh et al. 2012, 2014; Sekar

et al. 2016; Sigamani et al. 2016). The contribution of total polychaete sequences from India deposited at GenBank (excluding the present study) from multiple loci (only 486 records) and mainly COI (only 215 records of 61 species) is meagre (accessed on 09.07.2018). Thus there exists a pressing need than ever before, to widen the taxonomic coverage and capture the information in the "universal taxonomic system" (Tautz et al. 2003), i.e. genetic database. The barcode reference library can ameliorate polychaete identification for ecological monitoring, biological invasions (Geller et al. 2010; Miralles et al. 2016) in the Indian waters and aid in high-throughput sequencing technologies (Mohrbeck et al. 2015; Dowle et al. 2016) in future.

The present molecular survey was executed under the government aided monitoring program, i.e. Coastal Ocean Monitoring and Prediction System (COMAPS) where the Indian coastal waters are monitored regularly for ecological status. This study is the first to cover a large spatial extent of northwest Indian coastal fringe for genetic characterization of Polychaeta. The polychaete COI sequences were acquired from ecological assessment at five transects including two marine protected regions and three active harbours along vast 2360 km long northwest Indian coast. The specific objectives of this study were: (1) to provide an inventory of COI barcodes for enriching the reference library of Indian polychaete fauna, (2) to examine the phylogenetic characterization of polychaete fauna studied using COI marker and (3) to evaluate the efficiency of molecular taxonomy in correctly identifying different morphotypes in the study area. It is expected that in the scenario wherein substantial addition to the DNA library of polychaete species is undertaken and efficiency of molecular techniques is proven to be superior to traditional techniques, it would facilitate speedier and accurate environmental monitoring.

Materials and methods

Sample collection

Sampling was conducted along the northwest coast of India at five transects (Fig. 1). Four stations were sampled perpendicular to the coast along the depth gradient at every transect. Station 1 was located closest to the shoreline at shallow depths (2–5 m), while station 4 was the farthest from the shore and in relatively deep waters (25–30 m). Stations 2 (8–10 m) and 3 (15–20 m) were located at intermediate depths. Vadinar (stations: VD1–VD4) and Veraval (stations: V1–V4) transects of Gujarat state and Mumbai (stations: MY1–MY4), Ratnagiri (stations: R1–R4), Malvan (stations: M1–M4) transects of Maharashtra state were surveyed. Vadinar (22°N; 69°E), situated on the southern flank of Gulf of Kachchh (GoK), is a Marine National Park and Sanctuary (MNPS), designated to conserve the coral reefs and the rich biodiversity (Sukumaran et al. 2013). Veraval (20°N; 70°E) is one of the biggest commercial fishing centres in Asia (Misra and Kundu 2005). Mumbai (18°N; 72°E) is India's premier port that has played a substantial role in escalating the country's trade and commerce (Gaonkar et al. 2010). Ratnagiri (17°N; 73°E) is one of the important harbours along Maharashtra coast (Ingole et al. 2009) and major marine fish landing centres in India. Malvan (16°N; 73°E) is the only coralline marine protected area along Maharashtra coast that sustains rich biodiversity (Sukumaran et al. 2016). The sampling was conducted at each transect during premonsoon, postmonsoon and monsoon seasons of the year 2013–2014 (Appendix A).

At each station, a sediment sample was collected using a van Veen grab of $0.04 \, \mathrm{m^{-2}}$ bite area. Altogether, 60 grab samples from 20 locations were collected during three sampling campaigns. The sediment samples were immediately sieved through 0.5-mm-mesh sieve, and all the animals retained were fixed in 96% ethanol. The polychaetes were then sorted in the laboratory with the aid of a stereozoom microscope (Leica S6E). The specimens were morphologically identified to the lowest possible taxonomic resolution using the available literature (Day 1967; Fauchald 1977) and updated identification keys (Nogueira et al. 2013; Blake 2016). Minimum of one specimen was selected for the molecular identification from each species identified.

DNA extraction, amplification and sequencing

Genomic DNA of the morphologically identified polychaete specimens were extracted with the DNeasy Blood and Tissue Kit (Qiagen) as per manufacturer protocol. The quality of the DNA was tested, and the barcode region of COI gene (COI-5P) was amplified
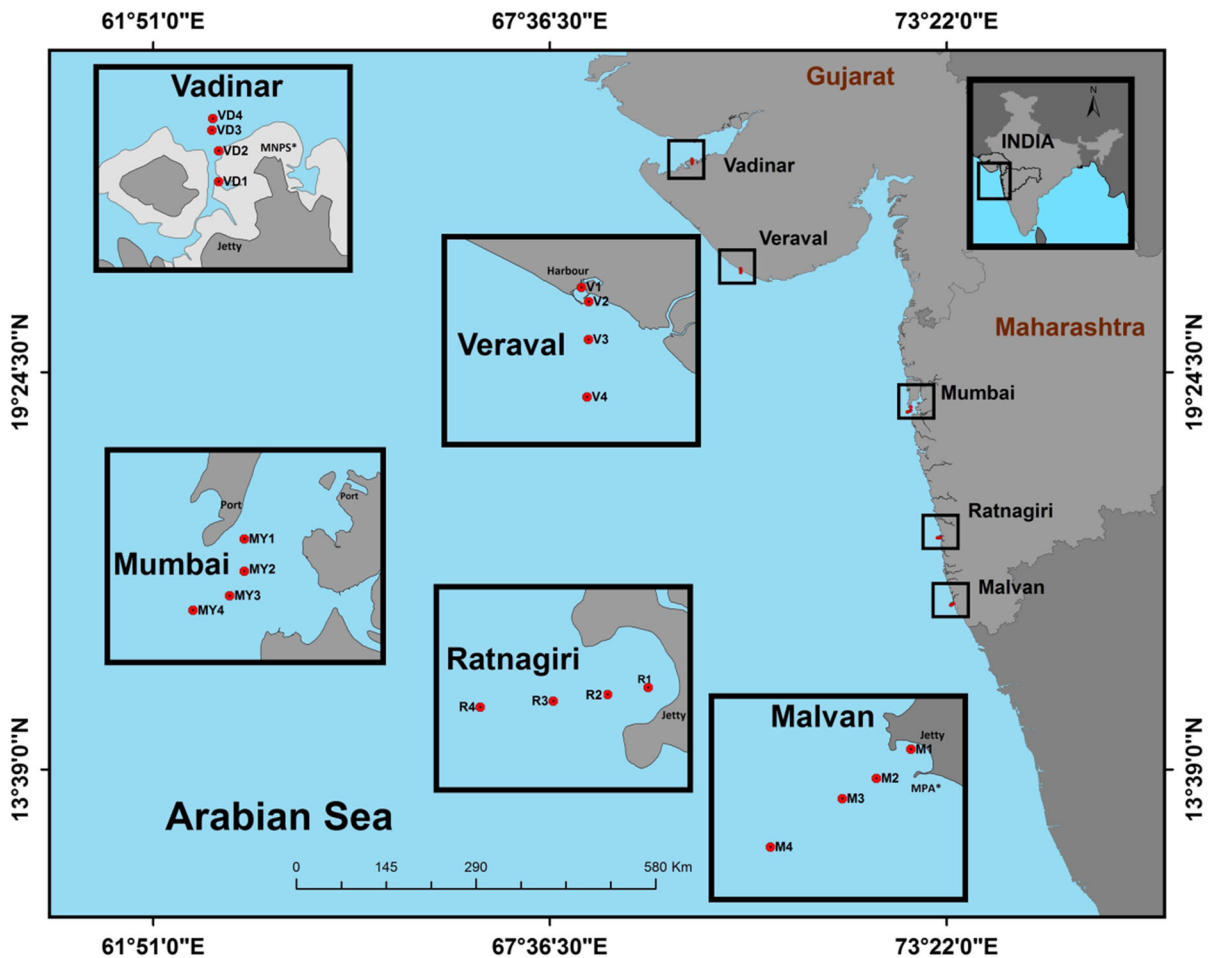
Fig. 1 Study area along northwest Indian coastline showing five investigated transects and the station locations. MNPS* Marine National Park and Sanctuary, MPA* Marine Protected Area

with specific primers (Table 1). Each PCR mixture contained 2.5 µl of Qiagen PCR buffer, 0.5 µl dNTPs, 19.5 µl ddH$_2$O, 0.5 µl of each forward and reverse primer (10 mM), 0.3 µl of (1.5 mM) MgCl$_2$, 0.2 µl Taq polymerase and 1 µl template DNA, making a total volume of 25 µl. Amplifications (for both primer pairs) were carried out in Eppendorf Mastercycler under the following thermal conditions: 94 °C for 3.0 min, 30 cycles of 94 °C for 1 min, 48 °C for 1 min

and 72 °C for 2 min. A final elongation at 72 °C was carried out for 10 min followed by cooling the reaction at 4 °C. Amplified products were tested on a 0.8% agarose gel. The PCR products were purified with gel extraction or column purification method with Invitrogen Purelink Quick Gel Extraction and PCR Purification Combo kit, following the manufacturer's protocols. The purified PCR products were sequenced bidirectionally by the ABi 3730XL Genetic Analyzer:

Table 1 COI primer sets used in the present study

Primer		References
HCO 2198	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al. (1994)
LCO 1490	GGTCAACAAATCATAAAGATATTGG	
LoboF1	KBTCHACAAAYCAYAARGAYATHGG	Lobo et al. (2013)
LoboR1	TAAACYTCWGGRTGWCCRAARAAYCA	

96 capillary sequencer. All the sequences generated in the present study have been deposited in GenBank.

Sequence data analyses

The study sequences were aligned using Clustal W with default settings in MEGA v. 7.0 (Kumar et al. 2016). Additional COI sequences were mined from NCBI database that closely related to the polychaete taxa present in this study. Phylogenetic analysis was performed using Neighbour-Joining (NJ) and Maximum Likelihood (ML) method using the Kimura-2-parameter (K2P) model (Kimura 1980). Bootstrap analysis with 1000 replicates was employed to estimate the node/clade support. The outgroup taxa selected to root the phylogenetic analyses were mollusc species. For the phylogenetic tree building, the study sequences were coded from P1 to P58 (Appendix A). The genetic distances were calculated using a Kimura 2-parameter (K2P) distance model at species and family levels, respectively.

An independent analysis was performed for *Streblospio* species complex following the steps mentioned above. For this study, a dedicated dataset which comprised of 15 *Streblospio* sp. from the present study with three congeneric species—*Streblospio gynobranchiata* Rice and Levin, 1998; *S. benedicti* Webster, 1879 and *S. shrubsolii* Buchanan, 1890 (Mahon et al. 2009) was employed.

Results

Taxa diversity

The current study contributed significantly to the existing Indian polychaete COI database in the GenBank by adding 58 sequences to the existing sequences. Thirty-one morphospecies belonging to 19 families were successfully barcoded. The maximum number of polychaete barcodes were generated from Ratnagiri ($n = 13$) followed by Vadinar ($n = 10$) Mumbai ($n = 8$) Malvan ($n = 7$) and Veraval ($n = 6$), respectively. Some species were present at more than one transect. In these cases, the specimens of the same species were barcoded at every transect it was present. *Leiochone* sp., *Glycera longipinnis* Grube, 1878; *Aglaophamus dibranchis* Grube, 1877, were barcoded from three transects, while

Paraprionospio patiens Yokoyama, 2007; *Heterospio indica* Parapar, Vijapure, Moreira and Sukumaran, 2016; *Sternaspis scutata* Ranzani, 1817; *Sthenelais boa* Johnston, 1833; *Gymnonereis* sp. and *Sigambra parva* Day, 1963, were barcoded from two transects. The rest of the species barcoded, occurred in a single transect (Appendix B). Each specimen obtained in the study area was barcoded to produce a single representative sequence, except that of a spionid worm *Streblospio* sp. The sequences from 15 specimens of *Streblospio* sp. were generated and analysed to ascertain the novelty of the morphotype considering that the specimens were morphologically similar to the well-described *S. gynobranchiata*. Majority of the sequences (96%) were amplified using a primer set LCO1490/HCO2198 while the remaining (4%) with LoboF1/LoboR1. The current study was the first to deposit the sequences of the monogeneric polychate family Longosomatidae. These sequences were used to flag a new species of *Heterospio indica* (Parapar et al. 2016) Also, the sequences of three potential new species (as per morphotaxonomy), i.e. *Streblospio* sp., *Polycirrus* sp. and *Kirkegaardia* sp., were submitted. The genus *Kirkegaardia* is a new record for the Indian west coast. Thus, the study added new records for four genera (*Dendronereides, Leonnates, Gymnonereis* and *Mysta*) and six species (*Scoloplos uniramus* Day, 1963; *Paraprionospio cristata* Zhou, Yokoyama and Li 2008; *Paraprionospio patiens* Yokoyama, 2007; *Glycera longipinnis* Grube, 1878; *Aphelochaeta filiformis* Keferstein, 1862; *Sigambra parva* Day, 1963). The details of the study sequences along with their accession numbers are presented in Appendix A.

Phylogenetic analysis

The DNA data matrix of 105 COI sequences that included study ($n = 58$) and reference ($n = 42$) alignments with outgroup taxa ($n = 5$) was used to construct a phylogenetic tree (Appendix A). Both ML and NJ trees were constructed and as they displayed similar topology, the ML tree was displayed along with the significant bootstrap values from NJ. The final phylogenetic consensus tree is depicted in Fig. 2. The database for family Longosomatidae did not have any prior sequence depositions, except our two *Heterospio* specimens (P22 and P23). Hence, no reference sequence could be included exclusively for this family. The other reference sequences selected

Fig. 2 Maximum likelihood (ML) tree of polychaeta cytochrome c oxidase subunit I sequences. Branch support values (higher than 50%) for maximum likelihood and neighbour-joining are displayed, respectively (_/_)
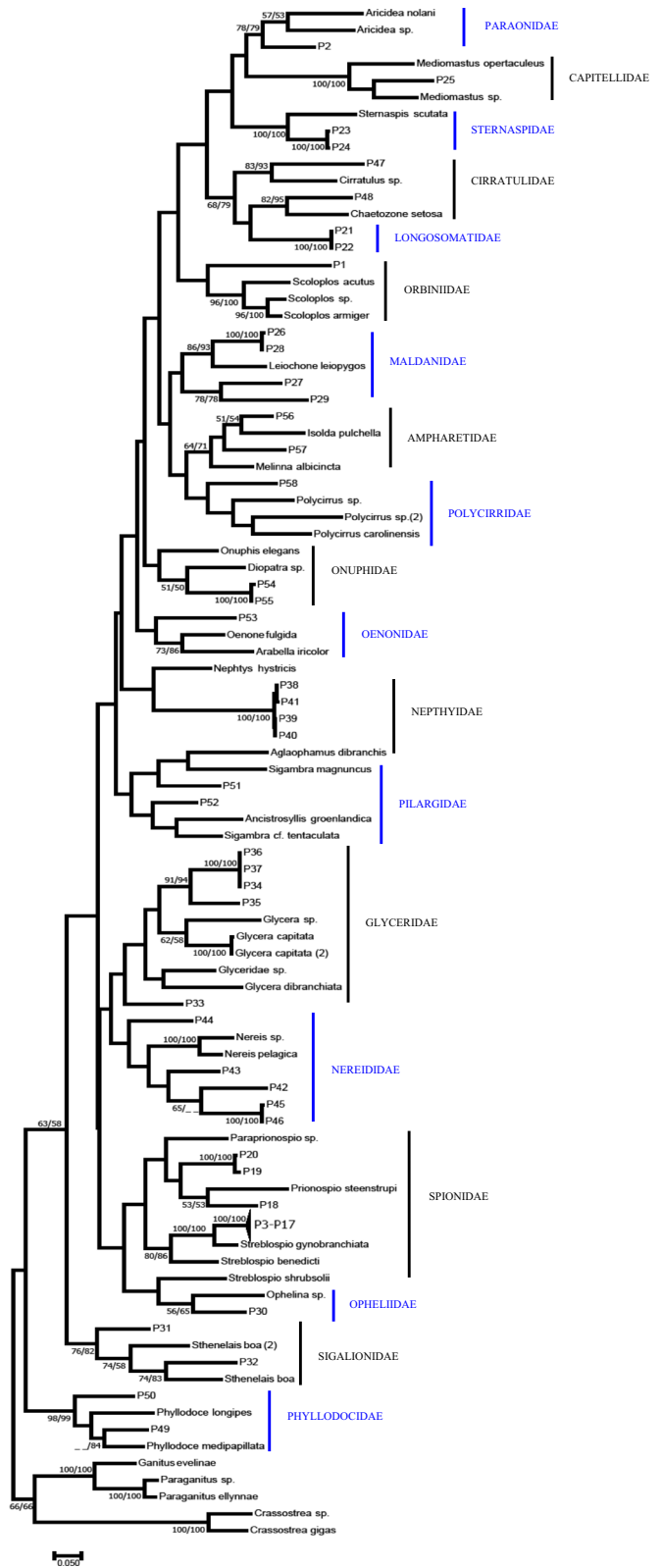
were of same genera of the studied taxa or members belonging to the corresponding family (Appendix A). All the sequences were clustered as per their taxonomic hierarchy into respective families, forming 19 clades. The analysis showed that the members of the errant polychaete taxa, i.e. Phyllodocidae (98/99), formed the most basal group followed by Sigalionidae (76/82). Apart from these two families, the entire tree was built with good support of 63/58. The sedentarian polychaete families formed the tree apex with nine clades (Paraoniidae to Polycirridae), while Spionidae and Opheliidae were placed immediately above the basal groups. The remaining six clades (Onuphidae to Nereididae) constituted of errant worms. Well-supported clades (100/100) found in the present study belonged to Capitellidae and Sternaspidae. The newly deposited Longosomatidae family was found to cluster with Cirratulidae with a good node support (68/79).

The minimum cofamilial distance noted (0.23) was for Polycirridae and Ampharetidae. The maximum divergence (0.50) recorded was for Phyllodocidae and Orbiniidae. Among the species, the lowest inter-specific distance observed (0.002) was due to low variance between *Streblospio* sp. (P7 and P9, P16) and *Gymnonereis* sp. (P45 and P46). *Kirkegaardia* sp. (P47) and *Phyllodoce longipes* Kinberg, 1866 (P49) displayed the maximum distance (0.55). The study sequences successfully nested with conspecific or congeneric reference sequences, forming 19 clades of respective families.

Streblospio species complex

The *Streblospio* species complex comprised of three known species i.e. Streblospio gynobranchiata, *S. benedicti* and *S. shrubsolii* along with *Streblospio* sp. from the present study. A comprehensive molecular analysis of all four species was performed with NJ and ML trees (Fig. 3). The outgroup taxa belonged to molluscan genera *Crassostrea*. The position of *Streblospio* sp. was well supported by a bootstrap value of 99/99. The analysis suggested that all four *Streblospio* species were reciprocally monophyletic. All the alignments (*n* = 15) obtained from the present study were closely related to *S. gynobranchiata* population with lowest inter-specific distance of 0.11, though forming a separate distinct cluster. Relatively greater distances were noted between *S. benedicti* (0.21), *S. shrubsolii* (0.29) and *Streblospio* sp. The *Streblospio*

sp. also differed from *S. gynobranchiata* based on distinct morphological characters like number of seminal receptacles in females and their arrangement in the chaetigers.

Discussion

Genomic monitoring of marine ecosystems has offered an innovative approach for the rapid evaluation of the ecostatus (Hajibabaei et al. 2011; Avó et al. 2017) as an alternative to the traditional biomonitoring (Aylagas et al. 2018). Among the biotic components, polychaetes are the potential candidates that are widely used for monitoring the benthic ecosystems (Lobo et al. 2016). However, the studies related to molecular characterization of polychaete fauna along the Indian waters are very limited. Satheeshkumar and Jagadeesan (2010) studied the phylogenetic position and genetic diversity of two Nereidae species, *Pseudonereis variegata* Grube, 1857 and *Namalycastis indica* Southern, 1921 from Pondicherry mangroves, based on 16S gene. Magesh et al. (2012) provided morphological and molecular phylogenetics of a new nereid species, *Namalycastis jaya* Magesh, Kvist and Glasby, 2012 from Kerala, southwest India by using both mitochondrial (COI, 16S) and nuclear (18S) markers. Magesh et al. (2014) redescribed a nereid, *Namalycastis glasbyi* Fernando and Rajasekaran, 2007 from Gorai Creek, Mumbai, west Indian coast, which was supplemented with COI sequences. Sekar et al. (2016) generated DNA barcodes for six polychaete species from Andaman and Nicobar Islands while Sigamani et al. (2016) barcoded 13 species from the Vellar–Coleroon estuarine system, east coast of India. As compared to the aforementioned studies, the present study investigated the molecular taxonomy of a considerable number of species from a broader geographic area. The survey resulted in a comprehensive reference library of 58 COI barcodes from 31 polychaete species, of which one family, four genera and six species were sequenced for the first time.

Advanced, sensitive and cost-effective DNA sequencing technology has lead to high-throughput techniques like metabarcoding that are recently being employed in monitoring studies (Lejzerowicz et al. 2015; Lobo et al. 2017; Aylagas et al. 2018). However, the success of these advanced techniques mainly

Fig. 3 Phylogenetic
relationships of *Streblospio*
species complex. Maximum
likelihood (ML) tree is
displayed where the branch
support values (higher than
50%) from maximum
likelihood and neighbour-
joining, respectively (_/_)
are depicted at the nodes
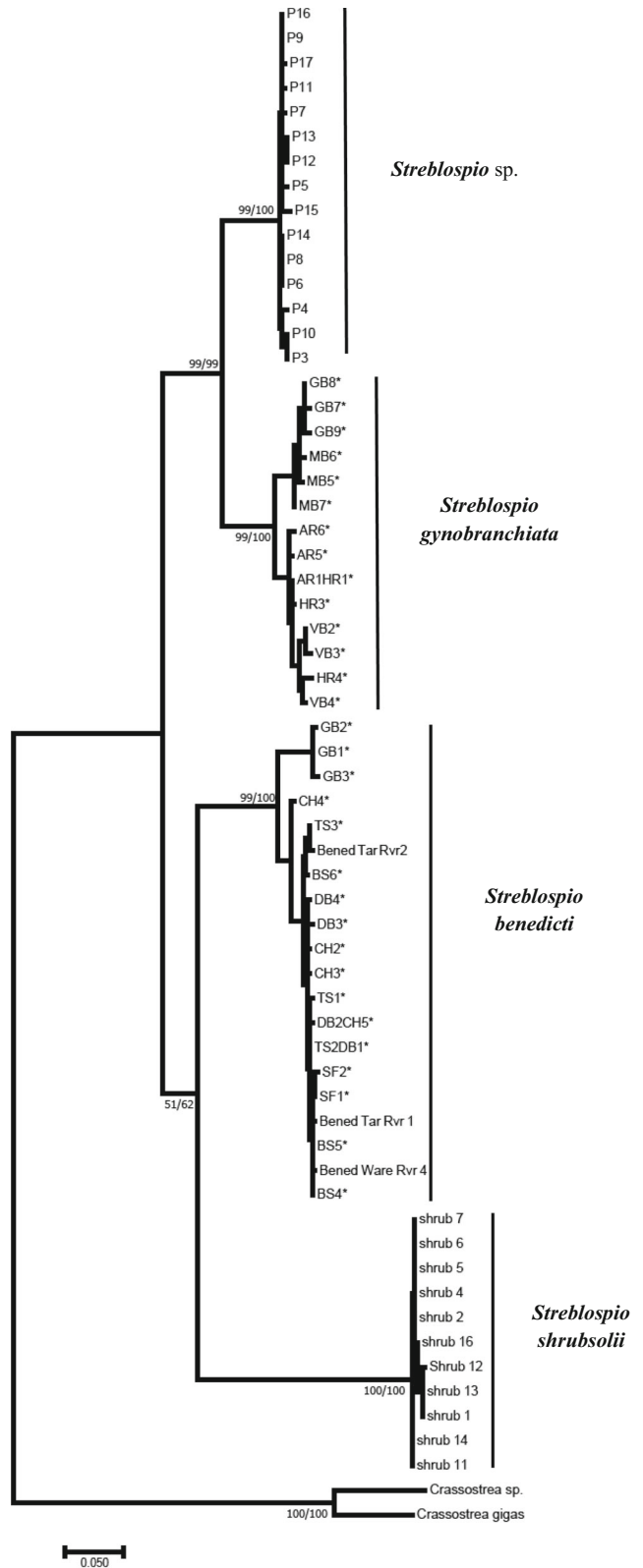
depends on the coverage of the reference sequence database (Larras et al. 2014). Dowle et al. (2016) pointed that the results of high-throughput techniques employed for monitoring of New Zealand rivers using macrozoobenthos would have been superior, if well-defined and validated reference library was available. As DNA barcoding and taxonomic identification complement each other, the lack of species or families reference sequences in the public database may not connect to a taxonomically known species (Brasier et al. 2016). Aylagas et al. (2014) suggested that at least the most common and abundant species encountered during the monitoring programmes should be sequenced with accurate taxonomic tagging to reliably assess the ecological state. Bourlat et al. (2013) opined that the investments made for building the DNA barcode reference libraries are worthwhile considering the accrued benefits. Future biomonitoring will be surely benefited from the constant improvement in the present sequence libraries (Curry et al. 2018). The current effort has expanded the DNA barcode repository of the Indian polychaetes with the sequence submission of polychaetes that are commonly found during ecological monitoring surveys.

Besides the technical advancements, efforts are now focussed towards modifying the traditional indices with genetic support for ecological monitoring purposes. The genomic-based biotic index like g-AMBI (Aylagas et al. 2014) has been developed recently as an alternative for harmonized monitoring, particularly for areas that lack taxonomic expertise and have monetary constrains. The index is based on AZTI's Marine Biotic Index AMBI (Borja et al. 2000) that is one among the plethora of indices successfully tested worldwide for ecostatus evaluation using benthic species. The AMBI index is also successfully applied in Indian water bodies (Sigamani et al. 2015; Feebarani et al. 2016; Mulik et al. 2017). However, the viability of g-AMBI is greatly dependent on the reference library. Sigamani et al. (2016) have urged to barcode the entire species spectrum available in the conventional AMBI list for the successful application of g-AMBI. Among the polychaete species barcoded in the present study, 79% were present in the AMBI list. Thus, the study has provided the genetic database of polychaete species widely used in ecological monitoring that will assist in the estimation of g-AMBI in Indian waters.

The traditional morphotaxonomic approach is now revitalized with molecular and phylogenetic analysis in monitoring programmes (Bhadury et al. 2006), especially in the case of polychaete fauna (Hutchings and Kupriyanova 2018). However, characterizing the phylogenetic relationships among the polychaete taxon is a complex exercise that requires the involvement of both classical and molecular taxonomists (Canales-Aguirre et al. 2011). The current work indicated congruence between the morphological and molecular data that was reflected in the topologies. In the present study, the sequences of morphologically identified taxa were appropriately nested in the clades of respective families with conspecific or congeneric reference sequences as also observed by Lobo et al. (2016). Thus, the phylogenetic analysis produced reliable conclusions in the present study.

The current work has provided the first entry of the monogeneric family Longosomatidae in the GenBank, flagging the species *Heterospio indica* (Parapar et al. 2016) with phylogenetic analysis. Longosomatidae had been placed in the clade Spionida according to the phylogenetic analysis based on morphological features (Rouse and Fauchald 1997) and reproductive, larval strategies (Blake and Arnofsky 1999). However, Wilson (2000) considered it more similar to Cirratulidae as both the families share conspicuous characters like expanded posterior segment and chaetal cinctures. Hence, Parapar et al. (2014) urged for a better clarification of the phylogenetic position of Longosomatidae. The current phylogenetic analysis is the first to clarify the existing ambiguity and reveal a close relationship between Longosomatidae and Cirratulidae, rather than with Spioniform polychaetes. An extensive phylogenetic analysis of the terebelliforms by Nogueira et al. (2013) incorporated eight different families that included Ampharetidae and considered Polycirridae as a separate family. These two terebelliform sedentarian families indicated the lowest cofamilial distance (0.23) in the present study. However, the maximum distance recorded was between the errant Phyllodocidae and the sedentarian Orbiniidae which is expected considering that both belong to distant clades of Scolecida and Aciculata, respectively (Rousset et al. 2007).

The species of the opportunistic genus *Streblospio* are known to dominate the polluted harbours and ports around the world (Çinar et al. 2012; Radashevsky and Selifonova 2013). In the present investigation,

Streblospio sp. was observed to be predominant in the polluted sediments of Veraval harbour having anoxic water conditions (Sundararajan et al. 2017; Majithiya et al. 2018). The *Streblospio* species are potential pollution indicators (Pelletier et al. 2010) that mainly thrive in stressed habitats. According to AZTI's classification (June 2017 species list; http://ambi.azti.es), the *Streblospio* species were categorized in Ecological Groups (EGs) III (tolerant species) and IV (opportunistic species). The *Streblospio* sp. encountered in the present study is a potential new species (MS in progress) that displayed close morphological resemblance with *S. gynobranchiata*. Though the genetic affinity between the two was evident with lowest inter-specific distance (0.11) as compared to other congeners, they were separated into two different clades representing separate species. Comparatively similar values were reported by Barroso et al. (2010) to distinguish congeneric pollution indicator species of *Eurythoe* (0.10). He emphasized on the genetic scrutiny of mainly the pollution indicator species as this method can reliably resolve the masked phenotypic differences. The presence of the existing robust genetic database for the invasive *Streblospio* species (Zenetos et al. 2005) could facilitate the identification of a prospective new *Streblospio* species in this study, thus preventing possible mis-identification. As the members of *Streblospio* species are expanding their biogeographic provinces, it becomes imperative to accurately identify the species for appropriate ecostatus evaluation and avoid the negative impacts of their invasions (Mahon et al. 2009).

Conclusion

Molecular characterization of organisms is advancement on the traditional microscopy-based taxonomical studies and is currently being employed in some areas leading to speedier and accurate species identification. One of the impediments to the progress of DNA taxonomy is the insufficiency of the DNA barcode library in many regions. The success of using genetic tools in an area depends mainly on a well-represented species library of DNA barcodes. Thus, the current study has added DNA sequences to the existing repository of the Indian polychaete species significantly ($\sim 21\%$). The study is first to sample over a large spatial extent of the Indian coastal waters and create a substantial genetic database of polychaete fauna. The study generated a comprehensive reference library of 58 COI barcodes of 31 polychaete species belonging to 19 families, of which one family, four genera and six species were sequenced successfully for the first time. The congruency between molecular and taxonomic tools has proved the efficacy of DNA barcoding in the identification of polychaete species. The molecular analysis facilitated the identification of a prospective new *Streblospio* species by segregating them from the closely related congeners. The utility of the genetic analysis over morphological identification was amply demonstrated in the study, where despite the masked phenotypic differences, the opportunistic *Streblospio* species was proven to be a new taxon thus preventing a possible mis-identification. The findings from the present study recommend a reliable integration of molecular approach in routine monitoring programmes for accurate and rapid identification of indicator taxon, Polychaeta. The enriched database of polychaete barcodes can act as a reference library for ecological monitoring studies that may include next generation marine monitoring programs in future.

Data availability The datasets generated during and/or analysed during the current study are available in the GenBank repository [https://www.ncbi.nlm.nih.gov/genbank/]. All data generated or analysed during this study is provided in this article as a supplementary information file [Appendix A and Appendix B].

Compliance with ethical standards

Human and animal rights This article does not contain any studies with human participants or animals performed by any of the authors.

Ethical approval All necessary permits for sampling and observational field studies have been obtained by the authors from the competent authorities.

References

Aguado-Giménez F, Gairín JI, Martinez-Garcia E, Fernandez-Gonzalez V, Moltó MB, Cerezo-Valverde J, Sanchez-Jerez P (2015) Application of "taxocene surrogation" and "taxonomic sufficiency" concepts to fish farming environmental monitoring, comparison of BOPA index versus polychaete assemblage structure. Mar Environ Res 103:27–35

Amini-Yekta F, Izadi S, Asgari M, Aghajan-Pour F, Shokri MR (2017) Higher taxa as surrogates for species richness in intertidal habitats of Qeshm Island in the Persian Gulf. Mar Biodivers 48(3):1421–1428

Avó AP, Daniell TJ, Neilson R, Oliveira S, Branco J, Adão H (2017) DNA barcoding and morphological identification of benthic nematodes assemblages of estuarine intertidal sediments: advances in molecular tools for biodiversity assessment. Front Mar Sci 4:66

Aylagas E, Borja Á, Rodríguez-Ezpeleta N (2014) Environmental status assessment using DNA metabarcoding: towards a genetics based marine biotic index (gAMBI). PLoS ONE 9(3):e90529

Aylagas E, Borja Á, Muxika I, Rodríguez-Ezpeleta N (2018) Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. Ecol Indic 95:194–202

Barroso R, Klautau M, Solé-Cava AM, Paiva PC (2010) *Eurythoe complanata* (Polychaeta: Amphinomidae) the 'cosmopolitan' fireworm consists of at least three cryptic species. Mar Biol 157(1):69–80

Bevilacqua S, Terlizzi A (2016) Species surrogacy in environmental impact assessment and monitoring: extending the BestAgg approach to asymmetrical designs. Mar Ecol Prog Ser 547:19–32

Bhadury P, Austen MC, Bilton DT, Lambshead PJD, Rogers AD, Smerdon GR (2006) Development and evaluation of a DNA-barcoding approach for the rapid identification of nematodes. Mar Ecol Prog Ser 320:1–9

Bik HM (2017) Let's rise up to unite taxonomy and technology. PLoS Biol 15(8):e2002231

Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK (2012) Sequencing our way towards understanding global eukaryotic biodiversity. Trends Ecol Evol 27(4):233–243

Blake JA (2016) *Kirkegaardia* (Polychaeta Cirratulidae) new name for *Monticellina* Laubier preoccupied in the Rhabdocoela together with new records and descriptions of eight previously known and sixteen new species from the Atlantic Pacific and Southern Oceans. Zootaxa 4166(1):1–93

Blake JA, Arnofsky PL (1999) Reproduction and larval development of the spioniform Polychaeta with application to systematics and phylogeny. Hydrobiologia 402:57–106

Borja A, Franco J, Pérez V (2000) A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Mar Pollut Bull 40(12):1100–1114

Borja A, Elliott M, Andersen JH, Berg T, Carstensen J, Halpern BS et al (2016) Overview of integrative assessment of marine systems: the ecosystem approach in practice. Front Mar Sci 3:20

Bourlat SJ, Borja A, Gilbert J, Taylor MI, Davies N, Weisberg SB et al (2013) Genomics in marine monitoring: new opportunities for assessing marine health status. Mar Pollut Bull 74(1):19–31

Brasier MJ, Wiklund H, Neal L, Jeffreys R, Linse K, Ruhl H, Glover AG (2016) DNA barcoding uncovers cryptic diversity in 50% of deep-sea Antarctic polychaetes. R Soc Open Sci 3(11):160432

Canales-Aguirre CB, Rozbaczylo N, Hernández CE (2011) Genetic identification of benthic polychaetes in a biodiversity hotspot in the southeast Pacific. Rev Biol Mar Oceanogr 46(1):89–96

Çinar ME, Katagan T, Öztürk B, Bakir K, Dagli E, Açik S et al (2012) Spatio-temporal distributions of zoobenthos in soft substratum of Izmir Bay (Aegean Sea eastern Mediterranean) with special emphasis on alien species and ecological quality status. J Mar Biol Assoc UK 92(7):1457–1477

Curry CJ, Gibson JF, Shokralla S, Hajibabaei M, Baird DJ (2018) Identifying North American freshwater invertebrates using DNA barcodes: are existing COI sequence libraries fit for purpose? Freshw Sci 37(1):178–189

Dasmahapatra KK, Mallet J (2006) DNA barcodes: recent successes and future prospects. Heredity 97(4):254–255

Day JH (1967) A monograph on the polychaete of southern Africa part I (Errantia) and part II (Sedentaria), Trustees of the British Museum Natural History London pp 1–878

De Jonge VN, Elliott M, Brauer VS (2006) Marine monitoring: its shortcomings and mismatch with the EU Water Framework Directive's objectives. Mar Pollut Bull 53(1–4):5–19

Del-Pilar-Ruso Y, De-la-Ossa-Carretero JA, Loya-Fernández A, Ferrero-Vicente LM, Giménez-Casalduero F, Sánchez-Lizaso JL (2009) Assessment of soft-bottom Polychaeta assemblage affected by a spatial confluence of impacts: sewage and brine discharges. Mar Pollut Bull 58(5):765–786

Dowle EJ, Pochon XC, Banks J, Shearer K, Wood SA (2016) Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: a case study using freshwater macroinvertebrates. Mol Ecol Resour 16(5):1240–1254

Erpenbeck D, Voigt O, Al-Aidaroos AM, Berumen ML, Büttner G, Catania D et al (2016) Molecular biodiversity of Red Sea demosponges. Mar Pollut Bull 105(2):507–514

Fauchald K (1977) The polychaete worms definitions and keys to the orders families and genera. Nat Hist Mus Los Angel Cty Sci Ser 28:1–190

Feebarani J, Joydas TV, Damodaran R, Borja A (2016) Benthic quality assessment in a naturally-and human-stressed tropical estuary. Ecol Indic 67:380–390

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mithochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Marine Biol Biotechnol 3(5):294–299

Gaonkar CA, Sawant SS, Anil AC, Venkat K, Harkantra SN (2010) Mumbai harbour India: gateway for introduction of marine organisms. Environ Monit Assess 163(1–4):583–589

Geller JB, Darling JA, Carlton JT (2010) Genetic perspectives on marine biological invasions. Annu Rev Mar Sci 2:367–393

Goldberg CS, Turner CR, Deiner K, Klymus KE, Thomsen PF, Murphy MA et al (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods Ecol Evol 7(11):1299–1307

Hajibabaei M, Shokralla S, Zhou X, Singer GA, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE 6(4):e17497

Halpern BS, Walbridge S, Selkoe KA, Kappel CV, Micheli F, D'agrosa C et al (2008) A global map of human impact on marine ecosystems. Science 319(5865):948–952

Hardy SM, Carr CM, Hardman M, Steinke D, Corstorphine E, Mah C (2011) Biodiversity and phylogeography of Arctic marine fauna: insights from molecular tools. Mar Biodivers 41(1):195–210

Hutchings P, Kupriyanova E (2018) Cosmopolitan polychaetes–fact or fiction? Personal and historical perspectives. Invertebr Syst 32(1):1–9

Ingole B, Sivadas S, Nanajkar M, Sautya S, Nag A (2009) A comparative study of macrobenthic community from harbours along the central west coast of India. Environ Monit Assess 154(1–4):135

Joydas TV, Damodaran R (2009) Infaunal macrobenthos along the shelf waters of the west coast of India Arabian Sea. Indian J Mar Sci 38:191–204

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16(2):111–120

Kobayashi G, Mukai R, Alalykina I, Miura T, Kojima S (2017) Phylogeography of benthic invertebrates in deep waters: a case study of Sternaspis cf williamsae (Annelida: Sternaspidae) from the northwestern Pacific Ocean. Deep Sea Res Part II Top Stud Oceanogr 154:159–166

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874

Kvist S (2013) Barcoding in the dark? A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. Mol Phylogenet Evol 69(1):39–45

Larras F, Keck F, Montuelle B, Rimet F, Bouchez A (2014) Linking diatom sensitivity to herbicides to phylogeny: A step forward for biomonitoring? Environ Science Technol 48(3):1921–1930

Lejzerowicz F, Esling P, Pillet L, Wilding TA, Black KD, Pawlowski J (2015) High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. Sci Rep 5:13932

Lobo J, Costa PM, Teixeira MA, Ferreira MS, Costa MH, Costa FO (2013) Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. BMC Ecol 13(1):34

Lobo J, Teixeira MA, Borges LM, Ferreira MS, Hollatz C, Gomes PT et al (2016) Starting a DNA barcode reference library for shallow water polychaetes from the southern European Atlantic coast. Mol Ecol Resour 16(1):298–313

Lobo J, Shokralla S, Costa MH, Hajibabaei M, Costa FO (2017) DNA metabarcoding for high-throughput monitoring of estuarine macrobenthic communities. Sci Rep 7(1):15618

Magesh M, Kvist S, Glasby CJ (2012) Description and phylogeny of Namalycastis jaya sp. n. (Polychaeta, Nereididae, Namanereidinae) from the southwest coast of India. ZooKeys 238:31–43

Magesh M, Glasby CJ, Kvist S (2014) Redescription of Namalycastis glasbyi Fernando, Rajasekaran 2007 (Annelida, Nereididae, Namanereidinae) from India. Proc Biol Soc Wash 127(3):455–465

Mahon AR, Mahon HK, Dauer DM, Halanych KM (2009) Discrete genetic boundaries of three Streblospio (Spionidae, Annelida) species and the status of S. shrubsolii. Mar Biol Res 5(2):172–178

Majithiya D, Yadav A, Ram A (2018) Behaviour of trace metals in the anoxic environment of Veraval harbour India. Mar Pollut Bull 129(2):645–654

Maturana CS, Moreno RA, Labra FA, González-Wevar CA, Rozbaczylo N, Carrasco FD, Poulin E (2011) DNA barcoding of marine polychaetes species of southern Patagonian fjords. Rev Biol Mar Oceanogr 46(1):35–42

Miralles L, Ardura A, Arias A, Borrell YJ, Clusa L, Dopico E et al (2016) Barcodes of marine invertebrates from north Iberian ports: Native diversity and resistance to biological invasions. Mar Pollut Bull 112(1–2):183–188

Misra S, Kundu R (2005) Seasonal variations in population dynamics of key intertidal molluscs at two contrasting locations. Aquat Ecol 39(3):315–324

Mohrbeck I, Raupach MJ, Arbizu PM, Knebelsberger T, Laakmann S (2015) High-throughput sequencing- The key to rapid biodiversity assessment of marine metazoa? PLoS ONE 10(10):e0140342

Mulik J, Sukumaran S, Srinivas T, Vijapure T (2017) Comparative efficacy of benthic biotic indices in assessing the Ecological Quality Status (EcoQS) of the stressed Ulhas estuary India. Mar Pollut Bull 120(1–2):192–202

Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403(6772):853

Neal L, Linse K, Brasier MJ, Sherlock E, Glover AG (2018) Comparative marine biodiversity and depth zonation in the Southern Ocean: evidence from a new large polychaete dataset from Scotia and Amundsen seas. Mar Biodivers 48(1):581–601

Nogueira JMM, Fitzhugh K, Hutchings P (2013) The continuing challenge of phylogenetic relationships in Terebelliformia (Annelida: Polychaeta). Invertebr Syst 27(2):186–238

Norlinder E, Nygren A, Wiklund H, Pleijel F (2012) Phylogeny of scale-worms (Aphroditiformia Annelida) assessed from 18SrRNA 28SrRNA 16SrRNA mitochondrial cytochrome

c oxidase subunit I (COI) and morphology. Mol Phylogenet Evol 65(2):490–500

Nygren A, Pleijel F (2010) Redescription of *Imajimaea draculai*—a rare syllid polychaete associated with the sea pen *Funiculina quadrangularis*. J Mar Biol Assoc UK 90(7):1441–1448

Parapar J, Aguirrezabalaga F, Moreira J (2014) First record of Longosomatidae (Annelida: Polychaeta) from Iceland with a worldwide review of diagnostic characters of the family. J Nat Hist 48(17–18):983–998

Parapar J, Vijapure T, Moreira J, Sukumaran S (2016) A new species of *Heterospio* (Annelida Longosomatidae) from the Indian Ocean. Eur J Taxon 220:1–17

Pelletier MC, Gold AJ, Heltshe JF, Buffum HW (2010) A method to identify estuarine macroinvertebrate pollution indicator species in the Virginian Biogeographic Province. Ecol Indic 10(5):1037–1048

Pleijel F, Rouse G, Nygren A (2009) Five colour morphs and three new species of *Gyptis* (Hesionidae Annelida) under a jetty in Edithburgh South Australia. Zool Scr 38(1):89–99

Prathapan KD, Dharma Rajan P, Narendran TC, Viraktamath CA, Subramanian KA, Aravind NA, Poorani J (2006) Biological Diversity Act 2002: shadow of permit-raj over research. Curr Sci 91(8):1006–1007

Radashevsky VI, Selifonova ZP (2013) Records of *Polydora cornuta* and *Streblospio gynobranchiata* (Annelida Spionidae) from the Black Sea. Mediterr Mar Sci 14(2):261–269

Rouse GW, Fauchald K (1997) Cladistics and polychaetes. Zool Scr 26(2):139–204

Rousset V, Pleijel F, Rouse GW, Erséus C, Siddall ME (2007) A molecular phylogeny of annelids. Cladistics 23(1):41–63

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, Arumugam S, Jose HPM, Veeraiyan B (2015) AMBI indices and multivariate approach to assess the ecological health of Vellar-Coleroon estuarine system undergoing various human activities. Mar Pollut Bull 100(1):334–343

Sigamani S, Perumal M, Thivakaran GA, Thangavel B, Kandasamy K (2016) DNA barcoding of macrofauna act as a tool for assessing marine ecosystem. Mar Pollut Bull. https://doi.org/10.1016/j.marpolbul.2016.07.017

Soares-Gomes A, Mendes CLT, Tavares M, Santi L (2012) Taxonomic sufficiency of polychaete taxocenes for estuary monitoring. Ecol Indic 15(1):149–156

Struck TH, Schult N, Kusen T, Hickman E, Bleidorn C, McHugh D, Halanych KM (2007) Annelid phylogeny and the status of Sipuncula and Echiura. BMC Evol Biol 7(1):57

Sukumaran S, Vijapure T, Mulik J, Rokade MA, Gajbhiye SN (2013) Macrobenthos in anthropogenically influenced zones of a coralline marine protected area in the Gulf of Kachchh India. J Sea Res 76:39–49

Sukumaran S, Vijapure T, Kubal P, Mulik J, Rokade MA, Salvi S, Thomas J, Naidu VS (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

Sun Y, Kupriyanova EK, Qiu JW (2012) COI barcoding of Hydroides: a road from impossible to difficult. Invertebr Syst 26(6):539–547

Sundararajan S, Khadanga MK, Kumar JPPJ, Raghumaran S, Vijaya R, Jena BK (2017) Ecological risk assessment of trace metal accumulation in sediments of Veraval Harbor Gujarat Arabian Sea. Mar Pollut Bull 114(1):592–601

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. Trends Ecol Evol 18(2):70–74

Thomsen PF, Willerslev E (2015) Environmental DNA—an emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv 183:4–18

Venkataraman K, Wafar MVM (2005) Coastal and marine biodiversity of India. Indian J Mar Sci 34(1):57–75

Vogler AP, Monaghan MT (2007) Recent advances in DNA taxonomy. J Zool Syst Evol Res 45(1):1–10

Wilson RS (2000) Family Longosomatidae. In: Beesley PL, Ross GJB, Glasby CJ (eds) Polychaetes and allies: the southern synthesis. CSIRO Publishing, Melbourne, pp 193

Zaiko A, Samuiloviene A, Ardura A, Garcia-Vazquez E (2015) Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. Mar Pollut Bull 100(1):53–59

Zenetos A, Çinar ME, Pancucci-Papadopoulou MA, Harmelin JG, Furnari G, Andaloro F et al (2005) Annotated list of marine alien species in the Mediterranean with records of the worst invasive species. Mediterr Mar Sci 6(2):63–118

Zhou H, Zhang Z, Chen H, Sun R, Wang H, Guo L, Pan H (2010) Integrating a DNA barcoding project with an ecological survey: a case study on temperate intertidal polychaete communities in Qingdao China. Chin J Oceanol Limnol 28(4):899–910