DNA BARCODES TO STUDY DIVERSITY AND SPECIES COMPOSITION OF CLAMS FOUND IN AND AROUND MUMBAI COAST.

Amrinder Kaur Walia and Bhavita Chavan.
Department of Zoology, The Institute of Science, 15 Madame Cama Road, Mumbai, India.

There are diverse aquatic animals living in broad range of habitats and niches and studying their diversity by morphological dataset has been a point of debate for many years because of their varied forms. However, DNA barcoding technique which originated in the last decade has helped for expanding the horizons of biodiversity science. In the present study, diversity of marine clams found in coastal area of Mumbai, India was studied and it was found that mitochondrial cytochrome c oxidase subunit I (COI) gene based identification of marine clam species has lower success rate (60%) in existing circumstances, which is, enough to disclose differently located species. Consequently, 4 species (Gafrarium divaricatum, Geloina erosa, Meretrix casta, Tegillarca granosa) and 1 family (Tellinidae) of clams collected from 6 locations were studied using platforms of the studied molecular marker and morphological database. Results showed that BOLD systems v4 and NCBI databases could generate 40.74% and 85.18% identification success of 54 clams under investigation, respectively. Moreover, the analyzed specimens formed distinct cohesive species-specific clusters using Neighbor Joining (NJ) tree. The maximum intraspecific genetic variation was 2.54% (SE 0.00%) which supported reliability of species delimitation and the barcode gap analysis using Automated Barcode Gap Discovery (ABGD) and BOLD analytical tool revealed explicit species distinction. A mixed prevalence was observed of investigated species of Mumbai coast with suspected threat of endangerment and extinction. It may be concluded that DNA barcode based analysis of Indian marine clams may be effectively used for robust identification of species with varied macro habitats in the presence of enough COI reference sequence library, thus assisting marine biodiversity investigation and conservation.

Introduction: -
DNA barcoding is a molecular taxonomic technique that employs short genetic markers in an organism's DNA to identify the species. It also helps in establishing genetic distances between various ecotypes. Various molecular markers have been identified for various species type and have been employed successfully to generate data bulk for species identification. It is a micro genomic identification system developed by Paul Hebert, from University of...
Guelph, Canada in 2003, which discriminates life through the analysis of small segment of genome to diagnose the biological diversity. This system was brought up by an analogy of supermarket scanner that uses black stripes of the Universal Product Code (UPC) barcode to identify the item. The DNA barcode has similar concept, where four alternate nucleotides at each position creates one billion alternate codes, a 100-fold surplus required to discriminate the diversity of life on our planet. DNA Barcoding has helped taxonomists sort specimens by highlighting divergent taxa that may represent new species and thus, offers taxonomists the opportunity to greatly expand and eventually complete a global inventory of life’s diversity.

There are diverse terrestrial, and aquatic (fresh water and marine) ecosystems, having their particular roles. Some of them such as marine ecological systems are poorly studied due to impediments of unfavorable transportation facilities. Marine environment harbours very rich biodiversity, richest among them is Mollusca (Vanmali and Jadhav, 2015). There are about 5070 known species in India including 3,370 marine species (Rao, 1991). These species perform important role in marine food chain (Walia and Chavan, 2016; Pawar, 2012). An unprecedented biodiversity crisis caused by human activities such as degradation of habitat (Vijay, et al., 2005) with its overexploitation (Zingde and M. D, 2002), pollution (Zingde and M. D, 1999) etc. has emerged in the past decades affecting biodiversity of clam species. Clams are utilized commercially in India (Breitburg et al, 2009) by food suppliers on larger scale since these are crucial as a source of food (Pawar, 2012) owing to their remarkable demand. As a result, their number and existence in natural ecosystems are under threat with risk of extinction (Sundaram and Deshmukh, 2011) and consequently, these may be declared as endangered biological species (Lizano and Santos, 2014) which suggests urgent need of their investigation, especially in those locations where clams are traded as food to a greater extent such as Mumbai, Maharashtra State of India. Therefore, it becomes necessary to analyse species biodiversity and asses risk extinction of clam species from the habitat and take relevant measures for their conservation (Vanmali and Jadhav, 2015).

Study of marine molluscs has been undertaken in tropical regions (Vanmali and Jadhav, 2015) and in India, 17 species of clams including 4 species of giant clams have been reported by Venkataraman and Wafar (2005). Nevertheless, understanding of biodiversity of these animals based on morphological taxonomy is difficult and requires adequate reference data meeting morphological taxonomic standards along with intra-specific and inter-specific distances is not available. Besides, the identification requires expert taxonomists who are rarely available (Bouchet, 2006) particularly, in case of clams. Further, damage to important morphological characters (shape, sculpture and hinge teeth), availability of immature (with complicated) larval stages (Drent et al, 2004; Marko et al, 2009) and damaged specimens, higher cost, time consuming data collection, more existence of cryptic groups (Drent et al., 2004, Marko et al., 2009), capacity to change as per their environments (Drent et al, 2004; Marko et al, 2009) along with presence of only tissue sample and unavailability of sufficient taxonomic literature particularly in case of cryptic species. Such restrictions in understanding of clam biodiversity have made their investigation process more challenging for researchers resulting into its lower degree study. Consequently, any type of species loss of these animals may remain undetected further leading to decrease in their biodiversity in marine environments.

In order to confront such type of problem, although various molecular markers for marine species identification were developed (Avise. et al, 1975) such as cytochrome b oxidase (cyt-b) for tuna (Bartlett, 1991), 16S rRNA for holothurian species (Bergene et al, 1995), cytochrome c oxidase subunit I (COI) for three genera of planktonic copepods (Bucklin et al, 1999),), the ground breaking approach was brought by Hebert et al. (2003) who proposed use of mitochondrial COI gene for species identification of animals and coined the term ‘DNA barcoding’. This mitochondrial COI gene is used as standard DNA barcode for species identification as it shows higher interspecific distance than intraspecific divergence among and within species in addition to higher conserved amino acid sequence and possesses broad array of phylogenetic indications aiding in differentiation among less divergent species and phylogeographic clusters in single species (Brown et al, 1986; Folmer et al, 1994; Hebert et al, 2003; Cox and Hebert 2001; Wares and Cunningham 2001). It can be applied for revealing cryptic taxa (Mikkelsen et al, 2007; Johnson et al, 2008; Zou et al, 2012) and genetic exchange among submarine clam species (Clague et al, 2011).

Later, this approach was widely followed by multiple researchers in the globe who tested this gene for its applicability for identification of marine fauna, Layton et al. (2014) and Mikkelsen et al. (2007) identified molluscs from Canada and four genera of Norwegian bivalves, respectively. However, since probabilities of contamination by mucopolysaccharides (Layton KKS et al, 2014; Skolov 2000) in marine clams are higher, this group of animals become quite complex for DNA barcode process and as a result, very less work has been done in molecular
biodiversity of such species, particularly, in India. Moreover, difficulties in collection of these species exist due to marine habitats that are not easy to reach generating poor understanding of clam biodiversity around coastal region especially in India, although it has got ample coastal area of 7,516 km\(^2\) (Venkataraman and Wafar, 2005), no efficient initiative has been taken for identifying and analyzing clam biodiversity (Pawar, 2012) through DNA barcoding resulting into inadequate specimen and sequence data repository and lack of robust species level identification data along with genetic lineage studies leading to low scale scientific understanding of cryptic and endangered species.

Therefore, it is necessary to study biodiversity of clams in India with a broad range perspective. By taking this view into consideration, the present study scratched up to fulfill this taxonomic gap and provided COI gene database of clam species for future identifications.

Materials and Methods:-
Ethical Statement:-
In India, clams are used as food on a commercial scale; however, the field studies do not involve the collection of endangered or protected species.

Sample Collection and Preservation:-
Total of 125 clam specimens were collected in various seasons at the time of low tides from various sites of Mumbai coast namely Haji Ali- Worli, Bandra Bandstand, Khardanda, Belapur, Vashi and Uran (Fig 1). The collected clams were fixed immediately in 100 % alcohol, preserved at \(4^\circ\)C. Further, the samples were processed for DNA barcoding analysis at Paul Hebert Centre for DNA Barcoding and Biodiversity Studies, Aurangabad. The samples were labelled and their images were taken by Nikon Digital Camera D5100 and the database of collection locations, images, sequences and traces were uploaded on BOLD systems v4 (www.v4.boldsystems.org).

Fig 1:-Collection locations of clam specimens from coastal region of Mumbai, India.
DNA Extraction, PCR and Sequencing:-
Foot muscle tissues of 125 clams were processed for DNA extraction by Promega Kit Method (Wizard® Genomic DNA Purification Kit, Promega Corporation, Madison, WI). The genomic DNA samples were amplified for their mitochondrial Cytochrome c oxidase subunit I (COI) gene using primer pairs LoboF1 (5'KBTCACHAAAYCAYAARGAYATHGG3') and LoboR1 (5'TAAACYTCWGGRTGWCCAAARAAYCA3') (Lobo et al., 2013) or LCO1490_t1 (5'-TGTAAAACGACGGCCAGTGGTCAACAAATCATAAATATTGG3') and HCO2198_t1 (5'CAGGAAACAGCTATGACTAAACTTCAGGGTGACCAAAAATCA3') (Folmer, et. al., 1994) for forward and reverse directions, respectively using Applied BiosystemsVeriti 96 well Thermal Cycler. The used cycling conditions of PCR were as follows: initial denaturation at 94°C for 3 minutes followed by 35 cycles of 94°C at 30 seconds, annealing at 46°C for 40 seconds followed by extension at 72°C for 1 minute and final extension at 72°C for 10 minutes or initial denaturation at 95°C for 2 minutes followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at 48 °C for 30 seconds, extension at 72°C for 45 seconds and final extension at 72°C for 7 minutes. After that, PCR products were checked on 1.2% agarose gel and successful amplicons were cleaned up by Exonuclease I and alkaline rSAP (Shrimp Alkaline Phosphatase) using Applied Biosystems GeneAmp PCR System 9700 and then, bi-directional sequencing of targeted molecular marker were carried out using Applied Biosystems Genetic Analyzer 3730 sequencer and the obtained sequences were aligned using Codon Code Aligner (Codon Code Corporation, USA) and examined for the presence of stop codons using MEGA 5 (Koichiro Tamura et al., 2011) and then submitted to NCBI with accession numbers from KX608941 to KX608991 and from KX792502 to KX792504.

Results:-
Species Identification Success:-
The success rate of species identification varies among individuals (Fig 2) depending on availability of required reference database. It was found that Gaffarium divaricatum, Meretrix casta and Tegillara granosa were identified with highest success (100%) by COI gene based identification platform using BOLD systems and NCBI databases, whereas clam species Geloina erosa was identified by DNA barcode data with least success in contrast to morphological taxonomy. Conversely, the individuals of family Tellinidae were not reliably supported by available identification datasets (Fig 2).

Automated Barcode Gap Discovery (ABGD) Analysis:-
The obtained sequences were analysed using Automated Barcode Gap Discovery (ABGD) (Puillandre, 2011) for investigation of partition (with prior maximal distance, P=2.78e-03 and Kimura (K80) TS/TV distance model with minimum slope of 1.500000) among clam sequences belonging to four species and one family (Table 1).
Table 1: The partition of four species and one family of clams obtained by ABGD analysis.

<table>
<thead>
<tr>
<th>Group</th>
<th>Sample ID</th>
<th>Name of Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group [2] n: 4</td>
<td>AMB6, SPOT2.2, SPOT2.6, SPOT2.7</td>
<td><em>Tellinidae</em></td>
</tr>
<tr>
<td>Group [3] n: 4</td>
<td>SPOT1.1, SPOT1.2, SPOT1.4, SPOT1.5</td>
<td><em>Geloina erosa</em></td>
</tr>
<tr>
<td>Group [5] n: 1</td>
<td>SPOT4.8</td>
<td><em>Tegillarca granosa</em></td>
</tr>
</tbody>
</table>

Additionally, a graph was plotted of obtained sequences for studying histogram distances (Fig 3) with P minimum 0.001, P maximum 0.1 with 10 steps along with X (relative gap width) 1.5 as well as 20 as Nb bins (for distance distribution) in addition with Kimura (K80) TS/TV as 2.0 and in order to study partition of clams, with respect to partition based analysis using threshold value (Fig 4). It was observed that prior intraspecific divergence with respect to number of species showed distinct separation between recursive and initial partitions. This suggests that the obtained sequences of 4 species and 1 family were precisely partitioned.

Fig 3: Investigation of histogram distances of clam specimens.

Fig 4: Partition based analysis of clam specimens using Automated Barcode Gap Discovery (ABGD).
**Phylogenetic Analysis:**

The 54 obtained sequences of clam specimens were studied for evolutionary distances by constructing their phylogenetic tree using Neighbor Joining (NJ) as a statistical method with Kimura 2 parameter as a distance model along with bootstrap support of 1000 replicates using MEGA5 (Fig 5) and found that individuals of similar clam species were grouped together with high bootstrap support and to further verify any possibility of disparity in species identifications, we downloaded reference sequences of related species under study and used for phylogenetic tree constriction in combination with obtained sequences (Fig 6) using the same parameters.

![Phylogenetic analysis of obtained clam sequences with Neighbor Joining method.](image-url)
Fig 6: Investigation of species identification through phylogenetic tree construction using assemblage of obtained sequences and sequences taken from NCBI.

Genetic divergence:-
The clams were analyzed for their distribution of sequence divergence at each taxonomic level using Kimura 2 parameter as a distance model, BOLD Aligner (Amino Acid based HMM) for sequence alignment and pair wise deletion as ambiguous base/gap handling tool available with analytical tool used for distance summary analysis available on BOLD systems v4 and found that there was increase in maximum K2P divergence from within species (2.54%, SE=0) to within families (48.68% SE=0.01) (Table 2) and there was no intrageneric distance with S.E=0. However, 234 comparisons among families showed minimum distance as 39.71%, mean divergence as 43.67% and maximum distance as 48.68% (S.E 0.01). In addition, within species distribution to lower bias in sampling at species level was normalized and found that mean distance within species was as 0.86% (SE 0.19%) with min. distance between species as 0.00%.

<table>
<thead>
<tr>
<th>Taxa</th>
<th>Comparisons</th>
<th>Min Dist (%)</th>
<th>Mean Dist (%)</th>
<th>Max Dist (%)</th>
<th>SE Dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within Species</td>
<td>49</td>
<td>762</td>
<td>0.00</td>
<td>0.73</td>
<td>2.54</td>
</tr>
<tr>
<td>Within Genus</td>
<td>0.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Within Family</td>
<td>45</td>
<td>234</td>
<td>39.71</td>
<td>43.67</td>
<td>48.68</td>
</tr>
</tbody>
</table>

Table 2: K2P Genetic distance comparisons of clam specimens under study.

Discussion:--
Species identification:-
40.74% and 85.18% identification success using COI sequence databases of BOLD systems v4 (although J Chen et al. (2011) proposed that it is comfortable infrastructure for taxonomic database) and NCBI, respectively. However, in our analysis, we found that there was considerable variation in morphological and molecular identifications of 4 specimens belonging to family Tellinidae. Using morphology based data, the specimens belonged to species *Semele cordiformis* of family Semelidae, whereas these were identified as members of Tellinidae family by COI gene based analysis. In this sequence, we propose that there may be error in morphological taxonomy due to absence of cryptic features such as dimorphism as well as larval and adult dissimilarity in organisms (such as marine water shrimps).
Similarly, failure of COI gene based taxonomy for clam species identification may be due to the fact that the BLAST searching of GenBank database is not sufficient for analysis of bivalve specimens for their taxonomy and there is need to develop better search strategies especially for identification using DNA barcoding technology (Nina T Mikkelsen et al, 2007). As a result, 4 individuals collected from Uran area of Mumbai coast were identified up to only genus level i.e. Geloina by their COI gene sequences and their species were identified by morphology based data irrespective of the fact that we could not identify Tellinidae species and advocacy of (Candek and Kuntner, 2015) for DNA barcodes for their capacitance for species identification without morphological characters. Besides, 32 records of clam specimens in BOLD database was not found and submission of these records in genetic databases has expanded horizons of global DNA based taxonomical science.

**COI discrepancy:-**
The DNA barcoding analysis of clam specimens collected from 6 areas of Mumbai coast (figure 3) revealed 4 species and 1 family and their genetic distance investigation showed 0.73% mean intraspecific variation, which was more than that of echinoderms (0.62%), polychaetes (0.38%), marine fishes (0.39%) and decapods (0.46%) (Ward et al, 2005, 2008), A. bullisi (0.68%), A. excavata (0.68%), A. sphoni (0.52%) as well as A. mori (Hertlein, 1952) and less than 0.85% for A. oophaga(Hertlein, 1952) and 17.36% reported by Nuryanto et al. (2007) for giant clam species. The aspect of intraspecific divergence for species distinction was supported by the present finding that the maximum intraspecific distance of clam species was found to be 2.54%. Also, analysis of 455bp sequences of T. crocea, T. squamosa, T. maxima, T. gigas(Giant clam species) and several sequences obtained from Genbank resulted into formation of monophyletic groups (Nuryanto et al. 2007). Such cases propose the value of integrating molecular biodiversity studies with morphological taxonomic identifications for investigating species boundaries. In the study, it was observed that identification of individuals belonging to Geloina erosa were achieved up to only genus level due to either insufficient or ambiguous sequence data available in DNA databases. However, this complexity was resolved through morphological taxonomic identification of the species in question (figure 5). The ABGD analysis of obtained sequences showed that they are clustered into distinct five groups showing that these are varied from each other giving five different lineages.

**Habitat-specific composition and comparison of species:-**
It is proposed that the dominance and distribution of clam species found across Mumbai coast using COI gene as a molecular marker supported with morphological features is valid in contrast to the opinion reported by Schmidlin et al. (2010) who believed on only morphological taxonomic identification system for analysis of clam species supremacy. Moreover, although Clague et al., (2012) claimed that investigation COI genes of 98 members of limid clam species (Acesta sphoni and Acesta mori) failed to show geographical division along sampling area, it is found that DNA barcoding tool has successfully enlightened geographical partition of individuals of clam species, which may spread over a wide area owing to either anthropogenic activities or natural resources like waterfall (Schmidlin et al, 2010) based on their genetic divergence pattern. With this approach, it was observed that Gafarium divaricatum species mostly dominated in KharDanda and Bandra bandstand sites of Mumbai coast (figure 5), whereas 3 members of unidentified species belonging to family Tellinidae were reported in Haji Ali site, while its single specimen was found in sea coast of Bandra bandstand indicating species distribution in two different locations of Mumbai coast (Ramkrishna and Dey, 2010), which may be due to transport by water tides or other natural means. Besides, four individuals of Geloina erosa were mostly found in Uran coast line along with one specimen of Meretrix casta, similar findings were reported by Pawar et al., (2012) and 5 members of this species in addition with many individuals of Tegillarca granosa were found in the muddy sea shore area of Vashi during low tides displaying habitat adaptation of the species in question in the same ecosystem (Rao S and N. V, 1991). Similar results were reported by Vanmali and Jadhav (2015) who stated that clam species of family Veneridae were the most dominant in Dativare Coast of Vaitarna Estuary belonging to Palghar district of Maharashtra constituting 23.33% of total family diversity.

**Phylogeny Analysis:-**
Separate clusters of giant clam species (Tridacna and Hippopus) can be obtained by Neighbor-Joining method of phylogeny construction using COI gene showing similar topology of constructed trees (Lizano et al, 2014) suggesting patterns of lineage formation. This aspect is supported by the present study which showed that Nearest-Neighbor joining method of phylogenetic tree construction with 1000 bootstrap replications is robust and reliable in
clustering the similar specimens in specific clades (Figure 5). In the analysis, 54 clam specimens of 4 species and 1 family were precisely clustered into species and family-specific clades according to their similarity and dissimilarity patterns like various species of genus *Crassostrea* belonging to class bivalvia that formed distinctive groups when their 44 barcode sequences belonging to 9 species were phylogenetically analyzed (Trivedi et al, 2015).

**Our contribution:-**
DNA barcoding is considered as an effective technique in describing species diversity and geographical distribution across diverse animal taxa and the present attempt was the first trial to verify its efficacy in identification, habitat-specific composition and comparison of clam species collected to Mumbai coast of India. Moreover, 32 new records (*Gafrarium divaricatum-18, Geloina erosa- 4, Tellinidae-4, Meretrix casta-5, Tegillarca granosa-1*) in COI gene database of BOLD systems v4 were established and hence, this finding would be important for comparison of genetic variation pattern of marine clams including freshwater as well as estuary clams in the presence of enough reference sequence library to understand how this pattern varies among different habitats. This view was exemplified by Paulay and Meyer (2002) who proposed that many species that live in the marine ecosystem of the tropical Pacific were more widely distributed and showed less variation than terrestrial species. Additionally, this research work may assist to generate the database required to examine species richness which is due to lower extinction rates in tropical habitats (Schemske, 2012).

Overall, the global conservation of marine biodiversity with particular emphasis to marine clams is challenging because of an increase in anthropogenic activities and extinction rate as these animals are used as food on a commercial scale. Our work has started generation of DNA barcode database of Indian marine clams for providing their identification baseline and geographical information and has crucial contribution in literature of DNA barcoding since it focuses on the efficiency of mitochondrial 648 bp COI gene fragment for genetic divergence analysis on local scale. Furthermore, the record of variation among species with respect to their genetic makeup leads to comparative investigation with other marine phyla, genera and species and this type of research is important to inspect speciation patterns in salt water ecosystems. In addition, knowing cryptic species in the oceanic environments has applications in the efforts of conservation of species diversity. The present investigation expands horizons of biological taxonomic science regarding with Indian maritime clams assisting upcoming national and international collaborative conservation efforts.

**Future Perspective:-**
Upcoming barcode studies should include detailed taxonomic identifications to solve complexities of misidentifications, cases of synonymy and cryptic species; otherwise there may be creation of misleading view for species identification. Also, it may focus on removal of mucopolysaccharides of both fresh water and marine clams belonging to diverse taxa during DNA extraction and development of universal primer sets for their PCR amplification resulting into greater sequence recovery. Besides, strategic plans should be developed for generation of COI gene based DNA libraries of molluscs for effective identification of their subspecies and cryptic species as well as investigation of genetic diversity. Although, the area of survey was limited, future research work should aim to reveal cryptic clam species by sampling deep sea areas (Archambault et al. 2010). Genetic comparative analysis of marine clams and freshwater clams will further help to understand how their diversity and speciation patterns vary among these diverse ecosystems along with broad area in clam genetic biodiversity research with sampling stations across entire Indian coast line which may reveal numerous cryptic and endangered species assisting their conservation programs.

**Conclusion:-**
DNA barcoding tool can be used for species identification and phylogeny analysis of Indian marine clams only when enough reference sequence library is available with collaborative approach of morphological identification strategies. Moreover, this viewpoint of species identification is not only useful for recording biodiversity, but also for enlightening infrastructures of habitat-specific specificity and composition of species across extensive taxa in given geographical areas. Unfortunately, currently available COI gene database is insufficient for species identification of Indian marine clams indicating need of their more comprehensive and collaborative taxon sampling and analysis on global scale to enhance identification success for future unknown or cryptic species assisting biodiversity investigation and conservation. Besides, the circumstances of molecular and morphological species identification complexities arrived due to insufficient reference sequences and unavailability of enough taxonomical literature respectively, may be improved by wide range of taxon sampling, covering diverse marine and fresh water clam species across the globe by both molecular and morphological taxonomic methods.
Acknowledgement:
Authors are thankful to Dr. G. D. Khedkar, Director of Paul Hebert Centre of DNA Barcoding studies at Aurangabad, India. Dr. A. S. Khemnar, Director, The Institute of Science, Mumbai and Dr A. P. Manekar, Head, Department of Zoology, The Institute of Science, Mumbai.

References:
32. Schmidlin S., DénesSchmera, Sylvain Ursenbacher and Bruno Baur (2010). Separate introductions but lack of genetic variability in the invasive clam Corbiculasp. in Swiss lakes Proceedings of the 17th International Conference on Aquatic Invasive Species, San Diego, USA.