

## Chapter

# Molecular Markers and Their Optimization: Addressing the Problems of Nonhomology Using Decapod COI Gene

*Deepak Jose and Mahadevan Harikrishnan*

## Abstract

Advancements in DNA sequencing and computational technologies influenced almost all areas of biological sciences. DNA barcoding technology employed for generating nucleotide sequences (DNA barcodes) from standard gene region(s) is capable of resolving the complexities caused due to morphological characters. Thus, they complement taxonomy, population analysis, and phylogenetic and evolutionary studies. DNA barcodes are also utilized for species identification from eggs, larvae, and commercial products. Sequence similarity search using Basic Local Alignment Search Tool (BLAST) is the most reliable and widely used strategy for characterizing newly generated sequences. Similarity searches identify “homologous” gene sequence(s) for query sequence(s) by statistical calculations and provide identity scores. However, DNA barcoding relies on diverse DNA regions which differ considerably among taxa. Even, region-specific variations within barcode sequences from a single gene leading to “nonhomology” have been reported. This causes complications in specimen identification, population analysis, phylogeny, evolution, and allied studies. Hence, the selection of appropriate barcode region(s) homologous to organism of interest is inevitable. Such complications could be avoided using standardized barcode regions sequenced using optimized primers. This chapter discusses about the potential problems encountered due to the unknown/unintentional/intentional use of nonhomologous barcode regions and the need for primer optimization.

**Keywords:** DNA barcodes, BLAST, homologous gene sequences, nonhomology, standardized barcode regions, primer optimization

## 1. Introduction

Deoxyribonucleic acid (DNA) is considered as the prime genetic material of the living world as it stores complete set of information for dictating the structure of every gene product. The order of nucleotide bases (viz. adenine, guanine, cytosine, and thymine) contains these instructions for genetic inheritance along DNA [1]. “DNA sequencing” refers to a technique for understanding the language of DNA by determining the order of nucleotide bases present within the genome of organism(s) of interest [2, 3]. During the 1970s, researchers utilized

two-dimensional chromatography for obtaining the first DNA sequence in laboratories. Later, dye-based sequencing methods with automated analysis were developed for easier and faster DNA sequencing. With the continued improvement in sequencing approaches, DNA sequence data derived from genes and genomes of organisms have become indispensable in basic research and allied fields.

Advancements in DNA sequencing and computational technologies influenced almost all areas of biological sciences. Taxonomy and systematics, the science for identifying organisms up to “species” level followed by classifying them based on their relationships, are also well complimented by DNA sequence database. Traditionally, “species,” the basic unit of taxonomy, is distinguished on the basis of certain unified external characters within a sufficient number of specimens termed as “morphological characters” [4]. Later, morphological-type specimens were complimented with molecular data from molecular markers (allozymes, nuclear DNA, mitochondrial DNA) specifically in morphologically problematic groups [5, 6]. As molecular markers, gene type sequences (referred to as DNA barcodes) are developed using a technology called “DNA barcoding.” Thus, fundamental information from conventional taxonomy is complimented with genetic information from molecular taxonomy for scientific inferences. More than a decade, this technique has been subjected for prime consideration in molecular research due of its capability to distinguish closely related species. It is also applicable to a broad spectrum of taxa for extensive biodiversity assessment studies. DNA barcoding remains as a standard method for specimen identification and allied studies, and DNA barcodes serve as an inevitable tool in understanding genetic relationships of organisms [7–11].

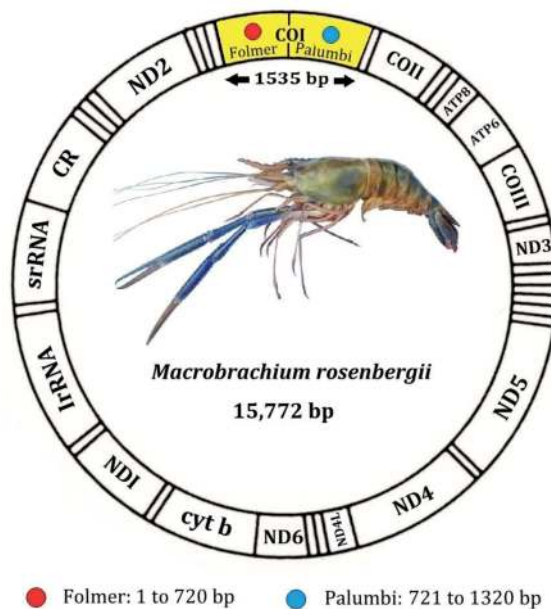
An ideal DNA barcode should possess certain qualities like higher universality and resolution. Since DNA barcoding relies on different DNA regions that vary between organisms (like bacteria, plants, animals, birds, etc.) [12–14], selection of barcode region is dependent on the selected sample type. DNA barcode sequences are normally compared with a DNA reference library of morphologically pre-identified vouchers to assess the rate of similarities/dissimilarities, followed by assignment of taxonomic names to unknown specimens according to the percentage of identity [15, 16]. Since homology relations are proportional to the origin and relations of taxa, focusing on molecular characters to examine homology relations is more direct than on morphology due to the discrete and “simple” nature of the characters in the latter. Thus, comparative sequence analyses are apparent to analyze the biological relationships of DNA sequences. Two major disciplines that work at both interspecific and intraspecific level are molecular phylogenetics and population genetics. Molecular phylogenetics deals typically with evolutionary relationships of different species, while population genetics is applied to characterize variations within and among populations of a single species [17, 18]. In short, DNA barcodes from standard gene region(s) compliment taxonomy, population analysis, and phylogenetic and evolutionary studies at genetic level. They are also utilized for identification of species, particularly for eggs, larvae, and commercial products [19, 20].

Among DNA barcodes, mitochondrial genes gained preference due to their higher stability, mutation rate, copy number per cell, and absence of introns that provide higher genetic information [21]. Among mitochondrial genes, cytochrome c oxidase I (COI) is considered as the primary barcode sequence for animal kingdom [10, 15]. Mitochondrial DNA (mtDNA) has been used for carrying out phylogenetic studies in a large number of animals including crustaceans, in a short span of time. Hitherto, numerous reports in support of broad benefits of DNA barcoding are available [15, 16, 22–25]. Even though DNA barcoding has completed a decade as one of the versatile techniques in addressing numerous concerns in the field of life science, authors like [26–28] have also pointed out many drawbacks with respect to this technique. A recent study [7] reported issues regarding the

usage of nonhomologous barcode sequences for molecular studies. This chapter discusses on the nonhomologous barcode regions of COI gene region, available in public database (like NCBI) and issues arising due to their unknown/intentional/unintentional use in molecular analyses. Molecular results inferred from mitochondrial COI gene sequences (amplified using “Folmer” and “Palumbi” primers) of *Macrobrachium rosenbergii* are used to demonstrate the combined effect of “nonhomologous” sequences over specimen identification, population analysis, and molecular phylogeny.

## 2. An overview of mitochondrial cytochrome oxidase subunit 1 (mtCOI) gene of *Macrobrachium rosenbergii*

DNA was first detected in mitochondria in the year 1963. It was found in association with proteins and lipids, localized to the mitochondrial matrix [29]. Almost all eukaryotic cells possess mitochondrial genome that contains genetic information utilized in systematic and population genetics for the past two decades [30]. Complete mitochondrial DNA (mtDNA) sequence having approximately 17,000 base pairs (bp) has been developed in many species, including humans [31]. Maternal inheritance, relatively rapid mutation rate, and lack of intermolecular recombination are considered as major characteristic features for their extensive use in population structure and phylogenetic studies at different taxonomic levels [21]. Hitherto, more than 1100 complete mitochondrial genome sequences or similar derivatives have been published [32]. However, crustaceans, one of the most morphologically diverse animal life forms, are represented only by limited number of complete mitochondrial sequences. Within crustaceans, decapods represent an extremely diverse group with many commercially important taxa including prawns, shrimps, lobsters, and crabs [7, 30]. Two major COI barcode regions are amplified for them using two sets of primers, namely Folmer (aka 5' COI; LCO-HCO) [33] and Palumbi (aka 3' COI; Jerry-Pat) [34], which are nonhomologous with limited



**Figure 1.** Gene map of *Macrobrachium rosenbergii* mitochondrial genome. Yellow color indicates cytochrome c oxidase subunit 1 (COI) gene having 1535 bp length with two nonhomologous regions (viz., Folmer and Palumbi).

overlaps [7, 35]. These two regions are widely used in decapod molecular taxonomy and associated research. In public database (e.g., NCBI), several decapod species possess COI sequences derived from both these regions. Among them, the giant freshwater prawn *Macrobrachium rosenbergii* (Crustacea: Decapoda: Palaemonidae) is having sufficient mtDNA data including its whole genome (**Figure 1**) and other marker gene sequences [7].

### 3. Impact of nonhomologous barcode regions in molecular taxonomy and allied studies

#### 3.1 Specimen identification

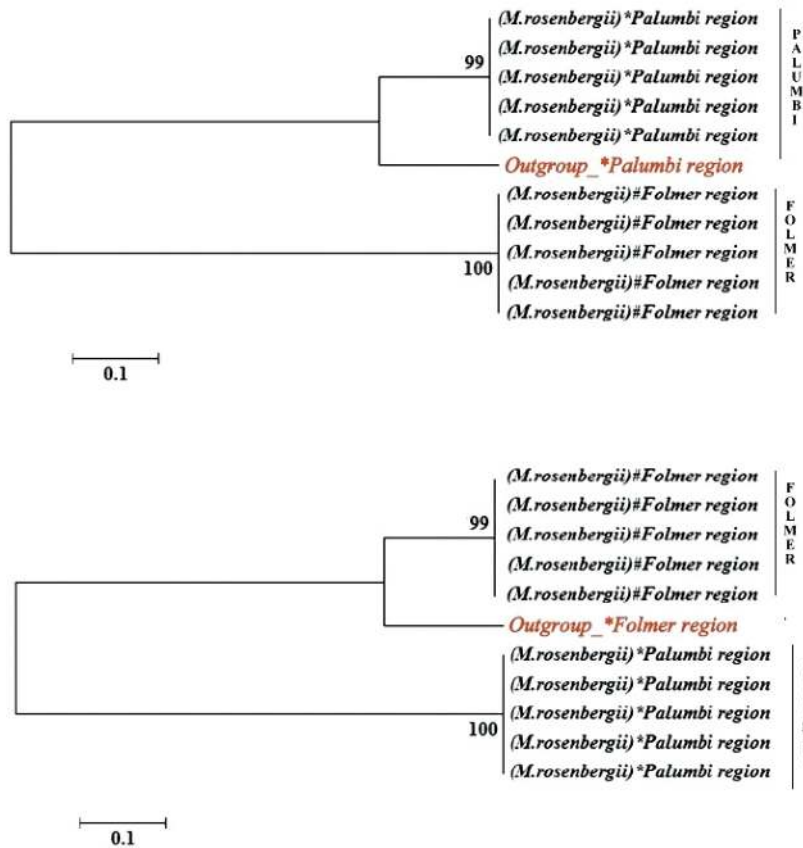
DNA-based taxon identification for recognition of known species and discovery of new species is reported in many studies [7, 15, 23]. Mitochondrial cytochrome oxidase I (COI) gene is recommended as an efficient DNA barcode for identifying all kinds of animals [15, 16, 23], including cryptic species [18, 24]. Pairwise comparison of COI sequences of congeneric species generates a divergence rate of >2% [23], reaching up to 3.6% in species complexes, and exceeds 5% in rare cases [15, 16, 23, 24]. The region of the 5' end of COI ("Folmer" portion) is considered as the "DNA barcode" sequence which might be no better than that of the 3' end of COI sequences, i.e., Palumbi sequence [7, 35, 36]. Even though these two regions are considered as related fragments, even within crustaceans [36, 37], the region-specific conservation for "Folmer" and "Palumbi" sequences creates nonhomology. This creates diversity within the same gene region, causing misinterpretations if it is used unknowingly.

Here, the results inferred from nonhomologous COI gene regions of *M. rosenbergii* are given for demonstrating the issues related to specimen identification. **Figure 2** depicts a phylogenetic tree constructed based on neighborhood joining (NJ) analysis from sequences of "Folmer" and "Palumbi" regions of *M. rosenbergii*. Tree topology could be expected to array these sequences as barcode regions of a single species (*M. rosenbergii*) within a major clade with sufficient bootstrap value corresponding to their monophyly and the selected out-group as another entity.

Results inferred from the NJ tree exhibited reciprocal monophyly in its array, differentiating "Folmer" and "Palumbi" regions as two different entities. Out-group species that was expected to have higher divergence than the rest showed affinity toward the "Palumbi" sequences of *M. rosenbergii* in the first tree. In the second case, a relationship was established between the "Folmer" sequences of *M. rosenbergii* and the out-group. This indicated a gene specific relationship between the barcode regions of the test and out-group organisms based on their homology. These results focus over the conservative nature of barcode region(s) of COI gene and its dominance over the species-level conservation within individual (genus or species).

Inferences from phylogenetic tree will also be reflected in genetic distance data since the substitution accounted for calculating intraspecific divergence within *M. rosenbergii* is higher than at interspecific level. Substitutions will be more among the nonhomologous sequences since they represent different regions within the same gene, accounting for higher distance. Out-group with homologous gene sequence could provide considerable genetic distance with the homologous sequences of species of interest (here it is *M. rosenbergii*). Further, the genetic distance provided by the nonhomologous sequences of species of interest will be more than that of the genetic distance provided by the homologous sequences of species





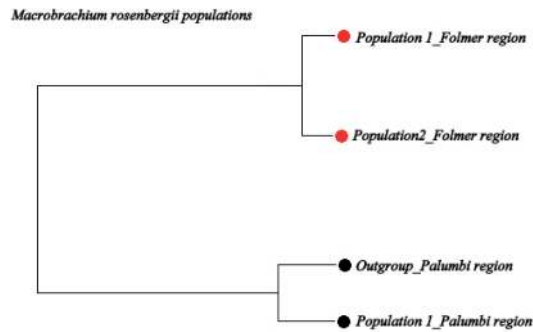
**Figure 2.**  
 NJ tree showing different clades for with respect to the “nonhomologous” regions present in COI region of *M. rosenbergii*.

of interest and out-group. It could be concluded that the existence of region-specific conservation within the COI barcode gene of decapod crustaceans could dominate species-level conservation causing serious errors in molecular results. Hence, the use of precise mitochondrial gene fragment(s) with respect to the homology of available nucleotide sequences is recommended for specimen identification and species confirmation for avoiding potential errors and erroneous results [7].

### 3.2 Population analysis

COI gene sequences are well considered for population analysis of many species including decapods. Here, the impact of nonhomologous regions in population studies is discussed using “Folmer” and “Palumbi” sequences of the genus *Macrobrachium*. Two populations were selected: both “Folmer” and “Palumbi” sequences were selected for Population 1, while for Population 2, only “Folmer” regions were considered. “Palumbi” sequence of an out-group organism was also considered.

The tree topology was expected to reveal only two highly diverged populations of *M. rosenbergii*, viz., Populations 1 and 2. However, the exhibited pattern showed three populations, differentiating Population 1 into two populations with regard to the nonhomologous barcode regions. The Folmer regions of Populations 1 and 2 were arrayed according to their population diversity, while the “Palumbi” region of Population 1 arrayed along with the “Palumbi” region of the out-group, indicating the presence of a third population, which is virtual (**Figure 3**) and was due to region-specific conservation (for “Folmer” and “Palumbi”) in COI gene.



**Figure 3.**  
NJ tree generated for *M. rosenbergii* populations using nonhomologous barcode sequences and out-group.

These findings were confirmed using AMOVA analysis using sequences of “Folmer” and “Palumbi” as two different populations which produced significant differences in support of the existence of two populations. This clarified that the nonhomology of barcode regions can lead to serious erroneous inferences.

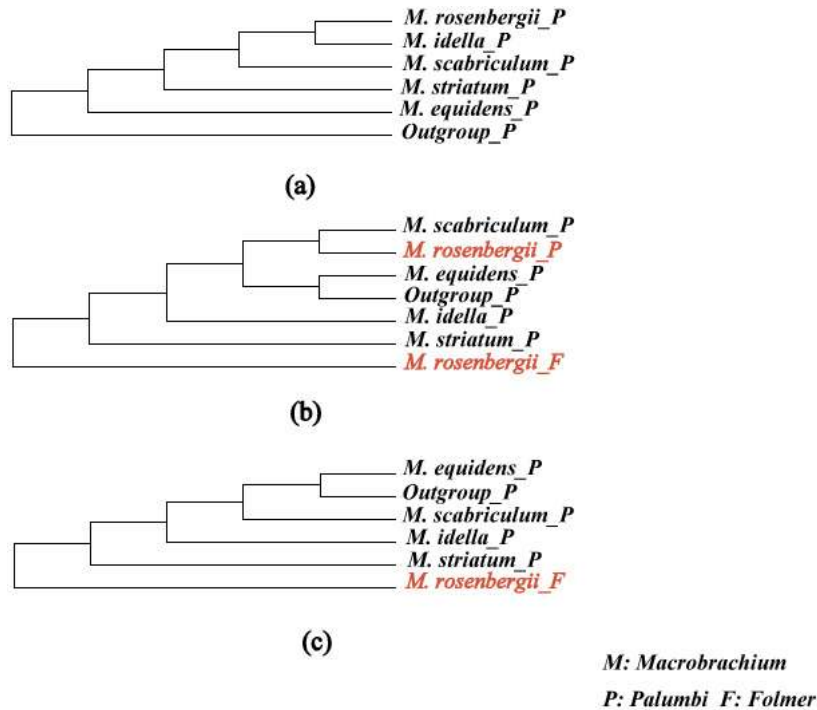
### 3.3 Molecular phylogeny

The influence of nonhomology in phylogenetic studies was examined using “Folmer” and “Palumbi” sequences of *M. rosenbergii* and other selected congeneric species. Three types of sequence selections were done: (i) incorporation of “Palumbi” region of *M. rosenbergii* along with the “Palumbi” regions of selected congeneric species and out-group (**Figure 4a**), (ii) incorporation of both “Folmer” and “Palumbi” sequences of *M. rosenbergii* along with the “Palumbi” sequences of all other individuals (**Figure 4b**), and (iii) incorporation of “Folmer” sequences of *M. rosenbergii* along with “Palumbi” sequences of other species (excluding “Palumbi” region of *M. rosenbergii*) (**Figure 4c**).

Tree topology exhibited cladistic array of selected organisms in accordance with the previous findings of specimen identification and population analysis, i.e., with respect to the region-specific conservation persisting within “nonhomologous” barcode regions of COI gene. Monophyly of *Macrobrachium* species was exhibited by the first NJ tree (**Figure 4a**) in which only homologous sequences of “Palumbi” region were used. The rest of the phylogenetic trees exhibited absence of monophyly and erroneous cladistic array due to the impact of nonhomologous barcode regions, i.e., “Folmer” and “Palumbi” regions (**Figure 4b** and **c**). These incongruences within the phylogenetic trees will be well reflected in pairwise distance data because of the impact of nonhomologous sequences. Due to the higher rate of substitution among “Folmer” and “Palumbi” regions (as they belong to different regions of same gene), the genetic distance was higher among them even though considerable distance was accounted among the congeners. These findings demonstrated problems in molecular phylogeny by incorporating nonhomologous barcode regions of COI.

## 4. Discussion

After the first discovery of mitochondrial DNA in 1963, more than 5300 complete mtDNA sequences of different taxa were submitted in NCBI till date [29, 31, 32]. These sequences are well utilized for addressing different fields of molecular taxonomy [38]. Among the preferred gene regions of mitochondrial



**Figure 4.** NJ tree based on Kimura two-parameter model (1000 bootstraps) generated using (a) “Palumbi” regions of *M. rosenbergii* and other selected congeners and out-group, (b) “Palumbi” and “Folmer” regions of *M. rosenbergii* and “Palumbi” region of other selected congeners and out-group, and (c) “Folmer” regions of *M. rosenbergii* and “Palumbi” region of other selected congeners and out-group.

DNA, cytochrome c oxidase subunit 1 (COI) remains as one of the most recommended molecular markers because of its ability to generate sequence data within a reasonable time in a cost-effective way. These data could be well utilized for sorting collections into identified species, biodiversity assessments, delineation of cryptic species, detection of population structure, gene flow pattern identification, phylogeographic studies, molecular phylogeny, evolution, etc. [7, 17, 39–41]. Altogether, this protein-coding gene mitochondrial gene has acquired great acceptance in large-scale projects of diverse taxa [42–44].

Usually, an ideal COI barcode region is reported to possess about 648–700 nucleotides that are used for similarity searches in nucleotide database for identification of known/unknown samples. Barcode of Life Data System (BOLD) refers to a freely available database which acquires analyses and releases DNA barcode data. Researchers interested in DNA barcoding and allied studies can submit sequence(s) to the public database (NCBI/DDBJ/EMBL) or the consortium for the Barcoding Life website. Similarity search using nucleotide BLAST (BLASTN) or BOLD search ([www.barcodinglife.org](http://www.barcodinglife.org)) is usually used for identifying the status of DNA sequence of interest. This will lead to the corresponding homologous nucleotide sequence(s) for your DNA sequence that has been sequenced previously or will give homologous sequences of its close relative(s). But there exists some hiding factors that could confuse a researcher to identify a species from the available database, i.e., nonhomologous sequences with region-specific conservation (e.g., COI gene) which may alter results to a great extent. Hence, these results are to be scrutinized carefully since there may be nonhomologous sequences for a taxa of interest that will not appear in the BLAST search because of their nonhomology. Altogether, similarity searches with the available database results in top species matches, where the name of the species having reference sequence accessioned in the database or the name

of the closest related taxa in the absence of reference sequence for that particular species, are enlisted [45].

This chapter has clearly discussed about the presence of “Folmer” and “Palumbi” regions of *M. rosenbergii* in its COI gene (1535 base pairs) within the 15,772 base paired complete mitochondrial DNA (NCBI accession no.’s AY659990 and NC\_006880) [30]. Within, the COI gene, first approximate 720 base pairs are amplified by “Folmer” primers and the rest (approximately 721–1535) by “Palumbi” primers. Being a recommended barcode region, “Folmer” region is recognized as a universal barcode fragment, and at the same time, the other COI fragment sequenced using “Palumbi” primers dated from the early 1990s is also well-known and utilized for DNA barcoding [34, 35]. Hence, these two fragments are sequenced for studying molecular aspects of crustaceans [36, 38].

However, the presence of these two barcode regions within a single target gene (COI) with “Folmer” region as the first and “Palumbi” as second barcode region for a broad class of organisms (particularly crustaceans) could result in severe problems with respect to molecular studies. It could be helpful if full COI sequences of taxa are included as a scaffold for containing the two fragments prior to molecular analysis. But many organisms are still lacking the whole genome mitochondrial sequence data, and instead they are having either “Folmer” or “Palumbi” sequence(s). In such a scenario, if a “Palumbi” region for a specimen is sequenced and the public database is having only the “Folmer” region for the same organism, BLAST search will fail to identify that particular sequence. It will indicate only the closest organism on the basis of sequence homology. For “Folmer” region also the case will be the same if database is having “Palumbi” sequences. This mainly affects specimen identification as it could be hard for a researcher to identify the sample, particularly for those specimens lacking major morphological characters. Even in specimens with little morphological variations from its type descriptions, there could be failures in identifying the existing species causing misinterpretation of the same as a novel species. Regarding the impact of nonhomology in population analysis, the chance of misinterpretation of “nonhomologous” fragments as a different population exists.

Molecular phylogeny could also be affected from severe errors due to dual barcode regions. Even if both barcode sequences are contained within the nucleotide database, BLAST search will enlist sequences according to the homology of our sequence(s) only. In such cases, the chance for errors could be minimized even though the full dataset is not explored. However, there could be possibilities of missing dataset of taxa supplied to the database, due to nonhomology. Another case is that, even among congeneric species, monophyly could not be established due to the impact of “nonhomologs” (refer to **Figure 4b** and **c**). Tree topology could be altered due to the impact of dual barcode regions. As a result, relationship between morphologically similar species and species groups could be altered.

## 5. Conclusions and recommendations

GenBank accounts for an enormous amount of molecular data within which more than 90% of mtDNAs belongs to metazoans and the remaining sequences represent fungi and terrestrial plants. About 3% of available mitochondrial genomes represents protists. Despite the usually discussed issues like misidentifications and pseudogenes, lack of primer pair(s) data, particularly for certain unpublished dataset, remains as a major drawback. Moreover, designing and using of multiple primer pairs for various objective of molecular taxonomy have generated multiple DNA fragment from the same gene. Hence, under a single species name, there could be multiple DNA sequences from a single gene, which are “nonhomologous” in nature.



It is very basic that selection of nucleotide sequences could be done on the basis of their homology but still, in the present scenario, it is hard to identify homologs by BLAST search. Most trace samples, used in forensic studies, remain undetected due to the lack of standardization in barcode regions. Even though “Folmer” region is considered as a universal barcode region, there are numerous reports regarding the use of “Palumbi” sequences as better barcodes based on species specificity. This issue could be well resolved with the use of complete genome sequences which, however, are developed only for limited taxa. A better way to resolve this problem of “nonhomology” is to provide data regarding primer pair(s) used along with the nucleotide sequence data. Even if someone is concerned about the privacy of research, they could opt the embargo period provided for releasing the nucleotide sequence data to public database. It is also true that researchers develop diverse primers for amplifying specific genes. One more recommendation is to update the nucleotide submission data after publication of the corresponding manuscript so that the entire research community could get the proper information regarding the background of the nucleotide without interfering one’s privacy. DNA barcoding has crossed the boundaries of academics and has made use of in food authentication, medical applications, forensic science, etc. Since DNA-based analysis has become an important part, region-specific issues related to gene sequences need to be addressed and resolved. This chapter has addressed the present nature of barcode regions derived from COI and the issues related to them so that the need for primer optimization [21] could be practiced at the earliest. Regardless of a single species (*M. rosenbergii*), identifying the nature of barcode regions on additional taxa in a broad spectrum could help to make the existing nucleotide database user-friendly, even to those who are beginning their research. “Error cascades” that occurred due to bad taxonomy in science made the research communities relying on advanced technologies like DNA barcoding for accurate species identification and taxonomic assignment. So, it could be beneficial if we are able to resolve or clarify these types of confusions so that DNA barcoding could be free from “error cascades” of molecular taxonomy. It is also recommended to have an integrative taxonomy in the case of morphologically recognizable organisms as suggested by Will et al. (2005) so that error-free results regarding a species could be drawn out only using both morphological and molecular approaches.

## **Acknowledgements**

Authors greatly acknowledge the Kerala State Council for Science, Technology and Environment (KSCSTE) for providing financial support. We gratefully acknowledge the director, School of Industrial Fisheries, Cochin University of Science and Technology, for providing necessary facilities and support for conducting this research.

## **Conflict of interest**

The authors declare responsibility for the entire contents of this chapter and have no conflicts of interest.

## **Author details**

Deepak Jose\* and Mahadevan Harikrishnan  
School of Industrial Fisheries, Cochin University of Science and Technology,  
Kerala, India

\*Address all correspondence to: [deepak140887@gmail.com](mailto:deepak140887@gmail.com)

## **IntechOpen**

---

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Travers A, Muskhelishvili G. DNA structure and function. The FEBS Journal. 2015;282(12):2279-2295. DOI: 10.1111/febs.13307
- [2] Kumar BR. DNA representation. In: Anjana M, editor. DNA Sequencing Methods and Applications. 2nd ed. ExLi4EvA; 2016. p. 3
- [3] Stranneheim H, Lundeberg J. Stepping stones in DNA sequencing. Biotechnology Journal. 2012;7(9): 1063-1073
- [4] Deshmukh VD. Principles of Crustacean Taxonomy. Manual on taxonomy and identification of commercially important crustaceans of India. 2013. pp. 28-39. Available from: <http://eprints.cmfri.org.in/9645/>
- [5] Lee MS. The molecularisation of taxonomy. Invertebrate Systematics. 2004;18(1):1-6
- [6] Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP. A plea for DNA taxonomy. Trends in Ecology & Evolution. 2003;18(2):70-74
- [7] Deepak J, Harikrishnan M. Non-homologous COI barcode regions: A serious concern in decapod molecular taxonomy. Mitochondrial DNA Part A. 2016;28(4):482-492
- [8] Deepak J, Nidhin B, Anil Kumar KP, Pradeep PJ, Harikrishnan M. A molecular approach towards the taxonomy of fresh water prawns *Macrobrachium striatum* and *M. equidens* (Decapoda, Palaemonidae) using mitochondrial markers. Mitochondrial DNA Part A. 2016;27(4):2585-2593
- [9] Deepak J, Rozario JV, Benjamin D, Harikrishnan M. Morphological and molecular description for *Glyphocrangon investigatoris* Wood-Mason & Alcock, 1891 emphasizing its phylogenetic relationship. Mitochondrial DNA Part A. 2016;27(3):2053-2057
- [10] Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessièrè J, et al. An in silico approach for the evaluation of DNA barcodes. BMC Genomics. 2010;11(1):434
- [11] Zhang J. Species identification of marine fishes in China with DNA barcoding. Evidence-based Complementary and Alternative Medicine. 2011;10:978253
- [12] Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, Bankvander M, et al. A DNA barcode for land plants. Proceedings of the National Academy of Sciences USA. 2009;106:12794-12797
- [13] Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology. 1994;3:294-299
- [14] Vences M, Thomas M, Bonett RM, Vieites DR. Deciphering amphibian diversity through DNA barcoding: Chances and challenges. Philosophical Transactions of the Royal Society, B: Biological Sciences. 2005;360(1462):1859-1868
- [15] Hebert PD, Cywinska A, Ball SL, Dewaard JR. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences. 2003a;270(1512):313-321
- [16] Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of

- the National Academy of Sciences. 2004a;**101**(41):14812-14817
- [17] Deepak J, Harikrishnan M. Evolutionary history of genus *Macrobrachium* inferred from mitochondrial markers: A molecular clock approach. Mitochondrial DNA Part A. 2018;**30**(1):92-100
- [18] Hajibabaei M, Singer GA, Hebert PD, Hickey DA. DNA barcoding: How it complements taxonomy, molecular phylogenetics and population genetics. Trends in Genetics. 2007;**23**(4):167-172
- [19] Deepak J, Harikrishnan M, Jenson VR, Pradeep PJ, Saswata M, Sameera S. Targeted species substitution in giant freshwater prawn trade revealed by genotyping (submitted)
- [20] Tedeschi R. DNA sequencing and crop protection. In: DNA Sequencing- Methods and Applications. IntechOpen; 2012. p. p63
- [21] Schubart CD. Mitochondrial DNA and decapod phylogenies: The importance of pseudogenes and primer optimization. Decapod Crustacean Phylogenetics. 2009;**47**:65
- [22] Hebert PD, Gregory TR. The promise of DNA barcoding for taxonomy. Systematic Biology. 2005;**54**(5):852-859
- [23] Hebert PD, Ratnasingham S, de Waard JR. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London. Series B: Biological Sciences. 2003b;**270**(suppl\_1):S96-S99
- [24] Hebert PD, Stoeckle MY, Zemplak TS, Francis CM. Identification of birds through DNA barcodes. PLoS Biology. 2004b;**2**(10):e312
- [25] Rougerie R, Haxaire J, Kitching IJ, Hebert PD. DNA barcodes and morphology reveal a hybrid hawkmoth in Tahiti (*Lepidoptera: Sphingidae*). Invertebrate Systematics. 2012;**26**(6):445-450
- [26] Collins RA, Cruickshank RH. The seven deadly sins of DNA barcoding. Molecular Ecology Resources. 2013;**13**(6):969-975
- [27] Hickerson MJ, Meyer CP, Moritz C. DNA barcoding will often fail to discover new animal species over broad parameter space. Systematic Biology. 2006;**55**(5):729-739
- [28] Will KW, Mishler BD, Wheeler QD. The perils of DNA barcoding and the need for integrative taxonomy. Systematic Biology. 2005;**54**(5):844-851
- [29] Nass MM, Nass S. Intramitochondrial fibers with DNA characteristics: I. fixation and electron staining reactions. The Journal of Cell Biology. 1963;**19**(3):593-611
- [30] Miller AD, Murphy NP, BurrIDGE CP, Austin CM. Complete mitochondrial DNA sequences of the decapod crustaceans *Pseudocarcinus gigas* (*Menippidae*) and *Macrobrachium rosenbergii* (*Palaemonidae*). Marine Biotechnology. 2005;**7**(4):339-349
- [31] Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;**290**:457-465
- [32] Smith DR. The past, present and future of mitochondrial genomics: Have we sequenced enough mtDNAs? Briefings in Functional Genomics. 2015;**15**(1):47-54
- [33] Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology. 1994;**3**:294-299

- [34] Palumbi SR, Benzie J. Large mitochondrial DNA differences between morphologically similar penaeid shrimp. *Molecular Marine Biology and Biotechnology*. 1991;1(1):27-34
- [35] Page TJ, Steinke D. No homology means there can be no analyses; a comment on Jose & Harikrishnan. *Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis*. 2018;29(2):220-221
- [36] Roe AD, Sperling FA. Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution*. 2007;44(1):325-345
- [37] Lefébure T, Douady CJ, Gouy M, Gibert J. Relationship between morphological taxonomy and molecular divergence within *Crustacea*: Proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution*. 2006;40(2):435-447
- [38] Wilson K, Cahill V, Ballment E, Benzie J. The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: Are malacostracan crustaceans more closely related to insects than to branchiopods? *Molecular Biology and Evolution*. 2000;17(6):863-874
- [39] Batista AC, Negri M, Pileggi LG, Castilho AL, Costa RC, Mantelatto FL. Inferring population connectivity across the range of distribution of the stiletto shrimp *Artemesia longinaris* Spence Bate, 1888 (*Decapoda*, *Penaeidae*) from DNA barcoding: Implications for fishery management. *ZooKeys*. 2014;457:271
- [40] Buhay JE. "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. *Journal of Crustacean Biology*. 2009;29(1):96-110
- [41] Liu MY, Cai YX, Tzeng CS. Molecular systematics of the freshwater prawn genus *Macrobrachium* Bate, 1868 (*Crustacea: Decapoda: Palaemonidae*) inferred from mtDNA sequences, with emphasis on east Asian species. *Zoological Studies*. 2007;46(3):272-289
- [42] Costa FO, DeWaard JR, Boutillier J, Ratnasingham S, Dooh RT, Hajibabaei M, et al. Biological identifications through DNA barcodes: The case of the *Crustacea*. *Canadian Journal of Fisheries and Aquatic Sciences*. 2007;64(2):272-295
- [43] Hogg ID, Hebert PD. Biological identification of springtails (*Hexapoda: Collembola*) from the Canadian Arctic, using mitochondrial DNA barcodes. *Canadian Journal of Zoology*. 2004;82(5):749-754
- [44] Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD. DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society, B: Biological Sciences*. 2005;360(1462):1847-1857
- [45] Ratnasingham S, Hebert PD. BOLD: The barcode of life data system (<http://www.Barcodinglife.Org>). *Molecular Ecology Notes*. 2007;7(3):355-364