World Applied Sciences Journal 15 (3): 304-310, 2011
ISSN 1818-4952
© IDOSI Publications, 2011

DNA Barcoding: A New-Fangled Global Standard of Taxonomical Frontier - A Review

Eby John, Linson Thomas, Arockiasamy Edward, Gabriel Melchias and Jeyapaul Antony Prabhu

Department of Biotechnology, St Joseph's College, Tiruchirappalli - 620002, TN, India

Abstract: The massive task of identifying species all over the earth has always been a major challenge in relations to biological monitoring and cataloguing. DNA barcoding is an avant-garde molecular technique, which promises to provide us with a universal key for the taxonomical identification and classification of species via miniature DNA sequences. DNA barcoding utilizes mitochondrial Cytochrome C oxidase subunit 1 gene sequences for characterising animals and plastid based Internal Transcribed Spacer (ITS) gene sequences for plants. It is thus an ambitious project, which presents a standardised locus of short sequences. It confers ample scope of variation involving different species yet allowing reasonably a little amount of variation within related species, all of which are showcased globally. This review defines some contemporary progress in the screening process, which produces a new level of phylogenetic and taxonomical opportunities assisted by classical taxonomy as well as the latest bioinformatics tools.

Key words: CBOL · ITS · Mitochondria · N-J Phenograms · Microcoding · Bio-Barcode

INTRODUCTION

The advent and utilisation of polymerase chain reaction (PCR) followed by the declining cost of DNA sequencing has catapulted molecular based genetics next to ubiquity all through the biological sciences. With the emergence of bioinformatics software tools, the mounting taxonomic coverage of sequences enclosed in online DNA sequence libraries (GenBank) and the simplicity with which these sequences can be searched and retrieved have become relatively simple [1]. DNA sequence data could be used to support organism identifications [2]. This review shows the process that was formalized as 'DNA barcoding', with the goal of extensive screening with one or few reference genes so as to assign or compare unfamiliar individuals to species and for the deliverance of the discovery of new species.

DNA barcoding is a technique for species identification, carried on by DNA sequences from a short segment of the whole genome, with the intention of advocating a broad range of environmental and conservation data in which conventional taxonomic classification is not relevant or applicable. The concept of DNA barcoding is so profound that it has created a storm of queries and controversies that fill the pages of many leading science journals [3, 4]. In other words, advocates of barcoding propose to use a fragment of DNA to describe and discriminate between all life forms on earth [5]. Hence it is a promising approach for the diagnosis of biological diversity, which differs from molecular phylogeny. The prime objective is to identify a novel sample based on the existing classification. Sequence differences in rRNA were used to discover species of archaea which consequently resulted into the redrawing of evolutionary based tree. Molecular markers (allozymes & rDNA) have been successfully employed in molecular systematics. Paul Hebert from the University of Guelph-Canada advocated the collection of a public library consisting of DNA barcodes thereby forming a database that would be linked to named specimens. The library would hence present a new global key for identifying species and will progress with the ever increasing taxon coverage with faster, cheaper and easier sequencing.

Classical barcoding makes use of the most fundamental phylogenetic processes accessible like simple pairwise distances calculated by phonetic clustering to create a tree-like illustration of species clusters (Neighbour-Joining phenograms). With the

Corresponding Author: Gabriel Melchias, Department of Biotechnology, St Joseph's College, Tiruchirappalli - 620002, TN, India. Tel: +91-9486752737.

application of molecular biomarkers, non-specialists could allocate the test specimens to known species - including the specimens that can amaze specialists (e.g., insect larvae, partial adults & eggs) [6]. Next-generation DNA sequencing systems will enable the rapid production of barcodes, thereby ultimately facilitating the assignment of unidentified individuals to identified species. DNA barcoding universally have spread out actively to novel experimentations as in forensic science, the biotechnology based food industries and animal diet. Ecologists, environmental scientists and health inspectors, public health groups and many other possible users with the requirement to categorize specimens, are hence focusing on barcoding as a new alternate to applied problems. Taxon classification using analytic single nucleotide polymorphisms (SNPs) and biodiversity evaluation from ecological samples (e.g., air, water and soil) can also be measured as DNA barcoding *sensu lato*. Thus the importance of sequence based studies presents a new dimension in the taxonomical frontier of organisms. The whole process from isolating the precise sequence to classification of the organism and finally the synchronization of the database library is the basis of this practice.

'Barcoding': an Encyclopaedia of Life: A famous science fiction drama known as Star Trek displayed a miniature handheld 'tricorder' device, which was featured as to scan and recognize the different alien life forms. Although it was just a play, it encouraged a technologically advanced species identification strategy. The very first international conference "Barcoding Life" was conducted during February 2005 at the National History Museum (London), attended by some 200 participants from nearly 50 countries. This international venture indulging in the foremost organizations has initiated a motivated project, to promote the 'Barcode of Life'- a procedure for the quick and inexpensive classification of the valued 11 million species across earth. It took more than two centuries for taxonomists to describe about 1.7 million species, but this figure presents only an approximate estimate of the original biodiversity on earth [3, 7]. The first conference held at the DNA Taxonomy Workshop at Munich (April 2002), was initiated and funded by the German Science Association with the participation of over 100 scientists mainly from European countries [8]. The primary focus of the conference was to elect out the best precise markers for DNA based taxonomy.

A team of scholars guided by Paul Hebert at University of Guelph implemented the use of a segment of the mitochondrial gene as a global 'identification' tag for all animal species. The use of these 'species barcodes' primarily requires the assemblage of an all-inclusive library that connects barcodes with organisms. Understanding the prospects of this technique, Alfred Sloan-Foundation initiated and funded two meetings at Cold Spring Harbour, during March - September 2003. The meetings arrived at the plan suggesting that all major natural history museums should play the lead in linking diagnostic DNA sequences to specimen in collections and to the present taxonomic system that is the so-called Linnean system. Consortium for the Barcode of Life (CBOL) was formed and coupled by major herbaria, natural history museums, plus private and other research organizations. The National Center for Biotechnology Information sealed a partnership with CBOL, thus barcoded DNA sequences and related information can be accessible via GenBank. Another facet of DNA barcoding is the assistance of fundamental biodiversity inventories. Hence, from the site of evolutionary relationships to constructing the tree of life [9], this method provides a better tool in the field of molecular phylogenetics. A locus for DNA barcoding was standardized, observable in maximum taxa of interest and which is sequencable devoid of species-specific primers for PCR. This segment should be short enough to be effortlessly sequenced with existing technology and also showcase a wide range of variation between different species yet a fairly small amount of variation within similar species. For animals and other eukaryotes, the mitochondrial *cox1* gene was employed and for land plants, the concatenation of the *rbcL* and *matK* chloroplast genes could be employed [10].

Mitochondrial DNA for ''Barcoding'': Almost all the eukaryotic cell contains mitochondria and the mitochondrial DNA (mtDNA) displays a relatively fast mutation rate, which marks in considerable variation in mtDNA sequences between varying species. This principle is used as a comparatively small variance within species. About 600-bp sequences isolated from the mitochondrial DNA have been selected as DNA based barcodes and have become one of the supreme-contentious and vibrant issues in the appliance of genetic data to biodiversity assessment and species identification universally. It is advocated that limited mtDNA sequence data can be used as a simple and economical means to scan and identify almost all forms of

life [2]. A 648-bp region of cytochrome c oxidase subunit 1 (CO1) gene present inside the mitochondria was proposed as a potential 'barcode'. GenBank announced a particular submission tool for DNA barcoding, labelling the sequences with the designation of BARCODE. The arrangement is particularly designed to acquire CO1 sequences, signalling the barcoding promise to this single gene. However, mtDNA is not the only source of data that can be used to defining species. Many limitations can arise while using the CO1 gene as barcode due to the following factors such as effective population size, maternal inheritance, recombination, mutation rate, heteroplasmy and compounding genetic factors. Hence the application of barcodes is vulnerable to technical challenges that are particularly related to mtDNA [11, 12].

DNA Barcodes for Identifying Flowering Plants: The proposal of DNA barcoding for identification of species and taxonomy has been seen with many controversies. Still a mounting scientific community has accepted DNA barcoding as a sensible technique for biodiversity studies, such as to aid the inventories of very dissimilar and taxonomically poorly understood regions [13]. The usage of mitochondrial *coxI* gene in DNA barcoding is now highly regarded for animals, but the pursuit for a common short DNA barcode in plants is still undecided [14]. The *coxI* gene is not suitable for most plants species due to slower rate of its evolution in advanced plants than in animals. The deviation of *coxI* coding sites amid families of flowering plants has been recorded only to be a few base pairs across 1.4 kb of sequence [15]. Additionally, plants rapidly change their mitochondrial genome structure; hence an alternate unique identifier has to be applied at the species level [13].

In the case of molecular systematics investigations of plants at the species point, ITS sites of the nuclear based ribosomal cistron (*i.e.* 18S, 5.8S and 26S) is the most frequently sequenced locus. The ITS has condensed species based inconsistency in few groups such as the diverged taxa on islands, deviating properties that need cloning of several copies and problems facing the secondary structure producing inferior sequence data [16]. The ITS region can be amplified in two minor proportions (ITS1 and ITS2) adjacent to the 5.8S locus, which confirms usefulness for degraded samples. The fairly preserved 5.8S region contains sufficient phylogenetics signals for differentiation at the order or phyla levels [17]. The benefit of preserved regions like 5.8S can produce a pool of bordering neighbours for advanced comparisons which will be vital for efficient data searches, mainly while comparing an unknown sequence in the sequence library [18].

For evolutionary based investigations, the plastid gene sequence has been more willingly explored than the nuclear genome and can be presented for plant barcoding. Similar to the mitochondrial genome that shows for animals the genome is uniparentally inherited, structurally firm and non-recombining. The plastid locus most generally used by plant systematists for evolutionary relations is *rbcL*, the *trnL-F* intergenic spacer, *matK* spacer and *atpB* [19]. The locus *rbcL* has been recommended as the prime contender for plant barcoding, although it has normally been used to establish phylogenetic relationships at the gene-level. Besides ITS, single-copy nuclear genes and their introns is gaining eminence in species level molecular systematics courses, for instance waxy leaf. In a search of GenBank, it was observed that the *trnH-psbA* spacer could be lucratively amplified in higher plants, mosses, ferns and liverworts, even though the degree of divergence between-species is not known. Hence further studies regarding the plastid region of the non- flowering plants must be deduced [20].

The Workflow with DNA Barcoding: Modern biology has branched and developed novel tools and applications engaged to assess biological relationships with DNA sequences - molecular phylogenetics and population genetics. Studies in molecular phylogenetics typically deal with evolutionary relationships among deeper clades, whereas those in population genetics target variation within and among populations of a single species. Species ID through barcoding is more often done by using a small DNA fragment - 'the barcode' - from a particular part of the specific genome region from the specimen under research. The barcode sequence from each unidentified specimen is then compared with a library or database of suggested barcode sequences copied from those of recognized identity. Various gene regions have been engaged for species featuring bio-systematics. Yet DNA barcoding promotes the approval of a 'universal standard', that is a 650 - base fragment of the 5-end of the mitochondrial gene known as cytochrome c oxidase 1 (CO1), which has acquired a label as the barcode site for animals. CO1 have also been advanced to be effective for the identification of specimens even with ruined DNA [21].

DNA Barcoding and Taxonomy: The job of identifying specimens relating to species level is strenuous, yet is an important aid for taxonomic workflow. Barcoding is no substitution for comprehensive taxonomic analysis. For example, when an unidentified specimen does not present a precise match to existing files in the barcode database, the barcode sequence hence does not meet the criteria of designating the unknown specimen as a new species [22]. DNA barcoding has also been used in well-studied groups such as Lepidoptera. For example, barcoding is now used routinely to recognize the biodiversity of the caterpillar fauna in north-western Costa Rica [23]. This project has over the past three years, employed barcoding to create a reference sequence library for more than 25000 specimens representing >2000 species of moths and butterflies and their parasitoids. Similarly DNA Barcoding approaches have been employed for precise identification and classification of Anamorphic or Deuteromycetes family of fungi [24]. The conventional taxonomic workflow, which usually requires the compilation of morphological and ecological data, can vary for different taxonomic assemblages (i.e. taxonomic identification of birds and fish require different methods and skills), whereas barcode analysis can be applied in a more or less standardized way across large domains of life (*i.e*. all animal taxa).

DNA Barcoding and Molecular Phylogenetics: A usual molecular phylogenetics project involves a primary decision in relation to the target group for analysis (e.g. family), the assembly of representative taxa, the acquisition of sequence information and the construction of phylogenetic trees by using optimality criteria such as Maximum Likelihood, Maximum Parsimony, or Bayesian analysis. With the advance of whole-genome sequencing projects, some researchers have even advocated the use of entire genomic sequences for phylogenetic inference [25]. While barcode libraries have similarities to molecular phylogenetic data (both are sequence information from assemblages of species), DNA barcodes do not usually have sufficient phylogenetic signal to resolve evolutionary relationships, especially at deeper levels [26]. DNA barcoding projects can support the construction of phylogenies by aiding the selection of taxa. Barcode of Life projects create a perfect taxonomic sampling environment for conducting phylogenetic studies on diverse branches of the Tree of Life.

DNA Barcoding and Population Genetics: The traditional analytical tool has approaches such as patterns of allozyme or restriction enzyme polymorphisms. They now are replaced by sequence-based analyses. However, the selection of an appropriate marker system for a population genetics survey requires careful thought of issues such as sensitivity of the questions being asked and practical measures for obtaining the information (i.e. ease of amplification by PCR). Molecular phylogenetic analysis and genetic diversity distance amid two species of marine polychaetes relationship was investigated based on 16S rRNA sequences [27]. Large-scale analyses of sequence-based markers are now underway in the framework of projects such as HapMap which aims to spot the genetic variation associated with human diseases [28].

Microcoding (Oligonucleotide Barcoding): In recent years, a mounting amount of techniques have been devised that can be predicated on the application of typical short nucleic acid sequences that is, small DNA barcodes in quick identification events. Microarrays based on silicon and nylon-membrane plus the original Luminex method of DNA-tagged polystyrene beads are used and sorted by flow cytometry. These systems contribute to the feature that fully sequenced genes are considered to find small, thermodynamically even and non-hairpin forming sites of high sequence distinctiveness that could be utilized as fixed single-stranded oligonucleotides on a classification platform [29]. The bound oligonucleotides then anneal explicitly to corresponding, complementary regions of DNA in a test solution constituting labelled single stranded amplicons from few or many anonymous strains. Fungal systems were among the first to be studied using silicon-platform microarrays. The use of microcoding techniques provides for conversion of many DNA ruined taxonomic specimens such as directly interrelated species complexes where species-level classification can only be completed by means of molecular techniques. These techniques allow forensic DNA analysis of materials that enclose DNA that has been injured but not utterly destroyed, for instance slightly burned, dried or heavily crushed material. In the end, the preparation of microcode with a variety of selections from an entire genome sequence databases may show potential for extensive genomic analysis from materials bearing only fragmented DNA.

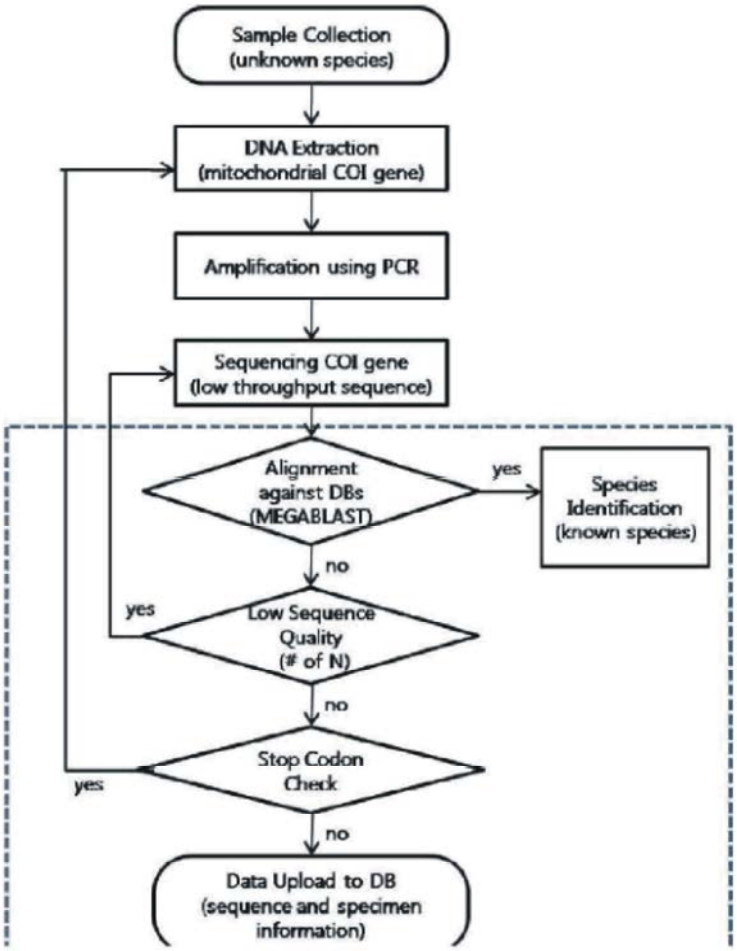

Fig. 1: DNA Barcoding for species identification and promotion of related database for future references (Courtesy: Jeongheui and Sang [15])

Bio-Barcode: Bio-barcode is used in bioinformatics as a database platform rather than an explicit DNA barcode server. The principle of bio-barcode is to be utilised by biologists and researchers, who have specific species data and who wish to put up a DNA-barcode database and server. It can support the collection, storage, investigation and eventually aids in the publication of high quality DNA-barcode paperwork. Bio-barcode has made use of the guidelines under CBOL and GenBank of the NCBI, which has to be furnished for the records to put on official barcode status. In addition, it can also be used for encouraging international association for constructing an Asia featured biodiversity system focusing to be the Asian biodiversity database or information server [30]. Bio-barcode provides the users with an option for the straight submission of their information, which carries the details related to specimens, sequences, other files and images. The data submitted, can be updated by an edit function which is directly available from the sequence and specimen pages.

CONCLUSION

DNA barcoding can contribute to refining species discovery and cataloguing. Once a set of barcodes is established for a group of organisms, unidentified specimens (i.e., query specimens) can be examined. Since the scientific basis of large-scale fabrication of molecular DNA barcodes are becoming more advanced and the value of the resultant database is ever more evident, DNA barcoding has now developed into a more comprehensive tool including assistance at the point for taxonomy, molecular phylogenetics and population genetics. In the coming frontiers, a ‘Life-Barcoder’ could be devised to

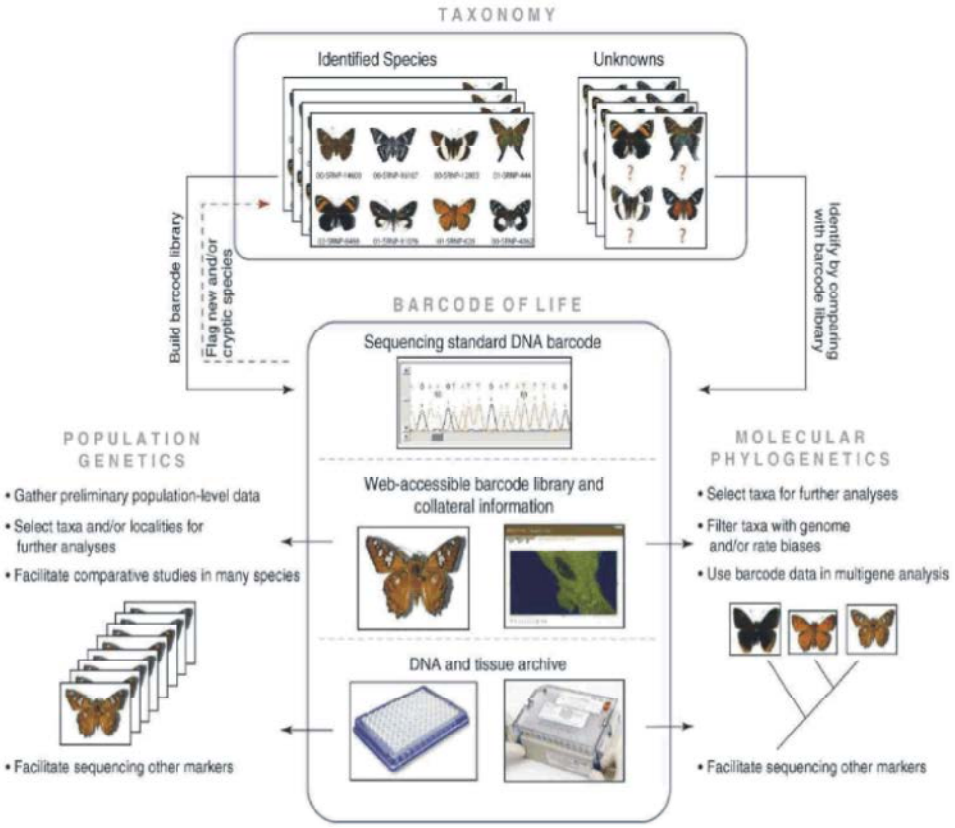

Fig. 2: The major components of the Barcode of Life projects and their contribution to taxonomy, reconstruction of molecular phylogenies and population genetics investigations. This diagram shows how DNA barcoding libraries can support the conventional taxonomic workflow by high-throughput identification of unknown specimens and by helping to draw attention to new and cryptic species. (BOLD: http://www.barcodinglife.org). This information can be useful in other contexts, such as phylogenetics (Tree of Life projects) and population-level studies. In addition, archival DNA and tissue specimens collected in barcoding projects provide an excellent resource for other investigations. Butterfly images are taken from the database of Daniel Janzen and Winnie Hallwachs. [Courtesy: http://janzen.sas.upenn.edu]

spot species, which can also be connected via the Internet to all other forms of biological data such as biodiversity, images or maintenance status. This could showcase direct access to all life form information to any individual in the society and thus enhancing the magnitude of preserving biodiversity. With all these efforts there still exist many complex problems with respect to biological conservation. However new initiatives must be invited and encouraged to secure prosperity for barcoding all the living forms on earth on an urgent basis.

ACKNOWLEDGMENTS

The authors are thankful to the management of St Joseph's College, Tiruchirappalli for their support.

REFERENCES

1. Bucklin, A., M. Guarnieri, R.S. Hill, A.M. Bentley and S. Kaartvedt, 1999. Taxonomic and systematic assessment of planktonic copepods using mitochondrial COI sequence variation and competitive, species-specific PCR, Hydrobiologia, 401: 239-254.
2. Hebert, P.D.N., E.H. Penton, J.M. Burns, D.H. Janzen and W. Hallwachs, 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*, Proc Natl Acad Sci. USA, 101: 14812-14817.
3. Blaxter, M., 2003. Counting angels with DNA. Nature, 421: 122-124.

4. Tsang, W.Y. and B.D. Lemire, 2002. Stable heteroplasmy but differential inheritance of a large mitochondrial DNA deletion in nematodes. Biochem Cell Biol., 80: 645-654.
5. Frey, J.E. and B. Frey, 2004. Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of *Thrips tabaci* (Thysanoptera: Thripidae): mitochondrial heteroplasmy or nuclear integration? Hereditas, 140: 92-98.
6. Will, K.W. and D. Rubinoff, 2004. The perils of DNA barcoding. Cladistics, 20: 47-55.
7. Zhang, D.X. and G.M. Hewitt, 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochem Syst Ecol., 25: 99-120.
8. Tautz, D., P. Arctander, A. Minelli, R.H. Thomas and A.P. Vogler, 2003. A plea for DNA taxonomy, Trends Ecol. Evol., 18: 70-74.
9. Cracraft, J. and M.J. Donoghue, (eds) 2004. Assembling the tree of life. New York: Oxford University Press.
10. CBOL Plant Working Group, 2009. A DNA barcode for land plants. Proc Natl Acad Sci. USA, 106: 12794-12797.
11. Holmes, R., 2004. Problems of using mtDNA in Barcoding. New Sci., 2453: 32-35.
12. Daniel, R., S. Cameron and K. Will, 2006. A Genomic perspective on the shortcomings of Mitochondrial DNA for ‘‘Barcoding’’ Identification, J. Heredity, 97(6): 581-594.
13. Kress, W.J., J. Kenneth, A. Wurdack Elizabeth, A. Zimmer Lee, H. Weigt and Daniel Janzen, 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. USA, 102: 8369-8374.
14. Wilson, E.O., 2003. The encyclopaedia of life: Trends Ecol. Evol., 18: 77-80.
15. Cho, Y., Y.L. Qiu, P. Kuhlman and J.D. Palmer, 1998. Proc. Natl. Acad. Sci. USA, 95: 14244-14249.
16. Alvarez, I. and J.F. Wendel, 2003. Taxonomical markers, Mol. Phylogenet. Evol., 29: 417-434.
17. Cullings, K.W. and D.R. Vogler, 1998. Use of DNA barcodes to identify flowering plants. Mol. Ecol., 7: 919-923.
18. Zuckerman, S.H., F.P. Gillespie, J.F. Solus, R. Rybczynski and J.M. Eisenstadt, 1986. Mitochondrial protein synthesis in interspecific somatic cell hybrids, Somat Cell Mol. Genet., 12: 449-458.
19. Gressel, J. and G. Ehrlich, 2002. Taxonomical applications of ITS in Plants, Trends Plant Sci., 7: 642-644.
20 Shaw, J., E.B. Lickey, J.T. Beck, S.B. Farmer, W. Liu, J. Miller, K.C. Siripun, C.T. Winder, E.E. Schilling and R.L. Small, 2005. The tortoise and the hare II: relative utility of 21 non coding chloroplast DNA sequences for phylogenetic analysis, Am. J. Bot., 92: 142-166.
21. Hajibabaei, M., D.H. Janzen and J.M. Burns, 2006. DNA barcodes distinguish species of tropical Lepidoptera, Proc. Natl. Acad. Sci. U.S.A., 103: 968-971.
22. Karl, S.A. and B.W. Bowen, 1999. Evolutionary significant units versus geopolitical taxonomy: molecular systematics of an endangered sea turtle (Chelonia). Conserv Biol., 13: 990-999.
23. Janzen, D.H., M. Hajibabaei, J.M. Burns, W. Hallwachs, E. Remigio and P.D.N. Hebert, 2005. Wedding biodiversity inventory of a large and complex *Lepidoptera* fauna with DNA barcoding, Phil. Trans. R. Soc. B., 360: 1835-1845.
24. Praveen Gehlot, H.A. Idress and Bahauddin Salleh, 2010. Anamorphic Fungi: An Overview, Middle-East J. Scientific Res., 6: 201-208.
25. Nardi, F., A. Carapelli, P.P. Fanciulli, R. Dallai and F. Frati, 2001. The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: evidence for heteroplasmy and tRNA translocation, Mol. Biol. Evol., 18: 1293-1304.
26. Hebert, P.D.N., A. Cywinska, S.L. Ball and J.R. DeWaard, 2003. Biological identifications through DNA barcode, Proc. R Soc. Lond Ser. B., 270: 313-321.
27. Satheeshkumar, P. and L. Jagadeesan, 2010. Phylogenetic position and genetic diversity of Neridae - Polychaeta based on molecular data from 16S rRNA sequences, Middle-East J. Scientific Res., 6: 550-555.
28. Janzen, D.H., 2004. Barcoding biodiversity and human diseases. Philos. Trans. R. Soc. London B., 359: 731-732.
29. Summerbell, R.C., C.A. Le Vesque, K.A. Seifert, M. Bovers, J.W. Fell, M.R. Diaz, T. Boekhout, G.S. DeHoog, J. Stalpers and P.W. Crous, 2005. Microcoding: the second step in DNA barcoding, Phil. Trans. R. Soc. B., 360: 1897-1903.
30. Jeongheui, L. and Y.K. Sang, 2009. BioBarcode: a general DNA barcoding database and server platform for Asian biodiversity resources. BMC Genomics 10(Suppl 3): S8.