Asian Journal of Biological and Life Sciences | Review Article

DNA Barcoding : Current advances and future prospects - a review

Prerna Vohra, K.S Khera*

* Professor, Department of Zoology, PAU, Ludhiana -141004, Punjab, India

E-mail : kskhera@pau.edu , vohra.prerna123@gmail.com Contact No : +91 9815909757

Submitted : 19.06.2013 Accepted :05.08.2013 Published : 31.12.2013

Abstract

DNA barcoding is a molecular-based identification system, recently introduced in the scientific community. This method is not new to science, but the real innovation is not in the discrimination system itself. DNA barcoding can be considered as the core of an integrated taxonomic system. It is established that the mitochondrial gene cytochrome c oxidase I (COI) can serve as the core of a global bioidentification system for animals. Now days, this approach has become increasingly popular and advances as well as limitations have clearly emerged as increasing amounts of organisms have been studied.

INTRODUCTION

Species identification is a fundamental part of recognizing and describing biodiversity. Traditionally, identification has been based on morphological diagnoses provided by taxonomic studies. At that time only taxonomists and trained technicians can identify taxa accurately, because it requires special skills acquired through experience [1]. Moreover, this approach to the task of routine species identification has four significant limitations. First, both phenotypic plasticity and genetic variability in the characters employed for species recognition can lead to incorrect identifications. Second, this approach overlooks morphologically cryptic taxa which are common in many groups [2]. Third, since morphological keys are often effective only for a particular life stage or gender, many individuals cannot be identified. Finally, although modern versions represent a major advance, the use of keys often demands such a high level of expertise that misdiagnoses are common [3]. Thus, researchers have been testing the idea that species could be identified easily and rapidly using only a short DNA sequence, which represents a standardized position in the genome and is called a DNA barcode [4-5]. The DNA barcode is analogous to the black stripes of the Universal Product Code, which are used to distinguish commercial products. The idea of a standardized molecular identification system has emerged with the development of PCR-based approaches for species identification [6]. Advances in DNA-sequencing technologies enabled researchers studying biodiversity to conduct simple, cost effective and rapid DNA analyses. This progress in biotechnology, and the taxonomy played a large role in the creation of DNA barcoding.

Overview of DNA Barcoding

DNA barcoding definition and Objectives

DNA barcoding is based on the premise that a short standardized sequence can distinguish individuals of a species because genetic variation between species exceeds that within species [7]

For animals, a 648-bp fragment of the mitochondrial gene cytochrome c oxidase subunit I (COI) has been chosen as the standard barcoding marker due to its high interspecific variation, low intraspecific variation, and relatively universal primers for taxonomic groups at the level of orders and even classes [8]. Hebert *et al* proposed a technique using a primer set to amplify a 648-base pair (bp) region of the mitochondrial cytochrome-c oxidase subunit 1 (COI) gene to ensure rapid and accurate identification of a broad range of biological specimens. They named this technique "DNA barcoding" [9-10]. Then, the Barcode of Life project was proposed to promote DNA barcoding as a global standard for sequence-based identification of eukaryotes. The International Barcode of Life consortium is an international initiative devoted to develop DNA barcoding as a global standard for the identification of biological species. Species identification of known specimens and discovery of overlooked species for enhancing taxonomy are the primary goals of barcoding [11]. With barcoding technique, a species can be identified from a tiny amount of tissue, from seeds, or from sterile, juvenile or fragmentary materials when morphological identification is difficult or even impossible [4].

Importance of DNA barcode

The cost and time-effectiveness of DNA barcoding enables automated species identification during large sampling campaigns [5]. In this way, DNA barcoding could also improve large surveys aiming at unknown species detection and identification of pathogenic species with medical, ecological and agronomical significance [12-13]. However, DNA barcoding has several advantages over previous methods. One advantage is its availability. The standard DNA barcode region COI is very efficient for species identification. This region has good discrimination power for most animal groups [14-15]. A 648-bp fragment has enough information and can be directly sequenced with a sequencer. These useful features are the reason why the *COI* region was selected as the standard DNA barcode. Thus, DNA barcoding can be a simple but powerful method for non-experts, especially those who routinely identify a large number of samples. Other advantage of DNA barcoding is the rapid acquisition of molecular data. As a contrast, morphological data gathering can be time consuming, in some cases totally confusing and in others, almost impossible [16-17]. Furthermore, in three important situations, relevant species identification must necessarily be molecular-based. First to determine taxonomic identity of damaged organisms or fragments of (e.g. goods, food and stomach extracts). Barcoding is thus potentially useful in the

food industry, diet analyses, forensic sciences and in preventing illegal trade and poaching of endangered species (e.g. fisheries, trees and bushmeat). Second, when there are no obvious means to match adults with immature specimens (e.g. fish larvae) [18-19]. Third is when morphological traits do not clearly discriminate species (red algal species [20-21] and field collected mosquito specimens [22])

What does accuracy of DNA Barcoding depend on? The barcoding "gap"

One of the critical issue in DNA barcoding is its accuracy. Accuracy mainly depends on the extent of, and separation between, intraspecific variations and interspecific divergence in the selected marker. The more overlap there is between genetic variation within species and divergence separating sister species, the less effective barcoding becomes [23]. (Fig 1.)

Role of barcoding in Biological sciences:

DNA barcoding is also of great interest to specialist besides as an identification tool for non specialists. It brings together taxonomy, molecular phylogenetics and population genetics [24] 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. DNA barcoding occupies a middle ground as it seeks comprehensive coverage for species, but focuses on their delineation rather than their relationships. Unlike other well-known sequence libraries (e.g. NCBI), BOLD is an interactive interface where deposited sequences can be revised and taxonomically reassigned. The compiling of sequences, from one or few common loci improves synergic studies at large geographic scales and across numerous genera. Such information on the global distribution of species, their genetic diversity and structure will enhance the speed and effectiveness of local population studies.

Current advances in barcoding

In the past 20 years DNA sequencing technology has greatly improved, from manual sequencing to automated sequencers. A single automated 96-capillary sequencer can provide more than 1000 sequences of 1000 base pairs (bp) per day. Clearly, the development of DNA barcoding is linked to these improvements. Public databases (GenBank, http://www.ncbi.nlm.nih.gov; EMBL, http://www.ebi.ac.uk/embl; DDBJ) has more and more sequence data for the accepted barcoding markers as sequencing facilities improve. The International Barcode of Life project (iBOL) is now under development by the new Canadian International Consortium Initiative (ICI). Researchers from 25 countries will be involved in this large-scale and collaborative program, which aims at building a comprehensive DNA barcode registry for eukaryotic life. The efficiency of DNA barcoding can be described in the detection and description of new cryptic species [25-26] and of sibling species [7]. The CBOL and iBOL have launched campaigns to build DNA barcode libraries of each group. The major targets are fish (Fish-BOL; Ward *et al*. 2009), birds (ABBI; Hebert *et al*. 2004a), mammals (Mammalia Barcode of Life), marine life (MarBOL) and insects. The Canadian Barcode of Life Network (BOLNET) was the first national network for DNA barcoding. Subsequently, the following regions or countries have also initiated projects as a part of the iBOL: Europe (ECBOL; http://www.ecbol.org), Norway (NorBOL; http:// dnabarcoding.no/en/), Mexico (MexBOL); http;//www.mexbol.org/) and Japan (JBOLI; http;//www.jboli.org/). JBOLI provides information and promotes collaborative projects on DNA barcoding in Japan (http://www. jboli.org/en/projects for relevant projects). Different campaigns of iBOL are shown in Table 1

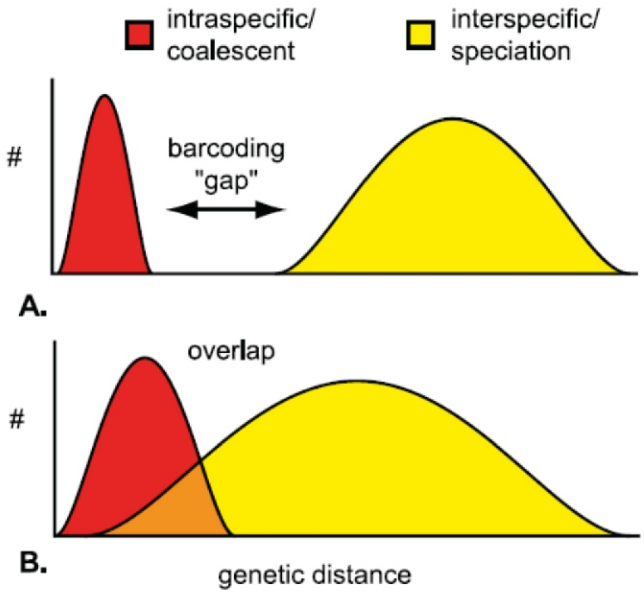

Fig.2 The distribution of intraspecific variation is shown in red and interspecific divergence in yellow

A) Ideal world for barcoding, with discrete distribution and no overlap

B) An alternative version of the world with significant overlap and no gap

Table 1. Current progress of international barcode of life campaigns

Name of campaigns	Total species number	Specimens barcoded	Species barcoded	Clusters recognized by barcodes
Formicidae barcode of life	12205	8495	792 (6%)	1697
Trichoptera barcode of life	13165	17823	2347 (18%)	654
Lepidoptera barcode of life	165000	438341	48676 (29%)	<4000
All bird barcoding initiative	9933	20246	3281 (33%)	31
Coral reef barcode of life	16807	28619	5431 (32%)	No data
Fish barcode of life	31220	60385	7882 (25%)	No data
Mammalia barcode of life	5426	19862	858 (16%)	305
Marine barcode of life	55451	37182	6199 (11%)	No data
Shark barcode of life	1160	4339	557 (48%)	No data

Success rate of barcoding

Various studies and analyses have been performed to determine the success of DNA barcoding for species identification. Meusnier *et al* report barcoding success levels over 97% in studies involving birds, mammals, fishes and arthropods [26]. Hebert *et al* (2003) created a profile of one hundred species from seven diverse animal phyla and then attempted to identify newly analyzed taxa using this profile [5]. This experiment resulted in a 96% success rate of correctly assigning the taxa to the appropriate phylum. Furthermore, each species had a different COI sequence for the barcoding region. This process was repeated with a different data set including eight orders of insects and 50 newly analyzed taxa were correctly assigned to each order. DNA barcoding has its share of flaws which are often more informative than the successes. DNA barcoding encounters problems common to any type of molecular analysis, degradation may make it impossible to amplify a sequence and primers can never be truly universal due to the potential to develop mutations in the primer binding regions [27]

Limitations of barcoding

The first limitation of the barcoding is its single-locus identification system. If several regions from these organelle DNAs are sequenced, this is still a single-locus approach because different genes of mitochondrial or chloroplast DNA are linked. It is known that identical mitochondrial or chloroplast DNA sequences can be present in different related species due to introgression, or due to incomplete lineage sorting since the time of speciation [28]. Another limitation of DNA barcoding lies in the length of the sequences used, usually greater than 500 bp, which prevents the amplification of degraded DNA. This is the case for all environmental samples where the target is DNA from dead animals or dead parts of plants. It is usually difficult to amplify DNA fragments longer than 150 bp from such samples [29]

Nuclear copies of mitochondrial DNA sequences are nuclear mitochondrial DNAs (NUMTs) that have been translocated into the nuclear genome [30]. In eukaryotes, the number and the size of NUMTs are variable, ranging from none or few in Anopheles, Caenorhabditis and Plasmodium, to more than 500 in humans, rice and Arabidopsis.

Technical advances in DNA barcoding

The main aim of the DNA barcoding is to assemble a reference library. Thus it is based on conventional and inexpensive protocols for DNA extraction, amplication and sequencing. This reference library will enable the rapid identification of low taxonomic level taxa with specific short- DNA sequences [29]. Other new molecular technologies used in bioengineering (eg. Siliconbased microarrays, nylon membrane-based macroarrays, etc) are becoming cheaper and may be integrated into the 'second step of DNA barcoding' [31]. Furthermore, new sequencing techniques such as pyrosequencing (454, Solexa, SOLID) enable rapid and representative analyses of mixed samples (e.g. stomach contents, food, blood or water columns). Largely used in the emerging field of metagenomics, this advance could be promising for future DNA barcoding initiatives [32].

Applications of barcoding for Entomology

The unique features of DNA barcoding also provide benefits to both basic and applied entomology. Identifications using molecular data can help elucidate the relationships of morphologically variable individuals of the same species, such as individuals in different developmental stages, castes in social animals and sexually dimorphic individuals [33] Insects, especially those of holometabolous orders, are extremely variable, and numerous attempts have been made to associate their life stages using molecular markers [34]. In addition to the features of typical non-barcode molecular markers, the advantages of DNA barcoding include primer universality, the accumulation of information on a wide range of taxonomic groups, and its association with taxonomy. These advantages may aid the study

of ecologically interesting insect phenomena, such as host plant alternation among aphids, extreme sexual dimorphism and heterotrophic heteronomy of Strepsiptera, as Kathirithamby *et al*. [35] investigated using non-barcode molecular markers.

CONCLUSION

DNA barcoding has become increasingly common since it was proposed in 2003. Currently, more than one million records are available in the BOLD system, which is the official depository of DNA barcode data. The new large-scale project, iBOL, will accelerate the creation of reference barcode libraries and will facilitate the application of this simple identification method.

The BOLD data system is central to the DNA barcoding approach. The specificities of BOLD are (i) to assemble standard information on voucher specimens using common description fields (DNA tag, specimen taxonomy, specimens details, collection information, voucher pictures), and, (ii) its dynamic status that allows taxonomic revisions and reassignment of the deposited Sequences.

REFERENCES

1. Kress JK, Erickson DL. DNA barcodes: genes, genomics, and bioinformatics. PNAS. 2008: 105: 276162.

2. Knowlton N. Sibling species in the sea. *A. Rev. Ecol. Syst.* 1993: 24: 189216.

3. Jarman SN. & Elliott NG. DNA evidence for morphological and cryptic Cenozoic speciations in the Anaspididae, 'living fossils' from the Triassic. *J. Evol. Biol.* 2000: 13, 62433.

4. Frezal L and Leblois R. Four years of DNA barcoding: Current advances and prospects Infection, Genetics and Evolution. 2008: 8: 72736.

5. Hebert PDN, Cywinska A, Ball SL, DeWaard J R. Biological identifications through DNA barcodes Proc R Soc Lond B Biol Sci The Royal Society. 2003: 270: 313-21.

6. Floyd R, Abebe E, Papert A, Blaxter M. Molecular barcodes for soil nematode identification. Mol Ecol. 2002: 11: 83950.

7. Bucklin A, Wiebe PH, Smolecack SB, Copley NJ, Beaudet JG, Bonner KG, Farber-Lorda, J, Pierson JJ. DNA barcodes for species identification of euphausiids (Euphausiacea, Crustacea). J. Plankton Res. 2007: 29: 48393.

8. Johnsen A, Rindal E, Ericson GP, Zuccon D, Kerr KCR, Stoeckle MY, Lifjeld J T . DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species J Ornithol. 2010: 3: 565-78.

9. Hebert PDN, Gregory TR. The promise of DNA barcoding for taxonomy. Syst Biol .2005: *54*: 85259.

10. Jinbo U, Kato T and Ito M. Current progress in DNA barcoding and future implications for entomology. The Entomol Sci. 2011: *14: 107-24.*

11. 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: 29499.

12. Ball SL, Armstrong KF. DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: Lymantriidae). Can. J. For.Res. 2006: 36: 337-50.

13. Ball SL, Hebert PDN, Burian SK, Webb JM. Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. J. North. Am. Benthol. Soc. 2005: 24: 508-24.

14. Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. Identification of birds through DNA Barcodes. *PLoS Biology.* 2004: 2: 312.

15. Valentini A, Pompanon F, Taberlet P. DNA barcoding for ecologists. Trends Ecol Evol. 2008: 24:11017.

16. Litaker RW, Vandersea, MW, Kibler, SR, Reece K.S, Stokes N.A, Lutzoni FM, Yonish BA, West MA, Black, MND, Tester PA. Recognizing dinoflagellate species using ITS rDNA sequences. J. Phycol. 2007: 43: 34455.

17. Evans KM, Wortley AH, Mann DG. An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in Sellaphora (Bacillariophyta). Protist. 2007: 158: 34964.

18. Pegg GG, Sinclair B, Briskey L, Aspden WJ. MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia. Sci. Mar. 70, 712 Pennisi, E., 2007. Wanted: a barocde for plants. Science. 2006: 318: 19095

19. Jaklitsch WM, Komon M, Kubicek C P, Druzhinina IS. Hypocrea crystalligena sp nov., a common European species with a white-spored Trichoderma anamorph. Mycologia.2006: 98: 499513.

20. Saunders GW. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. Philos. Trans. R. Soc. Lond. B .2005: 360: 187988.

21. Randrianiaina RD, Glaw F, Thomas M, Glos J, Raminosoa N, Vences M. Descriptions of the tadpoles of two species of Gephyromantis, with a discussion of the phylogenetic origin of direct development in mantellid frogs. Zootaxa. 2007: 1401, 5361.

22. Kumar NP, Rajavel AR, Natarajan R, Jambulingam P. DNA barcodes can distinguish species of Indian mosquitoes (Diptera: Culicidae). J. Med. Entomol. 2007: 44, 17.

23. Meyer Christopher P and Gustav Paulay. DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biology.* 2005: 3: 2229-38.

24. Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics Trends in Genetics 2007: 23: 167-72.

25. Handfield D, Handfield L. A new species of Plusia (Lepidoptera: Noctuidae) from North America. Can. Entomol. 2006: 138: 85359.

26. Meusnier Isabelle, Gregory AC Singer, Jean-François Landry, Donal A Hickey, Paul DN Hebert, and Mehrdad Hajibabaei. A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics.* 2008: 9: 214.

27. Mitchell A. DNA barcoding demystified. *Aus J Entomol.* 2008: 47: 169-73.

28. Crautlein MV, Korpelainen H, Pietilainen M, Rikkinen J. DNA barcoding: a tool for improved taxon identification and detection of species diversity Biodivers Conserv. 2011: 20:37389.

29. Ballard JWO and Whitlock MC The incomplete natural history of mitochondria. Mol. Ecol. 2004: 13: 729-44.

30. Willams ST, Knowlton N. Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus Alpheus.Mol. Biol. Evol. 2001: 18: 148493.

31. Deagle BE. et al. Quantification of damage in DNA recovered from highly degradedsamples a case study on DNA in faeces. Front.Zool. 2006: 3:11

32. Hudson ME. Sequencing breakthroughs for genomic ecology and evolutionary biology. Mol. Ecol. Resour. 2008: 8: 317.

33. Alcaide M, Rico C, Ruiz S, Soriguer R, Munoz J, Figuerola J. Disentangling vector-borne transmission networks: a universal DNA barcoding method to identify vertebrate hosts from arthropod blood meals. PLoS ONE. 2009: 4: 7092.

34. Ball SL, Armstrong KF. DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: Lymantriidae). Can. J. For.Res. 2006: 36: 337-50.

35. Kathirithamby J, Hayward A, McMahon DP *et al*. Conspecifics of a heterotrophic heteronomous species of Strepsiptera (Insecta) are matched by molecular characterization. *Systematic Entomology.* 2010: 35: 23442.