



# DNA barcoding: a way forward to obtain deep insights about the realistic diversity of living organisms

Rakeeb Ahmad Mir<sup>1</sup> · Kaisar Ahmad Bhat<sup>1,2</sup> · Gazanfer Rashid<sup>2</sup> · Leonard Barnabas Ebinezer<sup>3</sup> · Antonio Masi<sup>3</sup> · Randeep Rakwal<sup>4</sup> · A. A. Shah<sup>1</sup> · Sajad Majeed Zargar<sup>2,3</sup>

Received: 6 February 2020 / Accepted: 23 July 2020  
© Archana Sharma Foundation of Calcutta 2020

## Abstract

DNA barcoding is an innovative tool to rapidly identify and classify organisms based on highly conserved species-specific DNA sequences. This technique has developed in parallel with other genomic-based investigations, which share special emphasis on the acquisition of data pertaining to hotspots regions of DNA (usually short sequence of DNA up to 400–800 bps). These molecular techniques helped answer many queries that remain unanswered by the traditional approaches to accurately identify organisms across the kingdoms of life. Barcoding has narrowed the gap created between diverse range of ecological and traditional studies previously left by the traditional taxonomists. These internal DNA sequence tags, which serve as molecular operational taxonomic units, originate from both nuclear as well as cytoplasmic DNA, making it further convenient to barcode organisms from lower to higher organisms. The most commonly used DNA regions in molecular taxonomic approaches include the internal transcribed spacer regions in nuclear DNA, Cox1 region in mitochondrial DNA, rbcL and mat K gene in chloroplast DNA. Further, many other DNA regions suitable for barcoding have also been identified. It is increasingly evident from a large number of investigations that DNA barcoding has made Linnaean taxonomic system more accessible and convenient to other traditional biologists. Current understanding of the DNA barcoding will further help to understand the basic mechanism and also reveal applications of this approach in different fields. Primarily these approaches offer traditional taxonomists a great opportunity to expand the realistic inventory of the diversity of flora and fauna on planet earth. This review has been written to have better understanding of the mechanism of DNA barcoding, knowledge of DNA regions that can be used as targets, applications and implications of DNA barcoding.

**Keywords** DNA barcoding · MOTU · ITS · Cox1 · rbcL · matK · Molecular taxonomy

---

Corresponding Editor: Amita Pal.

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s13237-020-00330-3>) contains supplementary material, which is available to authorized users.

---

✉ Sajad Majeed Zargar  
smzargar@gmail.com

- <sup>1</sup> Department of Biotechnology, School of Biosciences and Biotechnology, BGSB University, Rajouri, Jammu and Kashmir 185234, India
- <sup>2</sup> Proteomics Laboratory, Division of Plant Biotechnology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Shalimar, Srinagar, Jammu and Kashmir 190025, India
- <sup>3</sup> Department of Agronomy, Food, Natural Resources, Animals and Environment, University of Padova, 35020 Padua, Italy
- <sup>4</sup> TIAS, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8574, Japan

## Introduction

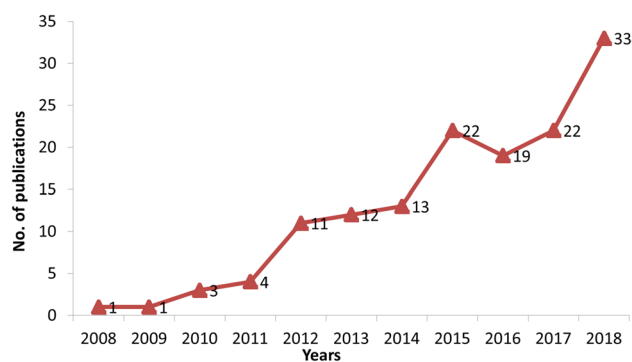
DNA barcoding is a molecular taxonomic technique used to identify the species based on variations in conserved DNA sequences among different species. The meaning of barcode is to analyze multiple samples accomplished by pooling the sequencing of unique identifier sequences. Barcoding has been employed to identify simple to complex living organisms. This approach is entirely different from other marker analysis like short tandem repeats (STR) usually employed for identifying organisms individually. The suitable species-specific DNA sequences are tedious to be identified across the taxa. The variation of these sequences among the species must be very high to clearly demarcate interspecific species variation. Barcode sequences can be retrieved from barcode of life data systems [13] database—comprises of online database of reference DNA sequences used to compare with

unknown DNA sequence of some organism. This database is a repository of species-specific nucleic acid sequences, images specimens and other trace files. The database also facilitates in identification of updated barcode libraries and monitors the number of barcode sequence records and species coverage. DNA barcoding has evolved in concert with genomics-based investigations. A massive online library of DNA barcodes are used as standard reference database that can be employed to identify the unknown samples from garden, forests and even marketed products for validation. In case of animals, barcoding has been previously used to resolve complex groups of organism such as nematodes and mosquitos [8, 48, 50]. The protocols for barcoding of different organisms are already published and available online on Census of Marine Life (CoML).

The number of conserved regions, which serve to distinguish organisms to set up a barcode, is significantly increasing (Supplementary Table 1). The plastome locus frequently used for phylogenetic analysis is *rbcL* and other sequences like *trnL-F* intergenic spacer, *matK*, *ndhF*, and *atpB* have also been reported to be used in several studies [71]. The DNA barcoding using *CoxI* gene has been well established in case of animals, even though it also clearly demarcates plant species [39, 66, 81]. The *trn-H-psb-A* regions in plastid have been proposed as universal barcode for terrestrial plants [48]. A series of other conserved genomic regions have been established for molecular taxonomy by a “plant working group” [16]. Several studies have authenticated the use of these diverse classes of markers to be employed for the DNA barcoding. However, critical evaluation of authenticity of DNA barcoding markers to assess the biological diversity investigation of species is still scanty. In the last few decades, DNA barcoding has been largely employed in order to ensure the accuracy or taxonomic classification of a species, which is quite evident from increasing number of research publications [65]. Based on publications on DNA barcoding in public domain by using specific keyword-based search on CabDirect (<http://www.cabdirect.org/>); an increased trend in the research on DNA barcoding has been observed (Fig. 1).

## DNA-based markers as taxonomic tools

1. *Cytochrome oxidase 1 (COI) gene* The cytochrome oxidase 1 (COI) gene has been very effectively employed to barcode animals. This is an important enzyme involved in electron transport chain system. COI has evolved as the standard DNA based marker to evaluate animal phyla and has been used as universal marker for barcoding of metazoans [39]. The important characteristics like size of gene is conserved and approximately 1600 bp shows heterogeneous substitution patterns as the regions harboring different functional regions [25, 55]. In addition,



**Fig. 1** Trend of publications related to DNA barcoding over the last 10 years (2008–2018). Information was acquired by keyword-based search on CabDirect (<http://www.cabdirect.org/>)

this marker is easy to amplify due to their haploid character and high copy number and evolve at much faster rate as compared to nuclear genes due to the absence of DNA proofreading process [3, 49]. The COI, universal primers are very robust in enabling a wide coverage of barcoding the most representatives of animal phyla [31, 85]. It also possesses inordinate number of properties to be used for phylogenetic studies as compared to other mitochondrial genes. The rate of mutation in COI gene is frequent such that it can help to discriminate the animals that are closely related and also help to resolve the phylo-geographic groups within a single species [39]. Hebert et al. [39] suggested that COXI gene (650 bp) may be enough to resolve the organisms at species and phylum level. The COI gene facilitates the species discrimination because it has higher rate of molecular evolution and it also helps in reconstruction of phylogenetic relationships, recognition of cryptic species and to trace the gene flow patterns [39]. Moreover, it must be kept in consideration that COI gene may not be well suited for characterization of certain nematode species due to high levels of recombination, editing by insertion and multi-partitioning [2, 55, 76]. Due to these complications, amplification of COI gene for barcoding is challenging in marine nematodes [22].

The Fish Barcode of Life campaign (FISH-BOL) is a collaborative international research effort, which seeks to establish a reference library of DNA barcodes for all fish species derived from voucher specimens with authoritative taxonomic identifications [81]. FISH-BOL is a database to record the barcode of maximum species of fishes, as they include half of all vertebrate species. This group includes approximately 15,700 marine and 13,700 freshwater species. The universal primer having a length of 650 bp have been found to flank the ‘Folmer region’ is approved by international DNA barcoding initiative for molecular taxonomy of

metazoans [31]. The ‘Folmer region’ is also tested for barcoding of Odonates [7]. Rach and associates [62] explored COI gene for resolving 51 populations of 23 damselfly and dragonfly species, validating the application of this marker to resolve intra- and inter species specific conflicts to have deep insight into the realistic diversity of metazoans. Further studies reveal that COI cluster help in resolving cryptic species which include *Arctia* sp., *Amnemopsyche* sp., *Hodebertia* sp., *Deinypena* sp., *Palpita* sp., and *Otroeda* sp. [54].

2. *Internal transcribed spacer (ITS)* The internal transcribed spacer (ITS) regions are non-coding sequences found in the ribosomal DNA of nuclear genome. It has been found widely distributed in the photosynthetic organisms. ITS in general is a variable region that surrounds the 5.8S ribosomal RNA gene, which is targeted for amplification. In addition, this region includes the variable DNA sequence areas of the intervening ITS regions called ITS1 and ITS2. Although ITS regions are not translated into proteins, they have a critical role in production of functional rRNA transcripts for ribosome biogenesis. The ITS regions are organized in a stretch of repetitive sequences as 18S, 5.8S and 28S/26S respectively in rDNA of nuclear genome as an operon. ITS1, a non-coding region is found between the 18S and 5.8S coding regions whereas ITS2 is found in between the 5.8S and 28S regions of the operon. The sequences of these coding regions are highly conserved throughout all the kingdoms of life. These conserved regions are targeted for designing the sequence complemented universal primers for amplification by PCR [29]. The primers are easily designed for 18S and 28S genes because they are highly conserved regions [14, 29]. ITS region is highly repetitive, and hence yields higher rate of species identification and reconstruction of phylogenetic trees [48]. It has also been used to resolve the phylogeny of lower taxonomic order and reconstruction of phylogenetic tree of plant species [5]. Several studies have validated and confirmed the application use of ITS as a potential DNA barcoding region [72]. The successful discrimination of 1757 seed plant species was carried out by using ITS regions [49]. ITS2 has been demonstrated to barcode 4800 medicinal plants up to 90 to 93% level [20]. Species identification is accomplished by ITS based DNA barcoding [45, 77], and phylogenetic analysis of plants due to its abundance in genome [43]. These intervening sequences are bearing complex mutations as compared to the rDNA genes, and the sequence heterogeneity within this area has been useful for the separation of both genera and species. Databases of species identification through ITS based barcoding is increasing in developing Genebank [1, 5]. Even though a wide group of scientists believe that ITS exists as potential

marker to barcode for several plant species, recent findings negate its success rate in plants due to prevalence of fungal contamination, paralogous sequences and the difficulty in recovering these sequences [42].

3. *rbcL (ribulose 1,5, biphosphate carboxylase) gene* The chloroplast protein *rbcL* is a large subunit consisting of 476 amino acid residues. The structural and mutational analysis have led to the elucidation of functional sites and conserved regions of this protein. The *rbcL* gene is found in chloroplast DNA of photosynthetic organisms and has been used for constructing the phylogeny of plants. It was successfully used to construct the phylogenies of photosynthetic plants at higher and lower taxonomic level [19, 58]. This gene codes for a large subunit of ribulose-1.5-bisphosphate carboxylase/oxygenase (rubisco) and is most abundant protein on earth [33]. The *rbcL* gene is used to assess the genetic similarities and differences in photosynthetic organisms. The consortium for Barcode of Life [15] plant working group in 2009 proposed that *rbcL* and *matK* genes can be used for the core barcode of most plant species; in addition intergenic sequences ITS and *trnH-psbA* can act as supplement barcodes for further taxonomic resolution [15]. Among these markers *rbcL* has been found to be universal. Amplification of this region is easy and is widely used for phylogenetic analysis of plant species of angiosperms within family and subclass level and other diversified seed plants [18, 42]. The drawback of *rbcL* lies in the existence of variation above species level, and lack of variation at species level makes it a poor marker to discriminate the organisms at species level [26, 48, 59, 67].
4. *matK gene* The *matK* (an open reading frame within the group II intron *trnK*) is commonly used in barcode studies. The *matK* core region locates adjacent to the chloroplast lysine tRNA (*trnK*) gene having a length of 1550 bps. This region encodes a protein called as maturase K which assists RNA editing [83]. The *matK* protein is a degenerate form of a reverse transcriptase enzyme and is called a maturase. This marker is one of the fast evolving among protein coding regions of the plastome [82]. In comparison to *rbcL* the evolution rate is 2–3 times higher in *matK* and lower in comparison to the nuclear ITS regions [32, 44]. Even though the amplification rate of *matK* gene is relatively low but this marker has been commonly used in systematic and evolutionary studies of wide range of plants [46, 58, 67].
5. *trnH-psbA gene* The application of *rbcL* and *matK* markers is limited with respect to phylogenetic analysis and is restricted mostly to forensic identification of species. However, these markers helped essentially to discriminate the huge number of land plants (445 angiosperms, 38 gymnosperms, 67 cryptogams), showing its reli-

ability in discriminating angiosperms [42]. In addition, these markers can only identify species up to genus or family level. Some investigators recommend adding *trnH-psbA* marker to further refine the resolution of species in plants. The small gene *rpS16* intron identify the sterile plant, snakeroot (*Rauwolfia serpentina*) for barcoding rather than using other complex markers [56]. These markers were also found to easily differentiate the mixed root in related species of snakeroots. This gene is located in the non-coding regions of the chloroplast DNA and has a faster rate of evolution. The sequences of this region bear about 75 bp conserved regions flanked at two ends. These end sequences can be employed for designing the universal primers [74]. These regions have been successfully amplified and have been found to be feasible for greater discrimination of plants species [74]. Snakeroot is an endangered species due to its overexploitation as a potent source of reserpine and its use as an anti-psychotic and anti-hypertensive drug. The *rpS16* intron has proven useful in various studies as well [79]. The *rpl16* intron region, and *rps16* intron region was used to discriminate between *R. serpentina* and closely related species in the genus of *Rauwolfia* in order to identify sterile Indian snakeroot samples [55].

6. *tufA* (*EF-Tu*) gene Green macro algae lack *matK* gene, therefore alternatively for barcoding purpose *tufA*, a chloroplast gene, coding for elongation factor *Tu* (*EF-Tu*) is being used [41]. It is sometimes difficult to go through DNA barcoding by using a single barcode, as species may share identical barcodes. So in these cases multiple barcoding regions are used to obtain sufficient distinction between closely related species. An evaluation of *tufA* as a potential DNA barcode marker for the marine green macroalgae was studied and it served well compared to *matk* and *rcbL* [68].

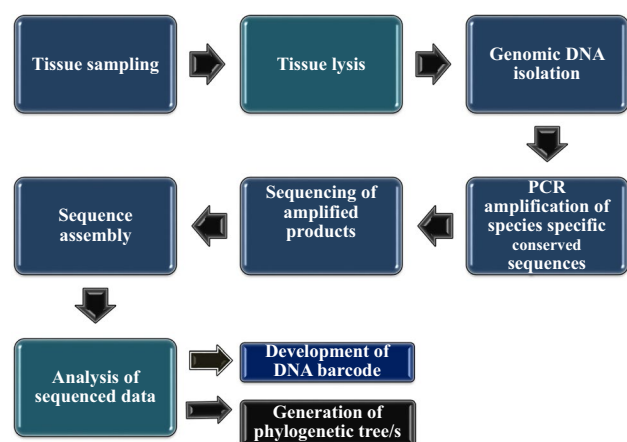
## Methodological steps followed for DNA barcoding of organism/s

**Sampling and preservation** The survey and sampling is done as applicable to the type of organism. Several methods are available for proper survey and sampling of the organism under examination. The collected samples should be adequately preserved to carry out morphological identification. Simultaneously the samples are processed for DNA extraction or direct PCR amplification. Most of the samples are preserved in 4% formalin, which keeps structures intact [40]. Nevertheless, formaldehyde creates hindrances to extract the DNA; formation of formaldehyde-DNA complexes mediates crosslinks between DNA and proteins and promotes fragmentation [35, 70]. Against the backdrop of these limitations, cryopreservation and

ethanol preservation of samples are followed [51, 64]. Several chemicals used for cryopreservation of samples include dimethyl sulfoxide (DMSO), EDTA and NaCl.

**Extraction of DNA** DNA extraction is important step for amplification of suitable DNA segments through PCR. Several protocols are available for extraction of DNA from microorganism, plants and animals [10, 32]. Certain protocols are based on use of alkaline solution (NaOH) and in addition freeze thawing steps and modified steps of these protocols are followed [10, 28]. Other protocols include use of lysis buffers aided by Proteinase K [8, 9]. The basic principle of these protocols are almost prototypical and their aim is to obtain adequate amount of DNA in highly purified form without any contamination of metabolites, that can hinder the PCR amplification process [21].

**Choice of species: specific segment and its PCR-based amplification** The genomic segment of interest is of prime importance to distinguish one organism from other. It differs from organism to organism, but usually most of these regions are applicable to wide range of organism. Mutation rate in these regions must be at least > 2% such that sufficient difference will be established. The universal primers designed to these regions i.e., flanking regions must be conserved, as they must be similar in all species, as already discussed in preceding section. Most often the mitochondrial COI gene is preferred region employed for universal barcoding region [53]. In addition, there must be low rate of recombination and high rates of molecular evolution [4, 61]. A flow chart depicting steps of DNA barcoding from living organisms is given in Fig. 2, however, it needs modification depending on the organism under investigation.



**Fig. 2** Flow diagram representing the steps followed for DNA barcoding of living organisms

## Applications of DNA barcoding

Applications of these diverse classes of species-specific markers to target the specific DNA fragments of genome [6, 60] aim to assess community structure and phylogenetic diversity to further refine taxonomic hierarchy [27, 77]. Direct examination of certain ecosystem to assess the diversity through barcoding or newer sequencing methods are the current approaches to diversity analysis; these approaches are known as meta-barcoding and meta-genomics respectively [73, 75]. General procedure for accomplishing the meta-barcoding is similar to that of DNA barcoding of a species already explained, but only difference lies in analyzing different DNA samples collected from a habitat in threshold level and rest of steps are similar. The desired operational taxonomic unit (OTU) is amplified, sequenced and submitted to repositories such as BOLD or NCBI. Both of these databases are free and can be used to generate the barcode for species. The BOLD can be used to submit the vouchered and non-vouchered DNA sequences whereas NCBI can be used by taxonomists to submit non-vouchered DNA sequences. Both the programs employ the alignment programs to analyze the identity of unknown sequences searching against the data bases.

- a. *Realistic enumeration of biological diversity* The diversity of living organism is surprising and is higher than what is mentioned in the literature. For example, by the application of DNA barcoding it has been found that skipper butterfly (*Astraptes fulgerator*), identified in 1775, comprises of ten different species [38]. Due to intervention of DNA barcoding the classification of organisms has been expanded. It is evident from another example by the revising the classification of orchids, unearthing its great complexity and a widespread plant family with an estimate of > 20,000 members. At least 54 species of bees and 24 butterfly species were added to the community gardens in New York City [38]. Molecular methods have been employed to identify the exact diversity of shark species due to limitation of classical taxonomy [74]. In Australia, DNA barcoding led to the identification of 207 fish species which includes 3 species of chimaerid, 61 species of sharks and 143 species of telesteans [80]. DNA barcoding has been tested for effectiveness of biodiversity assessment of Moths from Nigeria, thus validating and identifying the news species [54]. The accurate depiction of phylogenetic tree reconstruction is accomplished by combined approach of DNA barcode of conserved sequence and super tree method of phylogenetic patterns [24, 47].
- b. *Detection of food adulteration* DNA barcoding can be used in restaurants to convince customers about the authenticity of fancy fishes and other high quality food items are genuine not serving them local breeds or adulterated food items. In addition, several reports have confirmed that DNA barcoding led to the identification of adulterated food products of conserved species (<http://www.dnabarcoding101.org>). The research carried by students from Trinity High School found that sea foods purchased from the grocery stores and restaurants are mislabeled as expensive food items (<http://www.dnabarcoding101.org>). In addition, findings also revealed that a mislabeled fish was the endangered species, the Acadian redfish. The food products sold as sushi food was found to contain whale species in food markets in California and Korea (<http://www.dnabarcoding101.org>). In case of adulterated spices, it is very daunting task to detect the adulterants, but it has been found in the literature that DNA barcoding has proven handy to detect the unethical substitutions in natural products. This technique has emerged an effective tool to trace the adulteration of processed sea and meat products [36]. The application of barcoding is further validated by the experiments carried out on analysis of traditional Chinese medicine (TCM) preparations, which is based on six-herb formula named Liuwei Dihuang Wan. In these formulations, differences in quality and safety concerns were authenticated by presence of potential risky plant products extracted from *Senna obtusifolia* [16].
- c. *Containment of agricultural pests* DNA barcoding helps in controlling the pests by accurate identification to save expenses mediated by infestation of crop pests. The identification of fruit flies and their spreading was stopped at border by the Global Tephritid barcoding initiative. The dependence of DNA barcoding is not on morphology, so one can identify the pests at any stage from egg to adult for initial control of these pests. Early invasions of pathogens can be detected by DNA barcoding for timely development of containment and suppression strategies before the populations of these pests become out of control [36, 69].
- d. *Disease vector identification* Traditional biologists can identify the vector species causing serious damage to crop plants, animals and humans in particular and to understand the mechanism of control. For example, the reference barcode of mosquitoes was developed under global mosquito barcoding initiative to circumvent the diseases spread by mosquitoes that may in turn help health officials to devise the effective control measures and lesser use of insecticides [30].
- e. *Sustaining natural resources* The initial step in any conservation strategy is proper identification of threatened species. The latter is essential to reach the estimated target set by convention on Biological Diversity 2020 for the improvement of biodiversity globally. There are

number of problems associated with the identification based on morphological characters. For this reason, DNA barcoding has been employed as potential technique to resolve this problem. Illegal trade of natural resources can be monitored and controlled by DNA barcoding. Scientists have developed the database as reference barcode library for barcoding hardwood trees to further improve management and conservation strategies of natural resources. [37].

- f. *Monitoring water quality* Investigation of potential microflora in drinking water is an important aspect to control the water borne diseases. DNA barcoding has helped to resolve the identification of microorganisms very quickly to reduce the cost and time required for identification through complex techniques. Environmental agencies use DNA barcoding to determine the quality of water and ensure the supplied water is safe for drinking.
- g. *Precise disease diagnosis* Accurate, reliable and routine diagnostic tools are the need of hour to reduce burden of any diseases. DNA barcoding has helped in detection of various infections and diseases. Among immuno-compromised people various fungal infections such as *aspergillosis*, *candidiasis* and *cryptococcosis* have drastically increased. Now a days the identification of these pathogens is routinely done by fungal DNA barcoding which previously relied on single barcoding region. However, the success rate of identifying the fungi was a bit low. Most recently this gap has been filled by introducing a secondary barcoding region, the translational elongation factor *1α* (*TEF1α*) which ensured efficient and reliable detection of invasive fungal infections [57].

## Conclusion

Barcoding has been an eye-opener to elucidate the realistic biodiversity since decades [12, 23, 28, 34, 39, 63, 78]. Its importance is being highlighted by the diversity examination of highly diverse eukaryotes, biodiversity assessment of which is dwindling due to lack of expertise and presence of frequently existing cryptic species [28]. Molecular taxonomic approaches helped to reclassify cryptic species (morphologically similar species) to circumvent the limitations of Linnaean classification system [11, 82]. Barcoding approaches are mainly established by the use of molecular operational taxonomic units to distinguish between intra and inter-specific species variations [39]. This powerful tool complements the classical taxonomic approaches to decipher the diversity of some taxa, but large majority of nematode species are yet to be unearthed. The sequences obtained from the species specific regions having 97% small subunit (SSU) sequence similarity are usually considered

as substitute for species in prokaryotes. Different studies have shown that ITS-2 acts as a universal barcode to differentiate species among different families of Asteraceae, Fabaceae and other related families [20, 52]. In addition, DNA barcoding has been applied by US Food and Drug Administration to quickly identify fish and fish products rather than using time consuming techniques like isoelectric focusing [84]. It is evident from the existing literature that not even a single technique is 100% accurate to validate all the species. However, one can use more than one marker to generate a barcode for accurately identifying the species as is commonly applicable in plants. Barcoding in conjugation with metagenomics approaches has been used in operational taxonomic units [28, 62] which are different from considering morphological parameters. This approach helps in determining the ecological value of eukaryotic microbiota where distinction for correct identification is a matter of concern [62]. The augmentation of DNA barcoding applications has enhanced the profile of taxonomists and integrated different scientific communities in a broader aspects including social, political and economical arena as is evident from its wide applications. Non experts can use barcoding beyond taxonomy as to interpret DNA barcodes because it is easy, affordable and reliable. DNA barcoding have helped in drug authentication [17] and enhancing biosecurity by identifying exotic species. At the species and population level specific barcoding is believed to become widely used and authentic identification tool in assessing the biodiversity status and conservation.

## Future prospectus of DNA barcoding

DNA barcoding has been widely used to barcode living organisms. Currently more than million barcodes have been generated in the BOLD system (DNA barcode data depository). The barcoding is further accelerated by iBOL to create huge libraries for facilitating proper identification of living organisms. In addition to its role in taxonomic studies, barcoding mediated strong coordination and communications are possible between different scientific communities in particular taxonomists, population geneticists, phylogeneticists and applied biologists.

**Acknowledgements** Sajad Majeed Zargar acknowledges National Mission on Himalayan Studies (NMHS) implemented by Ministry of Environment, Forest & Climate Change (MoEF&CC) for the financial support.

**Author's Contribution** RAM: collected most of the literature and has framed first rough draft. KAB, GR: they added few portions under different headings. LBE, AM, RR: they have reviewed the manuscript and editing the manuscript several times before submission. AAS: contributed in compiling Supplementary Table 1 and Figures. SMZ:

developed concept, framed outline contributed in preparation of first draft and finalizing table and figures.

**Funding** Funding was provided by NMHS (Grant No. GBPNI/NMHS-2017-18/SG24).

### Compliance with ethical standards

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

### References

- Alvarez I, Wendel JF. Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogenet Evol.* 2003;29:417–34.
- Armstrong MR, Blok VC, Phillips MS. A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. *Genetics.* 2000;154:181–92.
- Avise JC. Molecular markers, natural history, and evolution. 2nd ed. Sunderland: Sinauer Associates; 2004.
- Avise JC. Molecular markers, natural history and evolution. New York: Chapman and Hall; 1994.
- Baldwin BG, Sanderson MJ, Porter JM, et al. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Ann Missouri Bot Gard.* 1995;82:247–77.
- Bates ST, Berg-Lyons D, Caporaso JG, et al. Examining the global distribution of dominant archaeal populations in soil. *ISME J.* 2011;5:908–17.
- Bergmann T, Rach J, Damm S, et al. The potential of distance-based thresholds and character-based DNA barcoding for defining problematic taxonomic entities by CO1 and ND1. *Mol Ecol Resour.* 2013;13(6):1069–81.
- Bhadury P, Austen MC, Bilton DT, et al. Evaluation of combined morphological and molecular techniques for marine nematode [*Terschellingia* spp.] identification. *Mar Biol.* 2008;154:509–18.
- Bhadury P, Austen MC, Bilton DT, et al. Exploitation of archived marine nematodes: a hot lysis DNA extraction protocol for molecular studies. *Zool Scr.* 2007;36:93–8.
- Bhadury P, Bik H, Lamshead JD, et al. Molecular diversity of fungal phylotypes co-amplified alongside nematodes from coastal and deep-sea marine environments. *PLoS ONE.* 2011;6:e26445.
- Bickford D, Lohman DJ, Sodhi NS, et al. Cryptic species as a window on diversity and conservation. *Trends Ecol Evolut.* 2007;22:148–55.
- Bik HM, Thomas WK, Lunt DH, et al. Low endemism, continued deep-shallow interchanges, and evidence for cosmopolitan distributions in free living marine nematodes (order Enoplida). *BMC Evol Biol.* 2010;10:389.
- Bold Systems, Barcode of life data systems. <https://www.linkedin.com/company/barcode-of-life-data-systems> (2015).
- Carvalho G, Creer S, Allen M, et al. Genomics in the discovery and monitoring of marine biodiversity. In: Boyen C, Cock M, editors. An introduction to marine genomics. Berlin: Springer; 2010. p. 1–32.
- CBOL Plant Working Group. A DNA barcode for land plants. *Proc Natl Acad Sci.* 2009;106:12794–7.
- Chang CC, Lin HC, Lin IP, et al. *Mol Biol Evol.* 2006;23:279–91.
- Chase MW, et al. Land plants and DNA barcodes: short-term and long-term goals. *Philos Trans R Soc Lond B Biol Sci.* 2005;360:1889–95.
- Chase MW, et al. A proposal for a standardized protocol to barcode all land plants. *Taxon.* 2007;56:295–9.
- Chase MW, Soltis DE, Olmstead RG, et al. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *trnL*. *Ann Missouri Bot Gard.* 1993;80:528–80.
- Chen S, Yao H, Han J, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE.* 2010;5:e8613.
- Creer S, Fonseca VG, Porazinska DL, et al. Ultra-sequencing of the meiofaunal biosphere: practice, pitfalls, and promises. *Mol Ecol.* 2010;19:4–20.
- Derycke SJ, Vanaverbeke A, Rigaux T, et al. Exploring the use of cytochrome oxidase c subunit 1 (COI) for DNA barcoding of free-living marine nematodes. *PLoS ONE.* 2010;5:e13716.
- Derycke S, Remerie T, Vierstraete A, et al. Mitochondrial DNA variation and cryptic speciation within the free-living marine nematode *Pellioditis marina*. *Mar Ecol Prog Ser.* 2005;300:91–103.
- Erickson DL, Jones FA, Swenson NG, et al. Comparative evolutionary diversity and phylogenetic structure across multiple forest dynamics plots: a mega-phylogeny approach. *Front Genet.* 2014;5:358.
- Erpenbeck D, Hooper JNA, Worheide G. CO1 phylogenies in diploblasts and the ‘Barcoding of Life’ are we sequencing a sub-optimal partition? *Mol Ecol Notes.* 2006;6(2):550–3.
- Fazekas AJ, et al. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE.* 2008;3:e2802.
- Fierer N, Leff JW, Adams BJ, et al. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc Natl Acad Sci.* 2012;109:21390–5.
- Floyd RM, Abebe E, Papert A, et al. Molecular barcodes for soil nematode identification. *Mol Ecol.* 2002;11:839–50.
- Floyd RM, Rogers AD, Lamshead PJD, et al. Nematode-specific PCR primers for the 18S small subunit rRNA gene. *Mol Ecol Notes.* 2005;5:611–2.
- Foley DH, Rueda LM, Wilkerson RC. Insight into global mosquito biogeography from country species records. *J Med Entomol.* 2007;44:554–67.
- Folmer O, Black M, Hoeh W, et al. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol.* 1994;3:294–9.
- Fonseca VG, Carvalho GR, Nichols B, et al. Metagenetic analysis of patterns of distribution and diversity of marine meiobenthic eukaryotes. *Glob Ecol Biogeogr.* 2014;23:1293–302.
- Freeman S. Biological science. San Francisco: Pearson/Benjamin Cummings; 2008.
- Gadek PA, Alpers DL, Heslewood MM, et al. Relationships within Cupressacea sensulato: a combined morphological and molecular approach. *Am J Bot.* 2000;87:1044–57.
- Gagna CE, Lambert WC, Kuo HR, et al. Localization of B-DNA and ZDNA in terminally differentiating fiber cells in the adult lens. *J Histochem Cytochem.* 1997;45:1511–21.
- Haack RA. Exotic bark and wood-boring Coleoptera in the United States: recent establishments and interceptions. *Can J For Res.* 2006;36:269–88.
- Hartvig I, Czako M, Kjær DE, et al. The use of DNA barcoding in identification and conservation of rosewood (*Dalbergia* spp.). *PLoS ONE.* 2015;10:e0138231.
- Hebert PD, Penton EH, Burns JM, et al. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci USA.* 2004;101(41):14812–7.
- Hebert PDN, Cywinska A, Ball SL, et al. Biological identifications through DNA barcodes. *Proc R Soc Lond Ser B Biol Sci.* 2003;270:313–21.
- Heip CHR, Vincx M, Vranken G. The ecology of marine nematodes. *Oceanogr Mar Biol Annu Rev.* 1985;23:399–489.

41. Henriques H, Bagatini L, Marques GC, et al. *tufA* gene as molecular marker for freshwater Chlorophyceae. *Algae*. 2016;31:155–65.
42. Hollingsworth PM, Li DZ, Van der B, et al. Telling plant species apart with DNA: from barcodes to genomes. *Philos Trans R Soc B*. 2016;371:20150338.
43. Hřibová E, Čížková J, Christelová P, et al. The ITS1-5.8 S-ITS2 sequence region in the Musaceae: structure, diversity and use in molecular phylogeny. *PLoS ONE*. 2011;6:e17863.
44. Johnson LA, Soltis DE. Phylogenetic inference in Saxifragaceae sensu strieto and Gilia (Polemoniaeeae) using *matK* sequences. *Ann Missouri Bot Gard*. 1995;82:149–75.
45. Kalivas A, Ganopoulos I, Xanthopoulou A, et al. DNA barcode ITS2 coupled with high resolution melting (HRM) analysis for taxonomic identification of *Sideritis* species growing in Greece. *Mol Biol Rep*. 2014;41:5147–55.
46. Khidir WH, Lawrence AA. Evolutionary implications of *matK* indels in Poaceae. *Am J Bot*. 1999;86:1735–41.
47. Kress WJ, Erickson DL, Swenson NG, et al. Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot. *PLoS ONE*. 2010;5:e15409.
48. Kress WJ, Erickson DL, Swenson NG, et al. Use of DNA barcodes to identify flowering plants. *PNAS*. 2005;102:8369–74.
49. Li DZ, Gao LM, Li HT, et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *PNAS*. 2011;108:19641–6.
50. Lin CP, Danforth BN. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Mol Phylogenet Evol*. 2004;30:686–702. [https://doi.org/10.1016/S1055-7903\(03\)00241-0](https://doi.org/10.1016/S1055-7903(03)00241-0).
51. Liu BZ, Song LS, Xiang JH. DNA isolation and RAPD analysis of the *Argopecten irradians* tissue preserved with different methods. *Mar Sci*. 2001;25:51–3.
52. Liu J, Michael M, Gao L, Zhang D, et al. DNA barcoding for the discrimination of Eurasian yews (*Taxus L.*, Taxaceae) and the discovery of cryptic species. *Mol Ecol Resour*. 2011;11:89–100.
53. Lorenz JG, Jackson WE, Beck JC, Hanner R. The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philos Trans R Soc Lond B*. 2005;360:1869–77.
54. Lotanna MN, Adeniyi CA, Yun-Yu WA, et al. Testing the effectiveness of DNA barcoding for biodiversity assessment of moths from Nigeria. *Diversity*. 2020;12:85. <https://doi.org/10.3390/d12020085>.
55. Lunt DH, Hyman BC. Animal mitochondrial DNA recombination. *Nature*. 1997;387:247. <https://doi.org/10.1038/387247a0>.
56. Marcel CME, Frederic L, Csilla P, et al. Forensic identification of Indian snakeroot (*Rauvolfia serpentina* Benth ex Kurz) using DNA barcoding. *J Forensic Sci*. 2013;58:3–8.
57. Hoang MTV, Irinyi L, Chen SC, Sorrell TC, Meyer W. Dual DNA barcoding for the molecular identification of the agents of invasive fungal infections. *Front Microbiol*. 2019;10:1647.
58. Mort ME, Soltis DE, Soltis PS, et al. Phylogenetic relationships and evolution of Crassulaceae inferred from *matK* sequence data. *Am J Bot*. 2001;88:76–91.
59. Newmaster SG, Fazekas AJ, Ragupathy S. DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach. *Can J Bot*. 2006;84:335–41.
60. Orgiazzi A, Bianciotto V, Bonfante P, et al. Pyrosequencing analysis of fungal assemblages from geographically distant, disparate soils reveals spatial patterning and a core mycobiome. *Diversity*. 2013;5:73–98.
61. Piganeau G, Gardner M, Eyre-Walker A. A broad survey of recombination in animal mitochondria. *Mol Biol Evol*. 2004;21:2319–25.
62. Porazinska DL, Giblin-Davis RM, Sung W, et al. Linking operational clustered taxonomic units (OCTUs) from parallel ultra-sequencing (PUS) to nematode species. *Zootaxa*. 2010;2427:55–63.
63. Rach J, Bergmann T, Paknia O, DeSalle R, et al. The marker choice: unexpected resolving power of an unexplored CO1 region for layered DNA barcoding approaches. *PLoS ONE*. 2017;12:e0174842. <https://doi.org/10.1371/journal.pone.0174842>.
64. Reiss RA, Schwert PD, Ashworth AC, et al. Field preservation of Coleoptera for molecular genetic analyses. *Environ Entomol*. 1995;24:716–9.
65. Rob D, Paul G. Review and interpretation of trends in DNA barcoding. *Front Ecol Evol*. 2019;7:302. <https://doi.org/10.3389/fevo.2019.00302>.
66. Rubinoff D, Cameron S, Will K. Are plant DNA barcodes a search for the holy grail? *Trends Ecol Evol*. 2006;21:1–2.
67. Sass C, Little DP, Stevenson DW, et al. DNA barcoding in the cycadales: testing the potential of proposed barcoding markers for species identification of cycads. *PLoS ONE*. 2007;2:e1154.
68. Saunders GW, Kucera H. An evaluation of *rbcL*, *tufA*, UPA, LSU and ITS as DNA barcode markers for the marine green macroalgae. *Crypt Algal*. 2010;31:487–528.
69. Scheffer SJ, Lewis ML, Joshi RC. DNA barcoding applied to invasive leaf miners (*Diptera: Agromyzidae*) in the Philippines. *Ann Entomol Soc Am*. 2006;99:204–10.
70. Serth J, Kuczyk MA, Paeslack U, et al. Quantitation of DNA extracted after micro-preparation of cells from frozen and formalin-fixed tissue sections. *Am J Pathol*. 2000;156:1189–96.
71. Shaw J, Lickey EB, Beck JT, et al. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am J Bot*. 2005;92:142–66.
72. Sickel W, Ankenbrand MJ, Grimmer G, et al. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecol*. 2015;15:20. <https://doi.org/10.1186/s12898-015-0051-y>.
73. Simon C, Daniel R. Metagenomic analyses: past and future trends. *Appl Environ Microbiol*. 2011;77:1153–61.
74. Smith P, Benson P. Biochemical identification of shark fins and fillets from the coastal fisheries of New Zealand. *Fish Bull*. 2001;99:351–5.
75. Taberlet P, Prud'Homme SM, Campione E, et al. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Mol Ecol*. 2012;21:1816–20.
76. Vanfleteren JR, Vierstraete AR. Insertional RNA editing in metazoan mitochondria: the cytochrome-b gene in the nematode *Tera-tocephalus lirellus*. *RNA*. 1999;5:622–4.
77. Vogel TM, Simonet P, Jansson J, et al. TerraGenome: a consortium for the sequencing of a soil metagenome. *Nat Rev Microbiol*. 2009;7:252.
78. Wang XC, Liu C, Huang L, et al. ITS-1: a DNA barcode better than ITS2 in eukaryotes? *Mol Ecol Res*. 2015;15:573–86. <https://doi.org/10.1111/1755-0998.12325>.
79. Wang Y, Wang W, Tong W, Zahao W. Analysis of chloroplast subunit S16 (rpS16) intron sequences in *Morus* (*Urticales: Moraceae*). *Afr J Biotech*. 2011;10:17695–9.
80. Ward RD, Holmes BH, White WT, Last PR. DNA barcoding Australasian chondrichthyans: results and possible uses in conservation. *Mar Freshw Res*. 2008;59:57–71.
81. Ward RD, Zemlak TS, Innes BH, et al. DNA barcoding Australia's fish species. *Philos Trans R Soc Lond B Biol Sci*. 2005;360:1847–57.
82. Winker K. Sibling species were first recognized by William Derham. *Auk*. 2005;1718(122):706–7.
83. Wolfe KH. Protein-coding genes in chloroplast DNA: compilation of nucleotide sequences, data base entries, and rates of molecular



- evolution. In: *The photosynthetic apparatus: molecular biology and operation*; 1991. p. 467–82.
84. Yancy HF, Zemlak TS, Mason JA, et al. Potential use of DNA barcodes in regulatory science: applications of the regulatory fish encyclopedia. *J Food Prot.* 2008;71:210–7.
  85. Zhang DX, Hewitt GM. Assessment of the universality and utility of a set of conserved mitochondrial primers in insects. *Insect Mol Biol.* 1997;6:143–50.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.