Journal of Advances in Biology & Biotechnology

Volume 27, Issue 7, Page 709-720, 2024; Article no.JABB.118581
ISSN: 2394-1081

DNA Barcoding: Accelerating Insect Species Discovery and Biodiversity Documentation

Rupali J. S. [a], Vidya Madhuri E. [a], Gundreddy Raja Reddy [a], Voodikala S. Akhil [b], Ramya N. [c] and Sagar D. [d*]

[a] Division of Entomology, ICAR-Indian Agricultural Research Institute, New Delhi-110012, India.
[b] Division of Nematology, ICAR-Indian Agricultural Research Institute, New Delhi-110012, India.
[c] ICAR- Indian Institute of Agricultural Biotechnology, Ranchi-834010, India.
[d] ICAR- National Bureau of Agricultural Insect Resources, Bengaluru- 560024, India.

Authors' contributions

This work was carried out in collaboration among all authors. Author RJS conceptualized the work. Authors VME and VSA Conceptualized (supporting) the work. Authors RN, GRR, SD, and RJS wrote, reviewed, and edited the manuscript. Author RJS wrote the original draft (lead). Author SD wrote the original draft (supporting). All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/jabb/2024/v27i71030

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/118581

Review Article

Received: 14/04/2024
Accepted: 17/06/2024
Published: 20/06/2024

ABSTRACT

Species identification is essential for recognizing and describing biodiversity. Traditionally, this process has relied on morphological diagnosis through taxonomic studies, which have certain constraints such as subjectivity and time-consuming processes. With the advancement of modern molecular techniques, DNA barcoding has gained global attention. The term "DNA barcoding" refers to the technique of establishing species-level identification by sequencing a short fragment of the mitochondrial cytochrome c oxidase subunit I (COI) gene, the "DNA barcode," from a specimen

**Corresponding author: E-mail: garuda344@gmail.com;*

Cite as: J. S. , Rupali, Vidya Madhuri E., Gundreddy Raja Reddy, Voodikala S. Akhil, Ramya N., and Sagar D. 2024. "DNA Barcoding: Accelerating Insect Species Discovery and Biodiversity Documentation". Journal of Advances in Biology & Biotechnology 27 (7):709-20. https://doi.org/10.9734/jabb/2024/v27i71030.

that is taxonomically unknown and comparing it to a reference library of barcodes from known species. This review article explores the evolution of DNA barcoding, its universal marker, and its application in insect taxonomy, emphasizing its role in accelerating species discovery and biodiversity documentation. In India, DNA barcoding initiatives have made considerable progress, yet there remains a vast opportunity to barcode the country's rich insect diversity. Overall, DNA barcoding emerges as a powerful tool to address the urgent need for efficient species identification and biodiversity conservation in an ever-changing world.

Keywords: Barcode; COI; sequencing; biodiversity; taxonomy.

1. INTRODUCTION

Insects are the most abundant of all life on earth and have evolved into various forms. They represent about 66% of all identified species, making up over three-quarters of the planet's biodiversity. Approximately one million insect species have been documented although only 7% to 10% have been scientifically described [1]. Given the many undiscovered species, estimates suggest there could be around eight million insect species globally [2]. According to Mayr and Ashlock [3], it took nearly 200 years for taxonomists to describe 1.7 million species which is only 10 percent of the total number of species estimated. In this context identification of insects has been a monumental task where it calls for the availability of more specialists and funding. But with the dwindling interest in taxonomy and fund availability, classification and identification of various life forms particularly insects have been a major challenge to the scientific community. Naturalists developed the concept of categorizing living things based on taxonomy, a field of science that aids in categorizing a living entity based on morphological features to catalogue the enormous number of species [4,5].

A novel technique termed DNA barcoding, a tool of DNA-based taxonomy is currently being used to identify known and undiscovered species based on the pattern of nucleotide arrangement in a particular species' DNA fragment [6]. Several researchers have suggested the use of DNA barcoding in taxonomy as a method to achieve rapid species identification in the context of the current biodiversity crisis [7,8]. With a total land area of around 3,287,263 km^2, India ranks among the world's most biodiverse countries, home to a diversity of habitats from deserts to high mountains and tropical to temperate woods [9]. The current necessity to classify such huge diversity calls for a quick, efficient, and accurate solution. DNA barcoding is the use of a short, standardized fragment of DNA sequence to identify and assign unknown specimens to species identity [10]. Besides, it facilitates the detection of new species based on the differences in DNA barcodes. For insect species, a 658 bp section of the mitochondrial cytochrome c oxidase (COX I) gene is extensively utilized for DNA barcoding [7]. The different barcode libraries *viz*., NCBI, BOLD are gaining value due to the integration of information of a species through voucher specimens, their binomial names, type locality and other collection data, and morphology in the form of digital photographs [11]. This method is widely feasible to catalog all the species on our planet. Over time it is largely accepted by hard-core taxonomists and governmental and non-governmental organizations as well. Since the development of molecular biology and molecular tools, identifying various life forms, including insects, has become simple, fast, and accurate.

2. HISTORY OF DNA BARCODING

In 1960, Carl Woese gave this concept for the first time. He utilized rRNA and molecular markers like rDNA and mtDNA for discovering *Archaea* i.e., prokaryotes, and for drawing an evolutionary tree [12]. In 2003, Paul Herbert, a researcher at the University of Guelph, Ontario, Canada proposed “DNA Barcoding” to identify species. He is considered the “Father of DNA Barcoding”. The first article on DNA barcoding was published by Hebert in 2003 with the title “Biological identifications through DNA barcodes”.

3. UNIVERSAL DNA BARCODE REGION OR UNIVERSAL MARKER FOR DNA BARCODING

Mitochondria are energy-producing organelles, found in nearly every cell in almost every plant and animal species. The mitochondrial genome, which is present in all the eukaryotic organisms and evolves more quickly than nuclear DNA, has proven to be incredibly helpful in tracing evolutionary history. Different inheritance

patterns can be seen in nuclear and mitochondrial genomes. Since mitochondrial DNA (mtDNA) is hereditary from the mother, it evolves quickly, and most nucleotide alterations occur at neutral sites. Mitochondrial markers are employed to indicate phylogenetic relationships among related groups. Using the sequence information obtained from the COX 1 marker gene amplification, the intra- and inter-phylogenetic interactions regarding this genetic marker have been investigated. In humans, Cytochrome c oxidase I (COX1), also referred to as mitochondrially encoded cytochrome c oxidase I (MT-CO1), is a protein encoded by the MT-CO1 gene. In other eukaryotes, this gene is known as COX1, CO1, or COI. COX1 serves as the primary subunit of the cytochrome c oxidase complex. A region approximately 650 base pairs in length from the 5' end of the Cytochrome c oxidase subunit 1 (COI) gene has been proposed as the universal barcode for animals [13].

Barcodes that are shorter than the full-length barcode are often referred to as “mini barcodes.” They have the additional advantage that they more easily can be amplified when the DNA is damaged or fragmented, which is common in environmental DNA samples. In addition to the COI gene, some of the other markers such as 16S rRNA, 12S rRNA, and CytB are also being used for metabarcoding. However, the reference libraries for these alternative markers are small in comparison with those for COI [14].

4. NEED FOR NEW MARKERS

An ideal barcoding marker should contain highly conserved sequence regions that allow for the design of universal primers capable of amplifying all taxa of interest in the sample. These conserved regions should flank a highly variable region, which can then be used to differentiate between species. Because of the redundancy in the genetic code and the fact that COI is a protein-coding gene, the third position of most codons is highly variable. This variability complicates the design of primers for metabarcoding that can provide sufficient taxonomic coverage. The sample will inevitably contain a range of mismatches between the primers and the templates, which will result in variable primer affinities for various templates. Less mismatched primer-template combinations will amplify more readily with each cycle, which could lead to a severe overrepresentation of these sequences in the PCR output. These "universal" COI primer biases have been empirically verified in several investigations. LepF1/LepR1 primer biases have been reported [15-17]. Folmer primers fail to amplify many species of Hymenoptera [18]. Several primer pairs are associated with amplification bias resulting in an overrepresentation of Diptera and Lepidoptera sequences. The bias can be somewhat reduced by using degenerate primers [19,20]. The amplification performance of four COI primer pairs from Malaise trap samples was investigated for several taxonomic groupings with varying degrees of degeneracy. Degeneracy significantly impacted amplification success, ranging from 5% for primers with little degeneracy to 49% for primers with considerable degeneracy [21].

Many experiments have sought alternative markers due to the amplification bias associated with COI primers. For studies involving a broad taxonomic range (up to the phylum level), it is common to use a highly conserved and easily amplifiable marker, such as the nuclear small subunit ribosomal RNA (rRNA) gene (18S) [22]. There are several examples of 18S metabarcoding, most of which include eukaryotic microorganisms and soil/sediment biodiversity evaluation. The mitochondrial large subunit rRNA gene (16S) has been evaluated for insect metabarcoding, yielding promising results. When applied to a set of 315 species of insects (representing 264 genera and 23 orders), In silico analyses demonstrated that 200 bp mini-barcodes of the 16S gene identified slightly more species compared to mini-barcodes of the COI gene of the same length. Moreover, the taxonomic coverage was higher for 16S (75%-90%) compared to COI (only 50%) [19]. However, longer COI mini barcodes enhanced the taxonomic resolution between closely related species to nearly 100%, whereas the resolution for 16S reached a maximum of 85%. Surprisingly, the taxonomic coverage and resolution of 16S were constant across the 11 insect orders that were examined, but the best taxonomic coverage of COI was only between 0% and 47% in all other insect orders and slightly above 50% in Diptera and Lepidoptera. 16S amplified more species and more evenly through orders, improving biomass estimation. According to them, COI is still the best option if the objective is to identify the species present in the sample because there are numerous public reference databases available. However, 16S would be a better option if the objective is to assess the biodiversity in terms of numbers rather than species names. A further benefit of

16S metabarcoding over COI is that amplicons cannot be confused for nuclear pseudogenes or Wolbachia [20].

Nuclear rRNA markers present a different scenario. The rRNA sequences of the large and small ribosomal subunits (18S and 28S) include conserved regions that allow for the design of primers with broad coverage, much like mitochondrial rRNA markers. However, the resulting amplicons tend to be highly conserved, resulting in very low taxonomic resolution [23]. The internal transcribed spacer (ITS), a nuclear marker, might be the most useful for metabarcoding. It has the benefit of being flanked by conserved areas (subunits 5.8S and 28S), which makes primer design possible, and is known to provide strong taxonomic resolution. For fungi, the ITS reference database is comprehensive. Unfortunately, this is not the case for insects, which face an indistinguishable situation to that of some mitochondrial genes mentioned here. However, ITS is unquestionably a viable option for a survey when the separation of MOTUs suffices. In addition to ongoing efforts to develop reference databases for entire mitochondrial genomes and specific mitochondrial markers, the entomological community could significantly benefit from establishing reference databases for promising nuclear metabarcoding markers, such as ITS. With low levels of primer degeneracy and stringent PCR conditions, rRNA markers offer considerably broader taxonomic coverage while still resolving most genetically distinguishable species [24].

5. CURRENT STATUS OF INSECT DNA BARCODING IN INDIA

In India, approximately 62,429 insect species across 595 families have been described, but only 2,330 species from 264 families have DNA barcodes i.e., only 3.73%. BOLD contains barcodes for 852,657 different insect species from 12 mega-diverse nations. The country with the highest recorded number of barcodes is Costa Rica, which accounts for 77% of all reported barcodes. South Africa, China, and Mexico each contribute about 5%, while India barely makes up 1.53%. With 13,152 sequences (including 10,570 COI-5P and other suitable barcode markers supported by the Consortium for the Barcode of Life for Animals), representing 2330 species gathered from various geographic regions throughout India, India ranks seventh among the given megadiverse countries [25].

The species belong to 20 different orders, which are (in decreasing number of samples) Lepidoptera, Hemiptera, Diptera, Coleoptera, Thysanoptera, Hymenoptera, Odonata, Blattodea, Ephemeroptera, Orthoptera, Neuroptera, Siphonaptera, Embioptera, Mantodea, Psocoptera, Trichoptera, Zygentoma, Dermaptera, Strepsiptera and Phasmatodea. Among these orders, the largest number of sequences was generated for Lepidoptera (26.08%), followed by Hemiptera (25.87%), Diptera (16.09%), Coleoptera (12.66%), Thysanoptera (6.55%), Hymenoptera (6.40%), Odonata (2.29%), Blattodea (1.78%), Ephemeroptera (1.17%) and the rest of the eleven orders comprise <1%. The highest species coverage was achieved for, Lepidoptera with 687 species (29.48%), followed by Hemiptera with 391 spp. (16.78%), Coleoptera 373 spp. (16.01%), Diptera 332 spp. (14.25%), Hymenoptera 209 spp. (8.97%), Odonata 117 spp. (5.02%), Thysanoptera 89 spp. (3.82%), and Ephemeroptera with 44 spp. (2.22%), while the remaining 12 orders together comprise <3%. The substantial percentage of known insect species that have not yet been barcoded indicates that there is a tremendous opportunity to work on the barcoding of Indian insects [25].

6. ADVANTAGES OF DNA BARCODING

DNA barcoding offers numerous advantages in insect identification, including high precision, speed, and the capability to distinguish various species when traditional morphological methods fail. One of the key benefits is its accuracy, as DNA barcoding often surpasses the limitations of conventional morphological methods by providing a consistent and objective means of identifying insect species through genetic sequences [26]. Additionally, DNA barcoding is a rapid method, delivering results in a matter of hours to days, which makes it suitable for extensive ecological investigations and biodiversity assessments [27]. This method is particularly useful in pest management, as it provides rapid and accurate identification of pest species, enabling timely and effective control measures.

Furthermore, DNA barcoding facilitates the identification of various insect life stages, such as eggs, larvae, or even damaged specimens, which can be challenging to identify using conventional morphological techniques [28]. Importantly, DNA barcoding has also led to the discovery of previously unrecognized cryptic

species that were indistinguishable based on morphology alone, revealing hidden biodiversity [26]. The objectivity and reproducibility of DNA barcoding minimize the subjectivity in species identification that can arise with traditional taxonomy, especially when dealing with cryptic species [29].

7. LIMITATIONS AND DRAWBACKS OF DNA BARCODING

Despite its advantages, DNA barcoding faces several limitations. A significant challenge is the lack of universal primers, as using a single universal identifier does not enable the successful barcoding of all insect groups. This necessitates the use of multiple markers or specific primers for some taxa, complicating studies [7]. Incomplete reference databases also pose a challenge, as DNA barcoding relies on comprehensive databases that may not be complete for certain regions or insect groups, leading to difficulties in accurate species identification.

Hybridization and introgression between closely related species can produce mixed or false genetic signatures, complicating the identification process [30]. Additionally, high intraspecific genetic variation within some insect species can hinder accurate species identification, as DNA barcoding might not effectively differentiate between individuals within a single species [7]. Specimens collected from certain environments may have degraded DNA, making it challenging to obtain high-quality sequences for barcoding, which is particularly relevant for museum collections [31].

Convergent evolution can result in distinct species having similar or identical sequences, leading to incorrect identifications [29]. Biological anomalies, such as parthenogenesis or endosymbiont associations in some insects, can also affect the interpretation of DNA barcodes. The lack of reference sequences for rare, newly discovered, or undescribed species further complicates identification efforts. Moreover, DNA barcoding can be expensive, requiring specialized equipment and expertise in molecular biology techniques. Limited taxonomic coverage in reference databases for some insect groups, particularly non-model organisms, reduces the efficiency of DNA barcoding. These limitations highlight the need for the development of new markers, improved techniques, and broader reference databases. While DNA barcoding is an effective tool, it is essential to combine DNA data with other sources of information, such as morphology and ecology, for a comprehensive approach to insect identification and taxonomy [32].

8. ADVANCEMENTS IN DNA BARCODING

A contemporary technique for identifying several species in a mixed sample, such as bulk DNA or environmental DNA (eDNA), is DNA metabarcoding. High-throughput sequencing (HTS) of a specific DNA marker is necessary for this method to work. DNA metabarcoding, in contrast to traditional DNA barcoding, which often uses Sanger sequencing on individual specimens, makes use of the vast amounts of DNA sequence data produced by HTS to rapidly allocate taxonomic classes to a variety of species present in a sample [33]. It takes DNA barcoding to the next level by allowing the processing of bulk samples and environmental samples. Samples collected from traps such as yellow pan traps, and malaise traps can be processed in bulk without sorting them. Even in situations when the species are not physically present, DNA metabarcoding can be used to extract species DNA from samples such as soil, water, sediments, or other materials. Currently, the most popular high-throughput sequencing (HTS) platform for DNA metabarcoding research is the Illumina MiSeq, located in San Diego, California. DNA metabarcoding helps in the detection of invasive insects in a surveillance context and highlights the unique technical and regulatory challenges that must be considered when implementing high-throughput sequencing technologies into sensitive diagnostic applications. Large-scale species identification using this technology is rapidly becoming an affordable option, particularly when traditional morphology-based identification presents budgetary or logistical difficulties.

Mitochondrial metagenomics, emerging as a promising alternative to PCR-based metabarcoding, presents an innovative avenue for biodiversity studies. By bypassing the limitations associated with PCR amplification, such as the overestimation of species richness due to the amplification of nuclear mitochondrial pseudogenes (NUMTs), mitochondrial metagenomics offers a more comprehensive approach to characterizing mixed-specimen samples or environmental samples [34]. Through the assembly of complete mitochondrial genomes from shotgun sequencing of specimen

mixtures, mitochondrial metagenomics enables the acquisition of a wealth of data on species diversity and community composition [35]. This method, known as 'genome skimming,' entails sequencing libraries at low depth to assemble the high-copy portion of the mitochondrial genome, resulting in the assembly of dozens or even hundreds of mitogenomes from raw reads [36].

The application of mitochondrial metagenomics holds promise in enhancing our understanding of biodiversity by providing finer-scale resolution and insights into species diversity and community composition. Unlike PCR-based approaches, mitochondrial metagenomics does not introduce PCR-induced biases and allows for the recovery of whole mitochondrial genomes, strengthening conclusions drawn from species identification, phylogenetics, and phylogeography [37]. Moreover, the use of shotgun sequencing approaches suppresses the recovery of single-copy NUMTs, leading to better estimation of species richness in genomic mixtures [38]. This method also enables the simultaneous analysis of species composition, abundance, genetic variation, and evolutionary relationships from a single biotic sample, offering a comprehensive view of biodiversity. The success of mitochondrial metagenomics in analysing environmental samples of mixed specimens underscores its potential as a powerful tool for biodiversity research [39].

Before mitochondrial metagenomics can be widely adopted, it is essential to carefully examine key parameters and distinguish between read-based and contig-based analyses. While read-based analyses rely on matching sequence reads against known reference data, such as DNA barcodes or full mitochondrial genomes, contig-based analyses aim at de novo assembly of mitogenomes from sequence reads mixtures. While the read-based approach has been more commonly utilized, the contig-based approach remains less explored but shows promise in phylogenetics and proof-of-principle studies [40]. Mitochondrial metagenomics holds the potential to revolutionize biodiversity research, offering a fast, efficient, and accurate method for analysing species diversity and community structure across various taxa and ecosystems.

NGS technology was initiated for commercial use in 2005. It is also called massively parallel sequencing because it allows the sequencing of millions of DNA fragments from thousands of DNA templates in parallel. Sanger sequencing technology can generate the sole method for sequencing readings of up to 1000 bases for DNA.sequencing for nearly three decades, but next-generation sequencing (NGS) devices are now beginning to dominate the sequencing niche. Next-generation sequencing technologies allow the sequencing of millions of DNA fragments, from thousands of DNA templates in parallel and facilitate the generation of DNA barcodes more quickly and at a lower total cost [41]. As a result, numerous genomics facilities have transitioned to NGS from Sanger sequencers. By using NGS the entire genome can be sequenced in 2 days, and it has greatly revolutionized genomics.

Mobile DNA barcoding is possible through third-generation sequencing platforms such as Oxford nanopore technologies. MinION is the smallest and most user-friendly portable sequencer that can be run outside of the conventional laboratory such as in fields and forest areas. MinION sequencing offers a rapid and cost-effective approach for analyzing smaller samples, making it more suitable for day-to-day border detections. It allows for in-situ species monitoring without having to remove organisms from their habitat. It can produce full-length DNA barcodes, unlike Illumina and other second-generation sequencing methods. By using MinION sequencing technology, morphology-based identification will be supplemented, allowing for more informed biosecurity decision-making, and offering a vastly quicker and less expensive alternative to the current Sanger sequencing molecular identification method [42]. Nanopore sequencing is a unique, scalable technology that enables direct, real-time analysis of long DNA or RNA fragments. It functions by keeping an eye on variations in an electrical current that occur as nucleic acids go through a protein nanopore. To obtain the precise DNA or RNA sequence, the resultant signal is decoded. A single fragment of DNA can produce lengthy reads with about two million base reads.

9. THE CHALLENGE OF NUMTS IN DNA BARCODING

DNA barcoding relies on the assumption that the sequences used, particularly the COI region of mitochondrial DNA, are orthologous (i.e., they are equivalent across species). If this assumption is violated and paralogous sequences (non-orthologous copies) are mistaken for orthologs,

incorrect species identification can result [43]. The DNA barcoding initiative assumes that every organism has a unique, identifiable molecular tag in the COI region, which can be amplified using universal primers. This method assumes the COI sequences being compared are orthologous across different species [7]. However, several molecular evolutionary processes can disrupt this assumption, including gene duplication within the mitochondrial genome, heteroplasmy (multiple types of mitochondrial genomes in one individual), bacterial infections altering mtDNA variation, and nuclear integration of mtDNA (NUMTs) [44].

NUMTs are extremely common, with cases reported in numerous eukaryotic species. For instance, in humans and mice, nearly all mitochondrial sequences have NUMT counterparts in the nuclear genome [45]. Studies have identified thousands of possible NUMTs in species such as the honey bee and the flour beetle [46]. The widespread nature of NUMTs suggests that many species may have NUMTs for the COI gene in their nuclear genome [37]. The existence of NUMTs presents a serious challenge for DNA barcoding. While universal primers are designed to amplify mitochondrial DNA across different species, they can also inadvertently amplify NUMTs, especially if the nuclear versions have diverged significantly from the mitochondrial ones [47]. This can result in the amplification of paralogous sequences, complicating species identification [48]. Research has shown that species like grasshoppers and crayfish have NUMTs of the COI gene, which can lead to incorrect species identification. In these cases, barcoding might incorrectly suggest that single individuals belong to multiple unique species [49-51]. This demonstrates the importance of thorough data exploration and the need to consider NUMTs in DNA barcoding practices.

Several strategies can be employed to mitigate the impact of nuclear mitochondrial DNA segments (NUMTs) in DNA barcoding. First, using specific primer sets can significantly reduce the likelihood of co-amplifying NUMTs. Primers should be designed to target regions specific to the mitochondrial genome, thereby minimizing the chance of amplifying NUMTs inadvertently. These primers should be highly specific to the mitochondrial gene region to ensure accurate amplification of the intended sequences [34]. Bioinformatic filters are another essential tool in sequence analysis. These filters can identify and exclude NUMTs based on distinct characteristics such as insertions or deletions (indels), in-frame stop codons, and nucleotide composition. Implementing filters that use open reading frame length and nucleotide profiles, which are built via hidden Markov model analyses of large COI barcode datasets, can effectively remove NUMTs from the data [52].

A thorough examination of sequence characteristics is crucial in identifying and eliminating NUMTs. By carefully analysing the sequences, researchers can distinguish between mitochondrial and nuclear-origin sequences. Targeting longer amplicons in metabarcoding and environmental DNA (eDNA) studies is another effective approach, as it reduces the likelihood of encountering NUMTs. Longer sequences are less likely to contain NUMTs, thereby improving the reliability of the barcoding process [53]. Additionally, reverse-transcription PCR (RT-PCR) can be employed to discriminate between NUMTs and their mitochondrial counterparts. Since NUMT sequences are not transcribed, RT-PCR can be used to target only the transcribed mitochondrial sequences. Incorporating a repository for all identified NUMTs in databases like BOLD (Barcode of Life Data Systems) can help exclude these sequences from new data, ensuring that only genuine mitochondrial sequences are considered [34]. Improving informatics platforms is also vital. Enhanced filtering based on indels, protein-coding sequences (IPSCs), and alignments to known COI barcode datasets can help identify and remove NUMTs. Using curated databases of COI barcodes allows researchers to verify records that derive from genuine mitochondrial COI sequences, similar to the approach described by Andujar et al. These curated databases can provide a reliable reference for identifying and excluding NUMTs from barcoding data [52,53]. By implementing these strategies, the impact of NUMTs on DNA barcoding can be significantly reduced, thereby improving the accuracy and reliability of species identification.

10. BLAST & BOLD: INTEGRAL PLATFORMS FOR GENETIC ANALYSIS AND DATA MANAGEMENT

10.1 BLAST – Basic Local Alignment Search Tool

A bioinformatics tool called BLAST allows you to compare the sequences of two or more proteins or nucleic acid molecules and one sequence to a

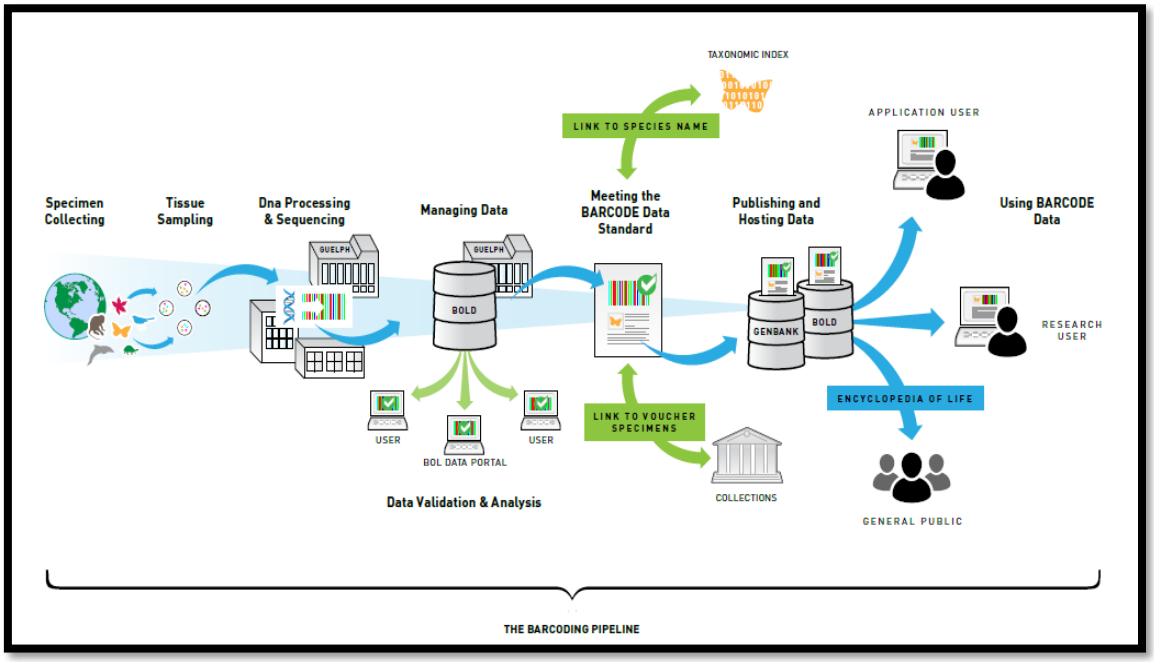

Fig. 1. The barcoding pipeline (www.barcodeoflife.org)

group of sequences in a database. The NCBI offers a common matching tool that looks for similarities between a query sequence and a sequence library. It breaks the query and database sequence into fragments and seeks matches between them. It can infer the evolutionary and functional links between sequences and identify the individuals who make up gene families.

10.2 BOLD – The Barcode of Life Data System

BOLD is an informatics workbench that makes it easier to collect, store, analyze, and publish DNA barcode records. It is a repository for specimen and sequence records. It helps with barcode data administration, quality control, and analysis. By combining flexible security and data entry features with web-based delivery, it offers a means of collaboration between geographically dispersed research communities.

11. INTERNATIONAL BARCODE OF LIFE CONSORTIUM

Established in 2008, the International Barcode of Life Consortium (iBOL) is a global research alliance dedicated to transforming biodiversity science. The consortium has initiated three major projects: Barcode 500K, BIOSCAN, and the Planetary Biodiversity Mission. The first major initiative, BARCODE 500K, ran from 2010 to 2015 with the primary goals of delivering DNA barcodes for 500,000 species and developing the necessary informatics tools and analytical protocols for DNA barcoding. Following this, the BIOSCAN project, spanning from 2019 to 2026, aims to deliver DNA barcodes for 2 million species and promote the various purposes of DNA barcoding. Looking further ahead, the PLANETARY BIODIVERSITY MISSION, scheduled from 2026 to 2045, seeks to complete a comprehensive census of all multicellular species, establish a global biosurveillance program, and construct a 'library of life' by preserving DNA extracts from all species. These initiatives reflect the consortium's commitment to enhancing our understanding and documentation of global biodiversity through advanced genetic tools.

iBOL Conference Series: The International Barcode of Life conference series is a biennial event that began in 2005. It serves as a major platform for the international community to discuss advancements and collaborations in DNA barcoding and biodiversity science. The series has grown significantly in participation and scientific scope over time.

12. CONCLUSION

DNA barcoding has arisen as a transformative tool in entomology, providing a standardized and powerful method for identifying insect species.

This technology has revolutionized insect taxonomy, biodiversity monitoring, and ecological research by offering rapid, accurate, and objective species identifications. DNA barcoding assists taxonomists in resolving complex species groups and enables a broader range of researchers, including citizen scientists, to participate in research on studies of insect diversity. This technique is pivotal in conserving endangered species and managing invasive pests, thereby supporting global biodiversity preservation and agricultural sustainability. The impact of DNA barcoding extends beyond traditional taxonomic boundaries. It facilitates the monitoring of insect populations, assists in ecological studies, and boosts our understanding of insect behaviour and interactions within ecosystems. The contributions of DNA barcoding to conservation efforts are particularly significant, as it helps in the identification and protection of endangered species, ensuring their survival in the face of environmental changes and human activities.

As technological advancements continue, the expansion of reference databases and the development of more efficient methods will further enhance the capabilities of DNA barcoding. The future of this technology in entomology looks promising, with the potential to reveal new insights into insect diversity, behaviour, and ecological roles. This versatile tool is set to remain a cornerstone of entomological research, playing a critical role in efforts to understand, protect, and manage insect populations in a rapidly changing world. Its capacity to offer comprehensive data on insect species will continue to support biodiversity conservation, pest management, and the study of ecological dynamics, making it an indispensable resource for scientists and conservationists alike.

13. FUTURE PERSPECTIVES

As technology and our understanding of DNA continue to advance, DNA barcoding will likely play an increasingly pivotal role in entomological research and insect management. One key area of development lies in the expansion and refinement of reference databases. Efforts to comprehensively catalog the DNA barcodes of insect species from diverse geographical regions will enhance the accuracy and applicability of this tool. This, in turn, can bolster biodiversity conservation and biosecurity efforts, aiding in the rapid identification of invasive species and the protection of endangered ones. Moreover, the incorporation of high-throughput sequencing technologies, for example, next-generation sequencing, will accelerate the DNA barcoding process, enabling researchers to analyze large-scale insect datasets more efficiently. This can be particularly valuable in monitoring and understanding complex insect communities and ecological interactions.

DNA barcoding will also continue to serve as a crucial tool for taxonomists, helping to resolve cryptic species complexes and providing an objective basis for species delimitation. Concerning disease vectors and agricultural pests, DNA barcoding will facilitate the development of targeted and effective control strategies, helping to mitigate the economic and health impacts of insect-borne diseases. Furthermore, the integration of machine learning and artificial intelligence with DNA barcoding data will allow for the development of automated identification tools that can be employed by non-experts, including citizen scientists and field workers. These tools will empower a broader community to contribute to insect biodiversity monitoring and research.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

ACKNOWLEDGEMENTS

The authors express their gratitude to ICAR-Indian Agricultural Research Institute, New Delhi, India, for granting the essential resources to pursue this review work.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Zhang ZQ. Animal Biodiversity: An Introduction to higher-level classification and taxonomic Richness. Zootaxa. 2011; 3148:7-12.
2. Samways MJ. Insect diversity conservation. Cambridge University Press, New York. 2005:25-29.

3. Mayr E, Ashlock PK. Principles of systematic zoology. McGraw-Hill, Inc., New York. 1991:475.
4. Tiwari, Awanindra Kumar. IPM Essentials: Combining biology, ecology, and agriculture for sustainable pest control. Journal of Advances in Biology & Biotechnology. 2024;27(2):39-47. Avaialble:https://doi.org/10.9734/jabb/2024/v27i2697
5. Wilson JJ. DNA barcodes for insects. DNA barcodes: Methods and Protocols. 2012;17-46.
6. Novotny V, Basset Y, Miller SE, Weiblen GD, Bremer B, Cizek L, Drozd P. Low host specificity of herbivorous insects in a tropical forest. Nature. 2002;416:841-844.
7. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proc R Soc Lond A. 2003;270:313-321.
8. Hebert PDN, Ratnasingham S, deWaard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Lond B. 2003;270(Suppl 1):96-99.
9. Stork NE. How many species of insects and other terrestrial arthropods are there on earth? Annu Rev Entomol. 2018;63:31-45.
10. Antil S, Abraham JS, Sripoorna S, Maurya S, Dagar J, Makhija S, et al. DNA barcoding, an effective tool for species identification: a review. Molecular Biology Reports. 2023;50(1):761-775.
11. Wilson JJ. DNA barcodes for insects. Methods Mol Biol. 2012;858:17-46.
12. Sapp J, Fox GE. The singular quest for a universal tree of life. Microbiol Mol Biol Rev. 2013;77(4):541-550.
13. Park DS, Suh SJ, Oh HW, Hebert PD. Recovery of the mitochondrial COI barcode region in diverse Hexapoda through tRNA-based primers. BMC Genomics. 2010;11:1-7.
14. Marquina D, Andersson AF, Ronquist F. New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. Mol Ecol Resour. 2019;19:90-104.
15. Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ. Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE. 2011;6:17497- 7.
16. Brandon-Mong GJ, Gan HM, Sing KW, Lee PS, Lim PE, Wilson JJ. DNA metabarcoding of insects and allies: An evaluation of primers and pipelines. Bull Entomol Res. 2015;105:717-727.
17. Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z. Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods Ecol Evol. 2012;3:613-623.
18. 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. Mol Mar Biol Biotechnol. 1994;3:294-299.
19. Clarke LJ, Soubrier J, Weyrich LS, Cooper A. Environmental metabarcodes for insects: In silico PCR reveals potential for taxonomic bias. Mol Ecol Resour. 2014; 14:1160-1170.
20. Elbrecht V, Leese F. Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Front Environ Sci. 2017; 5:11.
21. Moriniere J, Cancian de Araujo B, Wai Lam A, Hausmann A, Balke M, Schmidt S, Haszprunar G. Species identification in malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix. PLoS ONE. 2016;11.
22. Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, de Vargas C. CBOL protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol. 2012;10.
23. Wangensteen OS, Palacin C, Guardiola M, Turon X. DNA metabarcoding of littoral hard-bottom communities: High diversity and database gaps revealed by two molecular markers. PeerJ. 2018;6.
24. Schoch CI, Seifert KA, Huhndorf S, Robert V, Spouge JI, Levesque CA, Schindel D. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. Proc Natl Acad Sci USA. 2012;109:6241-6246.
25. Shashank PR, Naveena NL, Rajgopal NN, Elliott TA, Sreedevi K, Sunil S, Meshram NM. DNA barcoding of insects from India: Current status and future perspectives. Mol Biol Rep; 2022. Avaialble:https://doi.org/10.1007/s11033-022-07628-2.
26. Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. Identification of birds through

DNA barcodes (Charles Godfray, Ed.). PLoS Biol. 2004;2.
27. Hajibabaei M, Singer GAC, Hebert PDN, Hickey DH. DNA barcoding: How it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. 2007;23(4).
28. Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN. DNA barcodes and cryptic species of skipper butterflies in the genus Perichares in Area de Conservacion Guanacaste, Costa Rica. Proc Natl Acad Sci USA. 2008;105(17):6350-6355.
29. Meier R, Shiyang K, Vaidya G, Ng PK. DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. Syst Biol. 2006; 55(5):715-728.
30. Rubinoff D, Holland BS. Between two extremes: Mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst Biol. 2005; 54(6):952-961.
31. Meyer CP, Paulay G. DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol. 2005;3(12).
32. Geary J, Reay T, Bubel T. The impact of heterogeneity in a global knowledge commons: Implications for governance of the DNA barcode commons. Int J Commons. 2019;13(2):909–930.
33. Liu M, Clarke LJ, Baker SC, Jordan GJ, Burridge CP. A practical guide to DNA metabarcoding for entomological ecologists. Ecol Entomol. 2020;45:373-385.
34. Song H, Buhay JE, Whiting MF, Crandall KA. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc Natl Acad Sci USA. 2008;105:13486-13491.
35. Tang M, Tan M, Meng G, Yang S, Su X, Liu S, et al. Multiplex sequencing of pooled mitochondrial genomes—a crucial step toward biodiversity analysis using mito-metagenomics. Nucleic Acids Res. 2015; 43.
36. Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, et al. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. GigaScience. 2013;2:4.
37. Bensasson D, Zhang DX, Hartl DL, Hewitt GM. Mitochondrial pseudogenes: Evolution's misplaced witnesses. Trends Ecol Evol. 2001;16:314-321.
38. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol. 2012;21: 2045-2050.
39. Shokralla S, Spall JL, Gibson JF, Hajibabaei M. Next-generation sequencing technologies for environmental DNA research. Mol Ecol. 2012;21:1794-1805.
40. Gillett CPDT, Crampton-Platt A, Timmermans MJTN, Jordal BH, Emerson BC, Vogler AP. Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). Mol Biol Evol. 2014;31: 2223-2237.
41. Shokralla S, Gibson JF, Nikbakht H, Janzen DH, Hallwachs W, Hajibabaei M. Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. Mol Ecol Resour. 2014;14:892-901.
42. Abeynayake SW, Fiorito S, Dinsdale A, Whattam M, Crowe B, Sparks K, Campbell PR, Gambley CA. A rapid and cost-effective identification of invertebrate pests at the borders using MinION sequencing of DNA barcodes. Genes. 2021;12:1138.
43. Funk DJ, Omland KE. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Evol Syst. 2003;34:397–423.
44. Rubinoff D, Cameron S, Will K. A genomic perspective on the shortcomings of mitochondrial DNA for ‘‘barcoding’’ identification. J Hered. 2006;97:581–594.
45. Richly E, Leister D. NUMTs in sequenced eukaryotic genomes. Mol Biol Evol. 2004; 21:1081–1084.
46. Pamilo P, Viljakainen L, Vihavainen A. Exceptionally high density of NUMTs in the honeybee genome. Mol Biol Evol. 2007;24: 1340–1346.
47. Zhang D-X, Hewitt GM. Nuclear integrations: Challenge for mitochondrial DNA markers. Trends Ecol Evol. 1996;11: 247–251.
48. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly

Astraptes fulgerator. Proc Natl Acad Sci USA. 2004;101:14812–14817.
49. Bensasson D, Zhang D-X, Hewitt GM. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. Mol Biol Evol. 2000;17:406–415.
50. Williams ST, Knowlton N. Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus Alpheus. Mol Biol Evol. 2001;18:1484–1493.
51. Nguyen TTT, Murphy NP, Austin CM. Amplification of multiple copies of mitochondrial cytochrome b gene fragments in the Australian freshwater crayfish, Cherax destructor Clark (Parastacidae; Decapoda). Anim Genet. 2002;33:304–308.
52. Hebert PDN, Bock DG, Prosser SWJ. Interrogating 1000 insect genomes for NUMTs: A risk assessment for estimates of species richness. PLoS One. 2023 Jun 8;18(6). Avaialble:https://doi.org/10.1371/journal.pone.0286620.
53. Moulton MJ, Song H, Whiting MF. Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: A case study from Orthoptera (Arthropoda: Insecta). Mol Ecol Resour. 2010;10(4): 615-627.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/118581