DNA BARCODING AND MOLECULAR TAXONOMY: PRINCIPLES AND APPLICATIONS

Aniket Marathe[1], Omkar Yadav[2*]

[1]Department of Zoology, Shantarambhau Gholap Arts, Science and Gotirambhau Pawar Commerce College, Shivle, 421401
[2]Department of Zoology, Amdar Shashikant Shinde Mahavidyalay, Medha, Tal. Jaoli, Dist. Satara, 415012

Keywords

Voucher Specimen, Minibarcodes, Barcode Gap, Phylogeography, Integrative Taxonomy

Received - *24th June 2025*

Online Available - *18th October 2025*

Abstract

DNA barcoding is a standardized method for species identification using short, conserved gene sequences. The foundation relies on the "barcode gap" a measurable difference between intra and interspecific genetic variation and a suite of kingdom-specific markers: COI for animals, rbcL + matK for plants and ITS for fungi. The process follows a rigid workflow from non-destructive sampling and DNA extraction to PCR amplification, Sanger/NGS sequencing and bioinformatics analysis resulting in a validated sequence entry in BOLD (Barcode of Life Data System) with a BIN. Applications span three major areas: Ecological Monitoring (using metabarcoding and eDNA to assess biodiversity and trophic interactions), Forensics and Commerce (detecting food fraud, illegal wildlife trade, and authenticating herbal products) and Conservation Genomics (identifying cryptic species and Evolutionarily Significant Units (ESUs)). Limitations exist, notably the confusion caused by hybridization, high intraspecific variation, and the risk of amplifying pseudogenes (NUMTs). Ultimately, DNA barcoding is a powerful integrative tool driving rapid advances in biodiversity science and conservation with future improvements anticipated from genome skimming and automation.

I. Theoretical Foundations and Workflow of DNA Barcoding

A. The Conceptual Basis of DNA Barcoding

1. Definition and Historical Context

DNA barcoding is the process of using a short, standardized gene sequence as a unique molecular identifier for every species. The concept was formally introduced in 2003 by Paul Hebert and colleagues who proposed the 650 bp region of the mitochondrial Cytochrome c Oxidase subunit I (COI or COX1) gene as the universal marker for the animal kingdom (Hebert *et al.,* 2003b).

Rationale for COI: This gene is preferred in animals because it has a high mutation rate, particularly at the third codon position, which provides sufficient variation to distinguish closely related species, yet its flanking regions are highly conserved, allowing for the use of universal PCR primers across vast animal phyla (Folmer *et al.,* 1994). Moreover, it is a mitochondrial gene, meaning it is present in high copy numbers per cell, which is advantageous when working with degraded or minute samples.

2. The "Barcode Gap" Hypothesis

The efficacy of DNA barcoding relies entirely on the existence of a genetic discontinuity between species.

a. Principle of Genetic Distance (The Gap): The hypothesis posits that the distribution of genetic divergence values shows a gap between the maximum intraspecific distance (variation within a single species) and the minimum interspecific distance (variation between two sister species) (Meyer & Paulay, 2005). If an unknown sequence falls into a cluster below the gap, it belongs to that species. If it falls above, it likely represents a distinct species.

The Marker Genes: While COI is the animal standard, this single-locus approach faced challenges in plants due to a slow mutation rate in their mitochondrial genes. The scientific community, through the Consortium for the Barcode of Life

*Correspond author E-mail address: *sarpaveda@gmail.com*

Marathe et al., (2025) Biodiversity: Emerging Tools and Challenges, https://doi.org/10.5281/zenodo.17580937

(CBOL), adopted a two-locus combination for land plants: rbcL and matK (The CBOL Plant Working Group, 2009). For fungi, the Internal Transcribed Spacer (ITS) region of ribosomal DNA is the established primary barcode (Schoch *et al.,* 2012).

b. Threshold Setting: The determination of a reliable divergence threshold is a complex, empirical task. A commonly cited, though not universally applicable, threshold for COI in animals is approximately 2−3% divergence (Hebert *et al.,* 2203a). However, species that have recently diverged, or those undergoing hybridization and introgression, may exhibit incomplete lineage sorting (ILS), which can blur the barcode gap and lead to misidentification (Funk & Omland, 2003). For example, studies on certain insect taxa show high intraspecific divergence, suggesting the presence of cryptic species that require lower-level taxonomic refinement (Sperling, 2003).

3. Molecular Taxonomy vs. Traditional Taxonomy

A. DNA barcoding offers an integrative taxonomic approach (Padial *et al.,* 2010).

Complementary Strengths: Traditional taxonomy is crucial for providing the morphological context and formal names required for the reference library. DNA barcoding offers a diagnostic tool that bypasses the need for specialized knowledge of complex morphological characters, such as the male genitalia of many insects, which can be difficult to interpret or are missing in juveniles. This is particularly valuable for:

Identifying Cryptic Species: Barcoding often flags genetically distinct lineages that were previously grouped as a single species based on morphology (e.g., in the genus *Rheotanytarsus*) (Wang *et al.,* 2024).

Forensic and Commercial Applications: Verifying species identity in highly processed materials, like dried medicinal herbs or seafood products, where morphological features are absent (Baker *et al.,* 2021).

B. The Standard DNA Barcoding Workflow (The Lab Pipeline)

1. Field Sampling and Specimen Management

a. Non-destructive vs. Destructive Sampling techniques: The sampling strategy prioritizes the preservation of the voucher specimen—the physical specimen that links the molecular data to the formal taxonomy. Non-destructive methods include taking only a small tissue snip or a single leg, leaving the remainder intact for morphological analysis (Chase *et al.,* 2021).

b. Crucial Preservation Methods: Rapid preservation is essential to prevent the action of nucleases (enzymes that degrade DNA). 95% or 100% molecular-grade ethanol is the standard for fresh animal tissue, as it simultaneously dehydrates the tissue and inactivates nucleases (Fukatsu, 2009). Storing samples in silica gel is the preferred method for plant tissues because it removes water quickly, inhibiting degradation.

2. DNA Extraction and Quantification

a. Review of common methods: The choice of extraction method is often a trade-off between cost/yield and purity/speed.

CTAB-based methods (Cetyl Trimethyl Ammonium Bromide) are robust and high-yielding, specifically effective at removing high levels of polysaccharides and polyphenolic compounds common in plant tissues, which can inhibit PCR (Doyle & Doyle, 1987).

Commercial Kits (e.g., column-based kits) utilize spin columns to bind and purify DNA, offering a significantly faster, less labor-intensive workflow with reduced use of hazardous chemicals (Särkinen *et al.,* 2012).

b. Quality Control: DNA is typically quantified using a spectrophotometer. The A260/A280ratio is used to check for protein contamination, and the A260/A230 ratio assesses contamination by chemicals like guanidine salts, which are often residues from the extraction process (Sambrook *et al.,* 1989). High-quality DNA is mandatory for reproducible PCR.

3. Polymerase Chain Reaction (PCR) Amplification

a. Universal Primer Design: Universal primers for DNA barcoding are designed to target the highly conserved flanking regions of the target barcode gene, ensuring a broad taxonomic reach. For example, the Folmer primers are widely used for the COI region in animals (Folmer *et al.,* 1994).

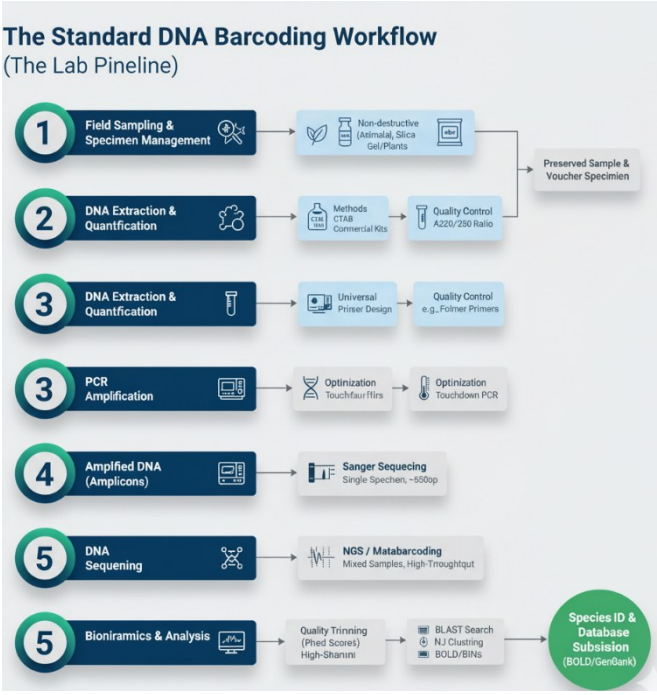
b. Optimization: PCR optimization, particularly the annealing temperature (Ta), is crucial. A "touchdown PCR" protocol is often employed, where the Ta is progressively lowered across cycles. This increases the specificity of primer binding at higher temperatures initially and boosts product yield later, ensuring amplification across a wide range of species

where the primer-binding site may have minor sequence differences (Roux, 1995).

4. DNA Sequencing

a. Sanger Sequencing: Sanger sequencing remains the "gold standard" for single-specimen barcoding, yielding a reliable, long read that covers the full ~650 bp barcode region with high accuracy (Sanger *et al.,* 1977). It is the backbone for building the global barcode reference libraries.

b. Next-Generation Sequencing (NGS) Integration (Metabarcoding): NGS platforms (e.g., Illumina)



have transformed the application of barcoding into DNA metabarcoding. This technique uses highly parallel sequencing to identify multiple species from a mixed environmental sample (e.g., soil, water eDNA, bulk insect trap catches) in a single run (Taberlet *et al.,* 2018). While NGS reads are shorter, its massive throughput allows for deep sequencing to resolve the complex diversity of an entire community.

5. Bioinformatics and Analysis

a. Sequence Assembly and Quality Trimming (Phred scores): Bioinformatics begins with cleaning the raw data. Phred scores are logarithmic measures of the base-call accuracy (Ewing & Green, 1998). Sequences are trimmed from both ends until the Phred score remains above a set threshold (e.g., Phred 20, meaning 99% accuracy). Forward and reverse reads are then aligned and merged to create a high-quality consensus sequence (a contig).

b. Database Submission: BOLD (Barcode of Life Data System) is the world's most comprehensive and dedicated public repository for DNA barcode records (Ratnasingham & Hebert, 2007). Data submission is tightly controlled, requiring validated taxonomic identification, GPS coordinates, and specimen images. Submitting to BOLD is often followed by submission to GenBank (NCBI).

c. Identification Methods:

BLAST Search: A similarity search against sequences in GenBank, providing a quick, initial indication of the closest match (Altschul *et al.,* 1990).

Neighbor-Joining (NJ) clustering: A distance-based phylogenetic method used within BOLD to group sequences based on their genetic distance, visually demonstrating the clustering of conspecifics and the separation of distinct species (Saitou & Nei, 1987).

Operational Taxonomic Units (OTUs) / Barcode Index Numbers (BINs): BINs are a system implemented by BOLD that uses an algorithm to automatically cluster sequences based on similarity, providing an objective, provisional species identifier. The BIN system is a crucial tool for estimating species richness, particularly for undescribed species (Ratnasingham & Hebert, 2013).

II. Molecular Markers in Taxonomic Resolution: COI, rbcL, ITS, matK, and others

The success of DNA barcoding hinges on selecting genetic regions that exhibit a critical balance: conservation in the flanking regions (for universal primer design) and divergence in the core region (for species discrimination).

A. The Standard for Animals: Cytochrome c Oxidase Subunit I (COI)

1. Characteristics: The Mitochondrial Advantage

The COI gene resides in the mitochondrial DNA (mtDNA), offering key advantages for barcoding (Hebert *et al.,* 2003a):

Maternal Inheritance: mtDNA is typically inherited solely from the mother, meaning there is no recombination, leading to clear, distinct lineages.

High Copy Number: Each cell contains hundreds to thousands of mitochondria, ensuring a high quantity of target DNA, which is vital for successful PCR amplification from degraded or minute samples (e.g., gut contents, museum specimens).

High Mutation Rate: Compared to nuclear genes, the animal mitochondrial genome evolves relatively quickly, particularly at the third codon position, providing the necessary interspecific variation for species separation.

2. The Folmer Region: The Standard Barcode

The ≈650 base pair (bp) region at the 5′ end of the COI gene, often referred to as the Folmer region, is the standard animal barcode. Its universality is largely due to the widely successful Folmer primers (Folmer *et al.,* 1994) which amplify this region across numerous invertebrate and vertebrate phyla. This ≈650 bp segment is the primary data source for the Barcode of Life Data System (BOLD) and its assignment of Barcode Index Numbers (BINs).

3. Exceptions and Adaptations: Secondary Markers

While highly effective across most Metazoa (animals), COI exhibits limitations in specific taxa, often necessitating a multilocus approach or the use of secondary markers (Hebert *et al.,* 2003b):

Cnidaria (Corals, Anemones): COI evolves too slowly in some groups (e.g., Anthozoa, including hard corals), providing insufficient resolution for recent speciation events. Alternative mitochondrial markers like mtMutS (a mismatch repair gene) or 16S rRNA are often used instead (Shearer *et al.,* 2002).

Nematodes (Roundworms): Difficulties in obtaining reliable COI sequences in some parasitic and free-living nematodes led to the adoption of the 18S or 28S ribosomal RNA (rRNA) genes as more consistent markers (Blaxter *et al.,* 2205).

Degraded DNA (Minibarcodes): In highly processed or old samples (e.g., forensic evidence, ancient specimens), the 650 bp region may be too long to amplify. Researchers use "minibarcodes"—shorter fragments (typically 100−200 bp) of the COI gene that are more readily amplified from fragmented DNA (Meusnier *et al.,* 2008).

B. Markers for Plants and Algae

The animal COI marker is ineffective in plants due to the slow evolutionary rate of COI in the chloroplast genome (Hollingsworth *et al.,* 2011). The search shifted to the plant's chloroplast (cp) genome and nuclear ribosomal DNA.

1.Ribulose-1,5-bisphosphate Carboxylase/Oxygenase Large Subunit (rbcL)

a. Characteristics: rbcL is a chloroplast gene encoding the large subunit of the crucial photosynthetic enzyme RuBisCO. It is highly conserved, making it exceptionally easy to amplify with universal primers across virtually all land plants. This high universality makes it an excellent tool for family- and genus-level phylogenetic studies, ensuring high success rates even in degraded samples (Kress *et al.,* 2009). However, its conservation means it often lacks the resolution to distinguish among closely related species.

2. Maturase K (matK)

a. Characteristics: matK is also a chloroplast gene (an intron maturase). It has one of the highest rates of nucleotide substitution among protein-coding chloroplast genes. This high divergence is essential for achieving species-level resolution, often outperforming rbcL in separating closely related plant species (Kress *et al.,* 2009). The main technical challenge lies in designing universal primers due to its high variability, which can sometimes lead to lower amplification success rates than rbcL.

3. The Two-Locus Plant Barcode

The CBOL Plant Working Group (2009) formally proposed the combination of rbcL and matK as the official core DNA barcode for land plants. This combination is a classic example of a multilocus approach, where:

rbcL provides universality (high PCR success and broad taxonomic coverage).

matK provides resolution (high species discrimination).

4. Secondary Plant Markers

To address species complexes or groups with low divergence in rbcL and matK:

trnH-psbA: An intergenic spacer (non-coding region) in the chloroplast genome known for its high variation and insertion/deletion (indel) events. While highly discriminating, indels can make sequence alignment difficult, reducing its suitability as a primary marker (Kress *et al.,* 2009).

ITS2: The Internal Transcribed Spacer 2 of the nuclear ribosomal DNA offers high resolution and is now widely used, especially for identifying medicinal and processed herbs (Chen *et al.,* 2010).

C. Markers for Fungi, Bacteria, and Protists

1. Fungi: Internal Transcribed Spacer (ITS)

a. The Official Barcode: The Internal Transcribed Spacer (ITS) region of the nuclear ribosomal DNA (rDNA) was formally accepted as the primary barcode for Fungi by the International Barcode of Life Project (Schoch *et al.,* 2012). The ITS region is a multi-copy gene cluster in the nuclear genome, consisting of two hypervariable spacers (ITS1 and ITS2) flanking the conserved 5.8S rRNA gene. Its high copy number and fast evolution rate make it ideal for resolving species boundaries.

b. The ITS1 vs. ITS2 Debate: The entire ITS region (≈600 bp) is often sequenced, but research suggests that the ITS2 sub-region alone, due to its conserved surrounding structure, offers sufficient resolution and is more easily aligned across diverse fungal groups. Some studies propose ITS2 as a powerful "minibarcode" for metabarcoding applications (Schoch *et al.,* 2012).

2. Bacteria: 16S Ribosomal RNA Gene (16S rRNA)

a. Role in Prokaryotic Taxonomy and Phylogeny: The 16S rRNA gene is the foundational molecular marker in prokaryotic (bacteria and archaea) taxonomy. It is highly conserved across all prokaryotes, allowing for universal primer design, while its evolutionary rate is suitable for inferring phylogenetic relationships at the kingdom-to-phylum level (Woese, 1987).

b. Hypervariable Regions (V1-V9): For high-resolution species identification, especially in complex community analysis (metabarcoding), researchers target specific Hypervariable Regions(V1 through V9) within the 16S gene (Clarridge, 2004). These shorter, highly divergent regions are amplified and sequenced en masse to rapidly characterize microbial diversity.

3. Protists and Other Groups

Marker selection is highly varied for other groups due to their vast phylogenetic diversity (Parfrey *et al.,* 2010):

Protists/Microeukaryotes: The 18S rRNA gene (the nuclear small subunit rRNA) is the most widely used marker for phylogenetics and metabarcoding, offering a balance of conserved and variable regions.

Algae: The tufA (elongation factor Tu) gene is commonly used as a primary barcode for green algae (Chlorophyta), while the COI gene, which functions poorly in land plants, is effective in many red algae (Rhodophyta).

General Eukaryotes: The 28S rRNA (the nuclear large subunit rRNA), like 18S, is often employed, particularly for its ability to resolve relationships at deeper (higher) taxonomic levels.

III. Conservation Genomics and Phylogeographic Applications

DNA barcoding, extended into the realm of genomics, provides powerful tools that move beyond mere identification to address complex questions in ecology, evolution, and conservation.

A. Biodiversity Assessment and Ecological Monitoring

1. Metabarcoding

Metabarcoding combines the short, standardized DNA barcode sequences with Next-Generation Sequencing (NGS) platforms, allowing for the simultaneous identification of multiple species from a single environmental sample containing a mix of DNA (Taberlet *et al.,* 2018).

Process and Output: Instead of sequencing one specimen at a time, metabarcoding amplifies a barcode region (e.g., COI for animals, 16S rRNA for bacteria) from an entire bulk sample. The NGS output is millions of reads, which are then clustered into Operational Taxonomic Units (OTUs) or Barcode Index Numbers (BINs), providing a comprehensive species list.

Applications: This technique is revolutionizing large-scale biodiversity surveys, replacing time-consuming morphological identification of samples like:

Bulk Invertebrate Samples: Identifying all insect species collected in a Malaise trap in a fraction of the time (Hajibabaei *et al.,* 2007).

Soil and Water: Characterizing complex microbial, fungal, and micro-eukaryotic communities.

2. Environmental DNA (eDNA) Barcoding

Environmental DNA (eDNA) is genetic material shed by organisms into the environment (feces, skin, mucus, gametes) that can be filtered and sequenced. eDNA metabarcoding is a non-invasive method for species detection.

Non-Invasive Detection: It is particularly effective for rare, elusive, or highly mobile species, such as endangered fish in a river, invasive mussels in a harbor, or large mammals in terrestrial habitats (e.g., amphibian chytridiomycosis surveys) (Rees *et al.,* 2014).

Applications in Ecosystems:

Aquatic Systems: Water samples are filtered, and eDNA from fish, amphibians, or invertebrates is captured. This is far more sensitive and reliable than electrofishing or netting for monitoring species presence.

Terrestrial Systems: Soil and snow tracks are sampled to detect mammals; air filtering can even detect airborne pollen and fungal spores (Taberlet *et al.,* 2018).

3. Food Web Analysis and Diet Composition

Metabarcoding is used to analyze the contents of an organism's diet or the entirety of a food web.

Analysis of Trophic Interactions: DNA extracted from gut contents or fecal matter (or regurgitated pellets) can be sequenced to identify the prey items using COI primers specific to the prey's kingdom (e.g., prey arthropods in a predator's stomach) (Pompanon *et al.,* 2012). This provides unprecedented detail on predator-prey dynamics, essential for ecological modeling and pest management.

B. Applications in Forensics, Health, and Commerce

1. Wildlife Forensics

DNA barcoding is an essential tool in the fight against illegal wildlife trade, especially when samples are highly processed, fragmented, or lack morphological features.

Identifying Illicit Products: Barcoding is used to verify the species identity of products such as bushmeat, ivory, rhino horn, shark fins, or processed poached timber (using cpDNA markers) (Baker *et al.,* 2021). This allows law enforcement to prosecute violations of international agreements like CITES (Convention on International Trade in Endangered Species).

2. Food Authenticity and Safety

Barcoding provides a rapid and legally defensible method to verify food source and detect contaminants.

Detecting Food Fraud: Barcoding has revealed widespread seafood mislabelling (e.g., cheap fish substituted for expensive snapper or cod) and meat substitution in commercial products (Baker *et al.,* 2021). The ability to use minibarcodes is key here, as food processing often severely degrades the DNA.

Identifying Pathogens and Contamination: Metabarcoding can quickly screen food products for the presence of microbial pathogens or unwanted allergenic species (e.g., peanuts in nut-free products) (Baker *et al.,* 2021).

3. Authentication of Herbal and Medicinal Products

Traditional Chinese Medicine (TCM) and other herbal products often contain ground or highly processed plant material, making morphological identification impossible.

Ingredient Validation: Barcoding (ITS, matK, rbcL) validates that the correct species is present, which is crucial since many closely related species have varying medicinal properties (Chen *et al.,* 2010).

Adulterant Detection: It is used to detect the presence of toxic or banned adulterants or cheap plant fillers that have been added to the product (Baker *et al.,* 2021).

C. Phylogeography and Population Genomics

1. Phylogeography

Phylogeography is the study of the processes determining the geographic distributions of genetic lineages within and among species (Avise, 2000).

Mapping Genetic Lineages: Barcode data, coupled with broader phylogenetic markers (e.g., Cytb or tRNA genes), helps to reconstruct the evolutionary history of species and map historical migration routes, glacial refugia and past barriers to gene flow. This understanding of deep genetic history is critical for defining management units.

2. Conservation Genomics

Barcoding data forms the foundation for more detailed conservation genomics studies, providing data to inform management policy.

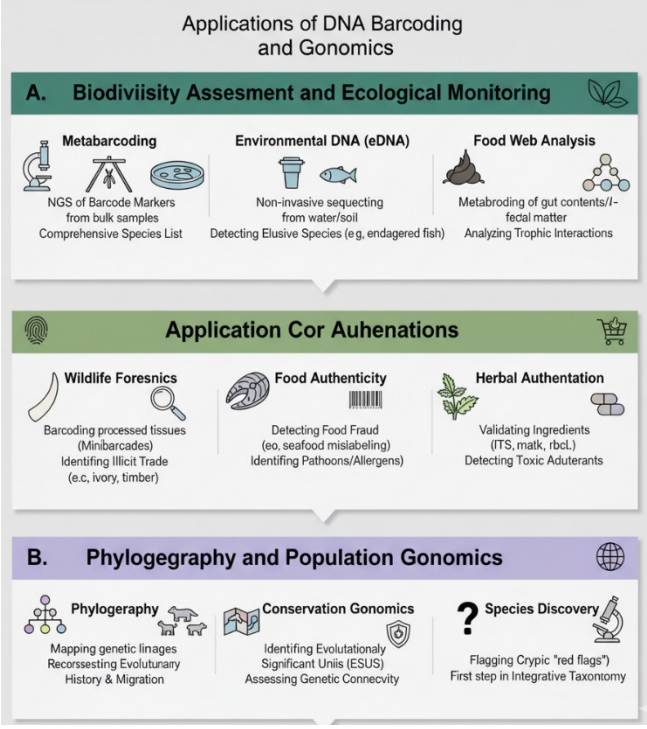
Identifying Evolutionarily Significant Units (ESUs): Barcode clusters that show deep, reciprocal monophyly (distinct lineages with no shared haplotypes) are often designated as ESUs. These units are recognized as highly distinct and may be on independent evolutionary trajectories, requiring separate and targeted conservation measures (Moritz, 1994).

Assessing Genetic Health: Barcode sequences can reveal low genetic diversity or population structure, which indicates isolated or fragmented populations. This data guides efforts to establish wildlife corridors or conduct translocations to restore genetic connectivity and health (Baker *et al.,* 2021).

3. Species Discovery and Delimitation

The most straightforward application of barcoding is the initial step in the discovery of new species.

Flagging Cryptic Species: Barcoding frequently reveals that what was thought to be one morphologically uniform species is actually a complex of several genetically distinct, cryptic species (Hebert *et al.,* 2004). Barcodes serve as "red flags," directing traditional taxonomists to re-examine the morphology and ecology of these distinct molecular clusters to formalize their description. This iterative process of molecular and morphological investigation is the core of integrative taxonomy.



IV. Limitations of DNA-Based Identification: Cryptic Diversity, Hybridization, and Degradation

While DNA barcoding is a revolutionary tool, it's not a panacea. Its limitations stem from inherent evolutionary processes, the constraints of using a limited set of genetic markers, and technical challenges in sample quality and data management.

A. Biological and Evolutionary Constraints

1. Cryptic Diversity

The Problem: Cryptic species are those that are reproductically isolated and genetically distinct but are morphologically indistinguishable (Bickford *et al.,* 2007). Barcoding is highly effective at flagging these cases because the genetic divergence often exceeds the typical barcode gap, even though traditional taxonomy failed to distinguish them (Hebert *et al.,* 2004). This leads to an initial overestimation of species richness by the barcode data.

Mitigation: The best solution is Integrative Taxonomy, which combines high genetic divergence from barcoding with rigorous re-examination of subtle morphological, ecological and bioacoustic data to formally describe the "new" species (Dayrat, 2005).

2. Intraspecific Variation and Species Complexes

In some taxa, particularly those with large effective population sizes and recent range expansions, the COI sequence variation within a recognized species can be unusually high, exceeding the interspecific threshold (Meyer & Paulay, 2005). This overlap of intra- and interspecific divergence—the blurring or absence of the barcode gap—can result in the current species being erroneously split into multiple BINs or OTUs, causing false splits and an over-estimation of diversity.

3. Hybridization and Introgression

The Problem: Hybridization (mating between species) followed by introgression (the movement of genes between species) can be a significant pitfall, especially when using mitochondrial or plastid markers. Since these markers are inherited as a single unit (haplotype), a single successful mating event followed by backcrossing can lead to the mitochondrial or plastid genome of one species spreading into the nuclear background of another (Tautz *et al.,* 2003). This gene flow homogenizes the barcode sequence, resulting in two truly distinct species having the same barcode. This causes false mergers and a severe under-estimation of species richness.

4. Mitochondrial Pseudogenes (NUMTs and KIPs)

Nuclear Mitochondrial Pseudogenes (NUMTs) are non-functional copies of mitochondrial DNA (mtDNA) that have been integrated into the nuclear genome. Similarly, Kinetoplastid Pseudogenes (KIPs) exist in some flagellates.

The Risk: Since NUMTs are nuclear, they lack the selective pressure applied to functional mtDNA and may accumulate random mutations. If universal mtDNA primers unintentionally amplify a NUMT

instead of the actual mitochondrial target, the resulting sequence will be an evolutionary fossil that is significantly different from all functional barcodes, leading to misidentification or an erroneous deep lineage split (Bensasson *et al.,* 2001).

B. Technical and Data Quality Limitations

1. DNA Degradation and Minibarcodes

Challenge: Samples exposed to heat, moisture, or chemical fixatives, such as old museum specimens, archaeological remains, or highly processed food and herbal products, contain DNA that is severely fragmented. The standard 650 bp barcode is often too long to be successfully amplified from these templates (Hajibabaei *et al.,* 2206).

Minibarcoding Approach: The solution is Minibarcoding, which uses a highly conserved, shorter fragment (e.g., 100−300 bp) of the standard barcode region. These fragments are more easily amplified from degraded templates, drastically increasing the success rate for difficult samples (Meusnier *et al.,* 2008).

2. Incomplete Reference Libraries

The ability to identify an unknown sample is entirely dependent on its match to a validated sequence in a public database like BOLD or GenBank.

The Taxonomic Impediment in the Digital Age: While millions of sequences exist, only a small fraction of the estimated ≈9 million eukaryotic species have a formally validated barcode (Costello *et al.,* 2013). For samples belonging to these unbarcoded species, the output will be a "No Match Found," which is a limitation of the database, not the technique. Filling this gap is a key ongoing global effort.

3. Data Quality Errors

The utility of public databases can be compromised by human error in the sequence submission process:

Contamination and Misidentification: A specimen that was morphologically misidentified by the collector, or a sequence contaminated during PCR, can be uploaded to a database. Subsequent users who rely on that erroneous sequence for identification may be led to a false or misleading result, perpetuating the error (Kress & Erickson, 2007).

C. Future Directions and Emerging Technologies

The future of DNA-based identification is moving toward larger genomic regions and massive scale processing:

1. High-Throughput Sequencing Platforms: The integration of NGS (e.g., Illumina or Oxford Nanopore) enables Metabarcoding, shifting the focus from identifying a single specimen to characterizing entire communities in a single, high-throughput assay (Taberlet *et al.,* 2012).

2. Genome-Skimming and Organellar Genomes: Instead of targeting short fragments, newer methods use shallow whole-genome sequencing (genome-skimming) to capture entire chloroplast and mitochondrial genomes (Organellar Genomes) (Nicholls *et al.,* 2015). These much longer sequences provide hundreds of potential gene regions, offering supra-barcoding resolution and better tools for resolving difficult species complexes and phylogeography.

3. Automation and Miniaturization: The entire barcoding workflow, from DNA extraction to library preparation, is being automated and miniaturized using microfluidics (Lab-on-a-chip), significantly reducing reagent costs, increasing throughput, and minimizing contamination risk.

Conclusion

DNA barcoding has unequivocally proven its worth as a standardized, rapid, and scalable platform for species identification, effectively addressing the "taxonomic impediment" by providing a reliable molecular tag for life. The chapter detailed the critical molecular markers—the COI gene as the universal standard for animals, the rbcL + matK combination for land plants, and the ITS region for fungi—highlighting the necessity of marker-switching across kingdoms to balance universality with discriminatory power.

The technology's reach extends far beyond pure taxonomy, becoming a foundational tool in conservation genomics and ecological monitoring. Techniques like metabarcoding and Environmental DNA (eDNA) analysis now allow for the high-throughput assessment of entire communities from bulk or non-invasive samples, revolutionizing food web analysis, invasive species monitoring, and biodiversity inventories. Furthermore, barcoding underpins critical applied fields, including wildlife

forensics (combating illegal trade) and food safety (detecting fraud and contamination).

However, the molecular approach is not without its challenges. Evolutionary phenomena such as hybridization and introgression threaten to homogenize barcode sequences, leading to the false merger of distinct species, while high intraspecific variation can cause false splits. The technical hurdles of DNA degradation in old samples and the current incompleteness of reference libraries also impose real limits on identification success.

Ultimately, DNA barcoding is best viewed not as a replacement for traditional methods, but as a crucial integrative tool. Its strength lies in its ability to flag cryptic diversity and deep evolutionary units (ESUs), providing the necessary genetic data that, when combined with morphological and ecological expertise, facilitates robust species delimitation and informed conservation policy. As sequencing technologies continue to advance toward high-throughput genome-skimming and full organellar sequencing, DNA barcoding is poised to become an even more powerful, precise, and automated pillar of biodiversity science.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410.

Avise, J. C. (2000). Phylogeography: The History and Formation of Species. Harvard University Press.

Baker, C. S., Hanner, R. H., & Steinke, D. (2021). Molecular identification of seafood: DNA barcoding and the seafood supply chain. *Trends in Food Science & Technology*, *108*, 337-347.

Bensasson, D., Zhang, D. X., Hartl, D. L., & Hewitt, G. M. (2001). Mitochondrial pseudogenes: evolution and implications for phylogenetics. *Trends in Ecology & Evolution*, *16*(6), 314-321.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., ... & Posa, M. R. C. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*(3), 148-155.

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., ... & Vanfleteren, J. R. (2005). The *Nematoda*: molecular insights into long-standing questions. *Invertebrate Biology*, *124*(1), 1-13.

Chase, M. W., Salamin, N., Wilkinson, M., Pridgeon, A. M., Källersjö, P., & Rudall, P. J. (2021). Molecular systematics of Lilianae: a re-examination of the order Asparagales and its subordinal structure. In *Monocots: Comparative Biology and Evolution (excluding Poales)* (pp. 109-127). CSIRO Publishing.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., ... & Zhu, Y. (2010). Validation of the ITS2 region as a novel DNA barcode for identification of medicinal plants. *PLoS One*, *5*(1), e8613.

Clarridge, J. E. (2004). Impact of 16S rRNA gene sequencing on clinical microbiology and infectious diseases. *Clinical Microbiology Reviews*, *17*(4), 840-862.

Costello, M. J., Wilson, S. P., & Houlding, B. (2013). Predicting total global species richness using a known to unknown ratio. *Global Ecology and Biogeography*, *22*(7), 868-877.

Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society*, *85*(3), 407-415.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*(1), 11-15.

Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, *8*(3), 186-194.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294-299.

Fukatsu, T. (2009). Biological laboratory methods (BLM)-95% ethanol as a better molecular sample preservative than 70% ethanol. *Entomological Science*, *12*(2), 177-180.

Funk, D. J., & Omland, K. E. (2003). Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, *34*, 397-423.

Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W., & Hebert, P. D. N. (2007). DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences*, *104*(11), 4791-4796.

Hajibabaei, M., Smith, M. A., Janzen, D. H., & Hebert, P. D. N. (2006). A minimalist barcode can identify a specimen accurately. *Molecular Ecology Notes*, *6*(4), 1159-1164.

Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds through DNA barcodes. *PLoS Biology*, *2*(10), e312.

Hebert, P. D., Cywinska, A., Shelley, L. B., & Jeremy, R. D. (2003a). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1525), 313-321.

Hebert, P. D., Ratnasingham, S., & Jeremy, R. D. (2003b). Towards a global system for identifying life. *Genome Biology*, *4*(6), 237.

Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and using a plant DNA barcode. *PloS One*, *6*(5), e19254.

Kress, W. J., & Erickson, D. L. (2007). A two-locus global DNA barcode for land plants: the coding genes rbcL and matK. *PLoS One*, *2*(6), e508.

Meusnier, I., Singer, G. A., Laframboise, B. M., Softic, S., Hebert, P. D., & Hajibabaei, M. (2008). A universal

mini-barcode for animal species identification. *BMC Genomics*, *9*(1), 1-9.

Meyer, C. P., & Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, *3*(12), e422.

Moritz, C. (1994). Defining evolutionarily significant units for conservation. *Trends in Ecology & Evolution*, *9*(10), 373-375.

Nicholls, J. A., Hatherell, P., Gurney, M., & Russell, S. (2015). Genome skimming: an approach for plant conservation genomics. *Methods in Ecology and Evolution*, *6*(5), 552-560.

Padial, J. M., Miralles, A., De la Riva, I., & Vences, M. (2010). The integrative future of taxonomy. *Frontiers in Zoology*, *7*(1), 1-14.

Parfrey, L. W., Barok, F., Latz, E., Patterson, D. J., & Katz, L. A. (2010). Broadly applicable high-throughput sequencing reveals a high diversity of unknown protists in aquatic environments. *PLoS One*, *5*(3), e9481.

Pompanon, F., Deagle, B. E., Symondson, A. J. C., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Molecular ecology: metabarcoding for diet analysis. *Molecular Ecology*, *21*(9), 1931-1934.

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, *7*(3), 355-364.

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patchett, R. J., & Gough, K. C. (2014). The application of eDNA for monitoring aquatic organisms. *Journal of Applied Ecology*, *51*(5), 1450-1461.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406-425.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: a laboratory manual*. Cold Spring Harbor laboratory press.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463-5467.

Särkinen, T., Staats, M., Richardson, J. E., & Cowan, R. S. (2012). DNA extraction from plant material. In *DNA barcodes: methods and protocols* (pp. 53-74). Humana Press.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, R. A., ... & Wang, X. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proceedings of the National Academy of Sciences*, *109*(16), 6241-6246.

Shearer, T. L., Gittings, P., & Coffroth, M. A. (2002). Molecular markers for Cnidarian identification. *Journal of Experimental Marine Biology and Ecology*, *281*(1-2), 163-171.

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press.

Tautz, D., Arctander, P., Minelli, A., Thomas, R. H., & Vogler, A. P. (2003). A plea for DNA taxonomy. *Trends in Ecology & Evolution*, *18*(2), 70-74.

The CBOL Plant Working Group. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, *106*(31), 12794-12797.

Wang, Y., Tang, Y., Wu, B., Li, X., Wu, X., & Liu, T. (2024). Species Delimitation and Cryptic Diversity in *Rheotanytarsus* Thienemann & Bause, 1913 (Diptera: Chironomidae) Based on DNA Barcoding. *Insects*, *15*(4), 370.

Woese, C. R. (1987). Bacterial evolution. *Microbiological Reviews*, *51*(2), 221-271.