Indian Farmer
Volume 13, Issue 02, 2026, Pp. 31-37
Available online at: www.indianfarmer.net
ISSN: 2394-1227 (Online)

Popular Article

DNA Barcoding in the Age of Bioinformatics: How Algorithms Identify Life

Anupama Roy[1,2], Parmila Bhukal[1], Sarika Jaiswal[1] and Mir Asif Iquebal[1]

[1]Division of Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012, India

[2]The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi-110012, India

**Corresponding author: roy.anupama11@gmail.com*

Received: 29/01/2026 *Published:04/02/2026*

ABSTRACT

DNA barcoding and its association with bioinformatics and data analysis have evolved to become a robust tool for species identification. This technique enables species identification using a short DNA marker using bioinformatics approaches. This concept has revolutionized species-identifying taxonomy and morphology and advanced by the usage of data and computing analysis. The article examines DNA barcoding and its utilization with bioinformatics and its analysis and gives a detailed explanation of DNA-raw data analysis and its various steps executed for species identification. Some popular and primary software and databases included in DNA barcoding and their usage in bioinformatics analysis systems are described and highlighted, along with biodiversity analysis and its association with food and human health analysis. The article finally examines primary and advanced bioinformatics concerns and its limitations and hindrances to data and DNA analysis.

keywords: DNA barcoding; Bioinformatics; Species identification; Biodiversity; Computational biology

INTRODUCTION

A leaf portion stuck on the shoe, a sliver of fish from a market, or a drop of water from a stream may look biologically insignificant to common man. But inside each of these samples lies DNA information that is transformed into digital data. In modern biology, identifying a species no longer depends only on the observations made by the naked eye or known by an expert in that specific field. Instead, it often begins in the laboratory and ends on a computer screen, calling it as in-silico analysis.

The scientist extracts the DNA from the sample and reads a short section (or portion) of the DNA to convert it to the nucleotide sequence (A, T, G, and C); thus, the segment becomes a digital representation of the original organism. The data can be stored, shared, and analysed by using computer-based tools, allowing users to manipulate the digital data as required. A piece of biological material that was believed to have very little value is now a digital "fingerprint" of that organism which can be further used in a search across large online databases.

Through DNA barcoding, we can link the living world with our digital world and demonstrate how we can use data to explore and analyse biodiversity (Figure 1). An example of a method used for identifying species is by a short segment of DNA. Rather than identifying a specimen by how it looks, behaves, or grows in nature, it is applied using the organism's genetic material as a unique identifier; this is done similarly to how we identify products in a store using the barcoding system (Hajibabaei *et al.* 2007).

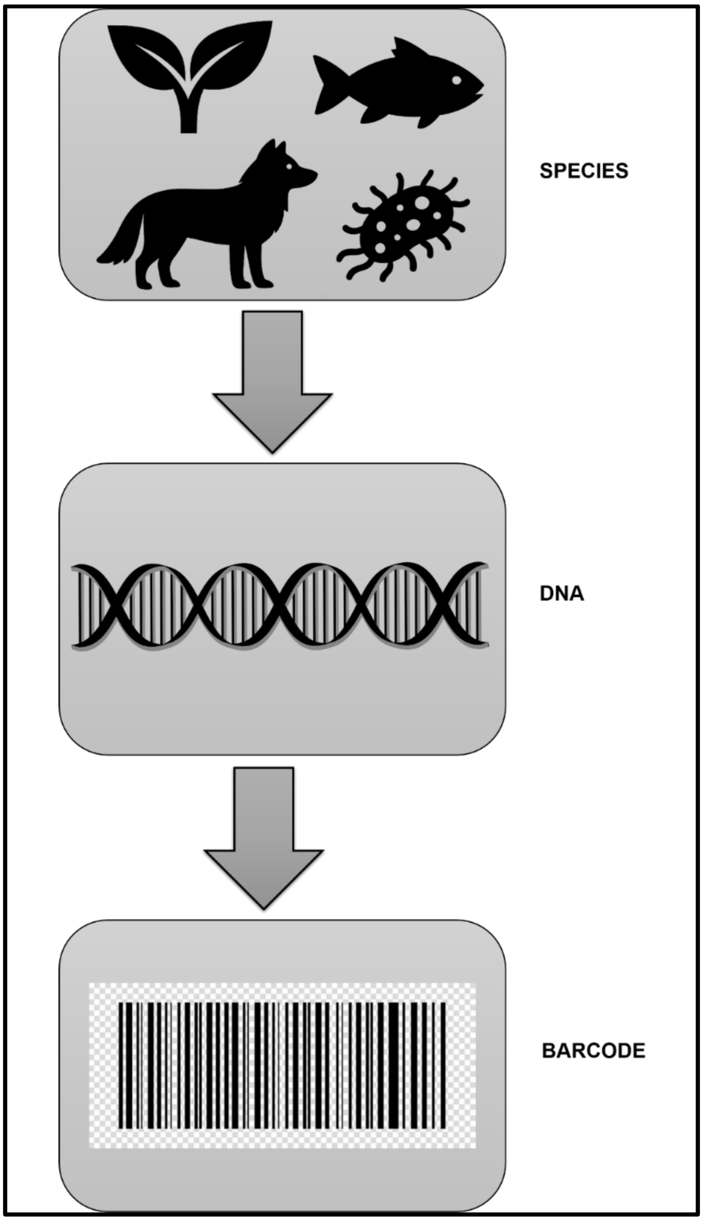


Figure 1. Conceptual illustration of DNA barcoding, showing how diverse biological samples are converted into a standardized DNA barcode for species identification.

One of the key aspects of DNA barcoding is its standardisation. For each major group of organisms, scientists have standardised a short DNA region, which frequently differs among species, but is usually similar among individuals of the same species. Hence, if DNA from two different organisms is analysed with respect to the same DNA region, we can compare those data, regardless of when or where it was collected, as well as the quality and quantity of the DNA.

DNA barcoding demonstrates how biological material can be transformed into digital data by generating a standardized DNA sequence from an individual organism sample and analysing it using computational tools (Kress *et al.*, 2008). Bioinformatics methods compare these sequences with large reference databases, assess genetic similarity, and assign the most likely species identity. This approach shifts species identification from reliance on visual observation to the use of sequence data, algorithms, and databases, allowing biodiversity to be studied as a data-driven system (Ali *et al.*, 2014).

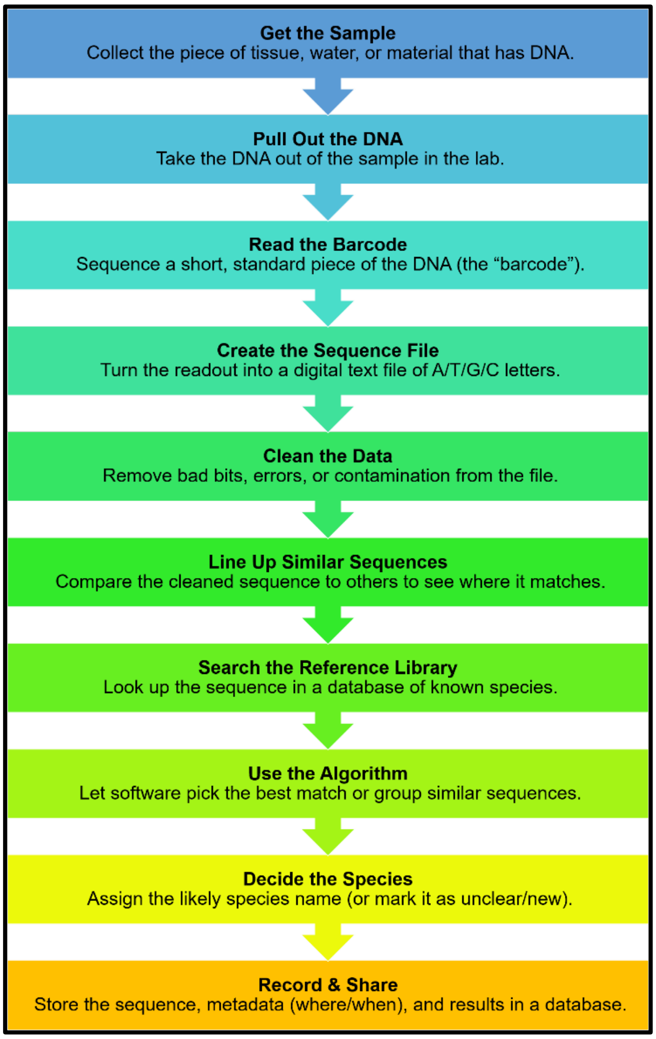
As DNA barcoding projects continue to expand, they generate vast numbers of sequences from diverse species, locations, and time periods. Managing and analysing such large datasets would be impractical without the bioinformatics support. Computational tools are used to clean raw sequence data, then detect errors, align sequences, and lastly compare them across thousands or even millions of records. These sequences, along with associated metadata, are stored in databases that enable efficient retrieval, analysis, and long-term study.

Bioinformatics also enhances the accuracy and scalability of species identification. Automated algorithms apply consistent criteria during sequence comparison, reducing subjectivity and human error. At the same time, computational pipelines allow rapid processing of large datasets, enabling high-throughput identification. Together, these advances have transformed DNA barcoding into a scalable and reliable system for identifying life on a global scale (Mahima *et al.*, 2022).

BIOINFORMATICS WORKFLOW

Species identification through DNA barcoding follows a structured bioinformatics workflow that transforms raw DNA sequence data into a species name. The process begins with sequencing a standard barcode region, generating a digital nucleotide sequence along with quality information. Bioinformatics tools then clean the data by trimming low-quality regions, correcting errors, and removing contaminants, ensuring that only reliable sequences are retained for analysis.

The curated sequence is aligned with reference barcode sequences to identify similarities and differences. Algorithms assess genetic similarity using distance measures or scoring methods and compare the unknown sequence against reference databases that serve as digital libraries of known species. Based on predefined criteria or statistical thresholds, a species identity is assigned, or the sequence may be classified as ambiguous or potentially novel. As illustrated in Figure 2, this workflow converts a short DNA fragment into a meaningful biological identity.



Get the Sample
Collect the piece of tissue, water, or material that has DNA.

Pull Out the DNA
Take the DNA out of the sample in the lab.

Read the Barcode
Sequence a short, standard piece of the DNA (the “barcode”).

Create the Sequence File
Turn the readout into a digital text file of A/T/G/C letters.

Clean the Data
Remove bad bits, errors, or contamination from the file.

Line Up Similar Sequences
Compare the cleaned sequence to others to see where it matches.

Search the Reference Library
Look up the sequence in a database of known species.

Use the Algorithm
Let software pick the best match or group similar sequences.

Decide the Species
Assign the likely species name (or mark it as unclear/new).

Record & Share
Store the sequence, metadata (where/when), and results in a database.

Figure 2. Stepwise bioinformatics workflow of DNA barcoding, from sample collection and DNA extraction to sequence analysis, database comparison, and final species identification.

ALGORITHMS AND DATABASES

DNA barcoding relies on sequence comparison methods and well-curated databases to link genetic data to species identities. Most identification algorithms compare an unknown barcode sequence with reference sequences to find the closest genetic match. Species assignment is based on measures of genetic similarity or divergence, often using predefined distance thresholds. Phylogenetic analysis provides an additional approach by placing barcode sequences within an evolutionary framework, helping resolve identification challenges among closely related species (Dasgupta *et al.*, 2005). To handle large and complex datasets, newer approaches such as machine learning and clustering methods have been developed to automate barcode classification.

All these methods depend on reliable reference databases that store barcode sequences along with their associated species information. Major repositories such as the Barcode of Life Data Systems, GenBank, and the European Nucleotide Archive serve as foundational resources for species identification using DNA barcodes (Ratnasingham *et al.*, 2024). Accurate matching also requires the integration of sequence data with taxonomic and geographic metadata, allowing more robust and context-aware comparisons. Together, sequence analysis algorithms and reference databases form an integrated system that enables precise and reliable species identification from DNA barcode data.

APPLICATIONS

The advancement of bioinformatics has expanded the scope of DNA barcoding far beyond basic species identification. Today, it represents a powerful integration of molecular biology and computational science that supports biodiversity research, ecosystem conservation, habitat monitoring, resource management, and public health protection (Ratnasingham *et al.*, 2024). By combining sequencing data with algorithms and large reference databases, bioinformatics enables molecular techniques to be applied at scales that were previously impossible, opening new opportunities in biodiversity, ecology, and health research.

Researchers can now analyze thousands of DNA barcodes simultaneously to monitor biodiversity and discover new species. Large datasets collected over time allow scientists to detect patterns in species distribution, track changes in biodiversity, and identify rare or endangered species across broad geographic regions. These insights are essential for understanding ecological change and supporting long-term conservation efforts (Lahaye *et al.*, 2008).

Environmental DNA (eDNA) analysis provides a non-invasive alternative to traditional sampling by detecting traces of DNA in water, soil, or air. These mixed samples generate large and complex datasets that require careful computational processing. Bioinformatics pipelines are used to remove noise, organize sequences, and compare them against reference databases, enabling researchers to reconstruct entire biological communities from environmental samples. This approach has become a valuable tool for monitoring ecosystems without disturbing natural habitats.

In food authentication and quality control, bioinformatics-supported DNA barcoding plays a crucial role in verifying species identity in processed products where visual inspection is not possible. By comparing barcode sequences from food samples with reference datasets, algorithms can detect mislabeling, substitution, or contamination, thereby supporting regulatory compliance, consumer protection, and transparency in the food supply chain (Antil *et al.*, 2023).

Public health is another important area where DNA barcoding and bioinformatics intersect, particularly in the identification of disease vectors such as insects or parasites. Rapid and accurate identification based on sequence data helps track vector distribution, detect outbreaks, and support effective control strategies. Automation of these analyses reduces reliance on specialized expertise, allowing DNA barcoding to function as a scalable and adaptable tool for health surveillance.

Across all these applications, bioinformatics is transforming DNA barcoding into a high-impact technology capable of handling the complexity, scale, and speed that traditional identification methods often struggle to achieve.

CHALLENGES

Despite its wide use, DNA barcoding supported by bioinformatics faces several important challenges. Many of these issues arise not from the barcoding technique itself, but from limitations in reference databases, including their quality, coverage, and interpretation (Mahima *et al.*, 2022). Incomplete or biased databases remain a major concern, as barcodes are unevenly distributed across taxa and geographic regions, which can reduce the accuracy of species identification.

Another key challenge is sequence error and contamination introduced during sample preparation or sequencing. Such errors can obscure true biological signals and lead to incorrect identifications, making strict quality control, filtering, and validation steps essential for reliable analysis.

Defining species boundaries also remains difficult, particularly for closely related species with very similar barcode sequences or for species showing high genetic variation across populations. Fixed similarity thresholds used by algorithms do not always reflect true biological relationships, leading to uncertainty in classification (Kundu *et al.*, 2020).

Finally, the rapid growth of DNA barcoding has created increasing computational and data management demands. Handling large volumes of sequencing data requires robust storage systems, efficient processing pipelines, and continuously updated databases, all supported by strong computational infrastructure.

CONCLUSION

DNA barcoding offers a clear example of how biology is becoming increasingly data-centric. A short stretch of DNA, once meaningful only in a laboratory context, is transformed into digital information that can be stored, analysed, and shared globally. Species identification no longer depends solely on direct observation or individual expertise, but on the ability to compare data across large, interconnected datasets. In this way, DNA barcoding serves as a model for modern biology, where insight emerges from the integration of molecular data and computational analysis.

At the heart of this transformation is bioinformatics, which acts as the bridge between molecules and biodiversity. Without bioinformatics, barcode sequences would remain isolated strings of letters; with it, they become tools for understanding life at scale. As biodiversity research continues to generate ever larger and more complex datasets, DNA barcoding demonstrates how computational approaches can connect the smallest units of life to the broad patterns that shape the living world.

REFERENCES

Hajibabaei, M., Singer, G. A., Hebert, P. D., & Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. TRENDS in Genetics, 23(4), 167-172.

Kress, W. J., & Erickson, D. L. (2008). DNA barcodes: genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences*, *105*(8), 2761-2762.

Ali, M. A., Gyulai, G., Hidvegi, N., Kerti, B., Al Hemaid, F. M., Pandey, A. K., & Lee, J. (2014). The changing epitome of species identification–DNA barcoding. *Saudi journal of biological sciences*, *21*(3), 204-231.

Mahima, K., Sunil Kumar, K. N., Rakhesh, K. V., Rajeswaran, P. S., Sharma, A., & Sathishkumar, R. (2022). Advancements and future prospective of DNA barcodes in the herbal drug industry. *Frontiers in pharmacology*, *13*, 947512.

DasGupta, B., Konwar, K. M., Măndoiu, I. I., & Shvartsman, A. A. (2005). DNA-BAR: distinguisher selection for DNA barcoding. *Bioinformatics*, *21*(16), 3424-3426.

Ratnasingham, S., Wei, C., Chan, D., Agda, J., Agda, J., Ballesteros-Mejia, L., ... & Hebert, P. D. (2024). BOLD v4: A centralized bioinformatics platform for DNA-based biodiversity data. In *DNA barcoding: Methods and protocols* (pp. 403-441). New York, NY: Springer US.

Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., ... & Savolainen, V. (2008). DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences*, *105*(8), 2923-2928.

Kundu, S., Lalremsanga, H. T., Rahman, M. M., Ahsan, M. F., Biakzuala, L., Kumar, V., ... & Siddiki, A. Z. (2020). DNA barcoding elucidates the population genetic diversity of venomous cobra species (Reptilia: Elapidae) in Indo-Bangladesh region. *Mitochondrial DNA Part B*, *5*(3), 2525-2530.

Antil, S., Abraham, J. S., Sripoorna, S., Maurya, S., Dagar, J., Makhija, S., ... & Toteja, R. (2023). DNA barcoding, an effective tool for species identification: a review. *Molecular biology reports*, *50*(1), 761-775.