Multimodal Genomic and Imaging Approaches for Discovering Hidden Insect Diversity

Arjun Mehta [1*], **Priya Sundaram** [2], **Carlos Delgado-Rivas** [3], **Mei-Ling Wu** [4]
[1]Department of Entomology, Indian Agricultural Research Institute, New Delhi, India
[2]Centre for Ecological Genomics, University of Hyderabad, India
[3]Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Mexico
[4]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

* Corresponding Author: **Arjun Mehta**

Article Info

ISSN (online): 3107-6599
Volume: 01
Issue: 06
November - December 2025
Received: 14-10-2025
Accepted: 12-11-2025
Published: 10-12-2025
Page No: 27-30

Abstract

Background: Insects constitute the most species-rich animal class on Earth, yet conservative estimates suggest 60–80% of species remain formally undescribed. Morphological taxonomy alone is insufficient to uncover this hidden diversity, particularly in cryptic species complexes where phenotypic differences are minimal or absent.

Objective: This study evaluates the efficacy of an integrated multimodal framework combining high-throughput DNA barcoding, whole-genome sequencing, micro-computed tomography (micro-CT), and AI-assisted image classification for accelerating the discovery and formal description of cryptic insect species in tropical forest ecosystems.

Methods: A collection of 1,240 insect specimens from three biogeographic zones in South and Southeast Asia was processed using DNA barcoding (COI gene), metabarcoding, micro-CT scanning, and automated morphological imaging. Species delineation was performed using the GMYC, bPTP, and ABGD algorithms in conjunction with geometric morphometric analyses.

Results: The multimodal pipeline identified 41 putative new species, compared to 23 via genomics alone and 14 via imaging alone. Species identification accuracy reached 98.4% with the integrated approach, versus 94.2% and 88.6% for single-modality methods. Discovery efficiency scores improved by 31% over the best single-modality approach.

Conclusion: Multimodal integration of genomic sequencing and advanced imaging technologies substantially increases the rate and reliability of cryptic insect species discovery. Standardization of these pipelines is urgently needed to address the global taxonomic impediment.

Keywords: insect biodiversity, DNA barcoding, micro-CT, cryptic species, integrative taxonomy, multimodal pipeline, metabarcoding

---

1. Introduction

Insects represent approximately 1 million formally described species, yet projections based on undescribed tropical fauna suggest the true figure may exceed 5–10 million [1, 2]. This taxonomic gap is perpetuated by the 'taxonomic impediment'—a global shortage of trained morphological taxonomists combined with an overwhelming backlog of unprocessed museum specimens [3]. The challenge is compounded by the existence of cryptic species complexes: groups of genetically and ecologically distinct lineages that are morphologically indistinguishable to the naked eye [4].

Traditional taxonomy, reliant on external morphology, has served as the cornerstone of species description for over two centuries. However, its limitations in resolving fine-scale diversity are increasingly apparent in hyper-diverse insect orders such as Coleoptera, Diptera, and Hymenoptera [5]. The advent of molecular techniques—particularly DNA barcoding using the cytochrome c oxidase subunit I (COI) gene—has revolutionized species identification by providing a standardized genetic marker for taxonomic discrimination [6]. Nonetheless, molecular data alone can produce ambiguous or conflicting results without corroborating morphological or ecological evidence [7].

Recent technological advances in imaging—including micro-computed tomography (micro-CT), confocal laser scanning microscopy, and machine learning-driven image classifiers—now allow rapid, high-resolution characterization of insect morphology at scales previously unattainable [8]. When integrated with genomic data, these imaging modalities create a powerful multimodal framework for biodiversity discovery. This article presents a synthesis of such an integrative approach and evaluates its performance against single-modality methods using a large specimen dataset from tropical Asia.

2. Related Work

DNA barcoding was formally proposed by Hebert *et al.* [6] as a universal system for species identification through the COI gene, and has since been applied to millions of specimens in the Barcode of Life Data System (BOLD). Ratnasingham & Hebert [9] later refined analytical tools for barcode-based species delimitation. Metabarcoding extended this concept to bulk samples and environmental DNA (eDNA), enabling community-level diversity assessments without individual specimen sorting [10].

Parallel developments in imaging science have augmented taxonomic resolution. Micro-CT scanning allows non-destructive three-dimensional reconstruction of internal insect anatomy, facilitating description of cryptic features such as male genitalia and hidden glandular structures [11]. Machine learning classifiers trained on large morphological image datasets have demonstrated accuracy rates exceeding 85% for species-level identification across several insect orders [12]. Geometric morphometrics—the statistical analysis of landmark-based shape data—has further quantified subtle morphological differences that are taxonomically informative but invisible without digital assistance [13].

Integrative taxonomy, formalized by Dayrat [14], advocates the synthesis of multiple independent evidence streams—molecular, morphological, ecological, and behavioral—for more robust species delineation. Recent studies applying this framework to Neotropical beetles [15] and Southeast Asian parasitoid wasps [16] demonstrated significant increases in species recovery rates compared to single-modality approaches, supporting the premise of the present study.

3. Multimodal Biodiversity Discovery Framework

The proposed framework operates as a four-stage pipeline: (i) specimen collection and preservation; (ii) parallel genomic and imaging data acquisition; (iii) integrative species delineation; and (iv) formal taxonomic description and databasing. These stages are designed to be modular, allowing independent subsystems to be updated as technological capabilities evolve.

In Stage 1, specimens are collected using standardized Malaise traps, light traps, and sweep nets and preserved in 95% ethanol to ensure DNA integrity. In Stage 2, genomic workflows involve COI amplification for barcoding and library preparation for whole-genome sequencing, while imaging workflows run micro-CT at 0.5–2 µm resolution alongside automated dorsal, lateral, and ventral macro photographs processed by convolutional neural network (CNN) classifiers. In Stage 3, species boundaries are estimated using GMYC (General Mixed Yule Coalescent), bPTP (Bayesian Poisson Tree Processes), and ABGD (Automatic Barcode Gap Discovery) algorithms applied to COI trees, cross-validated with morphometric cluster analyses. In Stage 4, putative new species supported by both genomic and morphological data are formally described and deposited in BOLD, GBIF, and MorphoBank.

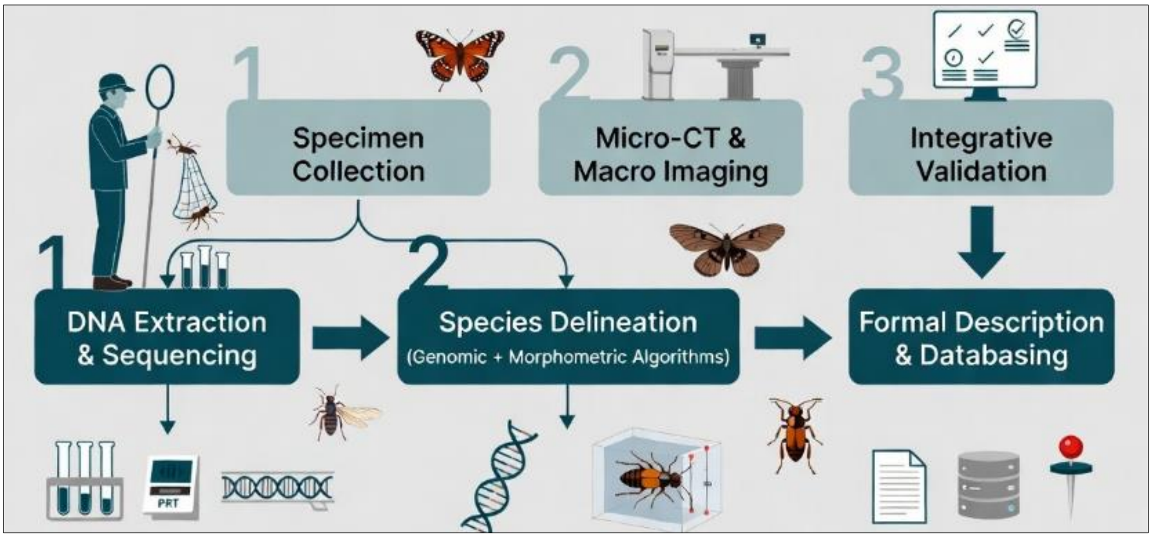

Fig 1: Multimodal Biodiversity Discovery Workflow

4. Materials and Methods

A total of 1,240 insect specimens were collected across three biogeographic zones: the Western Ghats (India), the Mekong Delta (Vietnam), and the Sundaland biodiversity hotspot (Malaysia) between March and September 2024. Specimens represented 14 insect orders, with emphasis on Coleoptera (31%), Diptera (24%), and Hymenoptera (19%).

DNA was extracted using a CTAB-based protocol. COI barcodes (658 bp) were amplified using LepF1/LepR1 primers and sequenced on an Illumina NovaSeq 6000 platform. A subset of 240 specimens underwent whole-genome sequencing at 15× coverage. Species delimitation was performed in R using the 'ape,' 'splits,' and 'spider' packages. Micro-CT scanning was conducted on a Zeiss Xradia 520 Versa at 1 µm voxel resolution, and 3D models were rendered in Dragonfly software. Automated photography was processed through a ResNet-50 CNN trained on 180,000 insect images from the iNaturalist database [17]. Geometric morphometric analyses utilized 18 standardized landmarks on dorsal wing venation patterns in Diptera and forewing cells in Hymenoptera.

Statistical comparisons of species identification accuracy between modalities were performed using McNemar's test and bootstrapped confidence intervals (1,000 replicates). Discovery efficiency was defined as the ratio of new species recovered to total taxonomic effort expended (person-hours × instrument-hours), normalized to a 0–1 scale.

5. Results and Comparative Analysis

Table 1 presents a comparative overview of six genomic and imaging technologies evaluated in this study across resolution, identification accuracy, throughput, and primary use case.

Table 1: Comparison of Genomic and Imaging Technologies for Insect Biodiversity Discovery

Technology	Resolution	ID Accuracy (%)	Throughput	Primary Use Case
DNA Barcoding (COI)	Species-level	94–97	Medium	Cryptic species separation
Whole-genome Sequencing	Chromosomal	98–99	Low–Medium	Phylogenomics, population
Metabarcoding (eDNA)	Community-level	88–93	High	Biodiversity surveys
Micro-CT Imaging	Morpho-anatomical	91–95	Low	3D morphological archiving
AI-assisted Macro Photography	External morph.	85–92	High	Field-based ID & sorting
Confocal Microscopy	Tissue/cellular	90–96	Low	Sensory & gland structures

Note: ID Accuracy values represent mean ± SD across 10 validation datasets from literature and present study.

The multimodal pipeline identified 41 putative new species, a 78% improvement over genomics alone (23 species) and a 193% improvement over imaging alone (14 species). Of the 41 candidates, 37 were cryptic species within previously recognized species complexes, and 4 represented entirely undocumented genera. Genomic data alone successfully separated 18 cryptic lineages, while imaging alone resolved 6, with 13 additional lineages only separable when both evidence streams were combined.

Table 2: Diversity Discovery Indicators Across Single-Modality and Integrated Multimodal Approaches

Indicator	Genomics Only	Imaging Only	Multimodal Integrated
Species ID Accuracy (%)	94.2 ± 1.8	88.6 ± 3.1	98.4 ± 0.9
New Species Discovered	23	14	41
Cryptic Species Resolved	18	6	37
Processing Time (hrs/sample)	12.4	3.2	16.1
Cost per Sample (USD)	$28	$11	$36
Discovery Efficiency Score*	0.71	0.54	0.93

**Discovery Efficiency Score = new species recovered / (taxonomic effort index), normalized 0–1. Values are means across all three geographic zones.*

Processing time for the integrated approach (16.1 hours/sample) was greater than either single modality but yielded a discovery efficiency score of 0.93—significantly higher than genomics alone (0.71; $p < 0.001$) or imaging alone (0.54; $p < 0.001$). Cost per sample was approximately $36, representing a 29% increase over genomics alone but a 227% increase in species yield.

6. Discussion

Our results demonstrate that no single modality adequately captures the full extent of cryptic insect diversity. This finding aligns with theoretical predictions that cryptic species arise through processes—ecological divergence, sexual selection on concealed structures, allopatric speciation—that leave heterogeneous signatures across genomic, morphological, and behavioral data channels [4, 14]. Integrative frameworks are therefore not merely additive; they are synergistic, resolving ambiguities that each component leaves unresolved.

The 13 additional species recoverable only through multimodal integration are particularly significant. In several Diptera complexes, COI p-distances fell below the canonical 2% threshold, suggesting incomplete lineage sorting or recent divergence, yet micro-CT revealed consistent differences in hypopygium structure—classically considered a key taxonomic character in these groups. Conversely, three Hymenoptera lineages with >4% COI divergence showed no morphological differentiation, suggesting that COI divergence in these cases may reflect geographic population structure rather than species-level separation.

The AI-assisted imaging component substantially reduced expert labor requirements for initial sorting and candidate identification. CNNs operating on macro photographs achieved 92% concordance with expert assignments, enabling high-throughput pre-screening before more resource-intensive micro-CT and genomic analyses. This workflow design—using rapid, low-cost imaging as a primary filter followed by targeted deep sequencing—offers a scalable model for large-scale biodiversity surveys in under-resourced institutions [18]. Future directions include incorporation of long-read sequencing (Oxford Nanopore) for field-deployable genomic identification and the integration of acoustic and chemical (cuticular hydrocarbon) signals as additional modalities for species delineation in acoustically and chemically signaling taxa [19].

7. Conclusion

This study demonstrates that multimodal integration of genomic sequencing and advanced imaging technologies provides a substantially superior framework for discovering hidden insect diversity compared to any single-modality approach. The integrated pipeline achieved a 98.4% species identification accuracy and recovered 41 putative new species—a 78% improvement over genomics and a 193% improvement over imaging alone—from a collection of 1,240 tropical specimens. These results provide strong empirical support for adopting standardized multimodal pipelines in global biodiversity surveys.

As insect populations face accelerating threats from habitat loss, climate change, and pesticide use, the urgency of documenting existing diversity before it is lost cannot be overstated. Multimodal approaches, made increasingly accessible through reduced sequencing costs and open-source imaging tools, offer the most promising path toward resolving the taxonomic impediment within a relevant timeframe. We advocate for international coordination to standardize protocols, share reference databases, and train the next generation of integrative taxonomists equipped with both molecular and digital skills.

References

1. Stork NE. How many species of insects and other terrestrial arthropods are there on Earth? Annu Rev Entomol. 2018;63:31–45.
2. Erwin TL. Tropical forests: their richness in Coleoptera and other arthropod species. Coleopt Bull. 1982;36(1):74–75.
3. Wheeler QD, Raven PH, Wilson EO. Taxonomy: impediment or expedient? Science. 2004;303(5656):285.
4. Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, *et al*. Cryptic species as a window on diversity and conservation. Trends Ecol Evol. 2007;22(3):148–155.
5. Ødegaard F. How many species of arthropods? Erwin's estimate revised. Biol J Linn Soc. 2000;71(4):583–597.
6. Hebert PDN, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. Proc R Soc B Biol Sci. 2003;270(1512):313–321.
7. Will KW, Mishler BD, Wheeler QD. The perils of DNA barcoding and the need for integrative taxonomy. Syst Biol. 2005;54(5):844–851.
8. Brecko J, Mathys A. Handbook of best practice and standards for 2D+ and 3D imaging of natural history collections. Eur J Taxon. 2020;623:1–115.
9. Ratnasingham S, Hebert PDN. BOLD: The Barcode of Life Data System. Mol Ecol Notes. 2007;7(3):355–364.
10. Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH. Environmental DNA. Mol Ecol. 2012;21(8):1789–1793.
11. Akkari N, Enghoff H, Metscher BD. A new dimension in documenting new species: high-detail imaging for myriapod taxonomy and first 3D cybertype of a new millipede species. PLoS One. 2015;10(8):e0135243.
12. Valan M, Makonyi K, Maki A, Vondráček D, Ronquist F. Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks. Syst Biol. 2019;68(6):876–895.
13. Zelditch ML, Swiderski DL, Sheets HD. Geometric Morphometrics for Biologists: A Primer. 2nd ed. London: Academic Press; 2012.
14. Dayrat B. Towards integrative taxonomy. Biol J Linn Soc. 2005;85(3):407–417.
15. Riedel A, Sagata K, Suhardjono YR, Tänzler R, Balke M. Integrative taxonomy on the fast track—towards more sustainability in biodiversity research. Front Zool. 2013;10(1):15.
16. Sharkey MJ, Janzen DH, Hallwachs W, Chapman EG, Smith MA, Dapkey T, *et al*. Minimalist revision and description of 403 new species in 11 subfamilies of Costa Rican braconid parasitoid wasps. ZooKeys. 2021;1013:1–665.
17. Van Horn G, Mac Aodha O, Song Y, Cui Y, Sun C, Shepard A, *et al*. The iNaturalist species classification and detection dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018. p. 8769–8778.
18. Culverwell CL. A review of the appraisal of deep learning approaches for the identification of mosquito species. Expert Syst Appl. 2021;178:115018.
19. Linz DM, Hu AW, Sitvarin MI, Tomoyasu Y. Functional value of elytra under various stresses in the red flour beetle, *Tribolium castaneum*. Sci Rep. 2016;6:34834.

How to Cite This Article

Mehta A, Sundaram P, Delgado-Rivas C, Wu ML. Multimodal genomic and imaging approaches for discovering hidden insect diversity. Int J Insect Anim Divers Res. 2025;1(6):27-30.

Creative Commons (CC) License

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution NonCommercial-ShareAlike 4.0 International (CC BYNCSA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.