Ecology & Evolution | Nov 13, 2026 | 5 min read | By Bryan White

# DNA Barcoding of One Million Specimens Reveals Over 60,000 Insect Species Across California, Doubling Prior Estimates

Ecology & Evolution * Entomology * DNA Barcoding

California is globally recognized as a biodiversity hotspot spanning alpine peaks, coastal redwood forests, Mediterranean chaparral, and hyper-arid deserts, yet the true richness of its most diverse animal class -- insects -- has long defied measurement [1, 2] . Traditional entomological checklists estimated roughly 28,000 to 30,000 described insect species across the state, leaving vast numbers of tiny, morphologically cryptic flies, wasps, and beetles uncataloged [1, 2] .

In a landmark biodiversity census published in the journal Ecography , the California Insect Barcode Initiative (CIBI) -- a 6 million USD collaboration uniting the Natural History Museum of Los Angeles County, UC Berkeley, UC Riverside, the San Diego Natural History Museum, the California Academy of Sciences, and the California Department of Food and Agriculture -- transformed that baseline [1, 2] . By sequencing DNA barcodes from more than one million individual insect specimens collected across the state, researchers demonstrated that California supports over 60,000 insect species , more than doubling prior estimates and revealing that at least one-third of the state's insect fauna remains scientifically undescribed [1, 2] .

## High-Throughput COI Barcoding and the "Dark Taxa" Bottleneck

Why did classical entomology miss more than half of California's insect species? The answer lies in the sheer taxonomic bottleneck created by hyper-diverse, miniature clades -- often termed dark taxa -- such as parasitoid wasps ( Hymenoptera ), gall midges and scuttle flies ( Diptera ), and micro-beetles (Coleoptera) [1, 3] .

* Standardized Mitochondrial Identification: Building on the global framework pioneered by Paul Hebert and colleagues, CIBI targeted a standardized ~648-base-pair region of the mitochondrial gene cytochrome c oxidase subunit I ( COI ) [1, 4] . Because mitochondrial COI evolves rapidly enough to accumulate diagnostic substitutions between closely related species while remaining conserved enough for universal primer binding, it serves as a high-resolution genetic fingerprint across animal phyla [4] .

* Barcode Index Numbers (BINs): Rather than waiting decades for specialist taxonomists to dissect genitalia under stereomicroscopes, researchers used non-destructive DNA extraction and multiplexed high-throughput sequencing to generate one million COI barcodes, clustering them via the Refined Single Linkage (RESL) algorithm on the Barcode of Life Data System (BOLD) [1, 3] . This assigns each genetically cohesive cluster a persistent Barcode Index Number (BIN) -- an operational taxonomic unit that closely mirrors biological species boundaries while preserving intact voucher specimens for future morphological description [3] .

* Illuminating Hyper-Diverse Guilds: While charismatic butterflies ( Lepidoptera )

and dragonflies (Odonata) are nearly 100% cataloged, the barcoding census revealed massive hidden radiations within Diptera (particularly Cecidomyiidae gall midges and Phoridae scuttle flies) and parasitoid Hymenoptera, whose host-specific life cycles drive extraordinary speciation [1, 3] .

## Asymptotic Richness Models and the Undiscovered Third

Even after processing more than one million specimens from Malaise traps, pitfall arrays, and pan traps, the statewide species accumulation curve failed to flatten into a plateau [1, 2] . Every new batch of samples continued to yield singleton and doubleton BINs -- species encountered only once or twice across the entire survey [1] .

| Biodiversity Metric | | Pre-CIBI Morphological Baseline | | CIBI Empirical Survey (>1M Specimens) | | Ecological & Taxonomic Significance |

| Estimated Statewide Species Richness | | ~28,000-30,000 described species | | >60,000 species (BINs) | | More than doubles historical estimates across California [1, 2] . |

| Undescribed / Undiscovered Share | | Unquantified ("dark taxa" blind spot) | | >=33% remain undescribed | | Statistical completeness models confirm thousands of unnamed lineages [1] . |

| Community Similarity (Beta-Diversity) | | Assumed broad ecoregional overlap | | Exceptionally high site turnover | | Even climatically matched sites host distinct insect assemblages [1, 2] . |

Applying asymptotic richness estimators (such as Chao1 and iNEXT extrapolation frameworks, which project unseen diversity from the ratio of rare singletons and doubletons to common species), the authors concluded that at least one-third of California's total insect biodiversity remains undiscovered by formal science [1] . Many of these cryptic lineages belong to ecological guilds -- such as fungal gnats, soil detritivores, and specialist parasitoids -- that perform indispensable nutrient-cycling and population-regulation functions in terrestrial food webs [1, 3] .

## Extreme Beta-Diversity: Why Every Valley Hosts a Distinct Fauna

Beyond raw species counts, the CIBI dataset exposed a striking spatial pattern: exceptionally high beta-diversity (species turnover between habitats) [1, 2] .

* Localized Micro-Endemism: Researchers found that nearly every trapping site contained its own unique assemblage of insect BINs, even when comparing locations separated by modest distances that share nearly identical climates, elevation, and dominant plant communities [1, 2] .

* Topographic and Geological Partitioning: California's complex mosaic of fault-bounded mountain ranges, isolated desert basins, serpentine soil outcrops, and rain-shadowed valleys creates fine-scale dispersal barriers for small, short-lived arthropods, driving rapid allopatric divergence across neighboring watersheds [1] .

* Phenological Specialization: In California's strongly seasonal Mediterranean climate, brief spring wildflower blooms and ephemeral vernal pools select for narrow

temporal emergence windows, isolating sympatric populations across seasons as well as across geography [1] .

## Conservation and Biosecurity Implications of the Genomic Baseline

The discovery that California harbors over 60,000 spatially partitioned insect species carries immediate consequences for both conservation biology and agricultural biosecurity [1, 2] :

* Rethinking Reserve Design: Because insect communities turn over sharply from site to site, protecting a handful of large, iconic national parks (such as Yosemite or Joshua Tree) is wholly insufficient to safeguard California's entomological diversity [1, 2] . Preserving regional insect biomass, specialist pollinators, and trophic stability requires maintaining a distributed network of smaller habitats across private rangelands, coastal canyons, urban greenbelts, and riparian corridors [1] .

* Rapid Invasive Pest Detection: With California producing more than a third of the United States' vegetables and over two-thirds of its fruits and nuts, distinguishing harmless native micro-insects from newly arrived agricultural pests is critical. The comprehensive CIBI COI reference library provides state agricultural inspectors with an authoritative genomic baseline, enabling real-time metabarcoding surveillance to flag invasive crop pests and disease vectors before outbreaks establish [1, 3] .

Crucially, because non-destructive barcoding preserves every sequenced insect as a physical museum voucher specimen, the CIBI archive creates a permanent, verifiable link between genomic BIN clusters and formal Linnaean taxonomy. As automated environmental DNA (eDNA) and Malaise-trap metabarcoding scale into routine ecological monitoring across North America, this million-specimen baseline transforms how scientists track insect population trends, climate range shifts, and ecosystem health in real time [1, 3, 4] .

## References & Primary Sources

* [1] Baker, A. J., Balukjian, B., Brown, B., Guzman, L. M., et al. (2026). DNA barcoding of more than one million insect specimens indicates that at least one third of California's insect biodiversity remains undiscovered. Ecography, e08572. -- https://doi.org/10.1002/ecog.08572

* [2] Bioengineer.org (2026). DNA Barcoding Identifies More Than 60,000 Insect Species Across California. -- https://bioengineer.org/

* [3] Ratnasingham, S. & Hebert, P. D. N. (2013). A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. PLoS ONE, 8(7), e66213. -- https://doi.org/10.1371/journal.pone.0066213

* [4] Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B, 270(1512), 313-321. -- https://doi.org/10.1098/rspb.2002.2218