Eurasia Specialized Veterinary Publication

Entomological Research Letters

ISSN: 3108-639X

2024, Volume 4, Issue 1, Page No: 125-135
Copyright CC BY-NC-SA 4.0
Available online at: www.esvpub.com/

Integrative Species Delimitation Should Weight Taxonomic Evidence According to Its Distinctive and Unequal Failure Modes

James Walker[1]*, Olivia Harris[1], George Wilson[2]

[1]Department of Evidence Weighting and Failure Modes in Taxonomy, Faculty of Life Sciences, University of Edinburgh, Edinburgh, United Kingdom.
[2]Department of Integrative Delimitation and Taxonomic Evidence, Faculty of Biosciences, University of Leeds, Leeds, United Kingdom.

***E-mail** ✉ james.walker@ed.ac.uk

ABSTRACT

Integrative species delimitation is commonly presented as a remedy for the limitations of single evidence classes, yet integration does not itself determine how conflicting observations should influence a taxonomic conclusion. Morphology, mitochondrial markers, genome-scale data, ecological differentiation and reproductive evidence can each fail for different biological or analytical reasons, and apparent agreement among evidence streams may reflect shared causes rather than genuine independence. This article develops a failure-mode perspective for integrative insect systematics. Instead of assigning evidence classes equal standing or privileging technologically intensive data, the proposed framework asks which processes can generate a misleading signal, whether those processes have been examined, whether sampling is adequate for the relevant spatial and biological scale, and whether nominally separate evidence streams contain genuinely independent information. Morphological evidence may fail through crypsis, polymorphism, sexual or maturational variation and inappropriate character selection; molecular evidence through marker inheritance, introgression, incomplete lineage sorting, threshold dependence and reference-library structure; and genome-scale evidence through population structure, gene flow, reticulation and model misspecification. Ecological and reproductive evidence add further information but remain conditional on geography, plasticity, contact opportunities and the scale at which isolation is measured. The resulting reliability-weighted model is explicitly proposed rather than empirically calibrated. Its purpose is to make taxonomic reasoning more transparent by linking the influence of each evidence stream to its diagnosed vulnerabilities, independence and unresolved uncertainty. Species hypotheses should therefore become more persuasive not merely by accumulating additional data, but by surviving targeted tests of the specific failure modes capable of producing the observed pattern.

Keywords: Integrative taxonomy, Species delimitation, Cryptic diversity, Gene flow, Genomic discordance, Evidence reliability

Received: 10 May 2024
Revised: 23 June 2024
Accepted: 25 June 2024

How to Cite This Article: Walker J, Harris O, Wilson G. Integrative Species Delimitation Should Weight Taxonomic Evidence According to Its Distinctive and Unequal Failure Modes. Entomol Res Lett. 2024;4(1):125-35. https://doi.org/10.51847/KedOwSMD1B

Introduction

Integrative taxonomy has become central to the resolution of difficult insect species complexes because morphology, molecular markers, life-history information and geographic evidence frequently illuminate different parts of the same taxonomic problem. Work on tephritid fruit flies demonstrates that multidimensional evidence can expose both synonymy and previously unrecognized diversity, but it also shows that integration does not remove disagreement among characters, markers or analytical approaches [1]. Large-scale integrative taxonomy

© 2024 Entomological Research Letters

similarly treats evidence acquisition as a staged process in which preliminary species hypotheses are tested with additional data where uncertainty persists rather than assuming that every available data type must be collected for every lineage [2]. The remaining methodological problem is therefore not whether integration is desirable, but how evidence should be interpreted when its component streams have unequal vulnerabilities.

That problem becomes acute because disagreement can arise in opposite directions. Morphology may under-recognize species when externally similar lineages contain cryptic diversity, a pattern documented across multiple insect groups [3]. Conversely, a barcode partition can provide useful identification structure without supplying the broader evidence needed to justify a formal species description or establish the biological meaning of the partition [4]. These are not equivalent shortcomings. The first concerns insufficient phenotypic discrimination; the second concerns an inferential transition from molecular partition to taxonomic conclusion. Treating both merely as "conflict" discards information about why a signal might be unreliable.

Genome-scale approaches create a related distinction between precision and biological relevance. Under the multispecies coalescent, population structure can be recovered with high statistical confidence even when the biological question is whether that structure corresponds to species rather than populations [5]. Gene flow adds another problem because models assuming strict isolation can misrepresent histories in which divergence and migration coexist [6]. Thus, increasing locus number or analytical sophistication cannot by itself resolve a mismatch between what a method estimates and what the taxonomist seeks to infer.

Direct insect studies reinforce this asymmetry. Nearctic *Phymata* ambush bugs show that sexual dimorphism, polymorphism, molecular disagreement and ecological association can contribute differently to delimitation depending on the taxon and character set [7]. In *Temnothorax* ants, genome-scale lineage hypotheses helped reveal diagnostic morphological differences that were not initially obvious, showing that morphology can become more informative after an independent lineage hypothesis has been formulated [8]. Neither case supports a universal hierarchy in which morphology, molecules or genomics should dominate. Instead, they suggest that evidence should be interpreted according to the mechanisms capable of generating false separation, false similarity or misleading concordance.

This article therefore develops a proposed reliability-weighted framework for integrative species delimitation. Reliability is not treated as a permanent property of an evidence class. It is treated as conditional on the failure modes relevant to a particular species hypothesis, the extent to which those failure modes have been tested, the adequacy of sampling, and the relative independence of supporting observations. The purpose is not to replace species concepts with a numerical scoring scheme, but to make explicit why one evidence stream should influence a difficult taxonomic decision more strongly than another in a given case.

Materials and Methods

The framework was constructed as a conceptual failure-mode analysis rather than an empirical performance study. Evidence was organized around four broad classes—morphological, molecular, genomic, and ecological or behavioral—but the analysis operated below those labels whenever individual characters, markers, models or sampling designs had distinguishable vulnerabilities. A failure mode was defined as a biological or analytical process capable of generating an observation that could be mistaken for stronger support for, or stronger opposition to, a species boundary than the observation warrants.

This definition is necessary because molecular delimitation methods do not show constant performance across divergence histories. Comparative simulations demonstrate that different procedures respond differently to population size, divergence depth, priors and gene flow, so method reliability cannot be inferred solely from the kind of data analyzed [9]. Likewise, increasing the amount of genetic information can strengthen support for structured populations without solving the conceptual problem of whether those populations should be treated as species [10]. Geographic sampling is therefore part of reliability assessment: widespread or parapatric taxa can appear discretely structured when intermediate populations and contact zones are insufficiently represented [11].

Sampling effects also occur within single-marker insect datasets. In leaf beetles, molecular-delimitation efficiency changed with haplotype representation, geographic sampling and methodological choice, indicating that "barcode evidence" cannot be assigned one constant level of confidence independently of dataset composition [12]. The proposed analysis therefore separates observational strength—for example, a well-supported genetic cluster—from inferential reliability, meaning how securely that observation can be connected to species-level independence after plausible failure processes are considered.

Model adequacy forms a second axis. Species-boundary estimates can change when coalescent models represent divergence processes differently, demonstrating that apparently precise outputs remain conditional on the assumptions encoded by the model [13]. High-dimensional integration does not eliminate the interpretive step: unsupervised machine-learning approaches can combine genetic, geographic, environmental and phenotypic features, but the recovered cluster remains a statistical structure that requires biological interpretation before it becomes a taxonomic conclusion [14].

Accordingly, the proposed reliability logic uses four questions rather than fixed numerical weights: What failure process could generate this signal? Has that process been examined? Is sampling adequate at the relevant biological and geographic scale? Is this information independent of evidence already counted? An evidence stream is not automatically discounted because it has a known failure mode; all evidence classes do. Reliability is reduced only when a plausible high-impact failure remains untested or when the inference exceeds what the observation directly measures. Conversely, an evidence stream becomes more persuasive when its most relevant failure processes have been explicitly challenged and when its contribution is not merely a replication of the same underlying biological signal.

Results and Discussion

Morphological evidence is often discussed as though "morphology" were a single data class, but individual characters can differ substantially in reliability. In *Psylliodes* flea beetles, qualitative characters, morphometric traits and genital features differed in diagnostic value, while maturity and sex affected the usability of particular observations [15]. The relevant failure mode is therefore sometimes not morphological conservatism at the species level but character selection at the specimen level. A structure visible only after complete sclerotization, a sexually dimorphic character, or a trait sensitive to preparation cannot be interpreted as equivalent to a stable diagnostic character expressed across the sampled population.

Morphology can also generate false separation. In the *Bactrocera frauenfeldi* complex, conspicuous morphotypes did not map straightforwardly onto species boundaries when evaluated against phylogenomic and biogeographic evidence [16]. Phenotypic discontinuity can therefore represent geographic variation, polymorphism or an incompletely sampled continuum rather than lineage independence. This is analytically distinct from crypsis, in which morphology produces too little apparent differentiation. A useful failure-mode scheme must consequently distinguish morphological under-resolution from morphological overpartitioning rather than simply assigning morphology a low or high global reliability.

Discordance can also reveal when morphology is comparatively informative. In a tiger beetle complex, mitochondrial evidence conflicted more strongly with genomic and morphological patterns than the latter conflicted with each other [17]. Such cases do not establish morphology as a superior evidence class. They show instead that a morphological signal should be interpreted in relation to its own vulnerabilities and to the diagnosed vulnerabilities of competing evidence. If mtDNA is susceptible to introgression or lineage sorting in the focal history, congruence between morphology and independently sampled nuclear evidence becomes more informative than a simple two-versus-one vote.

The proposed framework therefore treats morphology at the level of the character and biological context. Reliability increases where characters are stable across sex, maturity, preservation and geography, and where phenotypic discontinuities coincide with independently supported lineage boundaries. It decreases where a trait is developmentally contingent, environmentally labile, geographically clinal, polymorphic or selected because it happens to reproduce a pre-existing taxonomic expectation. Morphological similarity and morphological difference are thus both observations requiring an explicit failure-mode diagnosis before either is translated into species status.

Molecular failure modes

Single-locus and mitochondrial evidence provide rapid and often valuable information, but their vulnerabilities arise partly from their inheritance history. In rice seed bugs, mitochondrial markers failed to distinguish closely related taxa that were separated by nuclear SNP data and stable morphological differences [18]. The immediate observation is mito-nuclear discordance; attributing that discordance to introgression, incomplete lineage sorting or another historical process requires additional evidence. This distinction matters because reliability should respond first to the demonstrated mismatch and only second to a proposed mechanism.

A clearer mitochondrial-introgression problem has been reported in Chinese *Cletus*, where mitochondrial history obscured species boundaries inferred from nuclear evidence [19]. Such cases illustrate why mitochondrial divergence should not be treated as a direct surrogate for species history. They do not imply that nuclear history is automatically correct or that all mitochondrial discordance has the same cause. The appropriate inference is narrower: when cytoplasmic inheritance has plausibly crossed species boundaries, a mitochondrial partition should carry less taxonomic influence unless other evidence demonstrates that the partition corresponds to independently evolving lineages.

Barcode reliability is also conditional on reference structure. A large Dacini COI resource showed exceptions to fixed distance criteria and cases of species non-monophyly, limiting the portability of a single threshold across taxa [20]. A barcode gap or cluster can be highly useful for identification where the reference library is taxonomically representative and evolutionary history is uncomplicated. The same pattern is weaker as species-delimitation evidence when close relatives are absent from the reference set, when named species are non-monophyletic, or when the marker has known introgression or lineage-sorting problems.

Importantly, marker failure is not unidirectional. In the *Anopheles maculipennis* complex, molecular approaches produced both excessive subdivision in one taxonomic context and insufficient separation in another, while morphology provided additional information [21]. Molecular reliability therefore cannot be represented as a single tendency toward either splitting or lumping. It depends on marker inheritance, divergence history, taxon sampling, reference completeness and the inferential method applied to the sequence data.

The distinctions above are easier to recover when molecular observations are organized by the specific process capable of distorting the taxonomic inference rather than by marker name alone. The resulting cross-scale grid therefore separates what is observed from what can legitimately be inferred and from what must still be tested.

Table 1. Molecular signals become taxonomically unequal when marker history, reference structure and failure direction differ

	Signal under test		**Inheritance history**	**Resolution**	
Molecular vulnerability	**Observed conflict**	**Naive taxonomic error**	**Historical setting**	**Evidence that discriminates**	**Unresolved cause**
Mito-nuclear and cytoplasmic history					
Mito-nuclear decoupling	Mitochondrial grouping conflicts with nuclear genomic and morphological structure	Mitochondrial similarity may be mistaken for conspecificity, or mitochondrial divergence for separate species	Recently diverged or reticulate insect lineages	Test nuclear loci, geographic structure and alternative historical explanations	Discordance alone does not identify introgression versus incomplete sorting
Cytoplasmic-history capture	Mitochondrial genealogy follows a history inconsistent with broader lineage evidence	A transferred mitochondrial lineage may be treated as the species history	Lineages with plausible interspecific gene exchange	Use nuclear evidence and explicit tests of reticulation where feasible	Direction and timing of introgression may remain uncertain
Threshold and reference effects					
Threshold-dependent barcode partition	Species assignments change with distance rule or tree criterion	A convenient operational threshold may be mistaken for a biological boundary	Reference libraries containing close relatives, complexes or non-monophyletic taxa	Evaluate reference completeness, monophyly and within-versus-between variation	Threshold performance is taxon- and library-dependent
Reference-library incompleteness	Apparently distinctive haplotypes lack adequate close-relative comparison	Novel sequence clusters may be overinterpreted as novel species	Poorly sampled regions or species-rich complexes	Expand taxonomic and geographic sampling before taxonomic escalation	Unsampled diversity can alter apparent uniqueness
Direction of marker error					
Opposite-direction marker error	Overpartitioning and under-resolution	"Marker reliability" is incorrectly treated as one-dimensional	Complexes in which the same marker can split	Diagnose each boundary independently rather	Failure direction varies among lineage histories

	occur within one analytical system		one taxon yet merge another	than applying one global marker weight	
inheritance-history vulnerability (proposed)	Strong molecular structure with a plausible process that redirects marker history	Statistical support may exceed taxonomic reliability	Any molecular stream whose genealogy can decouple from species history	Down-weight only when the relevant failure process is plausible and insufficiently tested	Strength of signal is not equivalent to relevance of signal

Genomic failure modes

Genome-scale data reduce stochastic error from individual loci but do not remove biological processes that create discordance. Under gene flow, sequence relationships can violate the intuitive expectation that members of the same species will always be genetically more similar to each other than to members of another species [22]. The problem is therefore not insufficient information but an evolutionary process that changes what genetic similarity means. Genome-scale structure must consequently be interpreted through a model that is compatible with the demographic history being inferred.

Spatial structure creates a related risk. Work on geographically widespread taxa has shown that gene flow, environmental differentiation and population structure can inflate estimates of cryptic lineages when strong genetic subdivisions are read directly as species boundaries [23]. Although such demonstrations are not restricted to insects, the methodological implication is general: genomic clustering measures structured variation, whereas species delimitation requires an additional argument about the biological meaning of that structure.

Introgression further weakens any assumption that one bifurcating genomic history necessarily captures the relevant species history. Empirical work has shown that reticulation can complicate model-based delimitation where strict isolation is assumed [24]. Complementary analyses of clustering, species trees, demographic history and gene flow can therefore resolve different aspects of a difficult boundary rather than serving as redundant confirmations of a single result [25]. Multiple analyses are most useful when they test distinct failure hypotheses.

Direct insect phylogenomics makes the model issue especially clear. Within an important *Bactrocera* complex, concatenation, multispecies-coalescent and network approaches produced conflicting relationships, while gene flow was detected among lineages [26]. High node support under one model therefore did not eliminate uncertainty created by model mismatch. A topology can be statistically well supported conditional on a model and still remain taxonomically vulnerable if the evolutionary process violating that model is biologically plausible.

Migration-aware delimitation can reduce some forms of oversplitting, but its own thresholds and assumptions remain conditional. Hierarchical heuristic methods incorporating migration performed better than simpler alternatives in tested scenarios yet retained difficult cases involving structure and lineage history [27]. The appropriate lesson is not that one migration-aware method provides a universal species criterion, but that model compatibility must itself enter the reliability assessment.

These genomic failure modes are summarized visually as a conditional map linking evolutionary process, observed genomic pattern, analytical treatment and taxonomic inference. The figure emphasizes that the boundary between “strong genomic signal” and “strong species evidence” is crossed only after the processes capable of generating the signal have been examined.

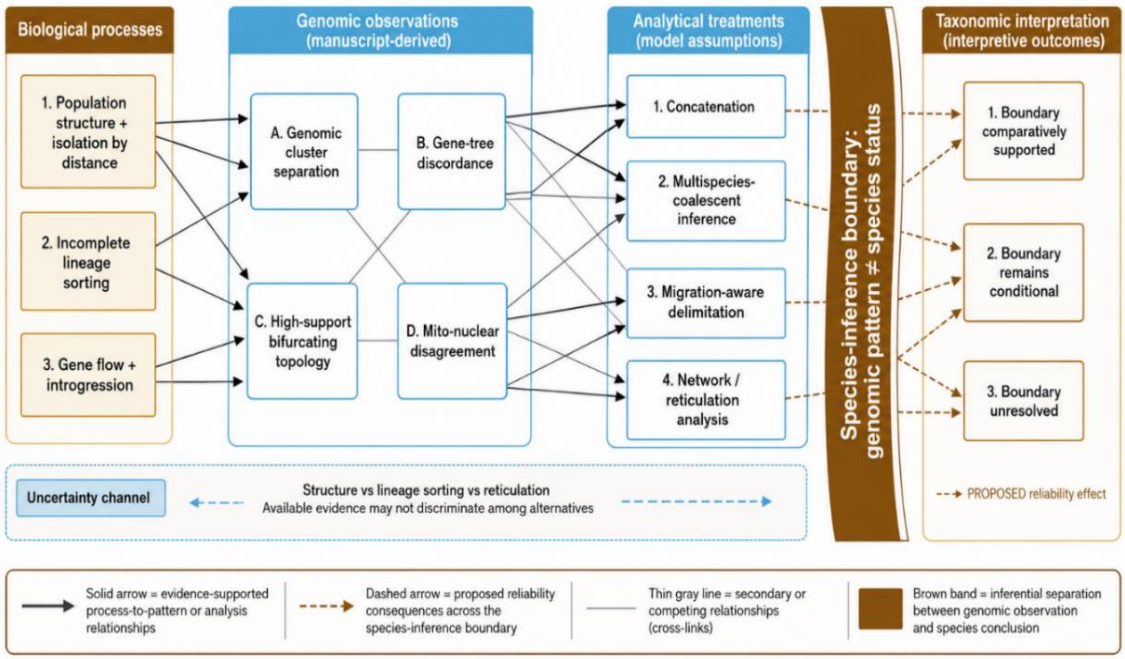


Figure 1. Genomic support changes meaning across structure, gene flow and model-adequacy boundaries

Ecological and behavioral failure modes

Ecological and behavioral evidence can provide information that genetic differentiation alone cannot, but neither constitutes a universal species criterion. In *Drosophila*, model-based coalescent delimitation and evidence of reproductive isolation showed substantial but incomplete concordance [28]. Reproductive evidence can therefore challenge or corroborate genomic hypotheses, yet its interpretation depends on what component of isolation was measured, whether opportunities for mating occurred, and whether the compared populations are sympatric, parapatric or allopatric. Failure to observe gene exchange where contact is impossible is not equivalent to demonstrating a reproductive barrier.

Geography creates a particularly important conditionality. Genomic work in allopatric Lepidoptera shows that substantial divergence can accumulate within and across continental separations, making spatial isolation part of the interpretation rather than independent proof of species status [29]. Deep allopatric structure may represent long-standing lineage independence, but it may also precede the evolution of reproductive or ecological differences that could be tested only after secondary contact. The relevant failure mode is therefore geographic opportunity confounding: absence of observed exchange can arise because exchange is biologically prevented or simply because populations do not meet.

Ecological association requires similar caution. Host use, habitat occupation and climatic separation can contribute strongly to delimitation when they are consistently associated with independently supported lineages, but ecological differences may also arise through plasticity, local availability or incomplete sampling of the species' environmental range. Conversely, apparent ecological similarity can conceal divergent reproductive or physiological processes when measured environmental variables are too coarse. The taxonomic value of ecology thus depends on whether the relevant ecological axis is measured at the scale at which lineage separation is hypothesized.

Behavioral evidence is likewise scale-sensitive. Mate choice, phenology, aggregation, host preference and signal differences may increase confidence in a species hypothesis when they restrict realized exchange, yet experimental behavior can fail to transfer to natural populations and field observations may confound preference with opportunity. These limitations differ from those affecting morphology or sequence data: they concern the translation from observed behavior or ecological association to persistent lineage independence.

The proposed framework therefore treats ecological and behavioral evidence as conditionally informative rather than intrinsically weaker. Its reliability increases when the measured variable is lineage-specific, temporally persistent, geographically replicated and connected to an independently testable consequence such as restricted contact or differential habitat use. It decreases when opportunity, plasticity, season, life stage or coarse environmental measurement can generate the same pattern.

Because ecological and behavioral observations become misleading under different combinations of opportunity, scale and plasticity, the relevant conditionalities are best compared as failure scenarios rather than compressed into a generic list of strengths and weaknesses.

Table 2. Ecological and behavioral evidence changes taxonomic meaning when opportunity, scale and persistence are unresolved

Opportunity to interact	
Geographic opportunity confounding	
Opportunity or persistence distinction	Allopatry versus biological isolation
No observed gene exchange between separated populations	Lack of exchange is treated as reproductive isolation
Observed pattern	**Taxonomic overreach**
Condition changing meaning	Populations may never encounter one another
Sample contact zones, natural or experimental secondary contact, or complementary reproductive evidence	**EVIDENCE BEFORE INFERENCE**
Partial and plastic divergence	
Partial reproductive concordance	
Opportunity or persistence distinction	Coalescent lineage versus measured reproductive barrier
Genomic boundary and reproductive evidence agree only incompletely	Either genomic delimitation or reproductive isolation is treated as an absolute gold standard
Observed pattern	**Taxonomic overreach**
Condition changing meaning	Reproductive isolation is multidimensional and may be incomplete
Test multiple barrier components and retain discordant pairs explicitly	**EVIDENCE BEFORE INFERENCE**
Plastic ecological divergence	
Opportunity or persistence distinction	Habitat or host association
Populations occupy different environments or resources	Environmentally induced differences are interpreted as fixed lineage divergence
Observed pattern	**Taxonomic overreach**
Condition changing meaning	Resource availability, developmental plasticity and local context may reproduce the pattern
Replicate associations geographically and across environmental conditions	**EVIDENCE BEFORE INFERENCE**
Scale and laboratory-to-field transfer	
Coarse-scale ecological similarity	
Opportunity or persistence distinction	Broad habitat overlap
Lineages appear ecologically equivalent at measured resolution	Absence of detected niche divergence is interpreted as absence of ecological differentiation
Observed pattern	**Taxonomic overreach**
Condition changing meaning	Relevant ecological axes may operate below the measurement grain
Increase environmental, temporal or host-resolution before dismissing differentiation	**EVIDENCE BEFORE INFERENCE**
Behavioral opportunity bias	
Opportunity or persistence distinction	Mate choice, host preference or phenology
Behavioral differences appear only where encounters occur	Encounter pattern is interpreted as intrinsic preference or isolation
Observed pattern	**Taxonomic overreach**
Condition changing meaning	Abundance, season and encounter probability alter observed behavior
Separate preference from opportunity experimentally or through replicated field design	**EVIDENCE BEFORE INFERENCE**
Laboratory-to-field transfer failure	
Opportunity or persistence distinction	Experimentally measured behavioral isolation
Strong laboratory discrimination between populations	Laboratory behavior is assumed to restrict gene flow in nature
Observed pattern	**Taxonomic overreach**

Condition changing meaning	Artificial density, cue presentation or mating opportunity may alter behavior
Compare laboratory results with field contact, mating or gene-flow evidence	**EVIDENCE BEFORE INFERENCE**
Persistence requirement	
persistence requirement (proposed)	
Opportunity or persistence distinction	Any ecological or behavioral evidence used to support delimitation
Repeated lineage-specific association across relevant conditions	A transient association is treated as durable species evidence
Observed pattern	**Taxonomic overreach**
Condition changing meaning	Temporal, life-stage or environmental stability remains untested
Increase reliability only when the association persists across the scale relevant to the species hypothesis	**EVIDENCE BEFORE INFERENCE**

Non-independence among evidence classes

Concordance is strongest when evidence streams are complementary rather than repeated views of the same signal. Independently generated genomic datasets can corroborate boundaries, yet shared demographic history still limits biological independence [30]. The proposed framework therefore separates data-generation, analytical and process independence. Agreement is down-weighted when several streams plausibly reflect the same underlying process; discordance instead triggers failure-mode diagnosis before averaging.

Reliability-weighted decision model

The proposed model treats evidence as conditionally influential rather than equally weighted. Integrative work in *Chanohirata* illustrates that molecular delimitation can inform, but not replace, broader systematic interpretation [31]. Each evidence unit is assessed for inferential relevance, sampling adequacy, diagnosed failure modes and relative independence. No numerical weights are assigned. A boundary is comparatively supported only when influential streams address the taxonomic proposition and principal vulnerabilities have been challenged; otherwise it remains conditional or unresolved.

Because these checks operate at different levels, **Table 3** separates what is evaluated from how it changes the decision.

Table 3. Reliability checks determining whether evidence should influence a species-boundary decision

Admission check		**Decision boundary**	
Decision gate	**Evidence unit presented**	**Stop condition**	**Consequence if passed**
Relevance and failure exposure			
Inferential relevance	Character, marker, ecology or genomic analysis	Structure or association may not equal lineage independence	Increase influence only when the observation addresses the proposition
Failure exposure	Any stream with a plausible confounder	Competing histories can generate the same pattern	Reduce influence while a high-impact alternative remains untested
Sampling and provenance			
Sampling adequacy	Individuals, stages, sexes, localities or reference taxa	Unsampled contact zones or close relatives may alter interpretation	Qualify claims exceeding sampled scale
Evidence provenance	Separate datasets or analyses	Shared data or processes can create apparent replication	Do not multiply non-independent support
Decision state			
Decision state	Integrated evidence	State describes reasoning robustness, not species probability	Comparatively supported, conditional or unresolved; no score

Figure 2 places these checks between evidence and taxonomic interpretation so that no stream connects directly to a species decision.

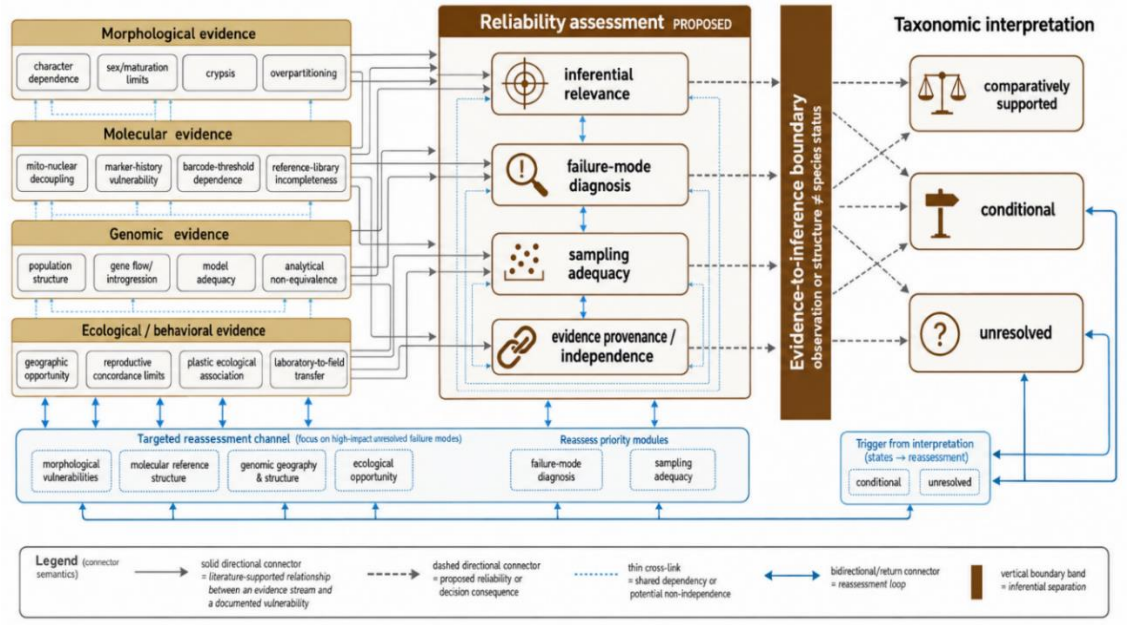


Figure 2. Evidence reaches taxonomic interpretation only through failure-mode and independence checks

Sensitivity analysis

Sensitivity analysis is conceptual rather than statistical. The decision is re-examined after down-weighting the most vulnerable stream, varying assumptions about geography or gene flow, and separating repeated analyses from complementary evidence. Method performance [9], sampling [12] and migration assumptions [27] can alter outcomes. Reversal indicates dependence on assumptions, not proof that the original hypothesis is false.

Failure-mode weighting changes integration from vote counting to diagnostic adjudication. It distinguishes evidence quantity from independence and statistical support from taxonomic relevance. Model formulation can alter inference [13], while multidimensional clustering still requires biological interpretation [14]. Future validation should test contact zones [11], character portability [15], introgression [19], marker disagreement [21] and allopatric divergence [29] rather than infer universal weights.

Conclusion

Integrative species delimitation is strongest when each observation is judged by how it can mislead. The proposed framework links evidential influence to relevance, failure-mode diagnosis, sampling adequacy and relative independence while avoiding fixed scores or automatic decisions. Its contribution is procedural: preserve unresolved alternatives, prevent concordance from being mistaken for independence, and direct additional evidence toward the failure mode most capable of changing the conclusion

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Schutze MK, Virgilio M, Norrbom A, Clarke AR. Tephritid integrative taxonomy: Where we are now, with a focus on the resolution of three tropical fruit fly species complexes. Annu Rev Entomol. 2017;62:147-64. doi:10.1146/annurev-ento-031616-035518

2. Hartop E, Srivathsan A, Ronquist F, Meier R. Towards large-scale integrative taxonomy (LIT): Resolving the data conundrum for dark taxa. Syst Biol. 2022;71(6):1404-22. doi:10.1093/sysbio/syac033
3. Li X, Wiens JJ. Estimating global biodiversity: The role of cryptic insect species. Syst Biol. 2023;72(2):391-403. doi:10.1093/sysbio/syac069
4. Zamani A, Fric ZF, Gante HF, Hopkins T, Orfinger AB, Scherz MD, et al. DNA barcodes on their own are not enough to describe a species. Syst Entomol. 2022;47(3):385-9. doi:10.1111/syen.12538
5. Sukumaran J, Knowles LL. Multispecies coalescent delimits structure, not species. Proc Natl Acad Sci U S A. 2017;114(7):1607-12. doi:10.1073/pnas.1607921114
6. Jackson ND, Carstens BC, Morales AE, O'Meara BC. Species delimitation with gene flow. Syst Biol. 2017;66(5):799-812. doi:10.1093/sysbio/syw117
7. Masonick P, Weirauch C. Integrative species delimitation in Nearctic ambush bugs (Heteroptera: Reduviidae: Phymatinae): Insights from molecules, geometric morphometrics and ecological associations. Syst Entomol. 2020;45(1):205-23. doi:10.1111/syen.12388
8. Prebus MM. Phylogenomic species delimitation in the ants of the Temnothorax salvini group (Hymenoptera: Formicidae): An integrative approach. Syst Entomol. 2021;46(2):307-26. doi:10.1111/syen.12463
9. Luo A, Ling C, Ho SYW, Zhu CD. Comparison of methods for molecular species delimitation across a range of speciation scenarios. Syst Biol. 2018;67(5):830-46. doi:10.1093/sysbio/syy011
10. Leaché AD, Zhu T, Rannala B, Yang Z. The spectre of too many species. Syst Biol. 2019;68(1):168-81. doi:10.1093/sysbio/syy051
11. Chambers EA, Hillis DM. The multispecies coalescent over-splits species in the case of geographically widespread taxa. Syst Biol. 2020;69(1):184-93. doi:10.1093/sysbio/syz042
12. Magoga G, Fontaneto D, Montagna M. Factors affecting the efficiency of molecular species delimitation in a species-rich insect family. Mol Ecol Resour. 2021;21(5):1475-89. doi:10.1111/1755-0998.13352
13. Douglas J, Bouckaert R. Quantitatively defining species boundaries with more efficiency and more biological realism. Commun Biol. 2022;5(1):755. doi:10.1038/s42003-022-03723-z
14. Pyron RA. Unsupervised machine learning for species delimitation, integrative taxonomy, and biodiversity conservation. Mol Phylogenet Evol. 2023;189:107939. doi:10.1016/j.ympev.2023.107939
15. Berrilli E, Biondi M, D'Alessandro P, Salvi D. Cryptic, sibling or neither of the two? Integrative species delimitation of Psylliodes flea beetles with overlapping ranges. Zool Scr. 2023;52(3):235-48. doi:10.1111/zsc.12574
16. Doorenweerd C, San Jose M, Geib S, Dupuis J, Leblanc L, Barr N, et al. A phylogenomic approach to species delimitation in the mango fruit fly (Bactrocera frauenfeldi) complex: A new synonym of an important pest species with variable morphotypes (Diptera: Tephritidae). Syst Entomol. 2023;48(1):10-22. doi:10.1111/syen.12559
17. Duran DP, Laroche RA, Roman SJ, Godwin W, Herrmann DP, Bull E, et al. Species delimitation, discovery and conservation in a tiger beetle species complex despite discordant genetic data. Sci Rep. 2024;14(1):6617. doi:10.1038/s41598-024-56875-9
18. Dong X, Yi W, Zheng C, Zhu X, Wang S, Xue H, et al. Species delimitation of rice seed bugs complex: Insights from mitochondrial genomes and ddRAD-seq data. Zool Scr. 2022;51(2):185-98. doi:10.1111/zsc.12523
19. Dong X, Zhang H, Zhu X, Wang K, Xue H, Ye Z, et al. Mitochondrial introgression and mito-nuclear discordance obscured the closely related species boundaries in Cletus Stål from China (Heteroptera: Coreidae). Mol Phylogenet Evol. 2023;184:107802. doi:10.1016/j.ympev.2023.107802
20. Doorenweerd C, San Jose M, Leblanc L, Barr N, Geib SM, Chung AYC, et al. Towards a better future for DNA barcoding: Evaluating monophyly- and distance-based species identification using COI gene fragments of Dacini fruit flies. Mol Ecol Resour. 2024;24(6). doi:10.1111/1755-0998.13987
21. Calzolari M, Bellin N, Dottori M, Torri D, Di Luca M, Rossi V, et al. Integrated taxonomy to advance species delimitation of the Anopheles maculipennis complex. Sci Rep. 2024;14(1):30914. doi:10.1038/s41598-024-81767-3
22. Jiao X, Yang Z. Defining species when there is gene flow. Syst Biol. 2021;70(1):108-19. doi:10.1093/sysbio/syaa052

23. Chan KO, Hutter CR, Wood PL Jr, Su YC, Brown RM. Gene flow increases phylogenetic structure and inflates cryptic species estimations: A case study on widespread Philippine puddle frogs (Occidozyga laevis). Syst Biol. 2022;71(1):40-57. doi:10.1093/sysbio/syab034
24. Kim D, Bauer BH, Near TJ. Introgression and species delimitation in the longear sunfish Lepomis megalotis (Teleostei: Percomorpha: Centrarchidae). Syst Biol. 2022;71(2):273-85. doi:10.1093/sysbio/syab029
25. DeRaad DA, McCormack JE, Chen N, Peterson AT, Moyle RG. Combining species delimitation, species trees, and tests for gene flow clarifies complex speciation in scrub-jays. Syst Biol. 2022;71(6):1453-70. doi:10.1093/sysbio/syac034
26. San Jose M, Doorenweerd C, Geib SM, Barr N, Dupuis JR, Leblanc L, et al. Interspecific gene flow obscures phylogenetic relationships in an important insect pest species complex. Mol Phylogenet Evol. 2023;188:107892. doi:10.1016/j.ympev.2023.107892
27. Kornai D, Jiao X, Ji J, Flouri T, Yang Z. Hierarchical heuristic species delimitation under the multispecies coalescent model with migration. Syst Biol. 2024;73(6):1015-37. doi:10.1093/sysbio/syae050
28. Campillo LC, Barley AJ, Thomson RC. Model-based species delimitation: Are coalescent species reproductively isolated? Syst Biol. 2020;69(4):708-21. doi:10.1093/sysbio/syz072
29. Joshi M, Espeland M, Huemer P, deWaard J, Mutanen M. Species delimitation under allopatry: Genomic insights within and across continents in Lepidoptera. Insect Syst Divers. 2024;8(5):7. doi:10.1093/isd/ixae027
30. Wootton LM, Forest F, Verboom GA. Consilience across multiple, independent genomic data sets reveals species in a complex with limited phenotypic variation. Syst Biol. 2023;72(4):753-66. doi:10.1093/sysbio/syad024
31. Wang D, Dietrich CH, Zhang Y. Phylogenetics, historical biogeography and molecular species delimitation of Chanohirata (Hemiptera: Cicadellidae: Deltocephalinae: Penthimiini). Syst Entomol. 2024;49(1):173-92. doi:10.1111/syen.12612