Eurasia Specialized Veterinary Publication

Entomological Research Letters

ISSN: 3108-639X

2022, Volume 2, Issue 1, Page No: 118-127
Copyright CC BY-NC-SA 4.0
Available online at: www.esvpub.com/

Treating Species Delimitation as Model Choice Can Make Integrative Insect Systematics More Transparent and Reproducible

Gabriel Costa[1]*, Lucas Ribeiro[1], Ricardo Alves[2], Mariana Lopes[3]

[1]Department of Model-Based Species Delimitation, Faculty of Biosciences, Federal University of Minas Gerais, Belo Horizonte, Brazil.
[2]Department of Transparent Systematics and Model Choice, Faculty of Life Sciences, University of Coimbra, Coimbra, Portugal.
[3]Department of Reproducible Taxonomy and Integrative Methods, Faculty of Sciences, University of São Paulo, São Paulo, Brazil.

***E-mail** ✉ gabriel.costa@ufmg.br

ABSTRACT

Integrative insect systematics increasingly combines mitochondrial barcodes, multilocus or genomic data, morphology, geography, ecology and reproductive information, yet species-delimitation outputs are still often presented as if they directly reveal species boundaries. This methodological framework argues that delimitation is more transparent when treated as model choice: investigators specify competing species hypotheses, state the assumptions connecting each evidence class to those hypotheses, compare evidential support, preserve discordance, test sensitivity to priors and thresholds, and report uncertainty before taxonomic action. The framework does not privilege a single species concept or algorithm. Instead, it separates observed evidence, process assumptions, statistical support and taxonomic interpretation. Insect case studies show why this separation matters: coalescent partitions can track population structure, gene flow can blur boundaries, mitochondrial and nuclear signals can disagree, and apparently concordant evidence can still remain conditional on sampling and analytical choices. The proposed architecture therefore treats species delimitation as an auditable sequence of competing explanations rather than an automatic conversion of clusters into names. Its purpose is methodological reproducibility, not universal taxonomic resolution.

Keywords: Species delimitation, Integrative taxonomy, Multispecies coalescent, Model comparison, Insect systematics,Taxonomic uncertainty

Received: 12 May 2022
Revised: 26 June 2022
Accepted: 29 June 2022

How to Cite This Article: Costa G, Ribeiro L, Alves R, Lopes M. Treating Species Delimitation as Model Choice Can Make Integrative Insect Systematics More Transparent and Reproducible. Entomol Res Lett. 2022;2(1):118-27. https://doi.org/10.51847/AZSzOenOnW

Introduction

Species delimitation is often described as the discovery of boundaries that already exist in nature, but most operational delimitations are inferences produced by data, models and decision rules. This distinction is especially important under the multispecies coalescent (MSC), because strong genetic structure can be interpreted as species-level divergence even when the model is primarily detecting population structure [1]. Empirical work in *Drosophila* further shows that coalescent species hypotheses and reproductive isolation can agree substantially without being identical, so statistical support and an external biological criterion cannot be treated as interchangeable [2].

Insect systematics provides unusually clear demonstrations of this problem because competing evidence classes can be collected at different biological and analytical levels. In Haeterini butterflies, alternative taxonomic

© 2022 Entomological Research Letters

opinions were converted into competing prior structures and species hypotheses, allowing their support to be compared explicitly rather than left implicit in a preferred taxonomy [3]. Other integrative studies of ambush bugs, *Phoebis* butterflies and Central European bees show that morphology, mitochondrial data, nuclear data and ecology can produce non-identical partitions or interpretations [4–6]. Such disagreement is not a defect to be averaged away; it is information about which biological process, sampling regime or analytical assumption remains unresolved. Parapatric *Melitaea* butterflies provide a further boundary case because gene flow and genomic continuity can make a crisp species partition difficult to justify even with dense genomic evidence [7].

The methodological problem, therefore, is not simply how to combine more evidence. It is how to make explicit the inferential bridge between evidence and taxonomic conclusion. This article proposes treating integrative species delimitation as model choice. Candidate species models are defined in advance, evidence classes remain analytically separable, statistical support is distinguished from external corroboration, discordance is adjudicated through competing explanations, and taxonomic action is delayed until model sensitivity and residual uncertainty are visible. The framework is proposed rather than empirically validated; its intended contribution is an auditable architecture for insect systematics in which disagreement can be inspected and reproduced.

Materials and Methods

The framework is constructed from methodological and insect-systematics evidence that exposes how alternative models encode different biological assumptions. A first distinction concerns migration. Species-delimitation models that permit gene flow address a different data-generating process from strict no-migration MSC models, so migration cannot be treated as a minor post hoc complication when it is biologically plausible [8]. Simulation comparisons likewise show that the relative behavior of delimitation procedures varies across speciation scenarios, population sizes and divergence conditions; there is no context-free basis for naming one algorithm the universally superior choice [9].

A second distinction concerns the level at which evidence is generated. Gene-tree discordance can arise under incomplete lineage sorting, but the MSC does not by itself explain every source of discordance, including introgression, hybridization or other departures from its assumptions [10]. Genomic MSC methods such as BPP provide probabilistic inference conditional on that model family and its priors [11]. By contrast, mPTP partitions a single-locus input tree according to branching-process assumptions [12], while ASAP ranks candidate partitions from pairwise sequence distances [13]. SODA uses quartet frequencies and explicit zero-branch tests, with the significance threshold α affecting how aggressively lineages are split [14]. These procedures therefore do not represent interchangeable measurements of one latent species boundary.

The proposed method treats an analysis as a set of candidate species models M1…Mk rather than as a sequence of software outputs. Each model contains four explicit components: the proposed partition of sampled lineages; the process assumptions under which that partition is meaningful; the evidence classes judged relevant to the comparison; and the decision settings, including priors, thresholds, guide-tree assumptions or preprocessing choices. Evidence is then recorded without being collapsed into a single score. A mitochondrial genealogy, nuclear genomic structure, morphology, geography, reproductive compatibility and ecology may all inform the same candidate model, but they remain distinct observations because they can respond differently to introgression, incomplete lineage sorting, phenotypic conservatism, local adaptation, spatial isolation or sampling.

This architecture also separates discovery from adjudication. Distance- or tree-based procedures can efficiently generate candidate partitions, whereas MSC-based or other probabilistic methods may compare more explicit lineage models. Neither stage automatically authorizes formal taxonomic action. The key methodological requirement is that every transition—from observation to process interpretation, from process interpretation to candidate species model, and from model support to taxonomic conclusion—be stated rather than implied.

Results and Discussion

The principal result of the reconstruction is that species-delimitation disagreements can often be restated as disagreements about assumptions rather than as simple conflicts among datasets. MSC analyses illustrate the point sharply: under population structure, increasing genetic information can increase confidence in a partition without establishing that the partition corresponds to separately evolving species, creating the “too many species” problem [15]. A related process-based approach distinguishes the formation of population lineages from completion of

speciation, showing why lineage divergence and species status should not be treated as the same inferential state [16].

Geography is another assumption-bearing variable. Spatial separation and isolation by distance can generate genetic structure that resembles lineage discontinuity, so geography must be represented within the interpretation of evidence rather than added only after a molecular partition has been chosen [17]. This is particularly important for allopatric or sparsely sampled insects, where gaps in collection localities can mimic biological discontinuities. Empirical benchmarking in leaf beetles reinforces the model-dependence of the problem. Across more than 90 COI datasets, the efficiency of molecular delimitation varied among methods and with dataset properties, demonstrating that performance is contingent on both analytical procedure and the structure of the evidence being analyzed [18]. These observations justify a proposed "assumption register" for each candidate model: whether gene flow is excluded or allowed; whether geography is treated as explanatory structure; whether the input represents one locus, multiple loci or genomic regions; what process links gene trees to species histories; and which observations are considered independent corroboration. The register is descriptive rather than a scoring device. Its purpose is to make visible the conditions under which a preferred partition could change.

Model comparison and evidential support

Once candidate species models are explicit, evidential support can be compared without equating support with taxonomic truth. Marginal likelihoods and Bayes factors offer one formal route for comparing specified Bayesian models, but their interpretation remains conditional on the candidate model set, priors and estimation procedure [19]. Quantitative boundary models likewise make delimitation parameters explicit, improving auditability while still depending on the biological and statistical assumptions built into the model [20]. The relevant question is therefore not "Which method found the species?" but "Which candidate model is best supported under the stated assumptions, and which independent observations corroborate or challenge it?"

Concordance matters, but it should be interpreted as a pattern of support rather than as automatic proof. In *Gnaptorina* beetles, molecular delimitation and morphology provided an empirical example in which multiple approaches could converge on a coherent taxonomic interpretation [21]. Such cases are methodologically valuable because they show what triangulation can look like, yet shared sampling design, correlated characters or a common geographic history can reduce the independence of apparently separate evidence streams.

The proposed adjudication rule consequently separates two dimensions: internal model support and external evidential corroboration. A candidate can be statistically favored but weakly corroborated by morphology, geography, ecology or reproductive information; conversely, external evidence can be coherent while competing genetic models remain difficult to distinguish. These are different inferential states and should not be compressed into an unsupported confidence score. **Table 1** is therefore used to distinguish combinations of model support and corroborating evidence that require different interpretations and follow-up tests, rather than reducing them to a ranked scale.

Table 1. Adjudicating candidate species models without converting support into taxonomic certainty

Relative model support			**Taxonomic permission**	
Candidate-model relation	**Taxonomic setting**	**Support pattern**	**Admissible species hypothesis**	**Assumption challenge and corroboration**
Strong internal model support, weak external corroboration	Structured or poorly diagnosed lineages	High relative statistical support but little independent morphological, ecological, geographic or reproductive differentiation	Preferred statistical partition remains biologically provisional. Proposed adjudication state.	Model may be fitting population structure, sparse sampling or an unmeasured process. Acquire independent evidence targeted at the disputed boundary
Moderate model separation, convergent evidence classes	Integrative insect datasets with partially independent characters	Several models receive support, while morphology, geography or ecology favor the same partition	Concordance increases plausibility but does not remove model dependence. Proposed adjudication state.	Evidence classes may share historical or sampling biases. Test whether corroborating characters are genuinely independent

Strong model support under migration-excluding assumptions	Parapatric or admixing lineages	Coalescent support from models that do not represent gene flow	Support is conditional on an assumption likely to be biologically material. Evidence-derived condition; proposed response.	Introgression or ongoing migration can distort the inferred partition. Refit or compare models that permit migration where feasible
Competing models with similar support	Shallow or recently diverged lineages	Marginal likelihoods, posterior support or alternative partitions remain close	The evidence does not justify a uniquely preferred boundary. Proposed adjudication state.	Candidate set, priors and limited information may dominate. Report multiple plausible species models and seek discriminating evidence
Convergent molecular and morphological delimitation	Morphologically diagnosable lineages such as some beetle complexes	Independent analyses recover a compatible partition	Concordance strengthens a working species hypothesis. Evidence-derived pattern.	Concordance is not equivalent to causal evidence of reproductive isolation. Test additional biological criteria when taxonomic consequences are substantial
Model support changes after explicit assumption change	Lineages sensitive to priors, thresholds, geography or preprocessing	Preferred partition differs after modifying a declared analytical choice	Instability is itself a property of the inference. Proposed adjudication rule.	No single output should be presented as assumption-free. Carry instability forward into uncertainty reporting

Discordant evidence

Discordance is not a single phenomenon and should not be treated as a reason to choose whichever evidence class appears most decisive. Phylogenomic work on the *Temnothorax salvini* ant group shows that genome-scale data can refine species hypotheses while still requiring comparison among delimitation methods and morphological information [22]. The implication is not that genomic data are unreliable, but that greater marker density does not eliminate the need to specify how gene histories are translated into species hypotheses.

Biological processes can also generate genuine conflict among otherwise well-measured signals. In the *Adelges piceae* complex, genetic assignment, morphology, reproductive mode and analyses of hybridization were required to interpret boundaries shaped by a complicated invasion and reproductive history [23]. Hybridization is therefore one possible explanation for discordance, not a default diagnosis. In other systems, similar patterns could arise from incomplete lineage sorting, recent divergence, spatial structure, phenotypic conservatism, mitochondrial introgression or uneven sampling. The framework requires these alternatives to remain explicit until evidence can discriminate among them.

Discordance also operates across taxonomic scale. An integrative analysis of Epyrinae combined a large morphological character matrix with multiple molecular markers to resolve a higher-level systematic impasse [24]. This demonstrates that morphological and molecular evidence can be complementary, but it does not license direct transfer of genus-level phylogenetic conclusions to species-level delimitation. The level of inference must remain visible.

The proposed discordance protocol therefore distinguishes three nonexclusive states. Data discordance occurs when evidence classes support different partitions. Model discordance occurs when the same evidence supports different partitions under different assumptions or algorithms. Biological discordance occurs when a plausible process—such as gene flow, hybridization or incomplete lineage sorting—predicts different histories for different evidence classes. These categories are proposed organizational devices rather than validated natural kinds. Their purpose is to prevent “integrative” analysis from becoming an unreported majority vote. A reproducible study should record the competing partition, the evidence class producing it, the process assumptions required to interpret it, and the alternative explanations that remain viable before any taxonomic decision is made.

Sensitivity to priors and thresholds

Sensitivity analysis should begin before the final delimitation algorithm is run. In diploid phylogenomic data, treatment of heterozygous sites and haplotype phase can alter shallow multispecies-coalescent inference, whereas deeper divergences may be more robust under the conditions examined [25]. Preprocessing is therefore part of the inferential model rather than a neutral technical prelude. A species hypothesis that is stable only after one phasing, filtering or locus-selection decision should be reported differently from one that persists across biologically defensible alternatives.

The same principle applies to data design. Experiments on barcode length show that adding sequence length does not produce a universally monotonic improvement in species discovery or identification [26]. The relevant question is consequently not whether the longest available marker was used, but whether the information sampled is adequate for the biological divergence and analytical task under consideration. This distinction becomes particularly important when identification performance is used to justify delimitation. Model-based identification under the MSC concerns assignment of observations to already specified candidate species and should not be interpreted as direct validation of the process by which those species were delimited [27].

Priors and explicit decision thresholds create another source of conditionality. In Haeterini, alternative prior specifications contributed to different candidate coalescent partitions before model comparison [3]. SODA makes this dependence unusually transparent because its significance level is an explicit analytical setting and larger α values permit more aggressive splitting under the method's test logic [14]. Empirical leaf-beetle comparisons likewise demonstrate that algorithmic performance interacts with dataset characteristics rather than remaining constant across applications [18]. These examples do not imply that every prior, threshold or preprocessing decision will materially change every insect delimitation. They establish that such choices can matter and therefore should be tested when scientifically plausible.

The proposed framework defines decision stability as the persistence of a taxonomically relevant partition across a prespecified set of defensible analytical choices. Decision stability is not a numerical confidence grade and does not replace biological corroboration. It records whether changes in priors, thresholds, preprocessing, spatial assumptions or model family alter the preferred candidate model, leave only local boundary changes, or produce no consequential change. Instability is treated as information to be reported rather than as an inconvenience to be hidden.

Because sensitivity can enter at several different points in the analysis, prose alone obscures which perturbations affect data representation, model construction, candidate ranking or taxonomic interpretation. **Table 2** separates these sensitivity states and the responses appropriate to each without assigning an unsupported universal threshold or score.

Table 2. Sensitivity states that can alter the stability of an insect species hypothesis

Perturbation axis	Result instability	What sensitivity does not establish	Robustness response and standing
Prior-sensitive candidate partition. Bayesian MSC or related probabilistic delimitation	Preferred partition changes across biologically defensible prior settings. Support is conditional on prior structure	Prior sensitivity may reflect limited information or genuinely shallow divergence	Report alternative priors and whether taxonomically relevant boundaries persist. Evidence-derived condition; proposed response.
Threshold-sensitive splitting. Tests or clustering rules with explicit decision thresholds	Number or placement of candidate lineages changes when the threshold is varied. The delimitation contains an operational decision boundary	A statistical cutoff is not a biological species threshold	Examine a defensible range and disclose boundary changes. Evidence-derived condition; proposed response.
Preprocessing-sensitive inference. Diploid genomic or multilocus data	Phasing, filtering or locus treatment changes downstream topology or partition. Uncertainty enters before species-model comparison	Effects may be strongest at shallow divergence and need not be universal	Repeat inference under alternative defensible preprocessing choices. Evidence-derived condition; proposed response.
Marker-design sensitivity. Barcode or reduced-representation datasets	Information content changes with marker length or sampling design. Apparent delimitation stability may reflect data design	More sequence is not automatically equivalent to more relevant information	Evaluate data adequacy for the intended discovery or identification task. Evidence-derived condition; proposed response.

Method-sensitive candidate ranking. The same taxon analyzed with different model families	Alternative methods produce different partitions or rankings. Disagreement identifies model dependence	Methods may target different quantities rather than compete on identical terms	Compare assumptions before comparing outputs. Evidence-derived condition; proposed response.
Stable partition across defensible perturbations. Any sufficiently sampled integrative analysis	Taxonomically consequential boundaries remain unchanged across planned perturbations. Analytical robustness is strengthened	Stability does not establish biological species status	Retain external corroboration and taxonomic adjudication. Proposed decision-stability state.

Sensitivity analysis also provides the strongest basis for the article's visual synthesis because perturbations can enter at several analytical levels and propagate differently. **Figure 1** therefore maps how choices made at the data, model and decision layers can alter a candidate species hypothesis, while keeping biological corroboration and formal taxonomic action outside the same inferential boundary.

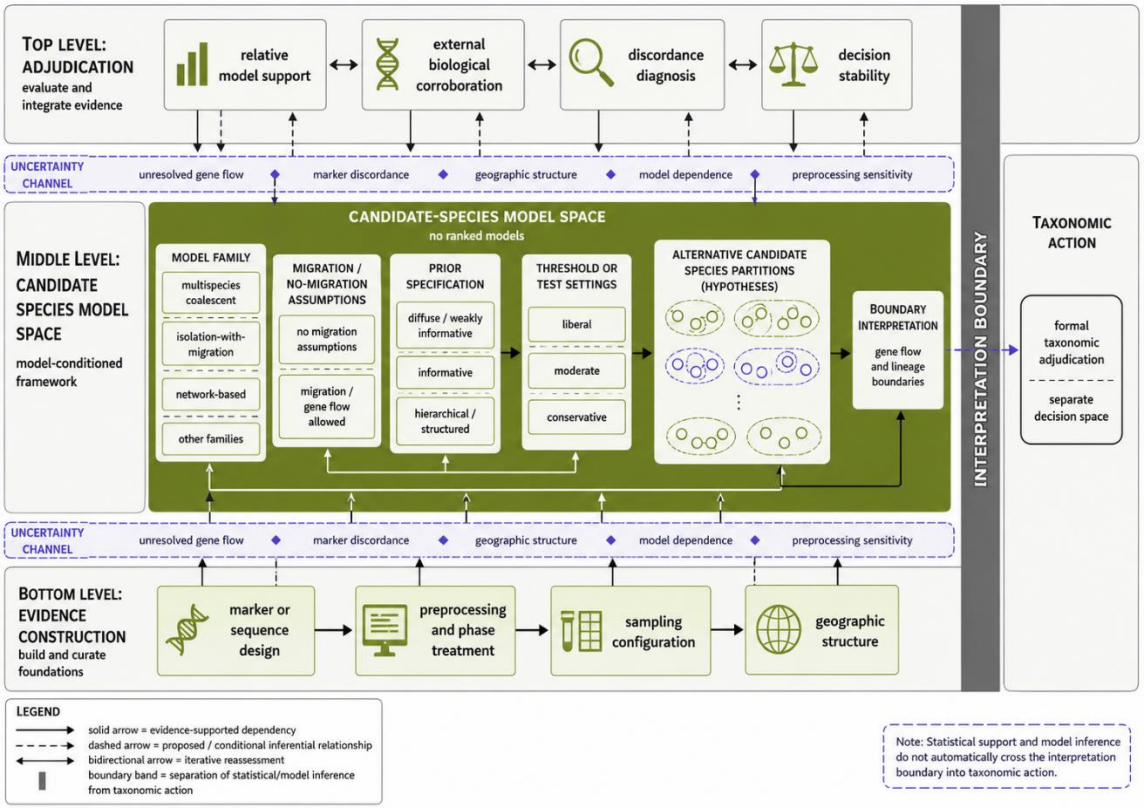


Figure 1. Assumption-sensitive species boundaries across evidence, model, and taxonomic decision layers

Reporting uncertainty

Reproducibility requires more than listing software names. Reanalysis of a large barcode-based taxonomic revision showed why cluster identifiers, algorithmic procedures, accessible data and the distinction between BINs and named species are consequential for independent reassessment [28]. The generalizable lesson is not that barcode clustering lacks value, but that a species claim becomes difficult to audit when the transformation from raw observations to named entities cannot be reconstructed.

The proposed framework separates uncertainty into four reporting domains. Evidence uncertainty concerns whether sampled markers, morphology, ecology, geography or reproductive information adequately represent the relevant lineage variation. Model uncertainty concerns whether plausible alternative models encode different evolutionary processes or candidate partitions. Decision-setting uncertainty concerns priors, thresholds, preprocessing or other analytical choices that can change outcomes. Taxonomic-action uncertainty concerns whether the supported evidence is sufficient to justify nomenclatural or classificatory action. These domains are proposed organizational categories; they are not validated certainty grades.

This separation prevents posterior probability from becoming a surrogate for total taxonomic confidence. Population structure can receive strong coalescent support without resolving the species-level interpretation [1]. Gene flow can maintain continuity that complicates a crisp boundary [7]. Marginal-likelihood comparisons remain conditional on the candidate models and priors evaluated [19], while shallow genomic inference can depend on preprocessing decisions such as phasing [25]. A transparent report should therefore state not only which model was favored, but which plausible alternatives were considered, which settings changed the conclusion, which evidence classes remained discordant and which taxonomic consequences remain conditional.

Because these uncertainty domains originate at different stages and demand different remedies, **Table 3** separates their mechanisms and reporting boundaries rather than collapsing them into a single confidence label.

Table 3. Mechanisms and reporting boundaries for uncertainty in integrative species delimitation

Where uncertainty arises			What may be reported		
Uncertainty domain	**Entry point**	**Manifestation**	**Bounded claim**	**Boundary retained**	**Reproducibility record**
Evidence uncertainty	Sampling and evidence construction	Sparse geography, limited loci, conflicting morphology, marker-specific structure or missing biological evidence	The available evidence supports only the sampled dimensions of divergence	Absence of evidence is not evidence of biological equivalence	Archive raw inputs, sampling design and evidence provenance. Proposed reporting category.
Model uncertainty	Translation of observations into lineage models	Plausible models produce different candidate partitions or explanations	Support is relative to the tested candidate set	Preferred model is not an assumption-free species boundary	Report candidate models and their biological assumptions. Proposed reporting category.
Decision-setting uncertainty	Priors, thresholds and preprocessing	Taxonomically meaningful boundaries shift under defensible settings	The result is setting-sensitive	Default settings are not biological standards	Record priors, thresholds, preprocessing and software versions. Proposed reporting category.
Discordance uncertainty	Comparison among evidence classes	Mitochondrial, nuclear, morphological, geographic or reproductive signals disagree	Competing explanations remain viable	Discordance does not identify its own cause	Preserve evidence streams separately and document adjudication. Proposed reporting category.
Taxonomic-action uncertainty	Transition from supported hypothesis to nomenclatural decision	Candidate model is supported but independent diagnosis or corroboration remains incomplete	A working lineage hypothesis may be reported	Statistical delimitation and formal naming are separate acts	State what additional evidence or judgement supports taxonomic action. Proposed reporting category.
Residual uncertainty after sensitivity analysis	Final adjudication	Partition remains stable but unresolved biological alternatives persist	Analytical stability can be claimed within tested conditions	Robustness is not equivalent to universal biological validity	Report what was tested and what remained untested. Proposed reporting category.

Reproducible species-delimitation workflow

Large-scale integrative taxonomy demonstrates why candidate discovery and taxonomic validation need not occur in a single analytical step. A scalable approach can use high-throughput evidence to generate initial hypotheses and then direct additional effort toward the boundaries where validation is most informative [29]. Conversely, methodological criticism of barcode-only descriptions emphasizes that molecular clustering should not

automatically be converted into formal species names without sufficient independent taxonomic justification [30]. Together, these positions support a staged workflow without requiring a universal evidentiary recipe.

The proposed workflow begins by defining the taxonomic question and specifying plausible candidate species models before selecting a preferred output. Evidence classes are then recorded separately, including their geographic and sampling context. Candidate-generation tools may identify partitions efficiently, as in distance-based approaches such as ASAP [13], but the resulting clusters remain hypotheses. Alternative model families are then fitted or compared where appropriate, with assumptions concerning migration, coalescent processes, priors and decision thresholds made explicit. When a method contains an adjustable decision rule, its sensitivity should be documented rather than hidden behind defaults [14].

Discordance initiates reassessment rather than forced consensus. The investigator asks whether the conflict is best characterized as evidence discordance, model discordance or a plausible biological process requiring additional analysis. Sensitivity testing then determines whether taxonomically consequential boundaries survive defensible perturbations. Reporting retains competing plausible models and all residual uncertainty. Reproducibility additionally requires sufficient access to underlying evidence and analytical procedures for an independent investigator to reconstruct how clusters or model outputs became candidate species hypotheses [28].

Only after these stages does formal taxonomic adjudication occur. That boundary is deliberately non-automatic. A statistically preferred species model can warrant additional collection, morphological reassessment, examination of reproductive biology or publication as a provisional hypothesis rather than immediate naming. Conversely, concordant evidence may justify action even when no single model supplies an overwhelming probability, provided the inferential basis and remaining limitations are transparent. The framework therefore standardizes disclosure and comparison, not the biological content of every species decision.

Treating species delimitation as model choice changes the unit of methodological transparency. Instead of asking readers to accept a final partition because a familiar algorithm generated it, the analysis exposes the candidate models that were considered, the assumptions that distinguish them and the evidence that changes their relative plausibility. The Haeterini example demonstrates that taxonomic alternatives and priors can be made explicit and formally compared [3], while general work on marginal likelihoods establishes the statistical logic for comparing specified Bayesian models without implying that the favored model exhausts biological reality [19].

The framework also clarifies why stronger statistical support need not equal stronger species evidence. MSC methods may confidently distinguish population structure [1], and the empirical relationship between coalescent boundaries and reproductive isolation is substantial but incomplete even in a system where both can be evaluated [2]. Process-oriented delimitation further emphasizes that population-lineage formation and species completion need not be the same event [16]. Thus, the proposed separation between model support and external corroboration is not an attempt to impose one species concept; it is a device for showing which biological interpretation is being asked of each dataset.

Insect studies demonstrate that both agreement and disagreement matter. Molecular and morphological evidence can converge, as in *Gnaptorina* [21], while phylogenomic analysis of *Temnothorax* still requires integration across models and character systems [22]. Hybridization and reproductive complexity can reorganize the interpretation of adelgid clusters [23], and combined morphological and molecular evidence can resolve higher-level systematic problems without implying that the same inferential rules transfer directly to species-level questions [24]. Integrative systematics should therefore preserve the taxonomic scale and causal modesty of each evidence class.

The framework is intentionally not a universal decision algorithm. Its applicability will vary with taxon biology, sampling density, life history, reproductive mode, geography and available character systems. High-throughput programs dealing with dark taxa may benefit particularly from separating candidate discovery from targeted validation [29], but the evidentiary demands of a well-known, sexually reproducing butterfly complex and a poorly sampled parasitoid lineage need not be identical. Likewise, criticism of barcode-only naming supports an inference–taxonomic-action boundary [30] without implying that DNA barcodes lack diagnostic or discovery value.

Validation of the proposed framework should therefore focus on reproducibility and decision stability rather than on claiming a universal accuracy benchmark. Prospective comparisons could ask whether explicit candidate-model declaration reduces unreported analytical flexibility, whether independent analysts can reconstruct the same decision path, and whether sensitivity analyses identify taxonomically consequential instability before nomenclatural action. Empirical benchmarking already shows that molecular-delimitation performance varies with method and dataset characteristics [18], preprocessing experiments demonstrate context-dependent

sensitivity [25], and independent reanalysis shows why accessible procedures matter for taxonomic reproducibility [28]. Those findings justify testing the architecture; they do not validate its performance in advance.

The main limitation is that transparency cannot eliminate biological indeterminacy. Recent divergence, introgression, incomplete lineage sorting, geographic structure, sparse sampling and phenotypic conservatism may leave several candidate models defensible. Under such conditions, the framework is designed to preserve rather than disguise uncertainty. Its success should be judged by whether another systematist can recover why a model was favored, what would have changed that decision and where statistical inference ended and taxonomic judgement began.

Conclusion

Species delimitation becomes more reproducible when its assumptions are treated as part of the result rather than as invisible properties of software. The methodological framework developed here separates observed evidence, candidate species models, comparative support, discordance, sensitivity, residual uncertainty and formal taxonomic action. It does not prescribe a universal species concept, evidentiary weighting scheme or decision threshold. Instead, it makes competing interpretations inspectable and requires instability to remain visible when biologically plausible analytical choices change the preferred boundary. For integrative insect systematics, this model-choice perspective offers a disciplined way to combine molecular, morphological, geographic, ecological and reproductive evidence without pretending that their agreement is automatic or that statistical support alone constitutes taxonomic proof. The framework remains a proposed methodological synthesis requiring prospective evaluation across contrasting insect systems and evidence regimes.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. 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
2. 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
3. Matos-Maraví P, Wahlberg N, Antonelli A, Penz CM. Species limits in butterflies (Lepidoptera: Nymphalidae): Reconciling classical taxonomy with the multispecies coalescent. Syst Entomol. 2019;44(4):745-56. doi:10.1111/syen.12352
4. 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
5. Núñez R, Genaro JA, Pérez-Asso A, Murillo-Ramos L, Janzen DH, Hallwachs W, et al. Species delimitation and evolutionary relationships among Phoebis New World sulphur butterflies (Lepidoptera, Pieridae, Coliadinae). Syst Entomol. 2020;45(2):481-92. doi:10.1111/syen.12408
6. Gueuning M, Frey JE, Praz C. Ultraconserved yet informative for species delimitation: Ultraconserved elements resolve long-standing systematic enigma in Central European bees. Mol Ecol. 2020;29(21):4203-20. doi:10.1111/mec.15629
7. Joshi M, Espeland M, Dincă V, Vila R, Tahami MS, Dietz L, et al. Delimiting continuity: Comparison of target enrichment and double digest restriction-site associated DNA sequencing for delineating admixing parapatric Melitaea butterflies. Syst Entomol. 2022;47(4):637-54. doi:10.1111/syen.12557
8. 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

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. Mirarab S, Nakhleh L, Warnow T. Multispecies coalescent: Theory and applications in phylogenetics. Annu Rev Ecol Evol Syst. 2021;52:247-68. doi:10.1146/annurev-ecolsys-012121-095340
11. Flouri T, Jiao X, Rannala B, Yang Z. Species tree inference with BPP using genomic sequences and the multispecies coalescent. Mol Biol Evol. 2018;35(10):2585-93. doi:10.1093/molbev/msy147
12. Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamatakis A, et al. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. Bioinformatics. 2017;33(11):1630-8. doi:10.1093/bioinformatics/btx025
13. Puillandre N, Brouillet S, Achaz G. ASAP: Assemble species by automatic partitioning. Mol Ecol Resour. 2021;21(2):609-20. doi:10.1111/1755-0998.13281
14. Rabiee M, Mirarab S. SODA: Multi-locus species delimitation using quartet frequencies. Bioinformatics. 2020;36(24):5623-31. doi:10.1093/bioinformatics/btaa1010
15. 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
16. Sukumaran J, Holder MT, Knowles LL. Incorporating the speciation process into species delimitation. PLoS Comput Biol. 2021;17(5). doi:10.1371/journal.pcbi.1008924
17. Hausdorf B, Hennig C. Species delimitation and geography. Mol Ecol Resour. 2020;20(4):950-60. doi:10.1111/1755-0998.13184
18. 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
19. Oaks JR, Cobb KA, Minin VN, Leaché AD. Marginal likelihoods in phylogenetics: A review of methods and applications. Syst Biol. 2019;68(5):681-97. doi:10.1093/sysbio/syz003
20. 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
21. Li XM, Bai XL, Kergoat GJ, Pan Z, Ren GD. Phylogenetics, historical biogeography and molecular species delimitation of Gnaptorina Reitter (Coleoptera: Tenebrionidae: Blaptini). Syst Entomol. 2021;46(1):239-51. doi:10.1111/syen.12459
22. 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
23. Havill NP, Griffin BP, Andersen JC, Foottit RG, Justesen MJ, Caccone A, et al. Species delimitation and invasion history of the balsam woolly adelgid, Adelges (Dreyfusia) piceae (Hemiptera: Aphidoidea: Adelgidae), species complex. Syst Entomol. 2021;46(1):186-204. doi:10.1111/syen.12456
24. Colombo WD, Tribull CM, Waichert C, Azevedo CO. Integrative taxonomy solves taxonomic impasses: A case study from Epyrinae (Hymenoptera, Bethylidae). Syst Entomol. 2022;47(3):504-29. doi:10.1111/syen.12544
25. Huang J, Bennett J, Flouri T, Yang Z. Phase resolution of heterozygous sites in diploid genomes is important to phylogenomic analysis under the multispecies coalescent model. Syst Biol. 2022;71(2):334-52. doi:10.1093/sysbio/syab047
26. Yeo D, Srivathsan A, Meier R. Longer is not always better: Optimizing barcode length for large-scale species discovery and identification. Syst Biol. 2020;69(5):999-1015. doi:10.1093/sysbio/syaa014
27. Yang Z, Rannala B. Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. Mol Ecol. 2017;26(11):3028-36. doi:10.1111/mec.14093
28. Meier R, Blaimer BB, Buenaventura E, Hartop E, von Rintelen T, Srivathsan A, et al. A re-analysis of the data in Sharkey et al.'s (2021) minimalist revision reveals that BINs do not deserve names, but BOLD Systems needs a stronger commitment to open science. Cladistics. 2022;38(2):264-75. doi:10.1111/cla.12489
29. 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
30. 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