Eurasia Specialized Veterinary Publication

Entomological Research Letters

ISSN: 3108-639X

2021, Volume 1, Issue 2, Page No: 51-60
Copyright CC BY-NC-SA 4.0
Available online at: www.esvpub.com/

Rapid Insect Radiations Expose Predictable Errors When Gene Trees Are Interpreted Directly as Species Histories

Kristian Eriksen[1]*, Solveig Berg[1], Magnus Dahl[2]

[1]Department of Phylogenomics and Rapid Radiations, Faculty of Life Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
[2]Department of Molecular Evolution and Tree Interpretation, Faculty of Biosciences, University of Bergen, Bergen, Norway.

***E-mail** ✉ kristian.eriksen@ntnu.no

ABSTRACT

Rapid insect radiations are difficult phylogenetic problems not simply because internal branches are short, but because several biological and analytical processes can produce the same visible symptom: discordant gene trees. This phylogenetic framework argues that a recurrent error occurs when an estimated gene tree is read directly as a species history. The framework separates four inferential objects—observed molecular characters, gene-copy genealogies, estimated gene trees, and species-history hypotheses—and classifies discordance by where it enters that chain. Incomplete lineage sorting and introgression are treated as biological generators of discordance; sampling, orthology, paralogy, sequence fragmentation, and upstream processing can distort the evidence supplied to tree estimation; and model misspecification can turn uncertainty into directional error. The framework therefore treats discordance as a diagnostic problem requiring competing explanations rather than as a label with a unique cause. Its principal contribution is a proposed error-source architecture for rapid insect radiations that preserves uncertainty, taxonomic boundaries, and the distinction between phylogenetic support and historical truth. It is intended as an interpretive framework, not a validated classifier of error causes.

Keywords: Insect phylogenomics, Gene-tree discordance, Species-tree inference, Incomplete lineage sorting, Introgression, Orthology

Received: 08 October 2021
Revised: 23 November 2021
Accepted: 26 November 2021

How to Cite This Article: Eriksen K, Berg S, Dahl M. Rapid Insect Radiations Expose Predictable Errors When Gene Trees Are Interpreted Directly as Species Histories. Entomol Res Lett. 2021;1(2):51-60.
https://doi.org/10.51847/DXFePla01s

Introduction

Genome-scale sequencing has transformed insect systematics, yet greater data volume does not remove the conceptual distinction between gene histories and species histories. Phylogenomic studies still depend on locus choice, homology assessment, gene-tree estimation, model adequacy, and the biological processes that create real genealogical heterogeneity [1]. The central problem addressed here is therefore not whether genomic data are informative, but when the path from molecular observations to species-history interpretation is being shortened illegitimately.

This distinction matters especially in insects because available molecular evidence is uneven across clades, markers, and taxonomic scales. Broad surveys of insect sequence resources show that data abundance is heterogeneous and that expanded databases do not produce uniform taxonomic or marker coverage [2]. A large matrix can therefore be information-rich overall while remaining sparse or unbalanced around the lineage split that matters most.

© 2021 Entomological Research Letters

Rapid radiations intensify the problem. In pimpliform parasitoid wasps, hybrid-capture data substantially improved resolution, yet very short branches associated with a rapid radiation remained difficult and conflict-prone [3]. That example supports a narrow conclusion: short internodes can preserve hard phylogenetic problems despite extensive targeted data. It does not establish incomplete lineage sorting as the unique cause of every difficult branch.

Analytical treatment can also alter what appears to be biological conflict. In Auchenorrhyncha, interrogation of gene conflict and alternative coding strategies changed the evidential interpretation of contentious relationships [4]. Such sensitivity means that disagreement among trees may partly reflect how signal is represented and modeled, not only how ancestral lineages diverged.

Finally, estimated gene trees themselves are uncertain objects. Empirical work on ultraconserved elements in *Pseudapis* bees showed that short loci can generate substantial gene-tree estimation error, so observed discordance can mix biological lineage history with finite-data uncertainty [5]. The article therefore develops one organizing claim: in rapid insect radiations, gene-tree discordance should be diagnosed across inferential levels before any topology is interpreted as a species history.

Materials and Methods

This article uses a conceptual classification procedure rather than a new empirical dataset. Error sources are organized by the stage at which they can alter the path from molecular evidence to species-history interpretation and by whether the source is primarily biological, data-generating, estimation-related, model-related, or interpretive. This structure is motivated by comparative phylogenomic work showing that ancient hybridization, incomplete lineage sorting, and gene-tree error can produce overlapping patterns that require competing explanations [6].

A second criterion is sensitivity to analytical representation. Reanalysis of conflicting phylogenomic branches has shown that concatenation and quartet-based approaches can disagree because gene-wise signal is heterogeneous and estimated gene trees contain error [7]. Method disagreement is therefore classified as a diagnostic observation, not as evidence that one biological process has already been identified.

A third criterion is response to bias-aware treatment. Studies of historically contentious relationships demonstrate that topology or support can change when data partitions, models, or systematic-error controls are altered [8]. In this framework, such instability elevates model or data error as a competing hypothesis; it does not by itself prove which alternative topology is historically correct. The resulting classification is explicitly proposed as an interpretive architecture and is not assigned sensitivity, specificity, or predictive accuracy.

Results and Discussion

The first result of the synthesis is an object distinction. A gene copy has its own history, an estimated gene tree is a statistical reconstruction of that history, and a species tree is a hypothesis about lineage divergence at another biological level. Paralogy-aware species-tree methods make this separation explicit by allowing gene-copy histories to differ from species relationships rather than assuming every gene tree is a one-to-one orthologous species-tree replicate [9].

Gene-family evolution adds another layer. Models that combine duplication, loss, and coalescent processes show formally that several genealogical histories can be nested within a species history [10]. This means that even a perfectly estimated gene-family tree need not have the same topology as the species tree, because non-equivalence can be biological rather than erroneous.

Large insect datasets reinforce the distinction between information quantity and inferential object. Whole-genome shotgun phylogenomics of swallowtail butterflies resolved many relationships, but locus recovery, filtering, and analytical choices remained relevant to difficult branches [11]. The species-history estimate is therefore an inference assembled from many conditioned data transformations, not a direct readout of the genome.

The same point begins even earlier in the pipeline. Anchored phylogenomics in Lepidoptera depends on explicit locus recovery, alignment, and homology procedures before any phylogenetic model is fitted [12]. I therefore propose a four-object separation for the remainder of the article: observed molecular characters, inferred gene-copy genealogies, estimated gene trees, and species-history hypotheses. Errors can enter between these objects, while biological discordance can arise because the objects legitimately encode different histories.

Incomplete lineage sorting

Incomplete lineage sorting is the clearest demonstration that gene-tree/species-tree disagreement need not be analytical error. Under short consecutive internodes and sufficiently strong ancestral lineage persistence, coalescent theory allows the most probable gene-tree topology to differ from the species-tree topology [13]. Rapid diversification therefore creates an ILS-compatible context, but rapidity alone does not establish that a particular insect clade occupies an anomaly-zone regime.

The diagnostic difficulty is compounded because gene-tree estimation error can interact with branch-length structure. Theory shows that bounded locus length and adverse branch configurations can create systematic species-tree bias rather than merely random noise [14]. A discordant gene-tree distribution may consequently contain both coalescent variation and estimation-induced distortion.

Data treatment can further change that mixture. Gene filtering may improve or reduce species-tree accuracy depending on the relationship among information content, gene-tree error, and the underlying coalescent regime [15]. Removing weak or discordant loci is therefore not a neutral cleaning operation; it changes the evidence distribution that a summary method is asked to interpret.

Fragmentary sequences create a related risk. Simulation and empirical reanalysis have shown that highly fragmentary loci can degrade gene-tree reconstruction and downstream species-tree estimation, while not all missing data have equivalent effects [16]. The framework consequently treats ILS as a biological hypothesis that must be evaluated alongside the quality of the gene trees used to infer it. The scientific question is not simply whether discordance exists, but whether its distribution remains ILS-compatible after estimation quality and data-treatment sensitivity are examined.

Because these scenarios can produce superficially similar conflict while implying different corrective actions, **Table 1** separates coalescent interpretation from estimation and data-treatment boundaries.

Table 1. When an ILS-compatible pattern is not yet an ILS diagnosis

Observed radiation		**ILS adjudication**		
Radiation regime	**Gene-tree signature**	**Biological permission**	**ILS diagnostic gate**	**Epistemic standing**
Coalescent regimes				
Short-internode coalescent compression. Rapid successive lineage splitting	Multiple gene-tree topologies around short internal branches	ILS is biologically plausible	Rapid branching alone does not establish anomaly-zone conditions. Estimate branch/coalescent conditions before assigning cause	Established condition; proposed diagnostic label
Anomaly-zone-compatible regime. Consecutive short internodes with strong ancestral lineage persistence	Dominant gene tree can differ from species-tree topology	Gene-tree majority may mislead	Applicability is conditional on the coalescent regime. Test whether estimated parameters support the required regime	Established theory; proposed diagnostic placement
Estimation and processing amplification				
Estimation-amplified discordance. Short or weakly informative loci	Discordance rises as gene trees become poorly determined	Observed conflict may exceed biological ILS	Biological and estimation variance are confounded. Reassess locus information and gene-tree uncertainty	Established interaction; proposed diagnostic label
Filtering-sensitive discordance. Selective inclusion or removal of loci	Species-tree result changes with gene filtering	Filtering is part of inference, not neutral cleanup	No universal filtering rule is supported. Compare defensible filtering regimes without imposing a fixed threshold	Established sensitivity; proposed organization
Fragmentation-sensitive discordance. Uneven or highly incomplete locus sequences	Unstable gene trees and downstream species-tree changes	Missingness can amplify apparent conflict	Effects depend on missing-data pattern and method. Test robustness to fragmentary loci and taxon-specific missingness	Established sensitivity; proposed organization

Note: The scenario labels organize established mechanisms and method sensitivities into the article's proposed diagnostic language; they are not a quantitative score and do not assign a unique cause to any observed radiation.

Introgression and hybridization

Incomplete lineage sorting is not the only biological reason that different genomic regions can carry different histories. In the *Heliconius* radiation, comparative genomic evidence showed extensive introgression and genomic heterogeneity, demonstrating that a rapid butterfly radiation can contain reticulate ancestry for which no single local genealogy summarizes the entire species history [17]. This is direct insect evidence for mosaic ancestry, but it does not imply that every rapid insect radiation is reticulate or that introgression erases species boundaries.

Reticulation can also be modeled jointly with lineage sorting. Bayesian implementations of the multispecies coalescent with introgression provide a formal alternative to forcing all discordance into a strictly bifurcating history [18]. Such models separate parameterized introgression from coalescent variation under their assumptions, yet model fit alone cannot establish a unique historical network.

Allele-sharing statistics offer another evidence class. D-statistic approaches can identify excess allele sharing compatible with admixture, but their interpretation depends on population sampling, topology, variant quality, and alternative demographic histories [19]. A significant asymmetry is therefore evidence for an introgression-compatible pattern rather than a complete reconstruction of direction, timing, or mechanism.

Network coalescent theory explains why this distinction is necessary: hybridization can alter expected gene-tree distributions in ways that overlap with patterns generated by lineage sorting [20]. The resulting identifiability problem makes “discordance equals introgression” as unsafe as “discordance equals ILS.” In this article, introgression is therefore treated as a competing species-history hypothesis that requires convergence among genomic distribution, population-level allele-sharing evidence, and model-aware comparison with lineage sorting and analytical error.

The key inferential boundary is best represented visually: **Figure 1** keeps observed genomic discordance, candidate biological histories, and downstream species-history interpretation in separate lanes so that allele sharing or local-tree conflict cannot cross directly into a reticulation conclusion.

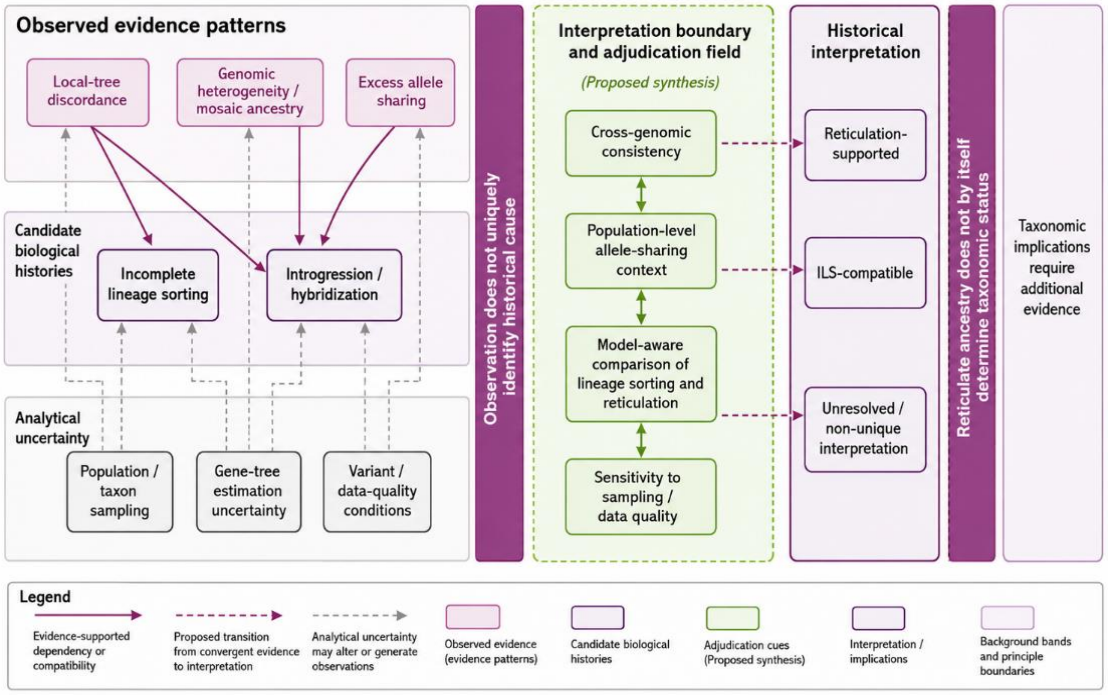


Figure 1. Reticulation signals cross an inference boundary before they become species-history claims

Sampling and orthology errors

Species-tree inference can also fail because the available data do not represent the relevant lineages or loci evenly. Large-scale Lepidoptera phylogenomics shows that expanding taxon and locus sampling can materially refine evolutionary reconstruction [21]. Sampling design is therefore part of the inferential model: missing lineages, uneven genomic coverage, or sparse representation around a rapid split can change which histories are recoverable.

The same principle applies to large transcriptomic datasets. Polyneopteran phylogenomics based on thousands of protein-coding genes resolved many relationships, yet difficult deep divergences still required explicit data processing and sensitivity to heterogeneous signal [22]. More genes increase evidence volume, but they do not guarantee that all taxa, genes, or histories are comparably represented.

Orthology errors are particularly consequential because they change the object being compared. In lice, compositional heterogeneity and alternative bias-mitigation treatments altered support for contentious relationships [23]. Although composition is not itself an orthology error, it illustrates a broader upstream rule: properties of the recovered sequence matrix can systematically redirect downstream topology before a species-history interpretation is made.

Upstream processing can introduce still another layer of error. Genotype-calling choices can propagate systematic differences into phylogenetic matrices and tree estimates under some data regimes [24]. The proposed framework therefore places taxon sampling, locus recovery, homology or orthology assignment, sequence completeness, and genotype construction before gene-tree estimation in the diagnostic chain. Their effects should be tested as evidence-generation problems rather than misread as direct signatures of speciation, introgression, or lineage sorting.

Table 2 is needed because these pre-tree failures differ in what is under-covered, what observable warning they create, and what kind of additional analysis can close the gap.

Table 2. Coverage gaps that can masquerade as species-history conflict before a gene tree is estimated

Evidence construction		Repair before tree comparison	
Pre-tree coverage fault	**Warning produced**	**Distortion path**	**Coverage repair**
Taxon-sampling sparsity. Lineages adjacent to a rapid split	Long isolated branches or unstable placements with added taxa	Alternative historical arrangements remain insufficiently sampled. Dense sampling can improve inference without guaranteeing correctness	Add strategically placed taxa and compare topology stability. Established sampling principle; proposed map placement.
Locus-recovery imbalance. Genes or genomic regions recovered unevenly among taxa	Conflict concentrates where matrix occupancy is poor	Apparent lineage conflict may partly reflect unequal evidence. Missingness pattern matters more than a single completeness percentage	Compare locus occupancy and topology across recovery strata. Proposed diagnostic organization.
Homology/orthology misassignment. Non-equivalent gene copies treated as comparable loci	Unexpected gene-family placement or persistent locus-specific conflict	Gene-copy history is mistaken for species history. Paralogy and orthology uncertainty can coexist with genuine ILS	Re-establish gene-family relationships before species-tree summarization. Established distinction; proposed map placement.
Sequence fragmentation. Partial loci with limited phylogenetic information	Weak or unstable estimated gene trees	Estimation error is propagated into species-tree methods. Not all missing data are equally harmful	Test fragment-sensitive loci separately and by taxon. Established sensitivity; proposed organization.
Composition-sensitive signal. Sequence classes with heterogeneous base or amino-acid composition	Topology or support changes under composition-aware treatment	Systematic signal can mimic corroborated history. Corrective treatments need not identify a unique true topology	Compare defensible composition-aware models or recodings. Established sensitivity; proposed map placement.
Genotype-construction sensitivity. Variant calls affected by caller, coverage, or heterozygosity	Tree changes across plausible genotype-processing choices	Error originates before phylogenetic estimation. Caller disagreement is not biological evidence	Repeat inference across justified calling pipelines. Established upstream error; proposed map placement.

Note: The map distinguishes coverage and representation gaps from biological discordance. Its row labels are an article-specific proposed organization of source-supported failure modes, not universal diagnostic categories or quality thresholds.

Model misspecification

Gene-tree conflict can also arise when an analysis repeatedly favors the wrong topology. Variation among mitochondrial gene-tree estimates has demonstrated that substantial topological heterogeneity may occur even among loci expected to share the same underlying genealogy, providing a control against interpreting all gene-

tree variation as biological [25]. The relevant distinction is therefore between stochastic uncertainty and systematic error: the latter can create reproducible structure in conflict.

Substitution-model choice is one point of vulnerability. Model-selection procedures can improve fit among candidate models, but the best-supported candidate is not equivalent to the true generative process [26]. Model adequacy must consequently remain separate from model ranking, particularly when rapid radiations combine short internal branches with heterogeneous sequence evolution.

Long-branch attraction illustrates how misspecification can become directional. Under particular model and branch-length conditions, systematic bias can favor incorrect groupings rather than merely reducing support [27]. The proposed framework therefore treats topology stability under one model as insufficient. A difficult branch should instead be challenged with defensible alternative models, data treatments, and composition-sensitive analyses before residual disagreement is assigned to biological lineage history.

For rapid insect radiations, this distinction changes the interpretation of reproducibility. Repeated recovery of one branch from the same alignment, coding scheme, and substitution assumptions demonstrates procedural stability, but it does not constitute an independent replication of the historical signal. The proposed error-source classification therefore treats cross-model and cross-representation persistence as more informative than repeated support calculated within a single analytical specification. Even this stronger persistence is a robustness observation, not proof that an alternative evolutionary process has been excluded.

Diagnostic error map for rapid radiations

The preceding sections imply that no single discordance statistic can identify its own cause. The proposed Diagnostic Error Map therefore begins with the observed failure pattern and moves through competing explanations while preserving the boundary between evidence and historical interpretation. High resampling support cannot close that process because bootstrap support may become large for a topology favored under misspecification [28]. Support is evidence about stability conditional on an analysis, not an independent validation of species history.

Improved bootstrap algorithms sharpen that conditional support estimate but do not remove the boundary. UFBoot2 improves ultrafast-bootstrap approximation and robustness, yet its output still depends on the alignment, model, and topology-search problem being evaluated [29]. The map consequently treats support as one diagnostic descriptor rather than a taxonomic threshold.

The final diagnostic dimension is signal concentration. Comparative phylogenomic reanalysis has shown that a small number of genes can disproportionately drive contentious relationships in otherwise large matrices [30]. This does not make influential genes erroneous; concentrated support can reflect strong historical signal, model sensitivity, or unusual locus properties. The proposed map therefore asks whether a conclusion survives four sequential challenges: evidence-generation quality, gene-tree estimation, competing biological histories, and model adequacy. Failure at any challenge reopens earlier explanations rather than assigning a score.

The map is deliberately non-compensatory. Strong performance at one challenge cannot cancel a serious failure at another: excellent bootstrap support cannot repair a homology mistake, a reticulation-compatible signal cannot excuse poor taxon coverage, and a well-fitting substitution model cannot make an estimated gene tree identical to species history. Conversely, a failed challenge does not automatically falsify the favored topology. It changes the interpretation from comparatively robust to unresolved until the specific competing explanation is examined. This rule is proposed as an inference safeguard rather than a validated decision algorithm.

A comparative adjudication matrix is needed because the same visible conflict can require different tests depending on where it first becomes unstable.

Table 3. Adjudicating look-alike discordance patterns across a rapid-radiation inference chain

Branch warning	Generating context	Observed basis	Histories still live	Premature-closure blocker	Adjudication route
Support-stable but model-sensitive branch	Strong resampling support under one analysis	Topology changes under defensible model or composition treatment	Systematic error remains a live explanation	High support is conditional on the fitted analysis	Compare model adequacy and topology across justified treatments. Proposed adjudication rule.

Gene-tree-dispersed branch	Many alternative estimated gene trees	Conflict persists across loci	ILS, introgression, and estimation error remain competing explanations	Discordance frequency does not identify mechanism	Examine gene-tree information, coalescent conditions, and reticulation-compatible evidence. Proposed adjudication rule.
Reticulation-compatible branch	Mosaic local genealogies or excess allele sharing	Genomic regions or populations show non-tree-like signal	Introgression is supportable if alternatives are challenged	Allele sharing alone does not specify direction, timing, or species status	Test sampling sensitivity and explicit ILS/reticulation models. Evidence-based condition; proposed adjudication rule.
Coverage-sensitive branch	Taxa, loci, or gene copies are unevenly represented	Topology changes with sampling, recovery, or homology treatment	Species history remains unresolved at the affected split	Added data are not equivalent to representative data	Restore coverage and reassess homology before tree comparison. Proposed adjudication rule.
Influential-locus branch	A small subset supplies disproportionate branch support	Gene-wise influence is highly concentrated	The favored topology may be real but is not uniformly corroborated	Influence is not synonymous with error	Inspect influential loci for model fit, homology, information, and biological coherence. Evidence-based condition; proposed adjudication rule.

Note: The matrix is a proposed qualitative adjudication device. It assigns neither probabilities nor universal thresholds and permits an unresolved outcome when competing explanations cannot be discriminated.

The architecture of this diagnosis is shown in **Figure 2** because its central contribution is the order in which competing explanations remain open, cross-linked, or closed rather than a linear sequence of preferred methods.

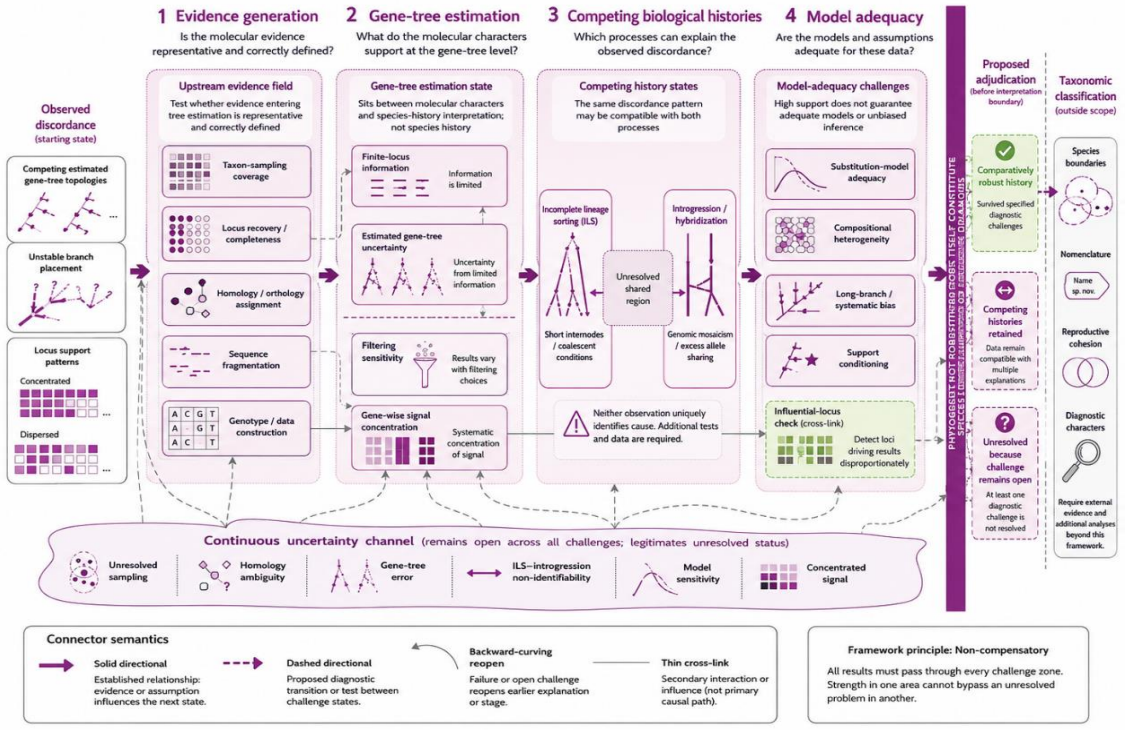


Figure 2. Discordance trajectories remain reversible until evidence crosses the species-history interpretation boundary

Implications for classification

Phylogenetic reconstruction and classification should remain connected but non-identical decisions. Broad Lepidoptera phylogenomics demonstrates that expanded sampling can materially revise evolutionary reconstruction [21], yet such reconstruction does not itself delimit species or authorize nomenclatural change.

Classification should therefore record which relationships are robust to the relevant error challenges rather than translating a preferred topology directly into taxonomic certainty.

Reticulation makes this safeguard especially important. Genomic analysis of the *Heliconius* radiation demonstrates extensive introgressed ancestry [17]. Evidence of gene flow should therefore alter the model of history before it is used to infer that named lineages lack taxonomic distinctness.

Support should be interpreted similarly. High bootstrap values can remain conditional on model assumptions [28], while contentious branches may be disproportionately driven by a few genes [30]. The proposed classification safeguard is thus qualitative: taxonomic action should be separated from topology support and should disclose unresolved model, sampling, orthology, ILS, or introgression alternatives where they remain material.

This separation also protects against a common scale shift. A branch may be well supported as an historical relationship among sampled lineages while species boundaries, diagnostic characters, reproductive cohesion, or nomenclatural consequences remain insufficiently evaluated. The framework therefore does not convert phylogenomic robustness into a universal species-delimitation criterion. Its taxonomic implication is procedural: state which inferential alternatives were tested, which remain open, and which evidence classes lie outside the phylogenomic analysis.

The framework reframes rapid-radiation difficulty as an inference-chain problem. Short internodes create conditions in which lineage sorting can separate genealogical and species histories, while finite loci make those genealogies themselves uncertain. The appropriate response is not to select ILS as a default explanation, but to ask whether observed conflict survives estimation-quality and branch-regime checks.

ILS and reticulation can generate overlapping genealogical patterns and require explicit model-aware comparison [6, 18, 20]. This is particularly consequential in insects because the direct examples span different scales: short-branch parasitoid-wasp radiations, bee UCE gene-tree error, butterfly introgression, and deep hemipteran or lepidopteran conflicts. Their heterogeneity argues against universal diagnostic shortcuts. Life stage and sex are rarely analytical dimensions in these phylogenomic studies, so the framework makes no life-stage or sex-specific predictions.

Systematic model or data bias can favor a topology while strong support remains conditional on the analysis [23, 27, 28]. Robustness should consequently mean persistence across relevant challenges, not merely high support under repeated applications of one pipeline. Even then, robustness is comparative rather than proof of historical truth.

The framework has three principal boundaries. First, it is organizational rather than prospectively validated; no sensitivity, specificity, or calibrated error probability is claimed. Second, non-insect theoretical and methodological sources establish general inferential mechanisms but do not estimate their prevalence in insects. Third, an unresolved diagnosis is legitimate. When sampling, gene-tree error, ILS, introgression, and model misspecification cannot be discriminated with available evidence, retaining competing histories is more defensible than forcing a taxonomically convenient bifurcation.

Future validation should consequently be comparative and failure-mode specific. Simulated radiations can test whether the proposed ordering detects known combinations of ILS, gene-tree error, paralogy, introgression, and model bias; empirical clades with independently characterized reticulation or gene-family histories can test whether the same diagnostics remain interpretable under realistic heterogeneity. Such work would evaluate the framework rather than merely illustrate it. Until then, the map should be used to structure questions and sensitivity analyses, not to assign calibrated confidence categories.

Conclusion

Rapid insect radiations expose a recurrent interpretive error because discordant gene trees are often treated as competing species histories before the sources of discordance have been separated. The framework developed here distinguishes molecular evidence, gene-copy histories, estimated gene trees, and species-history hypotheses, then subjects discordance to sequential but reversible challenges involving evidence construction, estimation, biological history, and model adequacy. Its proposed diagnostic map does not identify causes automatically or validate taxonomic decisions. Its purpose is narrower: to make uncertainty explicit, preserve competing explanations, and prevent analytical confidence from being mistaken for direct knowledge of species history.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Young AD, Gillung JP. Phylogenomics—principles, opportunities and pitfalls of big-data phylogenetics. Syst Entomol. 2020;45(2):225-47. doi:10.1111/syen.12406
2. Chesters D. The phylogeny of insects in the data-driven era. Syst Entomol. 2020;45(3):540-51. doi:10.1111/syen.12414
3. Klopfstein S, Langille B, Spasojevic T, Broad GR, Cooper SJB, Austin AD, et al. Hybrid capture data unravel a rapid radiation of pimpliform parasitoid wasps (Hymenoptera: Ichneumonidae: Pimpliformes). Syst Entomol. 2019;44(2):361-83. doi:10.1111/syen.12333
4. Skinner RK, Dietrich CH, Walden KKO, Gordon E, Sweet AD, Podsiadlowski L, et al. Phylogenomics of Auchenorrhyncha (Insecta: Hemiptera) using transcriptomes: examining controversial relationships via degeneracy coding and interrogation of gene conflict. Syst Entomol. 2020;45(1):85-113. doi:10.1111/syen.12381
5. Bossert S, Murray EA, Pauly A, Chernyshov K, Brady SG, Danforth BN. Gene tree estimation error with ultraconserved elements: an empirical study on Pseudapis bees. Syst Biol. 2021;70(4):803-21. doi:10.1093/sysbio/syaa097
6. Morales-Briones DF, Kadereit G, Tefarikis DT, Moore MJ, Smith SA, Brockington SF, et al. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. Syst Biol. 2021;70(2):219-35. doi:10.1093/sysbio/syaa066
7. Shen XX, Steenwyk JL, Rokas A. Dissecting incongruence between concatenation- and quartet-based approaches in phylogenomic data. Syst Biol. 2021;70(5):997-1014. doi:10.1093/sysbio/syab011
8. Roa-Varón A, Dikow RB, Carnevale G, Tornabene L, Baldwin CC, Li C, et al. Confronting sources of systematic error to resolve historically contentious relationships: a case study using Gadiformes (Teleostei, Paracanthopterygii, Gadiformes). Syst Biol. 2021;70(4):739-55. doi:10.1093/sysbio/syaa095
9. Zhang C, Scornavacca C, Molloy EK, Mirarab S. ASTRAL-Pro: quartet-based species-tree inference despite paralogy. Mol Biol Evol. 2020;37(11):3292-307. doi:10.1093/molbev/msaa139
10. Li Q, Scornavacca C, Galtier N, Chan YB. The multilocus multispecies coalescent: a flexible new model of gene family evolution. Syst Biol. 2021;70(4):822-37. doi:10.1093/sysbio/syaa084
11. Allio R, Scornavacca C, Nabholz B, Clamens AL, Sperling FAH, Condamine FL. Whole genome shotgun phylogenomics resolves the pattern and timing of swallowtail butterfly evolution. Syst Biol. 2020;69(1):38-60. doi:10.1093/sysbio/syz030
12. Breinholt JW, Earl C, Lemmon AR, Lemmon EM, Xiao L, Kawahara AY. Resolving relationships among the megadiverse butterflies and moths with a novel pipeline for anchored phylogenomics. Syst Biol. 2018;67(1):78-93. doi:10.1093/sysbio/syx048
13. Mendes FK, Hahn MW. Why concatenation fails near the anomaly zone. Syst Biol. 2018;67(1):158-69. doi:10.1093/sysbio/syx063
14. Roch S, Nute M, Warnow T. Long-branch attraction in species tree estimation: inconsistency of partitioned likelihood and topology-based summary methods. Syst Biol. 2019;68(2):281-97. doi:10.1093/sysbio/syy061
15. Molloy EK, Warnow T. To include or not to include: the impact of gene filtering on species tree estimation methods. Syst Biol. 2018;67(2):285-303. doi:10.1093/sysbio/syx077
16. Sayyari E, Whitfield JB, Mirarab S. Fragmentary gene sequences negatively impact gene tree and species tree reconstruction. Mol Biol Evol. 2017;34(12):3279-91. doi:10.1093/molbev/msx261
17. Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, et al. Genomic architecture and introgression shape a butterfly radiation. Science. 2019;366(6465):594-9. doi:10.1126/science.aaw2090
18. Flouri T, Jiao X, Rannala B, Yang Z. A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. Mol Biol Evol. 2020;37(4):1211-23. doi:10.1093/molbev/msz296

19. Malinsky M, Matschiner M, Svardal H. Dsuite—fast D-statistics and related admixture evidence from VCF files. Mol Ecol Resour. 2021;21(2):584-95. doi:10.1111/1755-0998.13265
20. Degnan JH. Modeling hybridization under the network multispecies coalescent. Syst Biol. 2018;67(5):786-99. doi:10.1093/sysbio/syy040
21. Kawahara AY, Plotkin D, Espeland M, Meusemann K, Toussaint EFA, Donath A, et al. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. Proc Natl Acad Sci U S A. 2019;116(45):22657-63. doi:10.1073/pnas.1907847116
22. Wipfler B, Letsch H, Frandsen PB, Kapli P, Mayer C, Bartel D, et al. Evolutionary history of Polyneoptera and its implications for our understanding of early winged insects. Proc Natl Acad Sci U S A. 2019;116(8):3024-9. doi:10.1073/pnas.1817794116
23. de Moya RS, Yoshizawa K, Walden KKO, Sweet AD, Dietrich CH, Johnson KP. Phylogenomics of parasitic and nonparasitic lice (Insecta: Psocodea): combining sequence data and exploring compositional bias solutions in next generation data sets. Syst Biol. 2021;70(4):719-38. doi:10.1093/sysbio/syaa075
24. Duchen P, Salamin N. A cautionary note on the use of genotype callers in phylogenomics. Syst Biol. 2021;70(4):844-54. doi:10.1093/sysbio/syaa081
25. Richards EJ, Brown JM, Barley AJ, Chong RA, Thomson RC. Variation across mitochondrial gene trees provides evidence for systematic error: how much gene tree variation is biological? Syst Biol. 2018;67(5):847-60. doi:10.1093/sysbio/syy013
26. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-9. doi:10.1038/nmeth.4285
27. Susko E, Roger AJ. Long branch attraction biases in phylogenetics. Syst Biol. 2021;70(4):838-43. doi:10.1093/sysbio/syab001
28. Huang J, Liu Y, Zhu T, Yang Z. The asymptotic behavior of bootstrap support values in molecular phylogenetics. Syst Biol. 2021;70(4):774-85. doi:10.1093/sysbio/syaa100
29. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518-22. doi:10.1093/molbev/msx281
30. Shen XX, Hittinger CT, Rokas A. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat Ecol Evol. 2017;1:0126. doi:10.1038/s41559-017-0126