Eurasia Specialized Veterinary Publication

Entomological Research Letters

ISSN: 3108-639X

2022, Volume 2, Issue 2, Page No: 57-66
Copyright CC BY-NC-SA 4.0
Available online at: www.esvpub.com/

Conflict Among Genomic Markers in Insect Radiations Often Contains Biological Signal Rather Than Merely Analytical Noise

Pieter Botha[1]*, Anel Van Wyk[1], Johan Marais[2]

[1]Department of Genomic Conflict and Biological Signal, Faculty of Biosciences, Stellenbosch University, Stellenbosch, South Africa.
[2]Department of Phylogenomic Discordance and Radiations, Faculty of Life Sciences, University of Pretoria, Pretoria, South Africa.

***E-mail** ✉ pieter.botha@sun.ac.za

ABSTRACT

Genome-scale datasets have increased phylogenetic resolution across insects, but they have also exposed persistent disagreement among loci, gene trees and genomic regions. Such discordance is often treated as an analytical defect to be filtered, down-weighted or overwhelmed by additional data. This perspective argues that this response is incomplete. In rapidly diversifying insect lineages, conflict can instead record incomplete lineage sorting, introgression, gene duplication and loss, or heterogeneous histories within genomic regions, although the same observable patterns may also arise from gene-tree estimation error, model misspecification, alignment problems, locus construction or contamination. We therefore distinguish observed discordance from explanations for discordance and from the stronger inference of species history. Evidence from aphids, bumblebees, drosophilids, butterflies, dragonflies, damselflies and other insect groups shows that genuine biological conflict can persist in large phylogenomic datasets, whereas methodological studies demonstrate that apparently convincing conflict may also be analytical. The central proposal is consequently not to privilege biological explanations, but to treat discordance as an inferential object requiring explicit validation. A staged interpretation of marker conflict can preserve biological information while reducing premature mechanistic attribution, indiscriminate filtering and unsupported taxonomic conclusions.

Keywords: Phylogenomics, Gene-tree discordance, Incomplete lineage sorting, Introgression, Paralogy, Insect radiations

Received: 15 October 2022
Revised: 13 December 2022
Accepted: 15 December 2022

How to Cite This Article: Botha P, Van Wyk A, Marais J. Conflict Among Genomic Markers in Insect Radiations Often Contains Biological Signal Rather Than Merely Analytical Noise. Entomol Res Lett. 2022;2(2):57-66. https://doi.org/10.51847/EaT9yRjUgL

Introduction

Genome-scale phylogenetics has transformed insect systematics by increasing both taxonomic coverage and the number of characters available for reconstructing evolutionary history, but larger datasets simultaneously increase exposure to heterogeneity in locus histories, model fit and data quality [1]. Insect phylogeny is therefore increasingly data-rich without becoming methodologically trivial [2]. Species-level synthesis across tens of thousands of insects demonstrates how extensive molecular resources can now be integrated into large evolutionary hypotheses [3]. Likewise, phylogenomic analyses have clarified major components of lepidopteran history [4] and beetle diversification [5]. These gains establish an important starting point: genomic scale is highly informative, but dataset size alone cannot determine whether disagreement among markers is error or history.

© 2022 Entomological Research Letters

The distinction becomes particularly important in rapid radiations. Protein-encoding ultraconserved elements have provided extensive information about oestroid fly relationships while retaining heterogeneous locus histories at difficult nodes [6]. Transcriptome-scale analysis of Auchenorrhyncha similarly showed that interrogation of gene conflict changes how controversial relationships are evaluated [7]. In Hawaiian *Drosophila* and *Scaptomyza*, substantial phylogenomic discordance persisted despite broad transcriptomic sampling, demonstrating that conflict can remain a defining property of the evolutionary problem rather than disappear with additional loci [8]. The inferential error is therefore to equate conflict either with useless noise or with a particular evolutionary process. Gene-tree discordance is an observation. Incomplete lineage sorting (ILS), introgression, duplication and loss, recombination, selection, estimation error and model inadequacy are competing or sometimes co-occurring explanations. This perspective develops that distinction across insect radiations and proposes that disagreement among genomic markers should first be characterized, then tested for analytical adequacy, and only subsequently interpreted biologically. The aim is not to make discordance inherently meaningful, but to prevent informative evolutionary heterogeneity from being erased before its cause has been examined.

Materials and Methods

This non-empirical perspective treats discordance as a multilevel inferential problem rather than as a single statistic. No new sequence dataset, species tree or statistical analysis is introduced. The literature is organized around four separable questions: what conflict is observed; whether the underlying loci and gene trees are analytically adequate; which biological processes could generate the observed pattern; and what level of species-history inference is justified.

The first distinction concerns measurement. Gene concordance factors and site concordance factors summarize different dimensions of agreement with a focal branch and can reveal conflict obscured by conventional branch support [9]. Low gene concordance, however, is not self-interpreting because it depends on the quality of the estimated gene trees. Empirical analysis of UCEs in *Pseudapis* bees showed that gene-tree estimation error can materially influence downstream species-tree inference, particularly when loci contain limited information [10]. Locus construction can add another layer: genomic characterization of UCEs showed that some nominally separate markers occur within the same gene, and merging such cogenic regions can improve gene-tree support and simulated accuracy [11].

A second distinction concerns evidential weight. Phylogenomic loci may contribute highly unequal amounts of information, with a small number exerting disproportionate influence on competing topologies [12]. Strongly discordant outlier genes also cannot be classified automatically as either erroneous or biologically exceptional because empirical investigation has recovered both analytical and potentially historical explanations [13]. Consequently, gene filtering is not a neutral cleaning step. Simulations show that filtering according to missing data or estimated gene-tree quality can improve, degrade or leave species-tree accuracy largely unchanged depending on the governing conditions [14].

A third distinction concerns model adequacy. Long-branch attraction can make both partitioned likelihood and topology-based summary approaches inconsistent under particular finite-locus conditions, showing that adding loci does not necessarily eliminate systematic bias [15]. Modern likelihood software makes increasingly rich model and sensitivity analyses computationally feasible [16], while automated model selection can reduce avoidable mismatch among candidate substitution models [17]. Yet even improved resampling procedures estimate support conditional on data and modeling assumptions; strong bootstrap support is therefore not equivalent to historical validation [18].

On this basis, discordance is classified here at three levels. Observed conflict describes disagreement among sites, loci or gene trees. Candidate explanation refers to a biological or analytical process capable of producing that disagreement. Historical interpretation concerns what the pattern permits us to infer about lineage relationships. Keeping these levels separate is the methodological premise for the sections that follow.

Results and Discussion

ILS provides the clearest reason why a discordant gene tree need not be a defective gene tree. In rapid radiations, ancestral polymorphisms can persist across short internal branches and sort differently among descendant lineages. Aphididae provide an instructive insect example: deep relationships are characterized by extensive gene-

tree discordance, with evidence relevant to both anomaly-zone conditions and introgression [19]. The significance of this case is precisely that the discordance cannot be reduced safely to one process.

Genus-wide genomic analysis of *Bombus* likewise recovered high gene-tree discordance around parts of the bumblebee radiation and interpreted ILS as a likely contributor [20]. The evidential wording matters. The pattern is consistent with rapid lineage divergence and ancestral polymorphism, but the observation of discordance alone does not uniquely demonstrate ILS.

The multispecies coalescent formalizes why such conflict can be expected under speciation occurring before ancestral alleles have completely coalesced [21]. Its value is therefore not merely to accommodate disagreement, but to supply expectations against which patterns of disagreement can be judged. Those expectations retain marker-specific boundaries. In Lepidoptera, multispecies-coalescent analyses of mitochondrial genomic data were broadly congruent with concatenated and nuclear relationships in the studied setting, but mitochondrial genes remain biologically linked components of one organellar genome and should not be treated automatically as independent nuclear loci [22]. ILS-compatible conflict is consequently meaningful only relative to the data structure and model under which it is evaluated.

Introgression

Introgression provides a second biological route to genomic disagreement because different segments of the genome can trace through different interspecific histories. Across 155 *Drosophila* genome assemblies representing 149 species, complementary analyses identified evidence of introgression across multiple radiations and at both relatively recent and substantially deeper evolutionary times [23]. Here, conflicting genealogies become potential records of reticulation rather than simply failed estimates, but only because the interpretation rests on evidence beyond the existence of alternative trees.

The same principle is visible in *Heliconius*, where whole-genome analysis showed that genomic architecture and introgression contributed to a mosaic radiation history [24]. Such a result makes a single bifurcating topology an incomplete summary of all genomic histories without implying that every conflicting region was introgressed. The theoretical problem is equally important: cross-species gene flow can bias species-tree estimation under particular timing and migration regimes, including conditions resembling rapid radiation, so a method assuming only ILS may return a confidently simplified history [25].

Reticulation also need not be recent. Phylogenomic analyses of dragonflies and damselflies support deep ancestral introgression at several positions in odonate history [26]. Temporal depth alone therefore cannot be used to classify conflict as ILS rather than gene flow. What changes across these examples is not simply whether introgression occurred, but the genomic scale, age and inferential context in which introgression-sensitive conflict is observed.

The key distinctions are easier to recover comparatively than from serial prose, so **Table 1** separates introgression-sensitive evidence patterns from the stronger historical claims that each pattern can legitimately support.

Table 1. Introgression-sensitive genomic conflict and the boundaries of reticulation inference

Marker conflict observed			Reticulation claim	
Reticulation scenario	**Radiation/genome setting**	**Genomic signature**	**Permissible history**	**Competing process and test**
Recurrent radiation-wide discordance	Multiple related species within a radiation	Alternative genealogical relationships recur across genomic data	Historical gene flow becomes a plausible contributor to lineage history. Evidence-supported distinction.	ILS and gene-tree error can generate overlapping disagreement. Require process-specific introgression tests rather than topology counts alone
Genomic mosaicism	Different chromosomal or genomic regions support different histories	Spatially heterogeneous genealogical support across the genome	A single species tree may not summarize all genomic ancestry. Evidence-supported distinction.	Recombination, selection, linkage and estimation error may structure local histories. Test whether discordance is geographically or genomically structured
Gene-flow-induced species-tree displacement	Rapid divergence with cross-lineage exchange	Species-tree inference changes under models	Apparent bifurcating history may be conditional on the assumed process	Model misspecification can mimic or exaggerate displacement. Compare tree and network/coalescent models under explicit sensitivity analysis

		allowing versus excluding gene flow	model. Evidence-supported distinction.	
Deep reticulation	Ancient internal branches of an insect radiation	Asymmetric or network-compatible genealogical signal persists at substantial temporal depth	Introgression cannot be rejected merely because divergence is old. Evidence-supported distinction.	Deep ILS and gene-tree uncertainty remain viable alternatives. Evaluate temporal depth jointly with coalescent expectations and gene-tree quality
Non-identifiable reticulation candidate	Conflict persists but mechanism-specific evidence is insufficient	Discordance survives basic quality checks without uniquely favoring introgression	Reticulation remains one competing explanation. Proposed synthesis.	ILS, GDL, intragenic conflict and analytical error remain unresolved. Retain an explicit unresolved state rather than forcing classification

Gene duplication and loss

Discordance can also arise because genes themselves do not share the same duplication and loss history as the species carrying them. Orthology is therefore an inference about gene-family history, not a property guaranteed by apparent single-copy status. Phylogenetic orthology methods such as OrthoFinder explicitly reconstruct orthogroups and duplication relationships rather than assuming that all superficially comparable sequences are orthologous [27].

The traditional response to paralogy has often been to restrict phylogenomic datasets to single-copy genes, but this can discard informative sequence data. ASTRAL-Pro demonstrates that multicopy gene-family trees can contribute to species-tree inference when paralogy is modeled explicitly [28]. Species-tree-aware reconciliation approaches such as GeneRax similarly make duplication and loss part of the inference problem rather than an unmodeled contaminant [29].

The implication is not that paralogy is harmless. Instead, its effect is conditional. Simulations and empirical analyses indicate that several species-tree methods intended for ILS can remain surprisingly robust in the presence of paralogs under tested regimes, whereas gene-tree estimation error may sometimes be more consequential [30]. Likewise, pseudoorthologs generated by duplication followed by differential loss are not predicted to produce misleading discordant topologies with equal probability across all species-tree shapes and parameter conditions [31].

The relevant distinction is therefore between gene-copy discordance, gene-tree error, and species-history discordance. Removing every multicopy family may suppress genuine gene-family history; retaining all copies without reconciliation may instead transfer duplication history into a species-tree problem. The appropriate evidential question is whether orthology, duplication/loss and gene-tree quality have been evaluated strongly enough to justify treating a particular gene history as evidence about lineage history.

Because these levels are easily collapsed in prose, **Figure 1** separates gene-copy evidence from the interpretation boundary at which marker histories become eligible to inform species history.

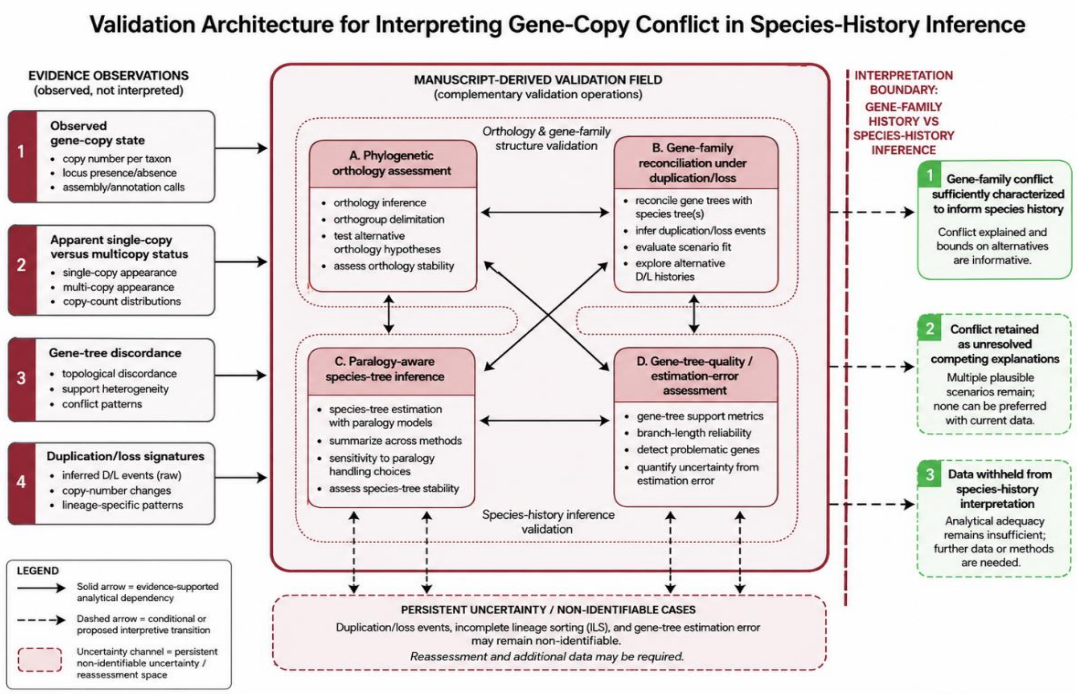


Figure 1. Orthology uncertainty and gene-family history as gates between marker conflict and species-tree inference

Recombination and selection

A further complication is that the gene itself may be an unstable unit of phylogenetic history. Analysis of multiple phylogenomic datasets found that supported intragenic topological conflict ranged from rare to pervasive depending on the dataset; alignment error explained part of this variation, while residual conflict remained compatible with biological or additional analytical causes [32]. A locus should therefore not automatically be assumed to possess one indivisible genealogy.

Recombination is an obvious biological mechanism capable of producing different genealogical histories along a sequence, but detecting a change in supported topology within a locus is not equivalent to demonstrating recombination. Assembly problems, alignment uncertainty, model inadequacy and stochastic gene-tree error can generate similar patterns. Conversely, genuine genomic mosaicism should not be erased simply because it complicates species-tree estimation.

Selection creates an even stronger inferential boundary. The heterogeneous genomic histories documented in *Heliconius* illustrate that different genome regions can participate differently in a reticulate radiation [24], but topology alone does not establish that selection generated, maintained or favored a particular local history. Such an interpretation requires independent evidence about selection, linkage or functional consequence. Thus, recombination and selection are best treated as candidate processes whose genomic predictions must be tested rather than as labels assigned to every within-locus or among-region disagreement.

Table 2 distinguishes the evidential scale at which heterogeneous histories are observed from the additional evidence required before recombination or selection can be invoked.

Table 2. Cross-scale diagnosis of within-locus and regional phylogenetic heterogeneity

Conflict scale	Topology pattern	Process compatibility—not cause	Alignment/recombination/selection challenge
Within-locus topology shift. Different segments of one nominal locus support different histories	Supported change in local topology or genealogical signal within a gene	The single-gene/single-history assumption may be inadequate. Evidence-supported distinction.	Alignment error, assembly error and stochastic GTEE can mimic the pattern. Test alignment stability, segment informativeness and recombination-compatible structure
Alignment-induced intragenic conflict. Ambiguous or erroneous sequence alignment	Conflict weakens or disappears after data-quality correction	At least part of the apparent heterogeneity is analytical.	Residual conflict may remain biological. Reassess homology and alignment before process attribution

		Evidence-supported distinction.	
Genome-region mosaicism. Different genomic regions retain different genealogical relationships	Spatially structured support across chromosomes or linked regions	Genome-wide history may be heterogeneous. Evidence-supported distinction.	ILS, introgression, recombination and linkage can overlap. Compare local histories with explicit process models
Recombination-compatible heterogeneity. Local histories change in a pattern consistent with exchange among ancestral sequence segments	Reproducible internal genealogical discontinuity after quality checks	Recombination becomes a plausible biological explanation. Proposed synthesis.	Topological discontinuity alone is insufficient. Require independent recombination-sensitive analysis
Selection-compatible concentration of local history. Particular genomic regions repeatedly support a distinct history and external evidence suggests non-neutral evolution	Concordant local genealogy plus evidence independent of topology	Selection may be considered as a modifier of genomic-history distribution. Proposed synthesis.	Topology itself does not demonstrate fitness effect or adaptation. Require independent population-genetic, functional or comparative evidence
Unresolved intragenic heterogeneity. Conflict persists after reasonable analytical checks but no process is uniquely supported	Residual local disagreement	Preserve competing explanations. Proposed synthesis.	Premature causal labeling. Report uncertainty explicitly and retain sensitivity analyses

Technical and modeling noise

Biological explanations are credible only if analytical alternatives have been made sufficiently implausible. In Adephaga, alternative modeling and trimming strategies produced materially different phylogenetic outcomes, demonstrating that model misspecification and excessive data reduction can create artefactual relationships rather than merely reduce precision [33]. Persistent conflict is therefore not intrinsically biological, and aggressive preprocessing is not intrinsically conservative.

Locus-level conflict must also be interpreted relative to information content. Likelihood-ratio analyses show that loci differ substantially in how strongly they discriminate among competing topologies [34]. A genome containing many nominally discordant loci may therefore mix decisive conflict, weakly informative disagreement and stochastic instability. Counting conflicting loci without evaluating their information can exaggerate the apparent breadth of an evolutionary signal.

Technical artefact can be still more direct. In Cicadidae, contamination affected phylogenomic inference and could generate misleading relationships despite strong apparent support; removing the affected signal changed interpretation [35]. High support is therefore compatible with analytical failure when erroneous information is reproduced consistently. Credibility instead requires checks of sequence identity, locus construction, alignment, model adequacy and gene-tree informativeness.

Distinguishing biological from analytical discordance

No single signature cleanly separates biological from analytical discordance. Gene and site concordance factors quantify different dimensions of agreement but do not identify the process generating low concordance [9]. Poorly estimated gene trees can themselves increase apparent conflict [10]. Strongly discordant outlier loci can reflect either analytical problems or unusual histories [13]. Contamination can even produce a stable, strongly supported alternative relationship [35]. Classification should therefore rely on convergence among evidence classes rather than one statistic.

Coalescent expectations provide a useful null. Quartet concordance-factor approaches can assess whether observed gene-tree discord is compatible with a simple multispecies coalescent; marked departure can indicate substantial gene-tree estimation error or the need for a more complex historical model [36]. Crucially, departure does not identify which alternative is correct. It marks an inferential boundary beyond which additional biological or analytical testing is required.

This article therefore proposes a conditionality principle: biological interpretation should strengthen only when discordance persists across increasingly independent adequacy checks while matching process-specific expectations. Aphid discordance illustrates that ILS-compatible and introgression-related explanations can coexist

[19]. *Drosophila* demonstrates that process-specific genomic evidence can support introgression beyond topology disagreement alone [23]. Paralogy can remain compatible with accurate species-tree inference under some tested regimes [30], while Adephaga shows that modeling choices can themselves produce conflict [33]. Model-departure diagnostics therefore define a need for further explanation rather than a unique causal label [36]. Between these cases lies a deliberately unresolved state in which several explanations remain viable.

The failure modes and evidential transitions are compressed in **Table 3** because the critical distinction is not a binary label but the conditions under which one interpretation gains or loses support.

Table 3. Conditional evidence states for separating biological history from analytical discordance

Conflict description			Biological permission	
Discordance state	**Signal**	**Allowed interpretation**	**Adequacy boundary**	**Next adequacy test and standing**
Model-consistent discordance	Conflict matches expectations under an adequate null model	A biological process such as ILS remains plausible	Gene-tree error may still contribute	Reassess gene-tree quality and process-specific predictions. Evidence-supported distinction.
Process-specific persistence	Conflict survives quality/model checks and matches an additional biological prediction	A named process becomes more credible	Alternative biological processes may coexist	Seek independent genomic or comparative corroboration. Proposed synthesis.
Information-poor conflict	Discordance concentrates in weakly informative loci	Historical interpretation should remain limited	Weak information can coexist with true history	Improve locus information or down-weight causal claims. Evidence-supported distinction.
Model-sensitive conflict	Topology changes under defensible model or trimming choices	Analytical inadequacy remains plausible	Sensitivity does not prove the alternative topology true	Compare richer models and data treatments. Evidence-supported distinction.
Contamination-compatible signal	Strong conflict maps to suspect sample or sequence identity	Technical artefact can explain apparent support	Genuine discordance may remain after cleaning	Remove or verify affected sequences and reanalyse. Evidence-supported distinction.
Unresolved competing explanations	Adequate analyses still fail to separate causes	Retain multiple explanations explicitly	Resolution cannot be manufactured from support alone	Add independent evidence or preserve uncertainty. Proposed synthesis.

The same logic is visualized in **Figure 2** as a state space rather than a forced decision tree.

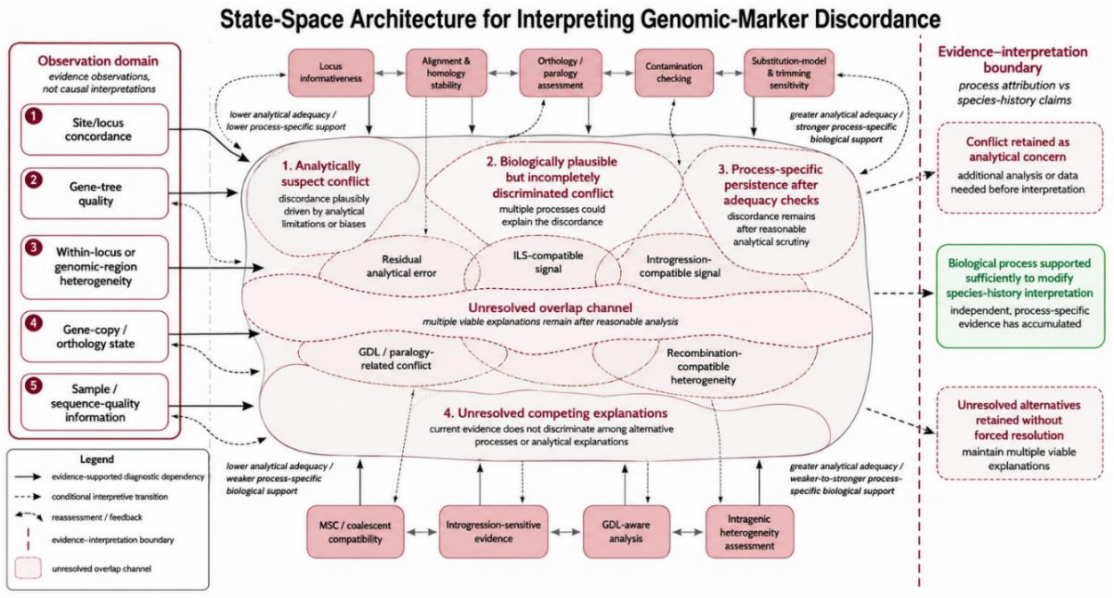


Figure 2. Discordance state space across biological credibility, analytical adequacy and unresolved overlap

Phylogenomic interpretation framework

The proposed framework treats interpretation as a sequence of evidential permissions rather than a decision tree that forces every discordant locus into one category. Concordance metrics first describe conflict without naming its cause [9]. Gene-tree quality [10], locus construction [11], model sensitivity [33] and contamination checks [35] then test whether the observation is analytically adequate. Biological processes are evaluated only after the relevant technical alternatives have been examined sufficiently for the question being asked.

Process inference remains separate from species-history inference. Aphid evidence shows that ILS-related and introgression-related signals can occur in the same radiation [19]. The multispecies coalescent formalizes one expected source of heterogeneous gene histories but has explicit assumptions [21]. Genome-scale *Drosophila* evidence shows that introgression can produce genuine reticulate histories [23], whereas gene-flow simulations demonstrate that such exchange can also mislead species-tree estimators when ignored [25]. Paralog robustness is conditional rather than absolute [30], and within-locus conflict can contain both analytical and biological components [32]. Even a departure from simple coalescent expectations does not itself identify which of these alternatives applies [36].

A final validation boundary precedes taxonomic or evolutionary interpretation. Phylogenetic orthology inference can clarify gene-family relationships [27], paralogy-aware species-tree approaches can retain multicopy information [28], and reconciliation can model duplication and loss explicitly [29], but none alone validates a species history. Likewise, contamination removal can repair a demonstrable technical failure [35] without proving that all remaining discordance is biological. This architecture is therefore proposed as a procedural framework, not a prospectively validated classifier, and it supplies no universal threshold or confidence score.

The central implication is that discordance should neither be removed reflexively nor celebrated reflexively. Genome-scale phylogenetics creates both inferential opportunities and pitfalls [1]. Filtering can help or harm depending on the underlying error regime [14]. Aphid phylogenomics shows that genuine biological processes can coexist in a discordant radiation [19], while Adephaga demonstrates that analytical treatment can also generate artefactual relationships [33].

Taxonomic breadth limits generalization. Gene conflict is informative but method-sensitive in Auchenorrhyncha [7]. Bumblebee genomic discordance has been interpreted as likely ILS [20]. *Drosophila* contains widespread introgression across sampled radiations [23], while deep introgression is supported in Odonata [26]. Cicadidae provide a technical counterexample in which strongly supported signal was misleading because of contamination [35]. These systems differ in marker type, taxonomic depth and available process-specific tests, so they should not be combined into a numerical estimate of how often discordance is biological.

Data exclusion requires similar restraint. Concordance factors expose disagreement that conventional support may obscure [9], whereas bee UCE analyses show why low-information gene trees require caution [10]. Removing loci because they disagree with a preferred topology risks deleting histories that reveal ILS, introgression or gene-family evolution; preserving every discordant sequence can instead propagate error. Sensitivity analysis should ask whether substantive conclusions survive defensible alternative treatments.

The framework also has limits. Gene flow can violate simple species-tree assumptions [25], paralogy has condition-dependent effects [30], and statistical departure from a simple coalescent model can remain causally non-identifiable [36]. Independent morphological, ecological or reproductive evidence may therefore aid downstream systematic interpretation, but genomic discordance alone does not establish a species boundary or justify nomenclatural action.

Future validation should test the architecture across deliberately heterogeneous insect radiations. Such work should combine explicit analytical contrasts, replicated quality audits, process-specific simulations, independent genomic partitions and comparisons among tree and network models. Success should be judged by robustness and explanatory discrimination rather than by whether more nodes become resolved. An unresolved result can be scientifically preferable when available evidence does not separate biological history from analytical failure.

Conclusion

Conflict among genomic markers is neither an embarrassment to be automatically discarded nor a biological signal to be accepted on sight. In insect radiations, discordance can preserve information about lineage sorting, reticulation, gene-family evolution and heterogeneous local histories, while indistinguishable-looking patterns can also arise from analytical inadequacy. The critical task is to preserve the boundary between observing conflict,

explaining its source and using it to infer species history. The proposed interpretation architecture makes that boundary explicit and retains unresolved alternatives when evidence is insufficient. Used in this way, discordance becomes a testable property of phylogenomic evidence rather than a residual category defined only by disagreement with a preferred tree.

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. Chesters D. Construction of a species-level tree of life for the insects and utility in taxonomic profiling. Syst Biol. 2017;66(3):426-39. doi:10.1093/sysbio/syw099
4. 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
5. McKenna DD, Shin S, Ahrens D, Balke M, Beza-Beza C, Clarke DJ, et al. The evolution and genomic basis of beetle diversity. Proc Natl Acad Sci U S A. 2019;116(49):24729-37. doi:10.1073/pnas.1909655116
6. Buenaventura E, Lloyd MW, Perilla López JM, González VL, Thomas-Cabianca A, Dikow T. Protein-encoding ultraconserved elements provide a new phylogenomic perspective of Oestroidea flies (Diptera: Calyptratae). Syst Entomol. 2021;46(1):5-27. doi:10.1111/syen.12443
7. 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
8. Church SH, Extavour CG. Phylotranscriptomics reveals discordance in the phylogeny of Hawaiian Drosophila and Scaptomyza (Diptera: Drosophilidae). Mol Biol Evol. 2022;39(3):msac012. doi:10.1093/molbev/msac012
9. Minh BQ, Hahn MW, Lanfear R. New methods to calculate concordance factors for phylogenomic datasets. Mol Biol Evol. 2020;37(9):2727-33. doi:10.1093/molbev/msaa106
10. 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
11. Van Dam MH, Henderson JB, Esposito L, Trautwein M. Genomic characterization and curation of UCEs improves species tree reconstruction. Syst Biol. 2021;70(2):307-21. doi:10.1093/sysbio/syaa063
12. Brown JM, Thomson RC. Bayes factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses. Syst Biol. 2017;66(4):517-30. doi:10.1093/sysbio/syw101
13. Walker JF, Brown JW, Smith SA. Analyzing contentious relationships and outlier genes in phylogenomics. Syst Biol. 2018;67(5):916-24. doi:10.1093/sysbio/syy043
14. 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
15. 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

16. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530-4. doi:10.1093/molbev/msaa015
17. 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
18. 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
19. Owen CL, Miller GL. Phylogenomics of the Aphididae: Deep relationships between subfamilies clouded by gene tree discordance, introgression and the gene tree anomaly zone. Syst Entomol. 2022;47(3):470-86. doi:10.1111/syen.12542
20. Sun C, Huang J, Wang Y, Zhao X, Su L, Thomas GWC, et al. Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits. Mol Biol Evol. 2021;38(2):486-501. doi:10.1093/molbev/msaa240
21. 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
22. Kim MJ, Kim I, Cameron SL. How well do multispecies coalescent methods perform with mitochondrial genomic data? A case study of butterflies and moths (Insecta: Lepidoptera). Syst Entomol. 2020;45(4):857-73. doi:10.1111/syen.12431
23. Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, et al. Widespread introgression across a phylogeny of 155 Drosophila genomes. Curr Biol. 2022;32(1):111-23.e5. doi:10.1016/j.cub.2021.10.052
24. 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
25. Jiao X, Flouri T, Rannala B, Yang Z. The impact of cross-species gene flow on species tree estimation. Syst Biol. 2020;69(5):830-47. doi:10.1093/sysbio/syaa001
26. Suvorov A, Scornavacca C, Fujimoto MS, Bodily P, Clement M, Crandall KA, et al. Deep ancestral introgression shapes evolutionary history of dragonflies and damselflies. Syst Biol. 2022;71(3):526-46. doi:10.1093/sysbio/syab063
27. Emms DM, Kelly S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238. doi:10.1186/s13059-019-1832-y
28. 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
29. Morel B, Kozlov AM, Stamatakis A, Szöllősi GJ. GeneRax: A tool for species-tree-aware maximum likelihood-based gene family tree inference under gene duplication, transfer, and loss. Mol Biol Evol. 2020;37(9):2763-74. doi:10.1093/molbev/msaa141
30. Yan Z, Smith ML, Du P, Hahn MW, Nakhleh L. Species tree inference methods intended to deal with incomplete lineage sorting are robust to the presence of paralogs. Syst Biol. 2022;71(2):367-81. doi:10.1093/sysbio/syab056
31. Smith ML, Hahn MW. The frequency and topology of pseudoorthologs. Syst Biol. 2022;71(3):649-59. doi:10.1093/sysbio/syab097
32. Smith SA, Walker-Hale N, Walker JF. Intragenic conflict in phylogenomic data sets. Mol Biol Evol. 2020;37(11):3380-8. doi:10.1093/molbev/msaa170
33. Vasilikopoulos A, Balke M, Kukowka S, Pflug JM, Martin S, Meusemann K, et al. Phylogenomic analyses clarify the pattern of evolution of Adephaga (Coleoptera) and highlight phylogenetic artefacts due to model misspecification and excessive data trimming. Syst Entomol. 2021;46(4):991-1018. doi:10.1111/syen.12508
34. Mount GG, Brown JM. Comparing likelihood ratios to understand genome-wide variation in phylogenetic support. Syst Biol. 2022;71(4):973-85. doi:10.1093/sysbio/syac014
35. Owen CL, Marshall DC, Wade EJ, Meister R, Goemans G, Kunte K, et al. Detecting and removing sample contamination in phylogenomic data: An example and its implications for Cicadidae phylogeny (Insecta: Hemiptera). Syst Biol. 2022;71(6):1504-23. doi:10.1093/sysbio/syac043
36. Allman ES, Mitchell JD, Rhodes JA. Gene tree discord, simplex plots, and statistical tests under the coalescent. Syst Biol. 2022;71(4):929-42. doi:10.1093/sysbio/syab008