Ecology & Evolution | Jan 13, 2026 | 9 min read | By Bryan White

What is a Species, Really? How Genomics is Solving Biology's Oldest Debate

Biology * Evolution * DNA

Introduction: The Epistemological Crisis of the Species Rank

The species is the fundamental currency of biology: the primary unit of conservation, the terminal node of phylogenetic analysis, and the core subject of evolutionary theory. Yet from Linnaean morphology to the reproductive isolation of Ernst Mayr's Biological Species Concept, taxonomists have lacked a universal yardstick for biodiversity. High-throughput sequencing has transformed species delimitation into a quantitative, process-oriented science.

A landmark synthesis in the Annual Review of Ecology, Evolution, and Systematics by Sonal Singhal, Adam D. Leach , Matthew K. Fujita, Carlos Daniel Cadena, and Felipe Zapata ("A Genomic Perspective on Species Delimitation") captures this conceptual shift.[1] Rather than treating speciation as an instantaneous bifurcation, modern systematics models an extended continuum where diverging populations gradually accumulate genetic, ecological, and reproductive barriers.[2] Navigating the resulting "grey zone" of speciation requires probabilistic models capable of resolving lineages ranging from desert reptiles and pelagic hydrozoans to the recombining microbes that sustain the biosphere.

Theoretical Foundations: The Multispecies Coalescent

Whereas single-locus phylogenetics -- frequently reliant on mitochondrial DNA -- assumed individual gene trees mirrored species histories, whole-genome datasets demonstrate that every genome is a mosaic of thousands of genealogies. Gene-tree discordance arises primarily from Incomplete Lineage Sorting (ILS), or deep coalescence, wherein ancestral polymorphisms segregate stochastically so that a gene copy in Species A coalesces more recently with Species C than with its sister Species B.[3]

The Multispecies Coalescent (MSC) -- the "censored coalescent" -- models how gene lineages coalesce backward in time within species-tree boundaries.[4] Parameterized by theta (effective population size scaled by the mutation rate) and tau (divergence time scaled by mutation rate), the MSC separates the stochastic noise of population genetics from lineage splitting.[5] Extensions including the Multispecies Coalescent with Introgression (MSci) and Multispecies Network Coalescent (MSNC) incorporate cross-species gene flow and reticulate hybridization directly into inference.[4]

Methodological Workflow: Discovery, Validation, and Over-Splitting

Operational genomic delimitation follows a two-step workflow: exploratory Discovery and model-based Validation.[1] Discovery algorithms such as Structure and Admixture assign single nucleotide polymorphisms (SNPs) to K clusters by minimizing deviations from the Hardy-Weinberg equilibrium, while PCA and phylogenetic networks visualize

genetic discontinuities.[1]

Validation tools then evaluate competing coalescent models. Bayesian Phylogenetics and Phylogeography (BPP) uses reversible-jump MCMC (rjMCMC) to collapse or split nodes across gene-tree uncertainty,[6][7] while SNAPP and Bayes Factor Delimitation (BFD*) compute exact species-tree likelihoods directly from biallelic SNPs.[7][8] However, as Sukumaran and Knowles (2017) demonstrated, the standard MSC detects any cessation of gene flow, often mistaking intraspecific population structure for species boundaries.[9][10] To prevent taxonomic inflation ("over-splitting"), systematists pair MSC probabilities with the continuous Genealogical Divergence Index (gdi), which scales divergence time by population size from 0 to 1 (where gdi > 0.7 indicates strong lineage independence and gdi < 0.2 indicates conspecific structure).[11]

| Method | | Algorithm / Basis | | Input Data | | Strengths | | Limitations |

| BPP | | Bayesian MSC (rjMCMC) | | Multilocus / SNPs | | Jointly estimates theta, tau, and delimitation. | | Computationally intensive; can over-split population structure. |

| SNAPP + BFD* | | Exact MSC Likelihood | | Biallelic SNPs | | Bypasses gene-tree estimation; ranks models via Bayes Factors. | | High compute cost; requires pre-defined candidate models. |

| gdi | | Divergence Index | | MSC Parameters (tau, theta) | | Continuous speciation metric; reduces taxonomic inflation. | | Heuristic threshold rather than formal likelihood test. |

| SODA | | Quartet Frequencies | | Gene Trees | | Rapid execution on large phylogenomic matrices. | | Sensitive to gene-tree error and introgression. |

| Structure / PCA | | Genotype Clustering | | Unlinked SNPs | | Visualizes admixture and hybrid zones de novo. | | Descriptive discovery tools; do not model coalescence. |

Empirical Case Studies Across the Tree of Life

Applying coalescent and genomics frameworks across disparate phyla reveals that morphological evolution and genomic divergence are frequently decoupled.[1]

* Cryptic Stasis in Xantusia Night Lizards: Inhabiting rock crevices across North American deserts, Xantusia populations appear morphologically uniform. Yet phylogenetic and coalescent analyses by Sinclair, Singhal, and colleagues uncovered ancient lineages separated by deep sequence divergence.[12][13] Low dispersal drove predictable allopatric speciation while stabilizing selection within rock crevices maintained phenotypic stasis -- an underestimation of biodiversity resolved via museum-vouchered phylogeography.[13]

* Permeable Avian Boundaries: In avian radiations such as Setophaga warblers and Scytalopus tapaculos, rapid sexual selection on plumage and song establishes reproductive barriers before genome-wide differentiation accumulates.[14][15] Recognized bird species frequently exchange alleles across porous boundaries, confining divergence to narrow genomic "islands" housing barrier loci while neutral regions remain homogenized -- requiring phenotypic integration alongside genomic

metrics.[14][16]

* Colonial Marine Hydrozoans (Siphonophora): Fragile pelagic siphonophores (Physalia, Nanomia) disintegrate in sampling nets. Using transcriptomics and chromosome-scale assemblies, researchers resolved deep Codonophora nodes, uncovering two-fold genome-size shifts within Nanomia (1.5-1.7 Gb in N. septata vs. ~0.7 Gb in N. bijuga) and partitioning the Portuguese man o' war (Physalia) into at least four distinct species.[17][18][19]

* Prokaryotic Recombination and Bacterial Ecotypes: Bacteria and other prokaryotes violate bifurcating tree models because Horizontal Gene Transfer (HGT) is decoupled from meiosis.[20][21] Under the Stable Ecotype Model, distinct thermal ecotypes of Yellowstone Synechococcus persist within a single <0.5% divergence cluster (well inside the standard 95% ANI threshold) via locus-specific selective sweeps, whereas Ensifer clades show tight genomic-ecological concordance.[22][23][24]

| Taxon Group | | Focal Genera | | Evolutionary Regime | | Genomic Delimitation Insight |

| Reptiles | | Xantusia | | Allopatry with morphological stasis | | Validates deep cryptic species masked by niche conservatism. |

| Birds | | Setophaga, Scytalopus | | Rapid sexual selection with introgression | | Divergence concentrates in barrier islands; requires phenotypic integration. |

| Siphonophores | | Physalia, Nanomia | | Coloniality and genome-size plasticity | | Splits Physalia into four species; tracks chromosomal rearrangements in Nanomia. |

| Bacteria | | Synechococcus, Ensifer | | Asexual reproduction with pervasive HGT | | 95% ANI clusters harbor multiple ecologically distinct ecotypes under selection. |

Machine Learning and Integrative Taxonomy 2.0

To scale beyond the computational limits of MCMC samplers, supervised Convolutional Neural Networks (CNNs) trained on simulated alignment tensors classify demographic splits with gene flow more accurately than Approximate Bayesian Computation.[25] Simultaneously, unsupervised Variational Autoencoders (VAEs) and Self-Organizing Maps (SuperSOMs) integrate SNP matrices with spatial, climatic, and morphometric layers without pre-labeled taxonomic priors.[26] Together, these tools address genome heterogeneity (conflicting local genealogies driven by introgression and linked selection) and divergence mismatch across rapid adaptive radiations.[1][27]

Conclusion: Toward a Process-Based Taxonomy

The genomic perspective synthesized by Singhal et al. (2025) advances an "Integrative Taxonomy 2.0" grounded in evolutionary independence rather than static typology.[1] By uniting coalescent models, gdi thresholds, and machine learning with phenotypic, ecological, and natural-history evidence, modern biology treats the species boundary as a dynamic, semi-permeable membrane shaped by the ongoing mechanics of evolution.

Updates... (Nov 20, 2026): Corrected Xantusia night-lizard phylogeography attribution and references (Sinclair et al. 2004; Singhal et al. 2024), unnumbered section headings, and converted publisher-restricted links to canonical DOIs with

verified publication dates.

Works Cited

* Singhal, S., Leach , A. D., Fujita, M. K., Cadena, C. D., & Zapata, F. (Nov 5, 2025). A Genomic Perspective on Species Delimitation. Annual Review of Ecology, Evolution, and Systematics. -- https://doi.org/10.1146/annurev-ecolsys-102723-055311

* Burriel-Carranza, B., Estarellas, M., Mochales-Ria o, G., et al. (May 30, 2023). Species boundaries to the limit: Integrating species delimitation methods is critical to avoid taxonomic inflation in the genomic era. bioRxiv. -- https://doi.org/10.1101/2022.12.28.522089

* Rannala, B., Edwards, S. V., Leach , A., & Yang, Z. (Mar 2020). The multi-species coalescent model and species tree inference. University of Washington. -- https://faculty.washington.edu/leache/wordpress/wp-content/uploads/2020/03/2020RannalaSpeciestree.pdf

* Wikipedia (2026). Multispecies coalescent process. -- https://en.wikipedia.org/wiki/Multispecies_coalescent_process

* Jiao, X., Flouri, T., & Yang, Z. (Dec 2021). Multispecies coalescent and its applications to infer species phylogenies and cross-species gene flow. National Science Review, 8(12), nwab127. -- https://doi.org/10.1093/nsr/nwab127

* bio.tools (2024). BPP: Bayesian Phylogenetics and Phylogeography. -- https://bio.tools/bpp

* Leach , A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (Mar 12, 2014). Species Delimitation using Genome-Wide SNP Data. Systematic Biology, 63(4), 534-542. -- https://doi.org/10.1093/sysbio/syu018

* Leach , A. D., & Bouckaert, R. R. (Jan 2018). Species Trees and Species Delimitation with SNAPP: A Tutorial and Worked Example. Evolution and Genomics. -- http://evomics.org/wp-content/uploads/2018/01/BFD-tutorial.pdf

* Sukumaran, J., & Knowles, L. L. (Jan 30, 2017). Multispecies coalescent delimits structure, not species. Proceedings of the National Academy of Sciences, 114(7), 1607-1612. -- https://doi.org/10.1073/pnas.1607921114

* Hausdorf, B. (Apr 2025). Species Delimitation Using Genomic Data: Options and Limitations. Molecular Ecology, e17717. -- https://doi.org/10.1111/mec.17717

* Jackson, N. D., Carstens, B. C., Morales, A. E., & O'Meara, B. C. (Dec 20, 2016). Species Delimitation with Gene Flow. Systematic Biology, 66(5), 799-812. -- https://doi.org/10.1093/sysbio/syw117

* Sinclair, E. A., Bezy, R. L., Bolles, K., et al. (Jan 1, 2004). Testing Species Boundaries in an Ancient Species Complex with Deep Phylogeographic History: Genus Xantusia (Squamata: Xantusiidae). The American Naturalist, 164(3). -- https://doi.org/10.2307/3473124

* Singhal, S., Davis, H. R., et al. (2024). Comparative approaches to understanding speciation: a case study in Xantusia lizards. Evolutionary Journal of the Linnean Society, 3(1), kzae024. -- https://doi.org/10.1093/evolinnean/kzae024

* Cadena, C. D., & Zapata, F. (Mar 1, 2021). The genomic revolution and species

delimitation in birds (and other organisms): Why phenotypes should not be overlooked. Ornithology, 138(2), ukaa069. -- https://doi.org/10.1093/ornithology/ukaa069

* Cadena, C. D., Cuervo, A. M., C spedes, L. N., et al. (2020). Systematics, biogeography, and diversification of Scytalopus tapaculos (Rhinocryptidae), an enigmatic radiation of Neotropical montane birds. The Auk / LSU Scholarly Repository. -- https://repository.lsu.edu/biosci_pubs/494/

* Deane-Coe, P. E., et al. (Dec 2019). Divergent selection and drift shape the genomes of two avian sister species spanning a saline-freshwater ecotone. Ecology and Evolution. -- https://doi.org/10.1002/ece3.5804

* Munro, C., Siebert, S., Zapata, F., et al. (Jun 22, 2018). Improved phylogenetic resolution within Siphonophora (Cnidaria) with implications for trait evolution. Molecular Phylogenetics and Evolution, 127, 823-833. -- https://doi.org/10.1016/j.ympev.2018.06.030

* Ahuja, N., Schultz, D. T., Destanovi , D., et al. (Jul 2, 2026). Siphonophore genome structure and the evolution of functional specialization. PLOS ONE. -- https://doi.org/10.1371/journal.pone.0351247

* Church, S. H., Dunn, C. W., et al. (Aug 2025). Population genomics of a sailing siphonophore reveals genetic structure in the open ocean. Current Biology, 35(15), 3556-3569.e6. -- https://doi.org/10.1016/j.cub.2025.05.066

* Doolittle, W. F., & Papke, R. T. (Nov 29, 2006). Genomics and the bacterial species problem / The bacterial species dilemma and the genomic-phylogenetic species concept. Philosophical Transactions of the Royal Society B. -- https://doi.org/10.1098/rstb.2006.1914

* Wright, E. S., & Baum, D. A. (Oct 3, 2018). Exclusivity offers a sound yet practical species criterion for bacteria despite abundant gene flow. BMC Genomics, 19, 724. -- https://doi.org/10.1186/s12864-018-5099-6

* Cohan, F. M. (Oct 2017). Transmission in the Origins of Bacterial Diversity, From Ecotypes to Phyla. Microbiology Spectrum. -- https://doi.org/10.1128/microbiolspec.MTBP-0014-2016

* Kuzmanovi , N., Fagorzi, C., Mengoni, A., et al. (Mar 2022). Taxonomy of Rhizobiaceae revisited: proposal of a new framework for genus delimitation. International Journal of Systematic and Evolutionary Microbiology. -- https://doi.org/10.1099/ijsem.0.005243

* Melendrez, M. C., Becraft, E. D., Wood, J. M., et al. (Sep 2017). Recombination-Driven Genome Evolution and Stability of Bacterial Species. Genetics. -- https://doi.org/10.1534/genetics.117.300061

* Perez, M. F., Bonatelli, I. A. S., Romeiro-Brito, M., et al. (Dec 24, 2020). Coalescent-based species delimitation meets deep learning: Insights from a highly fragmented cactus system. bioRxiv. -- https://doi.org/10.1101/2020.12.23.424219

* Pyron, R. A. (Oct 5, 2023). Unsupervised machine learning for species delimitation, integrative taxonomy, and biodiversity conservation. Molecular Phylogenetics and Evolution, 189, 107939. --

https://doi.org/10.1016/j.ympev.2023.107939

* Nosil, P., & Feder, J. L. (Feb 5, 2012). Genomic divergence during speciation: causes and consequences. Philosophical Transactions of the Royal Society B, 367(1587), 332-342. -- https://doi.org/10.1098/rstb.2011.0263