The Clarion Volume 8 Number 2 (2019) PP 1-9
The Clarion
International Multidisciplinary Journal
ISSN : 2277-1697

# Unraveling the tree of life: A grand challenge for Biology.

**Scott V. Edwards**
Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138 USA

**Abstract**

Building the Tree of Life is an ongoing activity of scientists around the world, one that combines information from both the genotype and phenotype of organisms. I review recent trends in this effort and describe a number of models, including the multispecies coalescent model, as means to achieve this end.

**Keywords**: Tree of Life, Genomes, Phenomes, phylogenetic, morphological

## 1. Introduction

Building and studying the Tree of Life (ToL) is one of the grand challenges of the biological sciences (ToL: Wolf et al. 2002, Cracraft et al. 2004, Delsuc et al. 2005, Pace 2009, Forterre 2015, Soltis *et al.* 2018). The ToL is a roadmap to understanding biological diversity. An understanding of the details of the branches of the tree of life allows researchers to understand macroevolutionary trends in diverse traits, from ecology (Webb et al. 2002, Geeta et al. 2014, Li 2016), behavior (Eisenstein et al. 2016), geography (Jetz et al. 2011), and adaptation to cancer (Aktipis et al. 2015), human health (Wolf et al. 2002) and many other issues of high societal relevance (McTavish et al. 2017). Some researchers have questioned the reality of the ToL, preferring to call it a 'net of life' or some other term that conveys the large amounts of horizontal gene transfer, hybridization and other reticulate phenomena that are now known to occur across diverse taxa (Doolittle 2003, Booth et al. 2016, Doolittle et al. 2016). Indeed, reticulation events in the form of hybridization, introgression or horizontal gene transfer are common across diverse clades, including eukaryotes and vertebrates and ultimately humans However, I and many others believe that there is enough consistency of phyologenetic signal across diverse data sets to suggest that there is a dominant or modal ToL, despite rampant discordance due to reticulations (Daubin *et al.,* 2003).

## 2. Genomes, Phenomes and the Tree of Life

The genomes of living and extinct organisms possess an abundance of data that can be used to build and unravel the Tree of Life. There is some disagreement among systematists with about the extent to which molecular data should dominate the search for the ToL, an argument that goes back several decades (Patterson et al. 1987). Suggestions have been made that molecular characters are the best choice for building the ToL, outperforming morphological traits in the number of characters, ease of delimitation of characters, consistency of signal or providing less biased sampling of characters (Scotland et al. 2003). However, some systematists, particularly those with backgrounds in museum science that traditionally have employed morphological characters in their classifications, favor continued use of morphology in building the ToL, especially with regard to inclusion of fossil taxa (Wiens 2004, O'Leary et al. 2013). Arguments for the use of morphology in building the ToL also derive from the philosophical mandate promoted by Arnold Kluge for 'total evidence' – using all available data to build phylogenetic hypotheses (Kluge 1989, Eernisse et al.

Corresponding author : sedwards@fas.harvard.edu
**DOI number:** 10.5958/2277-937X.2019.00011.X

1993). Indeed, there are many interesting and novel approaches to the analysis of morphological data, including analysis of continuous traits rather than the typical binary traits, some of which likely increase the amount of signal available for building the ToL (Wiens 2001, Parins-Fukuchi 2018). The challenges of morphological versus molecular data in building the ToL extend to microbial taxa (Harper *et al.* 2009).

My personal opinion about the molecules versus morphology debate in systematics is that the biggest challenge to morphological traits is not their numbers or their delimitation but their non-random choice: the act of consciously choosing specific morphological characters is often non-random, and can be biased for or against a given phylogenetic hypothesis in ways that make it difficult to study the distribution and strength of phylogenetic signal in a meaningful way. Some have argued that morphological characters show as much (or as little) homoplasy (conflicting phylogenetic signal) as do molecular characters, thereby suggesting that any biases in their choice may not compromise the purity of their signal (Sanderson *et al.* 1989). However, the fact that morphological characters are chosen consciously suggests that, unknowingly, their average signal could be inflated and their noise reduced compared to randomly chosen characters, a topic that deserves more study. As the scoring of morphological and other phenotypic traits becomes more automated, and the ontological relationships among traits becomes better defined, the possibility of scoring 'unbiased' morphological traits for phylogenetic ananlysis becomes more likely (Deans *et al.* 2015, Dececchi *et al.* 2015, Thessen *et al.* 2015, Edmunds et al. 2016, Wirkner *et al.* 2017). The era of "phenomics" has only just begun for animal diversity (O'Leary et al. 2013, Maddison 2016), is more mature for plant diversity (Tardieu *et al.* 2017, Ninomiya *et al.* 2019) and could result in a renaissance of the use of morphology in systematics. Some approaches to the unbiased selection of morphological characters, such as those envisioned in analyses of the fictious but scoreable 'Canimalcules' of James Rohlf (Rohlf *et al.* 1967), should be revisited with modern methods of phenomics.

For now, genomic characters – with significant help from fossil taxa and guidance (if not explicit characters) from morphology - seem to be moving towards a realization of the ToL at a rapid pace. For decades, DNA sequence data has been the workhorse of molecular systematics, a paradigm that was significantly accelerated in the 1980s and 1990s by the advent of the polymerase chain reaction (PCR: Kocher et al. 1989, Paabo et al. 1989). Today, whole genomes are the norm, dramatically advancing the field of comparative genomics. Even as whole genome data becomes commonplace, these genomes continue to be analyzed with reference primarily to the DNA sequence of the interrogated genomes.

## 3. Models for phylogenetic analysis of whole genome data

Phylogenetic analyses of genome-wide DNA sequences come in two principle flavors: either the various genes or other phylogenomic markers are concatenated into single "supergenes" for analysis (the supermatrix or concatenation method: de Queiroz *et al.* 2007) or, more recently, trees of each gene are constructed and subsequently combined according to biological models such as the multispecies coalescent model to achieve an overarching 'species tree' (Edwards 2009, Rannala *et al.* 2020). The term species tree is meant to refer to any demographic history that consists solely of divergence events, with little or no subsequent hybridization or other reticulation between lineages. The entities analyzed need not be species in the classical sense, but could instead simply be populations or representatives of higher taxa. The term 'species tree' was popularized by John Avise to distinguish the phylogenetic relationships of species from the phylogenetic patterns found in the underlying genes, which, early on, were found frequently to vary from gene to gene across the genome (Neigel *et al.* 1986, Avise 2000). The heterogeneity in gene trees sampled from a genome need not arise out of complex histories of hybridization or gene flow, but can instead be a simple consequence of rapid speciation accompanied by large population sizes, which permit only slow shifts in allele frequencies due to drift, with the result that the phylogeny of alleles sampled from a tree with species A, B and C may not reflect the branching sequence of populations A, B and C. Such heterogeneity is called incomplete lineage sorting (ILS) when thinking forwards in time, from past to present, or deep coalescence (DC) when thinking backwards in time, from present to past (Maddison 1997, Maddison *et al.* 2006). ILS is now known to be a ubiquitous aspect of genomic history and is likely the most common source of mismatches between gene trees and species trees (Pollard *et al.* 2006, Edwards 2009, Pease *et al.* 2013).

Phylogenetic methods employing the multispecies coalescent model (MSC: Degnan *et al.* 2009) arose out of early observations of abundant gene tree heterogeneity. The MSC is a mathematical framework describing the distribution of gene trees expected given a species tree with a set of phylogenetic relationships and branch lengths. The likelihood of a gene tree given a species tree, first written down by Rannala and Yang (2003) forms the cornerstone of most phylogenetic methods employing the MSC. The MSC allows researchers to take a set of rooted or unrooted gene trees and estimate an overarching species tree. Such methods provide a biological model of exceptional breadth, utility and elegance to combine data from different genes together to estimate phylogenetic tree, even when those gene trees differ from one another. MSC methods of tree inference are, to my mind, more satisfying than earlier 'supertree' methods of combining data from different genes or character sets, primarily because MSC models are informed by biological principles and population genetics, as opposed to consensus or supertree methods, which often combine trees due to arbitrary models (Steel *et al.* 2000, Bininda-Emonds 2005). As of this writing, there are numerous MSC methods for building phylogenetic trees, ranging from fast, clustering methods to likelihood and slower but more accurate Bayesian methods (Liu *et al.* 2009, Liu *et al.* 2010, Chifman *et al.* 2014, Rannala *et al.* 2017, Zhang *et al.* 2018, Liu *et al.* 2019). Although supermatrix methods are still used, primarily as a benchmark to acknowledge the traditional approach to molecular systematics, MSC methods are growing rapidly and generally considered superior and more appropriate for genome-scale data. In fact, theory suggests that supermatrix methods are actually a subset of MSC methods, and suffer from the same types of model violations that have been identified for MSC methods (Liu *et al.* 2015a, Liu *et al.* 2015b, Edwards 2016, Edwards *et al.* 2016). A recent study showed that, across 47 phylogenomic data sets of widely varying taxa, size and complexity, MSC methods did a better job at explaining characteristics of the underlying sequence data than did supermatrix methods, suggesting that the MSC model is more appropriate for building the ToL (Jiang *et al.* 2019). Although MSC methods were inspired by observations of gene tree heterogeneity, they do not require such heterogeneity to be applied appropriately. The key assumption is that recombination between loci is free, allowing gene trees with varying topologies or branch lengths to be observed at different loci.

Despite the wide applicability and versatility of MSC models, clearly more work is needed to develop models that capture even more subtle aspects of biological realities underlying the ToL (Bravo *et al.* 2019). The strict MSC model has a number of shortcomings – most notably, the inability to accommodate gene flow between species, an evolutionary force whose prominence has only increased in recent years (Mallet *et al.* 2015). New MSC models that can accommodate gene flow between lineages have been emerging and promise increased flexibility to account for reticulations in the ToL (Stenz *et al.* 2015, Yu *et al.* 2015, Solís-Lemus *et al.* 2016, Solis-Lemus *et al.* 2017). Another shortcoming is the assumption of no recombination within loci – a challenge for data sets, such as transcriptome data sets, whose loci sometimes span several tens of kilobases, if not greater genomic lengths (Gatesy *et al.* 2013). Marker choice will influence how strongly this assumption is violated (Costa *et al.* 2016, Jennings 2017). For example, the use of single nucleotide polymorphisms (SNPs) in phylogenetic analysis can circumvent this particular recombination assumption, because recombination cannot take place within a single site in the genome. However, the options available for analyzing SNPs data in a phylogenetic context are still limited, although some authors suggest success at analyzing such data across diverse taxonomic scales (Eaton *et al.* 2017, Leaché *et al.* 2017, Stange *et al.* 2018, Spriggs *et al.* 2019). Another challenge is the difficulty of combining molecular and morphological data in the framework provided by the MSC. It is not yet entirely clear how to model morphological characters using the MSC approach, although important inroads to this question are emerging (Mendes *et al.* 2018). Although the MSC has clearly brought the field of phylogenetics to a new 'local optimum' (Edwards 2009), we are still far from achieving a global optimum that will accommodate the many complexities inherent in the ToL (Bravo *et al.* 2019).

Another major area in need of further progress is the use of rare genomic changes in building the ToL (Rokas *et al.* 2000, Boore 2006, Rokas 2006, Boore et al. 2008, Rogozin *et al.* 2009). Rare genomic changes are structural or non-sequence-based molecular characters that usually have low homoplasy and high phylogenetic information content. Molecular characters like gene order, synteny, transposable

element insertions, insertions and deletions in protein-coding genes and chromosomal rearrangements fall into this category. The most common type of rare-genomic change used in phylogenetics are transposable element insertions (Shedlock *et al.* 2000), which have been used to great success in several clades across the ToL (Grover *et al.* 2008, Rogozin *et al.* 2009, Churakov *et al.* 2010, Wu *et al.* 2019). The emphasis by the community on DNA sequence data in phylogenetics – something I often refer to as the 'dance of nucleotides' – is limiting, because rare genomic changes often offer much more compelling evidence for the existence of a given clade. Rare genomic changes are not necessarily immune to the challenges posed by ILS, introgression or other reticulations (Hillis 1999, Suh *et al.* 2015), but they often have higher evidentiary power than do the accumulated signal of nucleotide variations in DNA sequences. Without reviewing the state of the field regarding rare genomic changes, I will simply note that in the past, these types of molecular characters were very hard won and therefore very rarely employed because of the large amount of benchwork required to discover and validate them (Ellegren 2007, Kaiser *et al.* 2007). Now that genome sequencing has become more routine, rare-genomic changes are much easier to discover and analyze, and are proving their worth in diverse phylogenetic contexts (Schmitz *et al.* 2016, Cloutier *et al.* 2019).

## 4. Conclusions

The future of phylogenomics and the quest for the ToL seems clear on at least one front: there will be a concerted effort to sequence the genomes of all life, including the genomes of complex eukaryotic orgnaisms (Lewin *et al.* 2018). Such an effort will not only provide a robust genomic foundation for building the ToL but will also present unprecedented opportunities for discovery in biology and for the conservation of biological diversity. Building the ToL will not only require large data sets but also intelligent analyses of such data sets. More data is not a panacea for phylogenetics (Delsuc *et al.* 2005, Philippe *et al.* 2011); rather, judicious use of available markers and appropriate models of evolution will be required. Increased dialogue between empiricists, theoreticians computer scientists and engineers and natural historians with deep knowledge of biodiversity will be essential to success. It's hard to imagine a time when the ToL is complete, but the journey towards its completion provides excitement enough.

## 5. Acknowledgements

I would like to thank Arup Hazarika for inviting me to write this perspective. My work and the work of my laboratory has been supported by the National Science Foundation (most recently NSF DEB-1831560 and DEB-1355343) and by Harvard University.

N.B: We are grateful to Prof. Scott V. Edwards and we take the previllage of reproducting this article in this volume of The Clarion which was published in Chapter form in the book Ecology, Environment & Conservation, published by CEEED for the benefit of the readers.

## *References*

Aktipis CA, Boddy AM, Jansen G, Hibner U, Hochberg ME, Maley CC, Wilkinson GS. 2015. Cancer across the tree of life: cooperation and cheating in multicellularity. Philos Trans R Soc Lond B Biol Sci, 370.

Avise JC. 2000. *Phylogeography: The History and Formation of Species*. Cambridge, MA, Harvard University Press.

Bininda-Emonds OR. 2005. Supertree construction in the genomic age. Methods Enzymol, 395:745-757.

Boore JL. 2006. The use of genome-level characters for phylogenetic reconstruction. Trends in Ecology & Evolution, 21:439-446.

Boore JL, Fuerstenberg SI. 2008. Beyond linear sequence comparisons: the use of genome-level characters for phylogenetic reconstruction. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 363:1445-1451.

Booth A, Mariscal C, Doolittle WF. 2016. The Modern Synthesis in the Light of Microbial Genomics. In: Gottesman S editor. Annual Review of Microbiology, Vol 70, p. 279-297.

Bravo GA, Antonelli A, Bacon CD, Bartoszek K, Blom MPK, Huynh S, Jones G, Knowles LL, Lamichhaney S, Marcussen T*, et al.* 2019. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. PeerJ, 7:e6399.

Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics, 30:3317-3324.

Churakov G, Sadasivuni MK, Rosenbloom KR, Huchon D, Brosius J, Schmitz J. 2010. Rodent Evolution: Back to the Root. Molecular Biology and Evolution, 27:1315-1326.

Cloutier A, Sackton TB, Grayson P, Clamp M, Baker AJ, Edwards SV. 2019. Whole-Genome Analyses Resolve the Phylogeny of Flightless Birds (Palaeognathae) in the Presence of an Empirical Anomaly Zone. Systematic Biology.

Costa IR, Prosdocimi F, Jennings WB. 2016. In silico phylogenomics using complete genomes: a case study on the evolution of hominoids. Genome Research, 26:1257-1267.

Cracraft J, Donoghue MJ. 2004. Assembling the Tree of Life. In: Cracraft J editor. New York, Oxford University Press, p. 468-489.

Daubin V, Moran NA, Ochman H. 2003. Phylogenetics and the cohesion of bacterial genomes. Science, 301:829-832.

de Queiroz A, Gatesy J. 2007. The supermatrix approach to systematics. Trends in Ecology & Evolution, 22:34-41.

Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M, Balhoff JP, Blackburn DC, Blake JA, Burleigh JG, Chanet B*, et al.* 2015. Finding our way through phenotypes. PLoS Biology, 13:e1002033.

Dececchi TA, Balhoff JP, Lapp H, Mabee PM. 2015. Toward Synthesizing Our Knowledge of Morphology: Using Ontologies and Machine Reasoning to Extract Presence/Absence Evolutionary Phenotypes across Studies. Syst Biol, 64:936-952.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference, and the multispecies coalescent. Trends in Ecology & Evolution, 24:332-340.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet, 6:361-375.

Doolittle WF. 2003. Some thoughts on the tree of life. Harvey Lect, 99:111-128.

Doolittle WF, Brunet TD. 2016. What Is the Tree of Life? PLoS Genet, 12:e1005912.

Eaton DAR, Spriggs EL, Park B, Donoghue MJ. 2017. Misconceptions on Missing Data in RAD-seq Phylogenetics with a Deep-scale Example from Flowering Plants. Syst Biol, 66:399-412.

Edmunds RC, Su B, Balhoff JP, Eames BF, Dahdul WM, Lapp H, Lundberg JG, Vision TJ, Dunham RA, Mabee PM*, et al.* 2016. Phenoscape: Identifying Candidate Genes for Evolutionary Phenotypes. Molecular Biology and Evolution, 33:13-24.

Edwards SV. 2009. Is a new and general theory of molecular systematics emerging? Evolution, 63:1-19.

Edwards SV. 2016. Inferring species trees. In: Kliman R editor. Encyclopedia of Evolutionary Biology. New York, Elsevier Inc., p. 236-244.

Edwards SV, Xi Z, Janke A, Faircloth BC, McCormack JE, Glenn TC, Zhong B, Wu S, Lemmon EM, Lemmon AR*, et al.* 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol Phylogenet Evol, 94:447-462.

Eernisse DJ, Kluge AG. 1993. Taxonomic Congruence Versus Total Evidence, and Amniote Phylogeny Inferred from Fossils, Molecules, and Morphology. Molecular Biology and Evolution, 10:1170-1195.

Eisenstein EM, Eisenstein DL, Sarma JS. 2016. An exploration of how to define and measure the evolution of behavior, learning, memory and mind across the full phylogenetic tree of life. Commun Integr Biol, 9:e1166320.

Ellegren H. 2007. Molecular evolutionary genomics of birds. Cytogenet Genome Res, 117:120-130.

Forterre P. 2015. The universal tree of life: an update. Front Microbiol, 6:717.

Gatesy J, Springer MS. 2013. Concatenation versus coalescence versus "concatalescence". Proceedings of the National Academy of Sciences of the United States of America, 110:E1179.

Geeta R, Lohmann LG, Magallon S, Faith DP, Hendry A, Crandall K, De Meester L, Webb CO, Prieur-Richard AH, Mimura M*, et al.* 2014. Biodiversity only makes sense in the light of evolution. J Biosci, 39:333-337.

Grover CE, Yu Y, Wing RA, Paterson AH, Wendel JF. 2008. A phylogenetic analysis of indel dynamics in the cotton genus. Mol Biol Evol, 25:1415-1428.

Harper JT, Gile GH, James ER, Carpenter KJ, Keeling PJ. 2009. The inadequacy of morphology for species and genus delineation in microbial eukaryotes: an example from the parabasalian termite symbiont coronympha. PLoS One, 4:e6577-e6577.

Hillis DM. 1999. SINEs of the perfect character. Proceedings of the National Academy of Sciences, 96:9979.

Jennings WB. 2017. Phylogenetic Data Acquisition: Principles and Practice. New York, CRC Press.

Jetz W, Rubenstein DR. 2011. Environmental uncertainty and the global biogeography of cooperative breeding in birds. Curr Biol, 21:72-78.

Jiang X, Edwards SV, Liu L. 2019. The Multispecies Coalescent Model Outperforms Concatenation across Diverse Phylogenomic Data Sets. in review.

Kaiser VB, van Tuinen M, Ellegren H. 2007. Insertion events of CR1 retrotransposable elements elucidate the phylogenetic branching order in galliform birds. Molecular Biology and Evolution, 24:338-347.

Kluge AG. 1989. A Concern for Evidence and a Phylogenetic Hypothesis of Relationships among Epicrates (Boidae, Serpentes). Systematic Zoology, 38:7-25.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceedings of the National Academy of Sciences (USA), 86:6196-6200.

Leaché AD, Oaks JR. 2017. The Utility of Single Nucleotide Polymorphism (SNP) Data in Phylogenetics. Annual Review of Ecology, Evolution, and Systematics, 48:69-84.

Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, Durbin R, Edwards SV, Forest F, Gilbert MTP*, et al.* 2018. Earth BioGenome Project: Sequencing life for the future of life. Proc Natl Acad Sci U S A, 115:4325-4333.

Li R. 2016. Using a mega-phylogeny of seed plants to test for non-random patterns of areal-types across the Chinese tree of life. Plant Divers, 38:283-288.

Liu L, Anderson C, Pearl D, Edwards SV. 2019. Modern Phylogenomics: Building Phylogenetic Trees Using the Multispecies Coalescent Model. In: Anisimova editor. Evolutionary Genomics: Statistical and Computational Methods. New York, Springer (Humana Press).

Liu L, Wu S, Yu L. 2015a. Coalescent methods for estimating species trees from phylogenomic data. Journal of Systematics and Evolution, 53:380-390.

Liu L, Xi Z, Wu S, Davis CC, Edwards SV. 2015b. Estimating phylogenetic trees from genome-scale data. Annals of the New York Academy of Sciences, 1360:36–53.

Liu L, Yu L, Edwards S. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evolutionary Biology, 10:302.

Liu L, Yu L, Pearl DK, Edwards SV. 2009. Estimating species phylogenies using coalescence times among sequences. Systematic Biology, 58:468-477.

Maddison DR. 2016. The rapidly changing landscape of insect phylogenetics. Curr Opin Insect Sci, 18:77-82.

Maddison WP. 1997. Gene trees in species trees. Systematic Biology, 46:523-536.

Maddison WP, Knowles LL. 2006. Inferring phylogeny despite incomplete lineage sorting. Systematic Biology, 55:21-30.

Mallet J, Besansky N, Hahn MW. 2015. How reticulated are species? Bioessays, 38:140-149.

McTavish EJ, Drew BT, Redelings B, Cranston KA. 2017. How and Why to Build a Unified Tree of Life. Bioessays, 39.

Mendes FK, Fuentes-Gonzalez JA, Schraiber JG, Hahn MW. 2018. A multispecies coalescent model for quantitative traits. Elife, 7.

Neigel JE, Avise JC. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: Karlin S, Nevo E editors. Evolutionary Processes and Theory. New York, Academic Press, Inc., p. 515-534.

Ninomiya S, Baret F, #xe9, #xe9, ric, Cheng Z-M. 2019. Plant Phenomics: Emerging Transdisciplinary Science %J Plant Phenomics, 2019:3.

O'Leary MA, Bloch JI, Flynn JJ, Gaudin TJ, Giallombardo A, Giannini NP, Goldberg SL, Kraatz BP, Luo ZX, Meng J*, et al.* 2013. The placental mammal ancestor and the post-K-Pg radiation of placentals. Science, 339:662-667.

Paabo S, Higuchi RG, Wilson AC. 1989. Ancient DNA and the polymerase chain reaction. Journal of Biological Chemistry, 264:9709-9712.

Pace NR. 2009. Mapping the tree of life: progress and prospects. Microbiol Mol Biol Rev, 73:565-576.

Parins-Fukuchi C. 2018. Use of Continuous Traits Can Improve Morphological Phylogenetics. Syst Biol, 67:328-339.

Patterson C, International Congress of S, Evolutionary B. 1987. Molecules and morphology in evolution : conflict or compromise? , Cambridge University Press.

Pease JB, Hahn MW. 2013. More accurate phylogenies inferred from low-recombination regions in the

presence of incomplete lineage sorting. Evolution; international journal of organic evolution, 67:2376-2384.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DT, Manuel M, Worheide G, Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol, 9:e1000602.

Pollard DA, Iyer VN, Moses AM, Eisen MB. 2006. Widespread discordance of gene trees with species tree in Drosophila: evidence for incomplete lineage sorting. PLoS Genet, 2:e173.

Rannala B, Edwards SV, Leaché AD, Yang Z. 2020. The multi-species coalescent model and species tree inference.

Rannala B, Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics, 164:1645-1656.

Rannala B, Yang Z. 2017. Efficient Bayesian species tree inference under the multispecies coalescent. Systematic Biology, 66:823-842.

Rogozin IB, Basu MK, Csürös M, Koonin EV. 2009. Analysis of Rare Genomic Changes Does Not Support the Unikont–Bikont Phylogeny and Suggests Cyanobacterial Symbiosis as the Point of Primary Radiation of Eukaryotes. Genome Biology and Evolution, 1:99-113.

Rohlf FJ, Sokal RR. 1967. Taxonomic Structure from Randomly and Systematically Scanned Biological Images. Systematic Biology, 16:246-260.

Rokas A. 2006. Genomics. Genomics and the tree of life. Science, 313:1897-1899.

Rokas A, Holland PWH. 2000. Rare genomic changes as a tool for phylogenetics. Trends in Ecology and Evolution, 15:454-459.

Sanderson MJ, Donoghue MJ. 1989. Patterns of variation in levels of homoplasy. Evolution, 43:1781-1795.

Schmitz J, Noll A, Raabe CA, Churakov G, Voss R, Kiefmann M, Rozhdestvensky T, Brosius J, Baertsch R, Clawson H*, et al.* 2016. Genome sequence of the basal haplorrhine primate Tarsius syrichta reveals unusual insertions. Nature Communications, 7:12997.

Scotland RW, Olmstead RG, Bennett JR. 2003. Phylogeny reconstruction: the role of morphology. Syst Biol, 52:539-548.

Shedlock AM, Okada N. 2000. SINE insertions: powerful tools for molecular systematics. Bioessays, 22:148-160.

Solis-Lemus C, Bastide P, Ane C. 2017. PhyloNetworks: A Package for Phylogenetic Networks. Mol Biol Evol, 34:3292-3298.

Solís-Lemus C, Yang M, Ané C. 2016. Inconsistency of Species Tree Methods under Gene Flow. Systematic Biology, 65:843-851.

Soltis DE, Soltis PS. 2018. The Great Tree of Life New York, Academic Press

Spriggs EL, Eaton DAR, Sweeney PW, Schlutius C, Edwards EJ, Donoghue MJ. 2019. Restriction-Site-Associated DNA Sequencing Reveals a Cryptic Viburnum Species on the North American Coastal Plain. Syst Biol, 68:187-203.

Stange M, Sánchez-Villagra MR, Salzburger W, Matschiner M. 2018. Bayesian Divergence-Time Estimation with Genome-Wide Single-Nucleotide Polymorphism Data of Sea Catfishes (Ariidae) Supports Miocene Closure of the Panamanian Isthmus. Systematic Biology, 67:681-699.

Steel M, Dress AW, Bocker S. 2000. Simple but fundamental limitations on supertree and consensus tree methods. Systematic Biology, 49:363-368.

Stenz NW, Larget B, Baum DA, Ané C. 2015. Exploring tree-like and non-tree-like patterns using genome sequences: An example using the inbreeding plant species Arabidopsis thaliana (L.) Heynh. Systematic Biology:syv039.

Suh A, Smeds L, Ellegren H. 2015. The Dynamics of Incomplete Lineage Sorting across the Ancient Adaptive Radiation of Neoavian Birds. PLoS Biology, 13:e1002224.

Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M. 2017. Plant Phenomics, From Sensors to Knowledge. Current Biology, 27:R770-R783.

Thessen AE, Bunker DE, Buttigieg PL, Cooper LD, Dahdul WM, Domisch S, Franz NM, Jaiswal P, Lawrence-Dill CJ, Midford PE*, et al.* 2015. Emerging semantics to link phenotype and environment. PeerJ, 3:e1470.

Webb CO, Ackerly DD, McPeek MA, Donoghue MJ. 2002. Phylogenies and Community Ecology. Annual Review of Ecology and Systematics, 33:475-505.

Wiens J. 2004. The role of morphological data in phylogeny reconstruction. Syst Biol, 53:653-661.

Wiens JJ. 2001. Character analysis in morphological phylogenetics: problems and solutions. Syst Biol, 50:689-699.

Wirkner CS, Gopel T, Runge J, Keiler J, Klussmann-Fricke BJ, Huckstorf K, Scholz S, Miko I, M JY, Richter S. 2017. The First Organ-Based Ontology for Arthropods (Ontology of Arthropod Circulatory Systems - OArCS) and its Integration into a Novel Formalization Scheme for Morphological Descriptions. Syst Biol, 66:754-768.

Wolf YI, Rogozin IB, Grishin NV, Koonin EV. 2002. Genome trees and the tree of life. Trends Genet, 18:472-479.

Wu C, Lu J. 2019. Diversification of Transposable Elements in Arthropods and Its Impact on Genome Evolution. Genes, 10.

Yu Y, Nakhleh L. 2015. A maximum pseudo-likelihood approach for phylogenetic networks. BMC Genomics, 16 Suppl 10:S10.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics, 19:153.

□□□