Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Insect Science

The rapidly changing landscape of insect phylogenetics

David R Maddison

Insect phylogenetics is being profoundly changed by many innovations. Although rapid developments in genomics have center stage, key progress has been made in phenomics, field and museum science, digital databases and pipelines, analytical tools, and the culture of science. The importance of these methodological and cultural changes to the pace of inference of the hexapod Tree of Life is discussed. The innovations have the potential, when synthesized and mobilized in ways as yet unforeseen, to shine light on the million or more clades in insects, and infer their composition with confidence. There are many challenges to overcome before insects can enter the 'phylocognisant age', but because of the promise of genomics, phenomics, and informatics, that is now an imaginable future.

Address
Department of Integrative Biology, 3029 Cordley Hall, Oregon State University, Corvallis, OR 97331, USA

Corresponding author: Maddison, David R
(david.maddison@oregonstate.edu)

Current Opinion in Insect Science 2016, **18**:77–82

This review comes from a themed issue on **Insect phylogenetics**

Edited by **Gregory W Courtney** and **Brian M Wiegmann**

For a complete overview see the Issue and the Editorial

Available online 20th October 2016

http://dx.doi.org/10.1016/j.cois.2016.09.007

Introduction

Our understanding of the evolutionary Tree of Life has grown dramatically over the last few decades, driven in part by the need for knowledge about the Tree. As the conduits along which genetic information has flowed, the branches of the Tree of Life are the pathways along which organismal diversity that we see today was formed. Knowledge of the branching (and anastomosing) patterns of the Tree form a vital underpinning to all of comparative biology: any study of multiple specimens is informed by the historical pathways of genes that connected the specimens. Our explanations of the similarities and differences of two specimens that belong to closely related species will differ from our interpretations if they are distantly related, on branches whose ancestors separated hundreds of millions of years ago. Although Darwin [1] understood the importance of considering phylogenetic relationships when interpreting evolutionary patterns ('We may often falsely attribute to correlated variation structures which are common to whole groups of species, and which in truth are simply due to inheritance'), it only came to be understood broadly by the biological community a few decades ago, beginning with the seminal paper by Felsenstein [2].

Much of the effort to reconstruct a complete picture of the Tree of Life will need to focus within insects. With nearly a million described species, and over half of all known species [3], hexapods include a significant portion of the Tree. A full view of their phylogeny will require inferring about a million clades, a number that will increase as more species are discovered.

Over the last few decades we have made tremendous progress in understanding the shape of deeper portions of insect phylogeny, and within scattered groups at finer scales. Most attention has focused on the ancient nodes of the tree, and we have resolved several key aspects in that region [4••]. For example, we are now confident that Polyneoptera are monophyletic [5••], and Strepsiptera are related to beetles, not flies [5••]. Substantial progress has been made in family-level relationships within several orders (e.g., in Lepidoptera [6,7]), and at various levels within scattered families. However, we are far from a full understanding of insect phylogeny, with many parts of the tree unresolved and unstudied. Some of these are ancient branches that have been well-studied, but without resolution; for example, relationships of Collembola, Protura, and Diplura are not yet clear [4••]. Most of the more recent divergences are unstudied, as many thousands of species have never been included in any phylogenetic analysis, morphological or molecular.

Progress has been made through a rapidly shifting landscape of methods and approaches that have changed profoundly over the last decade. While not perhaps reaching the depth of worldview-altering Kuhnian paradigm shifts [8], they have nonetheless significantly altered our current understanding of insect phylogeny, and will set the vision for future research. We are entering a time in which genetic and phenotypic data need not be limited except by inherent limits in the organism, in which the focus is shifting toward how we can best store, distribute, analyze, and present massive quantities of data and results. Even field and museum methods are undergoing notable shifts, which could enable or hinder progress. A decade ago we might not have been able to imagine a future in which we understand with confidence the phylogeny of insects, but today we can envisage such a future, although there are many challenges to overcome.

The genomics revolution

A major engine of this change has been development of relatively inexpensive methods to sequence significant portion of genomes: the cost of genomic sequencing has dropped to less than 0.0000026 of what it was in 2001 (Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP), https://www.genome.gov/sequencingcostsdata/). As sequencing methods are invented and compete (Glen T. The 2016 Next-Generation Sequencing Field Guide Preview: Zombie Systems and New Hope, http://www.molecularecologist.com/2016/03/2016-ngs-field-guide-preview/), the trend of reducing costs will surely continue, and will further increase the availability, worldwide, of genomic data.

The cost of sequencing a whole genome varies tremendously within insects, as their genome size ranges from less than 200 Mb in various flies, beetles, and wasps to 16.6 Gb in grasshoppers [9]. However, even for the largest insect genomes it is now possible to obtain 20× coverage for relatively low cost; the price for such sequencing at Center for Genome Research and Biocomputing at Oregon State University is about $8000. The cost is not yet trivial, and considering the millions of insects that will need to be sequenced to both understand the fine-scale phylogeny around the species boundaries and the deeper branches, from both developed and developing countries, some method of reducing the cost further is necessary.

Reduced-representation methods save costs by diminishing the portion of the genome to be sequenced. These include transcriptome sequencing [e.g., 5] and targeted enrichment or hybrid capture; the latter biases sequencing effort toward genomic sections that hybridize to designed probes [10]. These two methods provide data that can be compared in phylogenetically distant taxa, and thus can be incorporated more easily into future studies. They can also yield relatively long fragments whose genes trees might be more accurately inferred (although this has been challenged because of the consequent increase in probability of recombination points within the sequence [11]). By contrast, genotype-by-sequencing (GBS) [12] or restriction-site associated DNA sequencing (RADSeq) [e.g., 13] choose more or less random sites in the genome and sequence relatively small fragments at those sites. They have the advantage of easier and cheaper data generation from more regions spread throughout the genome, and have minimal setup costs [14]. However, GBS and RADSeq produce data from anonymous loci whose orthologs are less likely to be sequenced in phylogenetically distant taxa, making the methods more useful for studies within younger clades [15]. For the same reason, the data from GBS and RADSeq cannot as easily be incorporated into data from other taxa (at least not until full genomes are available as context) [14].

With sequence data limited only by the DNA content of the organism, we may hope to fully resolve with surety the phylogeny. This has been accomplished with some ordinal-level relationships in insects that were elusive, such as the placement of Hymenoptera, which has now been confidently determined [5••,16]. Extensive sequence data will not always suffice, however: even with 413 459 amino acid sites, the relationships of Odonata, Ephemeroptera, and Neoptera has not been confidently resolved, suggesting the need for metagenomic characters and additional taxon sampling [5••]. We should not always expect that the result of phylogenomic studies will be a well-supported dichotomous tree: in some cases the internodes will be nonexistent (i.e., there was effectively simultaneous divergence into three or more lineages) or are so short as to yield many conflicting gene trees [e.g., 17]; this is expected to be true even at old branch points within the tree [18].

Phenomics: the return of morphology

Genomics has not provided the only data windfall in insect phylogenetics. After an early reliance on morphology to supply markers needed for phylogenetic inference, and subsequent neglect with the rise of DNA sequencing technologies, structural data is undergoing a revival. A core driver of this is the need for a time-calibrated phylogeny by downstream consumers of phylogenetic knowledge, and the consequent necessity of incorporation of dated fossils into analyses. As morphological matrices are required to properly incorporate fossils in a phylogenetic analysis [19•,20], the task of gathering morphological data has been revived [21,22,23••,24•]. The discovery of important new fossil sites [25], including additional Cretaceous amber deposits, has also reinvigorated insect morphology.

The flourishing of phenomics is spurred by new methodological developments, both in data-gathering tools and in knowledge-representation methods. MicroCT and confocal microscopy, for example, offer unprecedented three-dimensional reconstructions of insect structure [22,23••], and efficient means of digitally capturing large quantities of data much more rapidly than traditional, eyes-to-the-microscope methods. Also significant is formalization of ways to describe and communicate about morphological traits through ontologies [26••], as embodied, for example, in the Phenoscape project (http://phenoscape.org). These developments have the potential to allow morphological data to enter the high-throughput world that to date has been the domain of genomics.

Efficiency and taxes in the field and museum

As most specimens in the world's museums were not killed and stored in ways designed to preserve DNA or RNA, additional collecting is often warranted. This effort is being transformed by new tools for collecting (e.g., LED-based UV light traps [27]) and documenting field sites (smartphones, drones). Extensive online data (e.g., aerial photos and street views) guide us to potential field sites; portable digital devices containing (or connecting to) the world's literature and images allow for informed in-the-field decisions. 'Third-generation' sequencing methods [28], through which extensive data can be gathered in the field via very small hand-held sequencing devices, will also revolutionize the efficiency of field work, allowing quick feedback about sampling strategies.

But not all changes have led to greater efficiency. The increasing regulation of collecting, exporting, and importing insects has led to a 'time tax', requiring an increasing commitment of time and funds to writing numerous and diverse permit applications and subsequent reports. Some faunas in biodiversity hotspots have become effectively impossible to sample because of bureaucratic burden or presence of absolute restrictions. To my knowledge there has not been a study of this problem and its effects on phylogenetics, but I suspect any active phylogeneticist who studies a widely distributed clade has encountered roadblocks that deter research.

A partial workaround is the discovery that many dried insects already in the world's museums contain sufficient DNA to yield sequence data of at least multi-copy genes (e.g., mitochondrial and ribosomal) when next-generation sequencing methods are applied [29,30•]. This should enable most known species of insects to be incorporated into molecular phylogenetic studies, although often with relatively few genes.

New methods enable more efficient decisions about specimens to examine in depth. These include image recognition systems [31], and sequencing of a diagnostic gene fragment for less than $1 per specimen [32]. The increasing digitization of the world's collections (see, e.g., iDigBio, https://www.idigbio.org) will also play a notable role in enabling better choices about specimens from which we should gather data.

Digital data, workflows, algorithms, and software

The massive quantities of data and results that are being produced about insects need efficient systems in which they can be safely stored, through which they can flow, and via which they can be interconnected and analyzed. The many data objects (about collecting events, specimens and their derivatives such as DNA extractions, morphological data, molecular data, and analytical results) that a modern lab needs to manage in the face of less-than-ideal digital tools puts an upper limit on throughput, and the number of species that can be studied phylogenetically is thereby limited. Progress on developing systems that will bring order to poorly organized, insufficiently linked and tracked data is being made. One tool that promises to help the data-gathering phase is TaxonWorks (http://taxonworks.org), which should increase pace and volume. Repositories such as GenBank, MorphBank (http://morphbank.net), MorphoBank (http://morphobank.org), and TreeBase (http://treebase.org) can store primary data long term. Data annotation and exchange standards (e.g., [33]) allow efficient data flow and reuse. New phylogeny inference tools are being continually developed (e.g., [34–36,37•,38]). More systems are being developed to take advantage of the Semantic Web. For example, OpenTree [39••] provides synthesis of resulting phylogenetic trees, built with scalability in mind, and with application programming interfaces (APIs) that allow easy harvesting by other elements in a workflow pipeline. There is an outburst of innovation in comparative biology, including methods synthesized in Arbor [40••], a scalable system for post-tree analysis of trees from OpenTree and elsewhere.

An efficient system to tackle all of insect phylogeny will require greater openness to data sharing and collaboration. The increasing awareness of the benefits of openly available data [41] are key to the effort. Cultural changes toward open and distributed science (http://opensciencefederation.com, [42]) and credit systems acknowledging the diversity of contributions in systematics [43] should increase the pace. Although formalizing credit for newer publication formats is vital for this enterprise, adoption may be slow; for example, this journal excludes websites, computer programs, and databases from formal citation in reference lists.

However, there are major challenges that must be overcome, each of which could serve as the impetus for a focused research effort. For example, a complete taxonomic name resolution service (e.g., [44]), which is needed to allow databases to share data about a taxon, is far from being fully implemented; without it there will be major barriers in the flow of information. Similarly, determining which sequences in one taxon are orthologs to those in a second taxon is a crucial step still requiring refinement [45]. Although we can quickly have thousands of genes at our disposal, we cannot appropriately model all of them, and yet methods, such as posterior predictive simulations [46•,47], to determine which genes we can properly analyze, are just starting to be developed. The complexity of gene tree diversity caused by recombination, and its subsequent effects on the meaning of the species tree [48] and our ability to infer it [11,49] will need to be more satisfactorily resolved. Once phylogenetic results are obtained, our ability to synthesize them into a more global view is limited, in part because of the lack of community involvement in depositing information into repositories such as TreeBase, and thus inspiring more extensive data deposition will be required. Once synthesized, we cannot easily visualize our results, as tree visualization has yet to produce an accepted focus-plus-context visualization [50] for a phylogeny the size of the hexapod tree. And, finally, as databases and software systems grow in complexity and require teams of people to develop and maintain, and the interdependence of these systems compels each to survive for decades, the need to develop communities and funding structures that enable sustainability will become more urgent [51].

The trajectory of insect phylogenetics

Phylogenetics of insects began with a head start relative to many other taxa, as the leader in early development of explicit phylogenetic inference methods was Willi Hennig [52], with his 1969 tome *Die Stammesgeschichte der Insekten* [53] standing as a major early milestone in phylogenetics. Although phylogenetic thinking has permeated systematic entomology for decades, the size of the insect branch means we have only completed a fraction of the task, even with the availability of new genomic, phenomic, and informatic tools. The size of the unknown parts of the insect clade also means that a significant fraction of systematic research on insects is still (appropriately) dominated by species discovery and documentation.

Insect phylogenetics has been both enabled and slowed down by the nature of insects. By contrast to other extremely diverse clades, such as fungi and bacteria, most insect specimens can be easily preserved in their original exoskeletal form (or nearly so) in museums. They exhibit complex, developmentally well-canalized structures that are of a size suitable for examination under common microscopes (and those of the 1800s and 1900s). These traits encouraged a rich history of morphological research that provided data for Hennig and other researchers. Their complex structures and our ability to see them have allowed us to uncover their diversity. This legacy of enormous known diversity and abundant morphological data and specimens has been tremendously valuable, but it may also have hindered the transition to the digital and ‘omics’ world. Microbial groups, similarly diverse (if not more so), but with fewer specimens in collections, and a much less extensive literature in morphological systematics, and many fewer described species, with much smaller genomes [54,55], have the potential to leapfrog insect phylogenetics by the necessity of embracing modern tools, in similar fashion to the leapfrogging observed in the more-thorough or rapid adoption of technological tools in developing countries [56]. Insect phylogenetics will need to examine its legacy, and determine which parts of our existing resources in collections and on paper can serve as a springboard rather than a burden as we embrace newer tools and data.

Will insects enter the phylocognisant age?

In smaller, more extensively studied clades, for which comprehensive phylogenies have been constructed through the phases (morphological, Sanger-sequencing-based molecular, and now genomic) of phylogenetics, the ‘phylocognisant age’ is at hand or on the horizon: an era in which we know with confidence most of the phylogeny from ancestral node to species. For example, in birds, this age appears close at hand: the B10K project seeks to sequence the genome of all 10 000+ species of living birds by 2020 [57], and thereby reconstruct the phylogeny of birds. One can envisage that Avian phylogeny may be resolved to a relatively fine resolution (even if not completely: [58]) within 10 years.

It will take significantly longer for insects to enter the phylocognisant age. Much of the community focus to date has been on deeper nodes of insect phylogeny, but most of the branches within insects are more recent, within families. Most studies in evolutionary processes, ecology, physiology, and so forth that require observations of repeated patterns on the phylogeny will be most informed by these abundant, shallower branchings. The first research project that deciphers the phylogeny of a 10 000-species clade in insects may be considered ‘transformative’, cutting edge research that is so coveted by funding agencies. Deciphering the phylogeny of the 90th such clade likely would not, and it is unclear how the extensive but necessary work to map the whole tree might be funded. However, a full tree of insect life will be vital if we are to avail ourselves of the fruits of comparative biology within this giant clade. One can hope that the need for this will lead to a rethinking of our approach to phylogenetics, perhaps by the development of radical new methods, and synthesis of old ones, into an effort that can tackle a tree with over a million branches.

Funding

This work was supported by the Harold E. and Leona M. Rice Endowment Fund at Oregon State University.

Acknowledgements

My thanks to Wayne Maddison, for discussions over the years on these topics, and for comments on the manuscript. I also thank Kojun Kanda, John S. Sproul, R. Antonio Gomez, and Wendy Moore for comments on the manuscript.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Darwin CR: *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray; 1872.

2. Felsenstein J: **Phylogenies and the comparative method**. *Am Nat* 1985, **125**:1-15.

3. Chapman AD: *Numbers of Living Species in Australia and the World*. edn 2nd. Australian Biological Resources Study; 2009.

4. •• Yeates DK, Meusemann K, Trautwein M, Wiegmann B, Zwick A: **Power, resolution and bias: recent advances in insect phylogeny driven by the genomic revolution**. *Curr Opin Insect Sci* 2016, **13**:16-23.

The authors provide an overview of the major advances in insect phylogenomics, with a discussion of the analytical challenges (including potential systematic biases) and insight into future research directions

5. Misof B, Liu SL, Meusemann K, Peters RS, Donath A, Mayer C,
•• Frandsen PB, Ware J, Flouri T, Beutel RG *et al.*: **Phylogenomics resolves the timing and pattern of insect evolution**. *Science* 2014, **346**:763-767.
This paper is the first attempt at a comprehensive, transcriptome-based phylogeny of all of the major lineages of hexapods, based upon 1478 orthologous single-copy nuclear protein-coding genes, with an extremely detailed phylogenomic analysis

6. Kawahara AY, Breinholt JW: **Phylogenomics provides strong evidence for relationships of butterflies and moths**. *Proc R Soc Lond, Ser B: Biol Sci* 2014:281.

7. Regier JC, Mitter C, Zwick A, Bazinet AL, Cummings MP, Kawahara AY, Sohn J-C, Zwickl DJ, Cho S, Davis DR *et al.*: **A large-scale, higher-level, molecular phylogenetic study of the insect order lepidoptera (moths and butterflies)**. *PLoS ONE* 2013, **8**:e58568.

8. Kuhn TS: *The Structure of Scientific Revolutions*. edn 2nd. The University of Chicago Press; 1970.

9. Hanrahan SJ, Johnston JS: **New genome size estimates of 134 species of arthropods**. *Chromosome Res* 2011, **19**:809-823.

10. Jones MR, Good JM: **Targeted capture in evolutionary and ecological genomics**. *Mol Ecol* 2016, **25**:185-202.

11. Springer MS, Gatesy J: **The gene tree delusion**. *Mol Phylogen Evol* 2016, **94**:1-33.

12. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE: **A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species**. *PLoS ONE* 2011, **6**:10.

13. Darwell CT, Rivers DM, Althoff DM: **Rad-seq phylogenomics recovers a well-resolved phylogeny of a rapid radiation of mutualistic and antagonistic yucca moths**. *Syst Entomol* 2016, **41**:672-682.

14. Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT: **Sequence capture versus restriction site associated DNA sequencing for shallow systematics**. *Syst Biol* 2016, **65**:910-924 10.1093/sysbio/syw036.

15. Rubin BER, Ree RH, Moreau CS: **Inferring phylogenies from rad sequence data**. *PLoS ONE* 2012, **7**:e33394.

16. Savard J, Tautz D, Richards S, Weinstock GM, Gibbs RA, Werren JH, Tettelin H, Lercher MJ: **Phylogenomic analysis reveals bees and wasps (hymenoptera) at the base of the radiation of holometabolous insects**. *Genome Res* 2006, **16**:1334-1338.

17. Pollard DA, Iyer VN, Moses AM, Eisen MB: **Widespread discordance of gene trees with species tree in drosophila: evidence for incomplete lineage sorting**. *PLoS Genet* 2006, **2**:1634-1647.

18. Oliver JC: **Microevolutionary processes generate phylogenomic discordance at ancient divergences**. *Evolution* 2013, **67**:1823-1830.

19. Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL,
• Rasnitsyn AP: **A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera**. *Syst Biol* 2012, **61**:973-999.
The authors provide a detailed presentation of an approach for dating phylogenetic divergences by incorporating fossils as terminal taxa in a Bayesian analysis combining DNA sequence data and a morphological matrix, as opposed to a sequential analysis whereby fossil data was condensed down into a simple constraint on the age of a node before phylogenetic analysis

20. Pyron RA: **Divergence time estimation using fossils as terminal taxa and the origins of lissamphibia**. *Syst Biol* 2011, **60**:466-481.

21. Lee MSY, Palci A: **Morphological phylogenetics in the genomic age**. *Curr Biol* 2015, **25**:R922-R929.

22. Deans AR, Miko I, Wipfler B, Friedrich F: **Evolutionary phenomics and the emerging enlightenment of arthropod systematics**. *Invertebr Syst* 2012, **26**:323-330.

23. Friedrich F, Matsumura Y, Pohl H, Bai M, Hornschemeyer T,
•• Beutel RG: **Insect morphology in the age of phylogenomics: innovative techniques and its future role in systematics**. *Entomol Sci* 2014, **17**:1-24.
This paper provides an overview of some of the novel tools available in microscopy and tomography for gathering extensive morphological data in insects

24. Giribet G: **Morphology should not be forgotten in the era of**
• **genomics — a phylogenetic perspective**. *Zool Anz* 2015, **256**:96-103.
The author argues for the value of phenomics in insect phylogenetics, and discusses the increasing relevance of morphological data and the new tools available to acquire and process it

25. Nicholson DB, Mayhew PJ, Ross AJ: **Changes to the fossil record of insects through fifteen years of discovery**. *PLoS ONE* 2015, **10**:e0128554.

26. Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M,
•• Balhoff JP, Blackburn DC, Blake JA, Burleigh JG, Chanet B *et al.*: **Finding our way through phenotypes**. *PLoS Biol* 2015, **13**:9.
The authors review the current status of the transformation of morphology into a rigorously defined, ontology-based field, and discuss the advantages in data integration and analysis this approach would entail

27. White PJT, Glover K, Stewart J, Rice A: **The technical and performance characteristics of a low-cost, simply constructed, black light moth trap**. *J Insect Sci* 2016:16.

28. Bleidorn C: **Third generation sequencing: technology and its potential impact on evolutionary biodiversity research**. *Syst Biodivers* 2016, **14**:1-8.

29. Staats M, Erkens RHJ, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT: **Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens**. *PLoS ONE* 2013:8.

30. Kanda K, Pflug JM, Sproul JS, Dasenko MA, Maddison DR:
• **Successful recovery of nuclear protein-coding genes from small insects in museums using illumina sequencing**. *PLoS ONE* 2015, **10**:e0143929.
This paper documents the ability of next-generation methods to sequence DNA from small, old, dried insects stored in museums, yielding accurate sequences from multicopy (ribosomal and mtDNA) genes as well as single-copy protein-coding genes

31. Favret C, Sieracki JM: **Machine vision automated species identification scaled towards production levels**. *Syst Entomol* 2016, **41**:133-143.

32. Meier R, Wong W, Srivathsan A, Foo M: **$1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples**. *Cladistics* 2016, **32**:100-110.

33. Panahiazar M, Sheth AP, Ranabahu A, Vos RA, Leebens-Mack J: **Advancing data reuse in phyloinformatics using an ontology-driven semantic web approach**. *BMC Medical Genomics* 2013, **6**:10.

34. Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F: **Revbayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language**. *Syst Biol* 2016, **65**:726-736 10.1093/sysbio/syw021.

35. Kozlov AM, Aberer AJ, Stamatakis A: **Examl version 3: a tool for phylogenomic analyses on supercomputers**. *Bioinformatics* 2015, **31**:2577-2579.

36. Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T: **Astral: genome-scale coalescent-based species tree estimation**. *Bioinformatics* 2014, **30**:i541-i548.

37. Chifman J, Kubatko L: **Quartet inference from snp data under**
• **the coalescent model**. *Bioinformatics* 2014, **30**:3317-3324.
The authors present a new phylogenetic inference method utilizing invariants, examining base patterns at single sites for quartets of sequences. This is an example of a novel method that avoids much of the model-dependence that pervades most inference methods

38. Yu Y, Dong JR, Liu KJ, Nakhleh L: **Maximum likelihood inference of reticulate evolutionary histories**. *Proc Natl Acad Sci U S A* 2014, **111**:16448-16453.

39. Hinchliff CE, Smith SA, Allman JF, Burleigh JG, Chaudhary R,
•• Coghill LM, Crandall KA, Deng J, Drew BT, Gazis R *et al.*:

Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc Natl Acad Sci U S A* 2015, **112**:12764-12769.
The authors present new methods of synthesizing inferred phylogenetic trees into a larger tree, with incorporation of taxonomies that enable almost all known species to be represented in an online database, which exposes the data through a well-defined API that allows the full Tree of Life to be part of a large-scale analytical pipeline

40. •• Harmon LJ, Baumes J, Hughes C, Soberon J, Specht CD, Turner W, Lisle C, Thacker RW: **Arbor: Comparative Analysis Workflows for the Tree of Life**. *PLOS Currents Tree of Life* 2013. Edition 1. doi:10.1371/currents.tol.099161de5eabdee073fd3d21a44518dc.
The authors describe an easy to use, scalable workflow system incorporating R and Python packages to allow large-scale comparative analysis of web-accessible data on phylogenies. Jun 21

41. Cranston K, Harmon LJ, O'Leary MA, Lisle C: **Best practices for data sharing in phylogenetic research**. *PLoS Currents Tree of Life* 2014 http://dx.doi.org/10.1371/currents.Tol.Bf01eff4a6b60ca4825c69293dc59645. jun 19. Edition 1 2014 Jun 19.

42. Maddison DR, Guralnick R, Hill A, Reysenbach AL, McDade LA: **Ramping up biodiversity discovery via online quantum contributions**. *Trends Ecol Evol* 2012, **27**:72-77.

43. McDade LA, Maddison DR, Guralnick R, Piwowar HA, Jameson ML, Helgen KM, Herendeen PS, Hill A, Vis ML: **Biology needs a modern assessment system for professional productivity**. *Bioscience* 2011, **61**:619-625.

44. Boyle B, Hopkins N, Lu ZY, Garay JAR, Mozzherin D, Rees T, Matasci N, Narro ML, Piel WH, McKay SJ *et al.*: **The taxonomic name resolution service: an online tool for automated standardization of plant names**. *BMC Bioinformatics* 2013, **14**:14.

45. Boeckmann B, Marcet-Houben M, Rees JA, Forslund K, Huerta-Cepas J, Muffato M, Yilmaz P, Xenarios I, Bork P, Lewis SE *et al.*: **Quest for orthologs entails quest for tree of life: in search of the gene stream**. *Genome Biol Evol* 2015, **7**:1988-1999.

46. • Brown JM: **Detection of implausible phylogenetic inferences using posterior predictive assessment of model fit**. *Syst Biol* 2014, **63**:334-348.
This paper describes one of the first attempts to crucially examine the fit of the models we use in phylogenetic inference to the data we have; model fit will be an increasingly important field of study was we seek to choose more carefully the data we use in phylogenetics

47. Duchene S, Di Giallonardo F, Holmes EC: **Substitution model adequacy and assessing the reliability of estimates of virus evolutionary rates and time scales**. *Mol Biol Evol* 2016, **33**:255-267.

48. Maddison WP: **Gene trees in species trees**. *Syst Biol* 1997, **46**:523-536.

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

50. Lamping J, Rao R: **The hyperbolic browser: a focus plus context technique for visualizing large hierarchies**. *J Visual Lang Computing* 1996, **7**:33-55.

51. Reiser L, Berardini TZ, Li DH, Muller R, Strait EM, Li Q, Mezheritsky Y, Vetushko A, Huala E: **Sustainable funding for biocuration: the Arabidopsis information resource (TAIR) as a case study of a subscription-based funding model**. *Database - J Biol Databases Curation* 2016:12.

52. Hennig W: *Grundzüge einer theorie der phylogenetischen systematik*. Deutscher Zentralverlag; 1950.

53. Hennig W: *Die stammesgeschichte der insekten*. W. Kramer; 1969.

54. Mohanta TK, Bae H: **The diversity of fungal genome**. *Biol Proced Online* 2015, **17**:9.

55. Koonin EV, Wolf YI: **Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world**. *Nucleic Acids Res* 2008, **36**:6688-6719.

56. Sharif MN: **Technological leapfrogging — implications for developing countries**. *Technol Forecast Soc Change* 1989, **36**:201-208.

57. Zhang GJ: **Bird sequencing project takes off**. *Nature* 2015, **522** 34-34.

58. Hahn MW, Nakhleh L: **Irrational exuberance for resolved species trees**. *Evolution* 2016, **70**:7-17.