

The future of supertrees: bridging the gap with supermatrices¹

OLAF R. P. BININDA-EMONDS

Abstract

The supertree and supermatrix frameworks have been cast as mutually exclusive approaches toward the problem of large-scale phylogenetic inference. Despite often coming under severe criticism, the supertree approach has proven superior to date at deriving comprehensive phylogenetic estimates for many groups (e. g., mammals as a whole) because by combining trees instead of characters, it is able to include more of the global phylogenetic database. The continued rapid growth in sequencing technologies, however, means that this advantage is time-limited given that abundant sequence information will rapidly become available for many groups. What then does such a future hold for the supertree approach? In this paper, I argue that the supertree framework could continue to have a place in phylogenetic inference, albeit altered to play a subordinate role as part of a divide-and-conquer heuristic search strategy for large molecular supermatrices. However, the divide-and-conquer approach has yet to realize its theoretical advantages (in terms of both speed and accuracy) over more conventional heuristic search strategies. I discuss two potential supertree-related bottlenecks that appear to be limiting the performance of the divide-and-conquer approach and which can be viewed as problems for which solutions need to be sought.

Key words: Supertree, supermatrix, phylogenetic inference.

Zusammenfassung

Supertree- und Supermatrix-Systeme entstanden als sich gegenseitig ausschließende Analysemethoden des Problems umfangreicher phylogenetischer Untersuchungen. Trotz zahlreicher Kritik hat sich die Supertree-Methode als überaus erfolgreich beim Erstellen sehr großer phylogenetischer Datensätze (z. B. Säugetiere) erwiesen, da durch die Kombination von Stammbäumen statt von Merkmalen Informationen aus der globalen phylogenetischen Datenbank einbezogen werden können. Dieser methodische Vorteil nimmt jedoch in dem Maß ab, wie die Datenmenge genetischer Sequenzen zunimmt, die in kürzester Zeit für immer mehr Gruppen zur Verfügung stehen. Was bedeutet diese Zunahme an Sequenzdaten für die Zukunft der Supertree-Analyse? Im Folgenden argumentiere ich, dass Supertrees für phylogenetische Analysen weiterhin relevant bleiben, wenngleich sie nur noch eine untergeordnete Rolle als Teil eines Divide-and-Conquer-heuristischen Verfahrens für große molekulare Supermatrices spielen werden. Die zumindest theoretisch vorhandenen Vorteile (wie z. B. Geschwindigkeit und Genauigkeit) dieser Divide-and-Conquer Verfahren gegenüber konventionellen heuristischen Methoden müssen sich jedoch in der Praxis noch bewähren. Im Folgenden diskutiere ich zwei mögliche, durch die Supertree-Analyse entstehende Probleme, welche die Effizienz von Divide-and-Conquer-Verfahren limitieren, und zu deren Lösung neue Strategien entwickelt werden müssen.

Contents

1. Introduction	99
2. Supertrees, supermatrices, and global congruence	101
3. Dividing and conquering: supertrees as a heuristic search tool	102
4. Conclusions	104
5. References	105

1. Introduction

Since the first large-scale supertree of Primates was published by PURVIS (1995), supertree construction has proven itself to be a highly useful method for constructing large, comprehensive phylogenetic trees for a wide variety of organismal groups. Indeed, many of the published supertrees represented the first complete phylogenetic es-

timates for the groups in question (for a now outdated list, see BININDA-EMONDS 2004a) and, in many cases, remain the only such estimates.

Despite their utility, supertrees nevertheless have been strongly criticized by many (e. g., NOVACEK 2001; SPRINGER & DE JONG 2001; GATESY et al. 2002; GATESY & SPRINGER 2004) and there has been much heated debate about the relative merits of the supertree and supermatrix

¹ Contribution to the WILLI-HENNIG-Symposium on Phylogenetics and Evolution, University of Hohenheim, 29 September – 2 October 2009.

frameworks of phylogenetic inference (see the exchange between myself and JOHN GATESY in particular: GATESY et al. 2002; BININDA-EMONDS et al. 2003; BININDA-EMONDS 2004b; GATESY et al. 2004). Criticisms of supertrees range from shortcomings in specific implementations of the method to perceived shortcomings that are fundamental to the method itself. Among the latter, the key concern is that supertrees, by combining and analyzing the trees derived from character data rather than analyzing the data directly, represent a meta-analysis one step removed from the real data (GATESY & SPRINGER 2004). The inherent loss of information this fact entails automatically translates for some to an inherent decrease in accuracy compared to a supermatrix analysis. Instead, the meta-analysis nature of supertrees can also be viewed as a potential strength. Because issues of character data combinability do not affect supertree construction, more of the total phylogenetic database can be used to derive the evolutionary trees. This fact, in large measure, accounts for the ability of super-

trees to obtain more comprehensive phylogenetic trees for most groups than is currently possible using a supermatrix approach. Even so, it is held by some that phylogenetic supertrees merely represent a stopgap measure for phylogenetic inference until sufficient molecular data become available to enable the more desirable supermatrix analyses. With the ever-increasing pace and ever-decreasing costs of high throughput sequencing, this opinion does have a certain validity to it, albeit much more so for charismatic groups (e. g., mammals or flowering plants) than for other far less well studied ones (e. g., rotifers and many other microfaunal taxa).

What then does the future hold for supertrees, if anything? As I argued some years ago (BININDA-EMONDS 2004a), the application of a supertree framework will gradually shift from its traditional application of combining source trees obtained from the literature to become more integrated with the supermatrix framework (see Fig. 1). The timing of this shift depends largely on the

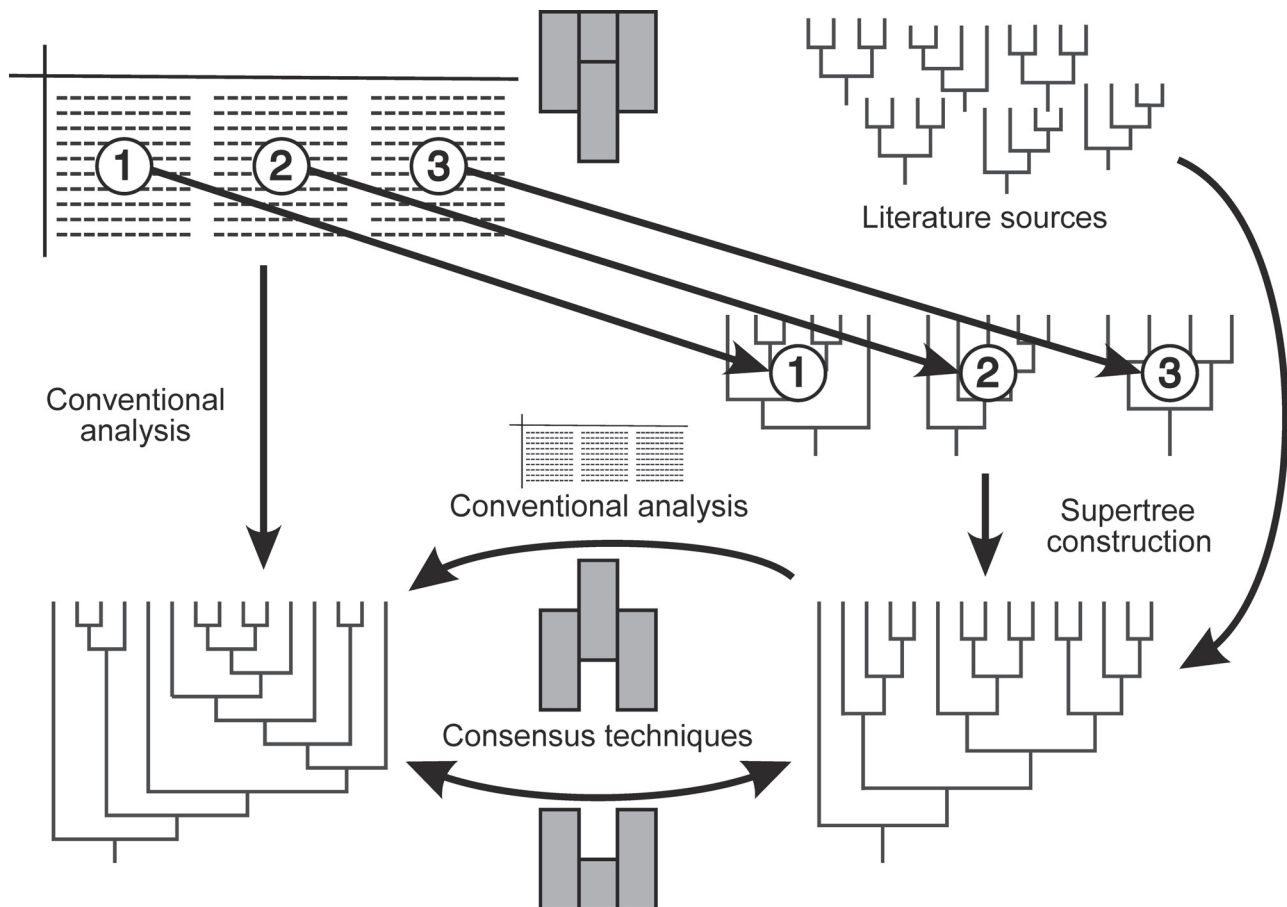


Fig. 1. Breaking down the wall between supermatrix (left) and supertree (right) analyses. Instead of being based on distinct data sets (top left and top right respectively), both frameworks will analyze the same (molecular) data set (top left) in the future. A supertree analysis of partitions in this data set will then be compared directly to the supermatrix solution in a global congruence framework (bottom middle) and/or used to seed a supermatrix analysis as part of a divide-and-conquer strategy (middle).

quantity of available molecular sequence data and so will occur sooner for the better studied groups. For the lesser studied groups, the traditional, literature-based application of supertree construction might represent the only viable option, if at all, to obtain reasonably complete phylogenetic estimates. For the remainder of this paper, I now briefly outline two ways that I see in which supertrees and supermatrices combine to help us in our goal of deriving the Tree of Life. In the first, the results of the two methods are juxtaposed in a sensitivity analysis, whereas, in the second, supertree analysis forms part of a search strategy that could potentially increase the speed and accuracy of supermatrix analyses.

Acknowledgements

I thank MARTIN BLUM, HANS-DIETER GÖRTZ, and JOHANNES STEIDLE for organizing the wonderfully diverse and interesting WILLI-HENNIG-Symposium on Phylogenetics and Evolution and also for inviting me to contribute to the diversity. I also thank JULIA GOCKEL for providing the German translation of the abstract.

2. Supertrees, supermatrices, and global congruence

Supertrees and supermatrices analyze the data at fundamentally different levels (BININDA-EMONDS 2004b). For supermatrix analyses, the fundamental unit of information is the individual character, be it a morphological character or a single nucleotide base or amino acid. By contrast, supertree analyses take entire phylogenetic trees as their fundamental units. Thus, in analyzing the same base molecular data set, the two frameworks place different methodological demands and assumptions on the analyses because of these different emphases. For a multi-gene data set, a supermatrix analysis looks for congruence among individual nucleotides or amino acids. Given the commonly held view that genes function at least in part as evolutionary units, the results of a supermatrix approach will tend to favour the larger genes because of the greater number of characters they provide. By contrast, the supertree approach will tend to weight the different gene regions (gene trees) more equally and the solution will reflect those clades that occur the most often (and/or are the most strongly supported) among the studied gene regions, regardless of how large each region is. As tempting as it is to argue which approach is better (or if one approach or the other is patently wrong), the reality is that the approaches are best viewed as simply being different.

The adoption of this latter perspective forms the basis for the global congruence approach (LAPOINTE et al. 1999), a variant on sensitivity analyses. For conventional character based analyses, for example, it is well accepted practice to analyze the data matrix using a variety of optimization techniques (e. g., maximum parsimony, neighbour joining

et al., maximum likelihood, and/or Bayesian inference). The different methods all have different biases, strengths and weaknesses, and can account for models of molecular evolution to different extents. Thus, where these methods provide similar solutions despite their methodological differences, we can have increased confidence in those solutions. Where the solutions conflict, more investigation is needed to determine the source of that conflict (e. g., inferiority of one method or another, lack of information in the data set, conflict in the data set, or an inherently difficult region of the tree because of a rapid adaptive radiation).

The global congruence approach applies exactly this same logic. Instead of using multiple optimization criteria (although this can still be done), the approach directly contrasts the results of supermatrix and supertree analyses of the same data set. Again, increased confidence can be placed in those parts of the tree that the two different methods indicate in common, with conflicting regions highlighting the need for further investigation. The end result is a more thorough exploration of both data and the phylogenetic hypotheses than would result from the application of only a single method.

Few global congruence analyses have been performed to date (e. g., GATESY et al. 2004; FULTON & STROBECK 2006; HIGDON et al. 2007) and then often to compare the methods rather than the results (e. g., GATESY et al. 2004). However, even these few examples have been instructive. For instance, our analysis of paired matrix representation with parsimony (MRP; BAUM 1992; RAGAN 1992) supertree and maximum likelihood supermatrix analyses of a multi-gene data set for pinnipeds (25 750 bp for 50 genes; HIGDON et al. 2007) yielded largely congruent answers, differing from one another by a total of eight nodes out of a total of 32. This result undermines assertions that the inherent loss of information in supertree analyses often leads to an undue loss of accuracy or resolution (DE QUEIROZ & GATESY 2007) and supports findings of roughly equal accuracy between the supertree and supermatrix approaches (BININDA-EMONDS & SANDERSON 2001).

Moreover, global congruence reveals that most of the conflict (six of the eight nodes) derives from alternative placements of only two species, the Gray Seal (*Halichoerus grypus*) and the Australian Sealion (*Neophoca cinerea*) (see Fig. 2). Both species, and the Gray Seal in particular, have been difficult to place historically (see HIGDON et al. 2007) and their conflicting placements in the respective trees simply reflect this situation and support calls for additional data. Finally, in comparing the two solutions, it is clear that those clades that are common between the two approaches are much better supported in each tree than are those clades that are in conflict. For instance, the 24 nodes that are congruent between the two trees display an average bootstrap support (FELSENSTEIN 1985) of 92.9% (± 3.3 % SE) in the supermatrix tree compared to a value of

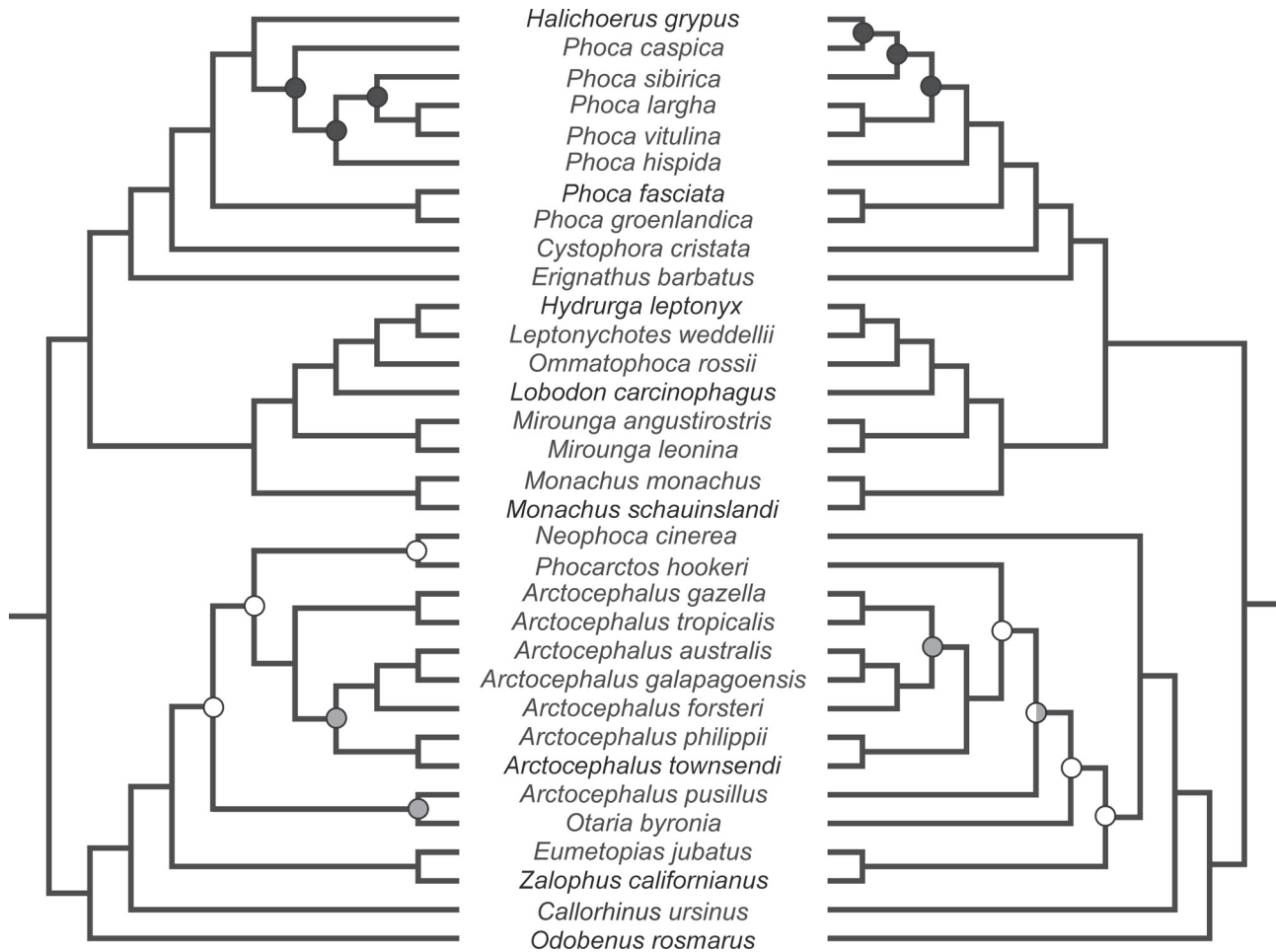


Fig. 2. Comparison of the maximum likelihood supermatrix tree (left) and MRP supertree (right) from the multigene analysis of HIGDON et al. (2007). Of the eight conflicting nodes on either tree (circles), three derive from conflicting placements of the Gray Seal, *Halichoerus grypus* (black) and three or four from the Australian Sealion, *Neophoca cinerea* (white); the remainder (grey) derive from other conflicts.

38.8% ($\pm 6.9\%$) for the eight conflicting nodes. This difference is highly significant (one-tailed Student's $t = 7.837$, $df = 30$, $P < 0.0001$). A similar result is found for the supertree: congruent nodes are significantly better supported (as measured by the supertree specific measure rQS; BININDA-EMONDS 2003; PRICE et al. 2005) than are conflicting ones (0.658 ± 0.065 versus 0.223 ± 0.149 , respectively; one-tailed Student's $t = 3.099$, $df = 30$, $P = 0.0021$). Thus, global congruence provides more information than either approach in isolation, indicating broader support for the strongly supported nodes in either phylogeny and that weakly supported nodes are attributable in large part to only two "rogue" species.

3. Dividing and conquering: supertrees as a heuristic search tool

The global congruence perspective, while bringing the supertree and supermatrix approaches closer together, still recognizes and indeed explicitly builds upon the differences in the two frameworks. Pending solving some outstanding problems, a promising future development could be to subsume supertree construction within a supermatrix analysis as a strategy to increase the speed of such analyses and so make ever-larger phylogenetic problems increasingly tractable. Indeed, even critics of the supertree approach reluctantly admit a fullscale assault on the Tree of Life probably will require supertrees at some stage of the problem (DE QUEIROZ & GATESY 2007).

Phylogenetic analysis is an inherently difficult computational problem because the number of possible solutions (trees) increases super-exponentially with the number of taxa in the analysis (FELSENSTEIN 1978). Indeed, most of the commonly used optimization criteria like maximum parsimony or maximum likelihood have been demonstrated to be NP-hard, meaning that there is no efficient solution known for them at present and that it is likely that none exists (for a more detailed explanation, see GAREY & JOHNSON 1979). Our current ability, therefore, to “solve” phylogenetic problems with tens of thousands of taxa lies with the development of increasing computer performance, parallel computation, clever heuristic search strategies, and efficient memory and data structuring, together with the realization that we have no guarantee that our optimal solution(s) truly are optimal. Ever-larger phylogenetic problems will demand increasingly clever computational strategies, one of which might include the divide-and-conquer approach.

This latter approach is a general computational strategy that seeks to solve a very large and difficult problem by dividing it into many smaller problems and then combining the solutions to these to derive the global answer. When applied to phylogenetic inference, a divide-and-conquer strategy amounts to breaking a larger data matrix in many smaller ones, with the trees derived from their analysis being combined (as a supertree) to obtain the final answer. The potential advantages of such an approach derive from two bases. First, smaller phylogenetic problems typically run proportionately faster than larger ones, representing the flip side to the NP-hardness of phylogenetic analysis. This fact is generally well accepted, having been documented in numerous simulation studies. What is perhaps less well appreciated is how great the savings could prove to be. For instance, BININDA-EMONDS & STAMATAKIS (2006) showed that in the time that was needed to analyze a single 4096-taxon data set comprised of 1000 bp sequences using maximum parsimony, it would be possible to analyze approximately 250 000 16-taxon data sets derived from it through pruning. This represents about 4 000 000 taxa in total or about 1000 times as much “taxon coverage” as the single global analysis. Second, it has been argued that the solutions to these smaller problems should be comparatively more accurate (ROSHAN et al. 2004b) because the faster analysis times enables the use of more robust search strategies and also because the problems themselves have a smaller phylogenetic diameter (i. e., the taxa tend to be more closely related, representing a less difficult problem). Evidence here, however, tends to be lacking. At the very least, there appears to be little appreciable change in accuracy under maximum likelihood with problem sizes ranging from four to 4096 taxa (BININDA-EMONDS & STAMATAKIS 2006) such that there is no advantage to analyzing the smaller problems in this regard.

The use of divide-and-conquer in phylogenetics is not new. In fact, the strategy underlies the method of quartet puzzling (STRIMMER & VON HAESLER 1996), which divides the global problem into its smallest possible subset of four taxa and combines the solutions to all possible quartets (weighted by their relative supports) to derive the global solution. Another variant of this method is provided by Quartets MaxCut (SNIR & RAO 2008). The potential fusion of the supertree and supermatrix methods is arguably best exemplified by the phylogenetic divide-and-conquer method Recursive-Iterative-DCM3 (Rec-I-DCM3; ROSHAN et al. 2004a). Given a large data matrix, this method uses disc-covering methods (HUSON et al. 1999a, 1999b; ROSHAN et al. 2004a) to identify the optimal subproblems that are then analyzed using conventional methods. The resulting solutions are then combined using the conservative supertree method Strict Consensus Merger (SCM; HUSON et al. 1999a), with the supertree being refined on the basis of the entire data matrix in a global supermatrix analysis. Thus, the supertree no longer represents an end result, but rather functions as a starting tree for the supermatrix analyses. In this way, supertree construction mirrors the use of neighbour joining in PHYML (GUINDON & GASCUEL 2003) or greedy Maximum Parsimony in RAxML (STAMATAKIS 2006) to deliver rapid, reasonable starting trees for more intensive maximum likelihood-based analyses.

However, the promise behind the divide-and-conquer approach has yet to be fully realized in practice. For example, despite combining apparently good performance with an existing maximum likelihood-based implementation in TREE-PUZZLE (SCHMIDT et al. 2002), quartet puzzling has not found a solid foothold in the phylogenetic community. The same holds true for Rec-I-DCM3, which appears to be especially effective in a maximum parsimony framework, showing large improvements with respect to both speed and especially accuracy (ROSHAN et al. 2004b). Results in a maximum likelihood framework are not as impressive, with virtually no improvement in accuracy and speed improvements being less than a factor of two (BININDA-EMONDS & STAMATAKIS 2006). So where has the promise gone?

For a divide-and-conquer approach to be truly effective and competitive in phylogenetic analysis, it is crucial to optimize the final two steps in the process, the construction of the supertree and the analysis of the entire data matrix. Because of the NP-hard nature of phylogenetic analysis, the last supermatrix-like step is especially important, being the most computationally intensive and representing the actual analytical bottleneck. The computational costs of the remaining steps, by contrast, are comparatively small (but see below). Speeding up of this final step can occur either through 1) delivering a more accurate starting tree for the analysis to build on and/or 2) reducing the search space of the analysis. Of these two strategies,

I would argue that the former is the more critical at this point. For instance, REC-I-DCM3 already implements an effective solution for the latter, using the SCM supertree not only as a starting tree for the subsequent analysis, but also as a constraint tree where the nodes that it contains cannot be contradicted but only refined (i. e., resolved). In so doing, the size of the global problem is effectively reduced compared to a conventional supermatrix analysis, thereby resulting in potential time savings.

The advantage to be gained here in terms of both speed and accuracy, however, is obviously highly dependent on the supertree being used to feed the global optimization analyses. A poorly resolved supertree will not contract the search space appreciably and an inaccurate one will reduce the accuracy of the final solution. An equally important consideration is that the supertree method being used is fast given that it too has to confront a problem comprising all taxa in the original data set. The majority of supertree methods, including those that are the most commonly used like MRP, are unsuitable here because they function by analyzing a matrix according to some optimization criterion (e. g., maximum parsimony). As such, they are also NP hard, and their use would string two NP-hard problems one after the other.

Fortunately, another class of less well known, but fast supertree methods does exist, with algorithms that run in so-called polynomial time according to the number of taxa in the analysis. SCM is such a method and seeks to deliver an analog of the strict consensus of a set of overlapping source trees by computing the strict consensus of pairs of trees at a time. As such, it is dependent on the input order of the trees (ROSHAN et al. 2004b) as well as prone to producing a poorly resolved supertree depending on the degree of conflict among the source trees. The latter is not a bad feature in and of itself. SCM does have a low false positive rate, meaning that the nodes that are present tend to be accurate (HUSON et al. 1999a). However, as mentioned above, a supertree that is too poorly resolved (= false negatives), even if it is accurate, will not realize sufficient time savings for the subsequent global analysis. It might be for this reason that SCM-based divide-and-conquer analyses in a maximum likelihood setting do not show the savings in analysis time compared to traditional supermatrix analyses needed to make divide-and-conquer approaches a viable alternative for large-scale phylogenetic analysis. However, research is currently being undertaken to address this limitation (SHEL SWENSON, unpubl. results).

Additional fast supertree methods do exist and also continue to be in development. The BUILD-based MinCutSupertree (SEMPLE & STEEL 2000) and Modified MinCutSupertree (PAGE 2002) are not to be counted among them, however. Although both are fast and fulfil a number of desirable properties (SEMPLE & STEEL 2000; STEEL et al. 2000), they deliver Adams consensus like topologies (see

ADAMS 1972, 1986) that are not guaranteed to be interpretable phylogenetically. Instead, PhysIC (RANWEZ et al. 2007) appears to be a promising candidate. PhysIC is a so-called veto supertree method that delivers a supertree that does not contradict the information in any source trees (either singly or jointly) as well as contains only those relationships that can be inferred from the source trees (again, singly or jointly). As such, it is akin to SCM, but without the input-order dependence. It too, however, can yield poorly resolved supertrees. This latter issue is addressed in the extension of PhysIC, PhysIC_IST (SCORNAVACCA et al. 2008), which engages a round of preprocessing to identify and discard those parts of individual source trees that conflict the most with the remaining information and so might be leading to a loss of resolution in the final supertree.

Thus, although the promise of a divide-and-conquer approach to large-scale phylogeny construction has not been realized to date, the theoretical basis of this approach combined with continued active development in the field represents cause for hope. In developing a viable divide-and-conquer framework, particular focus needs to be placed on developing fast supertree methods that yield accurate and well resolved solutions despite the final supermatrix step representing the computational bottleneck. This latter problem will always be present, but can be ameliorated by reducing the search space as much as possible through supertrees.

4. Conclusions

In the intervening six years since I last reviewed what the future holds for supertrees (BININDA-EMONDS 2004a), the perspectives have not changed appreciably. Improved sequencing technologies and increased efforts will eventually make the traditional application of supertrees to combine trees out of the literature obsolete. Such has not been the case to date, however, for all but a handful of very small taxa. It seems unlikely that the direct comparison of supertree and supermatrix approaches in a global congruence framework will tend to find much favour in the phylogenetic community given the contentious history between the two techniques. Even so, this approach is also ultimately limited computationally. As the problem sizes continue to grow, separate supertree and supermatrix analyses will become increasingly intractable computationally, even with continued advances in computer technology and algorithms. Instead, it is the merger of the techniques in a divide-and-conquer framework that appears to be the way forward, potentially expanding the problem sizes that would be tractable beyond what can be achieved through advances in hardware alone. As promising as this framework would appear to be, its true potential will only

be realized with the development of a supertree method (possibly in combination with another method) that is fast, accurate, and delivers well resolved solutions.

5. References

- ADAMS, E. M., III (1972): Consensus techniques and the comparison of taxonomic trees. – *Systematic Zoology*, **21**: 390–397.
- ADAMS, E. M., III (1986): N-trees as nestings: complexity, similarity, and consensus. – *Journal of Classification*, **3**: 299–317.
- BAUM, B. R. (1992): Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. – *Taxon*, **41**: 3–10.
- BININDA-EMONDS, O. R. P. (2003): Novel versus unsupported clades: assessing the qualitative support for clades in MRP supertrees. – *Systematic Biology*, **52**: 839–848.
- BININDA-EMONDS, O. R. P. (2004a): The evolution of supertrees. – *Trends in Ecology and Evolution*, **19**: 315–322.
- BININDA-EMONDS, O. R. P. (2004b): Trees versus characters and the supertree/supermatrix “paradox”. – *Systematic Biology*, **53**: 356–359.
- BININDA-EMONDS, O. R. P. & SANDERSON, M. J. (2001): Assessment of the accuracy of matrix representation with parsimony supertree construction. – *Systematic Biology*, **50**: 565–579.
- BININDA-EMONDS, O. R. P. & STAMATAKIS, A. (2006): Taxon sampling versus computational complexity and their impact on obtaining the Tree of Life. – In: HOTKINSON, T. R. & PARNELL, J. A. N. (eds.): *Reconstructing the Tree of Life: taxonomy and systematics of species rich taxa*: 77–95; Boca Raton (CRC Press).
- BININDA-EMONDS, O. R. P., JONES, K. E., PRICE, S. A., GRENYER, R., CARDILLO, M., HABIB, M., PURVIS, A. & GITTLEMAN, J. L. (2003): Supertrees are a necessary not-so-evil: a comment on GATESY et al. – *Systematic Biology*, **52**: 724–729.
- DE QUEIROZ, A. & GATESY, J. (2007): The supermatrix approach to systematics. – *Trends in Ecology and Evolution*, **22**: 34–41.
- FELSENSTEIN, J. (1978): The number of evolutionary trees. – *Systematic Zoology*, **27**: 27–33.
- FELSENSTEIN, J. (1985): Confidence limits on phylogenies: an approach using the bootstrap. – *Evolution*, **39**: 783–791.
- FULTON, T. L. & STROBECK, C. (2006): Molecular phylogeny of the Arctoidea (Carnivora): effect of missing data on supertree and supermatrix analyses of multiple gene data sets. – *Molecular Phylogenetics and Evolution*, **41**: 165–181.
- GAREY, M. R. & JOHNSON, D. S. (1979): *Computers and intractability: a guide to the theory of NP-completeness*. XII + 338 pp.; San Francisco (W. H. Freeman).
- GATESY, J. & SPRINGER, M. S. (2004): A critique of matrix representation with parsimony supertrees. – In: BININDA-EMONDS, O. R. P. (ed.): *Phylogenetic supertrees: combining information to reveal the Tree of Life*: 369–388; Dordrecht, the Netherlands (Kluwer Academic).
- GATESY, J., MATTHEE, C., DESALLE, R. & HAYASHI, C. (2002): Resolution of a supertree/supermatrix paradox. – *Systematic Biology*, **51**: 652–664.
- GATESY, J., BAKER, R. H. & HAYASHI, C. (2004): Inconsistencies in arguments for the supertree approach: supermatrices versus supertrees of Crocodylia. – *Systematic Biology*, **53**: 342–355.
- GUINDON, S. & GASCUEL, O. (2003): A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. – *Systematic Biology*, **52**: 696–704.
- HIGDON, J. W., BININDA-EMONDS, O. R. P., BECK, R. M. & FERGUSON, S. H. (2007): Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. – *BMC Evolutionary Biology*, **7**: 216.
- HUSON, D. H., NETTLES, S. M. & WARNOW, T. J. (1999a): Disk-covering, a fast-converging method for phylogenetic tree reconstruction. – *Journal of Computational Biology*, **6**: 369–386.
- HUSON, D. H., VAWTER, L. & WARNOW, T. J. (1999b): Solving large scale phylogenetic problems using DCM2. – In: LENGAUER, T., SCHNEIDER, R., BORK, P., BRUTLAG, D., GLASGOW, J., MEWES, H.-W. & ZIMMER, R. (eds.): *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*: 118–129; Menlo Park, California (AAAI Press).
- LAPOINTE, F.-J., KIRSCH, J. A. W. & HUTCHEON, J. M. (1999): Total evidence, consensus, and bat phylogeny: a distance based approach. – *Molecular Phylogenetics and Evolution*, **11**: 55–66.
- NOVACEK, M. J. (2001): Mammalian phylogeny: genes and supertrees. – *Current Biology*, **11**: R573–R575.
- PAGE, R. D. M. (2002): Modified mincut supertrees. – In: GUIGÓ, R. & GUSFIELD, D. (eds.): *Algorithms in Bioinformatics*, Second International Workshop, WABI, 2002, Rome, Italy, September 17–21, 2002, *Proceedings*: 537–552; Berlin (Springer).
- PRICE, S. A., BININDA-EMONDS, O. R. P. & GITTLEMAN, J. L. (2005): A complete phylogeny of the whales, dolphins and even-toed hoofed mammals (Cetartiodactyla). – *Biological Reviews*, **80**: 445–473.
- PURVIS, A. (1995): A composite estimate of primate phylogeny. – *Philosophical Transactions of the Royal Society London, Series B*, **348**: 405–421.
- RAGAN, M. A. (1992): Phylogenetic inference based on matrix representation of trees. – *Molecular Phylogenetics and Evolution*, **1**: 53–58.
- RANWEZ, V., BERRY, V., CRISCUOLO, A., FABRE, P. H., GUILLEMOT, S., SCORNAVACCA, C. & DOUZERY, E. J. (2007): PhySIC: a veto supertree method with desirable properties. – *Systematic Biology*, **56**: 798–817.
- ROSHAN, U., MORET, B. M. E., WARNOW, T. & WILLIAMS, T. L. (2004a): Rec-I-DCM3: a fast algorithmic technique for reconstructing large phylogenetic trees. Paper presented at: *Proceedings of the IEEE Computational Systems Bioinformatics conference (CSB)*; Stanford, CA (IEEE Computer Society Press).
- ROSHAN, U., MORET, B. M. E., WILLIAMS, T. L. & WARNOW, T. (2004b): Performance of supertree methods on various data set decompositions. – In: BININDA-EMONDS, O. R. P. (ed.): *Phylogenetic supertrees: combining information to reveal the Tree of Life*: 301–328; Dordrecht, the Netherlands (Kluwer Academic).
- SCHMIDT, H. A., STRIMMER, K., VINGRON, M. & VON HAESELER, A. (2002): TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. – *Bioinformatics*, **18**: 502–504.
- SCORNAVACCA, C., BERRY, V., LEFORT, V., DOUZERY, E. J. & RANWEZ, V. (2008): PhySIC_IST: cleaning source trees to infer more informative supertrees. – *BMC Bioinformatics*, **9**: 413.
- SEMPLE, C. & STEEL, M. (2000): A supertree method for rooted trees. – *Discrete Applied Mathematics*, **105**: 147–158.
- SNIR, S. & RAO, S. (2008): Quartets MaxCut: a divide and conquer quartets algorithm. Paper presented at: *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 10 Dec 2008 (IEEE Computer Society).
- SPRINGER, M. S. & DE JONG, W. W. (2001): Phylogenetics. Which mammalian supertree to bark up? – *Science*, **291**: 1709–1711.

- STAMATAKIS, A. (2006): RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. – *Bioinformatics*, **22**: 2688–2690.
- STEEL, M., DRESS, A. W. M. & BÖCKER, S. (2000): Simple but fundamental limitations on supertree and consensus tree methods. – *Systematic Biology*, **49**: 363–368.
- STRIMMER, K. & VON HAESELER, A. (1996): Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. – *Molecular Biology and Evolution*, **13**: 964–969.

Address of the author:

OLAF R. P. BININDA-EMONDS, AG Systematik und Evolutionsbiologie, IBU – Fakultät V, Carl von Ossietzky Universität Oldenburg, 26111 Oldenburg, Germany
E-mail: olaf.bininda@uni-oldenburg.de

Manuscript received: 15 April 2010, accepted 15 June 2010.