What might Evolutionary Algorithms (EA) and Multi-objective Optimisation (MOO) Contribute to Phylogenetics and the Total Evidence Debate

Leon Poladian[1] and Lars S. Jermiin[2]

[1] School of Mathematics and Statistics
University of Sydney NSW 2006, Australia
L.Poladian@maths.usyd.edu.au
[2] School of Biological Sciences and the
Sydney University Biological Informatics and Technology Centre
University of Sydney NSW 2006, Australia
lsj@bio.usyd.edu.au

Abstract. Evolutionary relationships among species are usually (*i*) illustrated by means of a phylogeny and (*ii*) inferred by optimising some measure of fitness, such as the total evolutionary distance between species (given the tree), the parsimony (number of different assumptions required to fit the data to the tree), the likelihood of the tree (given an evolutionary model and a data set) or the posterior probability of the tree (given an evolutionary model, a data set, and the distribution of prior probabilities). A large variety of different types of data can and have been used in reconstructing evolutionary relationships including nucleotide sequences, anatomical features, metabolic processes, behaviour or even the words of languages. Difficulties arise when different sources of evidence provide conflicting information about the inferred 'best' tree(s). The *Total Evidence Debate* focusses on how to assess, combine, modify or reject different types of data. We begin with a review of evolutionary algorithms (EA) used for phylogenetic inference. Then we discuss whether the population-based searches that are an intrinsic attribute of EA and multi-objective optimisation (MOO) can provide a powerful new approach to this area.

1 Introduction

> *"The affinities of all beings of the same class have sometimes been represented by a great tree."* [1]

The first diagram in Darwin's notebook *Transmutation of Species* and the only diagram to occur in Darwin's *Origin of the Species* is a phylogenetic tree [2]. Phylogenetic inference is the construction of trees to represent the genealogical relationships between different species. Other entities, such taxonomic groups higher than species, languages, documents or texts, can also be studied. Hereafter, we only use species as a generic term to cover all of these. The difference

between phylogenetic classification (or cladistics) and other forms of taxonomy or systematics is that the classification should only be based on common ancestry and not on mere similarity of appearance or function. Thus, a phylogenetic tree is an *hypothesis* of the evolutionary history and relationships among species. Its construction depends on the quality and quantity of the evidence available (*i.e.*, the data) and the reliability of the algorithms used to analyse these data.

Phylogenetic inference begins with obtaining a data set comprising of characters for each species. These characters might be nucleotide and amino acid sequences, protein shapes, anatomical characters, embryo development, biosynthetic pathways, behavioural traits, and linguistic data. The judicious choice of appropriate characters is a critically important skill and disagreements have often arisen over the appropriate choice and relative weight of different characters. Characters that are likely to arise often and independently will not provide useful information about common ancestry. Characters that have experienced *too much* selection pressure may exhibit convergent evolution or no evolution at all. Characters that have undergone *too rapid* evolution will provide a low signal-to-noise ratio for the reconstruction algorithms. These problems can arise even if one arbitrarily restricts oneself to purely genetic sequences, since different parts of the genome undergo different rates of mutation and respond to selection pressures in different ways. Recombination and horizontal gene transfer can also render genetic data less useful for phylogenetic inference.

The availability of easy-to-use powerful algorithms has created a situation where evolutionary trees are often produced and published without due consideration of the underlying assumptions [3, 4]. The danger of *ad hoc* approaches is the human tendency to consciously or sub-consciously eliminate or downplay evidence that disagrees with expectations [2].

The choice of characters and weighting of evidence between different sources is precisely analogous to dealing with multiple criteria in MOO. Each data set, each model, each method can be regarded as an independent optimisation criterion. By using MOO, one is never forced to merge or discard criteria. We hope in this paper to open up discussion on the value of this approach to phylogenetics.

We first give an overview of phylogenetic algorithms, followed by a review of existing applications of EA to phylogenetics. We then discuss the total evidence debate and recent ideas about not focussing purely on just the optimal solution. We finally present some ideas about the relevance of MOO to these issues.

2 A Cursory Look at Phylogeny Algorithms

Salemi and Vandamme's [5] excellent book on Phylogenetic Methods gives theoretical details and case studies on most of the algorithms. Open source code for almost all the algorithms is freely available. There are three main types of methods: Maximum Parsimony, Maximum Likelihood (including Bayesian) and Distance-Based methods; and there are various implementations of all these methods. In this paper, we will only discuss those that can be regarded as optimising some measure of fitness of the trees.

The search for the best tree often includes a local search consisting of modifications to the best tree found so far. These local search operations are discussed together in a later subsection. Exhaustive search (or the Branch and Bound algorithm [6]) is the only algorithm guaranteed to find the correct tree topology. This is computationally feasible for up to 12 to 25 species depending on whether one uses maximum likelihood or maximum parsimony. Thus for larger sets, heuristic search algorithms, such as EA, are required.

2.1 Maximum Parsimony (MP) Methods

MP is based on the principle of Okham's Razor: in choosing between two hypotheses the one with fewer assumptions should be preferred [7–9]. Each species is assigned a set of characters and placed into an hypothesised evolutionary tree. For any given tree, there are fast algorithms based on dynamic programming [10] to infer the character traits of the common ancestors while minimising the total number of character changes (tree-length) occurring in the tree. Thus for any tree, the minimum length of that tree (a fitness measure) can be rapidly established. Because of its simple fitness function, MP is one of the fastest algorithms and a common fall-back when computational limits apply to other algorithms. Situations in which MP gives misleading results have been identified [11, 12].

2.2 Maximum Likelihood (ML) and Bayesian Methods

ML begins with a model of evolution that may contain various free parameters. These may include the rate and probabilities of various types of mutations, and other parameters related to the independence or correlation of the evolving characters. Alternatively, these parameters can be chosen based on empirical values. Each competing hypothesis now consists of three parts: the topology of the tree, the evolutionary distance or time along the edges of the tree, and the model parameters. Bayesian methods, in addition, include the distribution of the prior probabilities of all these parts. The likelihood of each hypothesis is calculated with respect to the model used, and the most likely tree is inferred by comparing the likelihood values. The most likely model parameters also emerge from this search. Although more complex than MP, for a given topology, the optimisation with respect to model parameters and edge lengths is not considered to be the most challenging part and can be regarded as part of the fitness calculation. The main criticisms of ML are that (i) the results are necessarily dependent on the choice of evolutionary model and (ii) that it is a far more computationally demanding and time-consuming process.

2.3 Distance-Based Methods

These methods have two separate stages. The first stage determines an evolutionary distance between all pairs of species. This distance can be as simple as the fraction of nucleotide sites that differ between two genes, or more elaborate

by incorporating assumptions about different probabilities of different types of mutations. In the second stage, the matrix of pairwise distances is used to construct a tree topology. Two methods are used: cluster analysis and minimum evolution. We will not discuss cluster analysis here because it is not amenable to interpretation as a fitness optimising approach; however, it could be useful in generating candidate trees. for subsequent analysis with local hill climbing algorithms. Minimum evolution uses the sum the edge-lengths in the tree as a fitness measure and seeks the tree that minimises this while still consistent with all the pair-wise distances between species [13].

2.4 Local Search Operations

Most searches through tree-space are in the form of local hill climbing algorithms. A local permutation is performed on the current best tree in the hopes of constructing a better tree. These permutations are also used later in EA as mutation operators and as the inspiration for some recombination operators. The three basic branch-swapping local-search operations are discussed here. Experience suggests these methods are effective for up to 100 species [14–16].

Nearest-Neighbour Interchange (NNI) Each internal branch of a binary tree is visited. The three topologically distinct trees that can be obtained by swapping a sub-tree connected to one end of the branch with a sub-tree connected to the other end of the branch are considered.

Subtree pruning and re-grafting (SPR) An arbitrary sub-tree is detached and re-attached at an arbitrary location. All possible sub-trees and insertion points may be considered. NNI is a special case of SPR.

Tree bisection and Reconnection (TBR) The tree is cut into two subtrees by cutting an arbitrary branch. The trees are reattached by selecting an arbitrary re-attachment point on *each* tree. All variants may be considered. SPR is a special case of TBR where the detachment/re-attachment point of one subtree is fixed.

Each method is more computationally expensive than the previous one, but provides a more complete exploration of the neighbouring tree-space. Extremely efficient algorithms now exist for all these methods based on problem-specific knowledge of the tree structures using 'incremental down-pass optimisation' [17].

2.5 Methods to Avoid being Trapped in Local Optima

The most commonly used strategy to avoid entrapment in a local optima is to restart any of the above search algorithm from a different starting point. With a sufficient number of distinct starting trees, the global optimum may eventually be uncovered. Though simple to implement, multiple restart is computationally burdensome. For heuristic algorithms, where the results depend on the order in which the species are analysed, restarts begin by randomising the order of the species. (There is an example of a genetic algorithm being used to search for the best starting order [18].)

The parsimony ratchet [19] is a conceptually different technique to escape a local optimum. It can be regarded as having three steps:

1. Find a local optimum by some method.
2. Change the fitness function by modifying the weights assigned to different characters and use local search to move to the local optimum of this new fitness function.
3. Restore the fitness function to its original form and re-apply local optimisation.

At the end of this process we are either back at the same local optimum (which may, in fact, be the global optimum) or we have jumped to a neighbouring local optimum. By repeating this process many times, one 'ratchets' from one local optimum to another, hopefully, on the way to the global optimum. Changing the weights in the fitness function provides a conceptual link to Pareto-optimisation techniques.

Goloboff [15] introduced the ideas from simulated annealing in his Tree-Drifting algorithm, where sub-optimal solutions are accepted during branch-swapping with some small probability.This provides a chance to escape from a local optimum, without seriously impairing the local hill-climbing effort.

3 Evolutionary Algorithms in Phylogenetic Inference

A brief overview of the application of EA to phylogenetic inference is presented. As mentioned above, the many mutation operators are based on the local search algorithms discussed earlier. The recombination operators are varied and innovative; they are listed and described at the end of this section.

The first application of EA to phylogenetic inference appears to be by Matsuda in 1996 [20]. The method used was maximum likelihood with fixed model parameters. The optimisation of the edge lengths was absorbed into the fitness calculation for each tree topology. The algorithm explored tree-space by using random mutation based on NNI and an extremely 'directed' recombination operator listed below.

Lewis [21] developed a computationally more efficient version of this algorithm by extracting the edge length optimisation from the fitness calculation, and using the edge lengths as additional 'genes' that were mutated. This placed the edge lengths on the same footing as the topology. A single free parameter was also included in the evolutionary model. A random proportion of branch lengths and the model parameter are mutated using a random gamma-distributed multiplicative factor. The 'mutation' operator was an exhaustive local search with SPR.

Moilanen [22] used an evolutionary optimisation combined with local search. The method has a roulette selection with initially low but increasing selection pressure to avoid premature convergence. No mutation is used. Unlike Matsuda's operator [20], this recombination operator is random.

Goloboff [15] has a tree fusing algorithm (see below) that contains a non-random recombination operator that does an exhaustive search.

More recently, Congdon [23] has developed an algorithm 'GAPhyl', building upon the well known phylogenetic program called Phylip[24]. Mutation was performed by random swapping of species at the tips of the tree. The recombination operator is random (see below). The idea of sub-populations and immigration is used to avoid premature convergence.

3.1 Recombination or Crossover Operators

Matsuda's Operator [20] Two random trees are selected to be parents. If $d_{ij}^{(1)}$ and $d_{ij}^{(2)}$ are the distance between species i and j in the two parent trees the crossover operation is performed by first identifying the pair of species for which $|d_{ij}^{(1)} - d_{ij}^{(2)}|/(d_{ij}^{(1)} + d_{ij}^{(2)})$ is largest. The smallest subtree containing these two species is located in the parent with the higher fitness and transferred *in toto* to the other parent in the hope of improving its fitness. All the species that were in the subtree are removed first from the second parent before attaching the transferred sub-tree.

Moilanen's Operator [22] Select a random sub-tree from each parent and exchange them followed by deleting duplicated species from the recipient trees.

Goloboff's Tree-Fusing Operator [15] Instead of just randomly exchanging sub-trees use the extremely efficient 'incremental down-pass optimisation' of Gladstein [17] mentioned earlier in the context of single-tree local searches to find the best sub-tree to exchange and the best point to attach it.

Congdon's Operator [23] From one parent choose an arbitrary sub-tree (A). In the other parent, identify the smallest sub-tree (B) that contains all the species occurring in (A). Replace (A) with (B) in the first parent and delete duplicated species elsewhere in the tree. Repeat process with the original parents swapped. This produces two new candidate trees.

Interesting questions are: what differences might arise between directed recombination and random recombination operators? How dangerous is it too build too much prior knowledge into a recombination operator? The same can be asked about some of the local mutation operators that are extremely Lamarckian.

4 The Total Evidence Debate

The debate about how to use different sorts of data has a long and controversial history. As early as 1950, Hennig [25, 26] proposed that any data used for phylogenetic inference will give very similar results because organisms are the outcome of many genetic, developmental and behavioural systems interacting dynamically throughout history. There is no *a priori* reason to believe one type of data is intrinsically superior to any other. Many studies of the published record over the last few decades support this view (see p. 80 of [2]). There are two modern schools of thought about how to integrate information from different data sets into a unified phylogenetic history: taxonomic congruence and total evidence.

4.1 Taxonomic Congruence

Taxonomic congruence involves a search for a consensus between results obtained by *independently* analysing different data sets. The data should not be combined because they may have evolved under different conditions. The philosophical objections to this school of thought are that the hypothesis that different data sets have evolved differently should not be assumed *a priori* but should be empirically tested by comparison to results of a simultaneous analysis [27]. The empirical objections are by focussing only on the commonality or consensus between different data sets, a lot of potentially useful information is being discarded [28]. In addition, most taxonomic congruence treats the trees emerging from different data sets as equally well supported, despite substantial differences in the size of data sets, and in the results of internal consistency tests on the data sets. In MOO, one obtains a Pareto set first and only then looks for *both* commonality and systematic variations across the set. This information is then correlated with variations in the optimisation criteria. The MOO approach thus helps to answer some of the criticisms above.

4.2 Total Evidence

Kluge [27] has advocated the concept of total evidence: use all available data in one single phylogenetic analysis. In other words, the hypothesis supported by the largest amount of data is preferable to the consensus hypothesis that is common to many smaller sets of data. Brooks and McLennan [2] quote many studies that find that total evidence trees tend to be more robust than analysis based on subsets of the same data. Philosophically, the desire to use all available data is admirable but the problems that arise are: how should the data be combined and how should less reliable data be weighted or compared to more reliable data. This problem becomes particularly relevant to cases where molecular and morphological data are combined. The approach is fraught with danger: once we begin to manipulate the data, we can get almost any result. Farris [29] has shown that it is possible, in principle, to invent character weighting schemes that yield any desired tree. Practitioners of MOO will recognise this as analogous to applying different weights to conflicting criteria to turn the problem into a single objective optimisation, thus making *a priori* judgements about their relative importance.

4.3 The Need for Both Consensus and Conflict

In their early review [30] of this debate, de Queiroz *et al.* presented a conceptual framework based on the *reasons* that different data sets may give conflicting results. The precise nature of the conflict (or areas of consensus) gives the expert practitioner useful knowledge of the appropriate algorithms to use, where to look for more data and which model assumptions need more attention. To use a cliche, conflict should be seen as an opportunity to improve the analysis rather than as

a threat. The methods discussed in [30] attempt to avoid information loss whilst simultaneously coping with heterogeneity in data sets.

We have found no evidence in the literature that practitioners from either school of thought have contemplated the techniques and philosophy of MOO.

5 When the Best is not Good Enough

A common trend in many phylogenetic analyses is the desire to encapsulate the phylogenetic result in a single best tree, or in a consensus tree that is based on a small set of equally best trees. In so doing, a high level of confidence is placed on the data, the phylogenetic algorithms, and the phylogenetic results. Such over-confidence is often not justified because there are many cases of near optimal trees. All of these trees may contain elements of the 'true' tree that is commonly sought, and therefore they should not be ignored. Topological model uncertainty [31, 32] occurs whenever there are trees that do not differ significantly from the best tree. Topological model averaging [31, 32] builds on the idea that a tree-specific weight (based on its fitness) can be assigned to every tree in tree-space, and that a weighted consensus tree can be generated across the optimal and near optimal trees. An interesting, but not unexpected, outcome is that the weighted consensus tree sometimes differs from the 'best' tree; this is consistent with research by Nei et al. [33], who showed that the most likely tree sometimes is more likely than the 'true' tree. The method to account for topological model uncertainty has been used repeatedly since its inception and is currently being developed within the Bayesian framework.

6 Pareto Sets and Fitness Landscapes to the Rescue

Two important problems have been identified so far: how to deal with different data sets, and how to uncover more than just the best tree. MOO provides a compelling response to the first problem, and population-based search methods tell us how near-optimal solutions compare with the optimal ones. MOO analysis using EA achieves both goals in the same framework.

Each combination of phylogenetic method and data set used produces a fitness function that is used to evaluate and compare the competing hypotheses (tree topologies, branch lengths, model parameters etc.). Rather than attempt to merge the fitness functions, all are retained and the Pareto or non-dominated set of optima are produced. The dimensionality, shape and extent of the Pareto set reveals which data sets are in conflict and which are compatible. The position and density of second rank non-dominated sets reveals how close the sub-optimal solutions are to the optimal solutions. Naturally, such an approach is computationally more burdensome than single-objective optimisation problems, and it would only be used if the benefits were worth the effort or the problems with existing analyses remained unresolved.

7 Final Words

Of course, this sleight-of-hand has not removed the serious intellectual challenge of dealing with and interpreting multiple, possibly conflicting, evidence. What it has done is to move this activity from the beginning of the analysis, where the data has yet to be fully exploited, to the end of the analysis. This change of perspective opens many new opportunities and activities in both the fields of Phylogenetics and MOO:

- How do we visualise complex Pareto-sets, where the individual elements are tree topologies?
- How do we identify common properties and systematic trends within the Pareto-set and the neighbouring parts of the fitness landscape?
- If we desire a 'consensus' tree, how is it best extracted?
- How do we determine the relative weights of optimal and near-optimal trees when constructing consensus trees?
- How do we identify minimal subsets of common confounding factors (if they exist) that are responsible for conflicting fitness functions and remove them from the analysis?

No doubt there are many other exciting questions that will also arise out of this recombination of EA and MOO and the field of phylogenetic inference.

The authors acknowledge the support of the Australian Research Council and useful discussion with their colleagues Maryanne Large and Steven Manos.

References

1. Darwin, C.: Origin of the Species. 1st edn. John Murray, London (1859).
2. Brooks, D. R. and McLennan, D. A.: The nature of diversity: An evolutionary voyage of discovery. University of Chicago Press, Chicago (2002).
3. Jermiin, L.S., et al.: The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Systematic Biology (2004) In press.
4. Ho, S.Y.W. and Jermiin, L.S.: Tracing the decay of the historical signal in biological sequence data. Systematic Biology (2004) In press.
5. Salemi, M. and VanDamme, A.-M. (eds.): Handbook of phylogenetic methods. Cambridge University Press, Cambrudge (2003).
6. Hendy, M. D. and Penny, D.: Branch and bound algorithms to determine minimal evolutionary trees. Maths. Biosci. **59** (1982) 277–290.
7. Kluge, A. G. and Farris, J. S.: Quantitative phyletics and the evolution of anurans. Syst. Zool. **18** (1969) 1–32.
8. Farris, J.S.: Methods for computing Wagner trees. Syst. Zool. **19** (1970) 83–92
9. Fitch, W. M.: Toward defining the course of evolution: Minimum change for a specific tree topology. Syst. Zool. **20** (1971) 406–416.
10. Sankoff, D. and Rousseau, P.: Locating the vertices of a Steiner tree in an arbitrary metric space. Math. Progr. **9** (1975) 240–276.
11. Felsenstein, J.: Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. **27** (1978) 401–410.

12. Swofford, D. L., Olsen, G. J., Waddell, P. J. and Hillis, D.M.: Phylogenetic Inference In: Hillis, D. M., Moritz, C. and Mable, B. K. (eds.): Molecular systematics (2nd edn) Sunderland, Massachusetts (1996) 407–514.
13. Kidd, K. K. and Sgaramella-Zonta, L. A.: Phylogenetic analysis: Concepts and Methods. Am. J. Human Gen. **23** (1971) 235–252.
14. Swofford, D. L. and Olsen, G. J.: Phylogeny reconstruction. In: Hillis, D. M. and Moritz, C. (eds.): Molecular systematics (1st edn) Sunderland, Massachusetts (1990) 411–501.
15. Goloboff, P. A.: Analyzing large data sets in reasonable times: Solutions for composite optima. Clad. **15** (1999) 415–428.
16. Moilanen, A.: Simulated evolutionary optimisation and local search: Introduction and application to tree search. Clad. **17** (2001) S12–S25.
17. Gladstein, D.: Efficient incremental character optimisation. Clad. **13** (1997) 21–26.
18. Kim, Y.-H., Lee, S.-K. and Moon, B.-R.: Optimizing the order of taxon aAddition in phylogenetic tree construction using genetic algorithm. In: Genetic and Evolutionary Computation Conference (2003) Chicago, Illinois. 2168-2178
19. Nixon, K. C.: The parsimony ratchet, a new method for rapid parsimony analysis. Clad. **15** (1999) 407–414.
20. Matsuda, H.: Protein phylogenetic inference using maximum likelihood with a genetic algorithm. In: Hunter, L. and Klein, T. E. (eds.) Pacific Symposium on Biocomputing '96. World Scientific, London (1996) 512–523.
21. Lewis, P. O.: A genetic algorithm for maximum likelihood phylogeny inference using nucletide sequence data. Mol. Biol. Evol. **15** (1998) 277–283.
22. Moilanen, A.: Searching for most parsimonious tree with simulated evolutionary optimisation. Clad. **15** (1999) 39–50.
23. Congdon, C. B.: Gaphyl: an evolutionary algorithms approach for the study of natural evolution. In: Genetic and Evolutionary Computation Conference. 2002. San Francisco, California. 1057–1064.
24. Felsenstein, J.: 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle. http://evolution.genetics.washington.edu/phylip.html
25. Hennig, W.: Grundzüge einer Theory der phylogenetischen Systematik. Deutscher Zentralverlag, Berlin. (1950).
26. Hennig, W.: Phylogenetic systematics. University of Illinois Press, Urbana. (1966).
27. Kluge, A. G.: Testability and the refutation and corroboration of cladistic hypotheses. Clad. **13** (1997) 81–96.
28. Miyamoto, M. M.: Consensus cladograms and general classifications. Clad. **1** (1985) 186–189.
29. Farris, J. S.: A succesive approximations approach to character weighting. Syst. Zool. **18** (1969) 374–385
30. de Queiroz, A., Donoghue, M.J. and Kim, J.: Separate versus combined analysis of phylogenetic evidence: Ann. Rev. Eco. Syst. **26** (1995) 657–681.
31. Jermiin, L.S., et al.: Majority-rule consensus of phylogenetic trees obtained by maximum-likelihood analysis. Mol. Bio.Evol. **14** (1997) 1296–1302.
32. Wolf, M.J., et al.: TrExML: a maximum likelihood approach for extensive tree-space exploration. Bioinformatics. **16** (2000) 383–394.
33. Nei, M., Kumar, S. and Takahashi, K.: The optimization principle in phylogenetics tends to give incorrect topologies when the number of nucleotides or amino acids are small. Proceedings of the National Academy of Science of the USA **95** (1998) 12390–12397.