Phylogenetic Inference using Evolutionary Multi-objective Optimisation

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 phylogenetic tree and (*ii*) inferred by optimising some measure of fitness, such as the total evolutionary distance between species or the likelihood of the tree (given a model of the evolutionary process and a data set). The combinatorial complexity of inferring the topology of the best tree makes phylogenetic inference an ideal candidate for evolutionary algorithms. However, difficulties arise when different data sets provide conflicting information about the inferred 'best' tree(s). We apply the techniques of multi-objective optimisation to phylogenetic inference. We present results for the simplest model of evolution and an artificially constructed four species problem.

1 Introduction

Phylogenetic inference is the construction of trees that represent the genealogical relationships between different species. It begins with a data set consisting of characters for each species. In this paper we look at sequences of the nucleotides, A, C, G and T, although the method is equally applicable to other character systems (amino acid sequences, protein shapes, anatomical characters). Salemi and Vandamme's [1] recent book on phylogenetic methods gives theoretical details and case studies on many current phylogenetic algorithms. There are various implementations of many these methods but almost all can be regarded as optimising some measure of fitness of the trees. In this paper we use a maximum likelihood approach. Maximum likelihood begins with a model of evolution. Each competing hypothesis consists of three parts: the topology of the tree, the evolutionary distance or time along the edges of the tree, and any model parameters describing the evolutionary process. The likelihood of each hypothesis is calculated with respect to the evolutionary model used (i.e. the model of how nucleotide substitutions occur), and the most likely tree is inferred by comparing the likelihood of different trees.

The search for the best tree topology has been shown to be an NP hard problem [2] with the complexity scaling super-exponentially with the number of species. Evolutionary algorithms (EA) with their population based search strategies, fitness based selection, mutation and recombination operators is known to work well in many combinatorially complex situations.

In Section 2 we discuss the issues surrounding the use of multiple (possibly conflicting data sets in phylogenetic inference). In Section 3 we briefly review existing phylogenetic algorithms that use EA. In Section 4 we define the simplest possible phylogenetic inference problem and give the conventional solutions with conflicting data sets. The four species problem is investigated in this paper for two reasons: it is the simplest problem that exhibits alternate topologies; and the optimisation of four species trees is the first step in the popular quartet-puzzling algorithms [3]. In Section 5 we examine the richness of information obtained by applying a multi-objective optimisation (MOO) technique. Results of applying a simple evolutionary multi-objective optimisation algorithm (EMOOA) to the same problem are then shown in Section 6.

2 Multiple Data Sets

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. Taxonomic congruence involves a search for a consensus between results obtained by analysing different data sets *independently*. On the other hand, the concept of total evidence advocates the use of all available data in a single phylogenetic analysis.

Both schools have been criticised: the data should not be combined because they may have evolved under different conditions; the hypothesis supported by the largest amount of data is preferable to the consensus hypothesis that is common to many smaller sets of data. The problems that arise are: how should the data be combined or compared and how should less reliable data be weighted or compared to more reliable data. The latter strategy is fraught with danger: once we begin to manipulate the data, we can get almost any result. Farris [4] has shown that it is possible, in principle, given enough sets of data, to invent weighting schemes that yield any desired tree.

In their review [5] of the total evidence debate, de Queiroz *et al.* presented a conceptual framework that emphasises not the conflict itself, but 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. The methods discussed in [5] attempt to avoid information loss whilst simultaneously coping with heterogeneity in data sets.

A common trend in many phylogenetic analyses is the desire to encapsulate the phylogenetic result in a single optimal tree, or in the consensus of a small set of equally optimal trees. Consensus, is a form of summarising that replaces many trees by one tree. It is assumed that the information discarded is less

important than the information retained. In so doing, an extreme confidence is being placed on the data, the phylogenetic algorithms, and the phylogenetic results. This extreme level of confidence is often not justified because there are many analyses that yield many near-optimal trees [6, 7].

In EA the concept of a fitness landscape is well known. In addition to identifying the global peak in the landscape, it is also useful to know the location of nearby peaks and the 'width' of each peak. Thus, though it may be useful and valid to summarise an entire 'mountain' by the location of its peak and by a measure of variance around that peak; it is probably not valid to summarise two distinct peaks by their average. Thus, we make the rather obvious assertion that knowledge of the shape of the fitness landscape should precede any attempt to summarise (or apply other processes that discards information).

In MOO, one first obtains a Pareto set, and then looks for *both* commonality and systematic variations across the set. The patterns found are then correlated with variations in the optimisation criteria: this approach helps to answer the criticism made above against both taxonomic congruence and total evidence. Thus, for large sets of species and multiple conflicting data sets, heuristic search algorithms such as evolutionary algorithms combined with multi-objective optimisation techniques are ideal.

3 Current Evolutionary Algorithms

The first application of EA to phylogenetic inference appears to be by Matsuda in 1996 [8]. 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 mutations and recombinations based on swapping sub-trees.

Lewis [9] 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.

Moilanen [10] 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.

More recently, Congdon [11] has developed an algorithm 'GAPhyl', building upon the well known phylogenetic program called Phylip[12]. Mutation was performed by random swapping of species at the tips of the tree. The recombination operator was random swapping of sub-trees. The idea of sub-populations and immigration is used to avoid premature convergence.

4 The Simplest Phylogenetic Problems

The simplest model of evolution used in phylogenetic analysis is the Jukes and Cantor model of nucleotide substitutions [13]. This model assumes that mutations in nucleotide sequences occur randomly, that all possible mutations are

equally likely and that mutations at different sites in the nucleotide sequence occur independently and at the same rate. Thus there is only one free parameter in the entire model: the rate of mutation. However, if we measure evolutionary distances not in terms of time, but in terms of the average number of substitutions, then this parameter fades into the background.

Consider first the nucleotide sequences corresponding to two species. We assume the sequences are homologous (similar due to a common evolutionary history), that they have been aligned and that there are no gaps in either sequence. For example, the sequences may look like

```
....GGGATTTCCGATGACAGTTC....
....GGGAATGCCCATAACAGGCC....
```

Although at each site (column) there are $16 = 4 \times 4$ possible nucleotide pairs that might be observed, under the Jukes and Cantor model, the specific identity of the nucleotides or the location of the mutations is irrelevant. Only the number of similar and different sites is needed in the algorithm. Thus, one observes differences at 25% of sites. However, because two mutations may have occurred at the same site, the total number of mutations is always slightly larger then the observed number of differences. The most probable number of mutations is given by

$$d = -\frac{3}{4}\log(1 - \frac{4}{3}p) \tag{1}$$

where p is the observed fraction of sites with different nucleotides. The trees given in this paper are labelled with the inferred evolutionary distance, d.

For three species, at each site there are $64 = 4^3$ patterns of nucleotides, but under the Jukes and Cantor model, they can be grouped into five different patterns: all three nucleotides identical, all three nucleotides distinct, and three different ways of having exactly two nucleotides the same. The expressions for probabilities and inferred distances are not given here, but are available in many analyses and can be calculated most efficiently by a recursive method [14].

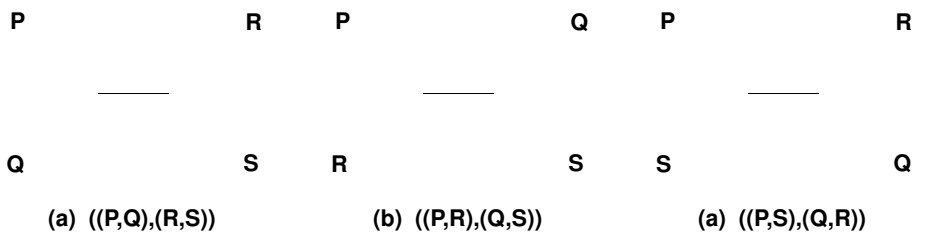

Fig. 1. Three different topologies for the phylogeny connecting four species P, Q, R and S. In Newick format they are (a) $((P,Q),(R,S))$, (b) $((P,R),(Q,S))$, and (c) $((P,S),(Q,R))$.

For four species, the $256 = 4^4$ patterns of nucleotides group into 15 distinct patterns, and an example of each is given in Table I. Thus, for a 4 species problem

with the simplest evolutionary model, the phylogenetic algorithm only requires a set of 15 frequencies or percentages. Four species is the smallest number of species for which the evolutionary tree can have a variety of topologies. If we call the species P, Q, R and S then there are three alternative evolutionary relationships: examples of each occur in Fig. 1. These three different tree topologies can also be indicated in the standard Newick [15] format as (a) $((P,Q),(R,S))$, (b) $((P,R),(Q,S))$, and (c) $((P,S),(Q,R))$.

We assume the pattern frequencies for two data sets have been analysed. An artificially constructed example is given in Table I. The two data sets might correspond to two different genes (a mitochondrial and a nuclear gene perhaps, or even to two separate parts of the same gene).

Table 1. The percentage pattern frequencies for the 15 possible nucleotide patterns occurring in two different data sets. The table is interpreted as follows. In the gene sequence corresponding to Data Set 1, pattern of type #10 occur 7% of the time. That is, in 7% of the nucleotides sites, Species P and R have the same nucleotide, and the other two species have two other distinct nucleotides.

Pattern #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Species P	A	A	A	A	C	A	A	A	A	A	A	C	C	C	A
Species Q	A	A	A	C	A	A	C	C	A	C	C	A	A	G	C
Species R	A	A	C	A	A	C	A	C	C	A	G	A	G	A	G
Species S	A	C	A	A	A	C	C	A	G	G	A	G	A	A	T
Set 1 (%)	11	9	5	9	14	1	5	5	3	7	5	9	7	6	4
Set 2 (%)	9	7	10	8	7	7	4	4	9	5	7	5	6	8	4

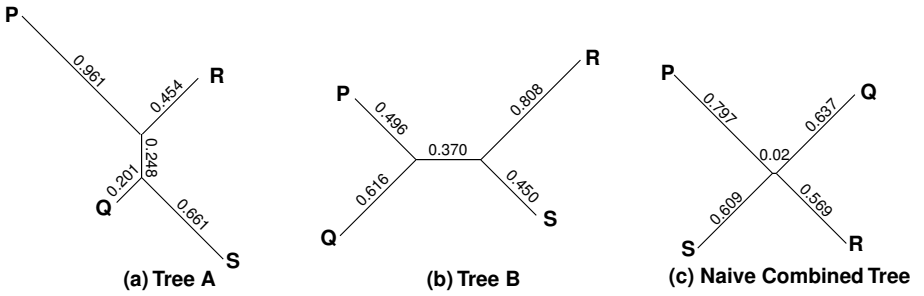

Fig. 2. Maximum likelihood trees obtained from (a) Data Set 1, (b) Data Set 2 and (c) a naive concatenation of the two data sets. The evolutionary distances are calculated with the Jukes and Cantor model.

4.1 Separate vs. Combined Analysis

If we perform a maximum likelihood analysis on Data Set 1 (ignoring Data Set 2) then the best tree is shown in Fig. 2(a). The data favours the evolutionary relationship $((P, R), (Q, S))$.

On the other hand, if we perform a maximum likelihood analysis on Data Set 2 (ignoring Data Set 1) then the best tree is shown in Fig. 2(b). This data favours the relationship $((P, Q), (R, S))$.

The two trees A and B, give clearly conflicting signals about the evolutionary relationship. If one naively assumes that larger data sets are intrinsically more reliable than smaller sets, one might be tempted to simple concatenate the sequences from the two data sets. If the sequences were of equal length, then the pattern frequencies for the combined set would be the averages of those in Table I. The maximum likelihood tree for the combined data set is shown in Fig. 2(c). The combined analysis does nothing to resolve the conflict between the two data sets and in fact suggests the third possible topology $((P, S), (Q, R))$. An additional problem is the extremely short internal edge length. Short internal edge lengths are problematic for many reasons [6]. As one might expect, they are not robust to perturbations in the data or in the underlying evolutionary model. In fact the combined data set provides almost no useful knowledge at all!

5 Multi-objective Optimisation

The naive combined analysis is a special case where the weighting given to the two data sets is equal. In multi-objective optimisation all possible weightings of independent objectives are considered: this corresponds to all possible weightings of the genetic data sets. This yields a family of trees instead of the single trees obtained by separate or combined analysis. This family of trees is called the Pareto set. The likelihood values corresponding to the Pareto set are shown on the Pareto set curve in Fig. 3. We also show curves for what we refer to as constrained Pareto Sets which are discussed later.

Table 2 summarises the topologies, edge lengths and likelihoods of the 12 special trees identified by the multi-objective analysis. Figure 3 summarises the Pareto sets and the smooth curves that connect these special trees. This is certainly a lot more information than either the separate or naive combined analyses. However, these are the 12 trees that will help the experienced biological practitioner interpret the conflict between the data sets and decide a plan of action. We now discuss what is important about this subset of trees.

The three different tree topologies from Fig. 2 occur along the Pareto set, although the edge lengths vary as one travels along the Pareto set. The edge lengths vary smoothly as one travels along the Pareto set, except at points where the topology changes. Thus, from A to E, the tree topology corresponds to $((P, R), (Q, S))$, the internal edge length decreases almost to zero, the length of the edges to P and S decrease slightly, the edges to Q and R increase. At the point E, a transition occurs: two different tree topologies give identical likelihoods

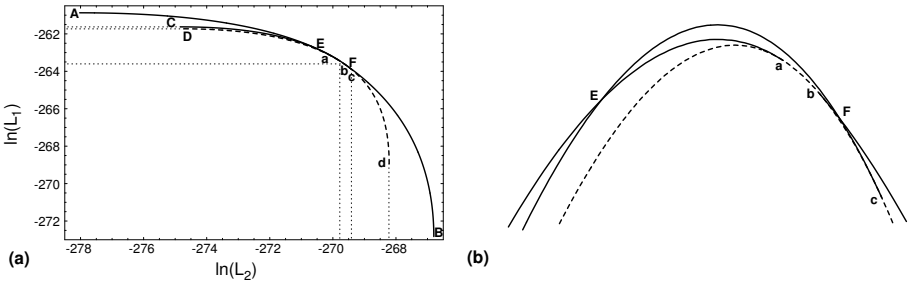

Fig. 3. (a) The Pareto set corresponding to maximising the likelihood with respect to both data sets. The vertical axis corresponds to data set 1, and the horizontal, to data set 2. (b) A close up of part of the Pareto set. The Pareto set has three parts and corresponds to the curves from A to E, E to F and F to B (abbreviated as AEFB). The diagram also shows the constrained Pareto sets: if the topology is forced to be $((P, R), (Q, S))$ then the Pareto set is the curve AEa; for the topology $((P, Q), (R, S))$, the Pareto set is the curve bFB; for the topology $((P, S), (Q, R))$, the Pareto set is CEFc. Finally, the dashed curve shows the Pareto set for the star topology with zero internal edge length and this is the curve Dabcd. The dotted lines show the constrained maximum likelihood values with respect to the two data sets.

with respect to both data sets. From E to F, the tree topology corresponds to $((P, S), (Q, R))$ and the edge lengths do not vary substantially. At the point F, another transition occurs between tree topologies. Finally, from F to B, the tree topology corresponds to $((P, Q), (R, S))$, the internal edge length increases substantially, the length of the edge to R also increases, and the edge lengths to P, Q and S decrease.

5.1 Constrained Pareto Sets

One can obtain additional insights by looking at constrained Pareto sets. The three curves AE, EF and FB in Fig. 3 are each parts of larger curves. If we apply a multi-objective optimisation analysis constrained only to trees with topologies $((P, R), (Q, S))$, then we obtain the curve AEa. One end of this curve corresponds to tree A, which we have already discussed. The other end, at tree a, corresponds to a limiting situation where an internal edge length has gone to zero.

Similarly, constraining the topology to $((P, Q), (R, S))$ yields the Pareto set along the curve bFB. Likewise, constraining the topology to $((P, S), (Q, R))$ yields the Pareto set along the curve CEFc. The tree at the other end of this constrained Pareto curve, is the best tree with topology $((P, S), (Q, R))$ when only data set 1 is considered.

In addition to constraining the topology, one can go further and constrain one or more internal edge lengths to be zero. This reveals the maximum likelihood values of simpler topologies; and the difference between the likelihood values of the simpler topologies and the more complex topologies gives a measure of confidence in the result.

Table 2. Variation of topology and edge length along the Pareto set. The internal edge-length is labelled i, the other edges are labelled by the species at the end. The critical trees on the Pareto sets identified by multi-objective optimisation. An asterisk next to the topology indicates that the tree is a limiting point of that topology as an internal edge length approaches zero.

Tree	Topology	P	Q	R	S	i	$\ln(L_1)$	$\ln(L_1)$
A	$((P,R),(Q,S))$	0.961	0.201	0.454	0.661	0.248	-260.9	-278.0
E	$((P,R),(Q,S))$	0.805	0.636	0.561	0.618	0.012	-263.2	-270.0
	$((P,S),(Q,R))$	0.805	0.636	0.561	0.618	0.024	-263.2	-270.0
F	$((P,S),(Q,R))$	0.783	0.650	0.603	0.612	0.007	-263.7	-269.6
	$((P,Q),(R,S))$	0.787	0.647	0.597	0.615	0.005	-263.7	-269.6
B	$((P,Q),(R,S))$	0.496	0.616	0.808	0.450	0.370	-272.8	-266.8
C	$((P,S),(Q,R))$	0.882	0.550	0.630	0.338	0.091	-261.6	-274.8
D	(P,Q,R,S)	0.943	0.553	0.348	0.698	0	-261.7	-274.8
a	$((P,R),(Q,S))$*	0.799	0.644	0.583	0.619	0	-263.4	-269.8
b	$((P,Q),(R,S))$*	0.791	0.648	0.595	0.617	0	-263.6	-269.6
c	$((P,S),(Q,R))$*	0.776	0.656	0.614	0.618	0	-263.9	-269.4
d	(P,Q,R,S)	0.582	0.668	0.927	0.668	0	-268.9	-268.2

In the four species problem, the only interesting choice of constraint is to force the one and only internal edge to be zero. In more complicated situations, the choice of which internal edge length(s) to constrain would be dictated by the nature of the full Pareto set. In this rather simple situation, this only generates one new constrained Pareto set curve, shown by the dashed curve Dabcd in Fig. 3.

With respect to data set 1, the four trees A, C, D, b represent in order the optimum tree for each possible topology. (Note that tree b is degenerate: an internal edge length is zero. It is a co-incidence of the simple 4 species case that b and D have the same apparent topology; in more complex situations it would be different internal edges that went to zero.) The gaps between the log likelihood values tell us what data set 1 has to say about how much confidence to place in the best tree, or in the order of the competing topologies. For data set 2, the corresponding set of best trees for each competing topology is B, d, c, a. The Pareto sets and the various curves in Fig. 3(a) and the close-up in (b) reveal how this ordering changes as we move from trusting one data set exclusively to trusting the other.

It is only at this point that it begins to be valuable to build in information about variances around these special trees, and to replace entire subsets of trees with single representative trees (or consensus trees) while maintaining the global information about the Pareto sets and the fitness landscape.

6 An EMOOA simulation

To complete this paper, we show how a very simple evolutionary algorithm performs on the same problem.

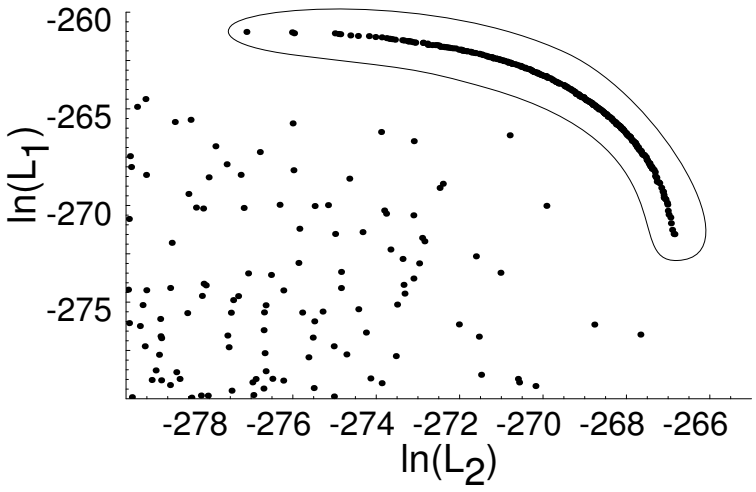

Fig. 4. The scattered dots show the likelihood values obtained from a random population of 100 trees. The encircled dots show the likelihood values for the trees in the non-dominated set after running the genetic algorithm for 50 generations. This non-dominated set is a good approximation to the exact Pareto set.

A very basic algorithm was used. The population size was 100. The genome consisted of a discrete gene with three possible states corresponding to the three distinct topologies, and 5 real-valued genes corresponding to the 5 edge lengths. Pairs of parents were selected randomly (without recourse to fitness) and children constructed by simple Mendelian segregation of the parental genes. A mutation operator was then applied to the genes. A uniformly distributed random mutation with range ± 0.01 was used for the edge lengths. (Negative edge lengths were avoided by taking the absolute value.) In this simple case, it was not found necessary to adapt the mutation rate. The initial population of candidate trees was generated with random topologies and random edge lengths. At each generation the non-dominated set was selected, and the entire non-dominated set became the parents for the next generation — obviously, such a trivial EA would not be suitable for a realistic problem and one should use NGSA-II [16] or an equivalent.

The initial population and a reasonably converged Pareto set is shown in Fig. 4. The evolution of the non-dominated set is shown for the first 4 generations in Fig. 5. The algorithm quickly yields a reasonable approximation to the Pareto set. However, Fig. 4 reveals that the trees are not uniformly distributed along

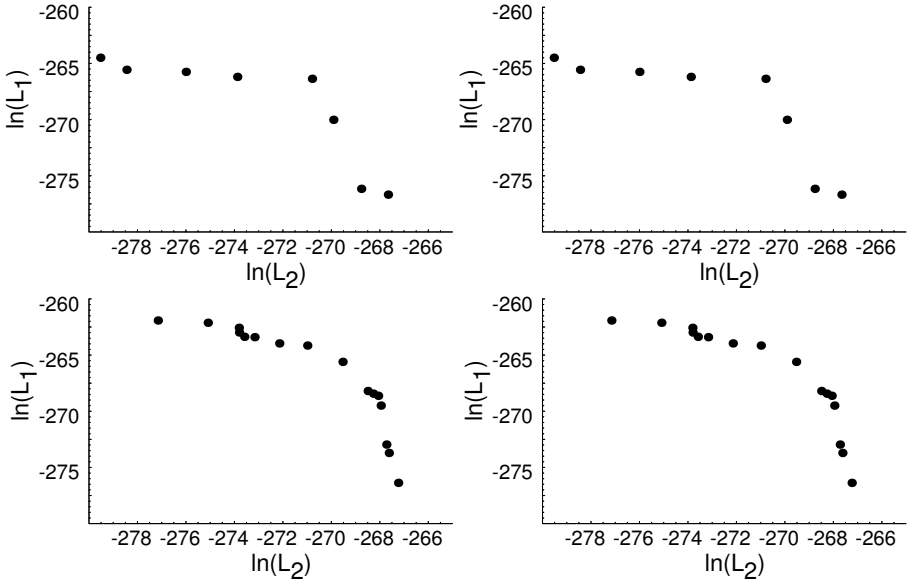

Fig. 5. Evolution of the non-dominated set for the first 4 generations.

the Pareto set curve: trees are missing from the extremities. More sophisticated EMOOA algorithms that incorporate niching would address this problem.

7 Discussion

We have used MOO to study the phylogenetic relationships between four species based on two conflicting data sets. Although, the importance of generalising this analysis to problems with many species is self-evident, it is also important to note that the four species problem itself has special significance. As mentioned in the introduction, Quartet-Puzzling [3] is an extremely popular phylogenetic algorithm. For a problem with a large number of species, the method begins by first looking at all possible quartets: subsets of four species. For each quartet, the optimal four species tree is constructed. The full phylogeny is then pieced together by combining information from all of these quartets. To our knowledge, all current implementations of quartet puzzling work on one data set at a time. Thus, one is either forced to make an *a priori* judgement about combining the data in the total evidence approach, or constructing consensus trees from many separate analyses. The MOO approach suggests a novel mode of attack. For each quartet, the Pareto set is constructed. For a large number of species, it may be that many quartets yield 'non-conflicted' Pareto sets with a common topology inferred from all data sets. The quartet analysis may reveal that only some species are problematic in generating highly 'conflicted' Pareto sets. Thus, it may be possible to work with a coherent or non-conflicting sub-

set of quartets, and puzzle-out a useful sub-phylogeny of the complete problem that is harmonious with all data sets. Alternatively, one can concentrate on the most 'conflicted' subsets of species and discover where to look for the underlying (possibly biological) mechanism that is producing the conflicting signals in the first place. Thus, techniques for rapid determination of the Pareto set for small numbers of species (or at least the 'special trees' in it) may be important.

The intrinsic significance of studying the four-species problem has also been pointed out in a recent paper that demonstrates that even with a single data set, the supposedly trivial four species problem may not have a unique global maximum [17].

Finally, we used the concept of a constrained Pareto set in this paper. The constrained Pareto sets can be seen as a stepping stone towards the unconstrained Pareto set. By working in a smaller dimensional space one can get good approximations to the optimal solutions for the larger dimensional space. Thus, one envisages an EA or EMOOA that starts with phylogenetic tress with many (if not all) internal edge lengths constrained to zero. Then, as the constrained Pareto sets are constructed, one gradually releases the constraints until the full problem is solved. This approach is analogous to a set of techniques known in phylogenetic inference as star-decomposition (also similar to neighbour-joining) [18].

In conclusion, the application of MOO techniques, and EMOOA in particular, to phylogenetic inference can resolve many of the issues concerned with analysing multiple data sets that may give conflicting signals about evolutionary relationships. The construction of the Pareto set moves the point at which human judgment and intervention is required: it moves from *a priori* decisions about the relative importance of data to *a posteriori* analysis of how the conflicting signals interact.

8 Acknowledgements

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

References

1. Salemi, M. and VanDamme, A.-M. (eds.): Handbook of phylogenetic methods. Cambridge University Press, Cambridge (2003).
2. Day, W. H. E.: Computational complexity of inferring phylogenies from dissimilarity matrices. Bull. Math. Biol. **49** (1987) 461–467
3. Strimmer, K. and von Haesler, A.: Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13** (1996) 964–969.
4. Farris, J. S.: A succesive approximations approach to character weighting. Syst. Zool. **18** (1969) 374–385.

5. de Queiroz, A., Donoghue, M.J. and Kim, J.: Separate versus combined analysis of phylogenetic evidence: Ann. Rev. Eco. Syst. **26** (1995) 657–681.
6. Jermiin, L.S., et al.: Majority-rule consensus of phylogenetic trees obtained by maximum-likelihood analysis. Mol. Bio.Evol. **14** (1997) 1296–1302.
7. Wolf, M.J., et al.: TrExML: a maximum likelihood approach for extensive tree-space exploration. Bioinformatics. **16** (2000) 383–394.
8. 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.
9. Lewis, P. O.: A genetic algorithm for maximum likelihood phylogeny inference using nucletide sequence data. Mol. Biol. Evol. **15** (1998) 277–283.
10. Moilanen, A.: Searching for most parsimonious tree with simulated evolutionary optimisation. Clad. **15** (1999) 39–50.
11. 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.
12. 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
13. Jukes, T. and Cantor, C. R.: Evolution of protein molecules. In: Mammalian protein Metabolism. Munro, H. N. ed. (1969) Academic Press, New York, pp. 21-132.
14. Felsenstein, J.: Evolutionary trees from DNA sequences: A maximum-likelihood approach. J. Mol. Evol. **17** (1981) 368–376.
15. http://evolution.genetics.washington.edu/phylip/ newicktree.html
16. Deb, K. and Goel, T. (2001). Controlled elitist non-dominated sorting genetic algorithm for better convergence. Lectures Notes in Computer Science, vol. 1993, pp. 67-81.
17. Chor, B., Hendy, M. D., Holland, B. R. and Penny, D.: Multiple maxima of likelihood in phylogenetic trees: An analytic approach. Mol. Biol. Evol. **17** (2000) 1529–1541.
18. 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.