

Running Head: Heuristic Phylogenetic Estimation

FROM ROLLING HILLS TO JAGGED MOUNTAINS: SCALING OF HEURISTIC SEARCHES
FOR PHYLOGENETIC ESTIMATION

Benjamin Kirkup¹

and

Junhyong Kim^{1,2,3}

¹Department of Ecology and Evolutionary Biology

²Department of Molecular, Cellular, and Developmental Biology

³Department of Statistics

Yale University

Send Proofs to:

Junhyong Kim
Department of Ecology and Evolutionary Biology
P. O. Box 208106
New Haven, CT 06520-8106
(203)-432-9917
(203)-432-3854 (fax)
junhyong.kim@yale.edu

Keywords: Objective function, landscape, Maximum Parsimony, Heuristic search

Abbreviations: Nearest-Neighbor Interchange (NNI), Subtree Pruning and Regrafting (SPR),

Tree Bisection and Reconnection (TBR)

Abstract

The analysis of large-scale data sets is rapidly becoming a central problem in phylogenetic biology and molecular evolution. The algorithms for phylogenetic inference are extremely difficult (*NP-hard*) for many popular estimation methods. Large datasets force the use of heuristic algorithms that are sensitive to the structure of how the function value relates to the function arguments. The relationship between the objective function value and function argument is called the objective function landscape. By examining landscapes constructed from real genetic data for the Maximum Parsimony function we characterized the variables relevant to the success or failure of heuristic algorithms. As expected, the structure of the objective function landscape became more complex and more difficult with increasing numbers of taxa for all search graphs. There were increasing numbers of local optima, decreasing size of the attraction basins, and decreasing correlation between the depth of the optima and their attraction basins. In particular, we found Nearest-Neighbor Interchange search strategy to be inappropriate for heuristic searches. The trees at different local optima can be quite different. We found a topological difference of 40-50% among local optima differing by only 0.5-1.0% in tree length. This indicates a need for different heuristic strategies for large-scale phylogenies such as stochastic search strategies, divide-and-conquer, or direct incremental estimation.

Introduction

Prior to recent times, computational limitations and data availability significantly restricted the use of phylogenetic techniques. However, the infusion of molecular data and the heightened general awareness of phylogenetic approaches are helping to rapidly expand the available data for phylogenetic estimations and allowing researchers to tackle increasingly larger problems. For example, Soltis et al. (1999) published an analysis of 560 species of seed plants using the DNA sequence of three genes. The Ribosomal Database Project (RDP) (Maidek et al. 2000) now contains more than 10,000 putative homologous sequences from almost 2500 species. DNA sequencing is routine, phylogenetically useful genes are continually being discovered, and whole genome comparisons are becoming popular (e.g., Adachi et al. 2000; Jobes et al. 1998; Ursing and Arnason 1998; Zardoya and Meyer 1998) with the result that large-scale data sets are becoming commonplace. The analysis of large-scale data sets is rapidly becoming a central problem in phylogenetic biology and molecular evolution.

Large-scale phylogenies present great computational challenges. The algorithms for phylogenetic inference are extremely difficult (*NP-hard*) for many of the popular estimation methods such as the maximum parsimony method or the maximum likelihood method (Day 1983; Foulds and Graham 1982; Graham and Foulds 1982) requiring heuristic algorithms. For smaller datasets, many heuristic procedures converge to the same solution. However, for larger datasets, the choice of particular heuristic strategies can be extremely important (e.g., Chase et al. 1993; Lewis 1998). Most of the algorithms involve optimization of an objective function of the form $O(D,T)$ where D is the set of datasets and T is the set of binary tree

topologies. For a fixed dataset, d , the problem is to find the value of t , that minimizes/maximizes $O(d, t)$ for all t in T . The difficulty of the problem and the success of the heuristic algorithms depend critically on the structure of the objective function with respect to the argument, t , of the function. Most methods for numerically optimizing a function start with some evaluations of the function over some initial set of argument values and then use the obtained information to evaluate the next set of argument values (Press et al. 1988). The information obtained at each evaluation of the objective function is usually a “local” information in the sense that the function value is evaluated around a small neighborhood of a given argument value. The ultimate success of the heuristic algorithms depends on how this local information relates to the global structure of the objective function. If the objective function is smooth with respect to the argument values—i.e., like rolling hills, then local information such as directions of steepest ascent invariably lead to globally optimal values. On the other hand, if the objective function is rugged like jagged mountain ranges, local information may be poorly related to globally optimal values and heuristic optimization procedures can become stuck on suboptimal solutions.

The critical structure is the neighborhood relationship of the objective function values with respect to the neighborhood relationship of the argument of the function. This structure is called the *landscape* of the objective function (cf. Fontana and Schuster 1998; Huynen et al. 1996; Rogers and Swofford 1999). Figure 1 shows examples of two objective function landscapes where one on the left might be considered an “easy” landscape while the one on the right might be considered “hard.” As mentioned, optimizing an objective function with a hard landscape is difficult because most algorithms use local information to find the global

optimal point. For example, standard hill-climbing algorithms evaluate the function at some initial point then interactively climb towards the steepest direction of ascent. Rough landscapes cause algorithms to become stuck at some local optimum rather than reaching the global optimum. The optimization landscape affects not only greedy hill-climbing algorithms but also many stochastic search algorithms (e.g., Larget and Donald 1999; Lewis 1998) since they also use the information of local structure to modulate the search strategies.

Tree topologies are combinatorial structures with no natural notion of landscape structure because the notion of neighborhoods of trees is not fixed. Search algorithms by necessity impose an ordering on the tree topologies to organize the computations for finding the optimum of the objective function. Ordering of the tree topologies is usually achieved by constructing a search graph. A search graph for tree estimation is a graph where the vertices are binary tree topologies and the edges are neighboring relationship between trees. Figure 2 shows a search graph for five-taxon trees constructed by connecting an edge between all tree topologies that differ by a single internal edge (Nearest-Neighbor Interchange relationship, Swofford et al. 1996). The optimization landscape for tree topologies is defined as the structure of the objective function over the search graph. The landscape is rough and hard if locally connected (by the edges of the search graph) tree topologies have disparate function values. The discrete nature of the search graph makes it a little difficult to intuitively picture the objective function landscape, but the analogies to continuous objective functions are useful and we will discuss our ideas in terms of continuous functions.

Since there is no canonical construction for the search graph for tree topologies, many different rules can be generated for its construction. For example, in PAUP* (Swofford 1998) three different search graphs are implemented: Nearest-Neighbor Interchange (NNI), Subtree Pruning and Regrafting (SPR), and Tree Bisection and Reconnection (TBR) (Swofford et al. 1996). These different methods produce increasingly more connected search graphs. The average number of edges connected to a vertex defines the degree of connection of a search graph. Increasing the connections of the search graph increases the number of tree topologies in the neighborhood of any given tree topology. Increasing the connectedness of a search graph is like taking a continuous objective function and smoothing the function by successive application of a moving window (i.e., averaging the values over some interval). Since smooth landscapes tend to be easier, highly connected search graphs are conducive to finding the global optimum of the objective function. However, there is an obvious tradeoff in terms of the amount of computation needed to determine the local neighborhood values of the objective function. In the extreme, a fully connected search graph has every tree as neighbors to each other and will guarantee finding the global optimum at the cost of having to evaluate all possible tree topologies.

Employing a completely connected search graph is not a heuristic procedure. A heuristic algorithm attempts to find the best optimum while evaluating the smallest number of tree topologies. Critical to the process this process is the structure of the objective function around a given optimum (local or global). There are three features of interest for the structure of the objective function: the depth, the slope, and the attraction basin. Depth simply refers to the function value at the optimum. The slope is a little bit harder to define for a discrete

structure like the search graph, but we use it to refer to how fast the function value changes as we approach the optimum point. The attraction basin refers to the total set of tree topologies from which we will reach the optimum using a greedy hill-climbing strategy. Figure 3 shows these features using a continuous function analogy. For a hill-climbing strategy, the size of the attraction basin is the most critical feature for achieving computational efficiency. An attraction basin means that evaluating any single tree in the attraction basin is equivalent to evaluating every tree in the entire basin because all trees lead to the same optimum. If the objective function landscape had the entire set of possible trees as the attraction basin, we would be guaranteed of reaching the global optimum and we would be able to do this in a very efficient manner. Some heuristic methods such as simulated annealing use the information in the slope of the function, in which case, it is important for the optimum to have a steep slope as well as having a large basin of attraction.

Larger datasets force the use of heuristic algorithms and those employing some kind of search strategy are sensitive to the structure of the objective function landscape. The number of binary tree topology grows extremely rapidly with respect to the number of terminal taxa (Felsenstein 1978), thus, the search problem grows exponentially hard. A critical question then is how the objective function landscape changes as the number of taxa increases. Ideally, features like the size of the attraction basin should keep up with the growth in the size of the possible binary trees; it most likely does not keep up—but then how does it scale? The presence of local optima is a source of problem. How do the number of local optima scale with the number of taxa? If a certain amount of computing effort was devoted to a heuristic search with a 20-taxon tree, how much increase in the effort is needed at 30-taxon

trees in order to have similar degree of satisfaction? In this paper, we attempt to answer these questions by examining the structure of the objective function landscape for the popular Maximum Parsimony (MP) function. Our goal is to characterize the changes in the features of the objective function landscape as the number of taxa increases. We ask the rather naïve question of whether currently popular heuristics using search strategies is a reasonable procedure for large-scale data sets.

Methods

Data Set

The structure of the objective function landscape is dependent on the dataset. Because of the intensive computational effort needed for this project we chose a single dataset for our investigations. This dataset is a 232 sequence/taxon rbcL dataset with 1850 aligned positions originally distributed as a challenge dataset from the Green Plant Phylogeny Coordination Group (<http://ucjeps.berkeley.edu/bryolab/greenplantpage.html>).

Software

PAUP* beta (4.0b1 for DOS, Fri Nov 13 19:02:03 1998) and custom scripts were used for all computations. In all of the heuristic searches described below we used the following settings in PAUP*: start=stepwise; addseq=random; randomize=trees; multrees=yes; savereps=yes. Steepest descent and rooting were not used.

Design of the experiments

We first randomly sub-sampled the 232-taxon dataset to generate smaller datasets ranging in size from 5 to 30 taxa. For each numbers of taxa and search graph, we sub-sampled 50 random replicate datasets. This project required over 2000 CPU hours (350 Mhz, Pentium) and we encountered various computational problems including inevitable system crashes. For NNI searches, too many local optima caused output overflow problems for larger numbers of taxa. Therefore, in some of the results presented below only a partial set of the replicates were analyzed (shown in Table 1). We initiated heuristic searches using each of the three different search graphs generated by NNI, SPR, and TBR. In the following, we will call each combination of dataset and search graph a landscape. For example, a particular 10-taxon dataset with NNI search will be one landscape and the same data set with TBR search will be a different landscape. We started the search at a random tree topology and used hill-climbing to reach the local optimum. This was repeated many times (between 5,000 to 10,000; Table 1) to assess how the different initial positions result in different optima. For a given optimum tree T, the size of the attraction basin for that optimum was defined as the number of initial positions resulting in the same tree T. For each landscape we kept track of the number of different optima, the tree topologies, and the size of the attraction basin of each optimum. Many of the optima contained multiple trees and were more of a “plateau” than a peak (cf. Maddison 1991). We also noted the number of trees in each plateau for these cases. All output was logged in a verbose format and processed for the various statistics. Table 1 shows the variables involved in the experimental design.

Results

In the following, local optimum refers to any minimum length value tree found as the result of a hill-climbing heuristic search starting from some random tree topology. We do not have a guaranteed global optimum for the heuristic searches. Therefore, we will call the shortest length found for a given landscape the minimum value and the corresponding tree(s) minimal optima. We will also use NNI, SPR, and TBR to refer to the search graphs formed by using these branch-swapping strategies. For a given combination of search graph and taxon size, the averages are computed over the different replicate landscape samples taken from the 232-taxon dataset.

Number of optima and length variation

Table 2a,b shows the average number of local optima and the number of minimal optima found for NNI, SPR, and TBR from 10-taxon problems to 30-taxon problems. (For NNI at 20 taxa, SPR at 25 and 30 taxa and TBR at 30 taxa, only 5,000 random positions were examined compared to 10,000 for the other treatments. Therefore, the numbers are doubled from the raw data to allow comparison.) The average number of local optima is an increasing function of numbers of taxa. For NNI at 20 taxa, 74% of the replicate searches resulted in a different local optimum. The average number of local optima for SPR and TBR did not increase as dramatically; a new local optimum was found for every ~65-80 replicates at 30 taxa. Nevertheless, given its continuous growth, the number of different local optima may also

become very large for SPR and TBR at higher number of taxa. The number of distinct optima at the minimal length value also increased with the numbers of taxa as well as the number of different tree topologies with the minimal length (Table 2b,c).

We also examined the variation in the tree length at different optima for a given landscape. We computed the length of the trees at the different optima, measured the interquartile range (75 percentile – 25 percentile) of this distribution and expressed it as a percentage of the average length of the optima. For NNI, the values are 4.24%, 3.60%, and 3.79% for 10, 15, and 20 taxa. Therefore, the % variation in the tree length at different optima remained reasonably constant regardless of the numbers of taxa. For SPR the values are 0.08%, 0.37%, 0.50%, 0.40%, and 0.33%, and for TBR the values are 0.03%, 0.32 %, 0.60%, 0.44%, and 0.37% for 10, 15, 20, 25, and 30 taxa, respectively. Again, for both SPR and TBR, the percent variation in the tree length at different optima is approximately constant irrespective of the numbers of taxa and the variation is relatively low.

Attraction basin for local and minimal optima

The attraction basin for a given optimum, T, refers to the percentage of the random initial trees ending up at T with a hill-climbing algorithm. Table 3 shows the summary statistic measuring the attraction basin size as a function of number of taxa and search graph. The first number in each cell of Table 3a shows the average size of the basins for all local optima while the second number (bold) shows the average size of the largest attraction basin. For all search graphs the average sizes of the basins decreases with increasing numbers of taxa

indicating an increase in the complexity of the objective function landscape with increasing numbers of taxa. The average size of the largest attraction basin also decreases with increasing numbers of taxa. As the average size of the attraction basins shrinks, it becomes increasingly important to test different random initial starting points. Table 3b shows two numbers characterizing the size of the minimal optima. In each cell of the table, the first number is the attraction basin of the minimal optima where the values were pooled if multiple minimal optima were present. The second number in bold is the percent of the time that the largest attraction basin was also the minimal optimum. For both SPR and TBR, the largest attraction basin was the minimal objective function value more than 90% of the time while for NNI the percentage decreased significantly with increasing numbers of taxa. If the objective function landscape is reasonably smooth, we expect a correlation between the size of the attraction basins and the depth of the optima. We measured this correlation for all landscapes where there were five or more local optima. Table 3c shows the results of this computation. The values in parenthesis show the total number of landscapes with five or more local optima. For TBR at 10 taxa, all landscapes resulted in less than five optima and we could not compute the correlation. As seen in the table, there is a high negative correlation (i.e., shorter the tree, larger the attraction basin) at smaller numbers of taxa that decreases with increasing numbers of taxa. The decrease in the correlation again shows that the complexity of the objective function landscape increases with increasing numbers of taxa.

Congruence of tree topology at different local optima

As shown in Table2, the number of different local optima range from a mean of 1.3 to ~7400. The large number of different local optima is computationally problematic but the situation may be still reasonable if the tree topologies at different local optima were similar to each other. We measured the similarities of the tree topologies at different local optima by computing the average of the partition metric (Robinson and Foulds 1981) between all pair-wise comparisons of the trees at all local optima. The partition metric was then normalized between 0 (all splits agree) and 1 (no splits agree) by dividing by the maximum possible splits. The average normalized partition metric for NNI was 0.877, 0.834, and 0.790 for 10, 15, and 20 taxa. The average normalized partition metric for SPR was 0.422, 0.400, 0.482, 0.483, and 0.473 and for TBR was 0.329, 0.408, 0.518, 0.420, and 0.472 for 10, 15, 20, 25, and 30 taxa, respectively. Thus, there did not seem to be an effect of the numbers of taxa on how different the trees are at disparate local optima. On the other hand, the differences in the tree topology are rather large with nearly 50% average difference in splits for SPR and TBR and more than 80% difference for NNI.

Discussion

As expected, the structure of the objective function landscape became more complex and more difficult with increasing numbers of taxa for all search graphs. There were increasing numbers of local optima, decreasing size of the attraction basins, and decreasing correlation

between the depth of the optima and their attraction basins. The decreasing correlation is especially troublesome. Unless the deeper optima also have larger attraction basins, they will be difficult to find. On one hand, we did find that the largest attraction basins were the most optimal solutions more than 90% of the time for SPR and TBR searches. On the other hand, this also was a decreasing function of the number of taxa.

The disagreement of the tree topologies at different optima is especially problematic. For SPR and TBR, the tree topologies at the different optima were different by at least 40-50% of the splits. With NNI search graph, the tree topologies were different by almost all of the splits. Our experimental design did not permit us to assess directly whether shorter trees are more similar to the shortest trees. However, in the SPR and TBR search graphs, the lengths of the trees in the local optima differed by less than 1% of each other. It is the nature of NP-hard problems that solutions with similar objective function values may be arbitrarily different from each other. However, since we have no guarantee of reaching the global optima, usual tree searches will always end up at some local optimum. If this local optimum bears no resemblance to the global optimum, the search problem will be hopeless. Therefore, we operate under the idea that there is at least a statistical relationship between similarities in the objective function value and the tree topology. Within the limited dataset we studied, there does not seem to be a very strong relationship between the objective function value and tree topology. This contrasts with simulation studies of large phylogenies (e.g., Hillis 1996) that suggest local solutions are reasonable solutions. One possibility is that simulation data is less complex than the real data set we explored here. It is also possible that at trees at local

optima will tend towards greater resemblance at the more optimal end of the objective functions. Future studies are needed to examine this possibility.

One of the goals of this study was to assess better heuristic strategies for large-scale phylogenies. Given limited computational resources, questions are typically asked whether to devote more time to examining different random initial positions, different search neighborhoods, or entirely different class of search algorithms. It is a little bit difficult to answer these questions since different search graphs produce different statistical characteristics of the objective function landscape. For example, at higher numbers of taxa NNI search resulted in a different local optimum for almost every random initial tree. In such cases, it may be computationally more efficient to simply evaluate random trees rather than use any hill-climbing strategy. At any rate, our study shows that NNI cannot be recommended as a heuristic strategy given the number of local optima, lack of correlation between the attraction basin and the objective function value, and the topological difference between the different optimal trees. The difference between SPR and TBR is minor. In our analysis, the difference in computational time was about two-fold, which suggest that SPR might be a reasonable generic search strategy.

Of the analysis presented here, the most important statistic is the decrease in the correlation between the size of the attraction basin and the depth of the objective function. It is likely that the “peaks” in the objective function landscape are becoming narrower as the number of taxa increases. One strategy to overcome this is to try many different initial positions to find the more optimal solutions, but with limited computation time, such a strategy will

increasingly fail since the probability of landing in the attraction basin of a large peak continues to decrease with the number of taxa. This suggests that more effort needs to be spent finding strategies to “smooth” the objective function rather than evaluating many different initial positions. Using an increasingly larger definition of neighborhoods such as in the progression from NNI to SPR to TBR is one way to smooth the landscape. Because of computational limitations, we were not able to thoroughly evaluate the efficacy of SPR or TBR at higher numbers of taxa, but we suspect that even SPR or TBR will show rapid degradation in landscape characteristics within a moderate increase of taxa. This seems to call for different heuristic strategies for large-scale phylogenies such as stochastic search strategies, divide-and-conquer, or direct incremental estimation such as in the Neighbor-Joining method. Extremely large phylogenetic problems already exist and many more are looming on the horizon (e.g., the much debated project to estimate the entire tree of life). Further understanding of the algorithmic problem and new solutions are desperately needed to meet these challenges.

Acknowledgements: This work has been supported in part by NSF DEB-9806570 and INTEL Education for 2000 grant to JK.

Table 1. Experimental design and computational time

# Taxa	NNI ^a	SPR ^b	TBR ^c
10	# ReArr ^d = 109.0	375.8	549.5
	Time/Land ^e = 0.0086	0.013	0.018
	# Land ^f = 50	50	50
	# Initial ^g = 10000	10000	10000
15	405.3	3175.4	6024.0
	0.0412	0.108	0.189
	25	50	50
	10000	10000	10000
20	1021.5	13661.5	30069.1
	0.350	0.564	0.736
	25	50	23
	10000	10000	10000
25	N/A ^h	46696.6	78486.0
		2.65	4.30
		30	50
		5000	10000
30	N/A ^h	71046.2	238942.3
		4.22	6.59
		28	42
		5000	5000

^aNearest-Neighbor Interchange search

^bSubtree Pruning and Regrafting search

^cTree Bisection and Reconnection search

^dAverage number of tree rearrangements to reach a local optimum from a given random initial tree.

^eAverage computational time (in seconds) per search on a 350Mhz Pentium computer.

^fNumber of different landscapes analyzed

^gNumber of different initial trees analyzed per landscape

^fNo data were collected for these treatments as NNI searches resulted in all different local optima and outflow error.

Table 2. Scaling of the number of optima with respect to tree size for NNI, SPR, and TBR search graphs

# Taxa	# Local Optima ^a			# Minimal Optima ^b			Total # Minimal Trees ^c		
	NNI ^d	SPR ^e	TBR ^f	NNI	SPR	TBR	NNI	SPR	TBR
10	133.9	1.8	1.3	1.6	1.3	1.1	3.0	3.9	3.3
	(10.93) ^g	(0.21)	(0.9)	(0.17)	(0.08)	(0.05)	(0.41)	(0.51)	(0.45)
15	1646.7	3.9	3.8	2.6	1.7	1.5	7.2	7.8	5.2
	(152.82)	(0.51)	(0.41)	(0.45)	(0.22)	(0.12)	(1.98)	(1.98)	(0.93)
20	7396.4 ^h	20.9	14.5	8.0	2.36	2.2	20.6	11.2	9.7
	(402.18)	(4.33)	(2.03)	(1.42)	(0.28)	(0.35)	(6.83)	(2.08)	(3.44)
25	N/A	124.0 ^h	19.6	N/A	5.2	4.6	N/A	26.9	28.3
		(33.40)	(2.94)		(0.92)	(0.44)		(8.85)	(5.56)
30	N/A	156.0 ^h	129.4 ^h	N/A	5.0	6.4	N/A	57.1	29.4
		(52.98)	(24.40)		(1.00)	(1.28)		(29.84)	(5.90)

^aLocal optimum is defined as any optimum tree reached from a random initial tree.

^bMinimal optimum is defined as the shortest tree(s) reached from a set of random initial trees.

^cThe number of equally shortest tree in the minimal optima

^dNearest-Neighbor Interchange search

^eSubtree Pruning and Regrafting search

^fTree Bisection and Reconnection search

^gStandard error

^hThese treatments only $\frac{1}{2}$ the random initial trees compared to the other treatments. The numbers for these treatments were rescaled to be comparable.

Table 3. Attraction basin statistics for NNI, SPR, and TBR search graphs

	Average size of all basins and the largest basin ^a			Average size of minimal basin ^b			Correlation between depth and basin size ^c		
	NNI ^d	SPR ^e	TBR ^f	NNI	SPR	TBR	NNI	SPR	TBR
10	1.62 ^g	74.90	87.17	38.5 ⁱ	94.7	96.9	-0.405	-0.894	N/S
	34.2^h	85.6	93.8	86.0^j	98.0	98.0	(50) ^k	(3)	
15	0.08	46.96	43.47	14.5	90.3	83.6	-0.210	-0.642	-0.685
	10.1	79.8	75.2	84.0	98.0	94.0	(25)	(15)	(17)
20	0.03	13.08	14.34	4.1	77.3	71.6	-0.177	-0.538	-0.619
	2.2	58.9	56.2	36.0	98.0	91.0	(29)	(42)	(20)
25		8.24	12.30	N/A	64.5	70.9	N/A	-0.458	-0.526
		48.6	54.8		100.0	92.0		(27)	(44)
30		3.27	4.95	N/A	57.2	62.4	N/A	-0.430	-0.433
		45.8	47.1		93.0	93.0		(28)	(45)

^{a,g}Average size of all basins over each landscape as well as the ^hlargest attraction basin in a given landscape. The values are given in terms of percent random initial trees ending at a particular optimum.

^{b,i}Average size of the basin with the minimal objective function value and the ^jpercent of the landscapes where the optimum with the largest attraction basin was also the minimal optimum.

^cThe correlation between the tree length at a given optimum and the size of its attraction basin. Only those landscapes with five or more local optima were included in the computation.

^dNearest-Neighbor Interchange search

^eSubtree Pruning and Regrafting search

^fTree Bisection and Reconnection search

^kThe number of landscapes included in the computation of the correlation.

Figure Legends

Figure 1. A schematic diagram of objective function landscape. The Y axes represent the objective function value while the X axes represent the neighborhood relationship of the search space. The landscape on the left is smooth and easier to optimize while the landscape on the right is rugged and difficult to optimize.

Figure 2. Nearest-Neighbor Interchange (NNI) search graph for five-taxon trees. All 15 possible binary trees for five-taxon trees are shown as vertices of a graph. The tree topologies are drawn in parenthesis format and labeled in the legend. The edges connect the tree topologies that differ by a single NNI operation.

Figure 3. Statistical characteristics of a local optimum in a landscape. The depth of the optimum is the difference of the optimal value from an average value. The slope of the optimum is how fast the function value changes along the direction of the local optimum. The attraction basin is the size of the base of the optimum where all trees within the attraction basin are guaranteed to converge to the local optimum by a hill-climbing search.

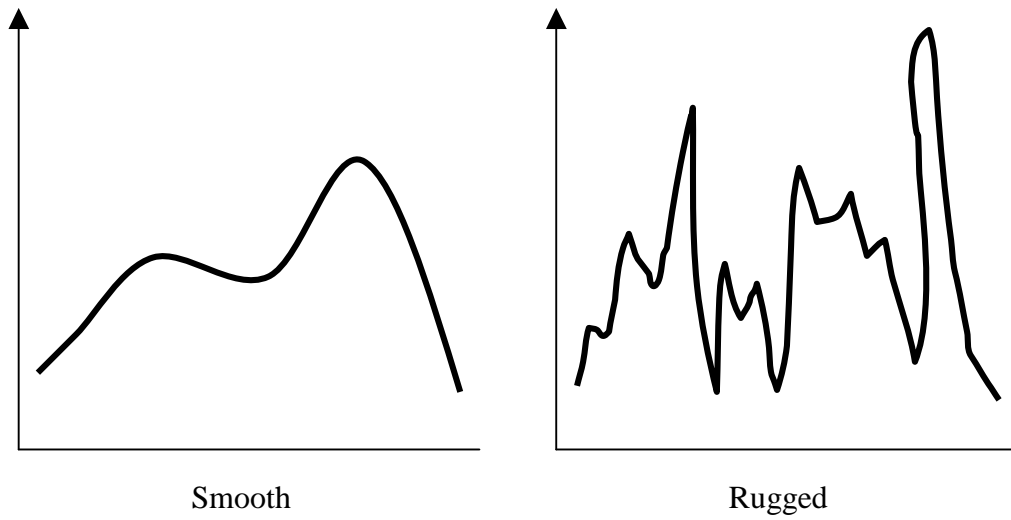
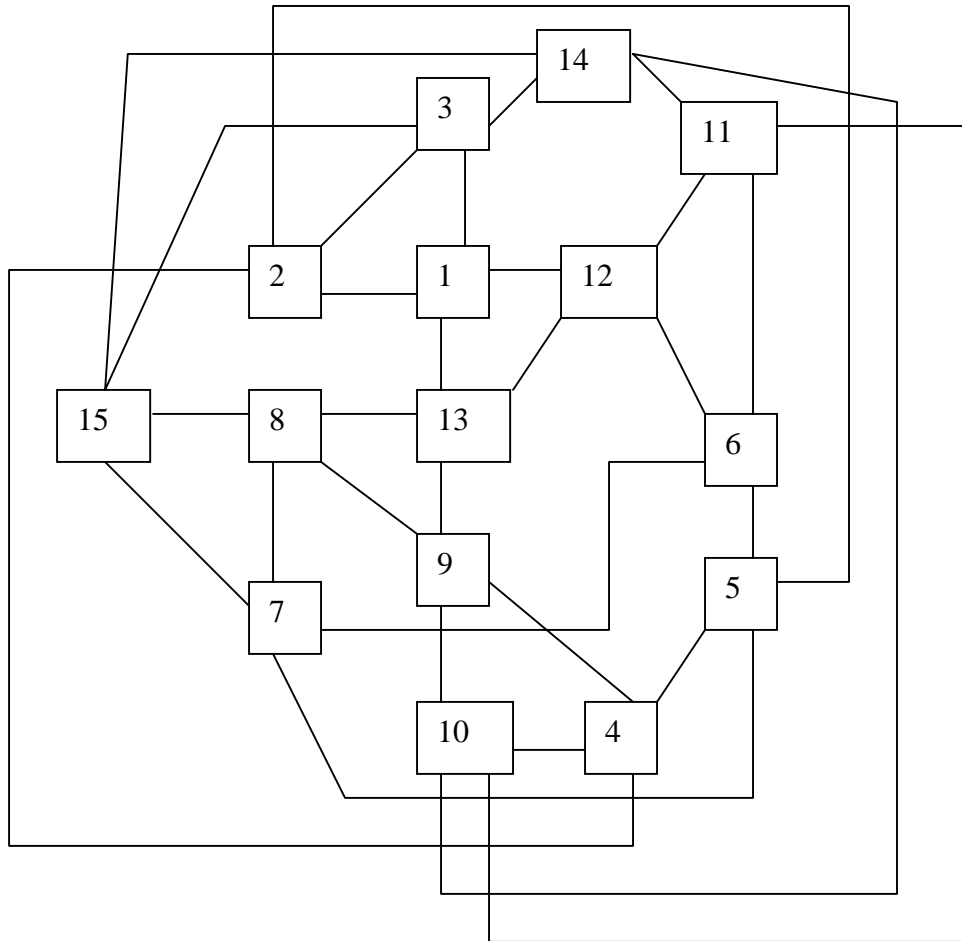


Figure 1



- | | | |
|----------------------|-----------------------|-----------------------|
| 1: ((AD),(C,(B,E))) | 6: ((A,E),(D,(C,B))) | 11: ((B,C),(A,(E,D))) |
| 2: ((B,E),(A,(C,D))) | 7: ((A,E),(C,(B,D))) | 12: ((A,D),(E,(B,C))) |
| 3: ((A,C),(D,(B,E))) | 8: ((B,D),(A,(E,C))) | 13: ((A,D),(B,(E,C))) |
| 4: ((A,B),(E,(C,D))) | 9: ((A,B),(D,(C,E))) | 14: ((A,C),(B,(E,D))) |
| 5: ((A,E),(B,(C,D))) | 10: ((A,B),(C,(E,D))) | 15: ((A,C),(E,(B,D))) |

Figure 2

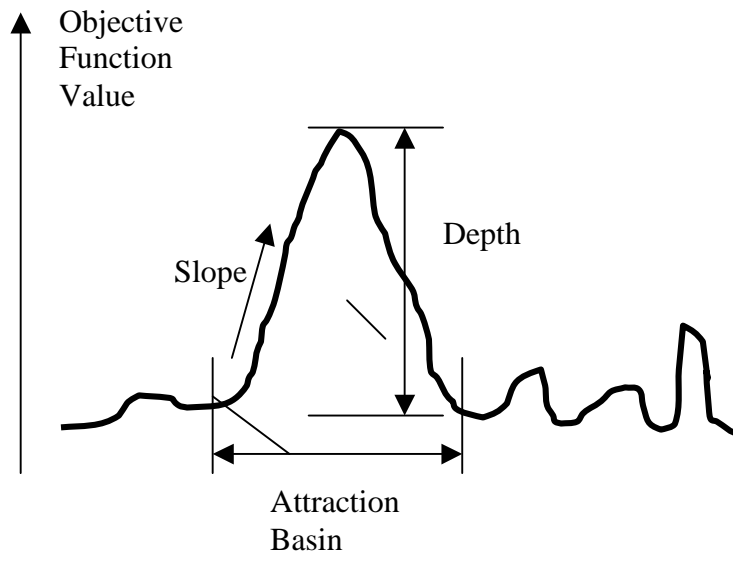


Figure 3

Literature Cited

- Adachi J, Waddell PJ, Martin W, Hasegawa M (2000) Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *J. Mol. Evol.* 50:348-358
- Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Y.-L.Qiu, al. e (1993) Phylogenetics of seed plants an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Gard.* 80:528-580
- Day WHE (1983) Computationally difficult parsimony problems in phylogenetic systematics. *J. Theor. Biol.* 103:429-438
- Felsenstein J (1978) The number of evolutionary trees. *Syst. Zool.* 27:27-33
- Fontana W, Schuster P (1998) Continuity in evolution: On the nature of transitions. *Science* 280:1451-1455
- Foulds RR, Graham RL (1982) The steiner problem in phylogeny in NP-complete. *Adv. Appl. Math.* 3:43-49
- Graham RL, Foulds LR (1982) Unlikelihood that minimal phylogenies for a realistic biological study can be constructed in reasonable computational time. *Mathematical Biosciences* 60:133-142
- Hillis D (1996) Inferring complex phylogenies. *Nature* 383:130-131
- Huynen MA, Stadler PF, Fontana W (1996) Smoothness and ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA* 93:397-401
- Jobes DV, Chima SC, Ryschkewitsch CF, Stoner GL (1998) Phylogenetic analysis of 22 complete genomes of the human polyomavirus JC virus. *J. Gen. Vir.* 79:2491-2498

- Larget B, Donald SL (1999) Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16:750-759
- Lewis PO (1998) A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data. *Molecular Biology and Evolution* 15:277-283
- Maddison DR (1991) The discovery and importance of multiple islands of most-parsimonious trees. *Systematic Zoology* 40:315-328
- Maidek BL, Cole JR, Lilburn TG, Parker Jr CT, Saxman PR, Stredwick JM, Garrity GM, Li B, Olsen GJ, Pramanik S, Schmidt TM, Tiedje JM (2000) The RDP (Ribosomal Database Project) continues. *Nucl. Acids Res.* 28:173-174
- Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1988) *Numerical Recipes in C*. Cambridge University Press, New York
- Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. *Math. Biosci.* 53:131-147
- Rogers JS, Swofford DL (1999) Multiple local maxima for likelihoods of phylogenetic trees: A simulation study. *Molecular Biology and Evolution* 16:1079-1085
- Soltis PS, Soltis DE, Chase MW (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402:402-404
- Swofford DL (1998) PAUP* 4.0 beta. Sinauer Assoc., Sunderland, MA
- Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK (eds) *Molecular Systematics*. Sinauer Associates, Sunderland, p 407-514
- Ursing BM, Arnason U (1998) Analyses of mitochondrial genomes strongly support a hippopotamus-whale clade. *Proc. R. Soc. Lond. B* 265:2251-2255

Zardoya R, Meyer A (1998) Complete mitochondrial genome suggests diapsid affinities of turtles. *Proc. Natl. Acad. Sci. USA* 95:14226-14231