Using Avida to Test the Effects of Natural Selection on Phylogenetic Reconstruction Methods

George I. Hagstrom
California Institute of Technology
Pasadena, CA 91125
hagstrom@caltech.edu

Dehua H. Hang
Charles Ofria
Eric Torng
Department of Computer Science and Engineering
Michigan State University
East Lansing, MI 48824
hangdehu@msu.edu
ofria@msu.edu
torng@cse.msu.edu

Abstract Phylogenetic trees group organisms by their ancestral relationships. There are a number of distinct algorithms used to reconstruct these trees from molecular sequence data, but different methods sometimes give conflicting results. Since there are few precisely known phylogenies, simulations are typically used to test the quality of reconstruction algorithms. These simulations randomly evolve strings of symbols to produce a tree, and then the algorithms are run with the tree leaves as inputs. Here we use Avida to test two widely used reconstruction methods, which gives us the chance to observe the effect of natural selection on tree reconstruction. We find that if the organisms undergo natural selection between branch points, the methods will be successful even on very large time scales. However, these algorithms often falter when selection is absent.

Keywords
Phylogeny reconstruction, digital evolution, molecular clock, natural selection, Avida

1 Introduction

The universality of the genetic code has led researchers to the conclusion that all known living organisms share a common ancestor from about 3.5 billion years ago. However, there is a lot more that we can say about ancestral relationships. For example, all animals have a single ancestor thought to be about 600 million years old, and all mammals have a more recent ancestor still (about 200 million years old). As Dobzhansky stated "Nothing in biology makes sense except in the light of evolution." Understanding the evolutionary relationships between organisms is a critical key to understanding biology. If we understand which organisms are similar to each other then we can, for example, examine poorly understood traits through their less novel homologues in other species. The field of *phylogenetics* [5] is aimed at revealing what the relationships between species are, and its most fundamental problem is constructing the tree of life, linking all species together. Moreover, research in phylogenetics has the potential to improve research in other fields such as epidemiology, where organisms are evolving rapidly enough to have large effects on society. As an example, it would be useful to learn how, in different situations, viruses and bacteria develop resistance to drugs and vaccines. Knowledge of phylogenetic trees also allows us to measure the statistical significance of evolutionary observations, specifically in cases when it is possible that a specific trait developed more than once. An understanding of the evolutionary relationships between the species is necessary for the advancement of biology.

Using mathematical algorithms, one can create phylogenetic trees from data sets containing the genomes of extant organisms. Here, we will focus on two methods that accomplish this: maximum parsimony and neighbor joining. Since we are using

a non-DNA-based system, we are currently limiting ourselves to studying methods that do not rely on nucleotides or proteins. Most notably absent is maximum likelihood, a well-used algorithm that requires a point accepted mutation (PAM) matrix, which does not currently exist for digital organisms. It is likely that such a PAM matrix will be available in the near future, which should further energize research in this vein.

Maximum parsimony is a reconstruction method heralded by biologists for its accuracy and its adherence to sensible scientific principles. Also known as *minimum evolution*, maximum parsimony uses *Occam's razor*,[1] the idea that the simplest explanation is the one most likely to be applicable, to reconstruct phylogenetic trees [3]. Maximum parsimony works by examining the space of all possible trees and producing the one that requires the fewest mutations. Although this is a reasonable method, it suffers from problems due to the exponential computational complexity of generating all of the possible trees for a given set of organisms [12]. This has prompted the creation of maximum parsimony methods that restrict the space of trees that are searched to a subspace that is likely to contain the minimum mutation tree. It is also important to note that maximum parsimony occasionally makes large scale topological mistakes [2].

Neighbor joining is one of the most intuitive methods for reconstructing phylogenetic trees. Neighbor joining relies on the creation of a distance matrix that contains the genetic distance between each genotype within the data set [14]. This genetic distance can be determined in a number of ways. One option that is widely used is the Hamming distance, which is simply the number of sites at which two genotypes are different [2]. This distance matrix is then used to reconstruct the tree, by using an algorithm that groups those genotypes into the clusters that are closest together in genetic distance, while allowing for differences in mutation rate between branches. This process is iterated to group clusters into superclusters based on their average distance to each other, until the tree is complete. This method has some advantageous properties, the most important being that its algorithmic complexity is $O(n^3)$ [12]. The other tree reconstruction schemes are generally unable to reconstruct trees in polynomial time. For many problems that involve sizable numbers of genotypes, neighbor joining is the only computationally tractable method feasible for use. However, prior research suggests that it is less accurate than maximum parsimony [2].

Because of the recent advancements in technology in the field of genetics, there are a large number of genes and even full genomes that have been sequenced and are available for analysis. In the absence of any analytical or a priori results (beyond those mentioned above) that distinguish the different phylogeny reconstruction algorithms, it is important to understand how accurate these methods are under different sets of circumstances [8]. This is a substantial problem, since there do not exist many phylogenies in nature that are known precisely, and fewer still on a fine scale within a species. In order to alleviate this difficulty scientists have generated test data sets representing phylogenetic trees where internal sequences are known but only the final data points are fed into the algorithms. This has been done in primarily two different ways: using actual organisms under the influence of a mutagen, or using a computer to generate strings of data that are subject to random mutations [9]. Both methods rely heavily on *neutral mutations*, that is, genetic changes that do not have any phenotypic effect. However, using a computer to create strings that randomly mutate appears to remove all of the complexity of biological macroevolution, which is due to *adaptive* (i.e., beneficial) mutations. The same is also largely true of the studies on actual biological organisms, since they involved populations that did not have to adapt to changing environmental conditions [9]. These studies are not without utility, considering the fact

1 *Pluralitas non est ponenda sine neccesitate*, or "Plurality should not be posited without necessity."

that some interesting phylogenies, such as those for tracking fast-spreading diseases, are governed mostly by neutral mutations.

In certain situations it seems reasonable to expect that these sequences of randomly changing nucleotides could become saturated with mutations. Saturation has dire implications for tree reconstruction. Any two saturated sequences will be separated by a random Hamming distance centered about the ratio of the sequence length to the number of different types of bases constituting the genetic code. Thus, saturated sequences are of no use to phylogenetic reconstruction algorithms. The reason that sequences become saturated is that there is no selective pressure on those sequences. Obviously, not all of the sequences of genes in the genome can become saturated; some of them have become fixed by selection. When reconstructing phylogenetic trees over long periods of time, the neutral sites in the genome will become saturated. This would force the tree reconstruction methods to deal with sections of the genome that have become fixed due to selective events. Using the model cases that researchers have used to generate sample phylogenies could provide erroneous results, since they by definition miss all of the complexity that selection could create.

In this article, we use Avida, a computer program that creates an environment in which digital organisms can evolve. Digital organisms are self-replicating computer programs that have genomes composed of strings of instructions (see [13] in this issue for details). Experiments studying the evolution of digital organisms have addressed a number of current issues in biological evolution (see [15] for a recent review). However, unlike biological evolution, evolution in digital organisms happens rapidly due to a short generation time, and the history of each run is known. Therefore, Avida is an ideal system to quickly create known phylogenies. Here, we demonstrate that over long time scales (i.e., once all neutral sites have reached saturation) tree reconstruction methods can group organisms accurately only if natural selection shaped the formation of the tree, since this will often cause genomic distinctions to be maintained. Specifically, groups of organisms that should be paired together must have had an adaptive event that distinguishes them from all other organisms in the population.

2 Materials and Methods

We performed all experiments using Avida version 2.0 (beta 3), freely available from http://www.sourceforge.com/projects/avida. We generated sample phylogenetic trees by starting with a single seed organism as the root. We caused speciation and divergence through competitive isolation, by choosing the dominant (most abundant) genotype at the end of a run and placing a single copy of it in two separate (isolated) populations. The ancestral seed organism resulted from prior evolution and has a genome with 100 instructions. The ancestor was evolved in the *standard* environment, which rewarded six logical operations: `NOT`, `NAND`, `AND`, `ORN`, `OR`, and `ANDN`, as shown in Table 1. In these experiments, we enforced a constant genome length by preventing insertions, deletions, or implicit mutations that would shift the positions of genetic in-

Table 1. A listing of the tasks that were rewarded in each environment. The alternate environment rewards three additional tasks beyond those in the standard environment.

Standard environment	Alternate environment
NOT, NAND	NOT, NAND
AND, ORN	AND, OR
OR, ANDN	OR, ANDN
	NOR, XOR, EQU

formation within the genome. This simplifies analysis substantially, because otherwise the sequences of the organisms would need to be aligned in order for the phylogenetic reconstruction techniques to be usable. The carrying capacity of each experiment was set to 2,500 organisms (on a 50 by 50 grid) and continued for 20,000 updates (about 2,000 generations.)

To generate a test phylogeny, we placed the ancestral seed organism in a population that we ran for 20,000 updates. The dominant genotype in this population was then placed in two isolated Avida populations, which were each run for an additional 20,000 updates, during which time further adaptation was expected. From each of these two populations an organism was selected from the dominant genotype to start two more populations, for a total of four simultaneous populations in this final round. Again, these populations were allowed to progress for 20,000 updates, and at the end four organisms (one from the dominant genotype in each population) were chosen to be used for the reconstruction. See Figure 1 for a diagram of this experimental method. Since that branch of the tree is produced by an individual Avida run, it is easy to control the environment along each branch, therefore allowing for the study of natural selection on the tree reconstruction algorithms.

We generated a total of five sets of phylogenies under different environmental conditions. Each phylogeny had exactly seven branches: one that formed the root (and

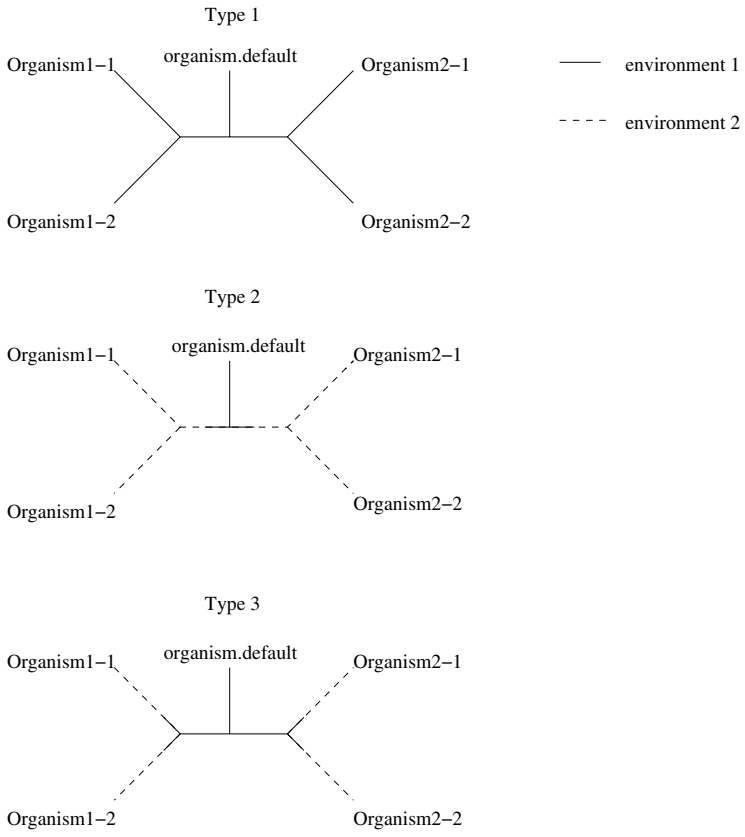

Figure 1. Diagrams of the environments used to create the test phylogenies. Type 1 uses the standard environment throughout; type 2 shifts to the alternate environment after the first branch point; type 3 shifts to the alternate environment only after the second branch point. Solid lines represent the standard environment (environment 1), and dashed lines the alternate environment (environment 2).

therefore cannot be reconstructed), two that branch off it at the first level, and four that branch off them at the second level. The ancestor was near a global fitness maximum for the standard environment. Phylogenies generated in set I used the standard environment all the way through (tree type 1), and thus had very little additional evolution. Set II changed the environment to the alternate environment from right after the first branch point (tree type 2), causing the organisms to be immediately subject to adaptive change. Set III was subjected to the alternate environment after the final branching (tree type 3), allowing new selective pressures only at that point. Set IV had conditions identical to those of set I, only we sterilized all organisms with non-neutral mutations in order to ensure artificially that no true selection occurred. We further enhanced the drift in set V, by using the same procedure as in set IV, but allowing the final populations to evolve for 100,000 updates (5 times the default).

Unless otherwise specified, all tree branches generated in Avida were allowed to progress for 20,000 updates with a mutation rate of 0.0075 per site, which we expect is more than enough time for all neutral sites to be randomized (that is, for the molecular clock to run out). This duration is equivalent to approximately 2,000 generations, and by the end of it, each site only has a $(1-0.0075)^{2000} = 2.89 \times 10^{-7}$ probability of not being mutated. We followed this up by testing a control version of Avida wherein all organisms were guaranteed to replicate every 10 updates no matter what their genome sequence. In such a random system, if all the sites that are similar between two organisms that are paired together by neighbor joining are different from the corresponding site in their common ancestor, in a kind of *homoplasy* of neutral sites, then it is purely by chance that they were paired together. Maximum parsimony is basically useless in these situations, because of the absence of informative sites.

To reconstruct the phylogenetic trees produced by Avida we used *Phylip*, which is a free software package of many different phylogeny programs distributed by Felsenstein [4]. We used the program `neighbor` to do the neighbor joining, and the program `dnapars` to obtain maximum parsimony. Since the genome of the Avida organisms contains 26 symbols, it was impossible to use `dnapars` in the form in which it was distributed. We modified it so that it could be used with a genome comprising every letter in the alphabet. To generate the distance matrices required for `neighbor` we used a very simple Python script.

3 Discussion

Under the hypothesis of neutral evolution it is assumed that within the genetic codes of organisms there are a number of sites that are not subject to selection. These *neutral* sites can be thought of as "junk" DNA, which is present in the genomes of many organisms. Mutations accumulate at these sites without the purging force of selection, allowing them to act as the molecular clock that most reconstruction methods rely on to determine how long ago the lineages of two species had diverged. However, given sufficient time, neutral sites will become saturated with mutations and the molecular clock runs out, indicating a maximal separation. Given additional time, new mutations only occur at sites of older mutations and are therefore uninformative. The phylogenies we are studying in this article were run for a sufficiently long period of time that we expect this maximum separation. In the absence of selection or a molecular clock, we expect reconstruction to be relatively unsuccessful.

When an environmental change occurs after a branch point, the organisms should adapt to that change, creating fundamental differences in their genomes that will be preserved in this new environment. All organisms that are born within one of these branches should have a higher similarity due to this new genomic structure. Essentially,

this creates two monophyletic groups, which by Hennig's principle contain organisms that share traits that were *uniquely derived* from their common ancestor [6].

It is rare for digital life forms [11], and probably biological ones as well [7], to reach the same genotypic solution twice. Convergent evolution is phenotypic, not genotypic. Thus all of the organisms that should be grouped together would have adapted in the same environment, and almost by necessity in the same manner. Organisms in the other branch, although adapting to the same environment, should have done so differently from the first pair. If one of these branches progresses for a long time with many subsequent branch points, there may be no similarity at the neutral sites, but the phylogenetic tree reconstruction methods would have enough information to correctly resolve this branch point on the tree.

4 Results

In set I we expected minimal additional selection, which would translate to a $\frac{1}{3}$ success rate in tree reconstruction. Maximum parsimony correctly reconstructed three out of five phylogenetic trees. Neighbor joining was perfect in these runs. The distance matrix for each run is an indication of how closely the actual tree matched the data that neighbor joining saw. In every single run, we found at least one organism that failed to be paired with its closest partner in terms of Hamming distance. Run 1 was an especially close example, where, as is shown in Table 2, organism1-2 and organism2-1 are slightly closer than any of the other pairs of organisms. This situation was ambiguous, and maximum parsimony failed by pairing together organism1-2 and organism2-1. In the other run in which parsimony failed, the distance data is more clear cut in favor of the correct tree, although in all cases the differences in the distance matrices were slight.

In spite of our expectation that our ancestor was near a fitness optimum, set I displayed minor fitness changes, which we believe accounts for the better than random performance of the reconstruction algorithms. Set IV was identical to set I, except that beneficial and detrimental mutations were explicitly removed. Under this condition, maximum parsimony again correctly reconstructed three out of five trees, while neighbor joining did marginally worse than in set I, reconstructing only four out of five trees. Set V expanded the duration of the random walk to 100,000 updates, making the drift more significant. In this case, maximum parsimony correctly reconstructed two out of five trees, and neighbor joining, three out of five.

Set II had the organisms subjected to a new environment after the first branch point. In this set of runs all of the methods were successful in every instance. The distance matrix data also tells this story compellingly, for the organisms that should have been paired together are substantially closer in terms of actual distance than in the first set of runs. Furthermore, the distance between two organisms that should not have been paired together in the second run is comparable to the distance between two organisms in the first run, while the distance between two that should have paired together is definitely not. This is shown in Table 3.

Table 2. An example distance matrix (Hamming distance) taken from a run in set I in which there should have been no additional natural selection after the first branch point. All values are very similar, making reconstruction difficult.

Organism	Organism1-1	Organism1-2	Organism2-1	Organism2-2
Organism1-1	0	57	55	58
Organism1-2	57	0	56	59
Organism2-1	55	56	0	57
Organism2-2	58	59	57	0

Table 3. An example distance matrix (Hamming distance) taken from a run in set II in which additional natural selection occurred right from after the first branch point. The much larger range in values significantly improves the probability of reconstruction.

Organism	Organism1-1	Organism1-2	Organism2-1	Organism2-2
Organism1-1	0	52	66	70
Organism1-2	52	0	62	69
Organism2-1	55	56	0	57
Organism2-2	70	69	55	0

In set III, the environmental change occurred after all of the divergences had already happened, and both methods succeeded three times and failed twice. In these runs, the distances we observed in the distance matrix are more spread out. The same result was also observed in the variation in which all beneficial and deleterious mutations were banned from the runs (before the final branching, obviously). In the completely random runs, neighbor joining was successful once and maximum parsimony could not be used due to a complete lack of informative sites.

5 Conclusions

Trees in set II were reconstructed accurately as expected, which is encouraging for phylogenetic tree reconstruction methods in general. It is important to note that in these runs adaptation was not always completed when the organisms branched for the final time. Thus, the fact that the organisms evolved together forced their genomes to have a structure that was similar enough so that whatever adaptations they made in the future would not pull them apart too far in terms of genetic distance. Most of their traits were derived from traits in the common ancestor. This is analogous to situations involving convergent evolution.

Since whenever selection separated the pair (organism1-1, organism1-2) from (organism2-1, organism2-2) the phylogenetic tree reconstruction programs were perfect, it seems safe to say that these trees would have been safely reconstructed even with extreme selection at the ends of each run. It has always been believed that using genetic data would allow for the detection and elimination of taxonomic errors that were based on convergent evolution, but this was often with the assumption that the time between the divergences of the organisms was small enough to allow for large statistical similarities between the parts of the genomes that had not been selected for. This assumption certainly makes sense when working with animals and higher organisms, but it seems that it could break down with organisms that reproduce and mutate faster, such as viruses and bacteria.

In all sets of runs where there was little or no adaptation, the phylogenetic tree reconstruction methods were not always perfect. In set I, where evolution was not strictly prevented, neighbor joining was able to reconstruct every tree. However, maximum parsimony was not able to do so (possibly due to a lack of informative sites). Even in the case of neighbor joining, the distance matrices did not clearly favor the correct phylogeny. We believe that in a more complex evolutionary situation these effects could fool neighbor joining.

In set III the environment was only changed after the final branching, and hence there was no adaptive differentiation between the two halves of the tree. Consequently, the methods had only limited success. Of course, one could have reasonably expected that they would perform no better than random in this situation.

A possible explanation for the partial reconstruction success in sets I and III is that epistatic interactions between mutations played a role in determining which neutral

mutations were possible, and hence could have caused correlations between sites, dramatically slowing down the molecular clock. There could have been enough neutral drift of each half of the tree away from the other half to make the final adaptation of the four organisms to their environments biased, that is, organism1-1 and organism1-2 would each find a solution more similar to the other's solution than to the solutions found by the organisms on the wrong half of the tree.

To explore the unexpected success of the methods in set III, we repeated the single run that led to organism1-1 in run 4 ten times, and computed the average distance from the resulting organism to organism1-2, organism2-1, and organism 2-2. For organism1-2 the average distance was 62.9 with a standard deviation of 3.24. The mean distance from organism2-1 was 65.2 with a standard deviation of 3.55. Finally, the distance from organism2-2 is 69.6 with a standard deviation of 1.91. However, it should be noted that in eight of the ten cases organism1-2 was closer than organism2-1, as might be expected. Furthermore, it is unclear how the statistics of this type of adaptive process should be computed. The results do indicate that somehow just the fact that organism1-1 starts closer to organism1-2 makes them diverge less in evolutionary distance, even though a matching site may contain a different instruction than the one found in the common ancestor. This strengthens the idea that epistatic effects limit the particular neutral mutations possible and must be taken into account when calculating the time needed for the molecular clock to run out.

Since it is possible to calculate the expected differences between two pieces of genetic code that are neutrally evolving at a certain mutation rate, it should be possible to determine to an order of magnitude what sort of time scale is involved for certain organisms. For a particular organism, let μ be the mutation rate, and N be the dissimilarity (in percent) between two sequences. We can say that for DNA-based life, the expected change in dissimilarity per generation is approximately (using the Jukes-Cantor model [10])

$$\frac{dN}{dg} = \mu\left(1 - \frac{4N}{3}\right). \tag{1}$$

Thus the number of generations that can elapse before we can expect the dissimilarity to be N is

$$g = \frac{-3\ln(1 - 4N/3)}{4\mu}. \tag{2}$$

At this point the question is one of determining a suitable N to produce problems for the reconstruction methods. Since the number of generations is based on the natural log of $1-4N/3$, making N very close to its maximum value of $\frac{3}{4}$ does not significantly affect the amount of time that this approximate calculation will yield. Thus we allow N to be $\frac{3}{4} - 1 \times 10^{-7}$. This allows us to create Table 4, which gives values of the randomization time for familiar organisms. We can also provide this data for an Avidian, for which we modify the calculation as

$$\frac{dN}{dg} = \mu\left(1 - \frac{26N}{25}\right). \tag{3}$$

This leads to a result analogous to the previous one for DNA-based organisms:

$$g = \frac{-25\ln(1 - 26N/25)}{26\mu}. \tag{4}$$

Table 4. A selection of example mutation rates and saturation times calculated using Equation 2 for some natural organisms [1], compared to the saturation time for Avidians, obtained with Equation 4.

Organism	μ (site^{-1}generation^{-1}) [1]	Genome (bp)	Saturation (years)
E. coli	$\sim 5 \times 10^{-10}$	5×10^{6}	$\sim 6.5 \times 10^{7}$
HIV	$\sim 2 \times 10^{-5}$	10,000	$\sim$ 1,500
Influenza A	$\sim 1 \times 10^{-4}$	12,000	$\sim$ 300
H. sapiens	$\sim 5 \times 10^{-11}$	8×10^{10}	$\sim 5 \times 10^{12}$
Standard Avidian	0.0075	100	$\sim$ 900 (generations)

However, one great source of debate among those who study molecular clocks is that for mammals it appears as though the rate of divergence is independent of the generation time of the organism. Of course this type of calculation has less applicability to higher organisms. It is evident from Table 4 that concerns about the effects of selection on the tree reconstruction methods should be much more prevalent in the case of organisms with high mutation rates and generation times than in more complex organisms like humans, where the statistical similarities approach should be fairly accurate. From a normative standpoint, care must be taken when working with organisms that might violate the statistical assumptions that are sometimes made in these situations.

A final curiosity is the relative success of the neighbor joining method in reconstructing the phylogenetic trees. It is a widely held notion that maximum parsimony and maximum likelihood are more accurate methods than neighbor joining. However, the data used to demonstrate this difference between methods tends not to based on real phylogenies, only on purely statistical ones. Although it is the case that the experiments in this study were not specifically designed to cause problems for neighbor joining, it is still worthwhile to point out that it appeared to be more effective than maximum parsimony. One should be cautious however, because the large number of characters in an Avidian genome in combination with the small number of organisms in a phylogenetic tree restricted the number of informative sites available to parsimony. However, it is encouraging that the two methods can be even comparable, since neighbor joining is almost immeasurably faster for reconstructing phylogenetic trees, and in many cases is the only practical method possible for use in real biological problems.

Acknowledgments

We thank C. Adami, D. A. Drummond, R.E. Lenski, D. Wagenaar, and C.O. Wilke for discussions. This work was supported in part by the National Science Foundation's Biocomplexity Program under Contract No. DEB-9981397. All data was obtained on a cluster of 160 processors at Michigan State University, supported by the NSF under the aforementioned contract. GIH acknowledges a Caltech SURF fellowship.

References

1. Drake, J. W., Charlesworth, B. C. D., & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics*, *148*, 1667–1686.
2. Durbin, R., Eddy, S., Krogh, A., & Mitchison, G. (1998). *Biological Sequence Analysis*. Cambridge, UK: Cambridge University Press.
3. Felsenstein, J. (1988). Phylogenies from molecular sequences: Inference and reliability. *Annual Review of Genetics*, *22*, 521–565.
4. Felsenstein, J. (1989). PHYLIP—Phylogeny inference package (version 3.2). *Cladistics*, *5*, 164–166.
5. Felsenstein, J. (2004). *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates.

6. Futuyma, D. (1998). *Evolutionary biology*. Sunderland, MA: Sinauer Associates.

7. Gould, S. J. (1998). *Wonderful life: The Burgess Shale and the nature of history*. New York: Norton.

8. Hillis, D. M. (1995). Approaches for assessing phylogenetic accuracy. *Systematic Biology*, *44*, 3–16.

9. Hillis, D. M., Bull, J. J., White, M. E., Badgett, M., & Molineux, I. J. (1992). Experimental generation of a known phylogeny. *Science*, *255*, 589–591.

10. Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules. In Munro, H. N. (Ed.), *Mammalian protein metabolism* (pp. 21–132). New York: Academic Press.

11. Lenski, R. E., Ofria, C., Pennock, R. T., & Adami, C. (2003). The evolutionary origin of complex features. *Nature*, *423*, 139–144.

12. Nakhleh, L., Moret, B. M. E., Roshan, U., St. John, K., Sun, J., & Warnow, T. (2002). The accuracy of fast phylogenetic methods for large datasets. In *Pacific Symposium on Biocomputing 2002* (pp. 211–222).

13. Ofria, C., & Wilke, C. O. (2004). Avida: A software platform for research in computational evolutionary biology. *Artificial Life*, *10*, 191–229.

14. Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*, 406–425.

15. Wilke, C. O., & Adami, C. (2002). The biology of digital organisms. *Trends in Ecology & Evolution*, *17*, 528–532.