

A PROPOSAL TO DO RESEARCH ON VISUALIZATION OF A PHYLOGENETIC TREE

By Kathleen Timmerman

1. INTRODUCTION

1.1 DESCRIPTION

For my research project, I will research and implement a three dimensional phylogenetic tree structure. Due to the large amount of data involved in phylogeny, the implementation will need to be optimized for speed. It will also need to have a flexible viewing of the tree, so that users can focus on what is most important to them. I will begin the project at viewing current visualization programs for phylogenetic trees and determining their limitations. I will also review several programming languages that cater to graphical user interfaces and determine which one best fulfills the needs of this project. After that I will implement a new program that displays a phylogenetic tree in three dimensions.

1.2 PURPOSE

I propose to research, implement, and produce a formal report on the visualization of phylogenetic trees to fulfill the requirement of the Distributed Research Experience for Undergraduates (DREU) and the Texas A&M Summer Undergraduates Research for Engineers (SURE), Summer 2009.

This research topic will allow me to gain a greater understand of a branch of Bioinformatics that I had limited prior knowledge of: phylogeny. I have an interest in pursuing a graduate degree in Bioinformatics, but have not chosen an area to specialize in at this point in time. This project will give me a chance to discover another area while at the same time building upon the programming skills that I already have obtained: a strong programming background and a basic understanding of phylogeny. I have not programmed graphical user interface before, but it has been something that I have been interested in attempting. For all these reasons, I believe this is a good project for myself.

1.3 PROJECT GOALS AND IMPLICATIONS:

A large issue in Bioinformatics is overcoming the communication barrier between computer scientists and biologists. A user interface that appears simple and straight forward to a computer scientist might be confusing and awkward to a biologist. A goal of this project is to make a user interface that is comfortable for both computer scientists and biologists.

A goal of this project is to better represent the relationships between taxa by creating the visualization of the phylogenetic trees in three dimensions. In two dimensions it is easy to give the illusion that two taxa are farther apart than they really are. In Figure 1.1 A and D have the same relationship as B and C, but due to the space differences between the two sets this is not intuitive. By using a three dimensional representation, issues such as this can be avoided.

By providing better representation of the phylogenetic tree it is hoped that new relationships can be discovered more easily than they currently are.

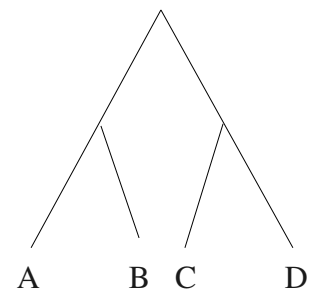


Figure 1.1 Basic 2D Tree

1.4 PERSONAL GOALS

I hope that this research project will give me a better understanding of what graduate school will be like and whether or not graduate school is something that I am interested in pursuing.

I am interested in discovering a different area of Bioinformatics. Previously most of my experience in Bioinformatics has dealt with sequence alignment issues. Another goal is to discover if I am more interested in genetics or the specie part of Bioinformatics.

I am also interested in completing a large project. During many classes, the programming assignments tend to be small, quick programs with limited data. Optimization tends to not matter. It will be a challenge to have to consider efficiency and memory while programming.

Finally, I am interested in getting the input from the biologists. A large part of my experience with Bioinformatics has been focused solely on the computer science perspective. I am looking forward to finding out how biologists view of current phylogenetic tree viewing programs differences from computer scientists, as well as dealing with the language barrier that can exists when people from different backgrounds communicate.

2. PROGRAM OUTLINE

2.1 APPROACH

Step	Task	Hours
1.	Research current phylogenetic trees and determine their limitations and advantages. Talk to biologists to get their perspective.	30
2.	Research different programming languages and determine which one would be best suited to this project.	20
3.	Outline how the program will be implemented.	12
4.	Determine any problematic areas.	8
5.	Adjust original outline to account for problem areas.	12
6.	Implement the program.	150
7.	Test the Program.	See 9
8.	Debug the program.	See 9
9.	Repeat steps 7-8 until program runs without issues.	75
10.	Complete any needed documentation for the program that was not created during the course of writing the program.	16
11.	Complete poster and Presentation	8
12.	Complete report	16
13.	Complete Website	3
	Total Hours:	350

2.2 APPROXIMATION OF WEEKLY SCHEDULE

<u>Week End</u>	<u>Steps Completed*</u>
End of Week 2	Steps 1 – 2 Completed
End of Week 3	Steps 3 – 5 Completed
End of Week 7	Step 6 & 11 Completed with some testing & debugging
End of Week 9	Steps 7 – 9 Completed
End of Week 10	Steps 10 & 12 – 13 Completed

*For steps please refer to section 2.1 Approach

The poster was moved up to week 7 in order to participate in SURE activities.

2.2 WORK SCHEDULE

My work schedule is as follows:

Monday – Friday 9:30 – 6:00 with half hour lunch.

That time will be spent in my office.

This will be adjusted slightly the weeks ending July 25 and August 1 due to being out of town for my sisters wedding. I will either work longer days the days I am here or work from home to ensure that the 40 hour criteria is met.

3. RESOURCES

3.1 METHODS AND MATERIALS

The university has provided everything that I will need for this project. The programs that I intend to use are open source and just need to be downloaded. The printing in Lab HRBB 210 should be adequate for the documentation and reports needed.

4. DELIVERABLES

4.1 ITEMS AND DATES

The following Items will be delivered on the dates specified below:

1. Research Abstract July 22, 2009 by Midnight
2. Research Poster August 3, 2009 by 8:00am
3. Poster Presentation August 4, 2009 at 4:00pm
4. Poster Presentation August 9, 2009 at 9:00am
5. Final Website August 19, 2009 by midnight
6. Research Report August 19, 2009 by midnight

5. CONCLUSION

5.1 SUMMARY

DREU and SURE require research to be completed and documented in the form of both a website and a final report. SURE also requires that a poster be created and presented. The create of a phylogenetic tree program will require me to do research in order to improve upon already existing programs. Off of the research and implementation that I do, I will be able to fulfill the requirements of DREU and SURE, Summer 2009.

5.2 CONTACT

For more information, please e-mail me at timmerman.16@tamu.edu, or phone 937-726-6717.

6. FACUALTY ADDITIONS FROM TIFFANI WILLIAMS

6.1 PROJECT GOALS

My expectations for this project is that life scientists will have a new way to view their evolutionary trees. At the end of the 10 week research period, there will a new application to view an evolutionary tree in three-dimensional space.

6.2 STUDENT MENTEE GOALS

My expectations for Katie include: (i) developing a three-dimensional representation for viewing evolutionary trees (ii) talking to biologists regarding their needs for viewing evolutionary trees and incorporating their needs into the design of the software (iii) writing up the results of her work for scientific publication, and most importantly, (iv) having fun.