



University
of Glasgow | School of
Computing Science

The effects of natural selection on the shapes of ancestry graphs

Adam Kurkiewicz

School of Computing Science
Sir Alwyn Williams Building
University of Glasgow
G12 8QQ

Level 4 Project — March 24, 2016

Abstract

Placeholder for the abstract, which will be written at the end.

Education Use Consent

I hereby give my permission for this project to be shown to other University of Glasgow students and to be distributed in an electronic format. **Please note that you are under no obligation to sign this declaration, but doing so would help future students.**

Name: _____ Signature: _____

Contents

1	Introduction	1
1.1	Context of the work	1
1.2	Research question	1
2	Review of the field	3
3	Report of research work	4
4	Report of practical work	5
5	Conclusion	6
6	References	7

Chapter 1

Introduction

Introduction. This should place the work in context and will be developed from the introduction produced for assessment.

1.1 Context of the work

The publication of "On the Origin of Species" is generally accepted to be a founding stone of still vigorously researched field of evolutionary biology. Over the years, as the understanding of the field deepened it became clear that certain scientific questions within the field can only be answered by obtaining progress in related fields, like establishing laws governing inheritance (genetics) [cite something from mendel or equivalent], understanding the role of organic molecules in the process of inheritance (molecular biology) [cite watson and crick's paper], understanding population dynamics of predator-prey system using both continuous and discrete methods (mathematical biology), understanding the character of population variability at molecular level (neutral theory of molecular evolution, Markov processes) [cite the controversial paper on neutral theory of molecular evolution, and PAM matrices], studying of historical populations (archaeology, anthropology, linguistics) [maybe java man?], proposing and verifying hypotheses about populations based on quantitative evidence (statistics) [bronze age Eurasia and steppe hypothesis], modeling and simulating biological systems *in silico* (computational biology), and many others.

While the research presented in this work falls within the realm of computational biology, one should understand the difficulty of both undertaking the research in the field and presenting the results, namely, the broadness of topics and tools involved, which need to be understood and taken into account if the research is to be of any significance. A part of the struggle in communication is one of vocabulary, as each of the related fields uses specialized jargon, and researches with expertise in one field might not understand the other fields' jargon. To ease this difficulty in the beginning of each chapter we shall include a glossary of terminology in use.

1.2 Research question

The scientific question that we will try to answer in this piece of work is directly connected with evolution by natural selection and perhaps might have even occurred to Darwin himself:

Research Question (First Attempt). *How much time does it take for an advantageous feature in an individual to spread across entire population?*

As we will approach the question on the grounds of computational biology, it is perhaps useful to rephrase it in a language more familiar to computer science, which will also serve us the purpose of disambiguating the question's meaning. Firstly, let us introduce the notion of the **ancestry graph**, which we will understand to be a directed acyclic graph, with vertices representing all individuals ever alive and directed edges representing the parent-child relationship between individuals. Additionally, each vertex will be labeled with the feature-status (present, not present) of the individual it represents. For any point in time all individuals alive at that point will be referred to as "population at time t ", or simply "population", if the time can be inferred from the context. Representing individuals in form of a graph has an advantage of allowing us to use a rich set of tools for analyzing the shape of the ancestry graph, such as graph centrality measures, or even simply most recent common ancestry of sets of vertices.

Finally, this allows us to reword the research question:

Research Question (Reworded). *How does natural selection affect the shape of the ancestry graph?*

Even after this linguistic juggling the question remains pretty vague. We never specify any parameters of population at any time t , such as the mating scheme, migration, competition for resource, survivability, number of offspring, etc. These parameters will be discussed in chapter 2, after we will have decided on appropriate methodology in chapter 1. We will discuss the architecture of the computer program which constitutes the solution, and is on its own a valuable contribution to the field in chapter 3. The results will be presented in chapter 2 and the conclusions will be drawn in chapter 4.

Chapter 2

Review of the field

This should be a critical survey of the relevant literature, adhering to normal academic conventions in citing references, etc.

Chapter 3

Report of research work

This should bring out the insights and new ideas produced by the review and may continue to suggest further practical work, which would test these ideas. Most importantly, the report should clearly state the overall value of the work which has been done - this should not be lost in a welter of detail.

Chapter 4

Report of practical work

This section should summarize the practical work done and its results, concentrating on any unusual or original features. Enough detail should be included to assure the reader that the work has been professionally and competently done, and that the results are trustworthy. Any design features, algorithms and data structures of special interest should be described. Documentation to full software engineering standards is not required.

Chapter 5

Conclusion

This section should summarize the work and its significance, comparing practical results with theoretical expectations where appropriate. Suggestions for possible extensions, or research projects arising from the work, may also be included here.

Chapter 6

References

A complete ordered list of any references which have been cited or consulted. Appendices. Full listings of software written, with selected test results, should be included in the accompanying CD. A summary log should also be appended, recording significant events in the progress of the project.