UNIVERSITY OF CALIFORNIA

Los Angeles

A New Computational Approach to the Ancient Greek Dialects: Phylogenetic Systematics

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Indo-European Studies

by

Christina Skelton

2014

ABSTRACT OF THE DISSERTATION

A New Computational Approach to the Ancient Greek Dialects: Phylogenetic Systematics

by

Christina Skelton

Doctor of Philosophy in Indo-European Studies

University of California, Los Angeles, 2014

Professor Brent Vine, Chair

Phylogenetic systematics, first used for reconstructing biological evolution, has become popular in historical linguistics for reconstructing the development of language families. This dissertation tests three approaches to phylogenetic analysis of the Ancient Greek dialects in order to determine which one best handles borrowing. For character weighting, only reweighting characters according to their CI (consistency index) improved the resolution of the final tree while still giving a plausible tree topology. NeighborNet captured the basic tree topology, but was not able to capture certain important types of borrowing. Combining taxa through a preliminary cluster analysis also produced a single tree with a plausible tree topology. However, all of these methods fail to capture a circular dialect continuum among the Greek dialects, indicating that tree methods alone cannot capture some important types of linguistic development.

The dissertation of Christina Skelton is approved.

H. Craig Melchert

Edward Stabler

Johanna Nichols

Brent Vine, Committee Chair

University of California, Los Angeles

2014

DEDICATION

To my unsung hero-mentors:

Marc Musick, Richard Meier, and Randy Diehl

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First, I would like to thank the people and organizations who helped with my dissertation work. My committee members, Brent Vine, Craig Melchert, Ed Stabler, and Johanna Nichols, provided helpful comments on my dissertation, and discussions with Richard Meier and Randy Diehl helped shape my vision of what the dissertation should be. The library at the Center for Hellenic Studies provided the best possible venue for my research, and I would like to thank everyone there, particularly Gregory Nagy, for their support and hospitality. Andrew Garrett made useful suggestions about my phylogenetic data matrix. Wally Gilbert first suggested to me the idea of cluster analysis. Matt Elkins, Rebecca Hale, and Matt Bufano helped with the coding. Rebecca Hale and Anna Pagé helped with the formatting, and Sherrylyn Branchaw helped with the proof-reading.

Funding for my graduate studies and for my dissertation research was provided by a National Science Foundation Graduate Research Fellowship, grant number DGE-0707424 (2009-2011), a UCLA Graduate Research Mentorship (2012), and a UCLA Dissertation Year Fellowship (2013).

I would also like to thank the people who did not help with the dissertation *per se*, but who made life much better while I was writing it. They include my parents, Mike and Charleen Skelton; my most excellent boyfriend, Matt Bufano; Éloïse Lemay, for her tireless support and for letting me stay at her place for extended periods of time; and Caitlin Caven and Chris Karper for being around to talk to during the day when I was lonely, which was often. I would also like to thank my other friends and mentors for their support: Jana Beck, Will Bibee, Christopher Colvin and Behice Kutay, Jessica DeLisi, Randy Diehl, Matt and Liliya Elkins, Female Science Professor, Mattyas Huggard, Joshua Katz, Josh and Hilah Kaufman, Bob King, Josh Lamiano and Megan Lai, Emilia Mahaffey, Chris and Christine McCann, Richard Meier, Frank Mikan, Abby Mooney, Anna Pagé, Na'ama Pat-El, Rick Pearson, Lydia Rieck, Adam and Laura Strickhart, Sally Thomason, Elizabeth Thornton, Patrick Wilson and Jessica Clark, and Cari Yavorcik and Adam Schenker.

Most of all, I would like to thank Marc Musick, because I never would have made it through graduate school without him.

VITA

Education

C.Phil. Indo-European Studies, December 2012, University of California, Los Angeles

Advisor: Prof. Brent Vine

B.A. Classics, with highest honors, August 2008, University of Texas at Austin

Minors: *Geology* and *Biology*

Advisors: Prof. Thomas G. Palaima (Classics), Prof. William D. Carlson (Geology)

Selected Additional Education

Visiting Graduate Student Researcher, Danish National Research Foundation

Center for Textile Research, July-August 2011. Advisor: Prof. Marie-Louise Nosch

Predoctoral Fellow, Max Planck Institute for the History of Science, Summer 2008. Advisor: Prof. Peter Damerow

Iklaina Archaeological Project, June 27-July 7, 2011

Archaeological Field School at Petsas House, Summer 2005

Graduate Student Seminar on Greek Dialects and Greek Historical Linguistics, Harvard University Center for Hellenic Studies

Selected Publications

In Press. with Richard Firth, “A study of the scribal hands of Knossos based on phylogenetic methods and find-place analysis,” *Minos*.

2011. “A Look at Early Mycenaean Textile Administration in the Pylos Megaron Tablets,” *Kadmos* 50, 101-121.

2009. “Re-Examining the Pylos Megaron Tablets,” *Kadmos* 48 (2009), 107-123.

2008. “Methods of Using Phylogenetic Systematics to Reconstruct the History of the Linear B Script,” *Archaeometry* 50, 158-176.

Selected Presentations

2013. “A Phylogenetic Analysis of the Ancient Greek Dialects,” at the International Conference on Historical Linguistics, University of Oslo, Norway, August 5-9.

2012. “Thoughts on the Initial Aspiration of HAKESANDRO,” at the conference “Panhellenes at Methone: Geometric and Protoarchaic Methone, Macedonia (ca 700 BCE),” Thessaloniki, Greece, June 8-10.

2012. “Spinning Fate– from Anatolia to Greece,” at the roundtable “Spinning Fates and the Song of the Loom,” Danish National Research Foundation’s Center for Textile Research, Copenhagen, Denmark, June 1-2.

2009. “How Did the World’s First Writing Evolve?” at the NSF IGERT PI meeting, Alexandria, Virginia, May 17th-19th. (One of ten winners for Best Poster).

Selected Honors, Awards, and Fellowships

2013. *University of California, Los Angeles* Dissertation Year Fellowship

2012. *University of California, Los Angeles* Graduate Research Mentorship

2009-2011. *National Science Foundation* Graduate Research Fellowship

2007. *George H. Mitchell Award for Academic Excellence*, $20,000 Award Winner

Chapter 1: Background

Overview

Phylogenetic systematics, the set of methods originally developed in the biological sciences for reconstructing the evolutionary histories of groups of organisms, is an increasingly popular tool in historical linguistics for studying the development of language families and dialect groups (e.g. Nichols and Warnow 2008, Forster and Renfrew 2006). This dissertation presents a phylogenetic analysis of the Greek dialects, or, rather, a series of phylogenetic analyses designed to address the major methodological problems a phylogenetic analysis of the Greek dialects would encounter. The Greek dialects provide the perfect opportunity for exploring some unresolved theoretical issues with the use of phylogenetic systematics in historical linguistics, and for providing new insight into the Greek dialects, in turn. Phylogenetic tree methods assume that evolution was perfectly treelike, with no transfer of information between unrelated entities. However, there are a variety of means by which linguistic features can be transferred from between unrelated languages, and they are very common and widespread. The Greek dialects show a variety of different types of non-treelike evolution, which has not yet been studied in detail. Thus, the Greek dialects provide an excellent opportunity to test how well different proposed phylogenetic methods for handling nontreelike evolution actually perform, as well as the opportunity to gain a more comprehensive understanding of nontreelike evolution in the Greek dialects. At the same time, phylogenetic systematics can shed additional light on the remaining controversial issues in the treelike evolution of ancient Greek.

Introduction to the Greek Dialects

The relationships among the Greek dialects and the history of the various dialects and dialect groups have been a source of curiosity since antiquity, and our understanding of the Greek dialects has continued to shape our understanding of history and historical linguistics.

For the ancient Greeks, dialect helped shape their conception of ethnic identity. In turn, this sense of ethnic identity shaped their understanding of the past through mythology and history, and continued to influence their perception of contemporary cultural, political, and religious ties (Colvin 2007, 21). For example, Hesiod attributes the Greeks' three-fold ethnic division between the Dorians, the Ionians, and the Aeolians to the three sons of Hellen (frag. 9, Merkelbach and West 1967):

Ἕλληνος δ᾽ ἐγένοντο φιλοπτολέμου βασιλῆος
Δῶρός τε Ξοῦθός τε καὶ Αἴολος ἱππιοχάρμης.

From Hellen, the king fond of battle, came
Doros and Xouthos and Aiolos, who loved horses.

Both Herodotos and Thucydides make several references to the Ionian, Dorian, and Aeolian ethnic groups when mentioning the settlement history of certain locations (Buck 1955, 3-6). Thucydides describes in detail how language and ethnic affiliation influenced different Greek city-states' decision to fight for or against Sicily. He pays close attention to whether each city-state fought against or alongside members of the same dialectal or ethnic group, and whether they were fighting for their side willingly or unwillingly (Thucydides 7.57). Strabo later expanded the Ionian, Dorian, and Aeolian ethnic groups into an all-inclusive classification scheme, in which all peoples who were not Ionians or Dorians were classified as Aeolians (Buck 1955, 6). On the other hand, grammarians such as Apollonius Dyscolus, Philoxenus, and Trypho used a dialectal classification based on the literary dialects which recognized the existence of a number of independent dialects, including Ionic, Attic, Aeolic (which referred exclusively to Lesbian), Boeotian, Doric, and *koine*. Doric could then be subdivided into local varieties, such as Syracusan, Argolic, or Laconian (Cassio 2007, 29-30).

These conceptions of Greek language and dialect began at least by the fifth century BCE, when we start to find evidence that the Greeks considered Greek a single language, but were aware of dialectal differences and the fact that they could make use of different dialects according to context, such as to show (or fake) ethnic affiliation or to follow the conventions of a literary genre (Morpurgo Davies 1987).

The discovery of Proto-Indo-European provided a new impetus for the linguistic study of ancient Greek. The Greek language provides an invaluable perspective on Indo-European reconstruction for many reasons, for example because it preserves the original Proto-Indo-European accent in many cases, because the original vowel qualities and stop system remain mostly intact, because each laryngeal shows a different reflex, and because the original aspectual distinctions in the verbal system are sometimes observed. The historical Greek dialects supply information for linguistic reconstruction that cannot be gained from Attic and Ionic alone. For example, Thessalian, Boeotian, Arcadian, and West Greek preserve /w/ in many instances, although /w/ was completely lost in Attic and Ionic by the historical period. Dialectal forms indicate the presence of original labiovelars before front vowels, thanks to different outcomes in Arcado-Cypriot, Aeolic, and West Greek and Attic-Ionic. Finally, dialectal inscriptions can preserve archaic forms, like the Arcadian optative form ἐξελάυνοια (Dubois 1988, 169-179). The verbal person/number ending –α is the reflex of the original first person singular secondary ending *–*m*, which Greek has generally replaced with the primary ending –μι in this particular optative category.

In 1951, the decipherment of the Linear B writing system revealed a new dialect of Greek, Mycenaean, which was spoken ca. 1400-1200 BCE, over half a millennium earlier than the earliest attested alphabetic inscriptions. The linguistic world of Mycenaean Greek came as a surprise. Mycenaean was most closely related to Arcadian and Cypriot, which only existed in remote, marginal areas in historical times. Yet, Mycenaean appeared to have been spoken widely over the Bronze Age world, at Knossos, Khania, Mycenae, Tiryns, and Pylos, where Doric dialects were spoken in historical times, and at Thebes, where Boeotian, an Aeolic dialect, was spoken in historical times. How the Greek dialects came to have their historical geographical distribution is a major unanswered question (Parker 2008). At the same time, the early date of Mycenaean has allowed us to establish the chronology of many of the changes which distinguish the Greek dialects from Proto-Indo-European. This new knowledge has led to a revised view of the early stages in the development of the major Indo-European language groups, perhaps a model for the early stages of differentiation in other language families (Garrett 2006).

Thus, understanding the Greek dialects clearly has far-reaching implications for historical linguistics, Indo-European linguistics, and Greek history and historiography. On the other hand, when it comes to understanding the Greek dialects on their own terms, the sorts of questions that ancient and modern scholars have sought to answer generally address one of two major issues. First, what are the major subgroups of the Greek dialects? In antiquity, scholars recognized four major dialects, Attic, Ionic, Doric, and Aeolic. Scholars of modern times have combined Attic and Ionic into a single dialect group, Attic-Ionic, and added another dialect group, Arcado-Cypriot, which came to be known as Achaean with the discovery and addition of Mycenaean. Doric has been combined with the Northwest Greek subgroup, consisting of Locrian, Phocian, and Elean, to form West Greek. Scholars have also recognized major disparities between the Aeolic dialects. Boeotian and Thessalian seem to have strong affinities with West Greek, while Lesbian seems to have strong affinities with Attic-Ionic. The debate as to whether Aeolic forms a true linguistic unity is still ongoing (e.g. Parker 2008).

Second, how did Proto-Greek develop into the major Greek dialect groups? Prior to 1955, it was thought that the first major split Proto-Greek underwent divided Proto West Greek from Proto East Greek, which later developed into Aeolic, Attic-Ionic, and Achaean. However, Risch argued that the first major split in Proto-Greek divided Proto North Greek, which developed into West Greek and Aeolic, from Proto South Greek, which developed into Achaean and Attic-Ionic (Risch 1955). The discovery of Mycenaean also raised the question of the geographic distribution of the Greek dialect groups in the second millennium BCE, since Mycenaean was found at sites where dialects of other dialect groups were spoken in historical times. The state of the Greek dialects in the second millennium BCE is still controversial.

Although questions about the linguistic history of the Greek dialects have predominantly been framed in terms of these two questions, it is important to recognize that the discussions of these questions tend to incorporate arguments based on more or less the same two premises. The first is that the early development of the Greek dialects into four major dialect groups was, at some level, fundamentally treelike, and that these branching events can be uncovered through the comparative method. The second premise is that the development of the Greek dialects was not completely treelike, such as when dialects of different dialect groups came to be located next to one another and exchanged linguistic features, and so linguistic features shared between different dialect groups are the result of some process other than descent with modification. These processes include parallel development, borrowing from one dialect to another, diffusion of linguistic change across several dialects, which can in extreme cases lead to areal convergence, and the formation of new dialects through mixing.

The premise that the comparative method can establish the branching of the Greek dialects is clearly sound, and needs no further discussion here, except to note a few of the many excellent studies of the Greek dialects which make use of the comparative method (e.g. Risch 1955, and, more recently, the many fine papers in Hajnal and Meier-Brügger, eds. 2007). On the other hand, very little thought has been devoted to the need to systematically understand the effect of parallel evolution, borrowing, diffusion, areal convergence, and geography on the development of the Greek dialects. Instead, these processes tend to be invoked in an isolated or *ad hoc* manner. For example, borrowing, diffusion, and areal convergence are frequently invoked in an inconsistent and contradictory fashion, sometimes in the same work or even on the same page, to explain the large number of linguistic features which are shared between the Aeolic dialects and their neighbors:

> The Aeolic dialects bear witness to a relatively brief period of common development followed by a much longer process of areal convergence (Boeotian and Thessalian have features in common with West Greek, Lesbian with East Greek). (Colvin 2007, 40)

> Boeotian is often described as a mixed dialect (West Greek and Aeolic), and in fact all three are in varying degrees fusions of disparate elements. (Colvin 2007, 40)

Even though these passage express the same fundamental idea, that the Aeolic dialects came to have linguistic features through a process other than descent with modification, and that these features came from neighboring dialects, the differing terminology implies different processes. The first quote cites areal diffusion as the means by which Boeotian and Thessalian came to have West Greek features, and Lesbian came to have East Greek features. This implies the existence of a defined geographic area which facilitated the spread of linguistic features. Some of these features may have been brought to the area by one or more of the dialects, while others may have arisen later within one of the dialects.

The second quote describes Boeotian as a "mixed dialect," which implies a very different process. If we accept Trudgill (2004)'s definition of mixing as "the coming together in a particular location of speakers of different dialects of the same language, or of readily mutually intelligible languages" (84), then a mixed dialect would be one that arose in a particular area through the juxtaposition of speakers of different dialects in that area. The process of dialect mixing, then, involves the creation of a single new homogeneous dialect by successive generations of child language learners (Trudgill 2004).

Thus, areal diffusion and dialect mixing involve very different geographic setups, namely two or more dialects with clearly defined boundaries versus two dialects juxtaposed in the same area. Areal diffusion and dialect mixing also involve very different processes of language change, namely diffusion of features presumably mediated by adult speakers versus the creation and normalization of a new dialect by child language learners. Therefore, it seems methodologically unsound to invoke them both as processes which led to the creation of the same dialect, for instance Boeotian, at least without further explanation as to what this situation entailed and which process affected which features.

Clearly, it is important to understand the markers of each of the processes involved in the transfer of linguistic features other than descent with modification, so that we may reach a true understanding of the dialects which seem to display extensive non-treelike evolution.

The lack of a systematic understanding of borrowing, diffusion, dialect mixing, and areal convergence in the Greek dialects does not necessarily invalidate existing analyses which invoke these processes. However, in order for these explanations to be compelling, we must develop a systematic understanding of how these processes shaped the history and development of the Greek dialects. This new understanding will also likely reveal the effects of these processes in previously unexplored areas, and open up new and productive avenues of investigation.

Introduction to Phylogenetic Systematics

Before the invention of phylogenetic systematics, it would have proved impossible to conduct such a detailed and comprehensive analysis of the Greek dialects, because it would have been, at best, too time-intensive, or, at worst, beyond the processing capabilities of the human mind. However, phylogenetic systematics, the method used in the biological sciences for reconstructing the evolutionary history of groups of organisms, can produce and analyze evolutionary trees with different types of phylogenetic analysis, as implemented in programs like PAUP* (Swofford 1998) and SplitsTree (Huson and Bryant 2006). Statistical software and even spreadsheets can be used for various types of probability and statistical analysis.

Phylogenetics has several major advantages over traditional historical linguistic methodology (cf. the discussion of phylogenetics versus traditional paleography in Skelton 2008). The data matrix that each phylogenetic analysis requires as input makes it clear exactly what data the analysis was based on. The phylogenetic method chosen, be it an algorithm or an optimality criterion and search strategy, makes it clear how the analysis arrived at an optimal tree or trees. Other analyses can be run after the fact to determine how well a given tree fits the data. Since the analysis is carried out by a computer, it is able to consider more data and run analyses faster than any human being ever could.

Overview of Phylogenetic Methods

Phylogenetic systematics encompasses a collection of methods, and choosing an appropriate phylogenetic method is essential for obtaining results which are reliable.

Several different classes of phylogenetic methods exist. Algorithm-based methods use an algorithm to build a tree from the data set. Examples include UPGMA (unweighted pair group method with arithmetic mean) (see Swofford et al. 1996, 486-487) and Neighbor Joining (Saitou and Nei 1987, Swofford et al. 1996, 488-490). Optimality criterion-based methods establish a standard for determining the best tree or trees, then generate a large number of possible trees and evaluate them against that standard. Examples include Maximum Parsimony (see Swofford et al. 1996, 415-430) and Maximum Compatibility (see Nakhleh et al. 2005b, 175-176). Statistical methods arrive at a tree by using an explicit statistical model of how evolution occurs. Examples include the methods developed in Nicholls and Gray (2008) and Gray and Atkinson (2003). All of the aforementioned are phylogenetic tree methods; their output is one or more strictly bifurcating trees.

Alongside phylogenetic tree methods, there exist phylogenetic network methods. These include both methods which produce explicit phylogenetic networks, like Perfect Phylogenetic Networks (Nakhleh et al. 2005b), where nodes represent ancestors and contact edges represent borrowing events, and methods which produce implicit phylogenetic networks, like Split Decomposition (Bandelt and Dress 1992) and NeighborNet (Bryant and Moulton 2002), where additional sets of parallel lines are introduced in order to represent both borrowing events and noise, and nodes do not represent ancestors (Nichols and Warnow 2008, 763-764).

Even though phylogenetic systematics has clear advantages, phylogenetic methods bring with them a number of unresolved theoretical issues. The set of issues that I aim to address in this dissertation pertain to non-treelike evolution, specifically the transfer of linguistic features between dialects, the linguistic equivalent of horizontal gene transfer. (As a shorthand, this lateral transfer of information will be referred to here as "borrowing," even though it does not correspond to the exact linguistic definition of borrowing.) Most phylogenetic methods assume a strictly bifurcating model of evolution, while linguistic development often involves borrowing between unrelated dialects or languages. Extensive borrowing may result in an unresolved phylogenetic tree or an incorrect tree topology.

The linguistic phylogenetic community has made use of two major approaches to handling borrowing, though the effectiveness of these approaches remains unclear.

The first approach is character weighting. A phylogenetic analysis may include characters based on phonological, morphological, syntactic, and lexical features, but some of these classes of features are more resistant to borrowing than others (Sankoff 2001). Giving these classes of characters more weight should improve the outcome of the phylogenetic analysis. This issue has already been addressed in some detail (e.g. Barbançon et al. 2013), but the Greek dialects provide another opportunity to test different possibilities.

The second approach is phylogenetic network models (Nichols and Warnow 2008, 762-764). If borrowing between related languages or dialects has been too extensive, it may no longer make sense to model these languages or dialects as a bifurcating tree, but instead, as a network. However, no study has yet tested how well they perform on linguistic data. Since most studies of the Greek dialects assume an initially treelike model of evolution with later borrowing between the dialects, my data should provide a good opportunity to test the performance of phylogenetic tree and network methods.

<u>Research Plan</u>

Thus, in my dissertation, I will test how well character weighting and phylogenetic network models improve the results of the phylogenetic analysis of the Greek dialects, and then propose my own approach to solving the difficulties presented by borrowing.

In Chapter 2, I will present a discussion of the phylogenetic data matrix used in the subsequent analyses. This includes a consideration of the general principles involved in developing a phylogenetic database and the particular challenges presented by the Greek dialects, as well as a discussion of each of the linguistic features used as phylogenetic characters.

In Chapter 3, I will test how well tree models perform on the Greek dialects under different weighting schemes, and test whether character weighting increases the accuracy of phylogenetic reconstruction. As mentioned above, character weighting is often put forward as a way to strengthen the influence of the characters that are the least likely to be borrowed and weaken interference from the characters that are the most likely to be borrowed. I will test two different methods of character weighting. First, I will test the standard practice of weighting based on the class of phylogenetic character, such as phonological, morphological, or lexical characters. Second, I will also test the practice of reweighting characters according to their consistency index (CI), a measure of how many times a character changed over the course of the tree.

I will discuss the results and offer an explanation for why these different weighting schemes performed the way they did, in the hopes that it will potentially be able to improve the effectiveness of character weighting schemes. I will use Maximum Parsimony for these analyses, since Maximum Parsimony performs well in simulation studies (Barbançon et al. 2013, 113). It would be worthwhile to test other types of tree models in future research, but that is unnecessary for the current work, because many of the problems and general conclusions will apply to all tree methods.

The performance of different weighting schemes will be evaluated by their accuracy at recovering known subgroups, the amount of resolution in the tree, and measures of noise such as the consistency index, retention index, and rescaled consistency index.

In Chapter 4, I will determine whether phylogenetic network models perform well on the Greek dialect data, since it is expected to be substantially non-treelike. As mentioned earlier, most phylogenetic methods assume that the process of evolution produces strictly bifurcating trees, and that there is little or no borrowing of features between branches. These assumptions are obviously at odds with the process of development we assume took place for the Greek dialects. Phylogenetic network models like NeighborNet are very popular within the linguistic community for situations such as these (e.g. Holden and Gray 2006, Bryand 2006, McMahon and McMahon 2006). However, very little attention has been devoted to the question of whether or not phylogenetic network models are well-suited to this task. A number of different processes can be responsible for the lateral transfer of linguistic features, and a network model which is well-suited to one process may not be well-suited to another. I will compare the results given by the network models against the results of the Maximum Parsimony analysis and what is known about the development of the Greek dialects. In particular, I will analyze how the patterns of homoplasy and borrowing found in the Maximum Parsimony analysis are reflected in the network models.

In Chapter 5, I will present a new solution to the borrowing problem. If a phylogenetic analysis consists of a set of taxa, a set of characters, and a phylogenetic method, weighting characters assumes that the problem lies with the set of characters, and phylogenetic network models assume that the problem lies with the phylogenetic method. But, no approach has yet been developed to address whether the problem lies with the taxa. My approach does just that. If there has been extensive borrowing between related taxa, it may be the case that the taxa have not been defined correctly, and what has been represented as several taxa should instead be represented as a single taxon in the phylogenetic analysis. To this end, I will perform a clustering analysis using Multidimensional Scaling on the single problematic branch of the Greek dialects, the West Greek dialects. I will use the results to condense a number of dialects into single taxa, and run the analysis again to see if the results have improved the resolution of the evolutionary tree.

For the character-weighting analysis and the MDS analysis, I have chosen to use Maximum Parsimony as my phylogenetic method. Maximum Parsimony assumes that the tree or trees which require the fewest number of evolutionary changes is likely to be the correct tree. In simulation studies of linguistic data, Maximum Parsimony appears to be the most accurate method (Barbançon et al. 2013). Maximum Parsimony has the additional advantage that it is possible to analyze the data afterwards to determine which phylogenetic characters support each branch of the tree. Time constraints prevented me from testing other methods.

Finally, I will compare these approaches to determine which approach was the most successful at handling the effects of borrowing among the Greek dialects. As far as I know, no-one has yet produced a side-by-side analysis of different approaches to handling borrowing for any language group. If this comparison is successful, it would be useful to perform similar side-by-side approaches on other problematic data sets to see whether any generalizations can be drawn about which set of approaches is better across data sets.

Though it does not relate to the thrust of the preceding chapters, one final chapter details the changes which support each branch of the phylogenetic tree, and explores a new phylogenetic dating method using relative chronology.

Chapter 2: Phylogenetic Data Matrix Discussion and Feature List

In this chapter, I will provide a detailed discussion of the data I will use as the basis for the phylogenetic analysis. The data matrix itself is provided at http://www.pies.ucla.edu/students/Greek_Dialects_14_Mycenaean03.nex.

Overview of Phylogenetic Data Matrix

A phylogenetic analysis requires as input a data matrix in which the rows (or columns) represent the objects under study, or *taxa*. In this case, the taxa would represent dialects of ancient Greek. The columns (or rows) represent similarities and differences between the taxa, or *phylogenetic characters*. In other words, phylogenetic characters represent linguistic features.

These linguistic features include phonological, morphological, lexical, and syntactic differences between the dialects. Some of these features reflect changes that affected some or all of the dialects at some point during the history of the Greek language. One example would be the outcome of the inherited labiovelar *k^w before *e*. Inherited k^w is still found in Mycenaean, for instance *–qe* 'and,' but we find, for instance, a dental outcome in Attic-Ionic and West Greek, and a labial outcome in Aeolic, e.g. Boet. πετράταν, Att.-Ion. τετράτην 'fourth' (acc. s. f.), based on PIE *k^w*etwer*- 'four.' Other features reflect variation which was present in Proto-Indo-European, where both variants were passed on to the Greek language. One example would be the preposition 'towards,' with the form πρός, from PIE **proti*, in Attic-Ionic, and the form ποτί, from PIE **poti,* in most Aeolic and West Greek dialects. Each phylogenetic character contains two or more *phylogenetic character states*, which represent the ways in which each linguistic feature is attested in each dialect. Each character state is assigned a number, and these numbers fill the cells in the data matrix. In this way, each cell in the data matrix represents the way in which each linguistic feature is attested in each dialect.

An illustration of these concepts is provided in Figure 1, a representation of a part of my phylogenetic data matrix of ancient Greek. For instance, phylogenetic character 25 from the data matrix is the outcome of labiovelars before front vowels. This phylogenetic character is coded as one column in the phylogenetic data matrix. In the figure below, the phylogenetic character "Development of the voiceless labiovelar /k^w/ before /e/" is represented in column 25. Phylogenetic character states represent whether labiovelars were retained, whether the outcome was a sibilant or an affricate, whether the outcome was a dental, or whether the outcome was a labial. Each of these outcomes is assigned a number, zero through three. In the figure below, the character states are described and numbered below the data matrix. Finally, the cells in the column representing labiovelars are filled by numbers, such that Arcadian and Cypriot are coded as 1, the Aeolic dialects as 2, and the Attic-Ionic and West Greek dialects as 3. In the figure, the taxa-- in this case, the various Greek dialects-- are listed in a column at the left. The cell representing Attic, for example, is filled with a 3, indicating that the outcome of the labiovelar /k^w/ in this environment was a dental.

When all of the relevant linguistic features distinguishing the Greek dialects are taken into account, the phylogenetic data matrix provides a comprehensive account of the linguistic variation in the Greek dialects.

Figure 1: Representation of phylogenetic data matrix showing the development of the voiceless labiovelar /k^w/ before /e/:

	24	**25**	26
Arcadian	0	**1**	2
Cypriot	0	**1**	2
Attic	0	**3**	2
West Ionic	0	**3**	2

<u>26: Development of the voiceless labiovelar kw before e:</u>

0 kw is retained
1 Outcome is a sibilant or affricate
2 Outcome is a labial
3 Outcome is a dental
? Unknown
- Not applicable

<u>Selection Criteria for Taxa</u>

The studies in this dissertation focus only on the dialects of the Greek mainland, the coast of Asia Minor, the Aegean islands, Crete, and Cyprus. All dialectal material attested from outside this area is excluded. This does exclude some dialects which are attested early and include a substantial amount of material, like Cyrenaean, which was spoken in modern-day Libya. The purpose of this project is to determine how the Greek dialects developed over time and populated the Greek mainland, the Aegean Sea, and the coast of Asia Minor, and omitting dialects outside of this area keeps the scope of the project feasible.

The exact choice of taxa more or less follows Buck (1955, xi-xiii), with some important differences, such as the inclusion of Mycenaean. Achaean is represented by Arcadian, Cypriot, and Mycenaean. For Attic-Ionic, Ionic is represented by three separate taxa, representing West Ionic, Central Ionic, and East Ionic. These varieties of Ionic are separated by enough important differences, like the outcome of the consonant clusters **t*(^{h}).#*y*, **k*#*y*, and **t.w*, and the third compensatory lengthening, that it was necessary to code them as separate taxa. Another taxon, of course, represents Attic. Aeolic is represented by four taxa, representing Lesbian, Thessalian, and Boeotian. Within Thessalian, the dialects of western and eastern Thessaly, namely Thessaliotis and Hestiaeotis versus Pelasgiotis, differ enough that they are worth dividing into two separate taxa, East and West Thessalian, representing the dialects of Pelasgiotis and Thessaliotis, respectively (Buck 1955, 150-151, Colvin 2007, 92). West Thessalian has a number of features in common with non-Aeolic dialects, most prominently West Greek, such as the genitive singular of o-stems in –ου, not –οι, and the present infinitive of thematic verbs in –ειν, instead of –εμεν.

Within West Greek, Northwest Greek is represented by Elean, Locrian, and Phocian. The area around the Saronic Gulf is represented by Megarian, Corinthian, and West and East Argolic. As with Ionic, there were enough important differences between West and East Argolic, such as the outcome of the second and third compensatory lengthenings, to necessitate splitting them into two separate taxa. The remainder of the Peloponnese is represented by Laconian. The islands are represented by Rhodian, Coan, Theran, and Cretan.

Pamphylian falls within the geographic area I outlined, so I included it in the data matrix for the sake of completeness, in case it might prove important to future scholars wishing to make use of my data matrix. However, Pamphylian was not included in the phylogenetic analyses. Several dialect groups, including Mycenaean/Arcado-Cypriot, West Greek, and Aeolic may have contributed to its genesis, and it may also show influence from neighboring Anatolian languages, such as Lycian, Sidetic, and Cilician. Given the circumstances, Pamphylian may represent dialect mixing, which is outside the range of problems this dissertation aims to address.

It is common for a phylogenetic analysis to include one or more taxa to serve as an outgroup. An outgroup would ideally consist of the language or languages most closely related to Greek without being Greek, in order to determine which character states were ancestral and which were innovations. The outgroup also serves to determine where the tree should be rooted. Unfortunately, there is no ideal outgroup for ancient Greek. Phrygian, which is probably the language most closely related to Greek, is poorly understood. Armenian is the next most closely related language to Greek, but would also not serve well as an outgroup because its phonology is very complex, there is very little data, and it is probably less closely related to Greek than previously thought (Clackson 1994). Vedic Sanskrit would be the next choice, but it is not as closely related to Greek as one would prefer, since innovations in Vedic Sanskrit may obscure ancestral forms. Macedonian may represent the language most closely related to Greek, or it may represent another dialect of Greek (Hatzopoulos 2007). In either case, the uncertain status of Macedonian makes it a poor candidate to include in a study which aims to test the accuracy of different phylogenetic methods. In light of these difficulties, I have chosen not to include an outgroup. As a result, this phylogenetic analysis of the Greek dialects will not be able to determine what was the initial split which divided the Greek dialects.

<u>Phylogenetic Characters</u>

There are a few general principles governing the selection of phylogenetic characters (after Skelton 2008):

First, phylogenetic characters must be heritable, that is, they must represent features which are able to be passed on from one generation to the next. As a general rule, this presents no problem for linguistic applications of phylogenetic systematics; language is passed from one generation to the next as children learn the language. However, when working with a dead language, we are at the mercy of what has been preserved in the written record. This fact gives rise to a number of circumstances under which variation should not be coded as phylogenetic characters. First, a variant which simply represents a mistake should not be coded as a phylogenetic character. However, when data are scarce, it may be difficult to tell the difference between a genuine variant and a mistake.

Second, a speaker of one dialect may attempt to imitate another to a greater or lesser degree. There are many examples of this, for instance, the spread of the *koiné* and the establishment of a single standard across Ionia based on the dialect of Miletus, despite the existence of regional dialects (Colvin 2007, 21). Naturally, only the native dialect should be used to code the phylogenetic data matrix, but this can sometimes be hard to determine. For instance, there exists systematic variation within Mycenaean known as ‘special’ Mycenaean (as opposed to ‘normal’ Mycenaean), but as yet, there is no consensus on whether this variation represents a different dialect of Mycenaean, influence from the scribe’s native dialect, or sound change in progress (e.g. Risch 1966, Chadwick 1983, Woodard 1986, Hajnal 1997, Thompson 2002-2003).

Phylogenetic character states must be homologous, that is, they must all share the same evolutionary origin. For example, for a phylogenetic analysis of mammals, it would be inappropriate to have a character, “Possesses a horn on its snout,” since, for instance, narwhals and rhinocerouses have evolved horns independently, adapting different aspects of their anatomy. For instance, for the coding of the thematic genitive singular, it would be inappropriate to have the character states –ου, -ον, -οιο, and –οι, since –ου and –οιο are different phonological outcomes of inherited *-*osyo*, but –οι is a variant of –οιο with apocope.

Phylogenetic characters and character states must be clear and unambiguous. This issue arises in the case of linguistic changes which are attested inconsistently in texts, such as examples of case usage with prepositions and case usage in general (e.g. Buck 1955, 108-110, 136ff.), or which are attested for a short time, and then disappear, such as rhotacism in West Ionic (Buck 1955, 56-57). Of course, there is no clear dividing line between features which were too inconsistent, and which ones were passable, so this is left to the discretion of the investigator. Cases of linguistic changes which were sporadic but were still included in the discussion are noted below.

A phylogenetic character must be able to change independently of every other character. In other words, linguistic changes which obligatorily took place together should not be coded as two different changes. For instance, it would be inappropriate to code one character representing the assibilation of the third person verbal ending from -τι to -σι, and another character representing whether the final consonant of the preposition 'towards,' Att.-Ion. πρός, WGk. ποτί, was *t* or *s*, since these two changes did not happen independently. The underlying assumption of the phylogenetic methods that I employ is that each phylogenetic character, each linguistic change, carries equal weight. If two characters changed in tandem, that effectively gives that single character a weight of 2. It is possible to give characters different weights, and, in fact, Chapter 3 deals with this very problem. However, it is best to do it explicitly, rather than hiding it in the data matrix.

A phylogenetic character must have chronological significance for the period of study in question. This study aims to uncover the pattern of development which led to the Greek dialects of historical times. Thus, changes which took place during historical times should not be useful for this purpose. However, this is a difficult principle to enforce in practice. Different dialects are first attested at different times, and not all dialects are attested well early on. Thus, two different changes in two different dialects may have taken place at the same time, but if one dialect is attested after the time of the change, and the other is first attested before the change, the first change would be included under this criterion, and the second would not. As a result, for the sake of consistency, I have included changes which did take place during historical times, up until ca. the 3rd century CE. For instance, in some dialects, ϝ (digamma) was lost during historical times (Buck 1955, 46-52).

A phylogenetic character should only be included if its state is known for a large enough number of taxa. What this number is depends on the investigator's preference, as well as the completeness of the data in general. For example, my phylogenetic data matrix of Linear B (Skelton 2008) contained a relatively large proportion of missing data, because we did not have examples of many Linear B signs in the repertories of most Linear B scribes. Thus, the standards for including a character were relatively low. However, there is enough documentary evidence for the Greek dialects that the values of many dialect features are known for most or all of the dialects. As a result, dialect features were only included as phylogenetic characters if they were attested for most of the dialects. This had the effect of excluding most lexical differences along the lines of English "pail" or "bucket." The handbooks frequently noted the variants in only a small handful of dialects. It is possible that these forms are not attested in most dialects, especially if they are not common words, or if the inscriptions in a given dialect simply never treat certain topics. In any case, it was beyond the scope of the project to hunt down variants in the other dialects.

A common restriction on which characters are included is that at least two of the taxa must have a different character state from the others. After all, one of the main purposes of a phylogenetic analysis is to group together subsets of taxa, and this is not possible if all or all but one of the taxa share the same character state. However, for studies of chronology, where branch length is important, not just branch arrangement, including such characters does serve a purpose. I do not address applications of phylogenetic systematics to chronological problems in my dissertation, but I want this data matrix to be as useful as possible to future researchers, so I have included phylogenetic characters that only affect one taxon.

<u>Sources</u>

In constructing this data matrix, I relied heavily on Colvin (2007) and Buck (1955). When these sources were insufficient, I consulted Thumb (1909), Bechtel (1921-4), in general, and Dubois (1988), Egetmeyer (2010), and Brixhe (1976) for Arcadian, Cypriot, and Pamphylian, respectively. When the discussion below notes that I was unable to determine a coding for a particular taxon for a given character, it typically means that I was unable to find it in these sources.

Concluding Remarks about Selection Criteria and Sources

At this point, it should be evident that the nature of the preserved material places a significant burden on the investigator to ensure that the phylogenetic character coding is accurate and consistent. Historical linguistics, especially the historical linguistics of an ancient language, is a historical science. As such, the quality of the data is highly dependent on the quality of the preservation and the quality of previous research in interpreting the existing evidence—it would be unreasonable to expect me to survey the whole of the inscriptional evidence, or to present a compelling argument for every dispute, in order to construct this data matrix. At the same time, the historical development and dialectology of ancient Greek have been very thoroughly studied. By primarily relying on two relatively recent and widely accepted handbooks, I have tried to adopt a relatively uncontroversial approach, though this often comes at the expense of comprehensiveness. However, one of the great strengths of phylogenetic analysis is that it is easy to update the data matrix as new data and new research become available, and to test what effect other intepretations might have on the outcome of the analysis. I welcome any such updates and inquiries.

For the following discussion, I have organized the phylogenetic characters into four classes: phonological, morphological, syntactic, and lexical features. Note that some phonological changes may be treated in the sections on morphological or lexical features if they exclusively or primarily affected a given morphological or lexical item.

Phonological Features

Sound Changes Affecting Vowels

Development of Proto Indo-European Syllabic Liquids (*_r̥_ and *_l̥_)

The standard view of the development of the Proto-Indo-European syllabic liquids is that *_r̥_ and *_l̥_ became αρ/ρα in Attic-Ionic and West Greek, but ορ/ρο in Aeolic, Mycenaean, and, to a lesser extent, Arcado-Cypriot, e.g. Att.-Ion. στρατός, Lesb. στρότος < _*str̥-to-_ ‘military foce’ (Colvin 2007, 11, 32, 41). Two alternative views have been put forth. One, first developed by Heubeck (1972), is that *_r̥_ is preserved in Mycenaean Greek. The second view, put forward by Morpurgo Davies (1968), is that the development of the syllabic liquids in Mycenaean and Arcado-Cypriot was a conditioned sound change, in which *_r̥_ and presumably *_l̥_ became _or_/_ro_ after _w_ and _ar_/_ra_ elsewhere. I found this view the most compelling when I was constructing the data matrix. However, Vine (Pers. comm. 3/14/2014) has brought it to my attention that counterexamples exist, and that Morpurgo Davies' theory was adequately refuted by García Ramón (1985). Unfortunately, I became aware of this too late to incorporate it in the data matrix. Therefore, the data matrix for the phylogenetic analysis notes whether the vocalism of the outcome was α, ο, or α except after _w_, where the result is ο. A detailed study of the outcomes of *_r̥_ and *_l̥_ in Aeolic would be useful in case they also show conditioning. I should note that the most thorough treatment of this material is now van Beek (2013), which came to my attention too late to incorporate in this dissertation.

Development of Proto Indo-European Syllabic Nasals (*m̥ and *n̥)

The Proto-Indo-European syllabic nasals **m̥* and **n̥* became α/αν in Attic-Ionic and West Greek, but sometimes vocalize with *o* in Aeolic, Mycenaean, and Arcado-Cypriot, e.g. Att.-Ion. δέκατος, Lesb., Arc. δέκοτος < **deḱm̥-to-* 'tenth' (Colvin 2007, 32, 41, Buck 1955, 20, Weiss 2010, 101). In Mycenaean in particular, the outcome is /a/, except that /a/ alternates with /o/ in the vicinity of labial consonants (e.g. Woodard 1986, Thompson 1997, Varias 1997). The data matrix for the phylogenetic analysis notes whether the outcome of **m̥* and **n̥* was always vocalized with /a/, or whether it was sometimes vocalized with /o/ as well.

Long Alpha Fronted in Attic-Ionic

In Attic and Ionic, original /a:/ was fronted, e.g. Att.-Ion. φημί versus φᾱμί 'I say' from the root *$b^h eh_2$*- 'say,' Lat. *fātur* 'he says.' This fronted /a:/ ultimately merged with /ę:/ (represented by η), but remained distinct for some time, as we can see from some Central Ionic inscriptions which write fronted /a:/ with η, but /ę:/ with ε (Buck 1955, 21). The phylogenetic data matrix records whether /a:/ has been fronted and merged with /ę:/, written as η.

The Attic Reversion

In Attic, fronted /a:/ became /a:/ after /i/, /e/, and /r/, while elsewhere fronted /a:/ merged with /ę:/. This process is known as the Attic Reversion (Buck 1955, 21, Colvin 2007, 36). For instance, Attic has οἰκία 'house' where Ionic has οἰκίη. The phylogenetic data matrix records whether the Attic Reversion has taken place.

Quantitative Metathesis and Prevocalic Shortening

In Attic and sometimes in Ionic, in the sequences ηο and ηα, η generally shortened to ε, while the ο or α lengthened to ω or ᾱ, for example, βασιλῆος > βασιλέως 'king (gen.)'. This process is called quantitative metathesis. In Ionic and most West Greek dialects, η is shortened in this sequence without lengthening of the second vowel, a process called prevocalic shortening (Buck 1955, 41). In Arcado-Cypriot, Aeolic, and Elean, the original sequence is retained. The details are, of course, more complicated (Lejeune 1972, 253). The phylogenetic data matrix records whether quantitative metathesis, prevocalic shortening, or no change has occurred.

E Raised to I Before N in Arcado-Cypriot

In Arcado-Cypriot and Pamphylian, ε is raised to ι before ν, for instance, ἰν for ἐν 'in' (Buck 1955, 23, Colvin 2007, 32, Egetmeyer 2010, 72-75, Dubois 1988, 17-22, Brixhe 1976, 17-18). The phylogenetic data matrix records whether this change has occurred.

O Raised to Y in Arcado-Cypriot

In Arcado-Cypriot, final and pre-nasal ο is raised to υ, e.g. gen. s. -ᾱυ for -ᾱο and middle verbal endings in -τυ and -ντυ instead of -το and -ντο (Buck 1955, 27, Colvin 2007, 32, Dubois 1988, 23-28, Egetmeyer 2010, 59-66). In Pamphylian, *o* is also raised to *u*, not just when final but also in final syllables ending in a consonant (Buck 1955, 27, Colvin 2007, 48, Brixhe 1976, 20-24).

The phylogenetic data matrix notes as three separate features whether *o* has been raised to *u* in final position, final syllables, or pre-nasal position.

<u>I Lowered to E after P (Rho) in Aeolic</u>

The phoneme ι was sometimes lowered to ε in Aeolic, e.g. Thess. κρεννέμεν 'to distinguish' (inf.) for κρίνω 'I distinguish' (Buck 1955, 25). This sound change is not commonly written, so the phylogenetic data matrix notes whether any examples are attested.

<u>EI Monophthongized and Raised to I in Boeotian</u>

The monophthongization of the diphthong ει to ι in Boeotian occurred in the fifth century (Colvin 2007, 41). The phylogenetic data matrix notes whether this change has occurred.

<u>AI Monophthongized and Raised to H in Boeotian</u>

In Boeotian, the diphthong αι began to transition to a monophthong starting at the end of the sixth century, and came to be written as αε and then η (Colvin 2007, 42). The phylogenetic data matrix records whether this diphthong was written as αι, αε, or η.

<u>OI Monophthongized and Raised to Y in Boeotian</u>

In Boeotian, the diphthong οι became a monophthong, coming to be written οε and then, in the middle of the fourth century, υ (Colvin 2007, 42). The phylogenetic data matrix records whether this diphthong was written as οι, οε, or υ.

<u>E Raised to EI or I Before Vowel in Boeotian</u>

In Boeotian, ε came to be written as ει starting in the sixth century, and ι starting in the fourth century. This indicates synizesis, the process by which two vowels which were both originally syllabic come to be pronounced as a single syllable (Colvin 2007, 42). The phylogenetic data matrix records whether this vowel was written as ε, ει, or ι.

H Raised to EI in Boeotian and Thessalian

In Boeotian and Thessalian, η came to be written as ει, e.g. Thess., Boeot. ἀνέθεικε for ἀνέθηκε 'he dedicated' (Buck 1955, 25). The phylogenetic data matrix records whether or not η was written as ει.

E Lowered to A Before P (Rho) in Northwest Greek

In Northwest Greek, ε is lowered to α before ρ, e.g. Locr. φάρειν for φέρειν 'to carry' (inf.) (Buck 1955, 23, Colvin 2007, 45). The phylogenetic data matrix records whether this sound change has occurred.

H Lowered to Ā in Elean

In Elean, the vowel originally represented by η had lowered to the point that it was sometimes represented by ᾱ, e.g. ϝράτρα for ῥήτρα 'an agreement' (Buck 1955, 25, Colvin 2007, 45). Since this spelling is not consistent, the phylogenetic data matrix records whether η is frequently written as ᾱ.

E Raised to EI or I Before Back Vowels

E is irregularly raised to ει or ι before back vowels in various dialects, including Attic and Ionic, Lesbian, Thessalian, Boeotian, Cypriot (Egetmeyer 2010, 72-74), and Argolic, and Laconian and Cretan except where ε had originally been preceded by intervocalic ϝ (Buck 1955, 21-22). One example would be θειός for θεός 'god.' This feature is included in the phylogenetic analysis, despite the irregularity with which it is attested in even the dialects which show it. It notes whether ε was raised to ει or ι before all back vowels, or whether ε was raised to ι except where it was originally followed by intervocalic ϝ.

Contraction of A or Ā Plus E, EI, and H

In Attic and Ionic, sequences of α or ᾱ plus ε, ει, and η contract to ᾱ, while in the other dialects, they contract to η, e.g. Att.-Ion. νικᾶν, Arg. νικῆν < νίκα-εν 'to win' (inf.) (Buck 1955, 37, Dubois 1988 30, Egetmeyer 2010, 113). There is no definitive evidence for Pamphylian, but it is likely that the outcome was ε̄ (Colvin 2007, 180), though this was brought to my attention too late to include in the data matrix. The phylogenetic data matrix notes whether the outcome of contraction, if it occurred, was η or ᾱ.

Loss of Coda Nasals and Nasalization of Vowels

Though the graphical evidence is difficult to interpret, Cypriot most likely shows nasalized vowels resulting from the loss of a coda nasal (Egetmeyer 2010, 97, Colvin 2007, 32-33). In Pamphylian, too, the evidence probably points to loss of nasals in the sequence -*VNC*- with accompanying nasalization of the vowel (Brixhe 1976, 64-68, Colvin 2007, 48). The phylogenetic data matrix records whether coda nasals were lost, most likely producing nasalized vowels.

Sound Changes Affecting Consonants

Loss of Initial F (Digamma)

Loss or retention of digamma (/w/, written ϝ), is one of the most distinctive differences among the Greek dialects, especially since it appears in common words like ϝοῖκος/οἶκος 'house,' and ϝέτος/ἔτος 'year.' Attic-Ionic and Lesbian lost initial ϝ, while Arcado-Cypriot, Thessalian, and Boeotian retained it (Colvin 2007, 33, 37, 42). Within West Greek, Theran, Coan, and Rhodian lost initial ϝ, while Laconian, Cretan, and Argolic (Thumb 1909, 107) retained it (Colvin 2007, 45). Corinthian retained initial ϝ in early inscriptions, but lost it in later times, so in the phylogenetic data matrix it has been coded as present (Thumb 1909, 114). In Megarian, handbooks are silent on the presence or absence of initial ϝ, so it has been coded as unknown (Thumb 1909, 118). Initial ϝ is retained in Elean (Thumb 1909, 174), Phocian, and Locrian (Thumb 1909, 190). Initial ϝ is also retained in Pamphylian (Brixhe 1976, 47). The phylogenetic data matrix records the presence or absence of initial ϝ.

Loss of Intervocalic F (Digamma)

Intervocalic ϝ was lost in Attic-Ionic and Lesbian (Colvin 2007, 33, 42), but retained consistently in Cypriot and Pamphylian (Brixhe 1976, 47), and in Arcadian through the 5th century in certain words (Dubois 1988, 57). Intervocalic ϝ was retained in early Thessalian material (Thumb 1909, 239), but lost early in Boeotian, though it is still occasionally written (Thumb 1909, 226). In early Phocian, Locrian, Laconian, Argolic, and Corinthian inscriptions, intervocalic ϝ was retained, e.g. Phoc. κλέϝος, Att. κλέος 'fame' (Buck 1955, 238, 249, 268, 283, 294). Intervocalic ϝ was lost in Rhodian, Coan, and Theran. Intervocalic ϝ was also retained in Elean (Thumb 1909, 174). I was unable to determine whether intervocalic ϝ was lost or retained in Megarian. The phylogenetic data matrix records whether intervocalic ϝ was retained at all, or was lost.

Loss of Initial Aspiration

Loss of initial aspiration, or /h/, called 'psilosis,' occurs in East Ionic, Lesbian, Elean, and Cretan, while Boeotian, Locrian, and Phocian show psilosis in the article only (Buck 1955, 52-55). West and Central Ionic lost aspiration early, but since aspiration is attested in the earliest inscriptions, I have coded them as showing aspiration (Buck 1955, 53). Other dialects, such as Arcadian, Locrian, and Argolic, show spelling irregularities which may show aspiration in the process of disappearing (Buck 1955, 54). These dialects have been coded as showing aspiration. The phylogenetic data matrix records whether initial aspiration is retained, or whether initial aspiration is lost completely or in the article only.

Loss of Secondary Intervocalic Σ

Secondary intervocalic σ, such as σ which resulted from the outcome of clusters of stop + yod, or σ which was restored under morphological pressure, such as in the σ-aorist, became aspiration or was lost completely in Laconian, Argolic, Elean, and Cypriot, e.g. Lac. νικάhας for νικάσας, 'having won.' (Buck 1955, 55-56). The loss of secondary intervocalic σ took place in different places at different times. For example, in Laconian the change is present from the earliest inscriptions, while in Elean, the change only appears in the middle of the fourth century. In Cypriot, the change only appears sporadically. In Argolic, intervocalic σ is present the earliest inscriptions. Therefore, in the phylogenetic data matrix, I have coded only Laconian as showing the loss of secondary intervocalic σ.

Development of the Voiceless Labiovelar /k^w/ Before /e/

The Proto-Indo-European labiovelars were retained in Mycenaean, but lost in the historical Greek dialects. Before /e/, the voiceless labiovelar /k^w/ became a sibilant or affricate in Arcado-Cypriot, a labial in Aeolic, and a dental in the other dialects, e.g. Boet. πετράταν, Att.-Ion. τετράτην 'fourth' from PIE *k^w*etwer*- 'four,' Arc. εἴ-σε, Att.-Ion. εἴ-τε 'either…or,' with the second member from PIE *-k^w*e*, Myc. -*qe*, Lat. -*que* 'and.' (Buck 1955, 62-63, Colvin 2007, 33, 42). The phylogenetic data matrix records whether labiovelars became sibilants, dentals, or labials in this environment.

Development of *$t(^h).y$

When the cluster *$t(^h)y$ contained a syllable boundary alone, it became σσ in West Greek, Lesbian and Thessalian, ττ in Boeotian and Central Cretan, and σ in Attic-Ionic and Arcado-Cypriot, e.g. *$med^h.yos$ > Att-Ion., μέσος, Lesb., Thess., WGk. μέσσος, Boeot., Cret. μέττος 'middle' (Weiss 2010, 178-179). The phylogenetic data matrix records whether the outcome of *$t(^h).y$ was σσ, ττ, or σ.

Development of *$t(^h).\#y$

When the cluster *$t(^h)y$ contained both a syllable and a morpheme boundary, it became ττ in Attic, West Ionic, Boeotian, and Central Cretan, but σσ elsewhere, e.g. the feminine of adjectives in -εις, PIE *-$wet.\#ih_2$ > Proto-Gk. *wet.#ya*, generally gives -ϝεσσα, but Att. οἰνοῦττα 'wine cake' and Boeot. χαρίϝεττα 'graceful' (Weiss 2010, 179). The phylogenetic data matrix records whether the outcome of this cluster was σσ or ττ.

Development of *ky*

When the cluster **ky* did not contain either a morpheme or a syllable boundary, its outcome was written with the *z*-series (the signs *za, ze, zo*) in Mycenaean, τ in Attic, but σ elsewhere, e.g. **kyāmeron* > Att. τήμερον, Ion. σήμερον 'today' (Weiss 2010, 179), Proto-Greek **kyāwetes* > Ion. σῆτες, Att. τῆτες, *za-we-te* 'this year.' The phylogenetic data matrix records whether the outcome of this cluster was τ or σ.

Development of *k#y*

When the cluster **ky* contained a morpheme boundary, it became an unknown sibilant in early East Ionic, ττ in Attic, West Ionic, Boeotian, and Central Cretan, and σσ elsewhere, e.g. *p^hulak-yō* > Ion. φυλάσσω, Att. φυλάττω 'I guard' (Weiss 2010, 180). The phylogenetic data matrix records whether the outcome of this cluster was a mystery sibilant, ττ, or σσ.

Development of *dy, *gy*

The clusters **dy* and **gy* became δδ in Boeotian and Central Cretan, but ζ elsewhere, e.g. **dyeus* 'sky god' > Att.-Ion. Ζεύς, Boeot. Δεύς (Weiss 2010, 180). The phylogenetic data matrix records whether the outcome of this cluster was δδ or ζ.

Development of *ly*

The cluster **ly* metathesized in Cypriot, but became the geminate λλ elsewhere, e.g. *ailos* for ἄλλος 'other' (Colvin 2007, 32, Weiss 2010, 181, Egetmeyer 2010, 123). The phylogenetic data matrix records whether the outcome of this cluster was metathesis or gemination.

Development of *_t.w_

When the cluster *_tw_ contained a syllable boundary, it became an unknown sibilant in early East Ionic, ττ in Attic and Boeotian, and σσ elsewhere, e.g. *_k_^{w}_et-wr̥-es_ > Ion. τέσσερες, Att. τέτταρες ‘four’ (Weiss 2010, 182). The phylogenetic data matrix records whether the outcome of this cluster was a mystery sibilant, ττ, or σσ.

First Compensatory Lengthening

The first compensatory lengthening affected several phonological environments, but all are united by the outcome that /s/ or /y/ was lost with compensatory lengthening of the preceding vowel, except in Lesbian and Thessalian, where gemination of the remaining consonant occurred. These environments include _V{m, n, w}sV_; _VLsV_ when the sequence was accented on the second vowel, and morphologically conditioned cases which were accented on the first vowel; _Vs{r, l, m, n, w}V_; and _V{r, l, n}yV_ (Weiss 2010, 141-145). The phylogenetic data matrix records whether the outcome of the first compensatory lengthening was compensatory lengthening of the preceding vowel or gemination of the consonant, e.g. Proto-Gk. *_e-krin-s-a_ ‘I judged’ > ἔκρῑνα, but Lesb., Thess. ἔκριννα.

Second Compensatory Lengthening

In the second compensatory lengthening, for the sequence _Vns_, /n/ is lost with compensatory lengthening of the preceding vowel, except in Lesbian, where /n/ is lost with the final vowel becoming a diphthong in /i/. The second compensatory lengthening did not occur in Central Cretan, West Argolic, Arcadian, and Thessalian (Buck 1955, 67-69, Weiss 2010, 146), e.g. Proto-

Greek *$pant$-ya > *pansa* > Cret., Arg., Thess., Arc. πάνσα, Lesb. παῖσα, Att. πᾶσα 'all' (nom. f. s.). The phylogenetic data matrix records whether the second compensatory lengthening occurred, and, if so, whether the outcome was lengthening or diphthongization.

Third Compensatory Lengthening

In the third compensatory lengthening, for the sequence *V{n, r, l, s, d}w*, /w/ is lost with compensatory lengthening of the preceding vowel in East and Central Ionic, Cretan, Theran, Coan, Rhodian, and West Argolic (Weiss 2010, 149). In Attic, /w/ is lost without compensatory lengthening. In Pamphylian (Brixhe 1976, 48), Cypriot (Egetmeyer 2010, 133-134), Corinthian, Elean, Arcadian, and Boeotian, /w/ is retained, e.g. Corinth. ξένϝος, Att. ξένος, Ion. ξεῖνος 'guest, host, stranger' (Buck 1955, 49-51). The handbooks are apparently silent on West Ionic, Lesbian, Thessalian, Locrian, Phocian, Laconian, Megarian, and East Argolic, so these dialects have been coded as unknown in the phylogenetic data matrix. Brent Vine brought it to my attention that the outcomes in some of these dialects are noted in Lejeune (1972, 159), a particularly egregious oversight on my part, but unfortunately, this came too late to redo the entire analysis. The phylogenetic data matrix records whether or not /w/ has been lost in this position, and, if so, whether the third compensatory lengthening has occurred.

Merger of the New Long Vowels

The loss of laryngeals created one set of long vowels, written using η and ω, such as Gk. ἵστημι 'I stand' from the PIE root *$steh_2$- 'stand,' and δίδωμι 'I give,' from the PIE root *deh_3-. Subsequent sound changes, such as the three compensatory lengthenings and contractions produced by loss of *y* and intervocalic *s*, created a new set of long vowels, such as Proto-Gk. *p^ht^her-$y\bar{o}$ >

φθείρω 'I destroy,' and Proto-Greek **treyes* > **trees* > τρεῖς 'three.' In some dialects, this new set of long vowels produced by compensatory lengthening merged either with the original long vowels and the long vowels produced by laryngeal loss, for instance, Proto-Greek **xenwos* > ξῆνος 'stranger, guest, host.' In other dialects, they merged with the new long vowels produced by the monophthongization of the diphthongs ει and ου, for instance, Proto-Greek **xenwos* > ξεῖνος 'stranger, guest, host.' The dialects which show a merger of the new long vowels produced by compensatory lengthening with the inherited diphthongs include Attic-Ionic, Northwest Greek except for Elean, Corinthian, Megarian, and East Argolic. The dialects which show a merger of the new long vowels produced by compensatory lengthening with the old long vowels include the Aeolic dialects, Arcadian, Laconian, and Cretan.

However, in some dialects, the new vowels generated by compensatory lengthenings have merged with the old long vowels, while new long vowels generated by contractions of the same two vowels have merged with the diphthongs. These dialects include West Argolic, Theran, Rhodian, and Coan. In Elean, the front vowels show the merger of the new long vowels with the diphthongs, while the back vowels show the merger of the new long vowels with the old long vowels (Ruijgh 2007, 396, Buck 1955, 28-30). In Pamphylian, the old long vowels and new long vowels which resulted from compensatory lengthening remained distinct in the front vowels, but merged in the back vowels. In the back vowels, the product of isovocalic contractions merged with the original diphthong /ou/ (Brixhe 1976, 28-31). The phylogenetic data matrix records which of these conditions has occurred.

<u>Rhotacism</u>

The change of σ to ρ occurred in various positions in various dialects. In Elean, instances of both rhotacism of final ς and preserved final ς occur in early inscriptions, while later inscriptions show rhotacism uniformly. Laconian also shows rhotacism of final ς, but only in late inscriptions, so this has not been coded in the phylogenetic data matrix. The phylogenetic data matrix records rhotacism of final s.

Assimilation of ΡΣ to ΡΡ

In Attic, West Ionic, Arcadian, Elean, and Theran, the sequence *rs* assimilates to *rr,* e.g. Ion. ἄρσην, Att. ἄρρην 'male.' (Buck 1955, 69). The phylogenetic data matrix records whether this sound change has occurred.

Gemination of Ρ (Rho) and Other Consonants Before Ι

In Lesbian and Thessalian, /r/ and other consonants sometimes geminate before /i/, e.g. μέτερρος for μέτριος 'moderate' (Buck 1955, 26). The phylogenetic data matrix records whether this sound change has occurred.

Development of Ζ

In Boeotian, Elean, Cretan, Laconian, and Thessalian (only Thessaliotis; otherwise unknown), ζ became δδ word-internally, and δ word-initially, e.g. Boeot. γραμματίδδω for γραμματίζω 'be a secretary' (Buck 1955, 71). The phylogenetic data matrix records whether ζ has become (δ)δ.

Morphological Features

Nominal Morphology

Dative Plural Endings in Ā Stems

The original dative-locative plural ᾱ-stem ending was *-*āsi*, continued in Mycenaean with regular aspiration of the intervocalic -*s*- as -*a-i* and (rarely) Cretan and early Attic -ᾱσι with restored -*s*-. Otherwise, it was remodeled after the thematic stems, producing Ionic -ηισι, Lesbian and Pamphylian -ᾱισι, and -αις in the remaining dialects (Buck 1955, 86, Rau 2010, 181). The actual situation is more complicated, as most dialects show a combination of forms. Mycenaean -*a-i* reflects /-āhi/ (intervocalic **s* > *h* by regular sound change, so the historical forms appear to have been remodeled). The phylogenetic data matrix records whether the dative plural a-stem ending is -αις, -ᾱισι, -ᾱσι, or a combination of those forms.

Masculine Genitive Singular in Ā Stems

Most Greek dialects, including Mycenaean (Willi 2008), have the masculine genitive singular ending -ᾱο in the long ᾱ stems, which can undergo quantitative metathesis to -εω, or contraction to -ᾱ or -ω. Attic, however, has taken the genitive singular -ου from the thematic stems. A few examples of the original masculine genitive singular ᾱ-stem ending -ᾱς survive in various dialects (Buck 1955, 87). The phylogenetic data matrix records whether the genitive singular of masculine long -ᾱ stems is -ᾱο or its phonological variants, or -ου.

Genitive Singular of Thematic Stems

There are two sets of genitive singular thematic endings in Greek, those derived from -οιο < *-*oyyo* (e.g. Homeric -οιο, Thessalian -οι), and those derived from -οο < -**ohyo* (e.g. -ου), in other words, differences in outcomes of the First Compensatory Lengthening (Willi 2008). However, it should be noted that this is only one conception of the material; for another view, see Haug (2002, Ch. 3). In Cypriot we also find -*ōn*, presumably on analogy to the genitive plural (Buck 1955, 88). The phylogenetic data matrix records whether the thematic genitive singular found in a given dialect is -οι, -ου, or -*ōn*.

<u>Dative Plural of Thematic Stems</u>

In most dialects, the dative plural of thematic stems is -οις, from the original instrumental plural ending, but in early Attic, Ionic, Lesbian, Pamphylian, and sometimes Cretan and Argolic, we find -οισι (Buck 1955, 88), which comes from the original locative ending. Mycenaean continues to distinguish the instrumental and dative-locative plural, and so preserves both forms (Rau 2010, 182). The phylogenetic data matrix records whether a dialect has used the original instrumental (-οις) or locative (-οισι) form for the new combined dative-locative.

<u>Dative Singular of Thematic Stems</u>

In most dialects, the thematic dative singular is -ωι, but in Arcadian, Elean, Boeotian, and later inscriptions from northern Greece, the dative singular is -οι or its phonological variants (Buck 1955, 88). The phylogenetic data matrix records whether the dative singular in a given dialect is derived from -ωι or -οι. Since the attestations of -οι from northern Greece are late, they have been coded as -ωι in the phylogenetic data matrix.

<u>Accusative Singular of Consonant Stems</u>

In most dialects, the accusative singular of consonant stems is -α, but in Cypriot and sporadically in other dialects, the accusative singular is -αν (Buck 1955, 89, Egetmeyer 2010, 404). The phylogenetic data matrix records whether the accusative singular of consonant stems is typically -α, or, only in the case of Cypriot, typically -αν.

Dative Plural of Consonant Stems

In most dialects, the dative plural of consonant stems is -σι, but in Lesbian, Thessalian, and Boeotian, and more sporadically in Pamphylian, Phocian, Locrian, and Elean, the dative plural is -εσσι (Buck 1955, 89, Morpurgo Davies 1976, 183). The phylogenetic data matrix records whether the dative plural of consonant stems is -σι, -εσσι, or a combination of the two.

Inflection of *i*-Stem Nouns

In most dialects, *i*-stem nouns are declined with the stem -*i*- throughout, that is, -ις, -ιος, -ῑ, -ιν, -ιες, -ιων, -ισι, -ινς. Lesbian shows this system, but shows a nominative plural in -ῑς. Cypriot is declined with -*i*- throughout, but adds a /w/ to the stem, for instance, /ptoliwi/ 'city' (dat. s.) (Buck 1955, 91, Egetmeyer 2010, 412-414). In Attic and occasionally Central and East Ionic, the inflection of *i*-stems shows ablaut, that is, -ις, -ηος, -ει, etc. In Proto-Indo-European, i-stems ablauted (Meier-Brügger 2003, 207-208). So, this phylogenetic data matrix first records whether i-stems are inflected with or without ablaut. Two additional characters record whether *i*-stems show -*w*- in the stem, and whether the nominative plural is in -ῑς.

Verbal Morphology

First Plural Primary Verbal Ending in -μες or -μεν

The first plural primary verbal ending is -μες in West Greek, but -μεν in Attic-Ionic, Aeolic, and Arcado-Cypriot (Buck 1955, 111). There are no attestations of the first plural verbal ending in Pamphylian (Brixhe 1976, 120-121). The phylogenetic data matrix records whether the first plural primary ending was -μες or -μεν.

Third Singular Middle Primary Verbal Ending

Most dialects, including Pamphylian, have -ται as the third singular middle primary verbal ending, but Arcadian, Mycenaean, and Cypriot preserve the original Proto-Indo-European ending, which is reflected in -τοι (Buck 1955, 113, Egetmeyer 2010, 514, Colvin 2007, 15, Brixhe 1976, 120). The phylogenetic data matrix records whether the third singular middle primary verbal ending was -ται or -τοι.

Third Singular Active Athematic Primary Ending

The original third singular active athematic primary ending was -τι, which is preserved in West Greek, Boeotian, and Pamphylian, but in the other dialects, it assibilated to -σι (Buck 1955 111, Woodard 1986, Colvin 2007, 48). The phylogenetic data matrix records whether the third singular athematic primary ending was -τι, -σι, or a mix of forms.

Athematic Inflection of Contract Verbs

Most dialects show thematic inflection of contract verbs, but Lesbian, Thessalian, Arcado-Cypriot, and perhaps Pamphylian generally show athematic inflection of contract verbs (Buck 1955, 123, Brixhe 1976, 118-119). There is some Mycenaean evidence which bears on this, but which I saw too late to incorporate in this dissertation; see Rau (2009). The phylogenetic data matrix records whether contract verbs are inflected with thematic or athematic inflection, or a mixture of the two.

Formation of the Future Tense

In most dialects, the future is formed with the suffix -σ-, but in West Greek, the future is formed with the suffix -σε- (Buck 1955, 115). Examples of the future in Pamphylian (Brixhe 1976, 115-116), Locrian, and Elean (Buck 1955, 115) are lacking. The phylogenetic data matrix records whether the future tense was formed with -σ- or -σε-.

Aorists and Futures in Verbs with –ζ-

Verbs with stems in -δ- and -γ- both form present stems in -ζ- (< *-*dy*-, *-*gy*-), while in the aorist, verbs in -δ- form stems in -σ-/-σσ-, while verbs in -γ- form stems in -ξ-. The ambiguity in the present system has led to the extension of aorist stems in -ξ- to dental stems in some dialects, including West Greek, Thessalian, Boeotian and Arcado-Cypriot. Other dialects, including Pamphylian, retain the inherited system (Buck 1955, 115, Brixhe 1976, 116). Arcadian and Argolic contain the additional restriction that -σ- is used instead of -ξ- when it is preceded by a guttural. Boeotian shows both -ξ- and -σ-, depending on location. This phylogenetic data matrix records whether the futures and aorists of verbs in -ζ- appear as -ξ- or -σ-, whether there is a mix of the two, or whether there are phonological conditions on where -ξ- is used.

Third Plural Active Imperfect and Aorist Ending

The original third plural active imperfect and aorist ending was originally *-*nt* > -*n*, e.g. ἔδον 'he gave' (aor.), which most dialects retain. Boeotian, Locrian, and Arcado-Cypriot recharacterized this ending with -αν from the aorist, Attic-Ionic recharacterized it with -σαν, and Thessalian recharacterized it with -εν, from an unknown source (Colvin 2007, 37-38, Buck 1955, 112-113). There are no examples from Pamphylian (Brixhe 1976, 120-121). The phylogenetic data matrix records whether the third plural active imperfect and aorist ending is -ν, -αν, -σαν, or -εν.

Formation of the Middle Participle of Verbs in –εω

In most dialects, verbs in -εω typically form their middle participles in -ε-ομενος, which gives -εομενος. However, in Northwest Greek and Boeotian, they form their middle participles in -ε-εμενος, which contracts to -ειμενος or its phonological variant -ημενος (Buck 1955, 124). The phylogenetic data matrix records whether verbs in -εω have participles in -εομενος, or -ειμενος and -ημενος.

Formation of the Perfect Active Participle

In most dialects, the perfect active participle is formed with the inherited suffix *-*wos*-, but in the Aeolic dialects, *-*wos*- was replaced by *-*ont*- from the present active participle (Colvin 2007, 43). It is also worth noting that 1st-millennium dialects using *-*wos*- have developed *t*-stem forms (in masc. and neut.), whereas Myc. retains the original *s*-stem inflection, though this distinction is only relevant for determining branch length. No perfect participles are attested in Pamphylian (Brixhe 1976, 123). The phylogenetic data matrix records whether the perfect participle is derived from *-*wos*- or *-*ont*-.

Formation of Athematic Infinitives

Attic-Ionic, Arcado-Cypriot, and Pamphylian form athematic infinitives in -ναι, Lesbian forms infinitives in -μεναι, Cretan and Rhodian form athematic infinitives in -μην or -μειν, and Thessalian, Boeotian, and the remainder of the West Greek dialects form infinitives in -μεν (Buck 1955, 122, Brixhe 1976, 123-124). The phylogenetic data matrix records whether athematic infinitives are formed in -ναι, -μεν, -μην/-μειν, or -μεναι.

Syntactic Features

Case Usage with Two-Case Prepositions

In most dialects, prepositions such as ἐκ and ἀπό take the genitive. However, in Arcado-Cypriot and Pamphylian, these prepositions take the dative instead of the genitive (Colvin 2007, 34). The phylogenetic data matrix and records whether these prepositions take the genitive or the dative.

Order of Particles

Attic-Ionic, Arcado-Cypriot, Lesbian, and Thessalian have the word order εἰ/αἰ ἀν/κα/κε τις/κις, while West Greek has αἰ τις κα. Boeotian shows both word orders (Buck 1955, 140). Examples from Pamphylian are apparently unknown, as Brixhe (1976) does not mention it. The phylogenetic data matrix records whether the order of these particles is αἰ τις κα, ἐάν τις, or both.

Patronymics

In Mycenaean and Aeolic, patronymics are formed with adjectives in -ιος, while in other dialects, they are formed from the genitive of the father's name (Colvin 2007, 43). There is an additional morphological patronymic formed with the suffix -δης (Att.-Ion.), which competes with -ιος in epic language. While it does appear in dialectal texts, it does not have a clear distribution among the dialects, so it has been omitted from the phylogenetic character coding. The phylogenetic data matrix records whether patronymics are formed using adjectives in –ιος or the genitive of the father's name.

Lexical Features

The Modal Particle

For the modal particle, Attic-Ionic and Arcadian have ἀν, while Cypriot, Lesbian, Thessalian, and Pamphylian have κε, and Boeotian and West Greek have κα (Buck 1955, 105-106). These most likely arose from an original particle **ken*, with zero grade **kn̥*, which would have produced **ka* before consonants and **kan* before vowels. **ke* could have been produced from **ken* on analogy to **ka*. An error in segmenting the words in the phrase **ou kan* as **ouk an* would then have produced the particle ἀν (Forbes 1958). For a different view, see Dunkel (1990, 100-130). The phylogenetic data matrix notes whether the modal particle is ἀν, κε, or κα.

The Emphatic Particle

In most Greek dialects, the emphatic particle is γε, but in Boeotian and West Greek, it is γα (Buck 1955, 24). I was unable to determine which form Pamphylian shows. The phylogenetic data matrix records whether the emphatic particle is γα or γε.

The Copulative (Coordinating) Particle

In most dialects, the copulative particle is δέ. However, in Thessalian, the copulative particle is μά. The phylogenetic data matrix records whether the copulative particle is δέ or μά.

The Conditional Conjunction

In Attic-Ionic and Arcado-Cyprian, the conditional conjunction is εἰ or phonological variants thereof, but in Aeolic and West Greek, the conditional conjunction is αἰ, or phonological variants thereof (Buck 1955, 105). I was unable to determine which form Pamphylian shows. The phylogenetic data matrix records whether the conditional conjunction is εἰ or αἰ.

The Copulative (Coordinating) Conjunction

In most dialects, including Pamphylian, the copulative conjunction 'and' is καί, but in Cypriot and some Arcadian, it is κάς (Buck 1955, 106), though see also Willi (2003). The phylogenetic data matrix records whether the copulative conjunction is κάς or καί.

The Preposition ἀνά

Most dialects, including Pamphylian and Mycenaean (Beekes 2010, 97) have this particular preposition as ἀνά, but Lesbian, some Thessalian, and Arcado-Cypriot have ὀν and its phonological variants (Buck 1955, 20). The phylogenetic data matrix records whether this preposition appears as ἀνά or ὀν.

The Preposition διά

Most dialects have this particular preposition as διά, but Thessalian has διέ (Buck 1955, 21). There seem to be no examples in Mycenaean, at least judging by Aura Jorro and Adrados (1999). The phylogenetic data matrix records whether this preposition is διά or διέ.

The Preposition 'with' (μετά)

In most dialects, including Mycenaean (Aura Jorro and Adrados 1999, 441-442), the preposition 'with' is μετά, but in Lesbian, Boeotian, possibly Thessalian, Arcadian, Argolic, Cretan, Theran, and Pamphylian, the preposition 'with' is πεδά or its phonological variants (Buck 1955, 107). It is also worth mentioning that in Mycenaean, *pe-da* is attested, at least once, on KN V 114, but probably in the meaning 'towards' instead of 'with,' in the phrase *pe-da wa-tu* /peda wastu/, 'to the town.' The phylogenetic data matrix records whether the preposition 'with' is μετά or πεδά.

The Preposition 'towards'

The preposition 'towards' has many variations. It is πρός in Attic-Ionic and Lesbian, προτί in Argolic and Cretan (πορτί, with metathesis), ποτί in West Greek, Thessalian, and Boeotian, πός in Arcado-Cypriot (Buck 1955, 107-108), περτί in Pamphylian, and *po-si* in Mycenaean (Aura Jorro and Adrados 1993, 155). An additional form, ποί, is found before dentals in Argolic, Phocian, and Locrian (see also the discussion in Ellsworth 2011 and Willi 2012). The phylogenetic data matrix notes whether the preposition 'towards' comes from inherited **poti* or **proti*. Whether or not **ti* has assibilated to **si* is noted in the character which also notes assibilation of verbal endings. The phylogenetic analysis notes whether there is an additional form, ποί, which appears before dentals (Beekes 2010, 1238, 1224, 1226).

The Preposition ‘with’ (σύν)

In most dialects, the preposition ‘with’ is σύν, but in early Attic, it appears as ξύν (Buck 1955, 108), and in Mycenaean as *ku-su-*, /ksu(n)/ (Beekes 2010, 1038). Dunkel (1982, but see also Vine 1999, 562-563) argues that σύν originated from Proto-Indo-European **sóm* ‘together, with,’ from the stem *sém-, ‘one, united.’ Greek also inherited the morpheme **kóm* ‘with,’ which cross-contaminated with σύν to produce ξύν. By our first attestation of Greek, **kóm* had lost out as a preposition, surviving only in forms like κοινός ‘common.’ By the time of our alphabetic texts, the same fate had more or less befallen ξύν, with forms only surviving in Attic in active use, but with the adjectival form ξυνός ‘common, public’ surviving. The phylogenetic data matrix records whether the preposition ‘with’ is σύν or ξύν.

The Preposition ‘in’

In Northwest Greek, Boeotian, Thessalian, and Arcado-Cypriot, the preposition ‘in’ is ἐν and its phonological variants, while in other dialects, including Pamphylian, this is replaced by the extended form ἐνς and its phonological variants (Buck 1955, 107). The phylogenetic data matrix records whether the preposition ‘in’ is ἐν or ἐνς.

The Preposition ‘from’

In most dialects, the preposition ‘from’ is ἀπό, but in Arcado-Cyprian, Mycenaean, Pamphylian, Lesbian, and some Thessalian, the preposition ‘from’ is ἀπύ, and sometimes exists alongside ἀπό (Colvin 2007, 34). The phylogenetic data matrix records whether the preposition ‘from’ is always ἀπό, or whether ἀπύ is found.

The Adjective ‘Holy’

In Attic-Ionic, Arcado-Cypriot, and Mycenaean (Aura Jorro and Adrados 1999, 273-276), the adjective for ‘holy’ is ἱερός. In West Greek, Pamphylian, and Boeotian, the form is ἱαρός. In Lesbian and sometimes Ionic, the form is ἶρος (Buck 1955, 24). The form in Thessalian is unknown. The phylogenetic data matrix records whether the form of ‘holy’ is ἱερός, ἱαρός, ἶρος, or a combination of forms.

The Noun ‘Zeus’

In most Greek dialects, including Mycenaean (Aura Jorro and Adrados 1999, 180-181) and Pamphylian (Brixhe 1976, 112), the genitive and dative stem of the noun ‘Zeus’ is Διϝ- and its phonological variants. However, East Ionic, Coan, Theran, Cretan, and Elean have a genitive and dative stem in Ζην- (Buck 1955, 93). This oblique stem Ζην- is actually based on the inherited acc. sg. Ζῆν, which is the regular result of Stang’s Law, cognate with Ved. acc. sg. *dyā́m* (Meier-Brügger 2003, 97, Buck 1955, 34). The phylogenetic data matrix records whether the oblique stem of ‘Zeus’ is Διϝ- or Ζην-.

The Noun ‘Apollo’

In Cretan, Laconian, Corinthian, Cypriot, and Pamphylian, the noun ‘Apollo’ is Ἀπέλλων and its phonological variants, while in Attic-Ionic, Phocian, and Locrian, it is Ἀπόλλων. In Thessalian, the form is Ἄπλουν, with syncope (Buck 1955, 46). The phylogenetic data matrix records whether the noun ‘Apollo’ shows an /e/ or an /o/.

The Numeral ‘one’

In most dialects, including Mycenaean, the numeral 'one' is expressed with Proto-Greek **hens*
(m. s. nom.), **mia* (f. s. nom.), but in Lesbian, Thessalian, Boeotian, and Cretan, it is expressed with a different form, **ios* (m. s. nom.), **ia* (f. s. nom.) (Buck 1955, 94). I am unsure what the form is in Pamphylian. In the phylogenetic data matrix, I note whether the numeral 'one' is expressed with **hens* or **ios* in the singular and plural.

<u>The Numeral 'twenty'</u>

In Attic-Ionic, Lesbian, and Arcadian (unattested in Cypriot), the numeral 'twenty' is εἴκοσι, but in West Greek, Boeotian, Thessalian, and Pamphylian, it is ϝίκατι and its phonological variants (Buck 1955, 96). The phylogenetic data matrix records whether the numeral 'twenty' is εἴκοσι or ϝίκατι.

<u>Second Person Singular Pronouns</u>

The PIE nominative singular second person pronoun **tu* would have remained *tu* in Greek through regular sound change, while the accusative singular **twe* would have become *se*. However, the dialects leveled the initial consonant of the second person pronoun in one direction or the other. Attic-Ionic, Lesbian, and Arcadian leveled the forms in favor of /s/, while West Greek and Boeotian leveled the forms in favor of /t/ (Buck 1955, 97). The Cypriot form is unattested (Egetmeyer 2010, 438-9). I am unsure what forms were used in Thessalian and Pamphylian; the Mycenaean form is unknown. The phylogenetic data matrix notes whether the second person pronouns were leveled in favor of /t/ or /s/.

<u>First and Second Person Plural Pronouns</u>

The expected outcome of the PIE first and second person pronouns **n̥sme* and **usme* was **amme* and **humme*. Most dialects left the accusative as is, and added an -*s* to the nominative. Attic-Ionic, however, added -*es* to the nominative form and –*as* to the accusative form, giving ἡμεῖς and ἡμᾶς (Colvin 2007, 38). The Mycenaean and Pamphylian forms (Brixhe 1976, 114) are unknown. The phylogenetic analysis records whether the first and second person pronouns have added -*s* in the nominative only, or -*es* in the nominative and -*as* in the accusative.

<u>Nominative Plural of the Article</u>

The inherited nominative plural of the article was τοι, ται, which is retained in most dialects. However, Attic-Ionic, Arcado-Cypriot, Lesbian, and eastern Thessalian have replaced τοι, ται with οἱ, αἱ on analogy with the nominative singular (Colvin 2007, 38). The plural article is not attested in Pamphylian (Brixhe 1976, 114). The phylogenetic data matrix records whether the nominative plural of the article is τοι, ται or οἱ, αἱ.

<u>Dissimilation of k^w to *k* in Pronominal Forms</u>

In East Ionic, the labiovelars of pronominal forms rarely show dissimilation and loss of their labial element, giving forms like ὅκως for ὅπως (Buck 1955, 63, Lillo 1991). The phylogenetic data matrix records whether these pronominal forms ever show κ or the regular outcome of a labiovelar in that environment.

<u>Indefinite Pronoun</u>

The indefinite pronoun is Иις in Arcadian (where the symbol И probably indicates some sort of sibilant or affricate), σις in Cypriot, κις in the Thessalian of Pelasgiotis, and τις in the Thessalian of Thessaliotis and the other dialects (Buck 1955, 63, Lillo 1991). The Pamphylian form is unknown (Brixhe 1976, 114). The phylogenetic data matrix records whether the indefinite pronoun shows a sibilant or affricate, τ, or κ.

The Deictic Pronoun

The particle added to the article (formerly a demonstrative pronoun) to show deixis varies among the dialects. Thessalian has ὅ-νε, Arcadian has ὅ-νι, Arcadian and Cypriot have ὅ-νυ, and Attic-Ionic and West Greek have ὅ-δε (Buck 1955, 100). The form in Pamphylian is unknown (Brixhe 1976, 114). The phylogenetic data matrix records whether the particle added to show deixis is -νε, -νι, -νυ, or -δε.

Temporal Adverbs

The particle added to form the temporal pronouns (Attic ὅτε and τότε) varies among the dialects. Attic-Ionic, Arcado-Cypriot, and Mycenaean (Aura Jorro 1993, 52) have -τε, Lesbian has -τα, and West Greek, Boeotian, and Pamphylian have -κα (Buck 1955, 104, Brixhe 1976, 146). I am unsure what the Thessalian form is. The phylogenetic data matrix records whether -τε, -τα, or -κα is used to form these temporal adverbs.

Chapter 3: Phylogenetic Character Weighting

This analysis is designed to test whether weighting phylogenetic characters improves the outcome of the phylogenetic analysis of the Greek dialects. It tests two different character weighting schemes. In the first, phonological, morphological, and lexical characters are weighted. In the second, phylogenetic characters are weighted according to their consistency index (CI). A phylogenetic analysis with all characters unweighted is provided for comparison.

Methods

Phylogenetic Data Matrix

As described in the previous chapter, the phylogenetic data matrix consists of 40 phonological, 20 morphological, 3 syntactic, and 22 lexical characters, all function words. There are 22 taxa, representing most dialects from the Greek mainland, the Aegean islands, Asia Minor, and Cyprus. The taxa include the Arcado-Cypriot dialects of Arcadian and Cypriot, as well as Mycenaean. The Attic-Ionic dialects include Attic, West Ionic, Central Ionic, and East Ionic. The Aeolic dialects include Lesbian, West Thessalian (the dialect of Pelasgiotis), East Thessalian (the dialect of Thessaliotis), and Boeotian. The West Greek dialects include Phocian, Locrian, Elean, Corinthian, East and West Argolic, Laconian, Cretan, Theran, Coan, and Rhodian. Megarian was omitted because it is essentially identical to Corinthian, except that several more character codings are unknown. Pamphylian was omitted from the analysis because it most likely represents a mixed dialect, not the product of descent with modification.

Phylogenetic Analysis

The phylogenetic method I used for these analyses was Maximum Parsimony. Maximum Parsimony works by assuming that the shortest tree, that is, the tree that implies the least amount of evolutionary change, is the one that is correct (Swofford et al. 1996, 415-416). Conceptually, to accomplish this, the analysis first generates a large number of possible tree topologies, or possible ways to arrange the taxa into a tree given that each branch can only split into two daughter branches. The analysis then takes one such tree, and the phylogenetic data matrix that was the input. It uses the data matrix to determine where on the tree each of the phylogenetic characters in the data matrix must have changed in order to produce that arrangement of taxa. It then sums the total number of character state changes the tree required. This is the *tree length*. If the phylogenetic characters are weighted, it sums the weights of all the character state changes. At the end, the analysis selects the tree or trees which had the lowest tree length.

As a computational problem, both the weighted and unweighted versions of Maximum Parsimony are NP-hard (Foulds and Graham 1982), which means that for all but very small numbers of taxa, it would take too long to compute the exact solution, that is, to find the best tree or trees of all possible trees. Thus, it is necessary to employ a non-exact search strategy. For a discussion of the types of search strategies used for Maximum Parsimony, see Swofford et. al. (1996, 478-485). The Maximum Parsimony analyses in this analysis were carried out in PAUP* 4.0b for Windows (Swofford 2001) using the default settings.

Phylogenetic Analysis

Unweighted Maximum Parsimony Analysis

The Maximum Parsimony analysis with all characters given equal weight produced seven optimal trees. Optimal trees are defined as having the lowest possible branch length, so the statistics which follow apply equally to all the trees. The tree length, the smallest number of character state changes needed to produce the tree, was 189. The consistency index (CI), retention index (RI), and rescaled consistency index (RC) are all measures of homoplasy, or deviation from a treelike ideal. Possible values range from 0 to 1. A phylogenetic character which was perfectly treelike, or a tree which all phylogenetic characters matched perfectly, would have a value of 1. Values below 1 show homoplasy, with smaller numbers indicating more homoplasy. For a more detailed discussion, see Farris (1989). For this tree, the consistency index (CI) excluding uninformative characters was 0.6243. The retention index (RI) was 0.5799. The rescaled CI was 0.4730. The tree recovered all four major dialect groups as well as the correct subgrouping within these groups, and the tree was fully resolved except for the relationships among certain West Greek dialects. The strict consensus tree, that is, the tree which shows only the bipartitions which appear in all trees, is given below. Below that, one of the seven trees is given. For this run and the other runs in this chapter, the first tree was chosen arbitrarily to be an example.

Figure 2: Strict Consensus of 7 Trees—Unweighted

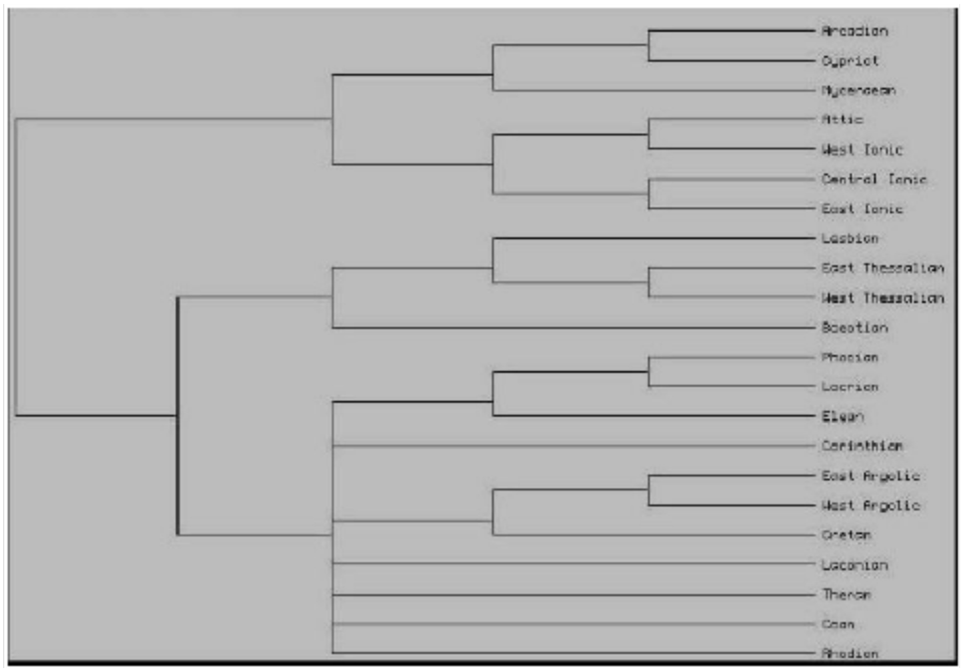
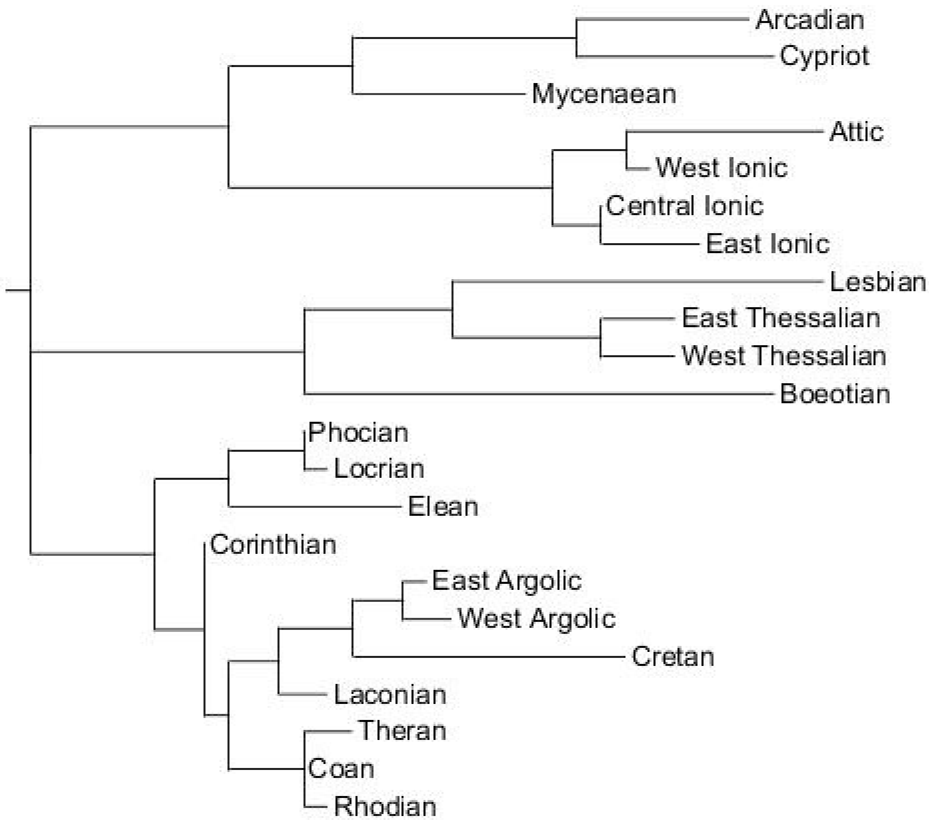


Figure 3: Tree 1-- Unweighted



Character Weighting—Phonological Characters

I performed two runs, one with phonological characters given weight 2, and one with phonological characters given weight 10. The other types of characters were given weight 1. I tried higher weights, but they produced results which were obviously wrong.

For the run with phonological characters given a weight of 2, there were two optimal trees, which only differ in the arrangement of certain West Greek dialects. The tree length was 281, the CI excluding uninformative characters was 0.6228, the RI was 0.7402, and the rescaled CI was 0.4610.

Figure 4: Strict Consensus of 2 Trees—Phonological Characters Weighted 2

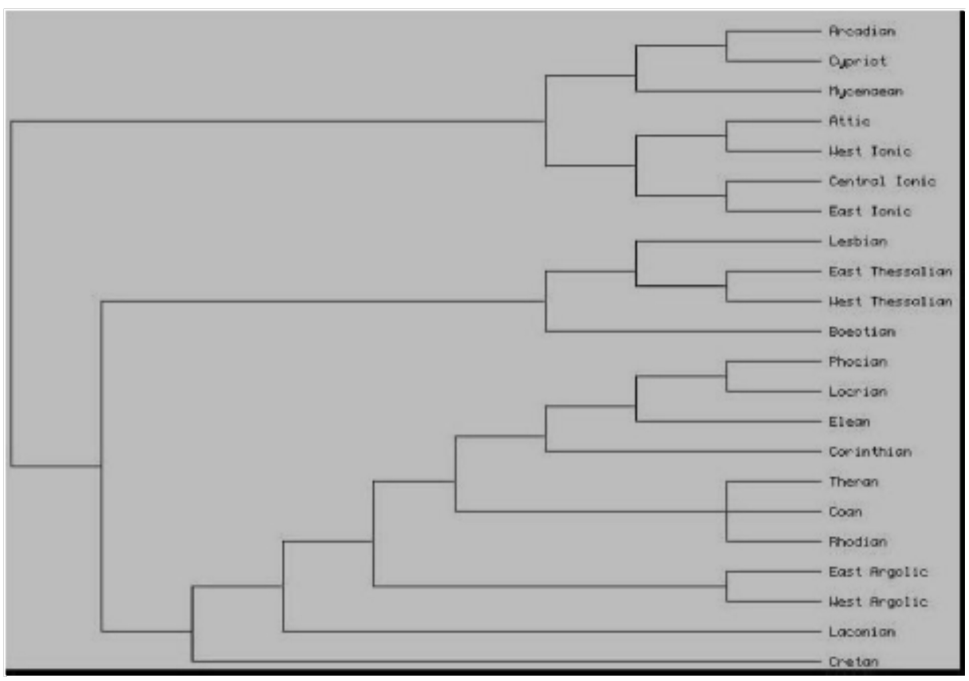
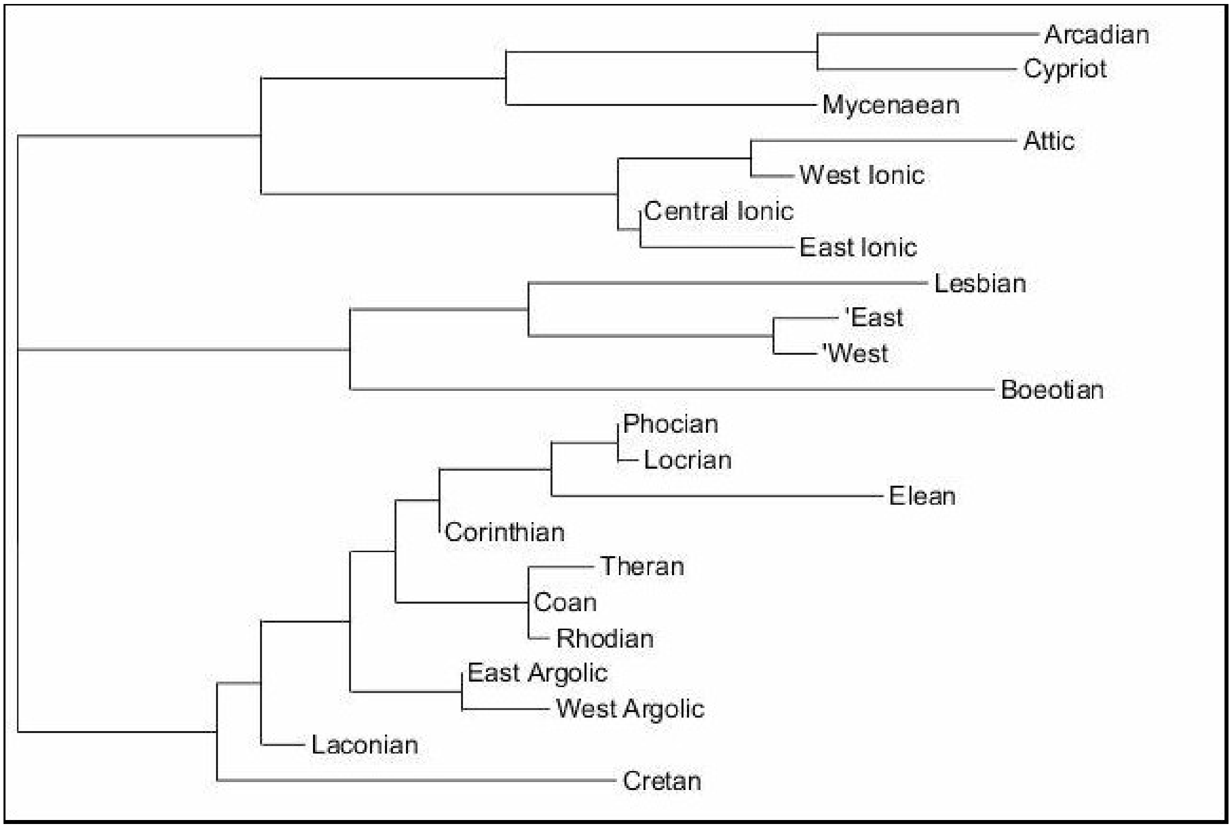


Figure 5: Tree 1-- Phonological Characters Weighted 2



For the run with phonological characters given weight 10, there were two optimal trees, which only differed in the arrangement of certain West Greek taxa. The tree topology is obviously wrong; the tree shows Aeolic as a sister taxon, that is, the closest relative, to Arcado-Cypriot, and Attic-Ionic as a sister taxon to Island Doric. The CI excluding uninformative characters was 0.5842, the RI was 0.7304, and the rescaled CI was 0.4660.

Figure 6: Strict Consensus of 2 Trees— Phonological Characters Weighted 10

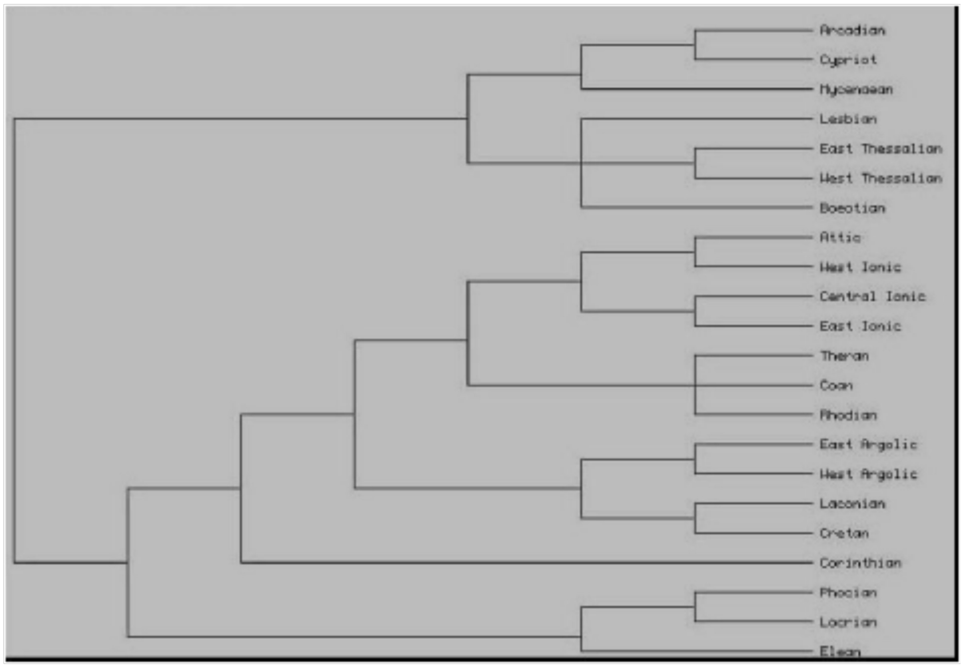
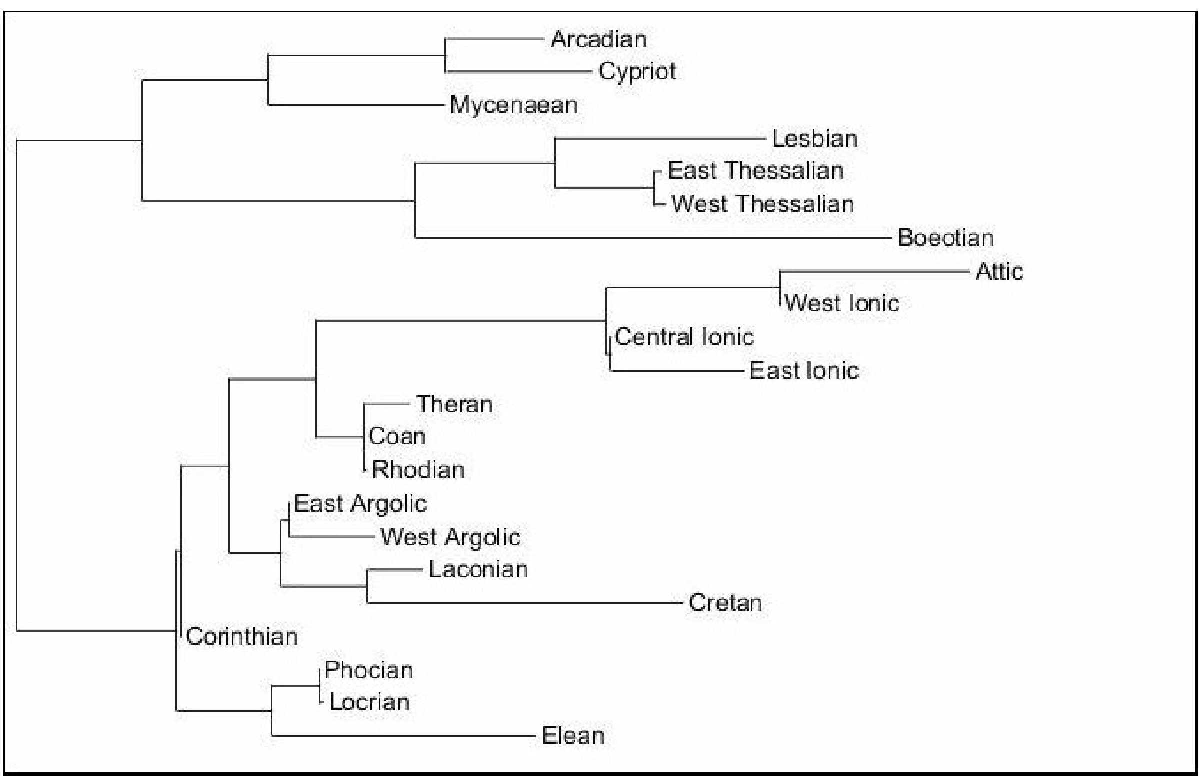


Figure 7: Tree 1--Phonological Characters Weighted 10



Character Weighting—Morphological Characters

Again, I performed two runs, one with morphological characters given weight 2, and one with morphological characters given weight 10, and the other types of characters given weight 1. I tried higher weights, but they produced results which were obviously wrong.

The run with morphological characters given a weight of 2 produced seven optimal trees, which only differed in the arrangement of certain West Greek taxa. The CI excluding uninformative characters was 0.5865, the RI was 0.7669, and the rescaled CI was 0.4851.

Figure 8: Strict Consensus of 7 Trees-- Morphological Characters Weighted 2

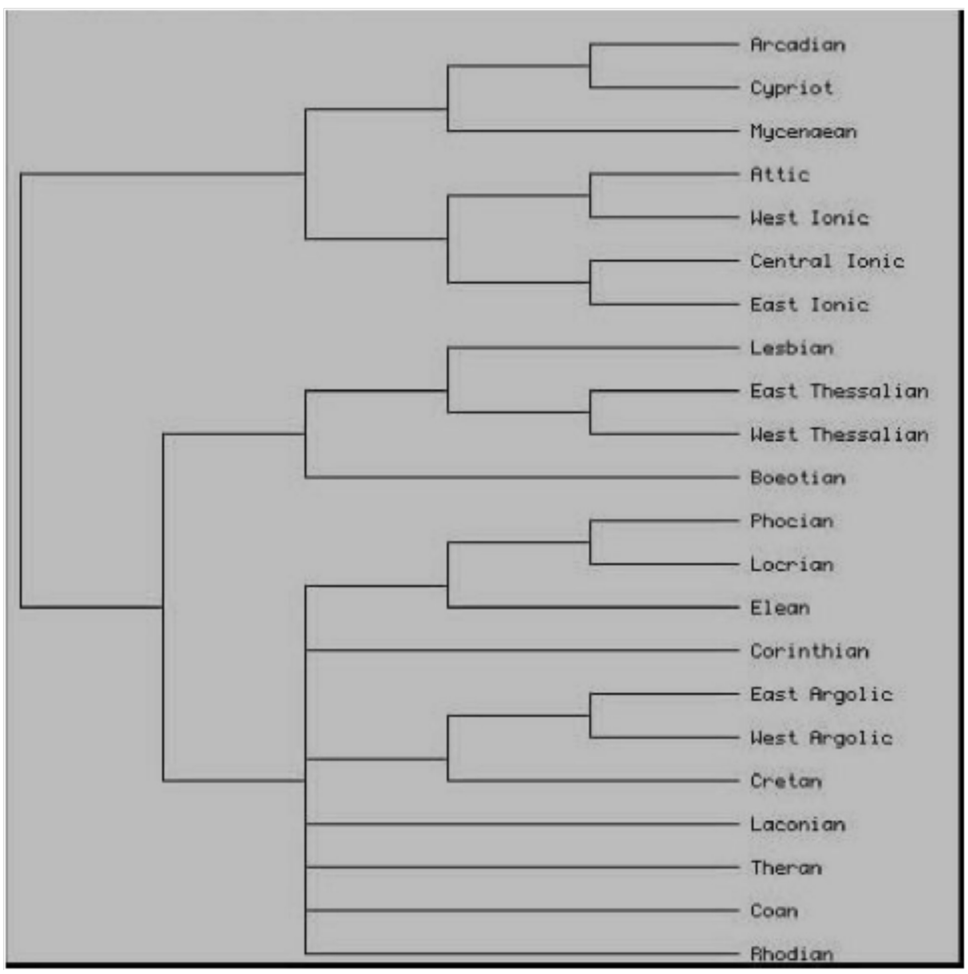
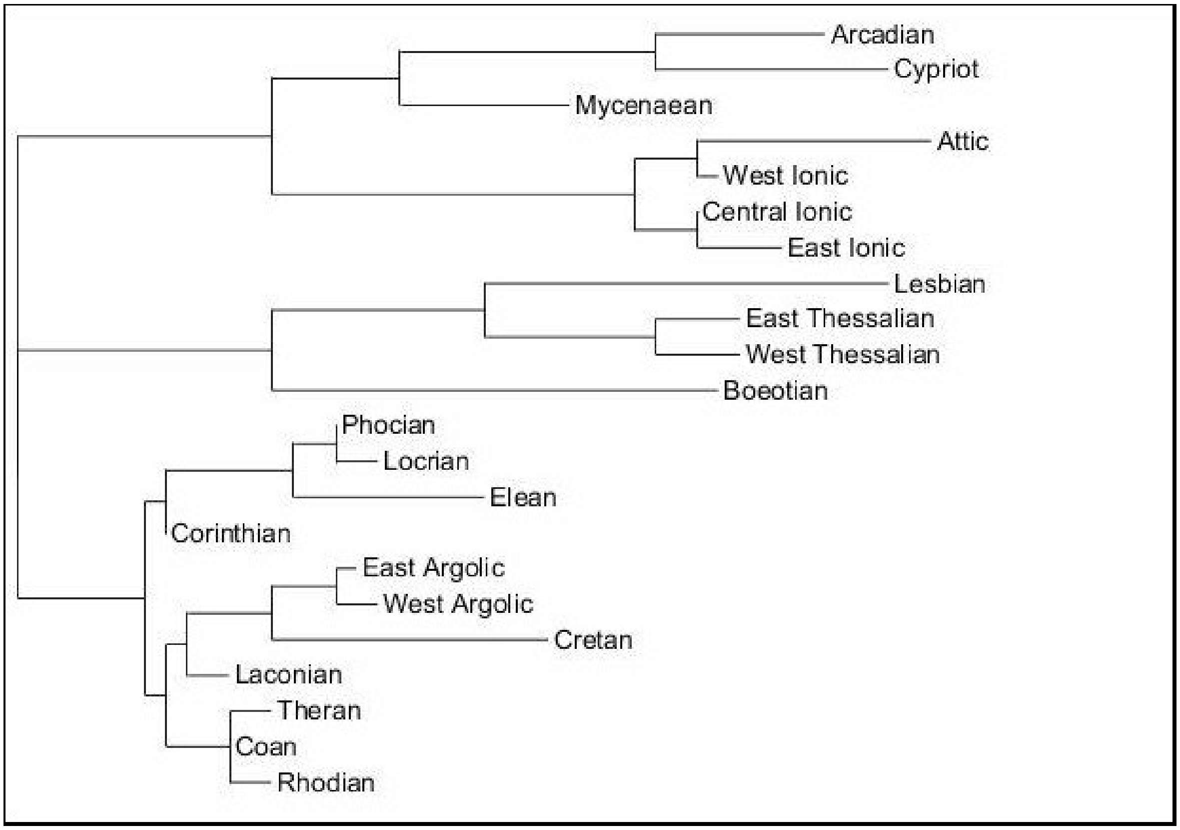


Figure 9: Tree 1-- Morphological Characters Weighted 2



The run with morphological characters weighted 10 produced 7 optimal trees, which only differ in the arrangement of the West Greek dialects. The tree length was 594, the CI excluding uninformative characters was 0.6038, the RI was 0.7892, and the rescaled CI was 0.5155.

Figure 10: Strict Consensus of 7 Trees-- Morphological Characters Weighted 10

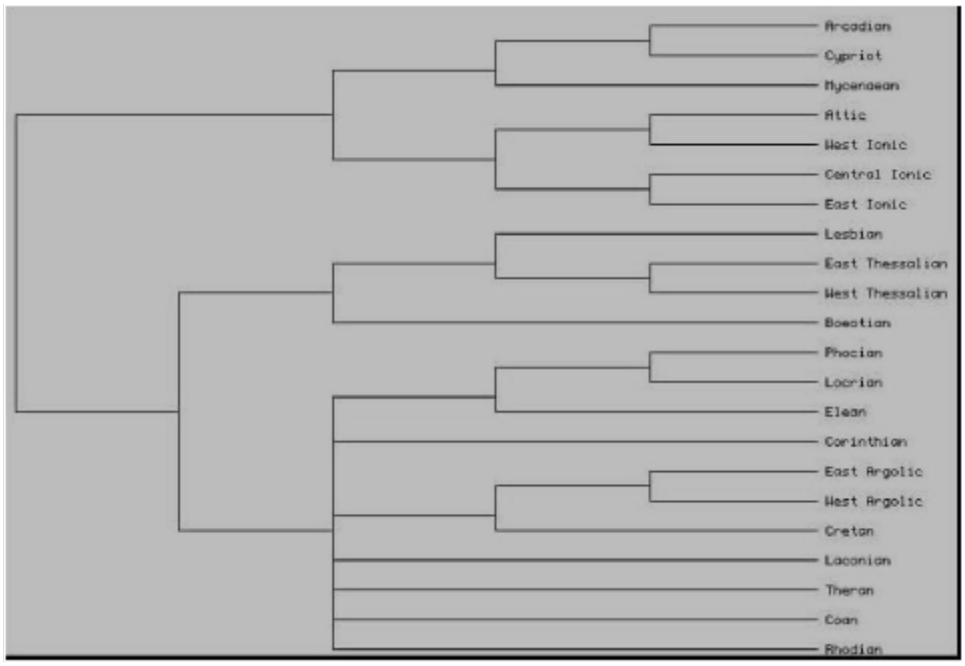
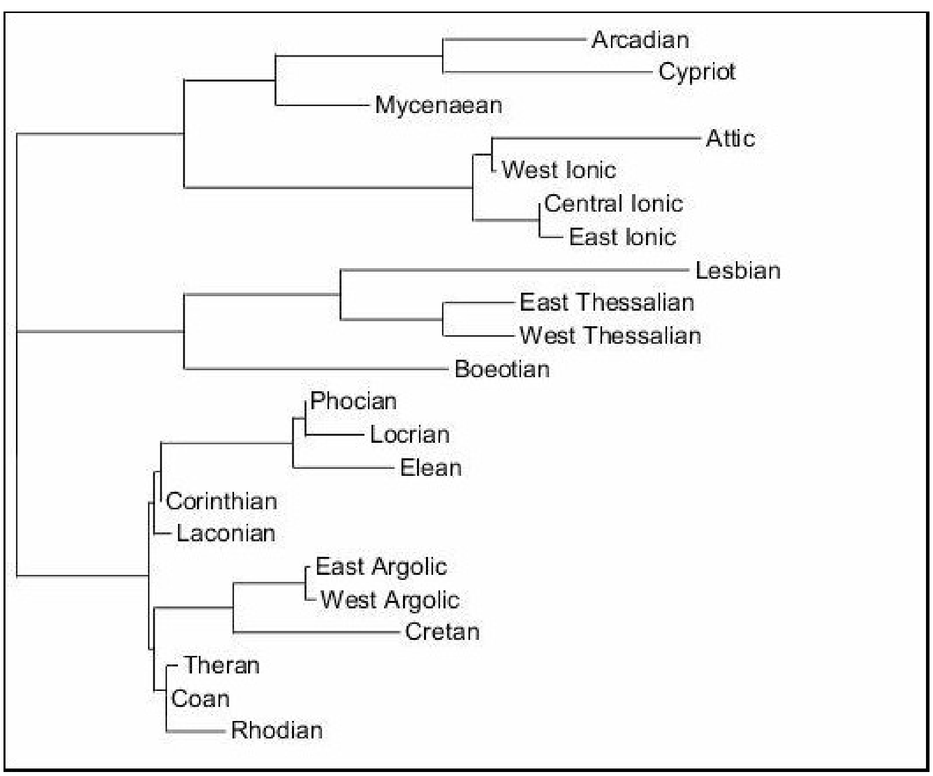


Figure 11: Tree 1-- Morphological Characters Weighted 10



Character Weighting—Lexical Characters

Again, I performed two runs, one with lexical characters given weight 2, and one with lexical characters given weight 10, and the other charatcter types given weight 1. I tried higher weights, but they produced results which were obviously wrong.

The run with lexical characters weighted 2 produced 7 optimal trees, which only differed in the arrangement of the West Greek taxa. These trees are basically correct as to the subclassification. The tree length was 213, the CI excluding uninformative characters was 0.5864, the RI was 0.7656, and the rescaled CI was 0.4816.

Figure 12: Strict Consensus of 7 Trees-- Lexical Characters Weighted 2

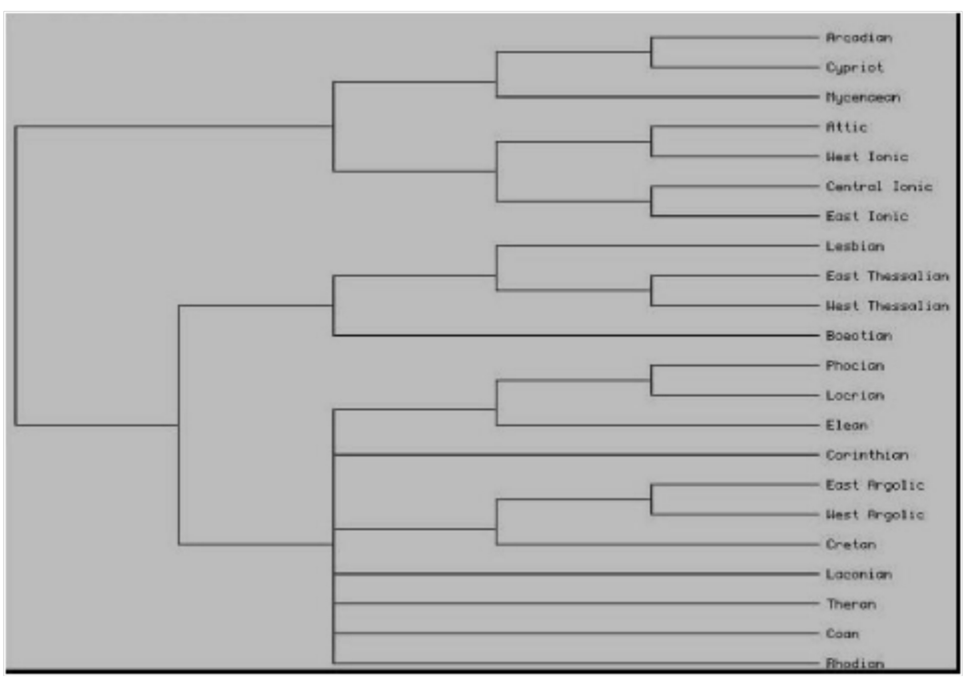
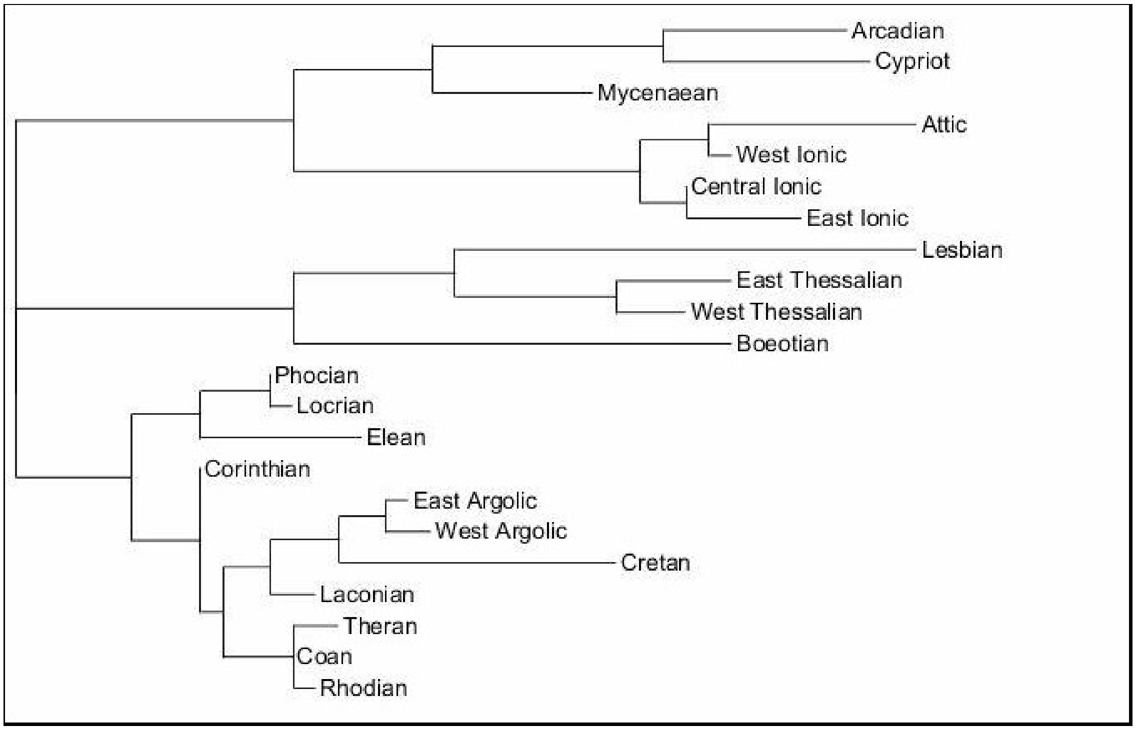


Figure 13: Tree 1-- Lexical Characters Weighted 2



The run with lexical characters given a weight of 10 produced 2 optimal trees, which only differ in the arrangement of the West Greek taxa. This particular tree is obviously wrong. The tree shows Lesbian as a sister taxon to Mycenaean and Arcado-Cypriot, and Elean as a sister taxon to the remainder of Aeolic. The tree length was 387, the CI excluding uninformative characters was 0.6418, the RI was 0.8186, and the rescaled CI was 0.5542.

Figure 14: Strict Consensus of 2 Trees-- Lexical Characters Weighted 10

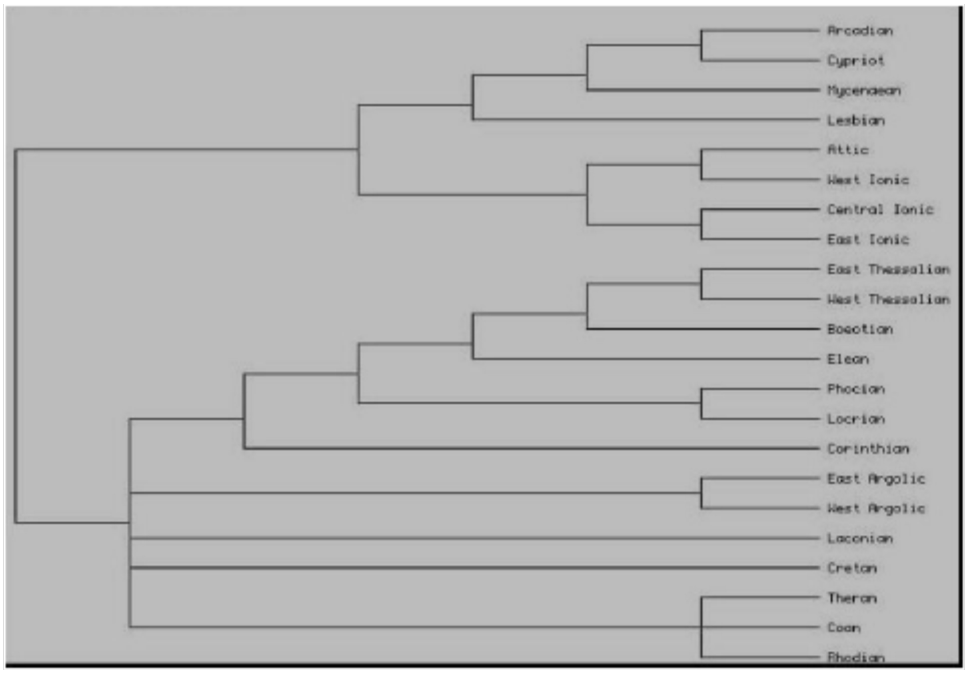
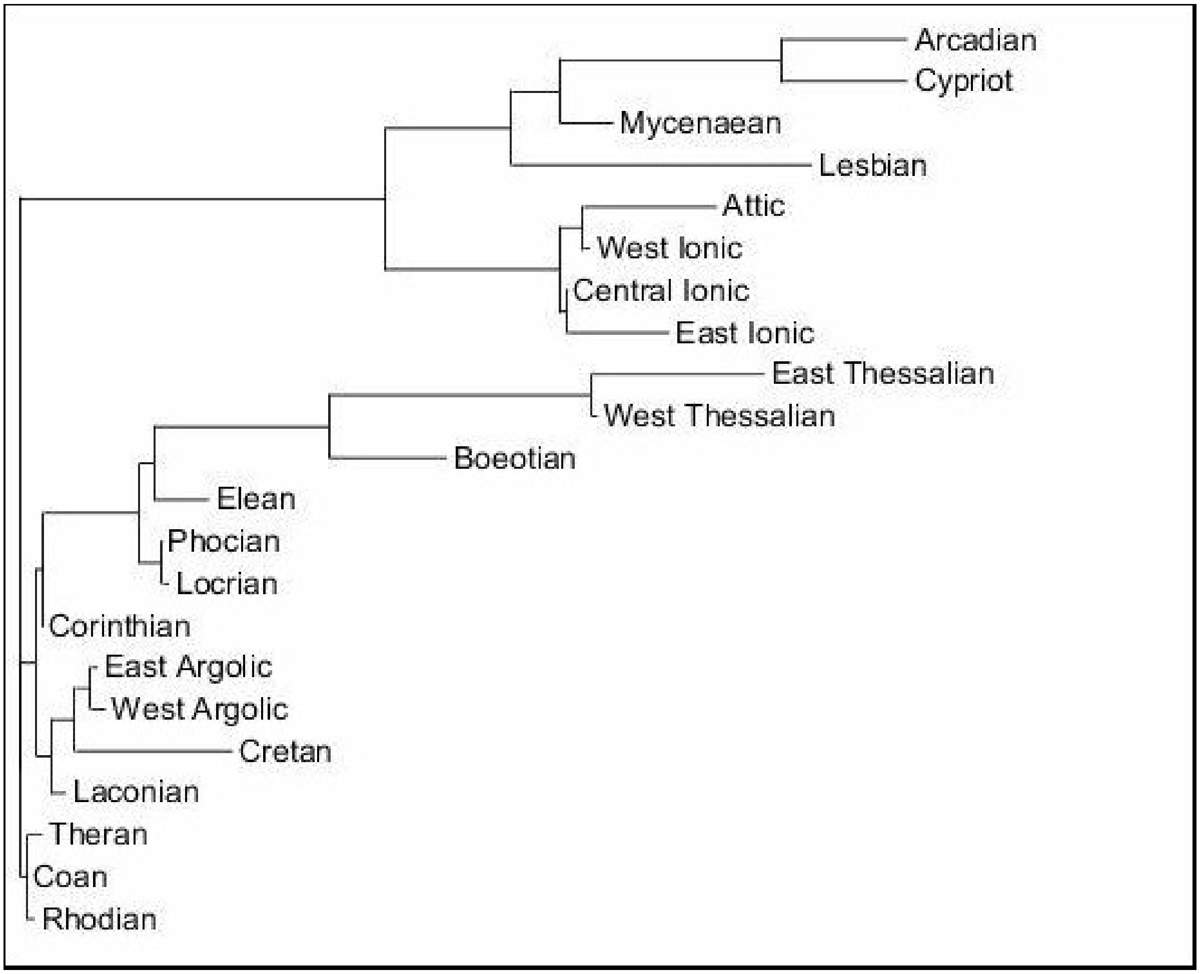


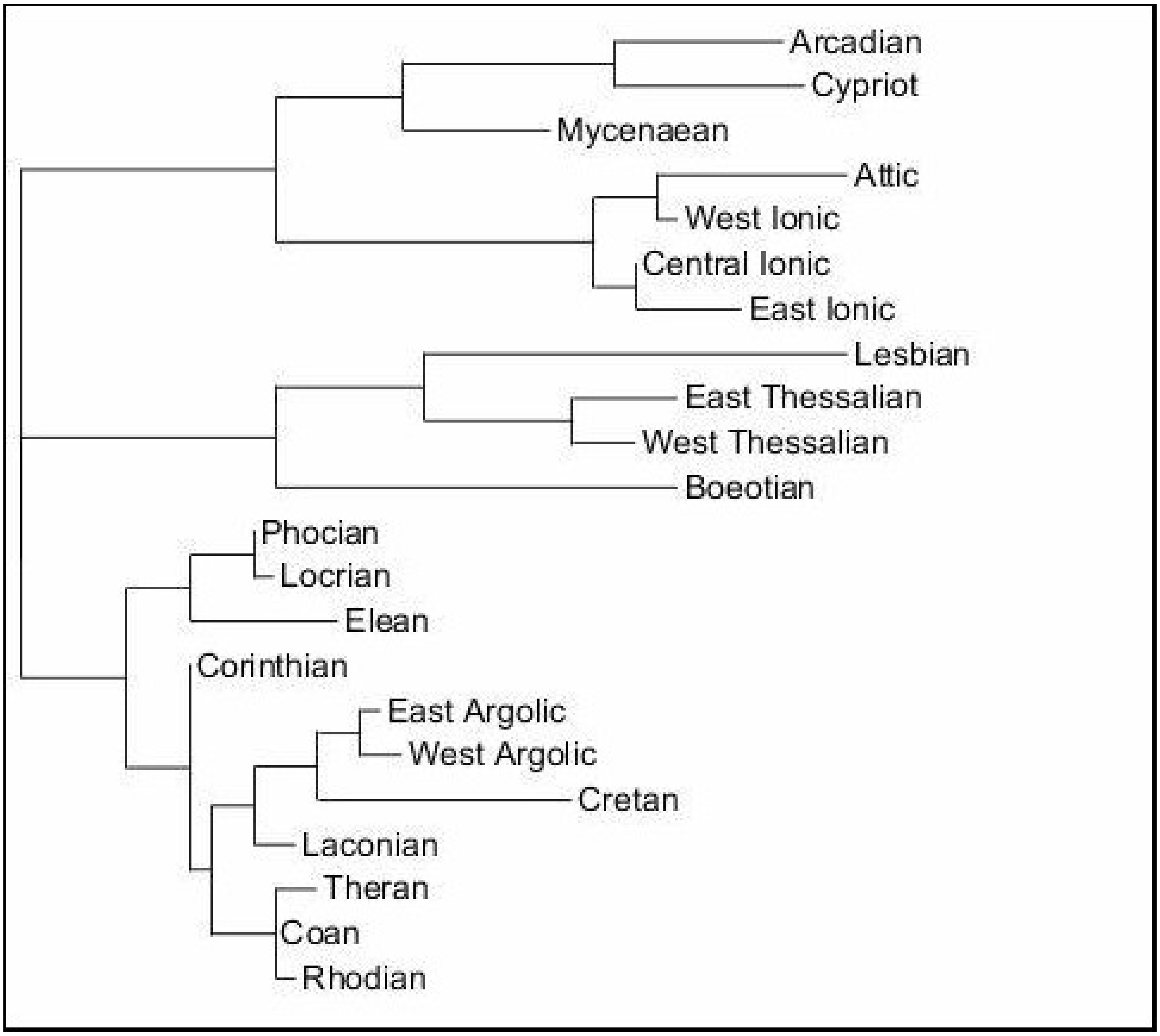
Figure 15: Tree 1-- Lexical Characters Weighted 10



Reweighting Characters Based on CI

I started out with a tree based on unweighted characters. I then reweighted characters based on their consistency index (otherwise using the default settings in PAUP*). This produced one optimal tree. The tree length was 118.91667, the CI excluding uninformative characters was 0.6940, the RI was 0.8338, and the rescaled CI was 0.6216.

Figure 16: Tree 1—Characters Weighted According to CI



<u>Analysis: Are Morphological, Phonological, or Lexical Characters More or Less Resistant to Borrowing?</u>

For the most part, character weighting did not improve the results of the phylogenetic analysis. Two character weighting schemes, those with phonological characters given a weight of 10 and those with lexical characters given a weight of 10, produced trees which were obviously wrong.

Weighting morphological characters, the type of character which is usually seen as most resistant to borrowing, did not change the tree topology or improve the resolution compared to the run with unweighted characters. The runs with phonological characters weighted 2 and with characters reweighted according to CI were the only runs which improved the resolution of the tree while still producing a plausible tree. These were also the runs which gave weight to the most characters, but it is unlikely that this alone is the reason for the result. After all, weighting a large number of misleading characters would tend to produce the wrong tree.

Character weighting did not substantially improve the CI or Rescaled CI over the analysis with unweighted characters, and even decreased the CI, in three cases. However, character weighting did significantly improve the RI. In only four cases, phonological characters weighted 2, phonological characters weighted 10, lexical characters weighted 10, and characters reweighted according to CI, the number of optimal trees (and, hence, the resolution of the tree) improved.

Table 1: Summary of Outcomes of Different Weighting Schemes

	Number of Optimal Trees	CI	RI	Rescaled CI
Unweighted	7	0.6243	0.5799	0.473
Phonological 2	2	0.6228	0.7402	0.461
Phonological 10	2	0.5842	0.7304	0.466
Morphological 2	7	0.5865	0.7669	0.4851
Morphological 10	7	0.6038	0.7892	0.5155
Lexical 2	7	0.5864	0.7657	0.4816
Lexical 10	2	0.6418	0.8186	0.5542
Reweighted CI	1	0.694	0.8338	0.6216

On the run with unweighted characters, roughly equal proportions of phonological, morphological, and lexical characters showed homoplasy.

<u>Consistency Indices of All Characters on the Unweighted Run</u>

As with the discussion of the consistency index as applied to trees above, CI is measured on a scale of 0 to 1, where 1 indicates a character which changed the minimum possible number of times on the tree, and numbers below that indicate characters which show some amount of homoplasy. U indicates "uninformative," a character which did not provide any information on the tree topology. C indicates "constant," a character which is the same across the entire tree.

Table 2: CI of Characters on Unweighted Run

Character	CI	U? C?
[1] 'Development of PIE syllabic liquids'	1	
[2] 'Development of PIE syllabic nasals'	0.41667	
[3] 'Long alpha fronted in Attic-Ionic'	1	
[4] 'Attic Reversion'	1	U
[5] 'Quantitative metathesis and prevocalic shortening'	0.65625	
[6] 'E raised to I before N in Arcado-Cypriot'	1	
[7] 'Final O raised to U in Arcado-Cypriot'	1	

[8] 'O raised to U in final syllables'	1	UC
[9] 'Pre-nasal O raised to U'	1	
[10] 'I lowered to E after R in Aeolic'	1	
[11] 'EI monophthongized and raised to I in Boeotian'	1	U
[12] 'AI monophthongized and raised to eta in Boeotian'	1	U
[13] 'OI monophthongized and raised to U in Boeotian'	1	U
[14] 'E raised to EI or I before a vowel in Boeotian'	1	U
[15] 'H raised to EI in Boeotian and Thessalian'	0.25	
[16] 'E lowered to A before R in Northwest Greek'	1	
[17] 'Eta lowered to long alpha in Elean'	1	U
	0.4090	
[18] 'E raised to EI or I before back vowels'	9	
[19] 'Contraction of A or long A plus E, EI or eta'	1	
[20] 'Loss of coda nasals and nasalization of vowels'	1	U
[21] 'Loss of initial pre-vocalic and pre-consonantal digamma'	0.2381	
[22] 'Loss of intervocalic digamma'	0.11111	
	0.0666	
[23] 'Loss of initial aspiration'	7	
[24] 'Loss of secondary intervocalic s'	1	U
[25] 'Development of the voiceless labiovelar kw before e'	1	
	0.5555	
[26] 'Development of *t.y'	6	
[27] 'Development of *t.#y'	0.11111	
[28] 'Development of *ky.'	1	U
[29] 'Development of *k#y'	0.2	
[30] 'Development of *dy, *gy'	0	
[31] 'Development of *ly'	1	U
[32] 'Development of *t.w'	0	

[33] 'First Compensatory Lengthening'	1	
[34] 'Second Compensatory Lengthening'	0.1	
	0.5333	
[35] 'Third Compensatory Lengthening'	3	
[36] 'Merger of the new long vowels'	0.2381	
[37] 'Rhotacism of final s'	1	U
[38] 'Assimilation of rs to rr'	0.0625	
[39] 'Gemination of R and other consonants before I'	1	
[40] 'Development of Z'	0	
[41] 'Dative plural ending of long A stems'	0.16	
[42] 'Masculine genitive singular long A stems'	1	U
[43] 'Genitive singular of thematic stems'	1	U
	0.5555	
[44] 'Dative Plural of Thematic Stems'	6	
[45] 'Dative singular of thematic stems'	0	
[46] 'Accusative singular of consonant stems'	1	U
[47] 'Dative plural of consonant stems'	0.5	
[48] 'Inflection of i-stem nouns'	1	
[49] 'Inflection of i-stem nouns (2)'	1	U
[50] 'Inflection of i-stem nouns (3)'	1	U
[51] 'First plural primary verbal ending in -mes or -men'	1	
[52] 'Third singular middle primary verbal ending'	1	
[53] 'Assimilation of ti > si in third singular active athematic primary	0.4444	
ending and prepositions'	4	
[54] 'Athematic inflection of contract verbs'	0.375	
[55] 'Formation of the future tense'	1	
	0.3333	
[56] 'Aorists and futures in verbs with -z-'	3	

	0.4285	
[57] 'Third plural active imperfect and aorist ending'	7	
	0.3333	
[58] 'Formation of the middle participle of verbs in -eO'	3	
[59] 'Formation of the perfect participle'	1	
	0.6428	
[60] 'Formation of athematic infinitives'	6	
[61] 'Case usage with two-case prepositions like apo and ex'	1	
[62] 'Order of particles'	1	
[63] Patronymics	0.11111	
	0.5714	
[64] 'Modal particle'	3	
[65] 'Emphatic particle'	0.4375	
[66] 'Copulative particle'	1	
[67] 'Conditional conjunction'	1	
[68] 'Copulative conjunction'	1	
[69] 'The preposition ana'	0.375	
[70] 'The preposition dia'	1	
	0.1562	
[71] 'The preposition "with"'	5	
[72] 'The preposition "towards" (1)'	0.2381	
	0.3333	
[73] 'The preposition "towards" (3)'	3	
[74] 'The preposition "with" (2)'	1	U
[75] 'The preposition "in"'	0.2381	
[76] 'The preposition "from"'	0.4	
[77] 'The adjective "holy"'	1	X
[78] 'The noun "Zeus"'	1	X

[79] 'The noun "Apollo"'	1	X
[80] 'The numeral "one"'	0.375	
[81] 'The numeral "twenty"'	0.4	
[82] 'Second singular personal pronoun'	0.4	
[83] 'First and second person plural pronouns'	1	
[84] 'Nominative plural of the article'	0.2381	
[85] 'Dissimilation of kw to k in pronominal forms'	1	U
[86] 'Indefinite pronoun'	1	
[87] 'Deictic pronoun'	1	
[88] 'Temporal adverbs'	1	

Table 3: Percentage of Homoplastic Characters for Each Character Type

	Informative	Nonhomoplastic	Homoplastic	% Homoplastic
Phonological	28	11	17	61%
Morphological	15	5	10	67%
Lexical	20	8	12	60%

The differences between these percentages are very close.

Furthermore, there is no obvious pattern as to which characters show homoplasy and which do not, even within the same class.

Table 4: Homoplasy in Phonological Characters

Nonhomoplastic Characters	Homoplastic Characters
[6] 'E raised to I before N in Arcado-Cypriot'	[15] 'H raised to EI in Boeotian and Thessalian'
[7] 'Final O raised to U in Arcado-Cypriot'	[18] 'E raised to EI or I before back vowels'
[10] 'I lowered to E after R in Aeolic'	[21] 'Loss of initial pre-vocalic and pre-consonantal digamma'
[16] 'E lowered to A before R in Northwest Greek'	[22] 'Loss of intervocalic digamma'
[19] 'Contraction of A or long A plus E, EI or eta'	[23] 'Loss of initial aspiration'
[25] 'Development of the voiceless labiovelar kw before e'	[26] 'Development of *t.y'
[33] 'First Compensatory Lengthening'	[27] 'Development of *t.#y'
[39] 'Gemination of R and other consonants before I'	[29] 'Development of *k#y'
	[30] 'Development of *dy, *gy'
	[32] 'Development of *t.w'
	[34] 'Second Compensatory Lengthening'
	[35] 'Third Compensatory Lengthening'
	[36] 'Merger of the new long vowels'
	[38] 'Assimilation of rs to rr'

[40] 'Development of *Z*'

Table 5: Homoplasy in Morphological Characters

Nonhomoplastic Characters	Homoplastic Characters
[48] 'Inflection of i-stem nouns'	[41] 'Dative plural ending of long A stems'
[51] 'First plural primary verbal ending in -mes or -men'	[44] 'Dative Plural of Thematic Stems'
[52] 'Third singular middle primary verbal ending'	[45] 'Dative singular of thematic stems'
[55] 'Formation of the future tense'	[47] 'Dative plural of consonant stems'
[59] 'Formation of the perfect participle'	[53] 'Assimilation of ti > si in third singular active athematic primary ending and prepositions'
	[54] 'Athematic inflection of contract verbs'
	[56] 'Aorists and futures in verbs with -z-'
	[57] 'Third plural active imperfect and aorist ending'
	[58] 'Formation of the middle participle of verbs in -eO'
	[60] 'Formation of athematic infinitives'

Table 6: Homoplasy in Lexical Characters

Nonhomoplastic Characters	Homoplastic Characters
[66] 'Copulative particle'	[64] 'Modal particle'
[67] 'Conditional conjunction'	[65] 'Emphatic particle'
[68] 'Copulative conjunction'	[69] 'The preposition ana'
[70] 'The preposition dia'	[71] 'The preposition "with"'

[83] 'First and second person plural pronouns'

[86] 'Indefinite pronoun'

[87] 'Deictic pronoun'

[88] 'Temporal adverbs'

[72] 'The preposition "towards" (1)'

[73] 'The preposition "towards" (3)'

[75] 'The preposition "in"'

[76] 'The preposition "from"'

[80] 'The numeral "one"'

[81] 'The numeral "twenty"'

[82] 'Second singular personal pronoun'

[84] 'Nominative plural of the article'

Conclusions

The Greek dialect data does not support the assertion that character weighting improves phylogenetic accuracy. It is true that weighted morphological characters performed better than weighted phonological or lexical characters, but weighted morphological characters did not perform substantially better than no weighting at all. An analysis of the numbers of homoplastic phonological, morphological, and lexical characters concluded that these character types did not show different levels of homoplasy. Weighting characters according to their CI improved the resolution of the unresolved portion of the tree, the West Greek dialects. However, in the absence of outside evidence confirming how the West Greek dialects developed over time, it is impossible to produce any outside proof showing that the phylogenetic analysis did, in fact, recover the correct tree. Resolution is a measure of precision, and as such, it does not necessarily indicate accuracy.

Prior studies of character weighting (Barbançon forthcoming, Nakhleh et al. 2005b) have studied the development of language families and unrelated languages, where we might expect strong structural constraints against certain types of borrowing. However, the Greek dialect data may show that between dialects of the same language, these structural constraints are weaker or nonexistent, thus rendering different types of characters roughly equally likely to be borrowed.

Chapter 4: Phylogenetic Network Models

Introduction

Phylogenetic network models, most prominently NeighborNet, are popular in phylogenetic analyses of linguistic data for determining evolutionary relationships when the relationships between the dialects or languages in question are expected to be non-treelike (e.g. Holden and Gray 2006, Bowern 2012).

We would expect various aspects of the relationships among the Greek dialects to be non-treelike, as well. For example, many of the Maximum Parsimony analyses of the Greek dialects had difficulty resolving the relationships among the West Greek dialects. Various scholars have proposed extensive borrowing between Boeotian and West Greek, and Lesbian and Attic-Ionic (Colvin 2007, 40-41). Thus, we might expect network models to be useful at revealing non-treelike relationships among the Greek dialects.

However, no one has yet tested network models to see how well they perform on linguistic data. Studies of the effectiveness of network models on phylogenetic data have shown that NeighborNet has an excellent false negative rate (less than 5%) on trees, and a very good false negative rate (less than 10%) on networks with one reticulation, with a long enough sequence, but a very poor false positive rate on both trees and networks with one reticulation, even when the sequences are very long (Nakhleh et al. 2005a, 345; see the paper for a discussion of the sequence lengths they used). In this context, a false positive represents a bipartition which was recovered, but is incorrect, and a false negative represents a bipartition which was correct, but was not recovered. A reticulation is a branch which connects two nodes which are not sisters. In other words, NeighborNet tends not to miss bipartitions which should appear, but tends to produce many bipartitions which should not appear. The most fundamental problem with using NeighborNet for linguistic data is that NeigborNet is an implicit, rather than an explicit phylogenetic network. In an explicit phylogenetic network, additional contact edges specifically model borrowing events; in other words, lines are drawn between nodes to represent where contact events took place. Implicit networks, on the other hand, seek to graphically represent how the input data differs from a phylogenetic tree. Thus, the representations of nontreelike information could represent contact events, but they could also represent parallel evolution (unrelated development of the same character state more than once), backmutation (changing from one character state to the previous character state on the evolutionary tree), or noise (unexplaned random variation) (Nichols and Warnow 2008, 763-764). The original authors of NeighborNet admit that the interpretation of the splits graphs produced by NeighborNet can be problematic, and therefore only advocate for its use as a tool for data representation and data exploration; the framework for interpreting splits graphs does not yet exist (Bryant and Moulton 2004, 265).

Therefore, this analysis aims to test how well phylogenetic network models perform at uncovering both treelike and non-treelike relationships. First, it will analyze the phylogenetic data matrix in detail to determine where unrelated dialects or dialect groups show higher than normal amounts of shared features. Then, it will run a network analysis of the phylogenetic data matrix using NeighborNet. Finally, it will compare the results of the borrowing analysis to the NeighborNet analysis in order to determine how well NeighborNet can detect borrowing between the different taxa.

Treelike and Non-Treelike Behavior Among the Greek Dialects

The basis for determining treelike behavior among the Greek dialects will be the Maximum Parsimony analysis with characters reweighted by CI.

For non-treelike behavior, I computed pairwise distances between the different dialects (the number of character state changes between two dialects), based on the phylogenetic data matrix. I then analyzed the pairwise distances between dialects of different dialect groups (that is, excluding the pairwise distances between dialects of the same dialect group). I checked for normality using the Shapiro-Wilk test. The data were not normal, p = .009898. I then calculated the median, first quartile, and third quartile (see table below).

Table 7: Analysis of Pairwise Distances Between Dialects

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
35.00	45.75	49.00	48.76	53.00	60.00

Pairs of dialects with pairwise distances above the third quartile were taken to have exceptionally high levels of borrowing, while those with pairwise distances below the first quartile were taken to have exceptionally low amounts of borrowing. Mycenaean was excluded from the calculation because it appears to be an outlier in the number of similarities it shares with other dialects, most likely because of its early date. The chart is shown below.

Table 8: Pairwise Distances Between Dialects

	Arcadian	Cypriot	Attic	W. Ionic	C. Ionic	E. Ionic	Lesbian	E. Thessalian	W. Thessalian	Boeotian
Arcadian										
Cypriot										
Attic	47	40								
W. Ionic	52	45								
C. Ionic	51	46								
E. Ionic	46	42								
Lesbian	50	43	47	52	54	51				
E. Thessalian	52	48	35	39	41	36				
W. Thessalian	54	49	38	42	44	39				
Boeotian	41	36	36	39	37	35				
Phocian	45	43	47	51	53	49	44	48	52	54
Locrian	46	44	47	51	53	49	43	48	52	55
Elean	47	41	41	46	46	45	41	45	50	55
Megarian	48	44	52	56	58	53	49	50	55	53

Corinthian	50	48	51	56	58	53	47	50	55	54
E. Argolic	48	41	51	55	57	52	50	49	54	51
W. Argolic	49	40	48	53	56	51	49	49	54	50
Laconian	47	45	49	53	55	50	47	51	57	54
Cretan	40	35	48	52	51	51	47	43	49	56
Theran	49	41	51	56	57	54	49	48	53	53
Coan	47	43	51	56	59	56	49	48	53	53
Rhodian	48	44	52	57	60	55	50	48	53	53

The pairs of unrelated dialects which have elevated levels of shared features, then, are the following:

Table 9: Pairs of Unrelated Dialects with Elevated Levels of Shared Features

Arcadian and West Thessalian

West Ionic with non-NW Greek West Greek, except Cretan

Central Ionic with Lesbian

Central Ionic with West Greek, except Elean and Cretan

East Ionic with Megarian and Corinthian

East Ionic with Island Doric

West Thessalian with non-NW West Greek, except Island Doric

Boeotian with all West Greek except East and West Argolic

We might then expect the network analysis to show a certain level of affinity between these unrelated dialects.

Network Models

There are two network models to consider, Split Decomposition (Bandelt and Dress 1992) and NeighborNet. However, Split Decomposition does not work well with more than a handful of taxa (Swofford et al. 1996, 492), so it is not appropriate to use here. NeighborNet basically works by using a distance matrix to create a set of weighted splits, and then using those splits to create a splits graph (Bryant and Moulton 2004, 255).

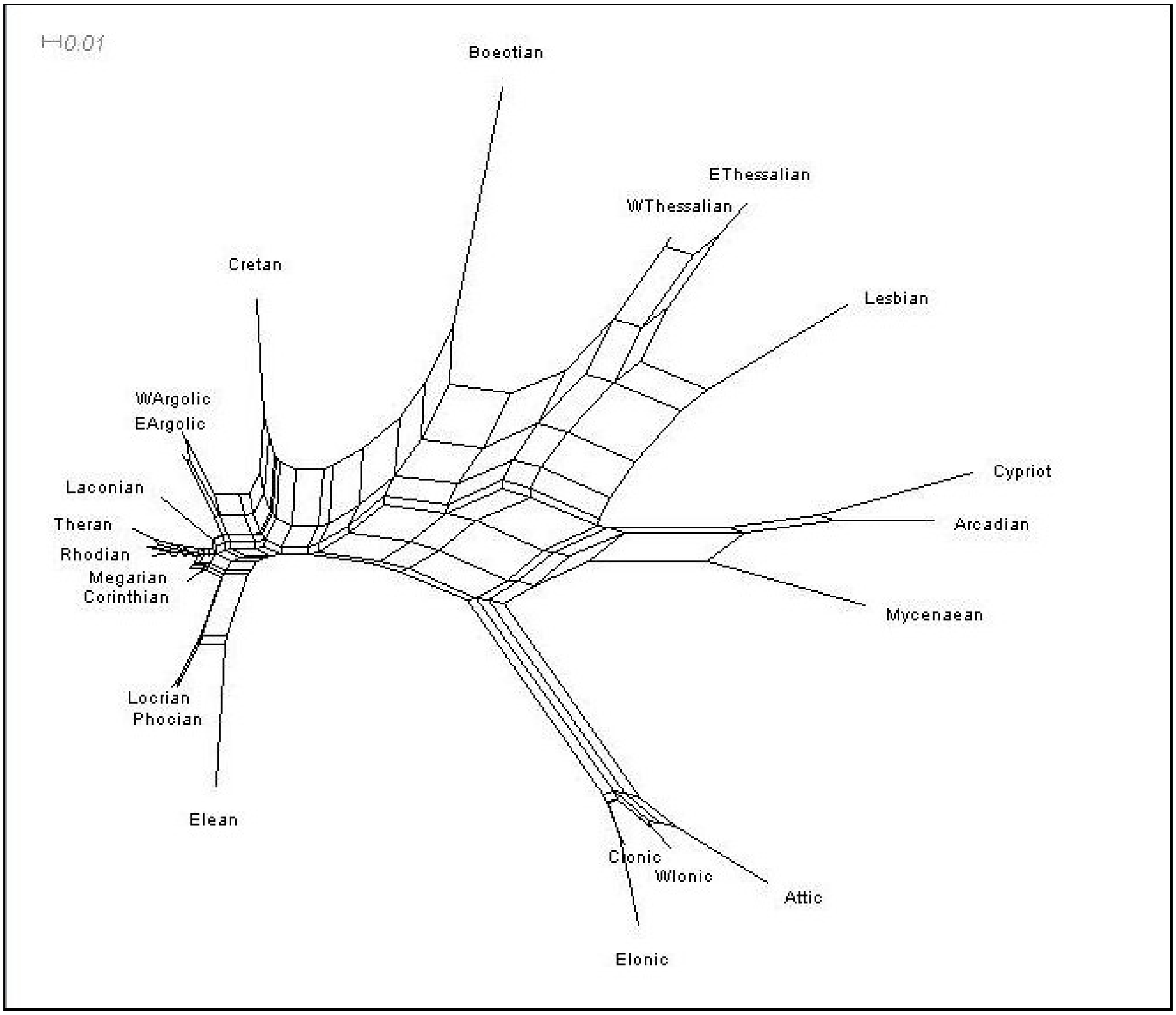
Methods-- NeighborNet

Phylogenetic Data Matrix

As with the analysis in the previous chapter, the phylogenetic data matrix consists of 40 phonological, 20 morphological, 3 syntactic, and 22 lexical characters. Unlike the previous analysis, however, there are 23 taxa, since this analysis includes Megarian. Pamphylian was also omitted from the analysis because it most likely represents a mixed dialect, a problem outside the scope of this analysis. The analysis was carried out using SplitsTree4, version 4.11.3 (Huson and Bryant 2006), also under the default settings.

The results of the NeighborNet analysis are as follows.

Figure 17: NeighborNet Splits Graph of the Greek Dialects



The interpretation of this splits graph is probably not intuitive, so how to read a splits graph deserves some explanation. Sets of parallel lines represent splits separating one group from another group. Since this splits graph is relatively complex and difficult to read, I will highlight the most interesting and significant splits.

First, we can identify each of the major dialect groups: Arcado-Cypriot and Mycenaean, Attic-Ionic, Aeolic, and West Greek.

Figure 18: Major Dialect Groups in the NeighborNet Splits Graph of the Greek Dialects

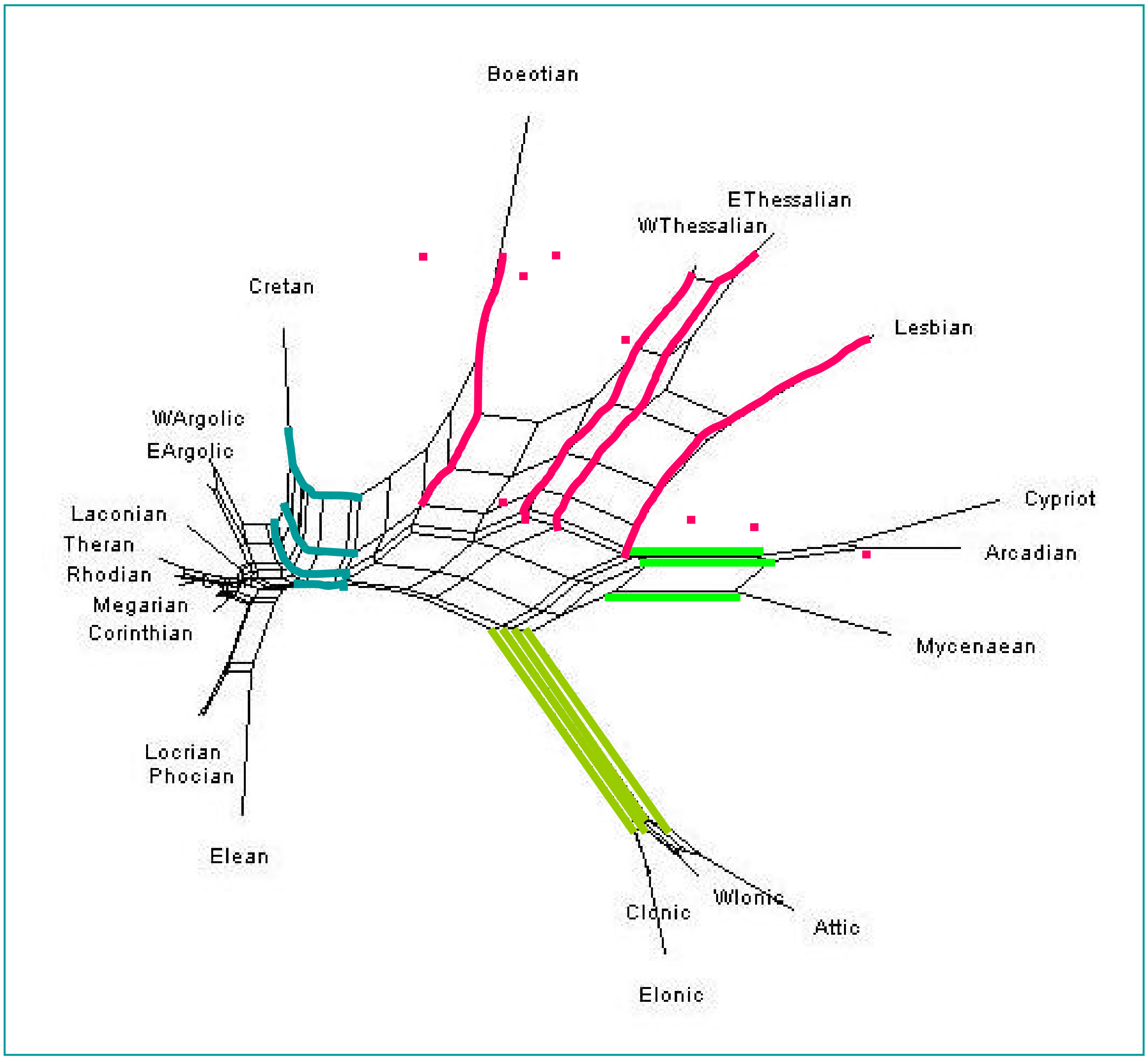
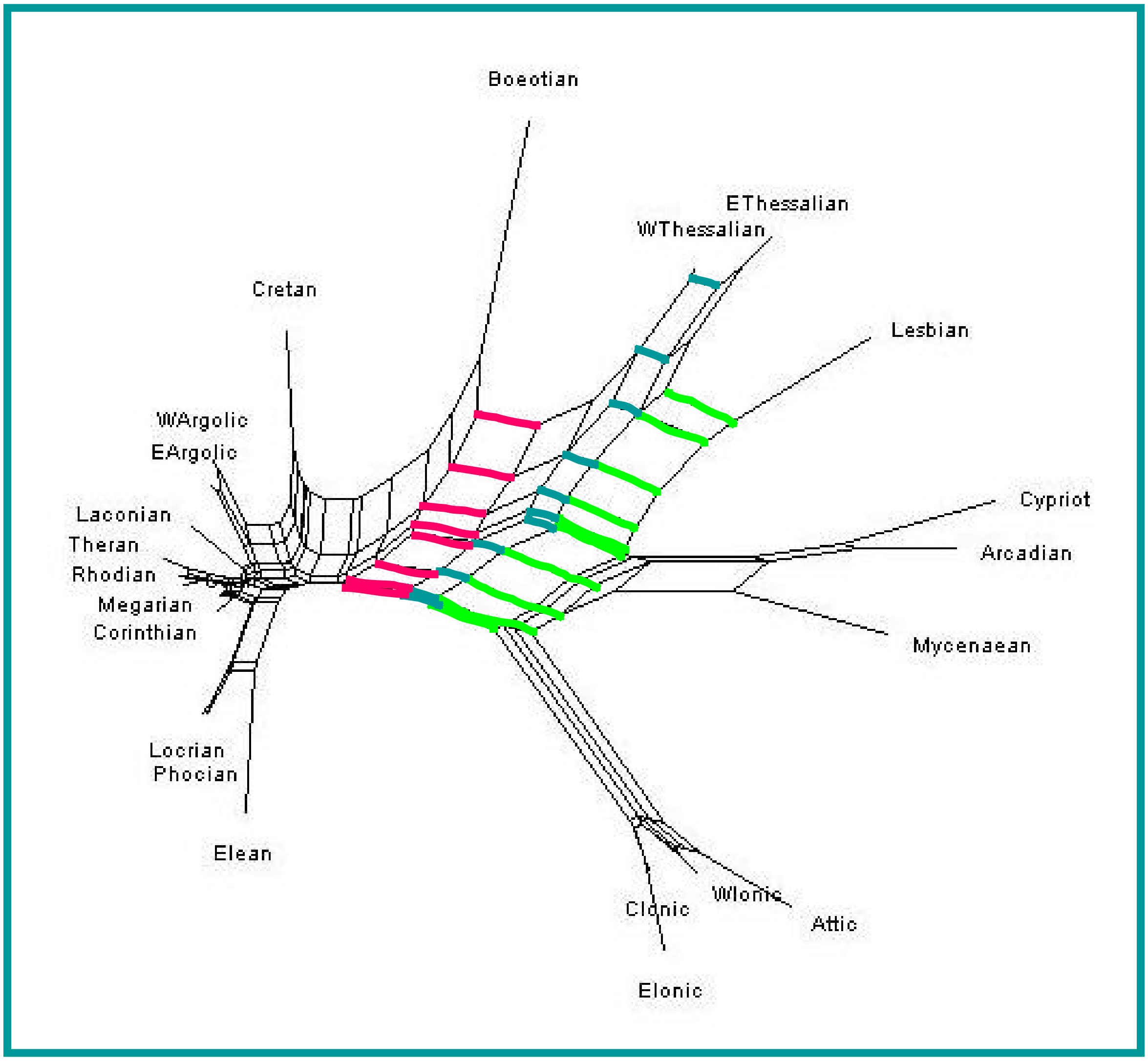
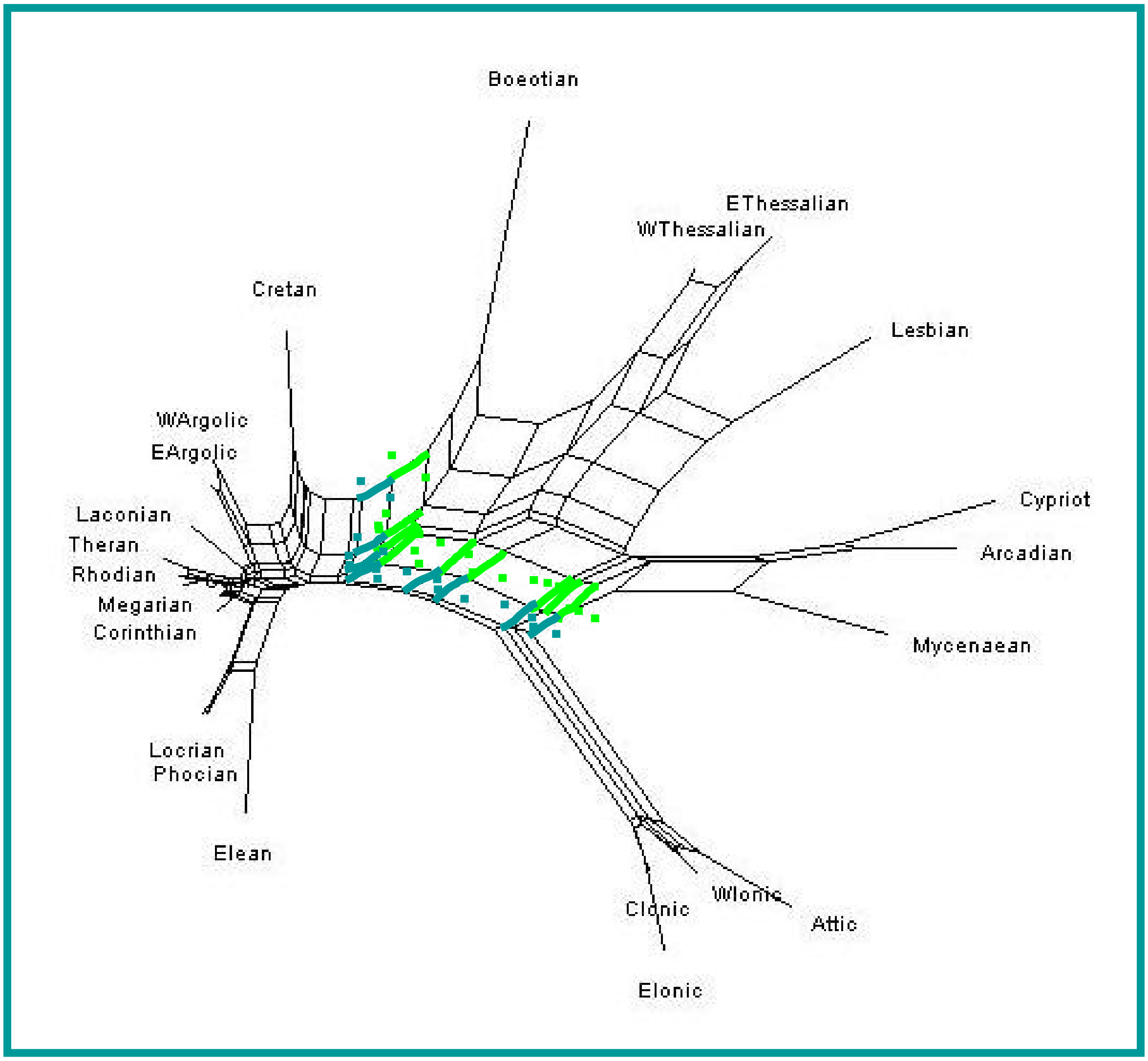


Figure 19: Additional Splits involving the Aeolic Dialects



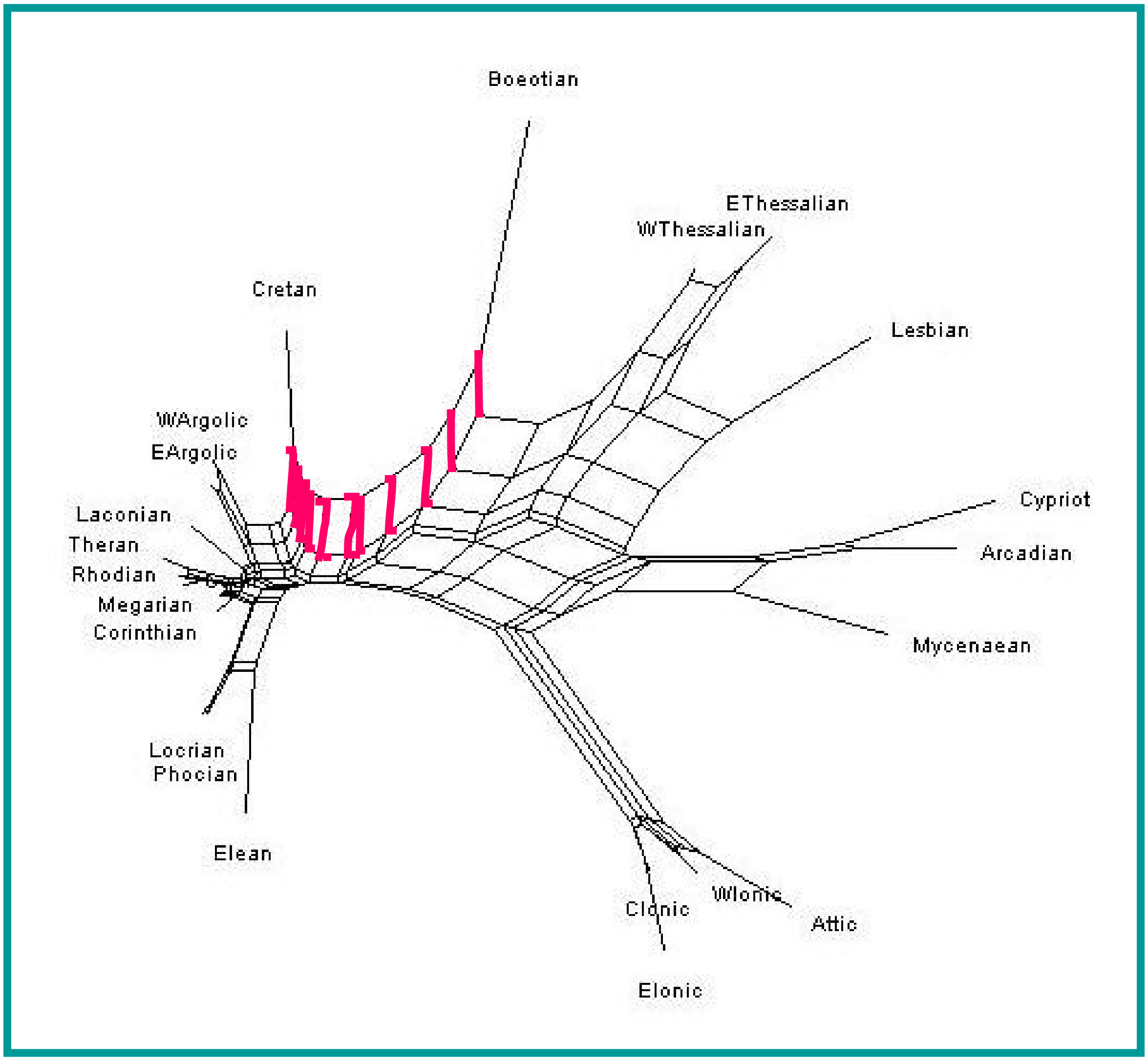
We can also identify splits grouping Lesbian with Attic-Ionic and Arcado-Cypriot; Thessalian, Lesbian, Arcado-Cypriot and Attic-Ionic; and Boeotian and West Greek (shown by the bipartition indicated in pink).

Figure 20: Additional Splits Involving the Major Dialect Groups



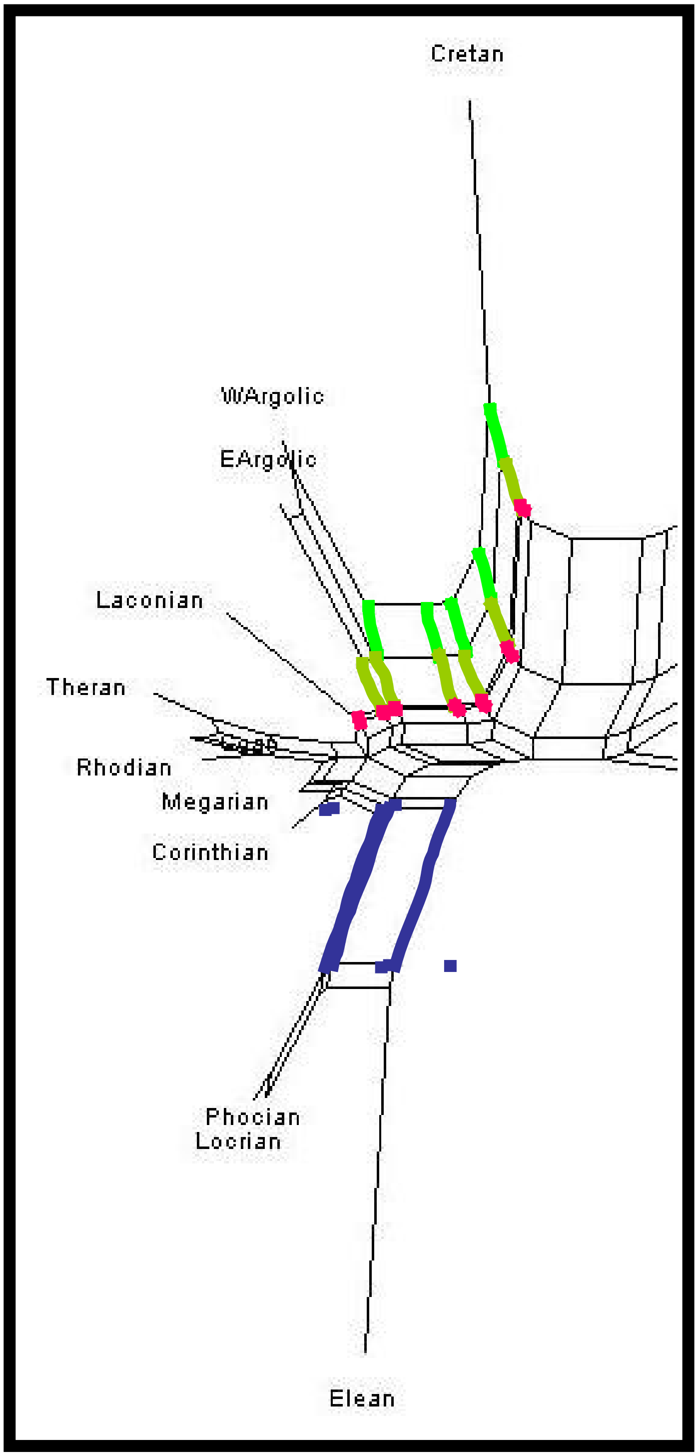
There is also a split between Arcado-Cypriot (excluding Mycenaean) and Aeolic, and then a split between Aeolic and Arcado-Cypriot and Mycenaean, and everything else.

Figure 21: Split Grouping Together Cretan and Boeotian



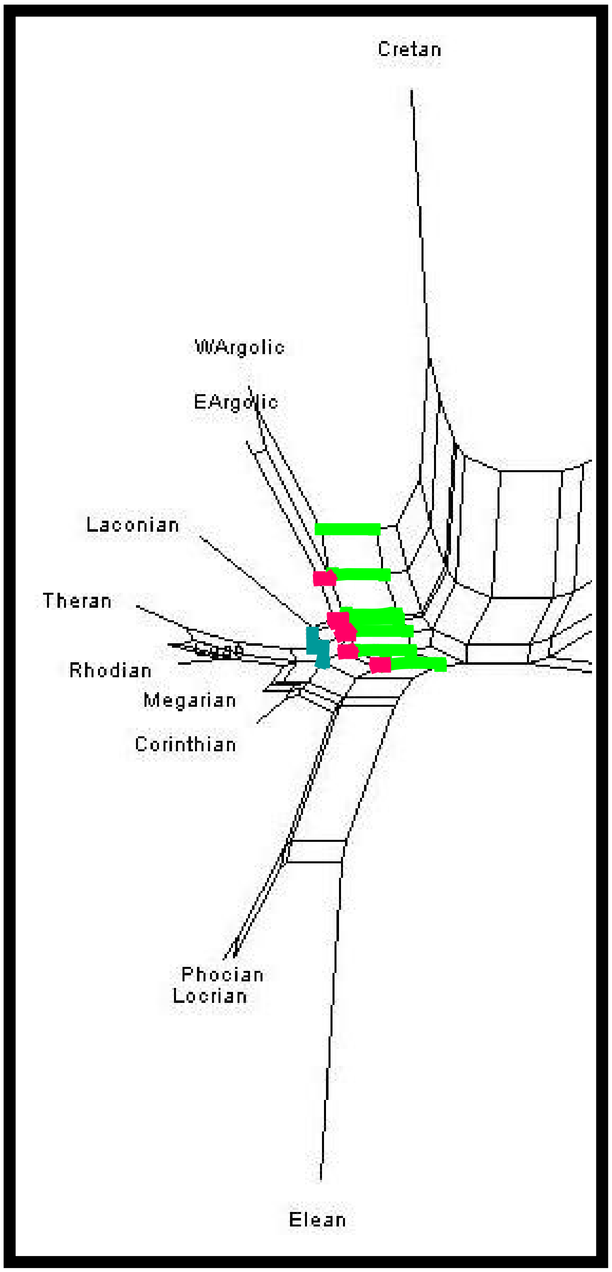
There is a split between Cretan and Boeotian, and everything else.

Figure 22: Splits Among the West Greek Dialects



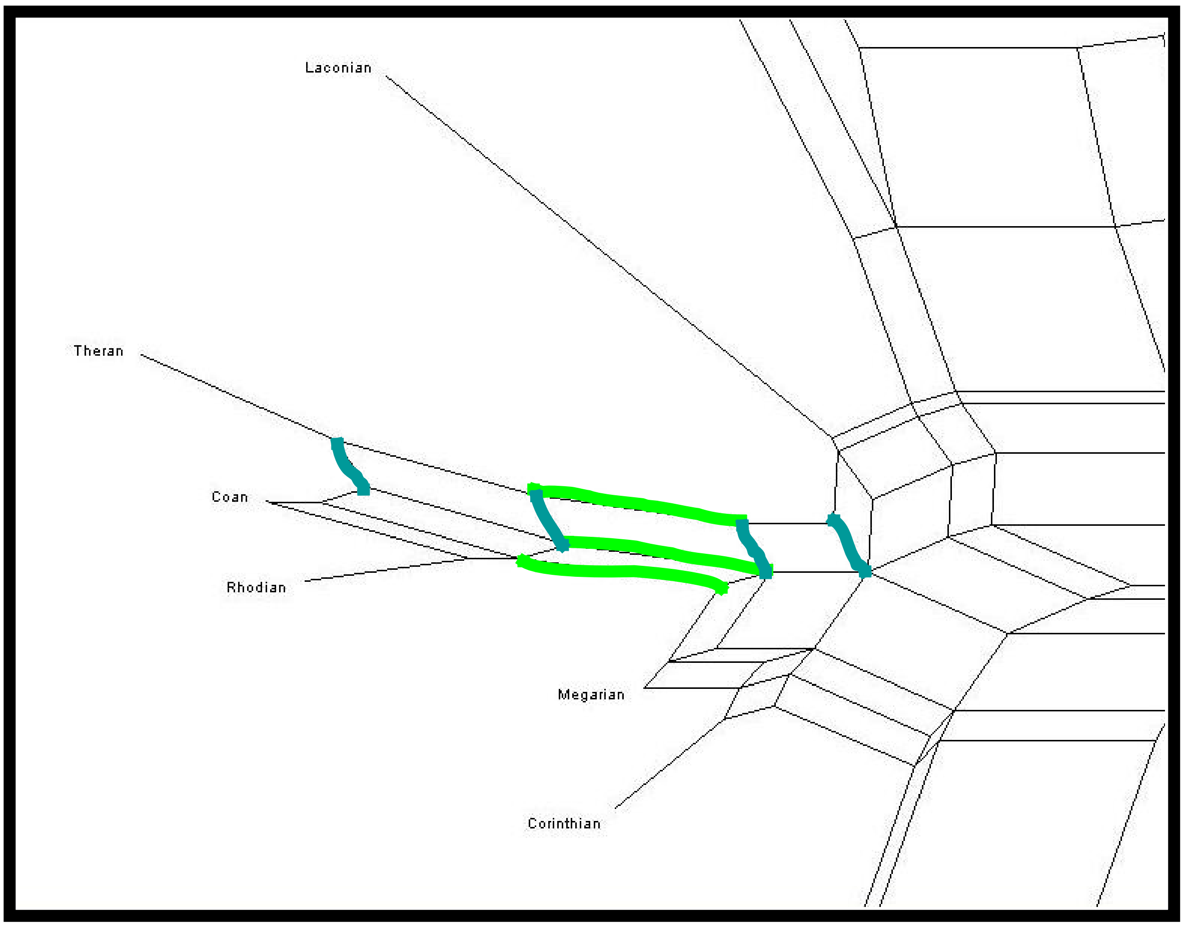
Within West Greek, there is a split separating Northwest Greek, and Elean is separated from Phocian and Locrian. A slit separates Cretan plus West Argolic; Argolic plus Cretan; and Cretan, Argolic, and Laconian.

Figure 23: Additional Splits Among the West Greek Dialects



There is another split separating Argolic, Laconian, Island Doric, Corinthian, and Megarian; another one with East Argolic, Laconian, Island Doric, Corinthian, and Megarian; and one defining Island Doric.

Figure 24: Splits Involving the Non-NW Greek West Greek Dialects



There is a split defining Island Doric. Then, within Island Doric, there is a split separating Coan and Rhodian from Theran.

The arrangement of the West Greek dialects will be discussed further in the concluding chapter.

Conclusions

The NeighborNet analysis largely agrees with the Maximum Parsimony analysis when it comes to reconstructing the treelike structure. NeighborNet correctly identifies Arcado-Cypriot plus Mycenaean, Attic-Ionic, Aeolic, and West Greek, and groups Arcado-Cypriot plus Mycenaean and Attic-Ionic closer together than it does Aeolic and West Greek. With the exception of West Greek, the groupings of taxa within each of the major dialect groups are the same.

However, there is almost no agreement between the Maximum Parsimony borrowing analysis and the additional splits in the NeighborNet analysis. As we can see from the table below, even though the MP borrowing analysis identified eight instances of borrowing, and the NeighborNet analysis identified seven instances of borrowing, the two analyses only agreed on two or maybe three of them, and even then, not completely.

Table 10: Agreement Between MP Borrowing Analysis and NeighborNet Analysis

MP Borrowing Analysis	NeighborNet Analysis
No, and Lesbian + Cypriot showed low levels of borrowing	Lesbian plus Arcado-Cypriot
Arcadian and West Thessalian	Yes, if we include Lesbian, East Thessalian, and Cypriot
No, only Arcadian and West Thessalian	Arcado-Cypriot and Aeolic
No, only Arcadian and West Thessalian	Mycenaean plus Arcado-Cypriot and Aeolic
West Ionic with non-NW Greek West Greek, except Cretan	No
Central Ionic with Lesbian	No

Central Ionic with West Greek, except	
Elean and Cretan	No
East Ionic with Megarian and	
Corinthian	No
East Ionic with Island Doric	No
West Thessalian with non-NW West	Yes, if we include NW Greek and Island
Greek, except Island Doric	Doric
Boeotian with all West Greek except	
East and West Argolic	Yes, if we include East and West Argolic
No	Boeotian and Cretan

Some of this disagreement likely stems from the different natures of the two analyses. The MP borrowing analysis is set up to make pairwise comparisons between taxa, and so is more poorly equipped to show comparisons between taxon groups. The NeighborNet analysis can only show borrowing through bipartitions, and so cannot show borrowing between two taxa if their two subgroups do not share a bipartition. For example, NeighborNet can show borrowing between Boeotian and one or more West Greek dialects because there is a bipartition which lumps Boeotian together with West Greek. However, it could not show borrowing between Lesbian and any dialect of Attic-Ionic, because there is no bipartition which groups Lesbian together only with Attic-Ionic. This is one strike against NeighborNet, since it is important to know whether borrowing has occurred between taxa in such situations.

The other major problem with NeighborNet is the difficulty in reading the resulting splits graphs. The splits graph was large enough that it was sometimes difficult to ascertain whether a given set of lines was truly parallel, and the graph was complex enough that it was impossible to be sure that I had identified all of the splits.

Overall, it seems that while NeighborNet is certainly appropriate as a first pass for gaining a basic level of familiarity with the data, it is a poor tool for a more comprehensive, detailed, and accurate assesment of where borrowing has occurred.

Chapter 5: Preliminary Cluster Analysis

Introduction

One fundamental assumption of phylogenetic analysis is that the data are treelike. Violation of this assumption may lead to results which are inaccurate or poorly resolved. Thus, instead of approaching the lack of resolution in the West Greek dialects as a problem to be solved through improvements in the phylogenetic methods, it may be more appropriate to work to improve the quality of the input data. More specifically, it seems worth reopening the question of whether the West Greek taxa as defined in the phylogenetic analysis do represent discrete dialects, rather than parts of a single dialect. There are two separate issues to address. The first is to understand the original historical linguistic basis for defining the West Greek dialects. The second is to apply an analytical approach to the problem using clustering methods, specifically Multidimensional Scaling.

Linguistic Basis for the Greek Dialects

There seems to be no explicit discussion of what basis was used to partition West Greek into individual dialects, but there are clearly grounds for considering an alternate division of the West Greek dialects.

A thorough discussion of the basis for dividing up the Greek dialects is lacking. Colvin (2007, 22) notes that the general classification of the dialects is “more or less inherited from the Greeks, and is therefore based on non-linguistic (cultural, political) as well as linguistic factors.” Buck discusses the individual Doric dialects in terms of geography and historical and political relations (1955, 12-14), and then in terms of their defining dialectal features (161-172). However, it is not clear from the text whether the original basis for classification was the geographical, historical, and political factors or the linguistic ones.

Buck's linguistic discussion of Cretan (1955, 171-172) offers more insight into the line of reasoning which was used to define the individual dialects. Buck lists the distinguishing linguistic features for Cretan, but then notes that these linguistic features primarily describe the dialect spoken in Gortyn, Knossos, Lyttus, Vaxus, and other areas of central Crete; the dialects of the eastern and western parts of the island are different. Buck discusses several differences, but ultimately rejects the idea that eastern and western Cretan represented different dialects. Thus, we can see that the starting point for the discussion was geography—the island of Crete constituted a single dialect, to be subdivided only if the linguistic data were strong enough. Bartoněk (1972, 91-92) takes a similar approach in that he begins with the assumption that Crete is a linguistic unity, but, in the course of his linguistic discussion, comes to the conclusion that there are enough differences between the different regions to treat them as separate dialects in his later analysis. Bile (1988, 10-12) notes that the conception of Cretan being divided into three parts geographically is a recent development, but that dividing the island based on physical geography and expecting the dialect geography to follow is misguided, and does not fit the linguistic situation on Crete.

The point of this discussion is not to resolve the dialect situation on Crete, but merely to point out the ways in which political divisions and physical geography have sometimes been taken as a shortcut for dialectal divisions, and that there is room to consider the possibility that the traditional divisions of the dialects may not be wholly accurate from a linguistic point of view. Therefore, we might expect a more comprehensive examination of the linguistic data to provide reasonable grounds for splitting some dialects which may have been seen as geographic, political, or cultural unities, and for combining other dialects from areas which were geographically or politically distinct, but do not have sufficient linguistic differences. For example, I split Argolic into two different taxa, East and West Argolic, because they differed in the outcome of the third compensatory lengthening, and I split East and West Thessalian because they differed in a number of regards. It is in this spirit of investigation that I would like to locate the following cluster analysis of the West Greek dialects.

Cluster Analysis and Multidimensional Scaling

The analytical tool we can use to test whether certain dialects should be grouped together as single dialects is cluster analysis. Cluster analysis is a general term for a large variety of methods, all of which attempt to group a set of entities based on some measure of overall similarity, so that the entities in a given cluster are more similar to each other than to the other entities. Cluster analysis is an extensive and diverse field; for an overview, see Osei-Bryson and Samoilenko (2014). The type of cluster analysis I have chosen to use here is Multidimensional Scaling. Multidimensional scaling uses the distances between a set of entities to create an *n*-dimensional map showing the relative locations of the entities. In essence, Multidimensional Scaling attempts to solve the inverse problem of taking a map and being asked to measure the distances between a set of cities; Multidimensional Scaling takes the set of distances and attempts to recreate the map (Kruskal and Wish 1978, 7-8). Thus, a Multidimensional Scaling analysis of the West Greek dialects will produce a 2D image graphically representing the level of similarity or differences between the dialects in question. The investigator can then decide which level of similarity warrants lumping together two or more dialects as a single taxa. Kruskal and Wish do note that MDS produces results which agree closely with hierarchical cluster analysis (1978, 45-46). By this, I mean that entities which were grouped closely together in the MDS analysis were also grouped together by the hierarchical clustering analysis.

One common use of MDS is to analyze data where subjects were asked to sort stimuli into mutually exclusive and exhaustive categories. A distance matrix can then be generated from the number of times each pair of stimuli was assigned to the same category (Kruskal and Wish 1978, 10). The use of MDS to analyze a phylogenetic data matrix would fall exactly under this type of useage, with phylogenetic characters representing the categories.

Because Mutlidimensional Scaling produces a plot which simply shows overall similarities as distances, it does not commit us to any particular solution. If all dialects in the input should be considered separate dialects, they should be spaced roughly equal distances from one another. However, if some dialects should be considered parts of the same dialect, they should appear much closer together. A dialect continuum would be represented by a line of dialects spaced close together.

Methods

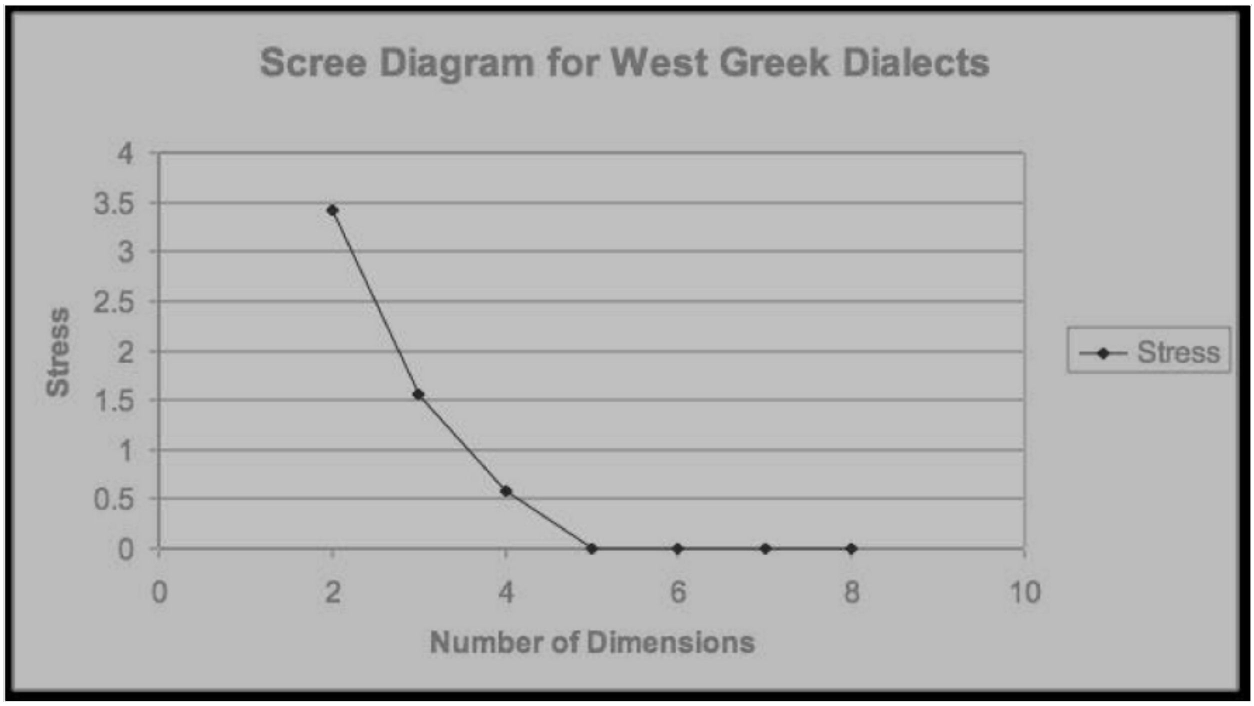
The input consisted of the West Greek portion of the phylogenetic data matrix. It was not necessary to provide a distance matrix because the MDS analysis computes one for itself. The symbol

“?” (missing data) in the phylogenetic analysis was treated as NA (missing data) in the MDS analysis. Megarian was removed from the matrix because it was essentially identical to Corinthian, which proved problematic for the analysis. I implemented the MDS analysis using the isoMDS() command of the MASS package in R.

I chose to use nonparametric MDS because there is less risk of inappropriate assumptions about the relationship between proximities and distances affecting the stress values (Kruskal and Wish 1978, 76).

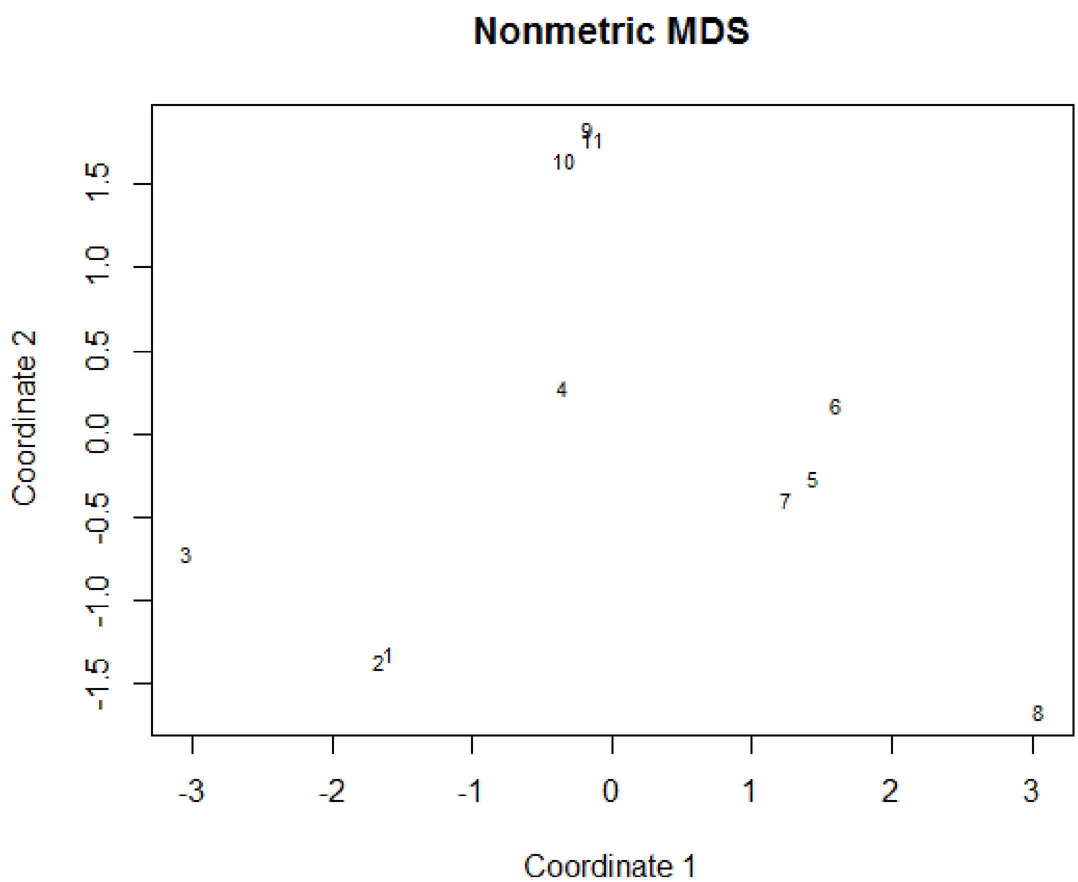
For MDS, how well or poorly the results fit the data is measured through a value called “stress.” Stress is essentially a measure of the badness of fit; the higher the number, the more poorly the analysis was able to fit the data. The stress values also allow us to determine the correct number of dimensions to use for the analysis. One way to test for the right number of dimensions is to create a scree diagram, which plots the number of dimensions against the resulting amount of stress (Holland 2008, 4). The point at which there stops being a significant improvement in the amount of stress is probably the correct number of dimensions. In this case, the correct number of dimensions is five.

Figure 25: Scree Diagram for the West Greek Dialects



The results of the nonparametric MDS analysis of the Greek dialects are as follows:

Figure 26: Non-Parametric MDS Analysis of the West Greek Dialects



The key to the diagram is as follows:

Table 11: Key to the MDS Analysis of the West Greek Dialects

1 Phocian

2 Locrian

3 Elean

4 Corinthian

5 East Argolic

6 West Argolic

7 Laconian

8 Cretan

9 Theran

10 Coan

11 Rhodian

The results strongly imply that the taxa have not been defined correctly. Instead, several sets of taxa should be grouped together as single taxa. These include Phocian and Locrian; East Argolic, West Argolic, and Laconian; and Theran, Coan, and Rhodian.

Redefining Taxa

The groups of taxa identified above were combined to produce single taxa. In general, when the dialects in a given group had different character states for a given phylogenetic character, I either selected the ancestral variant, or, if it wasn't clear what that was, I selected the majority variant. A more detailed discussion of the results of conflating the taxa is given below.

Phocian-Locrian

Character 55, Formation of Future Tense

Phocian has state 1, -σε-, while Locrian has state ?. State 1 has been restored for Phocian-Locrian since all other West Greek dialects have state 1.

Character 57, Third Plural Active Imperfect and Aorist Ending

Phocian has state 0, -ν, while Locrian has state 1, -αν. The other West Greek dialects have state 0, -ν, so state 0 has been restored for Phocian-Locrian.

Argolic-Laconian

<u>Character 24, Loss of Secondary Intervolcalic Σ</u>

Laconian is the only dialect which shows loss (state 1), so no loss (state 0) has been restored for Argolic-Laconian.

<u>Character 34, Second Compensatory Lengthening</u>

East Argolic and Laconian show state 2, compensatory lengthening, while West Argolic shows state 1, second compensatory lengthening did not occur. Since state 1 is the ancestral state, it has been restored for Argolic-Laconian.

<u>Character 35, Third Compensatory Lengthening</u>

West Argolic is the only dialect for which the character state of this character is known (state 2, third compensatory lengthening has occurred). This state has been restored for Argolic-Laconian.

<u>Character 36, Merger of the New Long Vowels</u>

East Argolic merged the new long vowels with the diphthongs in ει and ου; West Argolic merged the new vowels with the old vowels, while products of isovocalic contraction merged with diphthongs; and Laconian merged the new long vowels with the old long vowels. Given the disparate outcomes, it is not clear which state should be restored for Argolic-Laconian, so the character state has been coded as ?.

Character 40, Development of Z

ζ became δδ or δ in Laconian, but not most of the other West Greek dialects, including Argolic. Therefore, this character has been coded as 0, change has not occurred, for Argolic-Laconian.

Character 44, Dative Plural of Thematic Stems

East and West Argolic have both -οις and -οισι, while Laconian has only -οις. Given that most of the other West Greek dialects have only -οις, this character has been coded as 0, only -οις, for Argolic-Laconian.

Character 56, Aorists and Futures in Verbs With -Z-

Argolic has state 2, aorists in -ξ-, but -σ- when it is preceded by a guttural, while Laconian has state 1, aorists in -ξ-. Given that the rest of the West Greek dialects show state 1, this character has been coded as state 1 for Argolic-Laconian.

Character 71, The Preposition 'with' (μετά)

Argolic has state 1, πεδά, while Laconian has state 0, μετά. Given the variability within West Greek, this character has been coded as state ?, unknown, for Argolic-Laconian.

Character 72, the preposition 'towards' (1)

Argolic has state 1, preposition comes from inherited *proti, while Laconian and most of the rest of the West Greek dialects have state 0, preposition comes from inherited *poti. Thus, this character has been coded as state 0 for Argolic-Laconian.

Character 73, The Preposition 'towards' (3)

Argolic has state 0, which represents an additional form ποί before dentals, while Laconian has state 0, for no additional form. Given that no other dialect has this additional form, Argolic-Laconian has been given state 0.

Character 79, the noun 'Apollo'

Laconian has the form Ἀπόλλων, while the form for Argolic is unknown. Thus, the state for this character has been coded as 0, since it is the only form known for this group.

Theran-Coan-Rhodian

Character 38, Assimilation of ΡΣ to ΡΡ

Theran is the only dialect to show this change (state 1), so Theran-Coan-Rhodian has been restored to state 0 (change has not occurred).

Character 60, Formation of Athematic Infinitives

Theran and Coan both have -μεν, which is common among the West Greek dialects, while Rhodian has -μην, and is the only dialect aside from Cretan to do so. Thus, this character has been given as state 2, -μεν, for Theran-Coan-Rhodian.

Character 71, The Preposition ‘with’ (μετά)

Theran has state 1, πεδά, while Coan and Rhodian both have state 0, μετά. Given that the majority of the dialects have state 0, this character has been given as state 0 for Theran-Coan-Rhodian.

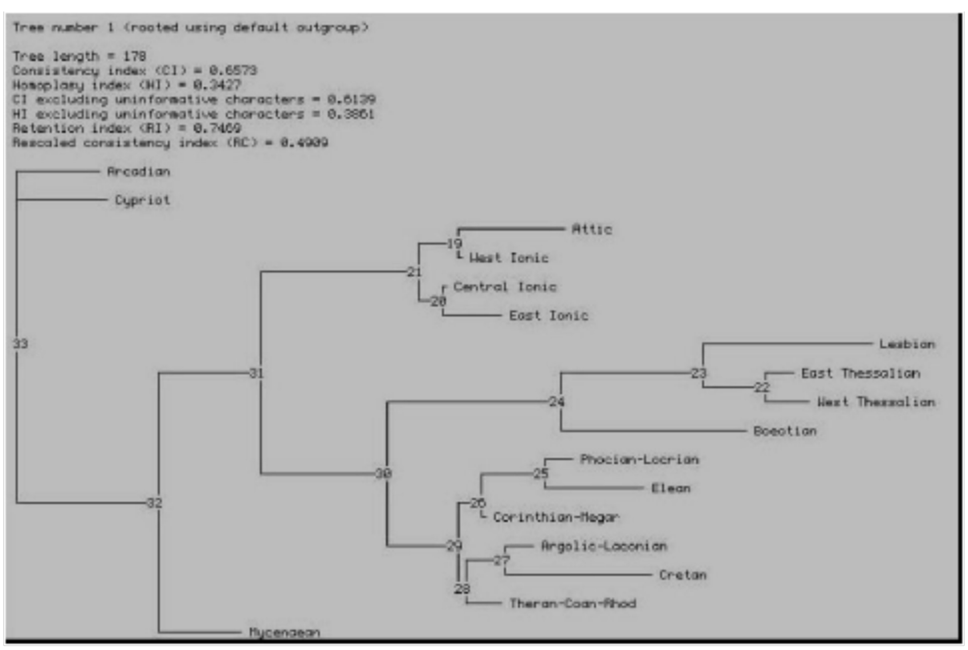
Character 78, The Noun 'Zeus'

Theran and Coan have state 1, a stem for the genitive and dative with Ζην-, while Rhodian has state 0, a stem for the genitive and dative with Διϝ-. Since the ancestral state is Διϝ-, which the majority of West Greek dialects have, this has been restored as state 0 for Theran-Coan-Rhodian.

Phylogenetic Analysis

The Maximum Parsimony analysis of the data matrix with West Greek taxa combined gave a single phylogenetic tree. The tree length was 178, the consistency index was 0.6573, the CI excluding uninformative characters was 0.6139, the retention index was 0.7469, and the rescaled consistency index was 0.4909.

Figure 27: Results of Maximum Parsimony Analysis With West Greek Taxa Combined



Thus, the clustering analysis was able to fix the problem of the nontreelike data in the West Greek dialects; the new phylogenetic data matrix gave a single completely resolved tree. It is interesting that the analysis suggests that the initial split in West Greek was between Corinthian and Megarian plus West Greek on the one hand, and Argolic-Laconian, Cretan, and Island Doric on the other hand. This would imply an initial split essentially between the area north and south of the Isthmus of Corinth, assuming Elean entered the Peloponnese from the north. The treatment of the West Greek dialects will be covered in more detail in the concluding chapter.

Chapter 6: Conclusions

Introduction

One purpose of these three sets of analyses, one comparing different types of character weighting, one testing NeighborNet, and one performing a cluster analysis prior to the phylogenetic analysis, was to determine which one of them, if any, offered the best way to resolve the lack of resolution produced by nontreelike data within the Greek dialects. Another purpose was to see if the subgrouping identified in the literature was supported by phylogenetic analysis. Each of the chapters offered preliminary results, which it seems worth recapping here. Then, we can proceed to assessing the performance of the different methods by comparing how they performed on the West Greek dialects, and then by drawing conclusions about how the results of the phylogenetic analysis support or refute the subgrouping of the Greek dialects established in the literature.

Findings From Previous Chapters

The first analysis chapter tested whether character weighting improves phylogenetic accuracy. The examples of character weighting included weighting phonological, morphological, and lexical characters, as well as reweighting characters based on CI. I then analyzed the resulting tree topologies and their statistics, as well as the percentages of each type of character that was homoplastic. I found that each type of character had roughly equal amounts of homoplastic characters. I also found that there was no significant benefit to weighting phonological, morphological, or lexical characters, and the only thing which offered increased resolution was reweighting characters based on the consistency index.

The second analysis chapter looked at NeighborNet. It found that while NeighborNet more or less accurately reconstructed the tree, the additional splits that NeighborNet picked up did not correspond to our analysis of elevated levels of features shared among different taxa from the phylogenetic analysis.

The third analysis chapter found that an MDS analysis of the West Greek dialects indicated that several West Greek taxa should be combined. This included Phocian and Locrian; East Argolic, West Argolic, and Laconian; and Theran, Coan, and Rhodian. When these taxa were combined and the phylogenetic analysis was run again, the result was a phylogenetic tree that was fully re-solved.

In short, each method could be said to have potentially resolved the problem. What is left is to compare them to see which, any, or all of them produce the best solution.

West Greek

We now explore the question of which of these models was the most successful, and by what criteria. If the criteria by which we judge the results is precision, i.e., whether or not we arrived at a single answer, then NeighborNet, reweighting characters based on CI, and preliminary cluster analysis were successful at producing what was a single tree (or network).

On the other hand, judging the results on accuracy, that is, whether the results match the absolute truth of the evolution of the dialects, then we have a problem—the results do not agree, and we have not yet reached a scholarly consensus on how the West Greek dialects developed. However, we can survey the major lines of thinking in the scholarship to date.

Bartoněk (1972) laid out a thorough view of the development of the West Greek dialects, partially based on traditional philological methods, and partially based on statistics. In his view, the development of the West Greek dialects is most significantly characterized by the development of the long vowel system after the first compensatory lengthening, which divided the Northwest Greek dialects and the dialects of the Saronic Gulf from the dialects of the Aegean islands and the remainder of the Peloponnese, with Elean having diverged earlier. Since discussions of the relationships among the West Greek dialects tend to hinge on the long vowel system, it might be useful to refer back to the discussion of the phylogenetic character which defines these differences, on pp. 38-39. There are also discussions of the three compensatory lengthenings on pp. 37-38.

According to Bartoněk, the change of *ē* to *ǣ* which distinguishes Elean happened very early, probably prior to 1000 BCE, and made Elean a separate West Greek dialect at that time (Bartoněk 1972, 110). Another early change was the development of the long vowel system, separating the dialects where the new long vowels arising from the first compensatory lengthening merged with the inherited long vowels and the long vowels arising from laryngeal loss, from the innovating dialects where these new long vowels did not merge with the old long vowels and, instead, became distinct phonemes. These innovating dialects, known as *mild Doric*, include Phocian, Locrian, Corinthian, Megarian, and East Argolic (Bartoněk 1972, 211-212); the non-innovating dialects, known as *severe Doric*, include Laconian, West Argolic, and Island Doric.

The second compensatory lengthening later served to differentiate Laconian, West Argolic, and Cretan from the remainder of Island Doric (Bartoněk 1972, 215), and then the third compensatory lengthening differentiated Island Doric and West Argolic. Meanwhile, certain morphological innovations separated Phocian and Locrian, as Northwest Greek, from the dialects of the Saronic Gulf: Corinthian, Megarian, and East Argolic (Bartoněk 1972, 219). Minor morphological innovations separate Corinthian and Megarian from East Argolic (Bartoněk 1972, 219-220).

However, Ruijgh (2007) showed that Bartoněk's interpretation cannot be correct because the long vowel system of severe Doric, which consisted of one set of long vowels, must have arisen from the long vowel system of mild Doric (which, in his view, roughly encompasses the dialects of the Peloponnese), or of *middle Doric*, which roughly encompasses the dialects of the Aegean islands, which both contained two sets of long vowels. For one example, we can explore Ruijgh's treatment of Laconian. Around 1600 BCE, the first compensatory lengthening produced a second set of long vowels in addition to the existing set of long vowels. After the Dorians arrived in Laconia, perhaps around the 11th century BCE, the new long vowels merged with the old long vowels, giving a vowel system with one set of long vowels. Around 1000, contractions of vowels in hiatus produced an additional new set of long vowels, which Laconian merged with the old long vowels, perhaps around 800 BCE (Ruijgh 2007, 435-436). These new findings call for a re-evaluation of the basis on which the West Greek dialects are grouped together.

I present the tree topologies of the three analyses below:

Figure 28: West Greek Dialects in Maximum Parsimony Analysis with Reweighting According to CI

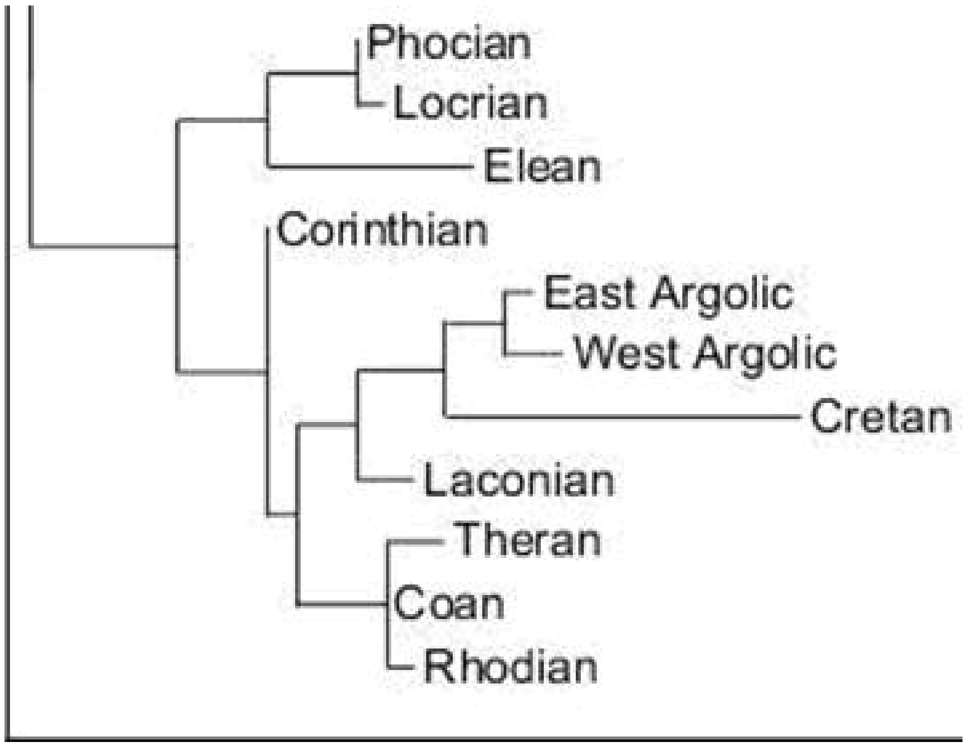


Figure 29: West Greek Dialects in NeighborNet Analysis

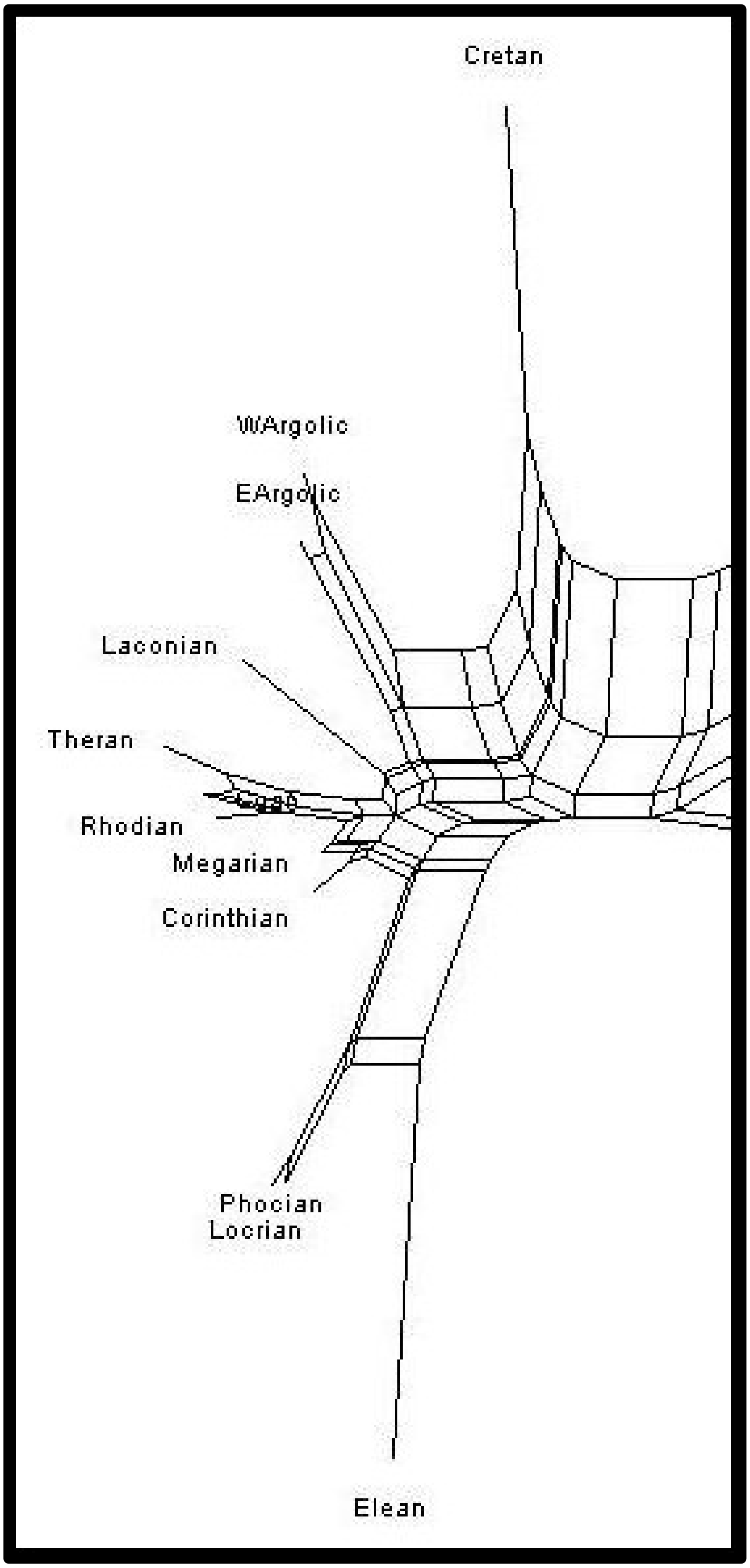
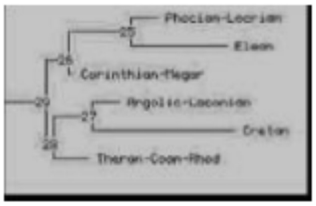


Figure 30: West Greek Dialects in Maximum Parsimony Analysis with Preliminary Cluster Analysis

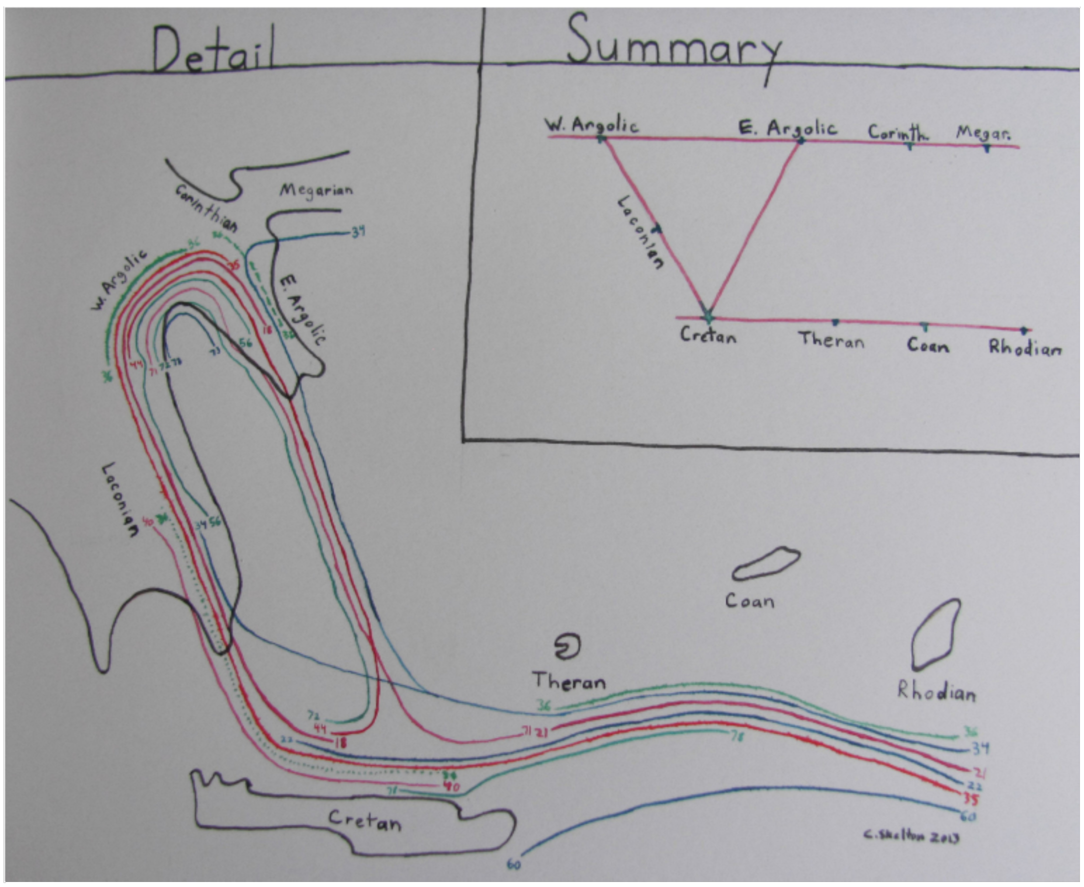


All of the analyses agree that Northwest Greek forms a clade, and that Theran, Coan, and Rhodian also form a clade. The problem is the arrangement of Corinthian/Megarian, East and West Argolic, Laconian, and Cretan, and then Island Doric and Northwest Greek. The analysis which reweighted characters according to CI posits an initial split between Northwest Greek and the other West Greek dialects. Of the remaining dialects, the first to diverge is Corinthian, then Island Doric, then Laconian, then Cretan, and finally Argolic. The cluster analysis, on the other hand, proposes a primary split between Corinthian and Northwest Greek on the one hand, and Argolic-Laconian, Cretan, and Island Doric on the other hand. Argolic-Laconian and Cretan are then sister taxa. The NeighborNet network is harder to read, but it appears that Northwest Greek represents its own clade, and then there is a series of clades progressively consisting of Cretan plus West Argolic, East Argolic, Laconian, and Island Doric. Megarian and Corinthian do not appear to form a clade within West Greek with any of the other West Greek dialects or dialect groups.

Clearly, these three scenarios are mutually contradictory, and cannot all be correct. Given the circumstances, it would be constructive to examine the data in more detail. Probably the best way to do this would be to take a map of the dialects in question and represent which dialects share features as lines connecting those dialects.

As we can see below, there is basically a dialect continuum which runs in a circular fashion around the Myrtoan Sea and Cretan Sea, with offshoots running to Corinthian and Megarian in the northeast, and Theran, Coan, and Rhodian in the southeast.

Figure 31: Dialect Map of West Greek, Excluding Northwest Greek



Given the circumstances, it is clear that the problem that all of the phylogenetic analyses encountered was that they could not represent data of this nature within their scope—neither a tree nor a network of bipartitions can represent a circular dialect network.

Conclusions about the Different Phylogenetic Approaches to Borrowing

There are two major conclusions and one question that come out of this analysis. First, when attempting to improve a tree-based analysis, better quality of data and better understanding of the data is just as important as a better method. While reweighting characters and performing a cluster analysis did produce a fully resolved tree, it was not necessarily the correct tree, and even this missed an interesting and potentially critical basic fact about the development of the West Greek dialects.

Second, when using a network analysis because your data is not expected to be fully treelike, it is just as important to understand what types of nontreelike evolution your network model can actually portray and whether these are the types of nontreelike evolution you expect to see, as it is to understand the basic assumptions of any type of analysis which produces a bifurcating tree. NeighborNet is commonly used for linguistic situations where the results are expected to be nontreelike, but very little thought goes into matching the type of network to the expected results. As both my analysis of borrowing within the Greek dialects and my analysis of the West Greek dialects showed, there are several situations involving borrowing which NeighborNet cannot represent.

Finally, we must confront the question of what method, if any, is appropriate in situations of borrowing such as these. Within phylogenetics, phylogenetic network methods which produce explicit networks and display borrowings as contact edges (branches connecting nodes which are not sisters) would be able to handle situations such as contact between Lesbian and several Attic-Ionic dialects, and possibly also the circular dialect continuum within West Greek. If we leave phylogenetic methods behind, various methods drawn from dialect geography may be more appropriate, though these methods will not work if the spatial relationships among the dialects are not well-understood.

Conclusions about Phylogenetic Systematics Versus Traditional Phylogenetic Methods

The results of the phylogenetic analysis closely match the results of traditional historical linguistic methods. All types of phylogenetic analysis identified the same four major dialect groups that had been previously identified by scholars, and they all agree that Attic-Ionic and Arcado-Cypriot, and West Greek and Aeolic, are more closely related to each other than the other possible arrangements. The phylogenetic analyses also had difficulties in the same area in which traditional historical linguistic analysis has also encountered difficulties: the arrangement of the West Greek dialects. It would certainly be a boon to phylogenetic analysis if it were able to definitively resolve this long-standing controversy, but in the end it the fact that they encountered problems in the same place only serves to confirm that phylogenetic methods closely match the sort of outcomes that can be obtained through traditional methods. If we wish to solve problems such as the development of the West Greek dialects, it seems as though traditional historical linguistics and phylogenetic systematics must both expand the scope of their methodology.

Bibliography

Aura Jorro, F. and F. Adrados. 1993. *Diccionario Griego-Español. Diccionario Micénico, Volume 2*. Madrid: Consejo Superior de Investigaciones Científicas.

Aura Jorro, F. and F. Adrados. 1999. *Diccionario Griego-Español. Diccionario Micénico, Volume 1*. Madrid: Consejo Superior de Investigaciones Científicas.

Bandelt, H. J., and A. W. Dress. 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Molecular Phylogenetics and Evolution 1(3)*: 242-252.

Barbançon, F., T. Warnow, S. N. Evans, D. A. Ringe, Jr, and L. Nakhleh. forthcoming. An experimental study comparing linguistic phylogenetic reconstruction methods. Proceedings of a conference on Language and Genes, University of California,
Santa Barbara, September 2006. http://www.cs.utexas.edu/users/tandy/nanterre-talk.ppt.

Bartoněk, A. 1972. *Classification of the West Greek dialects at the time about 350 B.C.* Amsterdam: Adolf M. Hakkert.

Bechtel, F. 1921–4. *Die griechischen Dialekte, 3 vols*. Weidmann: Berlin.

Beekes, R. 2010. *Etymological dictionary of Greek.* Leiden: Brill.

Bile, M. 1988. Le dialecte crétois ancien : étude de la langue des inscriptions, recueil des inscriptions postérieures aux IC. *Etudes crétoises 27*. Paris : Dépositaire, Libr. orientaliste P. Geuthner.

Bowern, C. 2012. The riddle of the Tasmanian languages. *Proceedings of the Royal Society 279*: 4590-4595.

Brixhe, C. 1976. *Le dialecte grec de Pamphylie*. Paris: A. Maisonneuve.

Bryand, D. 2006. Radiation and Network Breaking in Polynesian Linguistics. in P. Forster and C. Renfrew, eds., *Phylogenetic Methods and the Prehistory of Languages*. Cambridge: McDonald Institute Monographs, 111-118.

Bryant, D. and V. Moulton. 2004. NeighborNet: an agglomerative algorithm for the construction of phylogenetic networks. *Molecular Biology and Evolution 21(2)*: 255–265.

Buck, C. D. 1955. *The Greek Dialects, 2nd edn*. Chicago: The University of Chicago Press.

Cassio, A. C. 2007. Alcman's Text, Spoken Laconian, and Greek Study of Greek Dialects. in I. Hajnal and M. Meier-Brügger, eds., *Die altgriechischen Dialekte. Wesen und Werden*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.

Chadwick, J. 1983. Traditional spelling or two dialects? In Heubeck, A. and G. Neumann, eds. *Res Mycenaeae. Akten de VII. Internationalen Mykenologischen Colloquiums in Nürnberg vom 6.-10. April 1981*. Göttingen: Vandenhoeck and Ruprect. 78-88.

Clackson, J. 1994. *The Linguistic Relationship between Armenian and Greek*. Oxford: Blackwell.

Colvin, S. C. 2007. A *Historical Greek Reader: Mycenaean to the Koiné*. Oxford: Oxford University Press.

Dubois, L. 1988. *Recherches sur le dialecte arcadien, I-III*. Louvain-la-Neuve: Peeters.

Dunkel, G. E. 1990. J. Wackernagel und die idg. Partikeln **só*,**ke*,**kem* und **an*. In Eichner, H., and H. Rix, eds., *Sprachwissenschaft und Philologie. Jacob Wackernagel und die Indogermanistik heute*. Wiesbaden: L. Reichert, 100-130.

Egetmeyer, M. 2010. *Le dialecte grec ancien de Chypre. Tome I: Grammaire; Tome II: Répertoire des inscriptions en syllabaire chypro-grec*. Berlin: De Gruyter.

Ellsworth, M. 2011. The first palatalization of Greek. In Jamison, S. W. et al. (eds.), *Proceedings of the 22nd Annual UCLA Indo-European Conference*. Bremen: Hempen, 13-31.

Farris, J. S. 1989. The retention index and the rescaled consistency index. *Cladistics 5(4)*: 417-419.

Forbes, K. 1958. The relations of the particle ἄν with κε(ν) κα καν. *Glotta 37*: 179-182.

Forster, P. and C. Renfrew. 2006. *Phylogenetic Methods and the Prehistory of Languages*. Cambridge: McDonald Institute Monographs

Foulds, L. R., and R. L. Graham. 1982. The Steiner problem in phylogeny is NP-complete. *Advances in Applied Mathematics 3(1)* 43–9.

García-Ramón, J.-L. 1985. The Spellings "*Ta*" and "*Ta-ra*" for Inherited *"*Tr*" in Mycenaean. Sound Law, Phonetic Sequence and Morphological Factors at Work. *Minos 19(1,2)* 195-226.

Garrett, A. 2006. Convergence in the Formation of Indo-European Subgroups: Phylogeny and Chronology. in P. Forster and C. Renfrew, eds., *Phylogenetic Methods and the Prehistory of Languages*. Cambridge: McDonald Institute Monographs, 139–51.

Gray, R. D., and Q. D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origins. *Nature* 426.435–9.

Hajnal, I. 1997. *Sprachschichten des mykenischen Griechisch: Zur Frage der Differenzierung zwischen "*Mycénien spécial*" und "*Mycénien normal*."* Salamanca: Minos suppl. 14.

I. Hajnal and M. Meier-Brügger, eds., 2007. *Die altgriechischen Dialekte. Wesen und Werden*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.

Hatzopoulos, M. B. 2007. La position dialectale du macédonien à la lumière des découvertes épigraphiques récentes. In In Hajnal, I. and M. Meier-Brügger, eds., *Die altgriechischen Dialekte: Wesen und Werden: Akten des Kolloquiums Freie Universität Berlin 19-22 September 2001*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck, 157-176.

Haug, D. 2002. *Les phases de l'évolution de la langue épique*. Göttingen: Vandenhoeck & Ruprecht.

Heubeck, A. 1972. *Syllabic* r *in Mycenaean Greek*? In Ruipérez, M. S., ed., *Acta Mycenaea. Proceedings of the Fifth International Colloquium on Mycenaean Studies, Salamanca, 30 March-3 April 1970. II. Minos 12.* Salamanca: Universidad Salamanca, 55-79.

Holden, C. J. and Gray, R. D. 2006. Rapid radiation, borrowing, and dialect continua in the Bantu languages. In Forster, P. and C. Renfrew, eds., *Phylogenetic Methods and the Prehistory of Languages*. Cambridge, UK: McDonald Institute Monographs, 19-31.

Holland, S. M. 2008. *Non-Metric Multidimensional Scaling (MDS)*. unpublished ms., Department of Geology, University of Georgia, http://strata.uga.edu/software/pdf/mdsTutorial.pdf

Huson, D. H., and D. Bryant, Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution 23(2)*: 254-267, 2006. software available from www.split-stree.org

Kruskal, J. B., and M. Wish. 1978. *Multidimensional Scaling*. Thousand Oaks, CA: Sage.

Lejeune, M. 1972. *Phonétique historique du mycénien et du grec ancien*. Paris: Editions Klincksieck.

Lillo, A. 1991. Ionic κώς, ὅκως, ὅπως, Thessalian κις: A phonetic problem of analyzable compounds. *Glotta 69*(1-2), 1-13.

McMahon, A. and R. McMahon. 2006. Why Linguists Don't Do Dates: Evidence from Indo-European and Australian Languages. in P. Forster and C. Renfrew, eds., *Phylogenetic Methods and the Prehistory of Languages*. Cambridge: McDonald Institute Monographs, 153-160.

Meier-Brügger, M. 2003. *Indo-European Linguistics*. Berlin: De Gruyter.

Merkelbach, R. and West, M. L. 1967. *Fragmenta Hesiodeia.* Oxford: Clarendon Press.

Morpurgo Davies, A. 1968. The treatment of **r* and**l* in Mycenaean and Arcado-Cyprian. In consiglio nazionale delle ricerche, eds., *Atti e memorie del 1 Congresso Internazionale di Micenologia, vol. 2*. Roma: Edizioni dell'Ateneo, 791-812.

Morpurgo Davies, A. 1976. The -εσσι datives, Aeolic -*ss*-, and the Lesbian poets. In Morpurgo Davies, A. and W. Meid, eds., *Studies in Greek, Italic, and Indo-European Linguistics Offered to Leonard R. Palmer on the Occasion of His Seventieth Birthday.* Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft 16, 181-197.

Morpurgo Davies, A. 1987. The Greek Notion of Dialect. *Verbum* 10: 7–28. Repr.
T. Harrison, ed., *Greeks and Barbarians*. Edinburgh : Edinburgh University Press, 2002: 153–71.

Nakhleh, L., T. Warnow, C.R. Linder and K. St John, 2005a. Reconstructing reticulate evolution in species—theory and practice. *Journal of Computational Biology 12:* 796-811.

Nakhleh, L., T. Warnow, D. A. Ringe, Jr, and S. N. Evans. 2005b. A comparison of phylogenetic reconstruction methods on an IE dataset. *Transactions of the Philological Society 103*: 171–192.

Nicholls, G., and Gray, R. 2008. Dated ancestral trees from binary trait data and their application to the diversification of languages. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(3)*, 545-566.

Nichols, J. and T. Warnow. 2008. Tutorial on Computational Linguistic Phylogeny. *Language and Linguistics Compass* 2/5: 760–820.

Osei-Bryson, K. M., and S. Samoilenko. 2014. Overview on cluster analysis. In Osei-Bryson, K. M., and O. Ngwenyama, eds., *Advances in Research Methods for Information Systems Research*. New York: Springer, 127-138.

Parker, H. 2008. The Linguistic Case for the Aiolian Migration Reconsidered. *Hesperia* 77 (2008), 431-464.

Rau, J. 2009. Myc. *te-re-ja* and the athematic inflection of contract verbs. In Yoshida, K. and B. Vine, eds., *East and West: Papers in Indo-European Studies.* Bremen: Hempen, 181-188.

Rau, J. 2010. Greek and Proto-Indo-European. In Bakker, E. J., ed., *A Companion to the Ancient Greek Language*, Blackwell Companions to the Ancient World. Chichester, West Sussex: Wiley-Blackwell, 171-188.

Risch, E. 1955. Die Gliederung der griechischen Dialekte in neuer Sicht. *Museum Helveticum 12*: 61–75.

Risch, E. 1966. Les différences dialectales dans le mycénien. In Palmer, L.R. and J. Chadwick, eds., *Proceedings of the Cambridge Colloquium on Mycenaean Studies*. Cambridge University Press, Cambridge, 150-159.

Ruijgh, C. 2007. L'évolution des dialectes doriens jusqu'à la koina dorienne: le système des voyelles longues et la formation du futur. In Hajnal, I. and M. Meier-Brügger, eds., *Die altgriechischen Dialekte. Wesen und Werden*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck, 393-447.

Saitou, N. and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution, 4(4)*, 406-425.

Sankoff, G. 2001. Linguistic outcomes of language contact. in J. Chambers, P. Trudgill, and N. Schilling-Estes, eds. *The Handbook of Language Variation and Change*, Oxford: Blackwell, 638–668.

Skelton, C. 2008. Methods of Using Phylogenetic Systematics to Reconstruct the History of the Linear B Script. *Archaeometry 50*: 158-176.

Swofford, D. L. 1998 *PAUP*, Phylogenetic Analysis Using Parsimony (*and other methods), Version 4.0 beta 10*. Sunderland, MA: Sinauer Associates.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. In Hillis, D. M., C. Moritz and B. K. Mable, eds., *Molecular Systematics, 2nd edn*. Sunderland, MA: Sinauer Associates, 407–514.

Thompson, R. 1997. Dialects in Mycenaean and Mycenaean among the dialects. *Minos 31-32:* 313-333.

Thompson, R. 2002-2003. Special vs. Normal Mycenaean revisited. *Minos 37-38*: 337-369.

Thumb, A. 1909. *Handbuch der griechischen dialekte*. Heidelberg: C. Winter.

Trudgill, P. 2004. *New-dialect formation: The inevitability of colonial Englishes*. Oxford: Oxford University Press.

van Beek, L. 2013. *The Development of the Proto-Indo-European Syllabic Liquids in Greek*. Leiden: Leiden University. (dissertation)

Varias, C. 1997. A tentative analysis of dialectal differences in the Linear B texts from Mycenae. *Minos 29-30*: 135-157.

Vine, B. 1999. On "Cowgill's Law" in Greek. In Eichner, H. and H. C. Luschützky, eds., *Compositiones Indogermanicae in Memoriam Jochem Schindler*. Prague: Enigma, 555-600.

Weiss, M. 2010. *Outline of the Historical and Comparative Grammar of Greek*. unpublished ms., Department of Linguistics, Cornell University.

Willi, A. 2003. καί--mykenisch oder nachmykenisch? *Glotta(79)* 224-248.

Willi, A. 2008. Genitive problems: Mycenaean -*Ca-o*, -*Co-jo*, -*Co* vs. later Greek -ᾱο -οιο -ου. *Glotta 84*: 239-272.

Willi, A. 2012. Kiparsky's Rule, thematic nasal presents, and athematic *verba vocalia* in Greek. In Probert, P. and A. Willi, eds., *Laws and Rules in Indo-European.* Oxford: Clarendon, 260-276.

Woodard, R. D. 1986. Dialectal differences at Knossos. *Kadmos 25(1)*: 49-74.